Primordial-like enzymes from bacteria with reduced genomes

Matteo P. Ferla,1† Jodi L. Brewster,1 Kelsi R. Hall,1 Gary B. Evans2 and Wayne M. Patrick1*

1Department of Biochemistry, University of Otago, Dunedin, New Zealand.
2Ferrier Research Institute, Victoria University, Lower Hutt, New Zealand.

Summary
The first cells probably possessed rudimentary metabolic networks, built using a handful of multifunctional enzymes. The promiscuous activities of modern enzymes are often assumed to be relics of this primordial era; however, by definition these activities are no longer physiological. There are many fewer examples of enzymes using a single active site to catalyze multiple physiologically-relevant reactions. Previously, we characterized the promiscuous alanine racemase (ALR) activity of Escherichia coli cystathionine β-lyase (CBL). Now we have discovered that several bacteria with reduced genomes lack alr, but contain metC (encoding CBL). We characterized the CBL enzymes from three of these: Pelagibacter ubique, the Wolbachia endosymbiont of Drosophila melanogaster (wMel) and Thermotoga maritima. Each is a multifunctional CBL/ALR. However, we also show that CBL activity is no longer required in these bacteria. Instead, the wMel and T. maritima enzymes are physiologically bi-functional alanine/glutamate racemases. They are not highly active, but they are clearly sufficient. Given the abundance of the microorganisms using them, we suggest that much of the planet’s biochemistry is carried out by enzymes that are quite different from the highly-active exemplars usually found in textbooks. Instead, primordial-like enzymes may be an essential part of the adaptive strategy associated with streamlining.

Introduction
Over 40 years ago, Yčas and Jensen independently proposed similar scenarios for the evolution of metabolic pathways (Yčas, 1974; Jensen, 1976). They envisaged primordial organisms with small genomes and, therefore, small numbers of protein-encoding genes. They proposed that the only way such organisms could carry out the metabolic biochemistry required for life would be if many of the proteins were multifunctional. Each author reached the conclusion that primordial enzymes were likely to have exhibited broad substrate specificities and/or to have catalyzed classes of related reactions. From this starting point, gene duplication and divergence could give rise to the larger genomes and highly specific enzymes that epitomize modern organisms.

This model has gained considerable support from the finding that many, probably most, and perhaps all modern enzymes are promiscuous (Khersonsky and Tawfik, 2010). That is, modern enzymes have the ability (often weak) to catalyze reactions or to act on substrates that are different from those required physiologically. While the multiple activities of primordial enzymes were all assumed to be required for the viability of primordial cells, the promiscuous activities of modern enzymes are, by definition, not physiologically relevant (Copley, 2015). Nevertheless, these promiscuous activities are often assumed to be relics of their multifunctional ancestors, and the homologous members of enzyme superfamilies often share overlapping promiscuous activities (Glasner et al., 2006; Khersonsky and Tawfik, 2010; Copley, 2015). Beginning 16 years ago (Matsumura and Ellington, 2001), many practitioners of directed evolution have also observed ancestor-like intermediates, with broadened substrate specificities, on mutational trajectories that traverse enzymatic functions. More recently, phylogenetic approaches have been used to resurrect ancestral enzymes with broad specificities (Voordeckers et al., 2012; Risso et al., 2013). These approaches have provided enormous insight into enzyme evolution. However, they also suffer from either studying an activity that is no longer physiologically relevant (i.e., promiscuity), or studying a single enzyme outside of its metabolic context.
Our long-term goal is to assess the viability of the Yčas/Jensen model of primordial metabolism. Can a cell function with only a minimal set of multifunctional enzymes, and what would this set comprise? What are the structural, functional, dynamic and regulatory characteristics of the enzymes in such a cell? And can we recapitulate the evolutionary trajectory from this cell, to something that more closely resembles modern microbial life? As a first step toward answering these questions, here we have sought to discover primordial-like, multifunctional enzymes from extant bacteria. The defining characteristic of these enzymes is that a single active site is responsible for catalyzing two or more reactions, each of which are required for the growth of the organism. Further, the evolutionary model of Yčas and Jensen predicts that the multiple activities of a primordial-like enzyme are likely to be carried out by separate, specialized enzymes in most modern-day organisms.

Only a small number of primordial-like enzymes have been reported to date. Perhaps the best characterized is PriA, a bi-functional isomerase originally found in *Mycobacterium tuberculosis* and *Streptomyces coeli-color*, which catalyzes the reactions carried out by two separate enzymes (HisA and TrpF) in most bacteria (Barona-Gómez and Hodgson, 2003; Due et al., 2011). An example from archaeal and deep-branching bacterial lineages is the bi-functional fructose 1,6-bisphosphate aldolase/phosphatase, which remodels its active site to catalyze two consecutive steps in gluconeogenesis (Say and Fuchs, 2010; Du et al., 2011). The TrpF enzyme from *Chlamydia trachomatis* has also been shown to play a second physiological role in folate biosynthesis (Adams et al., 2014).

Recently, we showed that the cystathionine β-lyase (CBL) from *Escherichia coli* has promiscuous alanine racemase (ALR) activity (Soo et al., 2016). CBL is encoded by the *metC* gene and catalyzes the penultimate step in methionine biosynthesis, in which β-elimination of cystathionine yields homocysteine (Fig. 1A). ALR catalyzes the interconversion of L-alanine and D-alanine (Fig. 1B), with the latter being required for peptidoglycan biosynthesis. While both enzymes utilize the cofactor pyridoxal 5'-phosphate (PLP), they do not share any sequence or structural similarities. Of the seven protein folds that are known to bind PLP (Raboni et al., 2010), CBL adopts fold type I, while ALR adopts fold type III (Fig. 1). A parsimonious explanation of our previous results was that modern CBL enzymes are descended from a bi-functional CBL/ALR ancestor, but that the alanine racemase function has been rendered vestigial in lineages (such as the one leading to *E. coli*) that gained a non-homologous ALR specialist.

Fig. 1. Structure and function of CBL and ALR.
A. The CBL tetramer (PDB entry 1CL1; Clausen et al., 1996), which adopts fold type I. The β-elimination reaction catalyzed by CBL in methionine biosynthesis is shown below the structure.
B. The fold type III ALR dimer (PDB entry 2RJG; Wu et al., 2008) and the reaction it catalyzes to provide D-alanine for peptidoglycan biosynthesis. In each structure, the PLP cofactors are shown in space-filling format. Note: this figure was originally published in the *Journal of Biological Chemistry*. Soo, V.W.C, Yosaatmadja, Y., Squire, C.J., and Patrick, W.M. (2016) Mechanistic and evolutionary insights from the reciprocal promiscuity of two pyridoxal phosphate-dependent enzymes. *J Biol Chem* **291**: 19873–19887. © The American Society for Biochemistry and Molecular Biology.

Here, we have asked whether any contemporary bacteria retain multifunctional, primordial-like CBL enzymes. A search of >1000 sequenced genomes revealed a handful of species in which CBL was present, but ALR...
was absent. We have characterized the CBL enzymes from three of these species: *Pelagibacter ubique*, the *Wolbachia* endosymbiont of *Drosophila melanogaster* and *Thermotoga maritima*. In doing so, we have also shed new light on the pathways of methionine biosynthesis in these organisms, particularly *T. maritima*. Together, our results suggest that free-living bacteria with reduced genomes are the best models for studying primordial enzymology and metabolism.

**Results**

*The search for multifunctional CBL/ALR enzymes*

We began by searching for bacterial species with a requirement for peptidoglycan, a *metC* gene, but with no *alr* gene. We reasoned that these criteria would maximize the likelihood of identifying a *metC*-encoded enzyme acting either as the physiological alanine racemase, or as both CBL and ALR *in vivo*.

The annotated lists of proteins (feature tables) for each bacterial genome in the NCBI database were downloaded and parsed for the presence of *alr* and *metC* homologues, based on annotations in the Clusters of Orthologous Groups (COG) database (Galperin et al., 2015). While we were particularly interested in *metC* genes, it is difficult to distinguish *metC* from *metB* (encoding cystathionine γ-synthase, CGS) by sequence alone (Ferla and Patrick, 2014). As a result, these two genes are grouped into a single COG. At the time of our survey (August 2010), there were 1023 fully sequenced and annotated genomes to analyze. The presence or absence of *alr* (COG0787) and *metC/metB* (COG0626) in each genome was visualized on a tree that was based on taxonomy, in which the branches of taxa sharing the same pattern were collapsed (Fig. 2). Only six taxa (colored yellow in Fig. 2) met our criteria of possessing *metC/metB* but not *alr*. Of these, the Planctomycetes were thought to lack peptidoglycan (Fuerst and Sagulenko, 2011), although this view has recently been updated (Jeske et al., 2015; van Teeseling et al., 2015). *Thermomicrobium roseum* possesses a thin peptidoglycan layer with unusual features (Wu et al., 2009b), and the peptidoglycan status of the endosymbiotic gammaproteobacteria that lack *alr* but contain *metC/metB* is unclear.

We narrowed our search to the three remaining taxa: the alphaproteobacterial orders Pelagibacterales and Rickettsiales (*sensu* Ferla et al., 2013); and the genus *Thermotoga*. Members of all three taxa are characterized by having reduced genomes. The fine structure of the *Thermotoga maritima* peptidoglycan has been determined biochemically, and it is known to contain ω-alanine (Boniface et al., 2009). The well-characterized
member of the Pelagibacterales is *Pelagibacter ubique*, the genome sequence of which confirms that it synthesizes peptidoglycan (Giovannoni et al., 2005). Within the Rickettsiales, we chose to focus on the *Wolbachia* endosymbiont of *D. melanogaster* (wMel). Members of the genus *Wolbachia* are intracellular parasites of arthropods and nematodes. While they do not require a stress-bearing cell wall, it has recently been shown that they require the Lipid II component of peptidoglycan (containing α-alanine) for cell division (Vollmer et al., 2013). Most interestingly, the same authors also showed that the *Wolbachia* endosymbiont of *Brugia malayi* possesses a CBL that catalyzes alanine racemization, although no kinetic parameters were reported for this enzyme.

**Cloning, expression and purification**

The *metC* genes from *P. ubique* HTCC1062 (locus tag SAR11_RS04165) and wMel ( locus WD_RS04170; Wu et al., 2004) were synthesized with their codons optimized for heterologous expression in *E. coli*. The corresponding gene in *T. maritima* MSB8 ( locus TAMAR_I_RS06470) is variously identified as either *metB* (Latif et al., 2013) or *metC* (Pysz et al., 2004) so we amplified TMARI_RS06470 from genomic DNA and investigated both possible functions.

To begin, all three genes were cloned into the pBAD expression vector. The His₆-tagged *T. maritima* enzyme (*TmCBL*) and *Wolbachia* enzyme (wMelCBL) were produced solubly, with yields routinely exceeding 20 mg of purified protein per liter of culture medium. However, numerous attempts to optimize expression of the *P. ubique* enzyme (*PuCBL*) in a soluble form were unsuccessful. Eventually, we sub-cloned *P. ubique metC* into the pMAL vector, for expression fused to maltose binding protein (MBP). This greatly improved solubility and resulted in yields of >10 mg of purified fusion protein per liter of culture. However, proteolytic cleavage of the MBP fusion partner from *PuCBL* immediately led to its precipitation in any buffer system that we tested. Therefore, we conducted all functional tests on the MBP-*PuCBL* fusion protein.

**Complementation tests**

*In vivo* complementation assays were performed as an initial test for the CBL and ALR activities of the three *metC* gene products. The expression vectors for *TmCBL*, wMelCBL and *MBP-PuCBL* were each used to transform the *E. coli ΔmetC* strain from the Keio collection (Baba et al., 2006), and also the α-alanine auxotroph *E. coli* MB2795 (Δ*alr ΔdadX*) (Soo et al., 2016).

Given that *TmCBL* had been annotated as a cystathionine γ-synthase (CGS), we also tested each enzyme for its ability to complement the methionine auxotrophy of *E. coli ΔmetB*. Complementation tests were carried out at 28°C and 37°C, as the growth temperatures of wMel and *P. ubique* are below 30°C and we hypothesized that their enzymes may be thermolabile. However, we observed no differences in the rates of colony formation at the lower temperature.

At 37°C, expression of each enzyme rescued the alanine racemase knockout, *E. coli* MB2795, as quickly as expressing *E. coli* ALR itself (Table 1). Neither *TmCBL*, nor either of the other enzymes, was able to rescue *E. coli ΔmetB*. In contrast, all three enzymes were able to complement the methionine auxotrophy of *E. coli ΔmetC*, albeit by taking 5–14 days to effect colony formation (Table 1). These data provided the first qualitative indication that *T. maritima*, *P. ubique* and wMel all possess *metC* genes that encode bi-functional CBL/ALR (but not CGS) enzymes.

**Kinetic analysis of CBL and ALR activities**

The three enzymes were purified and assayed for CBL and ALR activity (in the physiologically relevant L-Ala → α-Ala direction), as described previously for the promiscuous *E. coli* CBL enzyme (Soo et al., 2016). In the case of MBP-*PuCBL*, cystathionine β-elimination was readily detectable, with an overall catalytic efficiency (\(k_{cat}/K_m\)) of 470 s⁻¹ M⁻¹ (Table 2). It is possible that this is an underestimate of the true activity, as the effect of the MBP fusion partner on the activity of the enzyme is unknown (but it is unlikely to be rate-enhancing). The ALR activity of MBP-*PuCBL* was 40-fold lower than its CBL activity, reflecting both a lower turnover number (\(k_{cat}\)) and a higher Michaelis constant (\(K_m\), i.e., the substrate concentration required for half the maximum reaction rate) for the ALR reaction (Table 2).

### Table 1. Days to form colonies for various *E. coli* strains grown at 37°C on selective media, supplemented with either arabinose (0.02%) for the pBAD plasmids or with IPTG (50 μM) for the pCA24N and pMAL plasmids

| Over-expressed protein | *E. coli* strain | \(ΔmetC\) | \(Δalr ΔdadX\) | \(ΔmetB\) |
|-----------------------|-----------------|------------|-----------------|------------|
| TmCBL | 10 | 2 | No growth |
| wMelCBL | 14 | 2 | No growth |
| MBP-PuCBL | 5 | 2 | No growth |
| *E. coli* CBL | 1 | NT* | NT* |
| *E. coli* ALR | NT* | 1 | NT* |
| *E. coli* CGS | NT* | 2 | NT* |
| None (empty pBAD vector) | No growth | No growth | No growth |

a. NT, Not tested.

© 2017 The Authors. Molecular Microbiology Published by John Wiley & Sons Ltd., *Molecular Microbiology*, 105, 508–524
The wMelCBL enzyme showed an opposite pattern of activities to MBP-PuCBL. Consistent with the auxotrophic complementation data (Table 1), its cystathionine β-elimination activity was barely detectable above the level of background noise. While extensive controls demonstrated that the activity was present, our estimates of the $k_{cat}$ and $K_M$ values for three batches of enzyme were consistently low, and somewhat variable (Table 2). In contrast, wMelCBL was more efficient as an alanine racemase ($k_{cat}/K_M = 580 \text{ s}^{-1} \text{ M}^{-1}$) and this ALR activity was readily quantified.

In spite of being sourced from a thermophile, we began by assaying TmCBL under the same conditions as the other two enzymes (i.e., 37°C, Tris-HCl buffer, pH 8.8). Under these conditions, it was the most active CBL, with a higher $k_{cat}$ and a lower $K_M$ than MBP-PuCBL, and a catalytic efficiency of $k_{cat}/K_M = 5800 \text{ s}^{-1} \text{ M}^{-1}$ (Table 2). Conversely, its ALR activity ($k_{cat}/K_M = 5.2 \text{ s}^{-1} \text{ M}^{-1}$) was even weaker than that observed for MBP-PuCBL (Table 2).

Next we sought to test the multifunctionality of TmCBL at a more physiologically-relevant temperature. Given the large effect of temperature on the pH of solutions buffered with Tris, we switched to Bicine buffer ($\Delta \text{pH} = -0.018$ per degree Celsius) and tested CBL specific activity over the temperature range 35–85°C, in solutions that were pH 8.0 at each temperature. The results showed that TmCBL has a temperature optimum of 70°C under these conditions (Fig. 3). To allow robust comparisons between the two different buffer conditions, we determined the steady state kinetic constants in Bicine, pH 8.0, at both 37°C and 70°C (Table 2). At 37°C, TmCBL was marginally less active as a CBL in Bicine than in Tris. As expected, the enzyme was more active at 70°C, although a threefold increase in $k_{cat}$ was offset by a small increase in $K_M$, meaning that the overall increase in $k_{cat}/K_M$ was only 1.6-fold (Table 2). We also implemented a discontinuous assay that allowed us to estimate ALR activity in Bicine buffered solutions at 37°C and 70°C. As before, this activity was weak, although a small decrease in $k_{cat}$ and a larger decrease in $K_M$ meant that catalytic efficiency increased eightfold from 0.79 s$^{-1}$ M$^{-1}$ at 37°C to 6.3 s$^{-1}$ M$^{-1}$ at 70°C (Table 2).

An unexpected third enzymatic activity

During our bioinformatics investigations, we noticed that murl, encoding glutamate racemase, was absent from the genomes of P. ubique, wMel and T. maritima. α-Glutamate is an essential component of peptidoglycan.
(Schleifer and Kandler, 1972; Vollmer et al., 2008). However, all characterized glutamate racemase (GLR) enzymes use two active site cysteine residues, and not the PLP cofactor, to effect catalysis (Tanner, 2002). Canonical GLR enzymes are unrelated in sequence, structure or mechanism to either CBL or ALR. Nevertheless our bioinformatics suggested that glutamate racemization must be catalyzed by an alternative enzyme, so we tested the three CBL enzymes for GLR activity as well.

MBP-PuCBL did not possess detectable GLR activity, but the other two enzymes did. Control reactions showed that the E. coli CBL also has no GLR activity. To the best of our knowledge, wMelCBL and TmCBL are the first PLP-dependent enzymes shown to racemize glutamate. The GLR activity of wMelCBL was very weak: approximately threefold less efficient than cystathionine β-elimination; and 28-fold less efficient than alanine racemization (Table 2). Conversely, TmCBL was found to be a substantially better GLR than an ALR (Table 2). As observed for CBL activity, both $k_{\text{cat}}$ and $K_M$ increased with temperature, such that these parameters are comparable for the CBL and GLR reactions at 70°C and the overall difference in $k_{\text{cat}}/K_M$ for these two activities is only 11-fold at this temperature.

**Methionine biosynthesis in T. maritima**

We started this work with a simple evolutionary and biochemical hypothesis: that some organisms may use their CBLs as bi-functional CBL/ALR enzymes. The discovery that TmCBL has very weak ALR activity but much stronger GLR activity, and the confusion around the gene encoding it (annotated as either metB or metC) caused us to examine the likely physiological role of this enzyme in more detail.

In the standard trans-sulfurylation pathway for methionine biosynthesis, as utilized by E. coli, cystathionine γ-synthase (encoded by metB) produces cystathionine which is then cleaved by CBL to produce homocysteine (Ferla and Patrick, 2014). Our complementation tests suggested that TmCBL is indeed a metC-encoded CBL, and not a metB-encoded CGS. Regardless, we set out to determine biochemically whether TmCBL possessed CGS activity in vitro.

A complication arose because the activated substrate of E. coli CGS is O-acetyl-L-homoserine (produced from succinyl-CoA by the protein product of its metA gene), whereas in most bacteria (Ferla and Patrick, 2014), including T. maritima (Goudarzi and Born, 2006), metA encodes an enzyme that produces O-acetyl-L-homoserine from acetyl-CoA. The T. maritima enzyme has 30-fold greater activity with acetyl-CoA than succinyl-CoA (Goudarzi and Born, 2006), so O-acetyl-L-homoserine would be the preferred substrate for any T. maritima CGS enzyme. However, O-acetyl-L-homoserine is not commercially available, so we synthesized it according to a previous scheme (Nagai and Flavin, 1971). We note that we are the first to report a complete set of analytical chemical data on the compound synthesized by this route, which is included in the relevant Methods sub-section (vide infra).

TmCBL was incubated with either O-acetylmethionine or O-succinylhomoserine, plus L-cysteine (the second substrate used by CGS enzymes), and the reaction products were analyzed for the presence of cystathionine by mass spectrometry. None was detected (Fig. 4A and B), even after a greatly extended incubation period of 16 h. We determined that the lower limit of detection for a cystathionine standard in our mass spectrometry protocol was 0.5 nmol (Supporting Information Fig. S1). From this it is possible to estimate that had CGS activity been present, we would have detected it if the TmCBL in the assay was catalyzing cystathionine formation at any rate greater than $1.7 \times 10^{-4}$ turnovers per active site per second (~0.6 turnovers per hour). This upper bound on the putative CGS activity of TmCBL is >50 000-fold lower than the turnover number associated with CBL activity under the same conditions ($k_{\text{cat}} = 9.2 \text{ s}^{-1}$; Table 2). Indeed, the fact that TmCBL did not produce any detectable cystathionine (using either O-acetylmethionine or O-succinylhomoserine as a substrate) makes it extremely unlikely that the enzyme has any CGS activity at all, and rules out a physiologically-relevant role for the enzyme in cystathionine biosynthesis.

Genome analysis suggests that T. maritima possesses a metY gene to compensate for the absence of metB. The enzyme encoded by metY, O-acetylmethionine thiolase, usually catalyzes the direct production of homocysteine from O-acetylmethionine and hydrogen sulfide (H$_2$S), thus bypassing CGS and CBL completely (Ferla and Patrick, 2014). We expressed and purified the T. maritima O-acetylmethionine thiolase, to test whether it possessed the expected activity. We also tested whether it might catalyze the closely related reaction, observed in Bacillus subtilis (Auger et al., 2002), in which cystathionine is produced from O-acety lmethionine and L-cysteine. Using mass spectrometry, we observed the enzyme-catalyzed formation of homocysteine from O-acety lmethionine and H$_2$S. However, we found no evidence for the formation of cystathionine from O-acety lmethionine and L-cysteine (Fig. 4C and D).

The absence of metB and the inability of O-acetylmethionine thiolase to utilize cysteine as a substrate means that cystathionine is never synthesized by T. maritima. In the absence of this substrate, there appears to be no physiological role for the CBL activity...
of TmCBL. Instead, methionine biosynthesis occurs via the one-step conversion of O-acetylhomoserine and H$_2$S into homocysteine, catalyzed by O-acetylhomoserine thiolase.

Is TmCBL the physiological ALR?

Our results implied that metC is being maintained in the genome of T. maritima because of its role in peptidoglycan biosynthesis, and not methionine biosynthesis. However, the weak alanine racemase activity of TmCBL (Table 2) led us to question whether such an inefficient enzyme was a reasonable candidate for performing such a critical physiological task. Specialist alanine racemase enzymes typically have $k_{cat}/K_M$ values of greater than $10^4$ s$^{-1}$ M$^{-1}$ (Patrick et al., 2002; Soo et al., 2016). At $k_{cat}/K_M = 65$ s$^{-1}$ M$^{-1}$ (Soo et al., 2016), even the non-physiological, promiscuous ALR activity of E. coli CBL is an order of magnitude greater than the ALR activity of TmCBL. Thus, we returned to bioinformatics to identify alternative alanine racemase candidates.

Potential alanine racemase genes were identified by performing BLASTP searches of known racemases against the fully sequenced genome (Latif et al., 2013). $\alpha$-Alanine can also be formed from $\alpha$-glutamate by the action of a $\alpha$-amino acid transaminase, so we also searched for homologues of this enzyme. These searches turned up yggS (locus tag TMARI_RS08860, a distant homologue of alanine racemase), ilvA (TMARI_RS01820, threonine dehydratase that is a homologue of human serine racemase), dapF (TMARI_RS07765, diaminopimelate epimerase that interconverts L, L- and meso-diaminopimelate), ilvE (TMARI_RS04255, branched chain amino acid aminotransferase that is a homologue of $\alpha$-amino acid transaminases) and TMARI_RS02240 (encoding a homologue of the larA lactate...
Primordial-like bacterial enzymes

Based on our previous knowledge of the promiscuous ALR activity of E. coli CBL (Soo et al., 2016), we started this study with the hypothesis that modern CBL enzymes are descended from a bi-functional CBL/ALR ancestor (PLP-dependent enzyme fold type I), but that the alanine racemase function has been replaced in most lineages by an alternate and non-homologous ALR specialist (PLP-dependent enzyme fold type III). However, our combination of biochemistry and bioinformatics provided evidence that CBL activity is not required for methionine biosynthesis in P. ubique, wMel or T. maritima. Instead, the physiological roles of wMel and TmCBL appear to be in acting as primordial-like, broad-specificity amino acid racemases.

To investigate the evolutionary history of CBL enzymes further, we inferred a phylogenetic tree of representative CBL sequences and compared it to a tree made with concatenated 16S and 23S rRNA sequences from the same species (Supporting Information Fig. S2). Simplified cladograms are shown in Fig. 5. The species and CBL trees are not congruent, but instead they are consistent with patterns of frequent metC loss and gain by horizontal transfer. For example, the Thermotogae CBL sequences cluster with those from the Bacteriodetes species Pontibacter roseus (Fig. 5B), whereas they are only more distantly related to sequences from Deinococcus-Thermus, which is a more closely-related phylum (Fig. 5A and Hug et al., 2016). The CBL sequences from the Alphaproteobacteria are also found in different clades (Fig. 5B). While rRNA sequences place P. ubique and wMel together in the subclass Rickettsiidae (Fig. 5A and Ferla et al., 2013), wMelCBL shared a more recent common ancestor with E. coli CBL than PuCBL (Fig. 5B). In contrast, a tree of alphaproteobacterial ALR sequences shows them to be vertically transmitted (Supporting Information Fig. S3). Thus, a parsimonious explanation is that gain of a multifunctional CBL by horizontal transfer has led to the subsequent displacement of ALR in the lineages leading to P. ubique, wMel and T. maritima.

Discussion

Primordial-like enzymes in non-canonical pathways

The goal of this study was to identify and characterize primordial-like enzymes, to shed light on primordial metabolism and processes of enzyme evolution. We discovered that the extant bacteria P. ubique, T. maritima and the Wolbachia endosymbiont of D. melanogaster have multifunctional CBL enzymes. However, our work has also highlighted the difficulties associated with assigning physiological functions to enzymes from non-model microorganisms with non-canonical metabolic pathways, in which gene knockouts are technically unfeasible.

The simplest situation arises in wMel. This obligately intracellular, parasitic bacterium has a heavily reduced genome of only 1.27 Mbp, which encodes 1270 proteins (Wu et al., 2004). The only met gene in its genome is metC, demonstrating that it is a methionine auxotroph. Our kinetics data (Table 2) show that metC has been gained and retained because of the alanine racemase and glutamate racemase activities of the enzyme it encodes, wMelCBL. Indeed, the vestigial CBL activity of wMelCBL has eroded to the point where it is now 10-fold weaker than the ALR activity. While the enzyme retains a Michaelis constant for cystathionine (K_M ~ 20 μM; Table 2) that is comparable to that of E. coli CBL for the same substrate (K_M= 39 μM; Soo et al., 2016), its ability to turn over the substrate has almost entirely disappeared (k_cat ~ 4 h^{-1}). Conversely, wMelCBL readily turns over L-alanine (k_cat = 2.3 s^{-1}) and the enzyme appears sufficiently active to provide the cell with the small amount of D-alanine it requires during cell division (Vollmer et al., 2013). The kinetic parameters of wMelCBL for the GLR reaction are poor – particularly k_cat, which is 135-fold lower than for the ALR reaction (Table 2). This decreased turnover number is likely to be offset somewhat by the relative intracellular abundance of the substrate L-glutamate, compared with L-alanine. The former is present at a 40-fold higher
concentration than the latter in *E. coli* (Bennett et al., 2009). Thus, it is highly likely that CBL is actually a primordial-like, bi-functional ALR/GLR enzyme in *wMel*, required for catalyzing two steps in the synthesis of the Lipid II component of peptidoglycan. Presumably the ancestor of *wMelCBL* was bi-functional when it was gained via horizontal transfer, such that it released *alr* and *murI* from selection and led to their loss.

Our attempts to characterize *P. ubique* were hindered by our inability to purify the enzyme without a maltose binding protein fusion partner. Nevertheless, the MBP-*PuCBL* fusion was active as a CBL and an ALR (but not as a GLR). As in *wMel*, the CBL activity of *PuCBL* is likely to be vestigial. *P. ubique* has the smallest genome known for any free-living organism, encoding 1354 open reading frames with its 1.31 Mbp chromosome (Giovannoni et al., 2005). A defining characteristic of this oligotrophic bacterium is its requirement for a reduced source of sulfur – either methionine or 3-dimethylsulfoniopropionate – for growth (Tripp et al., 2008; Carini et al., 2013). In cases where exogenous methionine is not available, it possesses enzymes for degrading 3-dimethylsulfoniopropionate to methanethiol (Reisch et al., 2011). *P. ubique* has an O-acetylhomoserine thiolase (encoded by *metY*), and it is known that these enzymes can catalyze the conversion of methanethiol, plus O-acetylhomoserine, directly into methionine (Ferla and Patrick, 2014). Therefore, *P. ubique* has no requirement for a CBL in methionine biosynthesis, and it is unsurprising that other unnecessary *met* genes, including *metB* and the methionine synthases (*metE* and *metH*), have been lost from its streamlined genome.

The implication is that *P. ubique* is now retaining CBL as its specialist alanine racemase. On one hand, the poor kinetic parameters of MBP-*PuCBL* for the ALR reaction ($k_{cat}/K_M = 12 \text{ s}^{-1} \text{ M}^{-1}$) make this explanation appear unlikely. On the other hand, its Michaelis constant ($K_M = 12 \text{ mM}$) is comparable to those reported for the fold type III alanine racemases from species such as *Pseudomonas fluorescens*, *Bacillus psychrosaccharolyticus* and *Shigella dysenteriae* (Yokoigawa et al., 2013).

**Fig. 5.** Comparison between the phylogenies of (A) the concatenated 16S and 23S rRNA genes; and (B) the CBL enzymes for 20 representative bacterial species. Species are color-coded at the level of phylum and/or class: Gammaproteobacteria – green; Betaproteobacteria – brown; Alphaproteobacteria – blue; Bacteroidetes – black; Deinococcus-Thermus – red; Thermotogae – purple. Note that the genome of *Yersinia pseudotuberculosis* encodes two CBL enzymes. Bootstrap values ($n = 500$) are shown at the nodes. Only tree topologies are shown; branch lengths do not represent genetic distances. Expanded versions of these trees, showing genetic distances and outgroups, are provided as Supporting Information Fig. S2.
The relatively poor turnover number of MBP-\(Pu\)CBL (\(k_{\text{cat}} = 0.15 \text{ s}^{-1}\)) may in part reflect the fusion to MBP. Further, these kinetics may not be maladaptive given the slow growth rate (and, therefore, low peptidoglycan requirement) of *P. ubique*. Even under optimized laboratory conditions, this bacterium completes one division per day (Tripp *et al.*, 2008; Carini *et al.*, 2013). In the absence of any more convincing candidates, we propose that the physiological role of *Pu*CBL is to act as an alanine racemase. Whether this is its only physiological role remains an open question. It is possible that *Pu*CBL catalyzes other (as yet undiscovered) reactions, especially as there are fold type I PLP-dependent enzymes that are known to catalyze transamination, \(\beta\)-replacement, \(\gamma\)-elimination, decarboxylation and side-chain cleavage reactions (Raboni *et al.*, 2010), in addition to \(\beta\)-elimination and racemization.

The third CBL we have characterized is from the obligately anaerobic thermophile, *T. maritima*. Like wMel and *P. ubique*, this bacterium has a small genome (1.87 Mbp; 1872 protein-encoding genes; Latif *et al.*, 2013). *Tm*CBL retains CBL activity, but in spite of its genomic annotation as *metB* (Latif *et al.*, 2013) it is not active as a CGS (Fig. 4A and B). *T. maritima* also has a *metY* gene, which we have shown encodes an active \(O\)-acytylhomoserine thiolsase (Fig. 4C), enabling CBL to be bypassed in methionine biosynthesis.

Like wMel, *T. maritima* appears to have gained a CBL and retained it as a bi-functional alanine/glutamate racemase. Our efforts to identify alternate alanine racemases were unsuccessful, although it remains possible that some other *T. maritima* protein may possess this activity (albeit while lacking detectable sequence similarity with any previously-described candidate). At the non-physiological temperature of 37°C, the three activities of *Tm*CBL are present in the ratio 5900:120:1 (CBL/GLR:ALR; ratio of \(k_{\text{cat}}/K_{M}\) values in Bicine buffer in Table 2). At the physiological temperature of 70°C, this ratio changes to 1200:110:1; that is, the enzyme is proportionately worse as a CBL and the ratio of ALR activity to GLR activity remains unchanged. In general the \(k_{\text{cat}}\) and \(K_{M}\) of enzymes both increase with temperature (Somero, 1995), and this is observed for the CBL and GLR activities of *Tm*CBL (Table 2). Unusually, \(k_{\text{cat}}\) and \(K_{M}\) for the alanine racemase reaction both decrease with increasing temperature, with the 22-fold decrease in \(K_{M}\) ensuring that *Tm*CBL has an overall catalytic efficiency (\(k_{\text{cat}}/K_{M} = 6.3 \text{ s}^{-1} \text{ M}^{-1}\)) that is comparable to that of MBP-*Pu*CBL. This dramatic change in \(K_{M}\) is largely responsible for changing the ratio of the three activities, and suggests a structural rearrangement that alters the ability of the enzyme to discriminate between its substrates. Structural biology and protein dynamics experiments will be required to explore this hypothesis further.

**Active site similarities and differences**

As expected, homology models show that the structures of *Pu*CBL, wMelCBL and *Tm*CBL are all fold type I enzymes, akin to the promiscuous *E. coli* CBL (Fig. 1A). The PLP cofactor is anchored in the active site of the *E. coli* enzyme by Lys210 (Clausen *et al.*, 1996). The incoming substrate (either cystathionine or \(L\)-alanine) displaces Lys210 and forms a Schiff base with the cofactor. Lys210 then acts as a catalytic base, and in the alanine racemase reaction Tyr111 acts as a catalytic acid (Soo *et al.*, 2016). These two catalytic amino acids, as well as the key active site residues Ser339 and Tyr56’ (the prime indicates a residue from the neighboring subunit), are conserved in *Pu*CBL, wMelCBL and *Tm*CBL.

Another important active site residue is Arg58’, which modulates the \(pK_a\) of Tyr111 via a hydrogen bonding interaction (Lodha *et al.*, 2010; Soo *et al.*, 2016). Arg58’ makes no interaction with Pro113 in wild type *E. coli* CBL, but when Pro113 is mutated to serine, a new hydrogen bond with Arg58’ is introduced and alanine racemase activity is enhanced (Soo *et al.*, 2016). The enzymes with poor ALR activity – *Pu*CBL and *Tm*CBL – both have an arginine in the position equivalent to Arg58’, with the former having a proline equivalent to Pro113 and the latter having tryptophan instead. In contrast, wMelCBL, which is the only enzyme more active as an ALR than a CBL, has asparagine and arginine respectively, in place of Arg58’ and Pro113. These observations suggest that the evolutionary route to enhancing ALR activity at the expense of CBL activity lies in optimizing the hydrogen bonding network around the catalytic acid, Tyr111.

Our discovery that wMelCBL and *Tm*CBL both possess glutamate racemase activity was unexpected, as comprehensive surveys have failed to ascribe this activity to any PLP-dependent enzyme (Percudani and Paccaloni, 2009; Raboni *et al.*, 2010). The PLP-dependent racemization of \(L\)-glutamate should proceed via the same mechanism as \(L\)-alanine racemization (deprotonation by the catalytic lysine and reprotonation by the catalytic tyrosine). On one hand, it is not surprising that glutamate can be accommodated in the CBL active site, given that it is intermediate in size between cystathionine and alanine (Fig. 1). On the other hand, neither the *E. coli* CBL nor MBP-*Pu*CBL possesses GLR activity. Experimentally determined structures (rather than homology models) will be required to better understand the source of GLR activity; to this end crystals of the wMelCBL and *Tm*CBL proteins have recently been grown in our laboratory.
Biochemical evolution in bacteria with reduced genomes

Our bioinformatics-based search for primordial-like CBL enzymes was not biased toward species with any particular life history trait or genome architecture. Nevertheless, it led us to wMel, *P. ubeique* and *T. maritima*, all of which have atypically small genomes (<2 Mbp). For endosymbionts such as wMel, rapid genome reduction is due to their ability to obtain metabolites from the host and, critically, also due to population structures that result in high levels of drift but weak selection (Wu et al., 2004; McCutcheon and Moran, 2011). Similarly, the population characteristics of endosymbionts—low effective population size, frequent bottlenecks during transmission and a lack of recombination—lead to rapid sequence evolution (McCutcheon and Moran, 2011). In the case of wMelCBL, this appears to have manifested as a particularly rapid erosion of its ancestral CBL activity (Table 2).

While there is clearly much more to be learned about the enzymology of endosymbionts, our primary goal is to understand the origins and evolution of metabolic networks. Logically, the first cells could not have been endosymbionts. Therefore, we are particularly interested in free-living bacteria with primordial-like enzymes, such as *P. ubeique* and *T. maritima*. In these species, unlike endosymbionts, selection may have driven genome reduction to favor cell architectures that minimize the resources required for growth (Wolf and Koonin, 2013; Giovannoni et al., 2014). The evidence we have presented suggests that the genes for several specialized enzymes (e.g., ALR and GLR) can indeed be lost from these bacteria, provided that a physiologically multifunctional—rather than promiscuous—enzyme, such as TmCBL, is gained. We have also provided evidence that options which shorten metabolic pathways (e.g., *O*-acetylhomoserine thiolase instead of CGS and CBL) may be favored in free-living organisms with reduced genomes.

Primordial-like enzymes have certainly been characterized from bacteria with larger genomes. For example, the bi-functional isomerase PriA was initially identified in *M. tuberculosis* (with a genome of 4.41 Mbp) and in *S. coelicolor*, which has one of the largest bacterial genomes known, at 8.67 Mbp (Barona-Gómez and Hodgson, 2003). Nevertheless, genome reduction is a pervasive mode of evolution (Wolf and Koonin, 2013), and two other examples of primordial-like enzymes are from endosymbionts: TrpF from *Chlamydia trachomatis* (Adams et al., 2014), which has a 1.04 Mbp genome; and IlvC from *Buchnera* (Price and Wilson, 2014), which has a 416 kbp genome. While the data on primordial-like enzymes from free-living bacteria are still limited, it is tantalizing to speculate that selection is reducing the genomes of many species, while simultaneously returning their metabolic networks to a state that is comparable to the one envisaged by Yčas and Jensen for primordial cells (Yčas, 1974; Jensen, 1976).

Primordial-like enzymes are under selection for multifunctionality. In the case of our CBLs, this has yielded enzymes with *K*<sub>M</sub> values that are in the range of more specialized enzymes. In their landmark survey of all the enzyme parameters in the comprehensive BRENDA database (Chang et al., 2015), Bar-Even et al. found that ~60% of enzymes have *K*<sub>M</sub> values in the range 0.01–1 mM (2011). None of the *K*<sub>M</sub> values we determined (Table 2) lie in the top 5% or bottom 5% of the Bar-Even dataset. Thus, our multifunctional CBLs show substrate recognition properties that are comparable to specialized enzymes. In contrast, the primordial-like CBL enzymes have poor *k*<sub>cat</sub> values; that is, they are slow at converting substrate into product. Of the 6,530 *k*<sub>cat</sub> values that were compiled by Bar-Even et al., only 73 (1.1%) are lower than the value we measured for TmCBL acting as an alanine racemase at its physiological temperature (*k*<sub>cat</sub> = 0.0024 s<sup>−1</sup>; Table 2). Similarly, the turnover number of wMelCBL acting as a glutamate racemase is in the lowest 3% of those examined by Bar-Even et al.

The general trend for enzymes from organisms without reduced genomes (i.e., >99% of the enzymes in BRENDA) is for those in central metabolism to be more efficient than those in intermediate or secondary metabolism (Bar-Even et al., 2011). In spite of their roles in central metabolism, the primordial-like CBL enzymes have kinetic parameters that place them among the least efficient enzymes inhabiting any part of the metabolic network. Nevertheless, bacteria with reduced genomes are the most abundant organisms in the biosphere (Wolf and Koonin, 2013; Giovannoni et al., 2014). As we learn more about the enzymes from these organisms, it seems likely we will conclude that most of the enzymes on the planet are substantially less efficient catalysts than the subset we have been studying for the past century.

Concluding remarks

This study provides evidence that the weakly active, multifunctional enzymes encoded by reduced genomes are excellent (but previously underappreciated) comparative models for studying primordial enzymes. Further, we suggest that the patchwork metabolic networks within free-living bacteria such as *P. ubeique* and *T. maritima* are ideal starting points for understanding the ancient origins of metabolic biochemistry. However, our data also emphasize the difficulty in assigning functions...
to many of the genes in these bacteria, based solely on sequence similarity. This is particularly true in the case of PLP-dependent enzymes, for which many different functions can be found on each different fold, with short mutational routes between these functions (Eliot and Kirsch, 2004; Raboni et al., 2010; Soo et al., 2016). For simplicity, we have persevered with calling the subjects of this study metC genes and CBL enzymes. However, we have presented evidence that these are misnomers, in spite of the sequence evidence to the contrary. We propose the new function-based annotation of aar (amino acid racemase) for the metC genes of wMel and T. maritima, and airX (alanine racemase, distinct from the alr genes that encode fold type III enzymes) for P. ubique metC.

A recent editorial highlighted the extraordinary richness and diversity of the 'esoteric, niche enzyme' that is largely absent from the textbooks (Tawfik and van der Donk, 2016). We have gone further, and suggested that esoteric enzymes — such as the poorly-active, multifunctional ones we have characterized in this study — represent the rule, and not the exception, in the biosphere. The emerging interest in esoteric enzymology highlights the need for an 'Enzymatic Encyclopaedia of Bacteria and Archaea', in analogy with the Genomic Encyclopaedia of Bacteria and Archaea (GEBA) project (Wu et al., 2009a). Continuing the analogy with phylogenomics, we suggest that the term 'phyloenzymology' encapsulates our approach of identifying enzymes to characterize, based solely on their phylogenetic novelty.

Experimental procedures

Materials

Oligonucleotides were from Integrated DNA Technologies (Coralville, IA). Restriction enzymes were from New England Biolabs (Ipswich, MA). Chemicals were from Sigma Chemical (St Louis, MO) unless noted otherwise.

Search for candidate species

A search to identify bacterial genomes without an alr homologue but with a metC homologue was conducted. A Perl script (available at https://github.com/matteoferla/Perly-scripts/tree/master/COG%20genome%20tool) first downloaded the protein tables of each fully sequenced bacterial genome from the NCBI http site (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi) and then parsed them based on the presence of the COG0787 (alanine racemase) and COG0626 (cystathionine β-lyase/γ-synthase) annotations. The species present were placed on a taxonomic tree generated by the iTOL server (itol.embl.de; Letunic and Bork, 2011) and annotated with the presence or absence of the two genes. This allowed groups of species with the same presence/absence pattern of alr and metC/metB to be collapsed into higher taxa, simplifying the tree. The list of species without alr was validated with BLASTP and TBLASTN searches in the NCBI database (Sayers et al., 2009).

Cloning of metC genes

The wMel metC gene was synthesized with codon optimization for expression in E.coli by GeneArt (Life Technologies), and modified to possess an N-terminal hexahistidine tag and a linker that contained a KpnI site. It was sub-cloned into pBAD/myc-His(B) (Invitrogen) with Ncol and HindIII. T. maritima metC was amplified from genomic DNA using the custom primers Tma_metC_KpnI_F (CAG GTA CCG AGA ACC TGT ATT TTC AGG GAA ACA CAG AGC ACA TTC TGT TTT CTG ACC G) and Tma_metC_XbaI_R (CTT GGT GTA TTA TTC TTT GTT AGT GCC TGA TCC TTA TCT GC). The product was cloned into the vector above with KpnI and XbaI. P. ubique metC was synthesized by DNA2.0, with codon optimization. The gene was amplified with primers Pub_metC_AvaI_F (AAC AAC CTC GGG ACT GAG GGA AGG ATG ACC AAA TCC TTT AAA ACC TTT C) and Pub_metC_AvaI_R (AAC AAC CTC GGG ACT GAG GGA AGG ATG ACC AAA TCC TTT AAA ACC TTT C) and cloned into the vector above with KpnI and XbaI. In vivo complementation

Complementation tests employed the E. coli ΔmetC and ΔmetB strains from the Keio collection (Baba et al., 2006), and the alanine racemase knockout strain, E. coli MB2795 (Δair ΔdadX) (Soo et al., 2016). In addition to the CBL expression vectors, an empty pBAD plasmid was used as a negative control. Positive controls for rescuing each strain were the plasmids pCA24N-metC, pCA24N-alr and pCA24N-metB from the ASKA collection of E. coli over-expression vectors (Kitagawa et al., 2005). Cells harboring the relevant plasmid were grown overnight in rich medium, pelleted, washed and re-suspended in 1 × 9 M salts. For testing complementation of the E. coli ΔmetC and ΔmetB strains, ~10^5 colony forming units were spread on M9 agar plates containing 0.4% (w/v) glucose, the relevant antibiotic for maintaining the plasmid (100 µg ml^-1 ampicillin or carbenicillin for pBAD and pMAL, or 34 µg ml^-1 chloramphenicol for pCA24N), and inducer (0.02% arabinose or 50 µM IPTG). In the case of E. coli MB2795, LB medium was used instead of M9 medium, as it lacks γ-alanine and the strain has multiple uncharacterized auxotrophies. Plates were incubated in airtight containers at 28°C or 37°C, for up to 4 weeks, and colony formation was monitored.

Expression and purification of CBL enzymes

TmaCBL and wMelCBL were expressed in E. coli strain LMG194 (Invitrogen) and purified by immobilized metal affinity chromatography (IMAC). MBP-PuCBL was expressed in E. coli ER2523 (New England Biolabs) and purified by amylose-affinity chromatography. For IMAC, the cell lysis buffer comprised 50 mM potassium phosphate, 300 mM NaCl and 10% (v/v) glycerol, pH 7.0. For amylose-
affinity chromatography, the lysis buffer contained 20 mM Tris HCl, 600 mM KCl, 1 mM EDTA, 10 mM β-mercaptoethanol and 10 μM PLP, pH 8.0.

Each expression strain was cultured in Terrific Broth (medium; Hunstanton, UK) supplemented with the appropriate antibiotic, at 37°C with shaking until an OD_{600} of 0.6 was reached. Cultures were transferred to 28°C and overnight protein expression was induced by the addition of 0.02% (w/v) arabinose or 0.5 mM IPTG. The cell pellets were harvested by centrifugation and then re-suspended in 0.02% (w/v) arabinose or 0.5 mM IPTG. The cell pellets were lysed by sonication on ice and cellular debris was separated from soluble protein by further centrifugation. After clarification through a 0.45 μm syringe-driven filter, the soluble lysate was applied to either Talon resin (Clontech; Mountain View, CA) for IMAC or to amylose resin (New England Biolabs) for amylose-affinity chromatography. After incubation at 4°C with rocking, for at least 1 h, the resins were washed extensively with lysis buffer, packed into BioSpin gravity flow columns (BioRad, Hercules, CA) and washed further. For IMAC, protein was eluted from the resin with lysis buffer supplemented with 100–500 mM imidazole. Fractions were pooled and exchanged into storage buffer (50 mM potassium phosphate buffer, 200 mM NaCl, 10% v/v glycerol, pH 7.0) using an Amicon centrifugal filter unit with a 50 kDa molecular weight cut-off (EMD Millipore; Billerica, MA). For amylose-affinity chromatography, the protein was eluted with 5 ml of lysis buffer containing 10 mM maltose. Fractions containing pure MBP-PuCBL were dialyzed extensively against storage buffer (50 mM Tris-HCl, 600 mM KCl, 10% v/v glycerol, pH 7.5) using a dialysis cassette with a 10 kDa molecular weight cut-off (Pierce Biotechnology; Waltham, MA). Enzyme concentrations were determined using molecular extinction coefficients that were calculated as described previously (Pace et al., 1985). All proteins were snap frozen in liquid nitrogen and stored at −80°C until use; control assays verified that this led to no significant loss of activity.

**Kinetic assays**

Continuous assays were performed using a Cary 100 UV-Vis spectrophotometer with temperature controller (Agilent Technologies; Santa Clara, CA). Cuvettes containing each reaction mixture were incubated at assay temperature for 2 min prior to the start of the assay. The absorbance was monitored for 4 min to establish a stable baseline and then for a further 8 min following the addition of enzyme. Progress curves were plotted and the initial rate was calculated by subtracting the slope of the baseline from the slope of the enzyme-catalyzed reaction. Except as noted in Table 2, technical triplicates and minimally two biological replicates were performed. Appropriate enzyme concentrations were determined on a protein-by-protein basis for each assay, depending on the rate of reaction. A minimum of 7 different substrate concentrations were measured to construct each kinetic curve. Initial rate data were plotted and fitted to the Michaelis-Menten equation using GraphPad Prism.

The continuous coupled assay for measuring cystathionine β-elimination (cystathionine → homocysteine + pyruvate) activity was performed as reported previously (Soo et al., 2016). The reaction mixture contained 0.8 mM 5,5-dithiobis(2-nitrobenzoic acid) (DTNB) and 10 μM PLP, buffered with either 50 mM Bicine (pH 8.0) or 50 mM Tris-HCl (pH 8.8). Concentrations of up to 5 mM L-cystathionine and 5 μM enzyme were used. The production of homocysteine was measured by monitoring the cleavage of DTNB at 412 nm (ε = 14 150 M⁻¹ cm⁻¹).

A continuous assay (Essaki and Walsh, 1986) was used to measure alanine racemization (L-alanine → D-alanine) by wMelCBL and PuCBL. The assay utilized coupled reactions converting the product, D-alanine, to pyruvate by D-amino acid oxidase and the subsequent reduction of pyruvate by lactate dehydrogenase, which is coupled to the oxidation of NADH. Reaction mixtures contained 50 mM Tris-HCl pH 8.8, 10 μM PLP, 0.2 mM NADH, 1 U ml⁻¹ D-amino acid oxidase (from porcine kidney; Sigma), 120 U ml⁻¹ lactate dehydrogenase (from bovine heart; Sigma), up to 100 mM L-alanine and up to 5 μM CBL. Assays were performed at 37°C and monitored the disappearance of NADH at 340 nm (ε = 6220 M⁻¹ cm⁻¹)

The alanine racemase activity of TmCBL (L-alanine → D-alanine) was measured using a discontinuous assay (Patrick et al., 2002). Reaction mixtures contained 10 μM PLP, up to 5 μM CBL and up to 40 mM L-alanine, buffered with either 50 mM Bicine (pH 8.0) or 50 mM Tris-HCl (pH 8.8). The reaction was incubated at 37°C or 70°C for up to 20 min. To stop the reaction, TmCBL was inactivated by incubation at 95°C for 5 min. Next, 10 μl of reaction mixture was transferred to a flat-bottomed Costar 96 well plate (Corning; Corning, NY) and 90 μl of color development reagent (100 mM sodium phosphate pH 7.0, 1.8 U ml⁻¹ D-amino acid oxidase, 20 U ml⁻¹ horseradish peroxidase from Sigma and 2 mg ml⁻¹ O-phenylenediamine) was added. Colour was allowed to develop by incubation at 37°C for 45 min and then stopped by the addition of 100 μl of HCl (3 M). The absorbance was measured at 492 nm in a Multiskan plate reader (Thermo Scientific; Waltham, MA) and quantified by reference to a standard curve (0–10 nmol D-alanine, freshly made for each assay).

Glutamate racemization (L-glutamate → D-glutamate) was measured using a continuous assay (Lundqvist et al., 2007) that utilized a coupled reaction catalyzed by glutamate dehydrogenase (converting L-glutamate to α-ketoglutarate), which in turn was coupled to the reduction of NAD⁺. The reaction mixture contained 10 μM PLP, 5 mM NAD⁺ and up to 5 μM D-glutamate, buffered with either 50 mM Bicine (pH 8.0) or 50 mM Tris-HCl (pH 8.8). For assays performed at 37°C, 0.4 U ml⁻¹ of bovine liver glutamate dehydrogenase (Sigma-Aldrich) was used as the coupled enzyme. For assays at 70°C, we expressed and purified the glutamate dehydrogenase from T. maritima, and used it at a concentration of 30 μM. CBL concentrations of 0.2–5 mM were used. The appearance of NADH was monitored at 340 nm.

**Synthesis and analysis of O-acetyl-L-homoserine**

O-acetyl-L-homoserine was synthesized with minor modifications to a previous protocol (Nagai and Flavin, 1971).
Molecular Microbiology, 105, 508–524

Mass spectrometry

To test whether TmCBL could act as a cystathionine γ-synthase and catalyze the formation of cystathionine a reaction mixture was set up containing: 10 mM Bicine pH 8.0; 5 mM L-cysteine, 10 μM PLP; 10 μM TmCBL; and either 5 mM O-acetylhomoserine or 5 mM O-succinylhomoserine. To determine whether the T. maritima O-acetylhomoserine thiolase could catalyze the production of homocysteine and/or cystathionine, from O-acetylhomoserine and either H2S or cysteine respectively, the following reaction mixture was set up: 10 mM Bicine pH 8.0; 5 mM O-acetylhomoserine; 10 μM PLP; 10 μM O-acetylhomoserine thiolase and either 5 mM Na2S or 5 mM L-cysteine. The reactions were incubated for 16 h at 70°C, and then for 5 min at 95°C to inactivate the enzyme, before being stored at −20°C prior to analysis by mass spectrometry.

For analysis, the samples were diluted 10-fold with 5% (v/v) acetonitrile and 0.2% (v/v) formic acid and directly injected at 800 nL min−1 into an Ultimate 3000 nano-flow UHPLC ( Dionex; Sunnyvale, CA) coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Scientific). MS-1 was produced under identical conditions (Supporting Information Fig. S1).

Phylogenies of CBL and ALR

The CBL tree was constructed as follows. The protein sequences were chosen from representatives of each major CBL-containing clade, as determined from a large reference tree that in turn had been assembled from >200 sequences found by concatenated BLASTP searches. The sequences were aligned with MUSCLE (Edgar, 2004) and the header names were changed with a Python3 script and nw_rename from the Newick utilities (Junier and Zdobnov, 2010). The tree was inferred with RAxML (Stamatakis, 2014) using fast bootstrap with 500 replicates under a WAG model with Π distribution and with NP_253712.1 (O-acetylhomoserine thiolase from Pseudomonas aeruginosa PA01) as the outgroup. Six iterations were done until a satisfactory tree was obtained. The rRNA tree for the species in the CBL tree was built using ARB SINA to obtain good quality alignments of the 16S and 23S rRNA sequences, which were then concatenated with a Python3 script. RAxML was used to infer the tree of the concatenated sequences, with a Π-distributed GTR model and 500 bootstraps.

Experiments with additional T. maritima enzymes

Expression vectors for the metY, ilvA, dapF, ilvE and TMARI_RS02240 genes of T. maritima were purchased from the DNASU Plasmid Repository (https://dnasu.org). These vectors were from the pMH series, constructed by the Joint Consortium for Structural Genomics (Lesley et al., 2002), and they facilitated arabinose-induced expression of His6-tagged enzymes. The yggS gene was amplified from T. maritima MSB8 genomic DNA using primers yggS_fwd (TAC CGA GAA CCT GTA TTT CCA AGG AGG ATT GAA AGA AAA CCT AAG AGG GG) and yggS_rev (TGA GAT GAG TTT TTG TTC TAG AAG CTC ACT TCC CTC CTT CCA ATA TGG CG). TMARI_RS08180 was amplified from genomic DNA with primers TM1597_fwd (TAC CGA GAA CCT GTA TTT CCA AGG AGT GTA TCC CAG GCT TCT GAT AAA TC) and TM1597_rev (TGA GAT GAG TTT TTG TTC TAG AAG CTC AGA TTG GAT GTT CAT ACA CTT TC). The two amplified inserts were cloned into pBAD using Gibson Assembly, after the vector had been amplified with pBAD_fwd (GCT TCT AGA ACA AAA ACT CAT C) and pBAD_rev_His (TCC TTG GAA ATA CAG GTT C).

For activity and mass spectrometry experiments, the proteins encoded by metY (O-acetylhomoserine thiolase) and TMARI_RS08180 were expressed and purified using IMAC, as described above. To test for alanine racemase activity, the vectors for expressing the proteins encoded by ilvA, dapF, ilvE, yggS, TMARI_RS02240 and TMARI_RS08180 were used to transform E. coli MAD2 (ΔΔαr ΔdadX ΔmetC). Complementation tests were carried out as described above, using LB supplemented with ampicillin (100 μg ml−1) and arabinose (0.1%). Plates were incubated in air-tight containers at 37°C, for 4 weeks.

© 2017 The Authors. Molecular Microbiology Published by John Wiley & Sons Ltd., Molecular Microbiology, 105, 508–524
Acknowledgements

We are grateful to Prof. Emily Parker (University of Canterbury) for providing *T. maritima* MSB8 genomic DNA, and to Dr. Torsten Kleffmann and Dr. Abhishek Kumar (Centre for Protein Research, University of Otago) for obtaining mass spectrometry data. *E. coli* strain MAD2 was a kind gift from Mark Lunzer and Prof. Tony Dean at the University of Minnesota. We thank Natasha Comer (Massey University) and James McKellar (University of Otago) for their expert technical assistance, and Bronwyn Carlisle (University of Otago) for her assistance with preparing figures. This work was supported by a grant from the Marsden Fund and by the award of a Rutherford Discovery Fellowship to W.M.P.

Author contributions

MPF and WMP conceived and designed the study. MPF, JLB, KRH and GBE acquired data. All authors analyzed data and contributed to the writing of the manuscript.

References

Adams, N.E., Thiaville, J.J., Proestos, J., Juárez-Vázquez, A.L., McCoy, A.J., Barona-Gómez, F., et al. (2014) Promiscuous and adaptable enzymes fill "holes" in the tetrahydrofolate pathway in *Chlamydia* species. *mBio* 5: e01378–e01314.

Auger, S., Yuen, W.H., Danchin, A., and Martin-Verstraete, I. (2002) The metIC operon involved in methionine biosynthesis in *Bacillus subtilis* is controlled by transcription antitermination. *Microbiology* 148: 507–518.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol Syst Biol* 2: 2006.0008.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., et al. (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50: 4402–4410.

Barona-Gómez, F., and Hodgson, D.A. (2003) Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep* 4: 296–300.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., and Rabinowitz, J.D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* 5: 593–599.

Boniface, A., Parquet, C., Arthur, M., Mengin-Lecreulx, D., and Blanot, D. (2009) The elucidation of the structure of *Thermotoga maritima* peptidoglycan reveals two novel types of cross-link. *J Biol Chem* 284: 21856–21862.

Carini, P., Steindler, L., Beszteri, S., and Giovannoni, S.J. (2013) Nutrient requirements for growth of the extreme oligotroph *Candidatus Pelagibacter ubique* HTCC1062 on a defined medium. *ISME J* 7: 592–602.

Chang, A., Schomburg, I., Placzek, S., Jeske, L., Ulbrich, M., Xiao, M., et al. (2015) BRENA in 2015: Exciting developments in its 25th year of existence. *Nucl Acids Res* 43: D439–D446.

Clausen, T., Huber, R., Laber, B., Pohlenz, H.-D., and Messerschmidt, A. (1996) Crystal structure of the pyridoxal-5'-phosphate dependent cystathionine β-lyase from *Escherichia coli* at 1.83 Å. *J Mol Biol* 262: 202–224.

Copley, S.D. (2015) An evolutionary biochemist's perspective on promiscuity. *Trends Biochem Sci* 40: 72–78.

Du, J., Say, R.F., Lu, W., Fuchs, G., and Einsle, O. (2011) Active-site remodelling in the bifunctional fructose-1,6-bisphosphate aldolase/phosphatase. *Nature* 478: 534–537.

Due, A.V., Kuper, J., Geerlof, A., von Kries, J.P., and Wilmanns, M. (2011) Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc Natl Acad Sci USA* 108: 3554–3559.

Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res* 32: 1792–1797.

Eliot, A.C., and Kirsch, J.F. (2004) Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem* 73: 383–415.

Esaki, N., and Walsh, C.T. (1986) Biosynthetic alanine racemase of *Salmonella typhimurium*: Purification and characterization of the enzyme encoded by the *alr* gene. *Biochemistry* 25: 3261–3267.

Ferla, M.P., and Patrick, W.M. (2014) Bacterial methionine biosynthesis. *Microbiology* 160: 1571–1584.

Ferla, M.P., Thrash, J.C., Giovannoni, S.J., and Patrick, W.M. (2013) New rRNA gene-based phylogenies of the *Alphaproteobacteria* provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS One* 8: e83383.

Fuerst, J.A., and Sagulenko, E. (2011) Beyond the bacterium: Planctomycetes challenge our concepts of microbial structure and function. *Nat Rev Microbiol* 9: 403–413.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucl Acids Res* 43: D261–D269.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309: 1242–1245.

Giovannoni, S.J., Cameron Thrash, J., and Temperton, B. (2014) Implications of streamlining theory for microbial ecology. *ISME J* 8: 1553–1565.

Glasner, M.E., Gerlt, J.A., and Babbitt, P.C. (2006) Evolution of enzyme superfamilies. *Curr Opin Chem Biol* 10: 492–497.

Goudarzi, M., and Born, T.L. (2006) Purification and characterization of the enzyme encoded by the *T. maritima* MSB8 genomic DNA, and to contribute to the writing of the manuscript.

Due, A.V., Kuper, J., Geerlof, A., von Kries, J.P., and Wilmanns, M. (2011) Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc Natl Acad Sci USA* 108: 3554–3559.

Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res* 32: 1792–1797.

Eliot, A.C., and Kirsch, J.F. (2004) Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem* 73: 383–415.

Esaki, N., and Walsh, C.T. (1986) Biosynthetic alanine racemase of *Salmonella typhimurium*: Purification and characterization of the enzyme encoded by the *alr* gene. *Biochemistry* 25: 3261–3267.

Ferla, M.P., and Patrick, W.M. (2014) Bacterial methionine biosynthesis. *Microbiology* 160: 1571–1584.

Ferla, M.P., Thrash, J.C., Giovannoni, S.J., and Patrick, W.M. (2013) New rRNA gene-based phylogenies of the *Alphaproteobacteria* provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS One* 8: e83383.

Fuerst, J.A., and Sagulenko, E. (2011) Beyond the bacterium: Planctomycetes challenge our concepts of microbial structure and function. *Nat Rev Microbiol* 9: 403–413.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucl Acids Res* 43: D261–D269.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309: 1242–1245.

Giovannoni, S.J., Cameron Thrash, J., and Temperton, B. (2014) Implications of streamlining theory for microbial ecology. *ISME J* 8: 1553–1565.

Glasner, M.E., Gerlt, J.A., and Babbitt, P.C. (2006) Evolution of enzyme superfamilies. *Curr Opin Chem Biol* 10: 492–497.

Goudarzi, M., and Born, T.L. (2006) Purification and characterization of *Thermotoga maritima* homoserine transsuclinylase indicates it is a transacylase. *Extremophiles* 10: 469–478.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., et al. (2016) A new view of the tree of life. *Nat Microbiol* 1: 16048.

Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 30: 409–425.
Jeske, O., Schüler, M., Schumann, P., Schneider, A., Boedecker, C., Jogler, M., et al. (2015) Planctomycetes do possess a peptidoglycan cell wall. *Nat Commun* 6: 7116.

Junier, T., and Zdobnov, E.M. (2010) The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 26: 1669–1670.

Khersonsky, O., and Tawfik, D.S. (2010) Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu Rev Biochem* 79: 471–505.

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., et al. (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): Unique resources for biological research. *DNA Res* 12: 291–299.

Latif, H., Lerman, J.A., Portnoy, V.A., Tarasova, Y., Nagarajan, H., Schrimpe-Rutledge, A.C., et al. (2013) The genome organization of *Thermotoga maritima* reflects its lifestyle. *PLoS Genet* 9: e1003485.

Lesley, S.A., Kuhn, P., Godzik, A., Deacon, A.M., Mathews, I., Kreusch, A., et al. (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc Natl Acad Sci USA* 99: 11664–11669.

Letunic, I., and Bork, P. (2011) Interactive Tree Of Life v2: Online annotation and display of phylogenetic trees made easy. *Nucl Acids Res* 39: W475–W478.

Lodha, P.H., Jaworski, A.F., and Aitken, S.M. (2010) Characterization of site-directed mutants of residues R58, R59, D116, W340 and R372 in the active site of *E. coli* cystathionine-β-lyase. *Protein Sci* 19: 383–391.

Lundqvist, T., Fisher, S.L., Kern, G., Folmer, R.H., Xue, Y., Newton, D.T., et al. (2007) Exploitation of structural and regulatory diversity in glutamate racemases. *Nature* 447: 817–822.

Ma, Y., Biava, H., Contestabile, R., Budisa, N., and di Salvo, M.L. (2014) Coupling bioorthogonal chemistries with artificial metabolism: Intracellular biosynthesis of azidohomoalanine and its incorporation into recombinant proteins. *Molecules* 19: 1004–1022.

Matsumura, I., and Ellington, A.D. (2001) *In vitro* evolution of β-glucuronidase into a β-galactosidase proceeds through non-specific intermediates. *J Mol Biol* 305: 331–339.

McCutcheon, J.P., and Moran, N.A. (2011) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10: 13–26.

Nagai, S., and Flavin, M. (1971) Synthesis of O-acetylhomoserine. *Methods Enzymol* 17B: 423–424.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., *et al.* (1999) Evidence for lateral gene transfer between *Archaea* and *Bacteria* from genome sequence of *Thermotoga maritima*. *Nature* 399: 323–329.

Okubo, Y., Yokoigawa, K., Esaki, N., Soda, K., and Kawai, H. (1999) Characterization of psychrophilic alanine racemase from *Bacillus psychrosaccharolyticus*. *Biochem Biophys Res Commun* 256: 330–340.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4: 2411–2423.

Patrick, W.M., Weisner, J., and Blackburn, J.M. (2002) Site-directed mutagenesis of Tyr354 in *Geobacillus stearothermophilus* alanine racemase identifies a role in controlling substrate specificity and a possible role in the evolution of antibiotic resistance. *ChemBioChem* 3: 789–792.

Percudani, R., and Peracchi, A. (2009) The B6 database: A tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. *BMC Biol* 10: 273.

Price, D.R.G., and Wilson, A.C.C. (2014) A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol* 12: 110.

Pyaz, M.A., Conners, S.B., Montero, C.I., Shockley, K.R., Johnson, M.R., Ward, D.E., *et al.* (2004) Transcriptional analysis of biofilm formation processes in the anaerobic, hyperthermophilic bacterium *Thermotoga maritima*. *Appl Environ Microbiol* 70: 6098–6112.

Raboni, S., Spyrakis, F., Campanini, B., Amadasi, A., Bettati, S., Peracchi, A., *et al.* (2010) Pyridoxal 5’-phosphate-dependent enzymes: Catalysis, conformation, and genomics. In Comprehensive Natural Products II: Chemistry and Biology, Townsend, C.A., and Ebizuka, Y. (eds). Kidlington, UK: Elsevier Ltd., pp. 273–350.

Reisch, C.R., Moran, M.A., and Whitman, W.B. (2011) Bacterial catabolism of dimethylsulfoxonipropionate (DMSP). *Front Microbiol* 2: 172.

Rioso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., and Sanchez-Ruiz, J.M. (2013) Hyperstability and substrate promiscuity in laboratory rescues of Precambrian b-lactamases. *J Am Chem Soc* 135: 2899–2902.

Sayf, R.F., and Fuchs, G. (2010) Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* 464: 1077–1081.

Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucl Acids Res* 37: D5–D15.

Schleifer, K.H., and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacterial Rev* 36: 407–477.

Somero, G.N. (1995) Proteins and temperature. *Annu Rev Physiol* 57: 43–68.

Soo, V.W.C., Yoshitadada, Y., Squire, C.J., and Patrick, W.M. (2016) Mechanistic and evolutionary insights from the reciprocal promiscuity of two pyridoxal phosphate-dependent enzymes. *J Biol Chem* 291: 19873–19887.

Stamatakis, A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Tanner, M.E. (2002) Understanding nature’s strategies for enzyme-catalyzed racemization and epimerization. *Acc Chem Res* 35: 237–246.

Tawfik, D.S., and van der Donk, W.A. (2016) Biocatalysis and biotransformation: Esoteric, niche enzymology. *Curr Opin Chem Biol* 31: v–vii.

Tripp, H.J., Kitner, J.B., Schwabach, M.S., Dacey, J.W., Wilhelm, L.J., and Giovannoni, S.J. (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* 452: 741–744.

© 2017 The Authors. Molecular Microbiology Published by John Wiley & Sons Ltd., *Molecular Microbiology*, 105, 508–524
van Teeseling, M.C., Mesman, R.J., Kuru, E., Espaillat, A., Cava, F., Brun, Y.V., et al. (2015) Anammox Planctomycetes have a peptidoglycan cell wall. Nat Commun 6: 6878.

Vollmer, W., Bivanot, D., and de Pedro, M.A. (2008) Peptidoglycan structure and architecture. FEMS Microbiol Rev 32: 149–167.

Vollmer, J., Schiefer, A., Schneider, T., Jülicher, K., Johnston, K.L., Taylor, M.J., et al. (2013) Requirement of lipid II biosynthesis for cell division in cell wall-less Wolbachia, endobacteria of arthropods and filarial nematodes. Int J Med Microbiol 303: 140–149.

Voordeckers, K., Brown, C.A., Vanneste, K., van der Zande, E., Voet, A., Maere, S., et al. (2012) Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. PLoS Biol 10: e1001446.

Wol, Y.I., and Koonin, E.V. (2013) Genome reduction as the dominant mode of evolution. Bioessays 35: 829–837.

Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., et al. (2004) Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: A streamlined genome overrun by mobile genetic elements. PLoS Biol 2: e69.

Wu, D., Hu, T., Zhang, L., Chen, J., Du, J., Ding, J., et al. (2008) Residues Asp164 and Glu165 at the substrate entryway function potently in substrate orientation of alanine racemase from E. coli: enzymatic characterization with crystal structure analysis. Protein Sci 17: 1066–1076.

Ycˇas, M. (1974) On earlier states of the biochemical system. J Theor Biol 44: 145–160.

Yokoigawa, K., Kawai, H., Endo, K., Lim, Y.H., Esaki, N., and Soda, K. (1993) Thermolabile alanine racemase from a psychrotroph, Pseudomonas fluorescens: purification and properties. Biosci Biotechnol Biochem 57: 93–97.

Yokoigawa, K., Hirata, K., Ueno, H., Okubo, Y., Umesako, S., and Soda, K. (2001) Gene cloning and characterization of alanine racemases from Shigella dysenteriae, Shigella boydii, Shigella flexneri, and Shigella sonnei. Biochem Biophys Res Commun 288: 676–684.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher’s web-site.