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

Running title: Cobamide biosynthesis predictions

Amanda N. Shelton[1], Erica C. Seth[1], Kenny C. Mok[1], Andrew W. Han[2], Samantha N. Jackson[3,4], David R. Haft[3], Michiko E. Taga[1*]

[1] Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, CA, USA

[2] Second Genome, Inc., South San Francisco, CA, USA

[3] J. Craig Venter Institute, Rockville, MD, USA

[4] Current address: Department of Biological & Environmental Engineering, Cornell University, Ithaca, NY, USA

[*] Address correspondence to Michiko E. Taga, taga@berkeley.edu

--

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 varies at the phylum level, with 57% of Actinobacteria, 45% of Proteobacteria, and 30% of Firmicutes, and less than 1% of Bacteroidetes containing the complete biosynthetic pathway. Cobamide structure 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 that 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 crossfeeding 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 crossfeeding, in which microbes that lack the ability to synthesize particular required nutrients (termed auxotrophs) obtain these nutrients from other organisms in their community (Seth and Taga, 2014). By understanding which organisms require nutrients and which can produce them, we can predict specific metabolic interactions between members of a microbial community (Abreu and Taga, 2016). 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 has long been appreciated. Only recently, however, has cobamide-dependent metabolism been recognized as a potential mediator of microbial interactions (Degnan *et al.*, 2014b; Seth and Taga, 2014). 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 (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 (Ragsdale and Pierce, 2008). Cobamides are also used for processes that are important for human use, such as reductive dehalogenation and natural product synthesis (Banerjee and Ragsdale, 2003; Broderick *et al.*, 2014).

De novo cobamide biosynthesis involves approximately 30 steps (Warren *et al.*, 2002), 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 crossfeeding 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 (Yan *et al.*, 2012; Mok and Taga, 2013; Degnan *et al.*, 2014a; Helliwell *et al.*, 2016; Keller *et al.*, 2018). Recent work has identified many of the genetic determinants for the biosynthesis of the benzimidazole class of lower ligands (Campbell *et al.*, 2006; Taga *et al.*, 2007; Gray and Escalante-Semerena, 2007; Hazra *et al.*, 2015; Mehta *et al.*, 2015) and attachment of phenolic lower ligands (Chan and Escalante-Semerena, 2011; Newmister *et al.*, 2012) (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 (Rodionov *et al.*, 2003; Zhang *et al.*, 2009), suggesting that cobamide crossfeeding may be widespread in microbial communities. Analyses of cobamide biosynthesis in the human gut (Degnan *et al.*, 2014a; Magnúsdóttir *et al.*, 2015) and in the phylum Cyanobacteria (Helliwell *et al.*, 2016) 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.

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) (Markowitz *et al.*, 2012) classified as "finished" (accessed January 11, 2017) or "permanent draft" (accessed February 23, 2017) were downloaded (Supplementary Table 1, Sheet 1). To assess genome completeness, we searched for 55 single copy gene annotations (Raes *et al.*, 2007; Brown *et al.*, 2015) 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 2,776 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 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.sbcs.qmul.ac.uk/iubmb/enzyme/), Pfam, TIGRFAM, Clusters of Orthologous Groups (COG), and IMG Terms (Cornish-Bowden, 2014; Finn *et al.*, 2016; Haft *et al.*, 2012; Galperin *et al.*, 2015; Markowitz *et al.*, 2012) for cobamide biosynthesis, cobamide-dependent enzymes, and cobamide-independent alternative functions were chosen. These included annotations used by Degnan *et al*. (2014a), 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, 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) (Supplementary Table 1, 2 sheet 2). The gene hits were downloaded as a list of gene unique identifiers, gene locus ID, function hit, and genome name (data available upon request).

For genes without functional annotations in the IMGer database, we chose sequences that were genetically or biochemically characterized to use as the query genes in one-way BLASTP (Altschul *et al.*, 1997) against the filtered genomes using the IMGer "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: very likely cobamide producer, likely cobamide producer, possible cobamide producer, tetrapyrrole precursor salvager, cobinamide (Cbi) salvager, likely non-producer, and very likely non-producer (Supplementary Table 7) and classified genomes accordingly (Supplementary Table 5).

To distinguish putative phenolic lower ligand attachment *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) (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 (Altschul *et al.*, 1997; Waterhouse *et al.*, 2009). 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 (Haft and Haft, 2017). 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 Files. Details for each protein are listed in the Supplementary Materials and Methods.

These protein sequences for 10591 of the filtered genomes were queried for each *bza* HMM using hmm3search (HMMER3.1). 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 (Hazra *et al.*, 2015).

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). Since the *C. saccharobutylicum* HemD is a fusion protein with both the UroIII synthase and UroIII methyltransferases domains, 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. 180 species remained

178 after this analysis (Supplementary Table 10). Genomes were classified as a particular type of
179 tetrapyrrole precursor salvager only if they were missing all genes upstream of a precursor.
180
181 **Strains and growth conditions**
182 *Clostridium scindens* ATCC 35704, *Clostridium sporogenes* ATCC 15579, and
183 *Treponema primitia* ZAS-2 (gift from Jared Leadbetter) were grown anaerobically with and
184 without added 5-aminolevulinic acid (1 mM for *C. sporogenes* and *T. primitia*, 0.5 mM for *C.*
185 *scindens*).
186 *Desulfotomaculum reducens* MI-1 (gift from Rizlan Bernier-Latmani), *Listeria*
187 *monocytogenes* (gift from Daniel Portnoy)*, Blautia hydrogenotrophica* DSM 10507, *Clostridium*
188 *kluyveri* DSM 555 (gift from Rolf Thauer), and *Clostridium phytofermentans* ISDg (gift from
189 Susan Leschine) were grown anaerobically. Details of the growth conditions are listed in the
190 Supplementary Materials and Methods.
191
192 **Corrinoid extraction and analysis**
193 Corrinoid extractions were performed as described (Yi *et al.*, 2012). For corrinoids
194 extracted from 1 L cultures of *C. sporogenes, C. scindens*, and *T. primitia*, high performance
195 liquid chromatrography (HPLC) analysis was performed with an Agilent Series 1200 system
196 (Agilent Technologies, Santa Clara, CA) equipped with a diode array detector with detection
197 wavelengths set at 362 and 525 nm. 50 to 100 µl samples were injected onto an Agilent Eclipse
198 XDB C18 column (5 µm, 4.6 x 150 mm) at 35 °C, with 0.5 ml min^{-1} flow rate. Samples were
199 separated using acidified water and methanol (0.1% formic acid) with a linear gradient of 18% to
200 30% 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 methanol for 2 min, 2%-10% acidified methanol in 0.1 min, and 10-40% acidified methanol over 9 min.

For *B. hydrogenotrophica,* corrinoids were analyzed as above with the following changes. 10 µl samples were injected onto an Agilent Zorbax SB-Aq column (5 µm, 4.6 x 150) with 1 ml/min flow rate at 30ºC. The samples were separated with a gradient of 25-34% acidified methanol over 11 minutes, followed by 34-50% over 2 min and 50-75% 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. Because most of these genes have been characterized and annotated, we used annotations from existing databases including Enzyme Commission (EC) numbers, Pfam, TIGRFAM, Clusters of Orthologous Groups (COG), and IMG Terms to identify most of these functions and used homology-based methods to identify those for which annotations were unavailable. After filtering 44,802 publicly available bacterial genomes to one genome per species and removing incomplete genomes, we had a working data set of 11,436 species (for details, see Materials and Methods).

Our results indicate that the capability to use cobamides is widespread in bacteria. 86% 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

224 family (Fig. 3A). This is consistent with previous analyses of smaller data sets (Rodionov *et al.*,
225 2003; Zhang *et al.*, 2009; Degnan *et al.*, 2014a). The four major phyla in the data set have
226 different distributions of the number of cobamide-dependent enzyme families per genome, with
227 the Proteobacteria and Bacteroidetes having higher mean numbers of enzyme families than the
228 Firmicutes and Actinobacteria (Fig. 3A). The most abundant cobamide-dependent enzymes are
229 involved in core metabolic processes such as methionine synthesis and nucleotide metabolism,
230 whereas processes such as reductive dehalogenation and mercury methylation are less abundant
231 (Fig. 3B, Supplementary Table 5). We also observe phylum-level differences in the relative
232 abundance of cobamide-dependent enzyme families (Fig. 3B), most notably the nearly complete
233 absence of epoxyqueuosine reductase in Actinobacteria. Nonetheless, the cobamide-dependent
234 methionine synthase (MetH) and, to a lesser extent, methylmalonyl-CoA mutase (MCM) and the
235 cobamide-dependent ribonucleotide reductase (RNR), are the most abundant cobamide-
236 dependent enzyme families in all of the four phyla (Fig. 3B).
237 For some cobamide-dependent processes, cobamide-independent alternative enzymes or
238 pathways also exist (Fig. 1, right side of arrows). For example, we find that the occurrence of
239 MetH is more common than the cobamide-independent methionine synthase, MetE, but that most
240 bacteria have both enzymes (Fig. 3C). In contrast, cobamide-independent RNRs are found more
241 commonly than the cobamide-dependent versions, and 30% of genomes have both cobamide-
242 dependent and -independent RNRs (Fig. 3C). The cobamide-dependent propionate (which uses
243 MCM), ethanolamine, and glycerol/propanediol metabolisms appear more abundant than the
244 cobamide-independent alternatives (Fig. 3C). However, the abundance of the cobamide-
245 dependent propionate function is overestimated because the MCM category includes mutases for
246 which cobamide-independent versions have not been found. The abundance of the latter two

247 cobamide-independent functions may be underestimated, as they were identified based on
248 similarity to a limited number of sequences. We did not observe dramatic phylum-level
249 differences in the relative abundances of cobamide-dependent and –independent processes
250 (Supplementary Figure 2).
251
252 **37% of bacterial species are predicted to produce cobamides *de novo***
253 We analyzed the filtered data set to make informed predictions of cobamide biosynthesis
254 to determine the extent of cobamide biosynthesis in bacteria and to identify marker genes
255 predictive of cobamide biosynthesis. A search for genomes containing the complete pathways for
256 anaerobic or aerobic cobamide biosynthesis, as defined in the model bacteria *Salmonella*
257 *enterica* serovar Typhimurium and *Pseudomonas denitrificans*, respectively (Warren *et al.*,
258 2002), revealed that few genomes contain all annotations for the complete pathway, but many
259 contain nearly all. Some bacteria that appear to have an incomplete pathway might nonetheless
260 be capable of cobamide biosynthesis because of poor annotation, non-homologous replacement
261 of certain genes (McGoldrick *et al.*, 2005; Gray and Escalante-Semerena, 2010), or functional
262 overlap of some of the enzymes. We therefore relied on experimental data on cobamide
263 biosynthesis in diverse bacteria to inform our predictions, using 63 bacteria that are known to
264 produce cobamides (Table 1, Supplementary Table 6), including 5 tested in this study (Table 1,
265 bold names, Supplementary Figure 3). We identified a core set of eight functions shared by all or
266 all except one of the genomes of cobamide-producing bacteria (Table 1, gray highlight). These
267 core functions include three required for corrin ring biosynthesis: *cbiL*, *cbiF* and *cbiC* in the
268 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 functions 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 (Schubert *et al.*, 1999).

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*) (Haas *et al.*, 2009) were excluded from these threshold-based definitions. We did not exclusively use the small set of core functions 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). 49% 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

292 also varies based on predicted cobamide biosynthesis, with predicted cobamide producers having
293 more cobamide-dependent enzyme families per genome than non-producers (Supplementary
294 Figure 4).

295 To assess whether the core functions could be used as markers, the threshold-based
296 assignments of cobamide biosynthesis were compared to the frequency of the three core corrin
297 ring functions. The presence of each core function alone is largely consistent with the threshold-
298 based assignments, as each is present in 99% of genomes in the producer categories and in less
299 than 1% of the non-producers (Table 2). The presence of two or all three marker functions
300 matches the threshold-based predictions even more closely (Table 2). The corrin ring markers
301 chosen in Table 1 are slightly more predictive of our threshold-based cobamide biosynthesis
302 classifications than *cbiA/cobB* (EC:6.3.5.11/EC:6.3.5.9), a previously selected marker used in
303 environmental DNA analysis (Bertrand *et al.*, 2011); although *cbiA/cobB* was found in 99% of
304 predicted cobamide producers, is it also present in 2.6% of predicted non-producers and 46% of
305 precursor salvagers (Supplementary Table 5).

306 As with the cobamide-dependent enzyme families, the four major phyla in the data set
307 have major differences in their predicted cobamide biosynthesis phenotypes (Fig. 4). Around half
308 of Actinobacteria (57%) and Proteobacteria (45%) and 30% of Firmicutes are predicted to be
309 cobamide producers. In contrast, only 0.6% of Bacteroidetes are predicted to produce cobamides
310 *de novo*, yet 96% have at least one cobamide-dependent enzyme, suggesting that most members
311 of this phylum must acquire cobamides from other organisms in their environment. In addition,
312 Bacteroidetes have the highest relative proportion of species predicted to salvage Cbi via a
313 partial cobamide biosynthesis pathway, and most of the tetrapyrrole precursor salvagers are
314 Firmicutes (see later section; Supplementary Table 10), whereas very few Actinobacterial

315 species are predicted to salvage precursors (Fig. 4). These divisions reveal potential cobamide
316 and cobamide precursor crossfeeding requirements across phyla.
317
318 **Predicting cobamide structure**
319 Lower ligand structure is determined by the intracellular production of lower ligand bases
320 as well as specific features of the lower ligand attachment genes *cobT* or *arsAB* (Campbell *et al.*,
321 2006; Taga *et al.*, 2007; Gray and Escalante-Semerena, 2007; Hazra *et al.*, 2013; Crofts *et al.*,
322 2014a; Hazra *et al.*, 2015; Yan *et al.*, 2018; Chan and Escalante-Semerena, 2011). We first
323 defined predictions for the biosynthesis of cobamides containing benzimidazole lower ligands
324 (benzimidazolyl cobamides), based on the presence of genes for the biosynthesis of
325 benzimidazoles. We used the presence of *bluB*, the aerobic synthase for the lower ligand of
326 cobalamin, 5,6-dimethylbenzimidazole (DMB), as a marker for cobalamin production (Campbell
327 *et al.*, 2006; Taga *et al.*, 2007; Hazra *et al.*, 2018) and found it in 25% of genomes in the data set,
328 including those without complete cobamide biosynthesis pathways. *bluB* is most abundant in
329 predicted cobamide-producing bacteria (Fig. 5A), particularly in Proteobacteria (Fig. 5B).
330 Anaerobic biosynthesis of DMB and three other benzimidazoles requires different
331 combinations of the *bza* genes as shown in Figures 2A and 5C (Hazra *et al.*, 2015; Mehta *et al.*,
332 2015). Because annotations for the majority of the *bza* genes were not available, we developed
333 profile HMMs to search for them (see Supplementary Materials and Methods, Supplementary
334 Files). 96 genomes contain one or more *bza* genes, and 88 of these contain either *bzaF* or both
335 *bzaA* and *bzaB*, the first step necessary for the anaerobic biosynthesis of all four benzimidazoles
336 (Fig. 5C, Supplementary Table 8). As seen with *bluB*, anaerobic benzimidazole biosynthesis
337 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 (Chan and Escalante-Semerena, 2011). 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 (Stupperich and Eisinger, 1989; Men *et al.*, 2014b) (Fig. 5A, B, Supplementary Table 9).

42% 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 α-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, α-ribazole salvaging); we found CblS in 363 species (3.2%), mostly in the Firmicutes phylum (Fig. 5 A, B, inner rings) (Gray and Escalante-Semerena, 2010; Mattes and Escalante-Semerena, 2017). A higher proportion of bacteria, 1,041 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 (Woodson and Escalante-Semerena, 2004) (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 (Gray and Escalante-Semerena, 2009a, 2009b; Men *et al.*, 2014a; Yi *et al.*, 2012). The reliance of some

bacteria on exogenous lower ligands or α-ribazoles produced by other organisms precludes prediction of cobamide structure in all cases.

17% 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) auxotrophs (15% of genomes), contain the nucleotide loop assembly steps but lack all or most of the corrin ring biosynthesis functions. As demonstrated in *Escherichia coli* (Di Girolamo and Bradbeer, 1976), *Thermotoga lettingae* (Butzin *et al.*, 2013), and *Dehalococcoides mccartyi* (Yi *et al.*, 2012), and predicted in human gut microbes (Degnan *et al.*, 2014a), Cbi auxotrophs can take up the late intermediate Cbi, assemble the nucleotide loop and attach a lower ligand in a process called Cbi salvaging.

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 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

383 cobamide or that they synthesize the precursors through a novel pathway. We therefore tested
384 three putative tetrapyrrole precursor salvagers for their ability to produce corrinoids (cobamides
385 and other corrin ring-containing compounds) in the presence and absence of a tetrapyrrole
386 precursor. *Clostridium scindens* and *Clostridium sporogenes*, which are predicted to require 5-
387 aminolevulinic acid (ALA), produced corrinoids in defined media only when ALA was supplied,
388 suggesting that they do not have a novel ALA biosynthesis pathway (Fig. 6B). We tested an
389 additional predicted ALA salvager, the termite gut bacterium *Treponema primitia* ZAS-2, for
390 which a defined medium has not been developed. When cultured in medium containing yeast
391 autolysate, *T. primitia* produced trace amounts of corrinoids, and corrinoid production was
392 increased by supplementing this medium with ALA (Fig. 6B). The ability of *T. primitia* to use
393 externally supplied ALA was further shown by its increased growth rate and cell density at
394 stationary phase when either cobalamin or ALA was added (Fig. 6C). Together, these results
395 support the hypothesis that predicted ALA salvagers synthesize cobamides by taking up ALA
396 from the environment.
397
398 **Discussion**
399 Vitamin B_{12} and other cobamides have long been appreciated as a required nutrient for
400 humans, bacteria, and other organisms due to their critical function as enzyme cofactors. Prior to
401 this work, the extent of cobamide biosynthesis and dependence across different bacterial taxa had
402 not been investigated. The availability of tens of thousands of genome sequences afforded us the
403 opportunity to conduct a comprehensive analysis of cobamide metabolism across over 11,000
404 bacterial genomes. This analysis gives an overview of cobamide dependence and cobamide
405 biosynthesis across bacteria, allowing for the generation of hypotheses for cobamide and

406 cobamide precursor crossfeeding in bacterial communities. Our work shows that cobamide use is
407 much more widespread than cobamide biosynthesis, consistent with the majority of previous
408 studies of smaller data sets (Rodionov *et al.*, 2003; Zhang *et al.*, 2009; Degnan *et al.*, 2014a).
409 The prevalence of cobamide-dependent enzymes in bacteria, coupled with the relative paucity of
410 *de novo* cobamide producers, underscores the ubiquity of both cobamide-dependent metabolism
411 and cobamide crossfeeding in microbial communities. Here, we additionally find that cobamide
412 production and use are unevenly distributed across the major phyla represented in the data set,
413 identify bacteria dependent on cobamide precursors, and predict cobamide structure. These
414 results underscore the widespread nutritional dependence of bacteria.

415 The most abundant types of cobamide-dependent enzymes in our data set are methionine
416 synthase, epoxyqueuosine reductase, RNR, and MCM. For all of these enzymes, cobamide-
417 independent alternative enzymes or pathways exist. (Note that the newly discovered alternative
418 to epoxyqueuosine reductase, QueH (Zallot *et al.*, 2017), was not included in our analysis.) The
419 prevalence of cobamide-dependent enzymes for which cobamide-independent counterparts exist,
420 particularly in the same genome, suggests that cobamide-dependent enzymes confer distinct
421 advantages. This view is supported by the observations that MetE is sensitive to stress and has a
422 100-fold lower turnover number than MetH (Hondorp and Matthews, 2004; Xie *et al.*, 2013;
423 Gonzalez *et al.*, 1992) and that cobamide-independent RNRs are limited in the oxygen
424 concentrations in which they are active (Taga and Walker, 2010; Fontecave, 1998)

425 In our analysis of cobamide biosynthesis, it was not possible to use a single definition of
426 the complete *de novo* cobamide biosynthesis pathway across all bacterial genomes because of
427 divergence in sequence and function. The use of experimental data gives confidence to our
428 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 (Roth *et al.*, 1996), and cobamide biosynthesis operons are commonly subjected to negative regulation by riboswitches (Nahvi, 2004; Rodionov *et al.*, 2003). The abundance of cobamide importers (Rodionov *et al.*, 2003; Zhang *et al.*, 2009; Degnan *et al.*, 2014a), 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 functions 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 (Croft *et al.*, 2005; Kazamia *et al.*, 2012), 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 crossfeeding relationships among different taxa.

The reliance of many bacteria on cobamide crossfeeding, coupled with the fact that structurally different cobamides are not functionally equivalent in bacteria (Yan *et al.*, 2012; Mok and Taga, 2013; Degnan *et al.*, 2014a; Helliwell *et al.*, 2016; Keller *et al.*, 2018), underscores the importance of cobamide structure in microbial interactions. The structural differences in cobamides are almost exclusively limited to variations in the lower ligand (Fig.

2B). Additional variation in the nucleotide loop was not considered here because of the absence of genes specific to norcobamide biosynthesis (Kräutler *et al.*, 2003; Keller *et al.*, 2016). We were able to predict lower ligand structure for 58% of predicted cobamide producers. The remaining bacteria may produce purinyl cobamides, which are abundant in some bacterial taxa and microbial communities (Helliwell *et al.*, 2016; Allen and Stabler, 2008). Further analysis of substrate specificity in CobT and other lower ligand attachment enzymes could lead to improved strategies for predicting purinyl cobamide production, as some CobT homologs appear to segregate into different clades based on lower ligand structure (Hazra *et al.*, 2013; Crofts *et al.*, 2013; Yan *et al.*, 2018). The presence of free benzimidazoles and α-ribazoles in microbial communities (Crofts *et al.*, 2014b; Johnson *et al.*, 2016; Wienhausen *et al.*, 2017) and the ability of bacteria to take up and incorporate these compounds into cobamides (Anderson *et al.*, 2008; Mok and Taga, 2013; Keller *et al.*, 2013; Crofts *et al.*, 2013) 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 cobamide-dependent bacteria 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

(Men *et al.*, 2014b), compared to cobalamin in bovine rumen (Girard *et al.*, 2009) and 2-methyladeninyl cobamide in human stool (Allen and Stabler, 2008). 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 (Degnan *et al.*, 2014a).

Dependence on biosynthetic precursors has been observed or predicted for amino acids, nucleotides, and the cofactors thiamin and folate (Sloan and Moran, 2012; Kilstrup *et al.*, 2005; Paerl *et al.*, 2016; de Crécy-Lagard *et al.*, 2007). Here, we describe genomic evidence for dependence on cobamide precursors, namely Cbi or tetrapyrrole precursors. The prevalence of Cbi-salvaging bacteria 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 (Men *et al.*, 2014b), and represented up to 12.8% of the total corrinoids detected in human stool samples (Allen and Stabler, 2008).

Our analysis defined five types of tetrapyrrole precursor salvagers and experimentally verified the ALA salvager phenotype in three species. Bacteria that lack tetrapyrrole precursor biosynthesis genes but contain the remainder of the cobamide biosynthetic pathway were overlooked in previous genomic studies of cobamide biosynthesis that considered only the corrin ring biosynthesis and nucleotide loop assembly the portions of the pathway (Rodionov *et al.*, 2003; Zhang *et al.*, 2009; Magnúsdóttir *et al.*, 2015; Helliwell *et al.*, 2016). Tetrapyrrole precursors have been detected in biological samples, suggesting that they are available for uptake in some environments. For example, UroIII was detected in human stool (Dobriner, 1937; Watson *et al.*, 1945) and ALA has been found in swine manure extract (Kanto *et al.*, 2013).

Although we confirmed experimentally the ALA dependence phenotype, we were unable to detect ALA in several biological samples using a standard chemical assay or bioassay, 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 (Sangwan and O'Brian, 1991; Lyell *et al.*, 2017) 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 (Qi *et al.*, 2008). Genome analysis suggests that Candidatus *Hodgkina cicadicola*, a predicted Uroporphyrinogen III (UroIII) salvager (McCutcheon *et al.*, 2009), 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. 17% of cobamide-requiring human gut bacteria lacked genes to make UroIII de novo from glutamate, suggesting they could be UroIII salvagers (Degnan *et al.*, 2014a).

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 (Morris *et al.*, 2012). 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 (Morita *et al.*, 2008; Lawrence and Roth, 1996). 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.

Acknowledgments

This work was supported by NIH grant AI117984 and a Hellman Family Faculty Fund award to M.E.T., a Grace Kase Graduate Fellowship to A.N.S., and an NSF Graduate Fellowship to E.C.S, and NSF grant No. 1458808 to the J. Craig Venter Institute.

We thank Alison Smith, Johan Kudahl, Jared Leadbetter, 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, and Susan Leschine for providing bacterial strains.

Contributions

A.N.S., E.C.S., and A.W.H. performed the bioinformatic analysis. S.N.J. and D.R.H. created the BzaABCDEF HMMs. K.C.M., E.C.S., and A.N.S. performed growth assays and corrinoid extractions.

Conflict of Interest

The authors declare no conflict of interest.

References

Abreu NA, Taga ME. (2016). Decoding molecular interactions in microbial communities. *FEMS Microbiol Rev* **40**: 648–663.

Allen RH, Stabler SP. (2008). Identification and quantitation of cobalamin and cobalamin analogues in human feces. *Am J Clin Nutr* **87**: 1324–1335.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, *et al.* (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Anderson PJ, Lango J, Carkeet C, Britten A, Kräutler B, Hammock BD, *et al.* (2008). One pathway can incorporate either adenine or dimethylbenzimidazole as an alpha-axial ligand of B_{12} cofactors in Salmonella enterica. *J Bacteriol* **190**: 1160–71.

Banerjee R, Ragsdale SW. (2003). THE MANY FACES OF VITAMIN B_{12}: CATALYSIS BY COBALAMIN-DEPENDENT ENZYMES. *Annu Rev Biochem* **72**: 209–247.

Bertrand EM, Saito MA, Jeon YJ, Neilan BA. (2011). Vitamin B_{12} biosynthesis gene diversity in the Ross Sea: the identification of a new group of putative polar B_{12} biosynthesizers. *Environ Microbiol* **13**: 1285–98.

566 Broderick JB, Du BR, Duschene KS, Shepard EM. (2014). Radical S - Adenosylmethionine
567 Enzymes. *Chem Rev* **114**: 4229–4317.
568 Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, *et al.* (2015). Unusual biology
569 across a group comprising more than 15% of domain Bacteria. *Nature* **523**: 208–211.
570 Butzin NC, Secinaro MA, Swithers KS, Gogarten JP, Noll KM. (2013). Thermotoga lettingae
571 can salvage cobinamide to synthesize vitamin B_{12}. *Appl Environ Microbiol* **79**: 7006–12.
572 Campbell GRO, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, *et al.* (2006).
573 Sinorhizobium meliloti bluB is necessary for production of 5,6-dimethylbenzimidazole,
574 the lower ligand of B_{12}. *Proc Natl Acad Sci* **103**: 4634–4639.
575 Chan CH, Escalante-Semerena JC. (2011). ArsAB, a novel enzyme from Sporomusa ovata
576 activates phenolic bases for adenosylcobamide biosynthesis. *Mol Microbiol* **81**: 952–967.
577 Cornish-Bowden A. (2014). Current IUBMB recommendations on enzyme nomenclature and
578 kinetics. *Perspect Sci* **1**: 74–87.
579 de Crécy-Lagard V, El Yacoubi B, de la Garza RD, Noiriel A, Hanson AD. (2007). Comparative
580 genomics of bacterial and plant folate synthesis and salvage: Predictions and validations.
581 *BMC Genomics* **8**. e-pub ahead of print, doi: 10.1186/1471-2164-8-245.
582 Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. (2005). Algae acquire vitamin
583 B_{12} through a symbiotic relationship with bacteria. *Nature* **438**: 90–3.
584 Crofts TS, Hazra AB, Tran JLA, Sokolovskaya OM, Osadchiy V, Ad O, *et al.* (2014a).
585 Regiospecific Formation of Cobamide Isomers Is Directed by CobT. *Biochemistry* **53**:
586 7805–7815.
587 Crofts TS, Men Y, Alvarez-Cohen L, Taga ME. (2014b). A bioassay for the detection of
588 benzimidazoles reveals their presence in a range of environmental samples. *Front*

Microbiol **5**: 592.

Crofts TS, Seth EC, Hazra AB, Taga ME. (2013). Cobamide structure depends on both lower ligand availability and CobT substrate specificity. *Chem Biol* **20**: 1265–1274.

Degnan PH, Barry NA, Mok KC, Taga ME, Goodman AL. (2014a). Human gut microbes use multiple transporters to distinguish vitamin B_{12} analogs and compete in the gut. *Cell Host Microbe* **15**: 47–57.

Degnan PH, Taga ME, Goodman AL. (2014b). Vitamin B_{12} as a Modulator of Gut Microbial Ecology. *Cell Metab* **20**: 769–778.

Dobriner K. (1937). Porphyrin Excretion in the Feces in Normal and Pathological Conditions. *J Biol Chem* **120**: 115–128.

Eddy, SR. (2015). HMMER: biosequence analysis using profile hidden Markov models (v3.1b2) [software]. Available from http://hmmer.org/

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, *et al.* (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res* **44**: D279–D285.

Fontecave M. (1998). Ribonucleotide reductases and radical reactions. *Cell Mol Life Sci* **54**: 684–695.

Galperin MY, Makarova KS, Wolf YI, Koonin E V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* **43**: D261–D269.

Girard CL, Santschi DE, Stabler SP, Allen RH. (2009). Apparent ruminal synthesis and intestinal disappearance of vitamin B_{12} and its analogs in dairy cows. *J Dairy Sci* **92**: 4524–4529.

Di Girolamo PM, Bradbeer C. (1976). Transport of vitamin B_{12} in Escherichia coli. *J Bacteriol*

612 **106**: 745–750.

613 Gonzalez JC, Banerjee R V., Huang S, Sumner JS, Matthews RG. (1992). Comparison of
614 Cobalamin-independent and Cobalamin-Dependent Methionine Synthases from
615 Escherichia coli : Two Solutions to the Same Chemical Problem. *Biochemistry* **31**: 6045–
616 6056.

617 Gray MJ, Escalante-Semerena JC. (2010). A new pathway for the synthesis of alpha-ribazole-
618 phosphate in Listeria innocua. *Mol Microbiol* **77**: 1429–1438.

619 Gray MJ, Escalante-Semerena JC. (2009a). In vivo analysis of cobinamide salvaging in
620 Rhodobacter sphaeroides strain 2.4.1. *J Bacteriol* **191**: 3842–51.

621 Gray MJ, Escalante-Semerena JC. (2007). Single-enzyme conversion of FMNH2 to 5, 6-
622 dimethylbenzimidazole the lower ligand of B_{12}. *Proc Natl Acad Sci* **104**: 2921–2926.

623 Gray MJ, Escalante-Semerena JC. (2009b). The cobinamide amidohydrolase (cobyric acid-
624 forming) CbiZ enzyme: A critical activity of the cobamide remodelling system of
625 Rhodobacter sphaeroides. *Mol Microbiol* **74**: 1198–1210.

626 Haas CE, Rodionov DA, Kropat J, Malasarn D, Merchant SS, de Crécy-Lagard V. (2009). A
627 subset of the diverse COG0523 family of putative metal chaperones is linked to zinc
628 homeostasis in all kingdoms of life. *BMC Genomics* **10**: 470.

629 Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. (2012). TIGRFAMs and
630 genome properties in 2013. *Nucleic Acids Res* **41**: 387–395.

631 Haft DR, Haft DH. (2017). A comprehensive software suite for protein family construction and
632 functional site prediction. *PLoS One* **12**: e0171758.

633 Hazra AB, Ballou DP, Taga ME. (2018). Unique Biochemical and Sequence Features Enable
634 BluB to Destroy Flavin and Distinguish BluB from the Flavin Monooxygenase

635 Superfamily. *Biochemistry* **57**: 1748–1757.

636 Hazra AB, Han AW, Mehta AP, Mok KC, Osadchiy V, Begley TP, *et al.* (2015). Anaerobic
637 biosynthesis of the lower ligand of vitamin B_{12}. *Proc Natl Acad Sci* **112**: 10792–7.

638 Hazra AB, Tran JLA, Crofts TS, Taga ME. (2013). Analysis of substrate specificity in CobT
639 homologs reveals widespread preference for DMB, the lower axial ligand of vitamin B_{12}.
640 *Chem Biol* **20**: 1275–1285.

641 Helliwell KE, Lawrence AD, Holzer A, Kudahl UJ, Sasso S, Kräutler B, *et al.* (2016).
642 Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B_{12}.
643 *Curr Biol* **26**: 999–1008.

644 Hondorp ER, Matthews RG. (2004). Oxidative stress inactivates cobalamin-independent
645 methionine synthase (MetE) in Escherichia coli. *PLoS Biol* **2**: e336.

646 Ivanova EG, Fedorov DN, Doronina N V, Trotsenko YA. (2006). Production of vitamin B_{12} in
647 aerobic methylotrophic bacteria. Microbiology **75**: 494–496.

648 Johnson WM, Kido Soule MC, Kujawinski EB. (2016). Evidence for quorum sensing and
649 differential metabolite production by a marine bacterium in response to DMSP. *ISME J*
650 **10**: 2304–2316.

651 Kanto U, Jutamanee K, Osotsapar Y, Chai-arree W, Jintanawich W, Promdang S, *et al.* (2013).
652 Quantification of 5-aminolevulinic acid in swine manure extract by HPLC-fluorescence.
653 *J Liq Chromatogr Relat Technol* **36**: 2731–2748.

654 Kazamia E, Czesnick H, Nguyen TT Van, Croft MT, Sherwood E, Sasso S, *et al.* (2012).
655 Mutualistic interactions between vitamin B_{12} -dependent algae and heterotrophic bacteria
656 exhibit regulation. *Environ Microbiol* **14**: 1466–76.

657 Keller S, Kunze C, Bommer M, Paetz C, Menezes RC, Svatoš A, *et al.* (2018). Selective

Utilization of Benzimidazolyl-Norcobamides as Cofactors by the Tetrachloroethene Reductive Dehalogenase of Sulfurospirillum multivorans. *J Bacteriol* **200**: e00584-17.

Keller S, Ruetz M, Kunze C, Kräutler B, Diekert G, Schubert T. (2013). Exogenous 5,6-dimethylbenzimidazole caused production of a non-functional tetrachloroethene reductive dehalogenase in Sulfurospirillum multivorans. *Environ Microbiol* **16**: 3361–3369.

Keller S, Treder A, Von Reuss SH, Escalante-Semerena JC, Schubert T. (2016). The SMUL_1544 gene product governs norcobamide biosynthesis in the tetrachloroethene-respiring bacterium Sulfurospirillum multivorans. *J Bacteriol* **198**: 2236–2243.

Kilstrup M, Hammer K, Jensen PR, Martinussen J. (2005). Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiol Rev* **29**: 555–590.

Kräutler B, Fieber W, Ostermann S, Fasching M, Ongania KH, Gruber K, *et al.* (2003). The Cofactor of Tetrachloroethene Reductive Dehalogenase of Dehalospirillum multivorans Is Norpseudo-B_{12}, a New Type of a Natural Corrinoid. *Helv Chim Acta* **86**: 3698–3716.

Lawrence J, Roth J. (1996). Evolution of coenzyme B_{12} synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. *Genetics* **142**: 11–24.

Lyell NL, Septer AN, Dunn AK, Duckett D, Stoudenmire JL, Stabb E V. (2017). An expanded transposon mutant library reveals that Vibrio fischeri δ-aminolevulinate auxotrophs can colonize Euprymna scolopes. *Appl Environ Microbiol* **83**: e02470-16.

Magnúsdóttir S, Ravcheev D, de Crécy-Lagard V, Thiele I. (2015). Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front Genet* **6**: 148.

Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, *et al.* (2012). IMG:

681 The integrated microbial genomes database and comparative analysis system. *Nucleic*
682 *Acids Res* **40**: D115–D122.
683 Mattes TA, Escalante-Semerena JC. (2017). Salmonella enterica synthesizes 5,6-
684 dimethylbenzimidazolyl-(DMB)-α-riboside. Why some Firmicutes do not require the
685 canonical DMB activation system to synthesize adenosylcobalamin. *Mol Microbiol* **103**:
686 269–281.
687 McCutcheon JP, McDonald BR, Moran NA. (2009). Convergent evolution of metabolic roles in
688 bacterial co-symbionts of insects. *Proc Natl Acad Sci* **106**: 15394–15399.
689 McGoldrick HM, Roessner CA, Raux E, Lawrence AD, McLean KJ, Munro AW, *et al.* (2005).
690 Identification and characterization of a novel vitamin B_{12} (cobalamin) biosynthetic
691 enzyme (CobZ) from Rhodobacter capsulatus, containing flavin, heme, and Fe-S
692 cofactors. *J Biol Chem* **280**: 1086–1094.
693 Mehta AP, Abdelwahed SH, Fenwick MK, Hazra AB, Taga ME, Zhang Y, *et al.* (2015).
694 Anaerobic 5-Hydroxybenzimidazole Formation from Aminoimidazole Ribotide: An
695 Unanticipated Intersection of Thiamin and Vitamin B_{12} Biosynthesis. *J Am Chem Soc*
696 **137**: 10444–10447.
697 Men Y, Seth EC, Yi S, Allen RH, Taga ME, Alvarez-Cohen L. (2014a). Sustainable growth of
698 Dehalococcoides mccartyi 195 by corrinoid salvaging and remodeling in defined lactate-
699 fermenting consortia. *Appl Environ Microbiol* **80**: 2133–2141.
700 Men Y, Seth EC, Yi S, Crofts TS, Allen RH, Taga ME, *et al.* (2014b). Identification of specific
701 corrinoids reveals corrinoid modification in dechlorinating microbial communities.
702 *Environ Microbiol* **17**: 4873–4884.
703 Mok KC, Taga ME. (2013). Growth inhibition of Sporomusa ovata by incorporation of

benzimidazole bases into cobamides. *J Bacteriol* **195**: 1902–11.

Morita H, Toh H, Fukuda S, Horikawa H, Oshima K, Suzuki T, *et al.* (2008). Comparative genome analysis of Lactobacillus reuteri and Lactobacillus fermentum reveal a genomic island for reuterin and cobalamin production. *DNA Res* **15**: 151–161.

Morris JJ, Lenski RE, Zinser ER. (2012). The black queen hypothesis: Evolution of dependencies through adaptive gene loss. *MBio* **3**: e00036-12.

Nahvi A. (2004). Coenzyme B_{12} riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* **32**: 143–150.

Newmister SA, Chan CH, Escalante-Semerena JC, Rayment I. (2012). Structural insights into the function of the nicotinate mononucleotide:phenol/p-cresol phosphoribosyltransferase (ArsAB) enzyme from Sporomusa ovata. *Biochemistry* **51**: 8571–8582.

Paerl RW, Bouget F-Y, Lozano J-C, Vergé V, Schatt P, Allen EE, *et al.* (2016). Use of plankton-derived vitamin B1 precursors, especially thiazole-related precursor, by key marine picoeukaryotic phytoplankton. *ISME J* **11**: 753–765.

Qi M, Nelson KE, Daugherty SC, Nelson WC, Hance IR, Morrison M, *et al.* (2008). Genomic differences between Fibrobacter succinogenes S85 and Fibrobacter intestinalis DR7, identified by suppression subtractive hybridization. *Appl Environ Microbiol* **74**: 987–993.

Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. (2007). Prediction of effective genome size in metagenomic samples. *Genome Biol* **8**: R10.

Ragsdale SW, Pierce E. (2008). Acetogenesis and the Wood-Ljungdahl Pathway of CO2 Fixation. *Biochim Biophys Acta* **1784**: 1873–1898.

Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. (2003). Comparative Genomics of the Vitamin B_{12} Metabolism and Regulation in Prokaryotes. *J Biol Chem* **278**: 41148–41159.

727 Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B_{12}): Synthesis and Biological
728 Significance. *Annu Rev Microbiol* **50**: 137–181.
729 Sangwan I, O'Brian MR. (1991). Evidence for an Inter-organismic heme biosynthetic pathway in
730 symbiotic soybean root nodules. *Science (80-)* **251**: 1220–1222.
731 Schubert HL, Raux E, Wilson KS, Warren MJ. (1999). Common chelatase design in the
732 branched tetrapyrrole pathways of heme and anaerobic cobalamin synthesis.
733 *Biochemistry* **38**: 10660–10669.
734 Seth EC, Taga ME. (2014). Nutrient cross-feeding in the microbial world. *Front Microbiol* **5**:
735 350.
736 Sloan DB, Moran NA. (2012). Genome reduction and co-evolution between the primary and
737 secondary bacterial symbionts of psyllids. *Mol Biol Evol* **29**: 3781–3792.
738 Stupperich E, Eisinger HJ. (1989). Biosynthesis of para-cresolyl cobamide in Sporomusa ovata.
739 *Arch Microbiol* **151**: 372–377.
740 Taga ME, Larsen NA, Howard-Jones AR, Walsh CT, Walker GC. (2007). BluB cannibalizes
741 flavin to form the lower ligand of vitamin B_{12}. *Nature* **446**: 449–453.
742 Taga ME, Walker GC. (2010). Sinorhizobium meliloti Requires a Cobalamin-Dependent
743 Ribonucleotide Reducatase for Symbiosis with its Plant Host. *MPMI* **23**: 1643–1654.
744 Warren MJ, Raux E, Schubert HL, Escalante-Semerena JC. (2002). The biosynthesis of
745 adenosylcobalamin (vitamin B_{12}). *Nat Prod Rep* **19**: 390–412.
746 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. (2009). Jalview Version 2-A
747 multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–
748 1191.
749 Watson CJ, Schwartz S, Hawkinson V. (1945). 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.

Wienhausen G, Noriega-Ortega BE, Niggemann J, Dittmar T, Simon M. (2017). The Exometabolome of Two Model Strains of the Roseobacter Group: A Marketplace of Microbial Metabolites. *Front Microbiol* **8**: 1985.

Woodson JD, Escalante-Semerena JC. (2004). CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme B_{12} precursor cobinamide in archaea. *Proc Natl Acad Sci* **101**: 3591–3596.

Xie B, Bishop S, Stessman D, Wright D, Spalding MH, Halverson LJ. (2013). Chlamydomonas reinhardtii thermal tolerance enhancement mediated by a mutualistic interaction with vitamin B_{12}-producing bacteria. *ISME J* **7**: 1544–1555.

Yan J, Bi M, Bourdon AK, Farmer AT, Wang P-H, Molenda O, *et al.* (2018). Purinyl-cobamide is a native prosthetic group of reductive dehalogenases. *Nat Chem Biol* **14**: 8–14.

Yan J, Ritalahti KM, Wagner DD, Löffler FE. (2012). Unexpected specificity of interspecies cobamide transfer from Geobacter spp. to organohalide-respiring Dehalococcoides mccartyi strains. *Appl Environ Microbiol* **78**: 6630–6636.

Yi S, Seth E, Men Y, Stabler SP, Allen RH, Alvarez-Cohen L, *et al.* (2012). Versatility in corrinoid salvaging and remodeling pathways supports corrinoid-dependent metabolism in Dehalococcoides mccartyi. *Appl Environ Microbiol* **78**: 7745–7752.

Zallot R, Ross R, Chen WH, Bruner SD, Limbach PA, de Crécy-Lagard V. (2017). Identification of a Novel Epoxyqueuosine Reductase Family by Comparative Genomics. *ACS Chem*

773 *Biol* **12**: 844–851.

774 Zhang Y, Rodionov DA, Gelfand MS, Gladyshev VN. (2009). Comparative genomic analyses of

775 nickel, cobalt and vitamin B_{12} utilization. *BMC Genomics* **10**: 78.

776

777

778 **Figure Legends**

779 **Figure 1: Functions carried out by cobamide-dependent processes.** Reactions carried out by
780 cobamide-dependent enzymes are shown on the left side of the arrows and cobamide-
781 independent alternative processes, if known, on the right. Annotations or query genes used for
782 searching for each function are listed in Supplementary Table 4.
783
784 **Figure 2: Cobamide biosynthesis and structure. A.** The cobamide biosynthesis pathway is
785 shown with each enzymatic step indicated by a white box labeled with the gene names and
786 functional annotation. Subsections of the pathway and salvaging and remodeling pathways are
787 bracketed or boxed with labels in bold. Orthologous enzymes that carry out similar reactions in
788 aerobic and anaerobic corrin ring biosynthesis are indicated by dashed lines. **B.** Structure of
789 cobalamin. The upper ligand R can be an adenosyl or methyl group. Classes of possible lower
790 ligand structures are also shown. Benzimidazoles: R1=H, OH, CH_3, OCH_3; R2=H, OH, CH_3,
791 OCH_3. Purines: R1=H, CH_3, NH_2; R2=H, NH_2, OH, O. Phenolics: R=H, CH_3.
792
793 **Figure 3: Cobamide dependence in bacteria. A.** Histogram of the number of cobamide-
794 dependent enzyme families (shown in Fig. 1, Supplementary table 4) per genome in the complete
795 filtered data set and the four most abundant phyla in the data set. The numbers are given for bars
796 with values less than 1%. The inset lists the mean, standard deviation, median, and mode of
797 cobamide-dependent enzyme families for each phylum. **B.** Rank abundance of cobamide-
798 dependent enzyme families in the filtered data set and the four most abundant phyla. The inset
799 shows an expanded view of the nine less abundant functions. **C.** Abundance of five cobamide-
800 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.

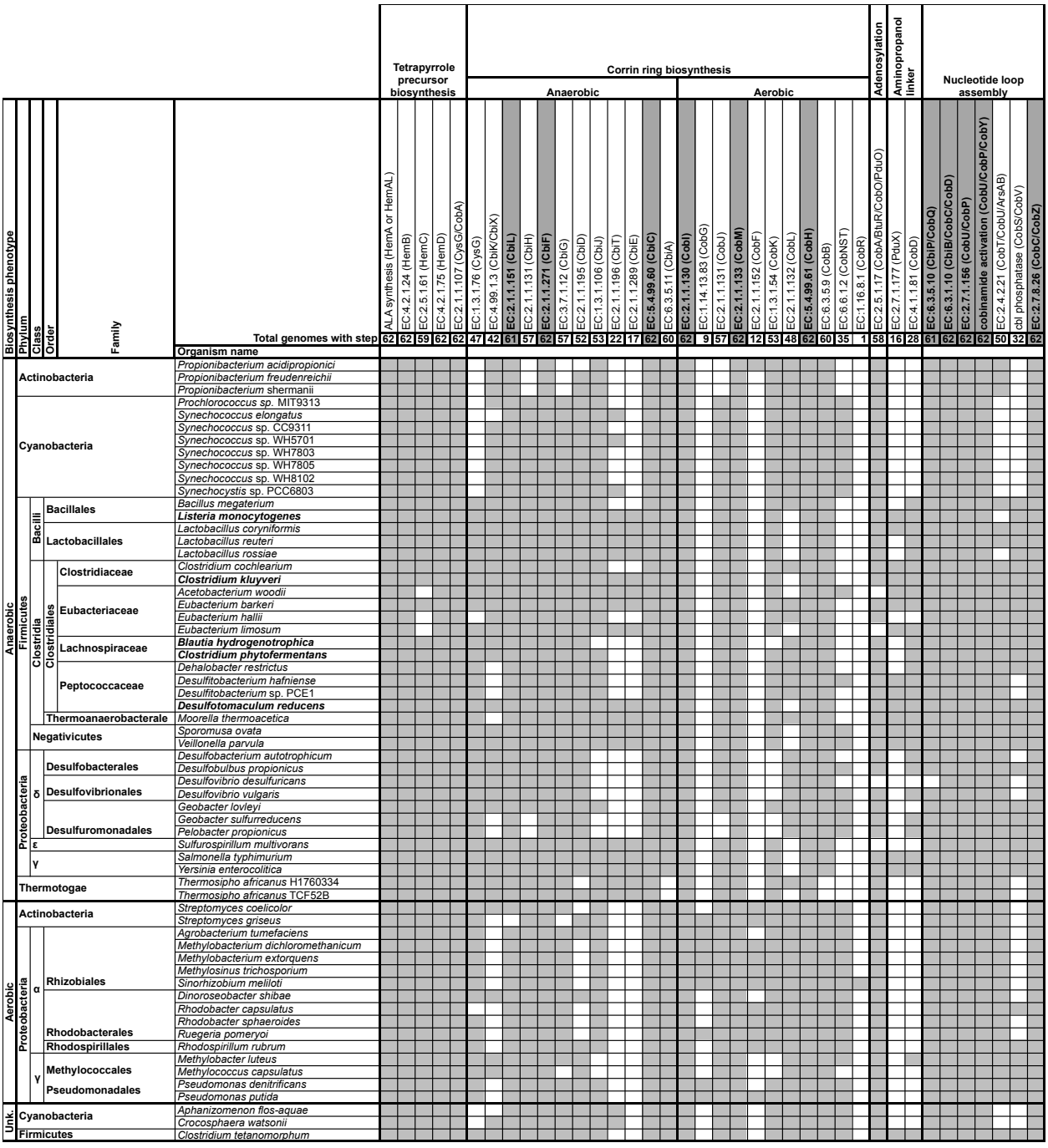
Figure 4: Predicted cobamide biosynthesis phenotypes in the complete filtered data set and the four most abundant phyla in the data set. Genomes were classified into predicted corrinoid biosynthesis phenotypes based on the criteria listed in Supplementary Table 7. The "Partial biosynthesis" category 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 5: 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. Abbreviations: AIR, aminoimidazole ribotide; 5-OHBza, 5-hydroxybenzimidazole; 5-OMeBza, 5-methoxybenzimidazole; 5-OMe-6-MeBza, 5-methoxy-6-methylbenzimidazole.

Figure 6: Characterization of putative tetrapyrrole precursor salvagers A. Early 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 tetrapyrrole precursor salvage category is on the left. Two genomes had cobamide biosynthesis pathways inconsistent with simple auxotrophy. Specific 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 (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 or ALA.

Tables

Table 1. Experimentally-verified cobamide producers and their cobamide biosynthesis annotation content.



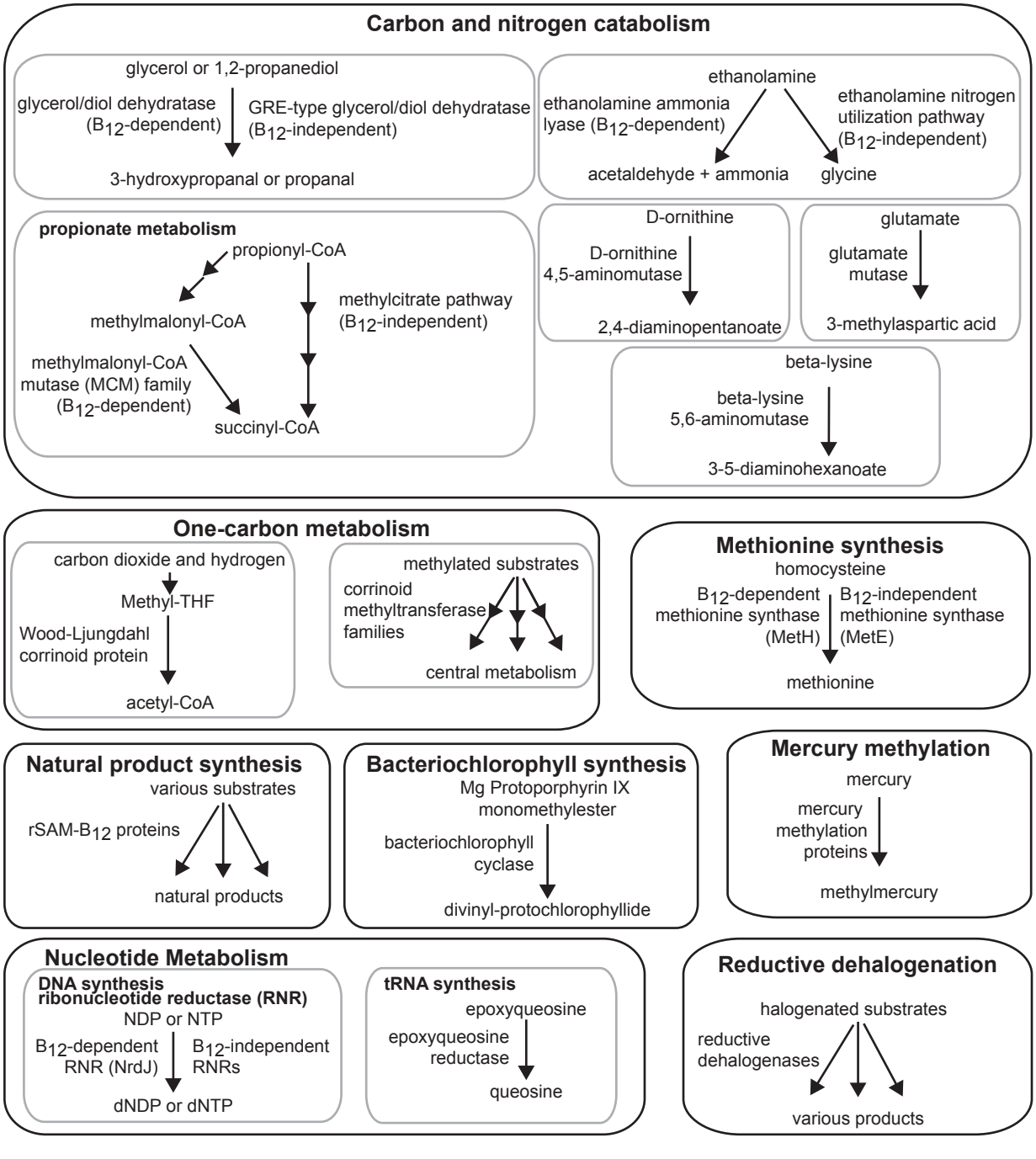
*_Methylophilus methylotrophus_* was also reported to be an aerobic corrinoid producer, but its genome only has 1 corrin ring biosynthesis annotation (CobH). The reported concentration of corrinoid it produced is 6-fold less than other strains in the study by Ivanova et al. (2006). We do not think this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

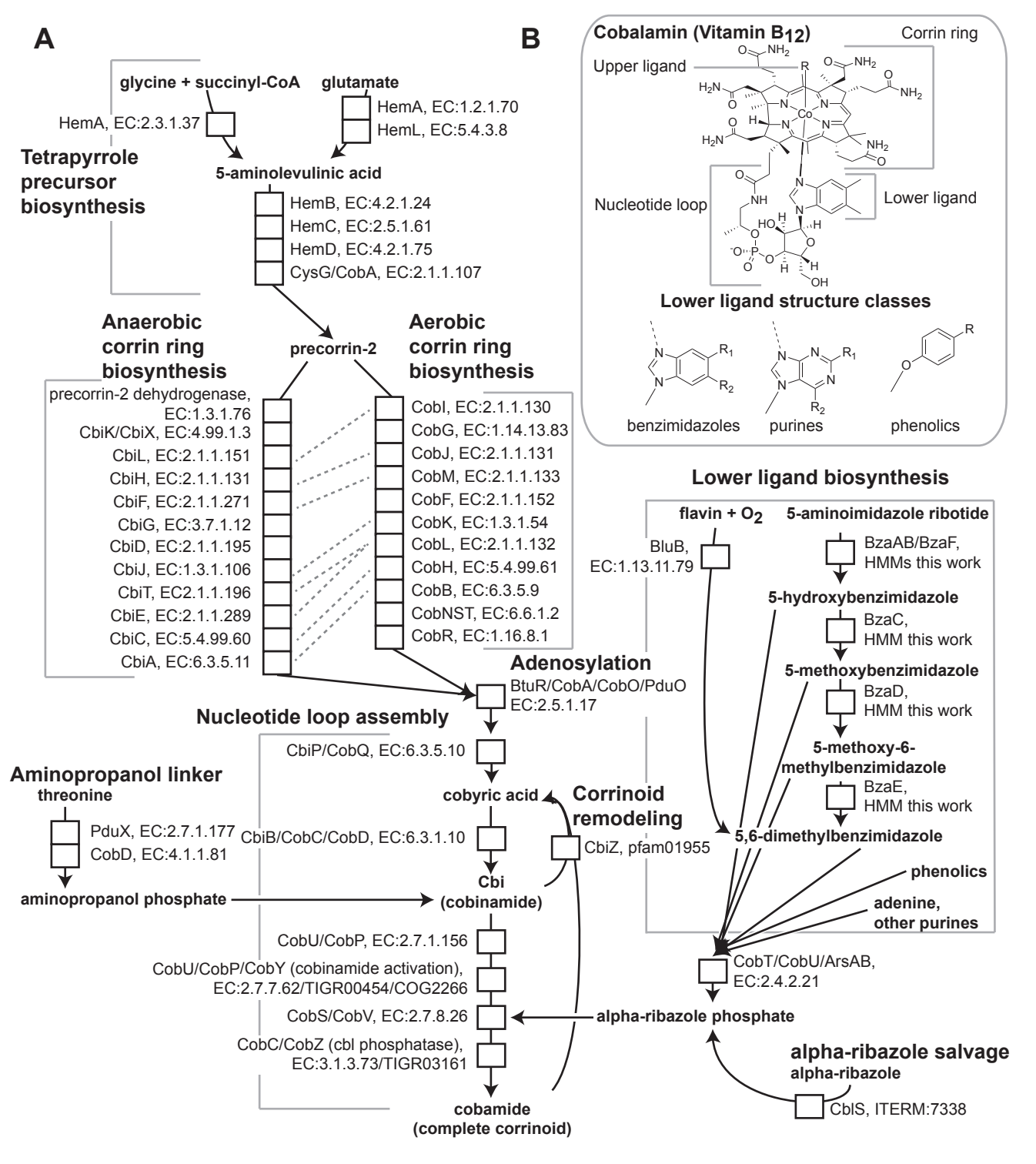
Bold species names were identified as cobamide producers in this study (Fig. S3).

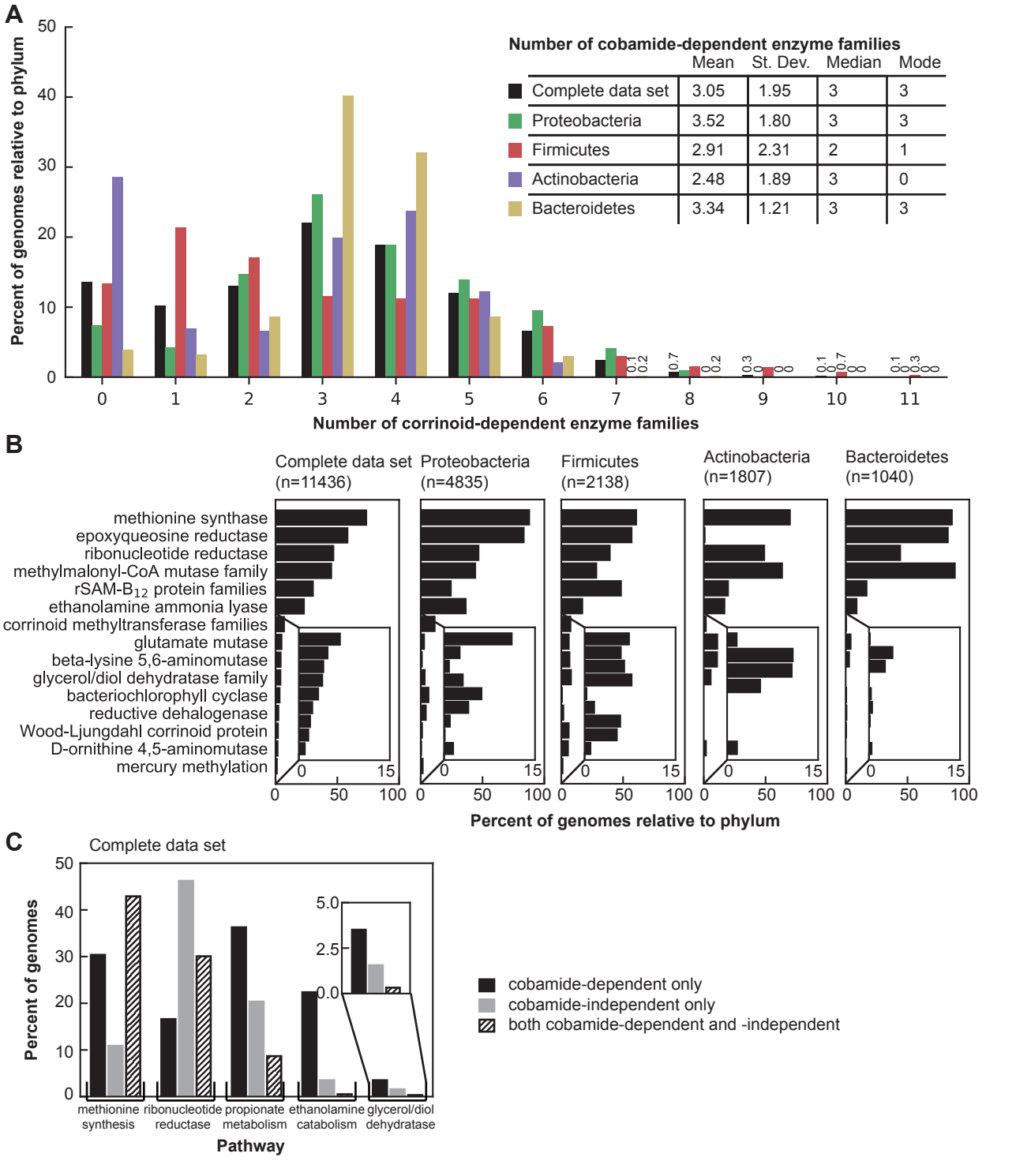
Table 2. Presence of corrin ring marker annotations in predicted cobamide biosynthesis categories

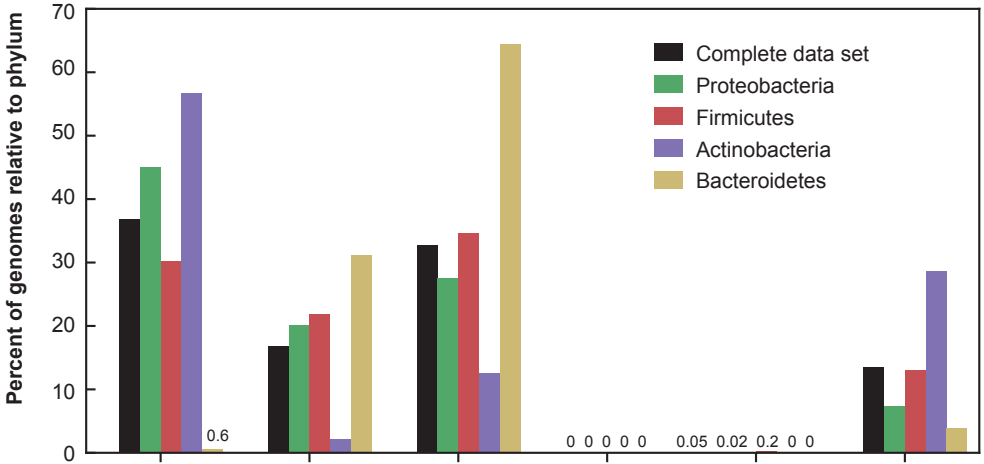
	Cobamide biosynthesis category								
	Cobamide producers				Partial biosynthesis		Non-producers		
	Very likely n=1016	Likely n=2361	Possible n=832	All cobamide producers n=4209	Tetrapyrrole precursor salvager n=201	Cbi salvage n=1734	Likely n=29	Very likely n=5263	All non-producers n=5292
CbiL/CobI	100.0*	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.

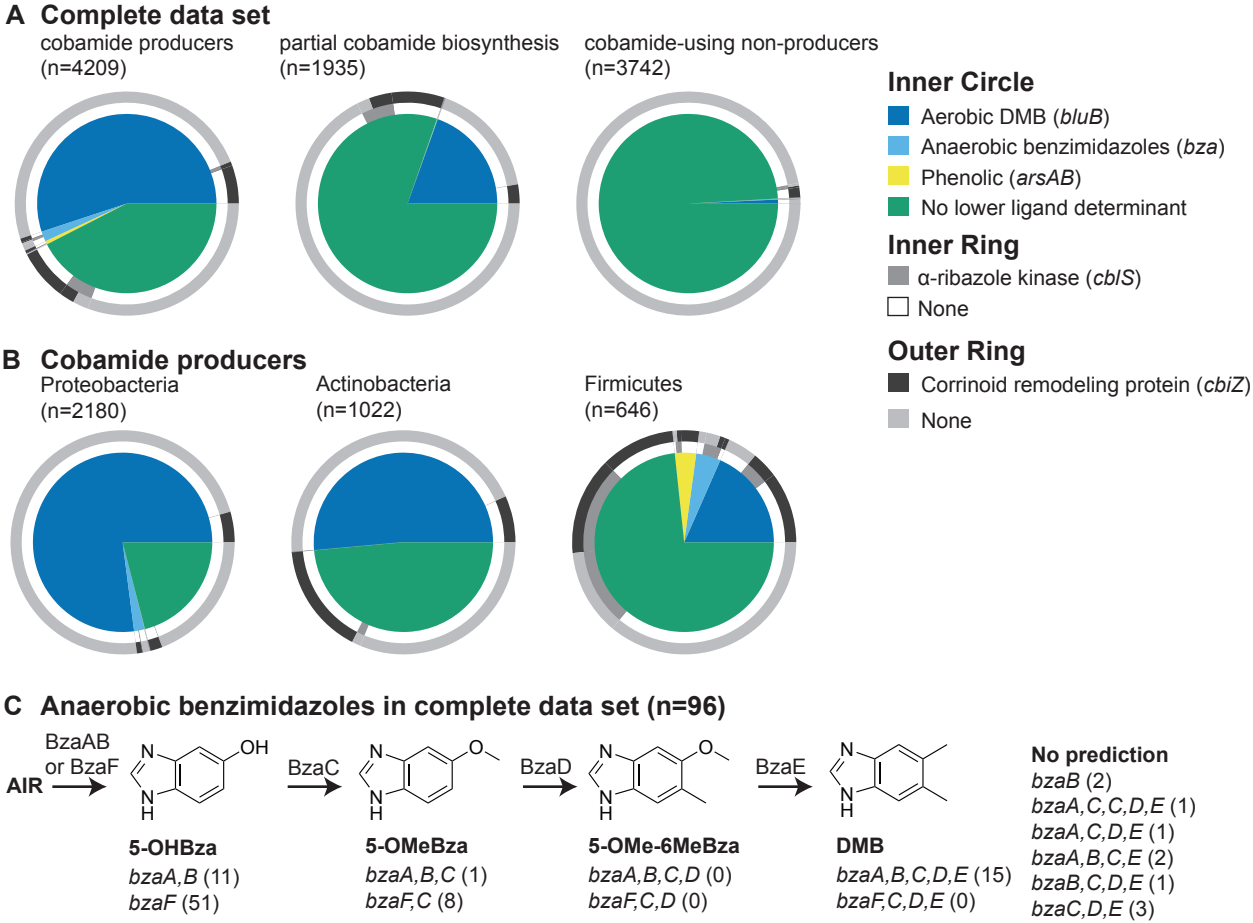

Carbon and nitrogen catabolism
glycerol or 1,2-propanediol
glycerol/diol dehydratase (B12-dependent)
GRE-type glycerol/diol dehydratase (B12-independent)
3-hydroxypropanal or propanal
ethanolamine
ethanolamine ammonia lyase (B12-dependent)
ethanolamine nitrogen utilization pathway (B12-independent)
acetaldehyde + ammonia
glycine
propionate metabolism
propionyl-CoA
methylmalonyl-CoA
methylcitrate pathway (B12-independent)
methylmalonyl-CoA mutase (MCM) family (B12-dependent)
succinyl-CoA
D-ornithine
D-ornithine 4,5-aminomutase
2,4-diaminopentanoate
glutamate
glutamate mutase
3-methylaspartic acid
beta-lysine
beta-lysine 5,6-aminomutase
3-5-diaminohexanoate
One-carbon metabolism
carbon dioxide and hydrogen
Methyl-THF
Wood-Ljungdahl corrinoid protein
acetyl-CoA
methylated substrates
corrinoid methyltransferase families
central metabolism
Methionine synthesis
homocysteine
B12-dependent methionine synthase (MetH)
B12-independent methionine synthase (MetE)
methionine
Natural product synthesis
various substrates
rSAM-B12 proteins
natural products
Bacteriochlorophyll synthesis
Mg Protoporphyrin IX monomethylester
bacteriochlorophyll cyclase
divinyl-protochlorophyllide
Mercury methylation
mercury
mercury methylation proteins
methylmercury
Nucleotide Metabolism
DNA synthesis
ribonucleotide reductase (RNR)
NDP or NTP
B12-dependent RNR (NrdJ)
B12-independent RNRs
dNDP or dNTP
tRNA synthesis
epoxyqueosine
epoxyqueosine reductase
queosine
Reductive dehalogenation
halogenated substrates
reductive dehalogenases
various products

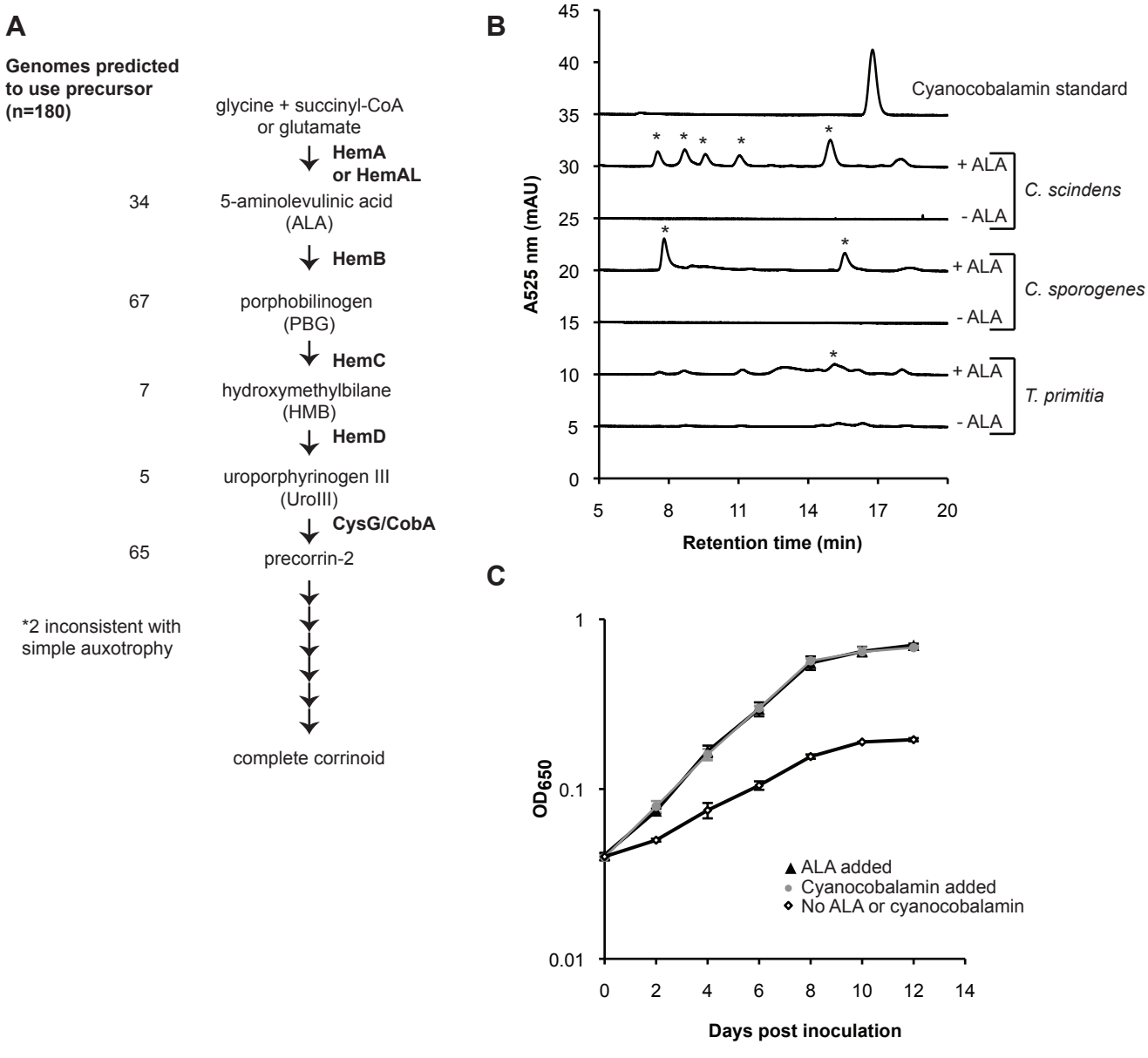

A
B
glycine + succinyl-CoA
glutamate
HemA, EC:2.3.1.37
HemA, EC:1.2.1.70
HemL, EC:5.4.3.8
Tetrapyrrole precursor biosynthesis
5-aminolevulinic acid
HemB, EC:4.2.1.24
HemC, EC:2.5.1.61
HemD, EC:4.2.1.75
CysG/CobA, EC:2.1.1.107
Anaerobic corrin ring biosynthesis
precorrin-2
Aerobic corrin ring biosynthesis
precorrin-2 dehydrogenase, EC:1.3.1.76
CbiK/CbiX, EC:4.99.1.3
CbiL, EC:2.1.1.151
CbiH, EC:2.1.1.131
CbiF, EC:2.1.1.271
CbiG, EC:3.7.1.12
CbiD, EC:2.1.1.195
CbiJ, EC:1.3.1.106
CbiT, EC2.1.1.196
CbiE, EC:2.1.1.289
CbiC, EC:5.4.99.60
CbiA, EC:6.3.5.11
CobI, EC:2.1.1.130
CobG, EC:1.14.13.83
CobJ, EC:2.1.1.131
CobM, EC:2.1.1.133
CobF, EC:2.1.1.152
CobK, EC:1.3.1.54
CobL, EC:2.1.1.132
CobH, EC:5.4.99.61
CobB, EC:6.3.5.9
CobNST, EC:6.6.1.2
CobR, EC:1.16.8.1
Cobalamin (Vitamin B12)
Corrin ring
Upper ligand
Nucleotide loop
Lower ligand
Lower ligand structure classes
benzimidazoles
purines
phenolics
Lower ligand biosynthesis
flavin + O2
5-aminoimidazole ribotide
BluB, EC:1.13.11.79
BzaAB/BzaF, HMMs this work
5-hydroxybenzimidazole
BzaC, HMM this work
5-methoxybenzimidazole
BzaD, HMM this work
5-methoxy-6-methylbenzimidazole
BzaE, HMM this work
5,6-dimethylbenzimidazole
phenolics
adenine, other purines
Adenosylation
BtuR/CobA/CobO/PduO EC:2.5.1.17
Nucleotide loop assembly
CbiP/CobQ, EC:6.3.5.10
Aminopropanol linker
threonine
PduX, EC:2.7.1.177
CobD, EC:4.1.1.81
aminopropanol phosphate
cobyric acid
CbiB/CobC/CobD, EC:6.3.1.10
Cbi (cobinamide)
Corrinoid remodeling
CbiZ, pfam01955
CobU/CobP, EC:2.7.1.156
CobU/CobP/CobY (cobinamide activation), EC:2.7.7.62/TIGR00454/COG2266
CobS/CobV, EC:2.7.8.26
CobC/CobZ (cbl phosphatase), EC:3.1.3.73/TIGR03161
cobamide (complete corrinoid)
CobT/CobU/ArsAB, EC:2.4.2.21
alpha-ribazole phosphate
alpha-ribazole salvage
alpha-ribazole
CblS, ITERM:7338


A
Percent of genomes relative to phylum
Number of corrinoid-dependent enzyme families
Number of cobamide-dependent enzyme families
Mean St. Dev. Median Mode
Complete data set 3.05 1.95 3 3
Proteobacteria 3.52 1.80 3 3
Firmicutes 2.91 2.31 2 1
Actinobacteria 2.48 1.89 3 0
Bacteroidetes 3.34 1.21 3 3
B
Complete data set (n=11436)
Proteobacteria (n=4835)
Firmicutes (n=2138)
Actinobacteria (n=1807)
Bacteroidetes (n=1040)
methionine synthase
epoxyqueosine reductase
ribonucleotide reductase
methylmalonyl-CoA mutase family
rSAM-B12 protein families
ethanolamine ammonia lyase
corrinoid methyltransferase families
glutamate mutase
beta-lysine 5,6-aminomutase
glycerol/diol dehydratase family
bacteriochlorophyll cyclase
reductive dehalogenase
Wood-Ljungdahl corrinoid protein
D-ornithine 4,5-aminomutase
mercury methylation
Percent of genomes relative to phylum
C
Complete data set
Percent of genomes
methionine synthesis
ribonucleotide reductase
propionate metabolism
ethanolamine catabolism
glycerol/diol dehydratase
Pathway
cobamide-dependent only
cobamide-independent only
both cobamide-dependent and -independent



Cobamide biosynthesis category	Complete biosynthesis	Partial biosynthesis	No biosynthesis	Complete biosynthesis	Partial biosynthesis	No biosynthesis
Uses cobamides	Yes	Yes	Yes	No	No	No


A Complete data set
cobamide producers
(n=4209)
partial cobamide biosynthesis
(n=1935)
cobamide-using non-producers
(n=3742)
Inner Circle
Aerobic DMB (bluB)
Anaerobic benzimidazoles (bza)
Phenolic (arsAB)
No lower ligand determinant
Inner Ring
α-ribazole kinase (cblS)
None
Outer Ring
Corrinoid remodeling protein (cbiZ)
None
B Cobamide producers
Proteobacteria
(n=2180)
Actinobacteria
(n=1022)
Firmicutes
(n=646)
C Anaerobic benzimidazoles in complete data set (n=96)
AIR
BzaAB
or BzaF
BzaC
BzaD
BzaE
5-OHBza
bzaA,B (11)
bzaF (51)
5-OMeBza
bzaA,B,C (1)
bzaF,C (8)
5-OMe-6MeBza
bzaA,B,C,D (0)
bzaF,C,D (0)
DMB
bzaA,B,C,D,E (15)
bzaF,C,D,E (0)
No prediction
bzaB (2)
bzaA,C,C,D,E (1)
bzaA,C,D,E (1)
bzaA,B,C,E (2)
bzaB,C,D,E (1)
bzaC,D,E (3)


A
Genomes predicted to use precursor (n=180)
glycine + succinyl-CoA or glutamate
HemA or HemAL
34
5-aminolevulinic acid (ALA)
HemB
67
porphobilinogen (PBG)
HemC
7
hydroxymethylbilane (HMB)
HemD
5
uroporphyrinogen III (UroIII)
CysG/CobA
65
precorrin-2
*2 inconsistent with simple auxotrophy
complete corrinoid
B
A525 nm (mAU)
Retention time (min)
Cyanocobalamin standard
+ ALA
- ALA
C. scindens
C. sporogenes
T. primitia
C
OD650
Days post inoculation
ALA added
Cyanocobalamin added
No ALA or cyanocobalamin