Vitamin B₁₂ Synthesis and Salvage Pathways Were Acquired by Horizontal Gene Transfer to the Thermotogales

Kristen S. Swithers[1], Amanda K. Petrus[1], Michael A. Secinaro[1], Camilla L. Nesbø[2,3], J. Peter Gogarten[1], Kenneth M. Noll[1], and Nicholas C. Butzin[1,*]

[1]Department of Molecular and Cell Biology, University of Connecticut

[2]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, Blindern, Canada

[3]Department of Biological Sciences, University of Alberta, Edmonton, Canada

*Corresponding author: E-mail: nicholas.butzin@gmail.com.

Abstract

The availability of genome sequences of Thermotogales species from across the order allows an examination of the evolutionary origins of phenotypic characteristics in this lineage. Several studies have shown that the Thermotogales have acquired large numbers of genes from distantly related lineages, particularly Firmicutes and Archaea. Here, we report the finding that some Thermotogales acquired the ability to synthesize vitamin B₁₂ by acquiring the requisite genes from these distant lineages. *Thermosipho* species, uniquely among the Thermotogales, contain genes that encode the means to synthesize vitamin B₁₂ de novo from glutamate. These genes are split into two gene clusters: the corrinoid synthesis gene cluster, that is unique to the *Thermosipho* and the cobinamide salvage gene cluster. The corrinoid synthesis cluster was acquired from the Firmicutes lineage, whereas the salvage pathway is an amalgam of bacteria- and archaea-derived proteins. The cobinamide salvage gene cluster has a patchy distribution among Thermotogales species, and ancestral state reconstruction suggests that this pathway was present in the common Thermotogales ancestor. We show that *Thermosipho africanus* can grow in the absence of vitamin B₁₂, so its de novo pathway is functional. We detected vitamin B₁₂ in the extracts of *T. africanus* cells to verify the synthetic pathway. Genes in *T. africanus* with apparent B₁₂ riboswitches were found to be down-regulated in the presence of vitamin B₁₂ consistent with their roles in B₁₂ synthesis and cobinamide salvage.

Key words: cobalamin, cobinamide, Thermotogales, vitamin B₁₂, horizontal gene transfer.

Introduction

The Thermotogales is defined by a unique outer envelope (Balows 1992) and is one of the deepest bacterial lineages in the "ribosomal tree of life" (Achenbach-Richter et al. 1987; Fournier and Gogarten 2010; Williams et al. 2011). As in many other prokaryote lineages, the genomes of the Thermotogales are subject to frequent gene transfers (Beiko et al. 2005), with the largest numbers of transfers involving the Archaea and the Firmicutes (Zhaxybayeva et al. 2009).

Vitamin B₁₂ (B₁₂), also termed cobalamin, is the generic name for a group of biologically active corrinoids that serves as a cofactor essential to cell growth across all branches of life. Only a few archaea and bacteria have the ability to synthesize this relatively large molecule. De novo synthesis starts with glutamate and requires over 30 gene products to produce an active form of B₁₂ (Martens et al. 2002). There are two pathways for de novo synthesis, an anaerobic pathway and aerobic pathway (Warren et al. 2002). The major difference between the two pathways is the timing of the addition of the central cobalt atom (early in the anaerobic pathway and later in the aerobic pathway). Although many prokaryotes require B₁₂ as a cofactor, many do not have genes for de novo synthesis. A report suggests that of the sequenced bacterial genomes, only half that utilize B₁₂ can synthesize it (Zhang et al. 2009). Some, instead salvage incomplete corrinoid rings called cobinamides and use these as precursors to synthesize an active form of B₁₂ (Escalante-Semerena 2007). Many other organisms acquire B₁₂ from the environment using a B₁₂/cobinamide BtuFCD ABC transporter (Woodson et al. 2005).

The B₁₂/cobinamide ABC transporter, B₁₂ synthesis, and succinyl-CoA (SucCoA) synthetase genes are regulated

by B$_{12}$ riboswitches (Rodionov et al. 2003). Riboswitches are small RNAs (~200 bp) that can regulate either at the transcriptional or at the translational level. A computational approach has been developed to predict the type of regulation of riboswitches (Vitreschak et al. 2003). If the riboswitch overlaps with the adjacent gene, then it may act at the translational level, and if the riboswitch does not overlap with the adjacent gene, then it may regulate at the transcriptional level (Vitreschak et al. 2003). On the basis of these criteria, they suggest that the BtuFCD ABC transporter in *Thermotoga maritima* is under transcriptional regulation (Vitreschak et al. 2003). We provide measures of transcription in response to the availability of B$_{12}$ that support that suggestion.

Since publication of the *Tt. maritima* genome (Nelson et al. 1999), representatives from several Thermotogales genera have been sequenced (Nesbø et al. 2009; Zhaxybayeva et al. 2009; Swithers et al. 2011a, 2011b) revealing the genomic diversity within the order. One example of this diversity is found in the B$_{12}$ biosynthesis pathway. The production and utilization of this molecule has been well studied in many bacteria and a few archaea (Escalante-Semerena 2007), but no studies have considered the deep branching Thermotogales lineage. This work explores the possible origins of B$_{12}$-related Thermotogales genes focusing on gene gain and loss. Here, we show that the cobinamide salvage pathway was likely the ancestral B$_{12}$ biosynthesis pathway for the order, and de novo synthesis was a later addition only to the *Thermosipho* (*Ts.*) lineage. We also show that *Thermosipho africanus* strains produce B$_{12}$.

Materials and Methods

Genome Sequences and Strains

Fourteen genome sequences were used in this study, 13 of which are completely sequenced and were downloaded from the NCBI GenBank database (Nelson et al. 1999; Nesbø et al. 2009; Zhaxybayeva et al. 2009; Swithers et al. 2011a, 2011b). The *Ts. africanus* H1760334 partial genome was submitted to the RAST server for annotation (Aziz et al. 2008). Gene functional names were cross-referenced between the KEGG (Moriya et al. 2007), SEED (Overbeek et al. 2005), and the MetaCyc (Caspi et al. 2012) databases. The final functional names were taken from the SEED database.

Proteins with B$_{12}$-Binding Domains

To determine which proteins contained B$_{12}$-binding domains and to cross check their putative function, Reverse Position Specific Blast v.2.2.23+ (rpsBlast) was used with an E-value cutoff of 1E-6 (Altschul et al. 1997). Each protein from each genome was used as a query against the CDD position scoring matrix database (Marchler-Bauer et al. 2011) proteins with hits to B$_{12}$-binding domains were saved.

Phylogenetic Trees

For each tree, homologs were gathered from the NCBI nonredundant database. An E-value cutoff of 1E-4 was used, and the number of sequences saved was set well above the number of possible sequences in the database to insure all homologs were gathered. For each organism with multiple hits, the hit with the lowest E-value was retained in the dataset; thus, one representative of each species was present in the dataset. This significantly reduced the datasets to a manageable size and retained the taxonomic sampling. Sequences were aligned using default parameters in MUSCLE v3.8.31 (Edgar 2004). ProtTest3 (Darriba et al. 2011) was used to assess appropriate parameters for phylogenetic reconstruction, which was the LG + I + G model for each dataset. After visual inspection of the alignment, trees were reconstructed using PhyML v3.0 with the model determined by ProTest and 100 bootstrap resamplings (Guindon et al. 2010).

Concatenated Protein Trees

The corrinoid synthesis gene cluster was divided into three biological parts: the siroheme synthesis portion, cobyrinate synthesis portion, and cobalt ABC transporter. To attempt to gather all homologs for each portion of the gene cluster present in the NCBI nr database, each gene was used as a query in a protein Blast search of the nr database. An E-value cutoff of 1E-4 was used, and the number of sequences saved was set well above the number of possible sequences in the database. To acquire a representative sampling of taxa for each organism with multiple hits, the hit with the lowest E-value was retained in the dataset. Therefore, one representative of each species was present in the dataset. Then overlapping taxa were concatenated and aligned using default parameters in MUSCLE v3.8.31 (Edgar 2004). ProtTest3 (Darriba et al. 2011) was used to assess the appropriate model for phylogenetic reconstruction, which was the LG + I + G + F model for each dataset. After visual inspection, trees were reconstructed using Phyml v3.0 with the model as determined by ProtTest and 100 bootstrap resamplings (Guindon et al. 2010). Individual gene trees were tested for compatibility against the concatenated alignments via the approximately unbiased (AU) test as employed in Consel (Shimodaira 2002). Each protein tree was rejected against the concatenated protein tree ($P = <0.01$). This rejection could be due to other HGTs within the tree that do not affect the local topology. Although individual gene trees were not compatible with the alignment, the local bootstrap support for the bipartitions that contain *Thermosipho* sequences increased for the cobyrinate synthesis and siroheme concatenations.

Riboswitch Identification

Riboswitches were identified using the RiboSW server (Chang et al. 2009).

Ancestral State Reconstructions

Three of the four gene clusters were treated as individual character states, and their evolution was traced over a 23S–16S rRNA gene tree. Ordered maximum parsimony and maximum likelihood approaches were carried out in Mesquite v2.75 (Maddison 2011). Parsimony reconstruction used an ordered constraint; the maximum likelihood analysis used the asymmetrical two-parameter Markov k-stat model, and forward (gain) and backward (loss) rates were estimated from the data (Pagel 1999).

Growth Studies

All cultures were grown in glassware that had been soaked in 65%–70% sulfuric acid (v/v) for at least 24 h for removal of trace levels of B_{12} and rinsed with distilled water (Gavin 1957). *Thermotoga* species were grown on cobalamin-free media, DB-B, which was modified from previously used media (Childers et al. 1992; Nanavati et al. 2002). This medium contained the following components per liter: 20 g NaCl, 12 g HEPES, 2.5 g sodium thiosulfate, 2.5 g KCl, 500 mg $MgSO_4 \bullet 7H_2O$, 250 mg NH_4Cl, 330 mg $Na_2WO_4 \bullet 2H_2O$, 50 mg K_2HPO_4, 50 mg $CaCl_2 \bullet 2H_2O$, 20 mg L-arginine, 14 mg DL-tryptophan, 1 mg resazurin, 0.05 mg pyridoxine-HCl, 0.05 mg calcium pantothenate, 0.05 mg nicotinic acid, 0.05 mg ρ-aminobenzoic acid, 0.05 mg riboflavin, 0.05 mg thiamine-HCl, 0.05 mg, 0.05 mg lipoic acid, 0.01 mg biotin, 15 mg nitrilotriacetic acid, 30 mg $MgSO_4 \bullet 7H_2O$, 5 mg $MnSO_4 \bullet H_2O$, 1 mg $FeSO_4 \bullet 7H_2O$, 1 mg $CaCl_2 \bullet 2H_2O$, 0.1 mg $ZnSO_4 \bullet 7H_2O$, 0.1 mg $CuSO_4 \bullet 5H_2O$, 0.1 mg $AlK(SO_4)_2 \bullet 12H_2O$, 0.1 mg H_3BO_3, and 0.1 mg $Na_2MoO_4 \bullet 2H_2O$. Trace amounts of cobalt is present in several media components at levels sufficient for biosynthesis of B_{12}. Media were made anaerobic as described previously (Nanavati et al. 2002). All Thermotogales were grown using maltose as the carbon source at a final concentration of 5 g/l. *Thermosipho* species were grown on DB-B media supplemented with 0.5 g/L vitamin-free casein (MP Biomedicals). The casein was autoclaved in the DB-B media.

All Thermotogales species were passed at least five times in B_{12}-free media, washed 3X times, and resuspended in DB-B (no casein). Media were inoculated with less than 0.2% starter culture, and growth rate and doubling times were calculated using an online calculator (Roth V. 2006, http://www.doubling-time.com/compute.php [cited 2012 Jun 1]).

Lactobacillus Bioassay

Thermotogales B_{12} production was measured using *Lactobacillus delbrueckii subsp. lactis* ATCC 7830 bioassay as described in the *Official Methods of Analysis of the AOAC International* (Feldsine et al. 2002; Taranto et al. 2003; Santos et al. 2009) using B_{12} Assay Medium from BD Diagnostic Systems (Sparks, MD). All reagents used to grow *Lactobacillus*, *Thermotoga*, and *Thermosipho* species were

tested for B_{12} contamination using the *Lactobacillus* bioassay. Thermotogales cell-free extracts were isolated as described (Noll and Barber 1988). Briefly, cells were centrifuged at $16,000 \times g$ for 3 min (Eppendorf benchtop centrifuge), washed three times, resuspended in B_{12}-free 50 mM HEPES (pH 7), incubated at 100°C for 30 min, and centrifuged for 20 min. The resulting supernatant was filtered through a 0.2 µm filter. The extract was stored at −20°C until use.

Transcriptional Analysis

Quantitative real-time polymerase chain reaction (qRT-PCR) was used to measure the expression of genes in the presence or absence of cyanocobalamin (B_{12}). Preparation of RNA, cDNA, and qRT-PCR were performed using commercially available kits as described by the manufacturers. RNA was isolated from cultures in mid-log growth phase using ZR Fungal/Bacterial RNA MiniPrepTM Protocol (Zymo Research, Irvine, CA) and treated with RQ1 RNase-Free DNase (Promega, Madison, WI). RNA concentration was determined using the Qubit® RNA Assay Kit (Invitrogen, Carlsbad, CA) and cDNA was generated using the cDNA GoScript™ Reverse Transcription System (Promega, Madison, WI). Transcriptional levels were determined using (BioRad Hercules, CA) SsoFast EvaGreen Supermix and CFX96 Real-Time PCR Detection System. CFX Manager™ Software (BioRad) was used for transcriptional analyses.

Expression was measured by qRT-PCR in cells grown with and without 92.5 mM cyanocobalamin. Values were normalized relative two to endogenous controls, deoD2 (*Ts. africanus* TCF52B, THA_957; *Ts. africanus* H1760334, H17ap60334_01651; *Tt. maritima*, TM1737), and *eno* (*Ts. africanus* TCF52B, THA_405; *Ts. africanus* H17ap60334_10105; *Tt. maritima*, TM0877). Both endogenous controls have been used previously for *Tt. maritima* qRT-PCR experiments (Nguyen 2004). Averages and standard deviations for transcriptional analyses were from three determinations and two biological replicas using both endogenous controls.

UPLC–MS/MS Instrumentation

Separation was performed on a Waters Acquity ultra-performance liquid chromatography (UPLC) system using an Acquity BEH C18 column (2.0 mm× 50 mm) held at 30°C. The method had a run time of 6 min with a flow rate of 0.6 ml/min. Solvent A contained 95:5 water:methanol. Solvent B contained 100% methanol. The initial ratio of solvents at $t = 0$ was 99.0% A and 1.0% B. From 2–3 min, the levels were adjusted to 45.0% A and 55.0% B. From 3–4 min, the flow was adjusted to 1.0% A and 99.0% B and then held at this percentage until 4.10 min. From 4.10–5 min, the concentrations were adjusted to a final concentration of 99.0% A and 1.0% B. This final concentration was then held constant to 6 min. Samples were maintained at 20°C within the instrument until injection. Injection volumes were 8.5 µl in partial
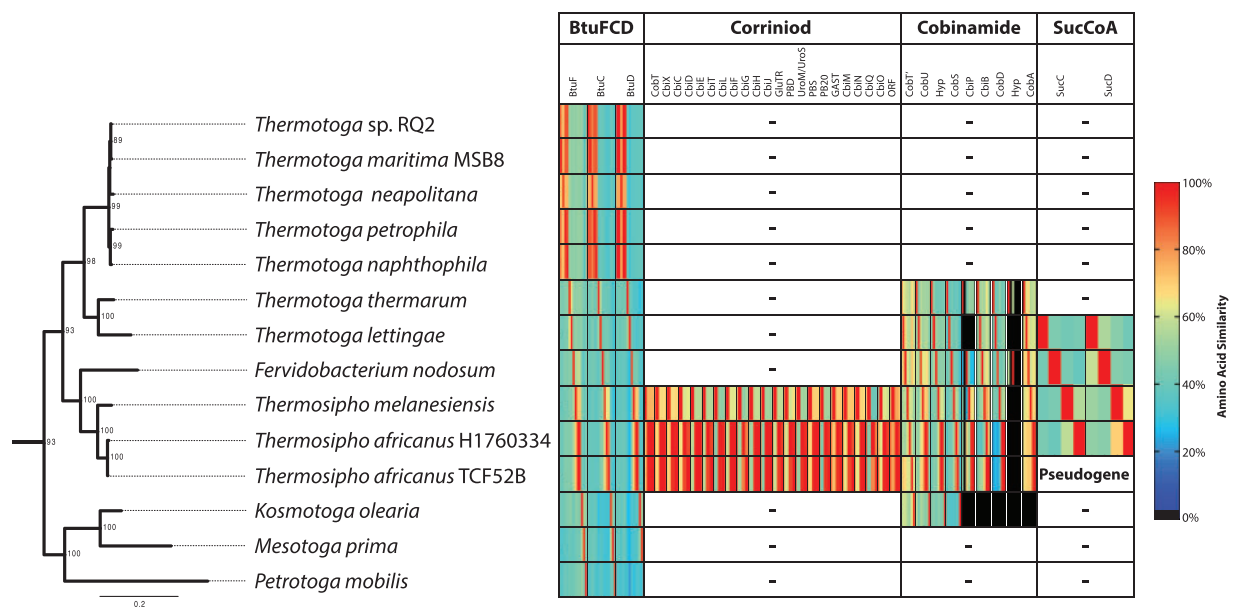
FIG. 1.—Presence/absence of the four gene clusters in the Thermotogales. A maximum likelihood 23S–16S rRNA concatenated gene tree is on left. The presence of a gene cluster is denoted by color, whereas the absence of a gene cluster is denoted by "-" (right). BtuFCD is the B$_{12}$/cobinamide ABC transporter genes, corrinoid is the corrinoid synthesis genes, cobinamide is the cobinamide salvage pathway genes, and SucCoA is the succinyl-CoA synthetase genes. For each gene cluster, the amino acid similarity of each gene is compared with its respective ortholog in that cluster and is colored according to the color bar on the right, adapted from Dagan and Martin 2007. The order in each box (left to right) is the same as in the table (top to bottom). Of the species that contain the cobinamide salvage pathway, the number of genes varies and missing genes within the cluster are colored black; K. olearia contains no archaea-derived genes in the cobinamide salvaging pathway. Bootstrap values ≥70 are shown in the tree.

loop mode. Stock solutions of cyanocobalamin were prepared to 10 μg/ml in solvent A (95:5 water/methanol). UV-Vis detection was carried out at full scan, 254 nm, and 360 nm.

The system was equipped with a triple quadrupole mass spectrometer (MS/MS), photodiode array detector (PDA), and evaporative light scattering detector. The PDA detector was a type UPLC LG 500 nm with a sample rate of 20 points/s, a range of 205–400 nm, and a resolution of 1.2 nm. Ion detection was carried out in selective ion recording mode.

Instrument parameters were set as follows: polarity was APPI+/APcI+. Repeller (kV) was 1.63, 0.00. Corona (μA) was 5.70, 0.83. Corona (kV) was 1.0, 1.1. APPI cone (V) was 58.00, 0.00. APcI cone (V) was 85.00, 87.67. Extractor (V) was 3.00, 3.17. RF (V) was 1.00. LM 1 resolution was 15.70 and HM 1 Resolution was 15.00. Ion energy 1 was 0.80. MS mode entrance was 20.00. MS mode collision energy was 3.00. MS mode exit was 20.00. MS/MS mode collision energy was 1.00. MS/MS mode exit was 20.05. LM 2 resolution was 16.00. HM 2 resolution was 15.70. Ion energy 2 was 3.00 and Gain was 1.0. Multiplier was set to −637.96.

Results and Discussion

Distribution and Phylogenetic Analysis of Putative B$_{12}$ Riboswitch-Regulated Genes in the Thermotogales

Putative B$_{12}$ riboswitches were identified near four gene clusters linked to B$_{12}$ metabolism. The phylogenetic distribution of

these clusters within the Thermotogales is depicted in figure 1. The evolutionary histories of these clusters were traced by ancestral state reconstruction using maximum parsimony and maximum likelihood (fig. 2) models. These analyses suggest the cobinamide salvage pathway was present in the ancestor of the Thermotogales and was lost in the lineages leading to the *Thermotoga* (excluding *Tt. lettingae* and *Tt. thermarum*), *Petrotoga mobilis,* and *Mesotoga prima* (Nesbø et al. 2012). The reconstructions also suggest that the corrinoid synthesis gene cluster was gained on the branch leading to the *Thermosipho* lineage. Both reconstructions support the SucCoA synthetase gene cluster as either being acquired twice, once at the base of the *Thermosipho/Fervidobacterium* clade and once in *Tt. lettingae,* or perhaps gained once by one of the clades and then transferred to the other.

Cobinamide Salvage Gene Cluster

The cobinamide salvage gene cluster was likely the ancestral state of B$_{12}$ biosynthesis for the Thermotogales, and the gene cluster is composed of both bacteria- and archaea-derived proteins. The gene cluster is present in the *Thermosipho* species, *Fervidobacterium nodosum, Kosmotoga olearia,* and *Tt. lettingae* and *Tt. thermarum* genomes, but their gene content is different (supplementary fig. S1, Supplementary Material online). The *Tt. thermarum* cluster is similar to that of *F. nodosum* except there have been rearrangements and an
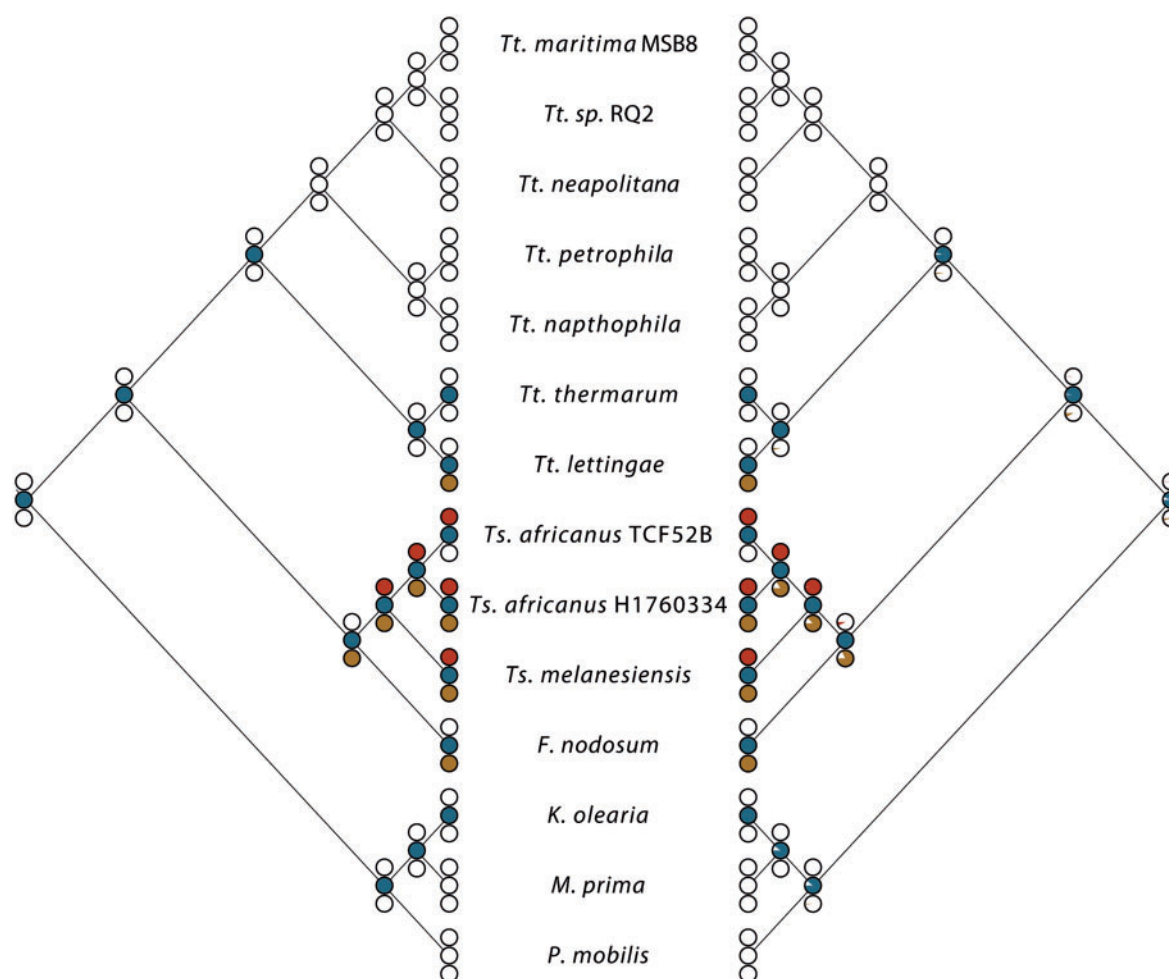
Fig. 2.—Ancestral state reconstructions of three gene clusters based on parsimony (left cladogram) and maximum likelihood (ML) (right cladogram). Colored circles represent the three gene clusters. From top to bottom, red is the corrinoid synthesis gene cluster, blue is the cobinamide salvage gene cluster, and brown is the SucCoA synthetase gene cluster. Likelihood ratio of presence to absence is denoted by color to white on the ML cladogram. Because SucCoA is present as a pseudogene, it is considered absent in this analysis.

inversion within the cluster as denoted by the lines. Only its *cobT'-cobU-hyp-cobS* region appears to be under riboswitch regulation. The *Tt. lettingae* cluster is similar to the *Thermosipho* species gene clusters except it is missing the *cbiP* gene. *K. olearia* appears to only contain *cobT'-cobU-hyp-cobS* and has no riboswitch, suggesting it is not under B_{12} regulation.

The top scoring Blast (TSB) hits for all the genes in this operon are to either bacteria or Archaea genes (supplementary table S1, Supplementary Material online), and phylogenetic reconstruction places the *cbiP*, *cbiB*, and *cobD* genes as sister groups to genes of members of the Thermococcales (supplementary fig. S2, Supplementary Material online). Furthermore, the *cbiB* and *cobD* genes are also found adjacent to each other in Thermococcales genomes. Taken together, these analyses suggest the cobinamide salvage pathway consists of bacteria- and archaea-derived proteins (fig. 3).

Corrinoid Synthesis Gene Cluster

The corrinoid synthesis gene cluster was likely gained in the -ancestor of the *Thermosipho* genus (fig. 2), from Firmicutes. The *Thermosipho* lineage is the only lineage within the Thermotogales that has this gene cluster, and 20 of its 21 genes show synteny with the clusters in *Thermoanaerobacter* genomes (supplementary fig. S3, Supplementary Material online). TSB hits for each of the proteins encoded in this cluster were from Firmicutes. Phylogenetic analyses of individual genes within this gene cluster led to poorly resolved gene trees, suggesting a lack of phylogenetic signal for individual genes. This is consistent with a previous report by Zhaxybayeva et al. (2009), where they showed that when Firmicutes are the closest phylogenetic neighbor to the Thermotogales, the phylogenetic signal is often insufficient to accurately determine the phylogenetic neighbor within the Firmicutes. One way to overcome weak phylogenetic
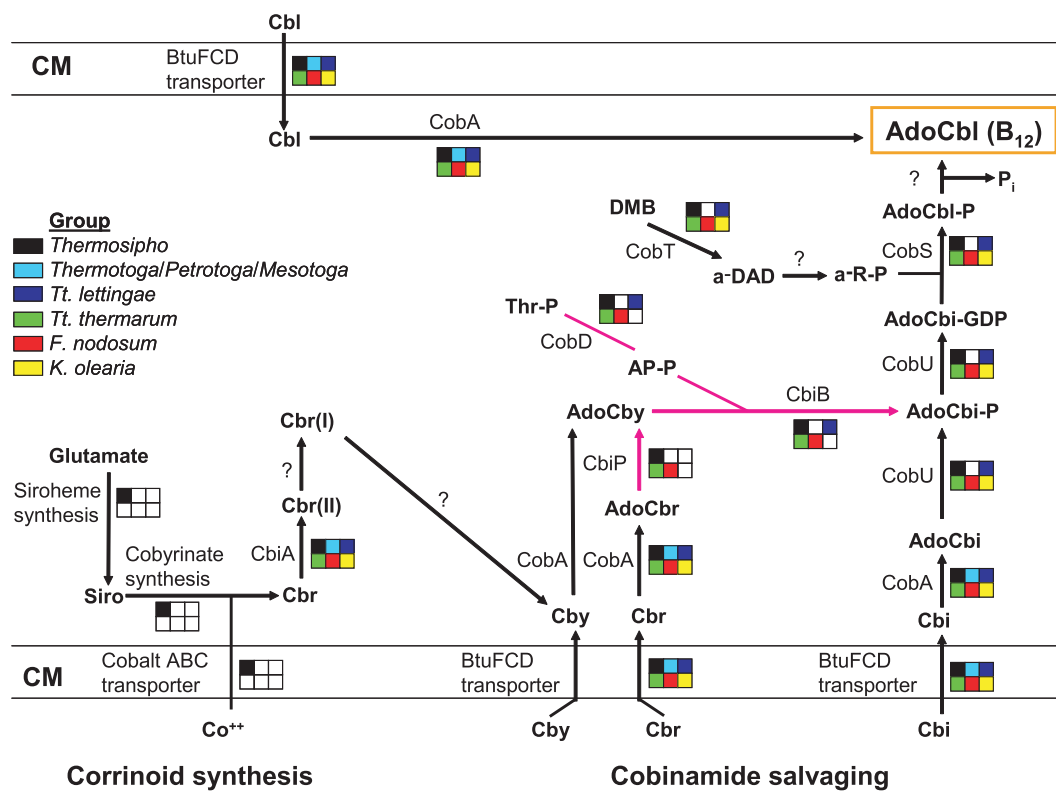
Fig. 3.—Proposed corrinoid synthesis and cobinamide salvage pathways in the Thermotogales. The color box indicates that the Thermotogales group contains a protein homolog, whereas an empty box indicates no protein homolog. Thermotogales groups: *Thermosipho* group (*Ts. melanesiensis*, *Ts. africanus* TCF52B and H1760334, black); *Thermotoga/P. mobilis/M. prima* (*Tt. maritima*, *Tt.* RQ2, *Tt.* cell2, *Tt. naphthophila*, *Tt. neapolitana*, *Tt. petrophila*, *P. mobilis*, *M. prima*, light blue), *Tt. lettingae* (dark blue), *Tt. thermarum* (light green), *F. nodosum* (red), and *K. olearia* (yellow). Reactions using archaea-derived proteins are indicated by pink arrows. In *Thermosipho* species, the B$_{12}$ de novo pathway is broken into two gene clusters that encode corrinoid synthesis and cobinamide salvage. Corrinoid synthesis consists of three parts: cobyrinate synthesis (from sirohydrochlorin to cobyrinate), siroheme synthesis (from glutamate to siroheme/sirohydrochlorin), and a cobalt ABC transporter for importing cobalt. The final steps in B$_{12}$ synthesis can also be used to salvage cobinamides. CM, cytoplasmic membrane; Cbi, cobinamide; AdoCbi, adenosylcobinamide; AdoCbi-P, adenosylcobinamide phosphate; AdoCbi-GDP, adenosylcobinamide guanosine diphosphate; Cbl, cobalamin; AdoCbl, adenosylcobalamin; Cby, cobyric acid; AdoCbr, adenosylcobyrinic acid; AdoCby, adenosylcobyric acid; Thr-P, threonine phosphate; AP-P, aminopropanol phosphate; a-R-P, α-ribazole; DMB, 5,6-dimethylbenzimidazole; a-DAD, α-5- ,6-dimethylbenzimidazole adenine dinucleotide; Cbr, cobyrinic acid; Cbr(II), cob(II)yrinate a,c-diamide; Cbr(I), cob(I)yrinate a,c-diamide; Siro, sirohydrochlorin. Proteins that have been identified in other microbes but have no homologs in the Thermotogales genomes are indicated by question marks. The reaction pathways were predicted based on evidence from other microbes (Escalante-Semerena 2007; Gray et al. 2008; Raux et al. 2000).

signals is to concatenate individual protein sequences (Delsuc et al. 2005; Mignard and Flandrois 2008; Sassera et al. 2011); therefore, concatenated protein trees of the three different portions of the corrinoid gene cluster (cobyrinate synthesis, siroheme synthesis, and cobalt ABC transporter) were constructed. The bootstrap supports for the cobyrinate and siroheme synthesis trees improved for the bipartitions that contain the *Thermosipho* species. The concatenation results suggest cobyrinate and siroheme synthesis genes group within the Firmicutes (fig. 4 and supplementary fig. S4, Supplementary Material online). The concatenation of the genes encoding cobalt ABC transporter did not lead to an increase of phylogenetic signal. The unique presence of the corrinoid cluster to the *Thermosipho* lineage within the Thermotogales and ancestral state reconstruction suggests that this cluster was

acquired horizontally. The close synteny of the corrinoid gene cluster with the *Thermoanaerobacter* species genomes (supplementary fig. S3, Supplementary Material online), the TSB hits to the Firmicutes, and the concatenated gene trees (fig. 4) together suggest transfers of the genes in this cluster from the Firmicutes to the *Thermosipho* lineage.

Two Other Gene Clusters under Putative Vitamin B$_{12}$ Riboswitch Regulation

Two other gene clusters may also be under the control of B$_{12}$ riboswitches, the B$_{12}$/cobinamide BtuFCD ABC transporter, and SucCoA synthetase. Due to weak phylogenetic signals, the origins of these clusters could not be determined. SucCoA synthetase is only present in a few Thermotogales genomes
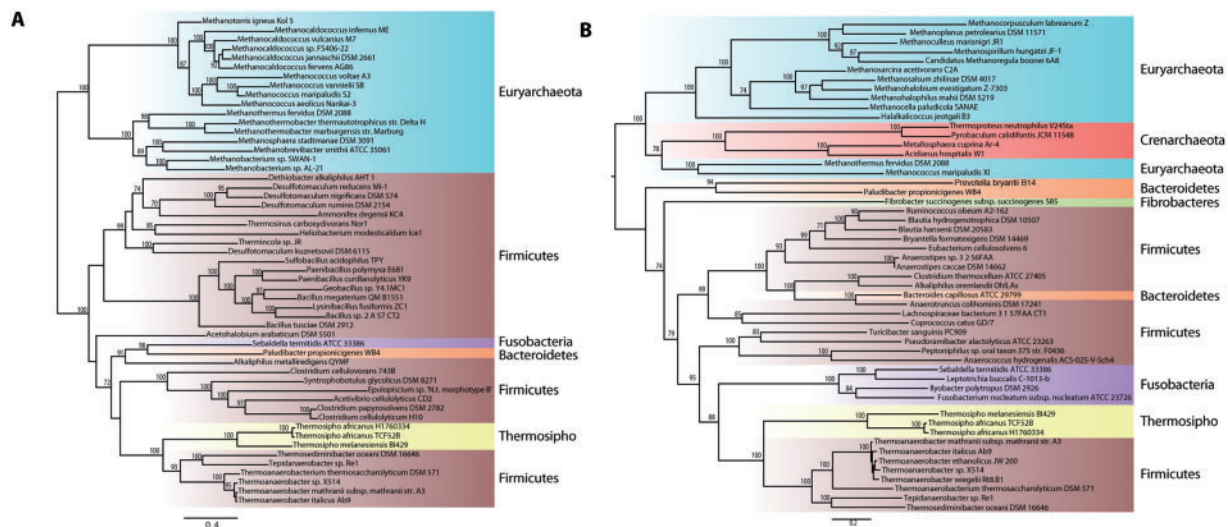
Fig. 4.—Concatenated maximum likelihood trees of proteins encoded by two different portions of the corrinoid gene cluster. (A) The cobyrinate synthesis portion. (B) A subtree of the siroheme synthesis portion (for full tree, see supplementary fig. S4, Supplementary Material online). Bootstrap values ≥70 are shown.

(fig. 1). The riboswitch in *Tt. lettingae* overlaps with the SucCoA synthetase β-subunit gene by 32 base pairs, which suggests that this gene cluster is regulated at the translational level rather than the transcriptional level (Rodionov et al. 2003) as for the other gene clusters under B_{12} riboswitch control in the Thermotogales. The SucCoA synthetase β-subunit appears to have acquired a premature stop codon in *Ts. africanus* TCF52B rendering it a pseudogene. One way of compensating for the lack of this enzyme would be through the B_{12}-dependent methylmalonyl-CoA mutase, which catalyzes the isomerization of methylmalonyl-CoA to succinyl-CoA. *Ts. africanus* TCF52B does indeed encode such a protein (THA_155, THA_1169). These genes have a patchy distribution among the Thermotogales genomes.

Expression of SucCoA synthetase and ABC transporter genes was tested in cultures of *Ts. africanus* strains H1760334 and TCF52B and in *Tt. maritima,* each grown with and without B_{12} (table 1). It has been shown in other studies (Santillan and Mackey 2005) that genes regulated at the transcriptional level by B_{12} riboswtiches are down-regulated in the presence of high amounts of B_{12}. In *Ts. africanus* H1760334, the SucCoA synthetase genes appear to be under B_{12} riboswitch regulation (table 1). However, in *Ts. africanus* TCF52B, no expression was detected, suggesting that the pseudogene also affects transcription or message stability (supplementary fig. S5, Supplementary Material online). Interestingly, the α- and β-subunit genes were amplified from the *Ts. africanus* type strain (strain Ob7[T]), and its gene carries even more deletions (supplementary fig. S5, Supplementary Material online).

The B_{12} ABC transporter genes appear to be under B_{12} riboswitch transcriptional regulation in *Tt. maritima* and *Ts. africanus* TCF52B (table 1). There was no change in

Table 1

Differential Expression of the B_{12}-Related Genes of *Ts. africanus* H1760334, *Ts. africanus* TCF52B, and *Tt. maritima* in Response to Cobalamin Availability

Function	Gene	Decreased Fold Expression		
		Ts. africanus H1760334	*Ts. africanus* TCF52B	*Tt. maritima*
Corrinoid synthesis	*cobT*	8.9 ± 1.6	30.9 ± 3.7	N/A
	cbiD	10.6 ± 4.0	23.9 ± 5.5	
	cbiJ	10.6 ± 2.6	14.1 ± 2.1	
	gluTR	14.5 ± 2.3	23.0 ± 2.3	
	pR20	15.8 ± 2.4	18.4 ± 4.3	
	cbiM	14.2 ± 4.1	13.7 ± 1.9	
	cbiO	12.3 ± 3.1	11.7 ± 2.6	
	ORF[a]	13.9 ± 2.5	10.8 ± 1.5	
Cobinamide salvaging	*cobT′*	3.8 ± 1.4	11.7 ± 1.5	N/A
	cbiP	4.5 ± 1.3	7.6 ± 1.1	
Succinyl-CoA synthase	*sucC*	4.1 ± 0.6	Not detected	N/A
	sucD	9.8 ± 1.8	Not detected	
B_{12}-ABC transporter	*btuF*	No change	4.4 ± 0.6	9.0 ± 1.8

NOTE.—Expression was measured by qRT-PCR in cells grown with and without 92.5 mM cyanocobalamin. Values were normalized relative to *deoD2* (*Ts. africanus* TCF52B, THA_957; *Ts. africanus* H1760334, H17ap60334_01651; *Tt. maritima*, TM1737), and *eno* (*Ts. africanus* TCF52B, THA_405; *Ts. africanus* H17ap60334_10105; *Tt. maritima*, TM0877). Values are averages of three determinations from two biological replicas and standard deviations are shown.

N/A, not applicable (contains no homologous gene).

[a]ORF, the gene adjacent to *cbiO* in the corrinoid synthesis gene cluster [*Ts. africanus* TCF52B, THA_1086; *Ts. africanus* H1760334, H17ap60334_02301].

expression observed for the ABC transporter in the H1760334 strain. A putative transposon is inserted between the riboswitch and the first transporter gene, *btuF*, of strain H1760334. This would disrupt the regulation of the transporter leaving these genes unaffected by the addition of B_{12};

thereby, explaining the differences in expression of this cluster in the two *Ts. africanus* strains.

Experimental Proof that *Ts. africanus* Strains Produce Vitamin B$_{12}$ and Support for the Riboswitch Model of Regulation

The *Thermosipho* species are the only Thermotogales that have both a corrinoid synthesis and a cobinamide salvage gene cluster. These two gene clusters together form a nearly complete B$_{12}$ de novo pathway, suggesting that they may allow de novo synthesis of B$_{12}$ (fig. 3). Thus, *Ts. africanus* strains H1760334 and TCF52B were tested for B$_{12}$ production using a *Lactobacillus* bioassay of cell extracts. *Tt. maritima* cell-free extract was tested as a negative control. *Ts. africanus* strain H1760334 produced 186 ± 14 ng/g cell paste B$_{12}$ (226 ± 20 ng/L culture), and strain TCF52B produced 112 ± 20 ng/g cell paste B$_{12}$ (153 ± 34 ng/L culture). *Tt. maritima* had no detectable B$_{12}$ when assayed using 12.5 times the weight of cell paste used for measures with *Ts. africanus* H1760334. *Ts. africanus* H1760334 was also shown to produce B$_{12}$ using UPLC and mass spectrometric analysis (supplementary fig. S6, Supplementary Material online). The UPLC method showed approximately 5-fold more B$_{12}$ in these cells, 938 ng/g cell paste B$_{12}$ (1,820 ng/l culture). Similar differences between bioassay and analytical methods have been reported previously (Taranto et al. 2003).

The genes that encode de novo synthesis (fig. 3) of B$_{12}$ that are thought to be subject to B$_{12}$ riboswitch regulation were tested for expression when cells were grown with and without B$_{12}$. In the presence of B$_{12}$, the corrinoid synthesis and cobinamide salvage genes are down-regulated (table 1). This is consistent with findings in other organisms (Santillan and Mackey 2005; Cheng et al. 2011) and suggests that these genes allow *Thermosipho* species to produce B$_{12}$ and are regulated at the transcriptional level. An open reading frame (ORF) adjacent to *cbiO* in the corrinoid synthesis island was also tested, and its expression was down-regulated along with the other genes (table 1). Although the function of this gene is unknown, it has a predicted ATP-binding domain and has homologs present within the cobinamide salvage island of *F. nodosum* (supplementary fig. S1, Supplementary Material online). The ORF is also found within gene clusters related to B$_{12}$ synthesis in other genomes (e.g., *Thermodesulfovibrio yellowstonii* and *Archaeoglobus fulgidus*). This suggests that this ORF is related to B$_{12}$ synthesis/cobinamide salvage.

An Evolutionary Model of Vitamin B$_{12}$ Synthesis and Usage in the Thermotogales

The ancestral state reconstruction (fig. 2) suggests the cobinamide salvage pathway was the ancestral state of B$_{12}$ biosynthesis for the Thermotogales, whereas the corrinoid synthesis pathway was acquired later in the *Thermosipho* lineage. The lower temperature Thermotogales, *K. olearia*

(optimum growth temperature [OGT] 65°C, grows as low as 20°C) (Dipippo et al. 2009), *P. mobilis* (OGT = 60°C, grows as low as 40°C) (Lien et al. 1998), and *M. prima* (OGT = 37°C, grows as low as 20°C) (Nesbø et al. 2012) contain fewer or no cobinamide salvage pathway genes (fig. 1 and supplementary fig. S1, Supplementary Material online). The loss of these genes may indicate the availability of B$_{12}$ rather than cobinamides in their environments.

The *Thermotoga* clade (excluding *Tt. lettingae* and *Tt. thermarum*) contains both B$_{12}$-independent and -dependent forms of methionine synthase and ribonucleotide reductase (supplementary table S2, Supplementary Material online). The loss of the cobinamide salvage genes in the *Thermotoga* clade may be due to the use of B$_{12}$-independent enzymes by these organisms. Methionine synthase (Huang et al. 2007) and ribonucleotide reductase (Jordan et al. 1997) are the only B$_{12}$-dependent enzymes in the *Thermotoga* clade, whereas the other clades have multiple B$_{12}$-binding proteins suggesting a possible need for B$_{12}$ by those species (supplementary table S2, Supplementary Material online). With few or no B$_{12}$-binding domain proteins, selection pressure to maintain a B$_{12}$ biosynthesis pathway should diminish. This relaxed selection pressure may have allowed for loss of the pathway. We have successfully grown *Thermotoga* species (*Tt. maritima*, *Tt. sp.* RQ2, *Tt. sp.* cell2, *Tt. naphthophila*, *Tt. neapolitana*, and *Tt. petrophila*) without the addition of B$_{12}$ on B$_{12}$-free media, showing B$_{12}$ is not essential for growth. These media were determined to be B$_{12}$ free using the *Lactobacillus* bioassay (Feldsine et al. 2002; Taranto et al. 2003; Santos et al. 2009). Both B$_{12}$-dependent methionine synthase (Huang et al. 2007) and ribonucleotide reductase (Jordan et al. 1997) from *T. maritima* have been previously shown in vitro to require B$_{12}$ as a cofactor for activity, and the expression data of *btuF* (BtuFCD transporter) suggests that B$_{12}$ is imported into the cell (table 1). This suggests that B$_{12}$ is not essential but can be utilized in this clade.

Conclusions

Here, we provide a comparison on the order level of the evolutionary history of genes related to biosynthesis, usage, and regulation of a vitamin. This comparison has only recently been made possible by the large number of sequenced genomes spanning several genera of the Thermotogales. The corrinoid synthesis pathway of *Thermosipho* species appears to have been acquired from the Firmicutes lineage, which is the lineage that the Thermotogales have been shown to exchange the most genes horizontally. The salvage pathway varies within the Thermotogales with some species containing both bacteria- and archaea-derived proteins. We have shown experimentally, by two separate methods, that *Ts. africanus* produces vitamin B$_{12}$ using genes that were likely acquired from the Firmicutes. We provide a model where the cobinamide salvage pathway was present in the ancestor of the

Thermotogales and lost in various lineages and that the corrinoid synthesis pathway was a later addition to the ancestor of the *Thermosipho* genus.

Supplementary Material

Supplementary figures S1–S6 and Supplementary tables S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Achenbach-Richter L, Gupta R, Stetter KO, Woese CR. 1987. Were the original eubacteria thermophiles? Syst Appl Microbiol. 9: 34–39.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 9:75.

Balows A. 1992. The prokaryotes: a handbook on the biology of bacteria: ecophysiology, isolation, identification, applications. Berlin: Springer-Verlag.

Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. Proc Natl Acad Sci U S A. 102:14332–14337.

Caspi R, et al. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 40:D742–D753.

Chang TH, et al. 2009. Computational identification of riboswitches based on RNA conserved functional sequences and conformations. RNA 15: 1426–1430.

Cheng J, Poduska B, Morton RA, Finan TM. 2011. An ABC-type cobalt transport system is essential for growth of *Sinorhizobium meliloti* at trace metal concentrations. J Bacteriol. 193:4405–4416.

Childers SE, Vargas M, Noll KM. 1992. Improved methods for cultivation of the extremely thermophilic bacterium *Thermotoga neapolitana*. Appl Environ Microbiol. 58:3949–3953.

Dagan T, Martin W. 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. Proc Natl Acad Sci U S A. 104:870–875.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet. 6:361–375.

Dipippo JL, et al. 2009. *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. Int J Syst Evol Microbiol. 59:2991–3000.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Escalante-Semerena JC. 2007. Conversion of cobinamide into adenosylcobamide in bacteria and archaea. J Bacteriol. 189:4555–4560.

Feldsine P, Abeyta C, Andrews WH. 2002. AOAC International methods committee guidelines for validation of qualitative and quantitative food microbiological official methods of analysis. J AOAC Int. 85: 1187–1200.

Fournier GP, Gogarten JP. 2010. Rooting the ribosomal tree of life. Mol Biol Evol. 27:1792–1801.

Gavin JJ. 1957. Microbiological process report: analytical microbiology. III. Turbidimetric methods. Appl Microbiol. 5:235–243.

Gray MJ, Tavares NK, Escalante-Semerena JC. 2008. The genome of *Rhodobacter sphaeroides* strain 2.4.1 encodes functional cobinamide salvaging systems of archaeal and bacterial origins. Mol Microbiol. 70: 824–836.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59:307–321.

Huang S, et al. 2007. Reactivation of methionine synthase from *Thermotoga maritima* (TM0268) requires the downstream gene product TM0269. Protein Sci. 16:1588–1595.

Jordan A, et al. 1997. B12-dependent ribonucleotide reductases from deeply rooted eubacteria are structurally related to the aerobic enzyme from *Escherichia coli*. Proc Natl Acad Sci U S A. 94: 13487–13492.

Lien T, Madsen M, Rainey FA, Birkeland NK. 1998. *Petrotoga mobilis* sp. nov., from a North Sea oil-production well. Int J Syst Bacteriol. 48(Pt 3):1007–1013.

Maddison, WPaDRM. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75. [cited 2012 Jun 1]. Available from: http://mesquiteproject.org.

Marchler-Bauer A, et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39: D225–D229.

Martens JH, Barg H, Warren MJ, Jahn D. 2002. Microbial production of vitamin B12. Appl Microbiol Biotechnol. 58:275–285.

Mignard S, Flandrois JP. 2008. A seven-gene, multilocus, genus-wide approach to the phylogeny of mycobacteria using supertrees. Int J Syst Evol Microbiol. 58:1432–1441.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–W185.

Nanavati D, Noll KM, Romano AH. 2002. Periplasmic maltose- and glucose-binding protein activities in cell-free extracts of *Thermotoga maritima*. Microbiology 148:3531–3537.

Nelson KE, et al. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature 399:323–329.

Nesbø CL, et al. 2009. The genome of *Thermosipho africanus* TCF52B: lateral genetic connections to the Firmicutes and Archaea. J Bacteriol. 191:1974–1978.

Nesbø CL, et al. 2012. *Mesotoga prima* gen. nov., sp. nov., the first described mesophilic species of the Thermotogales. Extremophiles 16:387–393.

Nguyen TN. 2004. Whole genome transcription profiling of Thermotoga maritima in response to growth on glucose, lactose and maltose. [dissertation]. [Storrs (CT)]: University of Connecticut.

Noll KM, Barber TS. 1988. Vitamin contents of archaebacteria. J Bacteriol. 170:4315–4321.

Overbeek R, et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33:5691–5702.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Raux E, Schubert HL, Warren MJ. 2000. Biosynthesis of cobalamin (vitamin B12): a bacterial conundrum. Cell Mol Life Sci. 57:1880–1893.

Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2003. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. J Biol Chem. 278:41148–41159.

Santillan M, Mackey MC. 2005. Dynamic behaviour of the B12 riboswitch. Phys Biol. 2:29–35.

Santos F, et al. 2009. Effect of amino acid availability on vitamin B12 production in *Lactobacillus reuteri*. Appl Environ Microbiol. 75:3930–3936.

Sassera D, et al. 2011. Phylogenomic evidence for the presence of a flagellum and cbb(3) oxidase in the free-living mitochondrial ancestor. Mol Biol Evol. 28:3285–3296.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51:492–508.

Swithers KS, et al. 2011a. Genome sequence of *Kosmotoga olearia* strain TBF 19.5.1, a thermophilic bacterium with a wide growth temperature range, isolated from the Troll B oil platform in the North Sea. J Bacteriol. 193:5566–5567.

Swithers KS, et al. 2011b. Genome sequence of *Thermotoga* sp. strain RQ2, a hyperthermophilic bacterium isolated from a geothermally heated region of the seafloor near Ribeira Quente, the Azores. J Bacteriol. 193:5869–5870.

Taranto MP, Vera JL, Hugenholtz J, De Valdez GF, Sesma F. 2003. *Lactobacillus reuteri* CRL1098 produces cobalamin. J Bacteriol. 185:5643–5647.

Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. 2003. Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. RNA 9:1084–1097.

Warren MJ, Raux E, Schubert HL, Escalante-Semerena JC. 2002. The biosynthesis of adenosylcobalamin (vitamin B12). Nat Prod Rep. 19:390–412.

Williams D, et al. 2011. A rooted net of life. Biol Direct. 6:45.

Woodson JD, Reynolds AA, Escalante-Semerena JC. 2005. ABC transporter for corrinoids in *Halobacterium* sp. strain NRC-1. J Bacteriol. 187:5901–5909.

Zhang Y, Rodionov DA, Gelfand MS, Gladyshev VN. 2009. Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization. BMC Genomics. 10:78.

Zhaxybayeva O, et al. 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc Natl Acad Sci U S A. 106:5865–5870.

Associate editor: Bill Martin