An L-threonine transaldolase is required for L-threo-β-hydroxy-α-amino acid assembly during obafluorin biosynthesis

Thomas A. Scott¹, Daniel Heine¹, Zhiwei Qin¹ & Barrie Wilkinson¹

β-Lactone natural products occur infrequently in nature but possess a variety of potent and valuable biological activities. They are commonly derived from β-hydroxy-α-amino acids, which are themselves valuable chiral building blocks for chemical synthesis and precursors to numerous important medicines. However, despite a number of excellent synthetic methods for their asymmetric synthesis, few effective enzymatic tools exist for their preparation. Here we report cloning of the biosynthetic gene cluster for the β-lactone antibiotic obafluorin and delineate its biosynthetic pathway. We identify a nonribosomal peptide synthetase with an unusual domain architecture and an L-threonine:4-nitrophenylacetaldehyde transaldolase responsible for (2S,3R)-2-amino-3-hydroxy-4-(4-nitrophenyl)butanoate biosynthesis. Phylogenetic analysis sheds light on the evolutionary origin of this rare enzyme family and identifies further gene clusters encoding L-threonine transaldolases. We also present preliminary data suggesting that L-threonine transaldolases might be useful for the preparation of L-threo-β-hydroxy-α-amino acids.

¹Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. Correspondence and requests for materials should be addressed to B.W. (email: barrie.wilkinson@jic.ac.uk).
Despite the great structural diversity of biologically important natural product scaffolds, a number of privileged structural motifs occur within them with high frequency. An important example is found in the \( \beta \)-hydroxy-\( \alpha \)-amino acids which are constituents and intermediates in the biosynthesis of many agriculturally and medicinally valuable bioactive natural products. These include the antibiotics chloramphenicol\(^1\) and vancomycin\(^2\), sphingofugin antifungal agents\(^3\) and the proteasome inhibitor salinosporamide\(^4\) among others. They are also of great importance as chiral building blocks for chemical synthesis and 1,3,4-\( \beta \)-hydroxyphenylserine (Droxidopa) is itself a (pro)drug used in the treatment of Parkinson’s disease\(^5\). Although many different synthetic methods have been reported for their chemical synthesis—including methods based on glycine enolates\(^6\), glycaminides\(^7\), glycine Schiff’s base\(^8\) and the aminohydroxylation of olefins\(^9\)—the production of \( \beta \)-hydroxy-\( \alpha \)-amino acids by enzymatic means is particularly attractive as these can set two stereocentres in a single reaction that can be performed in a one-pot process with minimal protection of substrates and under mild, aqueous conditions. One such example is found in the threonine aldolases (TAs) although their utility is limited due to low synthetic yields and modest diastereoselectivity (they are highly stereoselective for the \( \alpha \)-carbon)\(^10–12\). As such the discovery and characterization of new enzymes with utility for the synthesis of \( \beta \)-hydroxy-\( \alpha \)-amino acids is desirable to expand our toolkit for asymmetric synthesis. Given the presence of \( \beta \)-hydroxy-\( \alpha \)-amino acid derived moieties in many natural product structures we decided to target biosynthetic gene clusters (BGCs) encoding for the production of such molecules.

On this basis the antibacterial agent obafluorin\(^13,14\) (1; Fig. 1) produced by Pseudomonas fluorescens ATCC 39502 attracted our attention as classical feeding experiments suggested its \( \beta \)-lactone moiety derives from cyclization of an unusual \( \beta \)-hydroxy-\( \alpha \)-amino acid intermediate (2S,3R)-2-amino-3-hydroxy-4-(4-nitrophenyl)butanoate (2) via homologation of L-4-aminophenylalanine (3)\(^15–18\). Previously, Herbert and Knaggs proposed a biosynthetic pathway in which 2 derives from 4-aminophenylpyruvate (4; derived from transamination of 3) and glyoxylate via a decarboxylative, thiamine diphosphate (ThDP)-dependent mechanism as shown in Supplementary Fig. 1 (ref. 16). If correct, such an enzyme, when coupled with appropriate aminotransferases, would provide access to valuable substituted \( \beta \)-hydroxy-\( \alpha \)-amino- and \( \alpha \)-hydroxy-\( \beta \)-amino acids in enantiomerically enriched form.

Moreover, the biosynthesis of 1 is of additional interest as the \( \beta \)-lactone structural motif occurs infrequently in nature but is usually associated with molecules possessing potent and valuable biological activity\(^19,20\). For example, 1 exhibits narrow spectrum antibacterial activity\(^13,14\), despite being a substrate for \( \beta \)-lactamases and sensitive to acidic conditions, undergoing facile hydrolysis or ring opening in the presence of nucleophiles. It protects mice infected with a clinical isolate of Streptococcus pyogenes when dosed systemically, and microscopic examination of Escherichia coli grown at sub-lethal doses showed unusual cell elongation\(^14\). Taken together these observations suggest that 1 acts in a specific manner rather than as a general acylating agent (\( \beta \)-lactones are generally effective electrophiles able to form reversible covalent linkages with nucleophilic residues of target proteins). The molecular target of 1 remains elusive making studies on its biological activity and biosynthesis of particular interest: in an era of extensive antimicrobial resistance the identification of new antibacterial targets is of great potential significance.

Using genome mining, in conjunction with mutational and biochemical analysis, we report the pathway for the biosynthesis of 1 and show that the BGC includes genes involved in the biosynthesis of the precursors 4 and 2,3-dihydroxybenzoic acid (5). These co-localize with additional genes including those encoding for a ThDP-dependent phenylpyruvate decarboxylase (pPDC) ObaH and a putative serine hydroxymethyltransferase (SHMT)/L-TA ObaG, both of which are essential for the biosynthesis of 2 and therefore, ultimately, 1. Our data show that ObaG actually encodes for an L-threonine transaldolase (L-TTA), and we propose that this group of enzymes may offer utility as tools for the synthesis of L-\( \beta \)-hydroxy-\( \alpha \)-amino acids.

**Results**

**Identification of the 1 BGC.** The genome of P. fluorescens ATCC 39502 was sequenced at the Earlham Institute (Norwich, UK) using the Pacific Biosciences (PacBio) RSII platform, and assembly using the HGAP2 pipeline gave a single circular contig of \( \sim 6.15 \) Mb. The sequence was submitted to the open access genome mining platform antiSMASH 3.0 (ref. 21) which predicted sixteen BGCs, seven of which included nonribosomal peptide synthetase (NRPS) encoding genes. Functional assignments were made for each gene by comparing the deduced amino acid sequence with proteins of known function which allowed us to identify one BGC encoding the activities we anticipated would be required for the biosynthesis of 1 (Fig. 1; Supplementary Table 1). The oba BGC sequence has been deposited with GenBank under the accession no. KX931446.

To verify the oba locus we subjected the BGC to mutational analysis (see below). To facilitate genetic modification we generated pTS1 (Supplementary Fig. 2 and Supplementary Table 2; GenBank accession no. KX931445), a variant of the suicide vector pME3087 (ref. 22) which was first sequenced de novo by primer walking (Supplementary Table 2; GenBank accession no. KX931444). pTS1 incorporates an expanded multiple cloning site and the sacB gene which allows for positive selection of gene replacements after secondary recombination and plasmid loss\(^23\) (see Supplementary Note 1 for construction). Gene inactivation experiments were performed by generating in frame deletions using pTS1, and all resulting mutants were verified by PCR amplification across the deleted genomic region followed by sequencing of the amplicon; at least three independent mutants were checked. All mutants were functionally confirmed and checked for the lack of polar effects by genetic complementation through ectopic expression of the deleted gene under the control of the tac promoter using the expression plasmid pH10TS, based on pH10 previously reported by Thomas and co-workers\(^24\) (Supplementary Note 2 and Supplementary Table 2). As required, chemical complementation with putative pathway intermediates was carried out in parallel by the addition of exogenous material to growing cultures. All strains and their complemented derivatives were grown in triplicate under 1 producing conditions (alongside the wild type (WT) strain), extracted with ethyl acetate and the extracts subjected to high performance liquid chromatography (HPLC) and high performance liquid chromatography-mass spectrometry (HPLC-MS) analysis.

**Bioinformatics and mutational analysis of the 1 BGC.** Bioinformatics analysis suggests that three genes (obaJLN) encode for the well understood pathway to 5 from chorismate\(^25,26\) and deletion of either obaJ (an isochorismatase) (Supplementary Fig. 3) or obaL, a 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (Fig. 2a) led to the loss of 1 production and 4 to 5-fold elevated levels of 4-nitrophenylethanol (6) and 4-nitrophenylacetic acid (7), which are proposed shunt metabolites of the biosynthesis of 2 (ref. 17). The addition of exogenous 5 was able to restore the biosynthesis of 1 (to 71% (ΔobaJ) and 68% (ΔobaL) of WT titres), as was...
**Figure 1** | Gene cluster architecture and proposed biosynthetic pathway for obafluorin biosynthesis in *P. fluorescens* ATCC 39502. Obafluorin protein coding sequences *obaA*-N are labelled and coloured according to related biosynthetic functions, with operons indicated with black arrows beneath the gene cluster. The domain architectures of the aryl carrier protein ObaK and nonribosomal peptide synthetase Obal are represented (ArCP = Aryl carrier protein; C = Condensation domain; A = Adenylation domain; PCP = Peptidyl carrier protein domain; TE = Thioesterase domain; MbtH = MbtH-like domain), and the anticipated mechanism of TE-catalysed peptide release and β-lactone ring closure is also illustrated. 6 and 7 represent breakdown products of 10, and were found to accumulate when either the biosynthesis of 5, or the assembly of 1, was disrupted. Compounds 3 and 8 were used in chemical feeding experiments in this work, but were not shown to be intermediates in the biosynthesis of 1.

**Figure 2** | HPLC profiles for mutagenesis and complementation experiments in selected *oba* genes. (a) *obaL*, (b) *obaF*, (c) *obaH* and (d) *obaG*. Numbered peaks refer to key products and shunt metabolites identified in Fig. 1.
ectopic expression of the deleted gene (to 73% (obaI) and 85% (obaI) of WT titres). Moreover, by blocking production of the key intermediate 5 (and therefore the pathway endpoint 1), we anticipated that the predicted pathway intermediate 2 (refs 16, 17) might accumulate (see Fig. 1). However, after exhaustive LCMS analysis we could not observe any accumulation of 2, nor of the potentially related pathway intermediate (2-amino-3-hydroxy-4-(4-aminophenyl)butanoate (8)). Synthetic standards of both 2 and 8 were prepared to aid this analysis (see Materials and Methods).

The three genes (obaDEF) appear sufficient to encode for the production of 4, the biosynthesis of which is well understood from previous genetic studies on chloramphenicol27 and pristinamycin28 biosynthesis. Consistent with these predictions, deletion of the bifunctional gene obaF (a bifunctional 4-aminochorismate mutase/4-amino prephenate dehydrogenase) led to the loss of 1, 6 and 7 production (Fig. 2b). Addition of exogenous 3 was sufficient to restore efficient production of 1 (between 68%–71% of WT titres), as was the ectopic expression of obaF (between 73 and 90% of WT titres). We presume that 3 is converted to 4 by transamination in vivo. Surprisingly, no accumulation of the pathway intermediate 5 could be observed in the obaF mutant.

A putative type II 3-deoxy-d-arabino-heptulosonate-7-phosphate (DAHP) synthase is encoded by obaM. Homologues of this gene have been observed in a number of bacterial BGCs,28,29 where their products are believed to mitigate the aromatic amino acid-based negative feedback of primary metabolic DAHP synthases and drive flux into specialized metabolism. This indicated a likely role for ObaM in chorismate supply for the biosynthesis of both 2 and 5. Consistent with this deletion of obaM led to the almost complete abolition of 1 biosynthesis (Supplementary Fig. 4), but we were surprised to observe the accumulation of 6 and 7. Moreover, exogenous supply of 5 alone was sufficient to restore production of 1 (68–72% of WT titres), further indicating that biosynthesis of 2 is unperturbed by obaM deletion, which appears instead to only impact 5 biosynthesis. Consistent with this, exogenous supply of 3 alone could not recover 1 production but did lead to further elevated levels (five- to sevenfold) of 6 and 7. Ectopic expression of obaM recovered 1 production (50–55% of WT titres), precluding the possibility of polar effects on obaJLN.

The gene obaC encodes a putative non-heme di-iron monooxygenase related to AurF from the aurithrin pathway (73% query cover, 25% identity)30,31 and ClmL from the chloramphenicol pathway (72% query cover, 19% identity)32,33. Both enzymes are arylamine oxygenases which catalyse the conversion of an aromatic amine to a nitro group, and we propose ObaC catalyses the conversion of 4 to 4-nitrophenylpyruvate (9). Consistent with this role, deletion of obaC abolished production of 1, 6 and 7 (Supplementary Fig. 5). The biosynthesis of all three compounds was re-established after ectopic complementation with obaC (81% of WT titres of 1). Surprisingly, only the addition of exogenous synthetic 9 to growing cultures of this mutant was able to restore 1 biosynthesis (to 29% of WT titres), whereas addition of 2 did not. We believe this is likely due to failure of this metabolite to penetrate the Pseudomonas cell membrane based on subsequent data below.

At the centre of the BGC is the bimodular NRPS-encoding gene obaI, the product of which is predicted to catalyse formation of the amide bond between 2 and 5. The thiosterase (TE) domain of Obal is unusually located between the peptide carrier protein (PCP2) domain and first adenylation (A1) domain, and we predict that it catalyses release of the resulting enzyme bound pseudo-dipeptide with concomitant formation of the β-lactone moiety to yield 1. Analysis of the A domain sequences of Obal using NRSPredictor2 (ref. 34) identified 5 as the likely substrate of A1, whereas L-threonine is predicted to be the substrate of A2. This means that the domain architecture of Obal (Fig. 1) is further unusual as the putative A1 domain is located at the C terminus of the enzyme, rather than towards the N terminus as would be expected. Moreover, A1 is adjacent to an embedded MbtH-like protein domain. MbtH-like proteins are auxiliary proteins required for the activity of many, but not all NRPS A domains.28 Some A domains require MbtH-like proteins for their functional expression, suggesting they may have a chaperone-like function29. Discrete MbtH-like proteins are the most commonly observed, but the identification of MbtH-like domains embedded within an NRPS, as for Obal here, is rare and has been reported for only two other BGCs, those responsible for the biosynthesis of nikkomycin29 and streptolydigin40. While the amino acid residues critical for intra-domain interaction have been well characterized in the latter case, no biochemical data exists to verify that either of these in cis domains are functionally essential in vivo, and the exact role of MbtH-like proteins in NRPS catalysis is yet to be determined. Deletion of obaI abolished production of 1, which was recovered to 79% of WT levels on ectopic complementation (Supplementary Fig. 6). The obaI mutant accumulated the shunt metabolites 6 and 7 at similar levels to the WT strain, but, once again, exhaustive analysis of the fermentation extracts did not show any accumulation of 2 or 5, the putative substrates of Obal.

Obal contains only a single PCP2 domain where two such domains would be expected for its correct function. Further analysis of the BGC suggested that obaK encodes a discrete aryl carrier protein (ArCP), the presence of an ArCP is anticipated as this protein function is required for the tethering of 5 (following activation by the A2 domain) to participate in amide bond formation (Fig. 1). On the basis of these analyses we hypothesize that the unusual positioning of A domains within the assembly line provides access for both to the embedded MbtH-like domain, a likely prerequisite for efficient catalysis. Given the canonical structure and function of NRPS assembly line enzymes, this organization will preclude functional interaction of the A1 domain and the condensation (C2) domain, a consequence of which is the requirement for a discrete, and therefore mobile, ArCP (ObaK) to enable both interaction with the A1 domain for acylation with 5 and subsequent interaction with the C2 domain for peptide bond formation.

Deletion of obaK also abolished production of 1, which was recovered by ectopic expression (to 77% of the WT level) (Supplementary Fig. 7). The titres of 6 and 7 also increased significantly (approximately fourfold) but again neither 2 nor 5 was accumulated. Exogenous supply of 5 did not lead to production of 1 confirming that this phenotype is not linked to a lack of 5 biosynthesis. We queried this possibility as obaK appears to have evolved by the splitting of an ancestral entb-like gene, which usually encode for bidomain proteins involved in the biosynthesis of 5 (refs 25, 26). Indeed, the adjacent obaI gene corresponds to the N-terminal isochorismatase domain of EntB and is required for production of 5 during biosynthesis of 1. The biosynthesis of 5 does not require the ArCP domain of EntB11 consistent with our various results and biosynthetic hypothesis.

Our analysis thus left the biosynthesis of 2 as the remaining unknown, and only the products of obaG and obaH to be functionally assigned. obaG encodes a putative SHMT/T-A like domain, whereas obaH encodes a putative ThDP-dependent PDPC. Mutation of both genes abolished the production of 1, but the accumulation of 6 and 7 was only observed in the obaG mutant (Fig. 2c,d). Again, exogenous addition of 2 did not recover 1 production in these mutants whereas ectopic expression of the deleted genes did (to 62% (obaG) and 29% (obaH) of WT titres). These mutational data are consistent with the role of ObaG as the enzyme directly responsible for 2 production as proposed in Fig. 1, but did not
provide direct support for 2 as a biosynthetic intermediate. In contrast, the accumulation pattern for the shunt metabolites 6 and 7 strongly suggested that 4-nitrophenylacetaldehyde (10) might be a key intermediate. This reactive and potentially toxic molecule is likely to be catabolized to the shunt metabolites 6 and 7 via standard detoxification pathways before accumulating to any degree. Consistent with this hypothesis, addition of exogenous 10 to growing cultures of the ΔobaG mutant did not restore 1 production but did lead to elevated levels of 6 and 7. We also noted that chemical complementation of mutants deficient in 5 biosynthesis are sensitive to the concentration of exogenous material added. Production of 1 was re-established when 5 was added to give a final concentration of 0.2 mM, but at higher concentrations (for example, ≥0.5 mM) both growth of P. fluorescens and biosynthesis of 1 were significantly affected. As noted above we could not observe the presence of 2 or 5 in any of the mutants anticipated to accumulate them. Additionally, in mutants where 2 or its aldehyde precursor 10 might have accumulated, we always observed elevated levels of the shunt metabolites 6 and 7, breakdown products of aldehyde 10, a highly reactive and toxic molecule. Taken together these combined data suggest that the biosynthetic pathway is tightly regulated to avoid the accumulation of intermediates deleterious to cell growth and survival. Interestingly, the intermediary of free phenylacetaldehyde, a reactive aldehyde structurally similar to 10, is avoided in the recently reported ripostatin biosynthetic pathway via the activity of a pyruvate dehydrogenase-like complex which functions to decarboxylate phenylpyruvate and then retain the phenylacetaldehyde product as a phenylacetyl-S-carrier protein species.42 Finally, bioinformatics suggests the obafluorin BGC is regulated by quorum sensing under the control of the luxIR homologues obaAB.

2 is assembled by sequential activity of ObaH and ObaG. The essential nature of obaG and obaH, in conjunction with the accumulation pattern of 6 and 7 by their mutants, led us to hypothesize that the biosynthesis of 2 involves ThDP-dependent decarboxylation of 9 by ObaH to yield the aldehyde 10. This would then be used as a substrate by the pyridoxal-5′-phosphate (PLP) dependent enzyme ObaG acting as an aldolase to give 2. Both enzymes were readily expressed as soluble proteins in hexahistidine-tagged form using E. coli NiCo21(DE3) pLysS. They were purified using standard procedures and their identity verified by tandem MS.

Due to the original proposal of Herbert and Knaggs that a pPDC-like enzyme might produce an acyloin intermediate (Supplementary Fig. 1)6 we first tested ObaH with 9 plus glyoxylate as substrates using a discontinuous format coupled to independent LC/MS/MS and HPLC-UV assays. In none of these reactions could we detect any acyloin products whereas the ketoacid substrate was depleted and aldehyde 10 accumulated whenever ObaH was incubated with 9, clearly identifying it as a 9 decarboxylase (Fig. 3a). To examine the ability of the putative l-TTA-like enzyme ObaG to produce 2, it was incubated with varying concentrations of 10 and glycine using the same assay format as for ObaH. Under none of these conditions could we observe production of 2. At this point we considered the possibility that ObaG might actually be a transaldolase and glycine using the same assay format. Moreover, when [U-13C6,15N2]-l-threonine was used as substrate we were able to demonstrate, by 13C NMR spectroscopy and high-resolution (HR)-LC/MS/MS, the regiospecific transfer of three heavy isotope atoms to the appropriate positions in the product 2, in addition to the formation of [1,2,13C2,15N2]acetaldehyde (Figs 3c and 4). The structure and stereochemistry of isolated 2 was then verified by comparison of its NMR spectra and optical rotation to those of previously synthesized material.43 To confirm the biosynthetic relevance of l-threonine in vivo we fed [U-13C6,15N2]-l-threonine to growing cultures of the WT producer and examined the resulting 1 by HR-LC/MS/MS analysis. This indicated the highly efficient, site specific incorporation of a 13C615N2 unit as anticipated (Supplementary Fig. 8). In a final experiment, we were able to couple the ObaH decarboxylase and ObaG l-threonine transaldolase (l-TTA) reactions in vitro (Fig. 3c,d), to catalyse the formation of 2 using l-threonine and 9 as the only substrates. The proposed mechanism for the biosynthesis of 2 by ObaG is shown in Supplementary Fig. 9.

Having characterized the substrates and function of both ObaH and ObaG, we sought to further characterize ObaG by confirming its identity as a PLP-dependent enzyme and attempted to collect preliminary kinetic data. Despite being identified as a putative SHMT, ObaG shares relatively little sequence similarity to bona fide bacterial SHMTs (25% amino acid sequence identity), but does comprise several residues crucial for PLP-dependent activity, most importantly the lysine residue required for internal aldimine formation (Supplementary Fig. 10). ObaG displays a UV/Vis spectrum (Supplementary Figs 11a and b) characteristic of PLP-dependent proteins,44 comprising absorption maxima at 340 and 390 nm that correspond to the equilibrium established between the enolimine and ketoenamine forms of PLP respectively. ObaG is unusual in that it exhibits a distinct salmon pink colour in solution (Supplementary Fig. 11c), which creates an additional absorbance maximum at 512 nm (Supplementary Figs 11a and b). Treatment of ObaG in the absence of excess PLP with 10 mM L-penicillamine led to the anticipated formation of a thiazolidine adduct (Supplementary Fig. 11d) determined by UV absorbance at 340 nm (Supplementary Fig. 11a), with concomitant loss of the ketoenamine peak at 390 nm, as observed previously by Lowther et al.44 for serine palmitoyltransferase. ObaG was further incubated with NaBH4 (1 mM) resulting in the formation of a peak at 330 nm consistent with the formation of the reduced ObaG-PLP amine adduct (Supplementary Figs 11b and e). In both experiments a decrease in the 512 nm peak was also observed. Taken together these results strongly support ObaG being a PLP-dependent enzyme.

Having characterized the l-TTA reaction catalysed by ObaG, and its PLP-dependence, we acquired preliminary kinetic data. A time course was determined using standard l-TTA assay conditions showing that equilibrium between substrate 10 and product 2 was achieved within ~1 h after reaction initiation (Supplementary Fig. 12a). Unfortunately, the UV/Vis spectrum of 10 precluded the use of a continuous coupled assay based on acetaldehyde accumulation due to overlap with that of the measured side-product, and its inherent reactivity confounded discontinuous methods measuring product formation with varying 10. The best results were achieved by varying l-threonine concentration using a discontinuous HPLC-based approach. Single-substrate kinetic analysis of His6-ObaG revealed typical Michaelis–Menten kinetics with respect to varying l-threonine, yielding kinetic constants of $k_{cat} = 40.2 \pm 3.8$ mM and $K_{m} = 62.9 \pm 1.9$ min$^{-1}$ (Supplementary Fig. 12b).

**Preliminary investigation of the NRPS ObaI.** In a final biochemical experiment we were able to express the NRPS ObaI as the full length (211.5 kDa) protein in its apo-form and probe the ability of its two A domains to activate various amino
acids using a discontinuous hydroxamate formation assay \(^{46}\) (Methods section). While this approach did not allow us to dissect the activities of the two A domains independently, it did suggest that 2 is the preferred substrate for the A\(_2\) domain, but that it displays relaxed substrate specificity with activation of \(\varepsilon\)-lysine and \(\varepsilon\)-lysine being observed (Supplementary Fig. 13). It should be noted that NRPSpredictor2 (ref. 34) identified \(\varepsilon\)-lysine as the substrate for the A\(_2\) domain. The A\(_1\) domain, which is predicted to activate 5, also appears to have relaxed specificity as benzoic acid was activated in addition to 5. Despite the apparent flexibility of both A domains we could not observe the production of obafluorin congeners in the WT strain, although such molecules might be sensitive to proteolytic degradation. Most likely, and consistent with the observations of others, while the A domains of intact Obal show an inherent substrate flexibility, the C\(_2\) domain is likely to exert a gatekeeper function with regards to the selection of substrates for carrier protein acylation and subsequent amide bond formation \(^{47}\).

**Phylogenetic analysis of \(\varepsilon\)-TTAs.** \(\varepsilon\)-TTA activity like that of Obal here has previously been observed only in the biosynthesis of 4-fluoro-\(\varepsilon\)-lysine \(^{48}\) and in 5'-C-glycyluridine (GlyU), a key constituent of lipopeptidyl nucleoside natural products \(^{49}\) (Supplementary Fig. 14). Both 4-fluoro-\(\varepsilon\)-lysine transaldolase (FTase) and \(\varepsilon\)-lysine:uridine-5'-transaldolase of the LipK family were all identified as putative SHMTs, but have been shown to catalyse \(\varepsilon\)-lysine \(\beta\)-substitutions via the sequential breaking and forming of \(\text{Ca} - \text{C} \beta\) bonds. We have now undertaken detailed phylogenetic analyses of PLP-dependent SHMTs, \(\varepsilon\)-TAs and the \(\varepsilon\)-TTAs described here (Fig. 5) \(^{50}\). This reveals that bacterial \(\varepsilon\)-TTAs and SHMTs diverged from each other to form discrete clades, and that they and the TAs diverged from an earlier common ancestor. It is particularly notable that the common ancestor appears to have diverged into enzymes that all utilize \(\varepsilon\)-lysine, other than the SHMTs. While the evolutionary relationship between Obal and FTase is not so easily determined when comparing different phylogenies (Supplementary Figs 15 and 16), what is clear is that they are distinct from the \(\varepsilon\)-lysine:uridine-5'-transaldolase (LipK) clade. In addition, OrfA—identified as a glycine hydroxymethyltransferase and proposed as a biosynthetic enzyme required for the biosynthesis of alanylclavam in the cephamycin/clavulanic acid/5S-clavam producer *Streptomyces clavuligerus* \(^{51}\)—also consistently grouped within the \(\varepsilon\)-TTA clade. The biochemical characterization of this enzyme has not yet been reported, but based on our analysis we suspect that it has been incorrectly annotated and is in fact an \(\varepsilon\)-TTA using \(\varepsilon\)-lysine as a substrate rather than an \(\varepsilon\)-TA using glycine as previously proposed (Supplementary Fig. 14).
Discussion

On the basis of our bioinformatics analysis combined with mutational, biochemical and chemical feeding results we propose the biosynthetic pathway for 1 shown in Fig. 1. The pathway involves a unique, non-canonical NRPS which carries an integrated MbtH-like protein domain and contains a central TE domain which is likely to be responsible for β-lactone formation. Central to the pathway, however, is a new route for

Figure 4 | Characterization of the biosynthesis of 2 by ObAG using $^{13}$C NMR. (a) $^{13}$C NMR spectrum of [U-$^{13}$C$_4$,15N] L-threonine in methanol-d$_4$. (b) $^{13}$C NMR spectrum of the synthetic reference 2. (c) $^{13}$C NMR spectrum of enzymatically synthesized 10. Signals marked with X correspond to the two acetal forms of acetaldehyde and glycerol, being formed after release of acetaldehyde and therefore proving threonine L-TTA activity of ObAG.

Figure 5 | Maximum likelihood tree of ε-TAs, SHMTs and ε-TTAs. A set of 3-amino-5-hydroxybenzoate synthase (AHBAS) amino acid sequences serve as the outgroup. RAxML likelihood values at the root nodes for SHMT and ε-TA cluster clades are annotated. The clade comprising the ε-TTA ObAG characterized in this work has been expanded and annotated with RAxML likelihood values and the identity of amino acid sequences represented.
biosynthesis of the homologated 1-threo-\(\beta\)-hydroxy-\(\alpha\)-amino acid \(\Delta\) via sequential decarboxylation and transaldolase reactions. \(\Delta\)-TTA reactions are rare and have been reported previously only for the biosynthesis of 4-fluoro-\(L\)-threonine and Gly\(\U{1D480}\) from fluoroacetalddehyde\(^{48}\) and uridine-5'-aldehyde\(^{49}\), respectively (Supplementary Fig. 14), both of which exhibit the \(L\)-threo stereochemistry of \(\Delta\). Detailed phylogenetic analysis has shed light on the evolutionary origins of the \(L\)-TA (which use glycine as substrate) and \(L\)-TTA families (Fig. 5), and suggests that the two families diverged from a common ancestor with the transaldolase family subsequently splitting again into lineages that specifically utilize either \(L\)-TTAs or \(L\)-serine (SHMTs) as substrates. With this knowledge in hand we used a phylogeny-based, mechanism-guided genome mining approach to identify BGCs containing \(L\)-TTAs, and therefore biosynthetic pathways likely involving a \(\beta\)-hydroxy-\(\alpha\)-amino acid intermediate. This led to the identification of a putative \(L\)-TTA involved in the biosynthesis of a 5S-clavam metabolite, indicating phylogenetic analysis as a potentially general genome mining approach for the identification of further \(L\)-TTAs.

We also identified multiple BGCs likely to encode for the production of \(\Delta\) or very closely related compounds in the genomes of several \textit{Pseudomonas} environmental isolates\(^{-}\). \textit{Borrelia} strains (Genome sequences for \textit{Borrelia} species containing \textit{oba} like BGCs: \textit{Borrelia} bacterialis (NZ_LZGD01000041.1); \textit{Borrelia} diffusa (NZ_LOTC01000033.1), \textit{Borrelia} territii (NZ_LOSY01000044.1), \textit{Borrelia} ubonensis (NZ_CP0013463.1.) and a \textit{Chitiniphilus} strain (Genome sequences for \textit{Chitiniphilus} species containing \textit{oba} like BGCs (GenBank Accession: NZ_KBB895358.1). The relatively common occurrence of the \(\Delta\) BGC is consistent with the original report describing its discovery which found that numerous \textit{Pseudomonas} isolates, from a range of geographical locations, were able to produce \(\Delta\). In several of the sequenced genomes, but not all, a set of genes encoding a glycine cleavage system\(^2\) is located adjacent to the putative biosynthetic genes. We also identified an apparently intact glycine cleavage system (Supplementary Table 1) adjacent to the BGC for \(\Delta\).

The utility of \(\Delta\)-TAs for the enzymatic synthesis of \(L\)-threo-\(\beta\)-hydroxy-\(\alpha\)-amino acids from an aldehyde and glycine has been extensively investigated, but is limited due to low synthetic yields and modest diastereoselectivity\(^{10,11}\). The low yields are in part due to the reversible nature of the reaction (retroaldol cleavage of \(L\)-threo-\(\beta\)-hydroxy-\(\alpha\)-amino acids) and an equilibrium which favours aldehyde and glycine. Due to this synthetic reactions are commonly run in the presence of an excess of glycine to shift the equilibrium to the aldol side. The highly efficient reversible nature of this reaction has, however, been harnessed for the resolution of racemic mixtures to provide \(d\)-threo-isomers\(^{54}\).

In contrast Oba\(G\) gives \(\Delta\) as a single stereoisomer in a excellent 55–59% yield based on accumulation of \(\Delta\) (Supplementary Fig. 12a) without the need for an excess of \(L\)-threonine. On the basis of this work we undertook preliminary investigations into the ability of Oba\(G\) to accept the alternative aldolase substrates phenylacetaldehyde and benzaldehyde (Supplementary Figs 17a and b). Incubation with phenylacetaldehyde led to a single product in excellent yield (\(\sim 45\%\), based on consumption of starting material) whose LCMS profile was consistent with the expected product \(\textit{2S}\)-amino-(3R)-hydroxy-4-phenylbutanoate. The reaction with benzaldehyde was less efficient leading to two products in a 1:2 ratio and poor overall yield (\(< 20\%\)). The LCMS profile of these was consistent with the production of both \(L\)-threo and \(L\)-erythro-phenylserine, respectively (an authentic standard of the \(\textit{threo}\) diastereoisomer was utilized). These data suggest that \(L\)-TTAs may offer an alternative to \(\Delta\)-TAs for the synthesis of \(L\)-threo-\(\beta\)-hydroxy-\(\alpha\)-amino acids. The equilibrium appears to lie in the favour of the aldol reaction and might be improved further through the use of excess \(L\)-threonine and the addition of glyceraldehyde which reacts preferentially with acetaldehyde, shifting the equilibrium further in the direction of the target molecule. This is consistent with our observation that Oba\(G\) is a poor catalyst for the reverse transaldol reaction when incubated with acetaldehyde and racemic \(\Delta\) made synthetically (Supplementary Fig. 17c). Moreover, our preliminary data suggest that there is a potential for high diastereoselectivity depending on the substrate aldehyde which is utilized. Future work in our lab will include additional biochemical and kinetic analysis of Oba\(G\), and a more comprehensive investigation of its potential utility for the synthesis of enantiomerically pure \(L\)-threo-\(\beta\)-hydroxy-\(\alpha\)-amino acids.

In summary, we have delineated the entire biosynthetic pathway to \(\Delta\), and identified a distinct enzyme family for the synthesis of \(\beta\)-hydroxy-\(\alpha\)-amino acids, privileged chiral building blocks in a variety of pharmacologically and agriculturally important natural products and medicines. \(L\)-TTAs perform a unique transformation, incorporating two stereocentres in one overall step, with high diastereoselectivity and excellent yields. Thorough phylogenetic investigation revealed that the \(L\)-TTAs form their own discrete evolutionary lineage, distinct from \(\Delta\)-TAs and SHMTs, shedding light on their own evolutionary history.

**Materials**

**General.** Reagents and chemicals were purchased from Alfa-Aesar, Sigma-Aldrich, Santa Cruz Biotechnology, Inc., Amatek Chemical Co., Ltd., and BD Biosciences, and were used without further purification. All strains used in this study are listed in Supplementary Table 2. \textit{Pseudomonas fluorescens} ATCC 39502 was purchased from the American Type Culture Collection (ATCC). All strains were maintained on solid Luria-Bertani medium (with appropriate selection) at 37 °C (\(E. coli\)) or 28 °C (\(P. fluorescens\)). All solvents used for HPLC were obtained from Fisher Scientific at least of HPLC grade and were filtered before use.

**Instrumentation.** NMR spectra were recorded on a Bruker AVANCE III 400 MHz spectrometer. Chemical shifts are reported in parts per million (p.p.m.) relative to the solvent residual peak of chloroform-\(d_2\) (\(\delta\): 7.24 p.p.m., singlet; \(\delta\): 4.69 p.p.m., singlet; \(\delta\): 77.00 p.p.m., triplet), methanol-\(d_4\) (\(\delta\): 3.30 p.p.m., quintet; \(\delta\): 49.00 p.p.m., septet) or water-\(d_2\) (\(\delta\): 4.79 p.p.m., singlet). Semi-preparative and analytical HPLC was performed using an 1100 system (Agilent Technologies). Extracted samples from mutagenesis and complementation experiments were analysed using a Gemini 3 \(\mu\)m X-C18 110 Å, 150 \(\times\) 4.6 mm column (Phenomenex) with a gradient elution: MeCN/0.1% (v/v) TFA (H\(\text{O}\)) gradient from 10/90 to 90/10 for 15–16 min, gradient to 90/10–16.50 min and 90/10 for 16.50–20 min; flow rate 1 ml min\(^{-1}\); injection volume 10 μl. Biochemical assays were analysed using a Synergy 4 \(\mu\)m Fusion-RP 80 Å LC column 250 ∗ 10 mm with a gradient elution: MeOH/0/1% (v/v) TFA (H\(\text{O}\)) gradient from 10/90 to 100/0 0–14 min, 100/0 for 10–19 min; gradient to 10/90–18.50 min and 10/90 for 18.50–23.50 min; flow rate 1 ml min\(^{-1}\); injection volume 5 μl (15 μl for the reverse reaction experiments). UPLC-MS measurements were performed on a Nexera X2 liquid chromatograph (LC-30AD) LCMS system (Shimadzu) connected to an autosampler (SILO-30AC), a Prominence column oven (CTO-20AC) and a Prominence photo diode array detector (SPD-M20A). A Kinetex 2.6 \(\mu\)m C18 100 Å, 100 ∗ 2.1 mm column (Phenomenex) was used for LCMS. The UPLC-System was connected with a LCMS-IT-TOF Liquid Chromatograph mass spectrometer (Shimadzu). Solid phase extraction was carried out using Discovery DSC-18 SPE Tubes, filled with 1,000 mg of octadecyl-modified, endcapped silica gel (Supelco). The specific optical rotation of compounds was measured with a Model 341 Polarimeter (PerkinElmer, Inc.).

**Cloning.** All primers used in this study are reported in Supplementary Table 3. \(\Delta\)oBa strains were generated using the suicide vector pTS1, constructed as part of this work (Supplementary Note 1). Primers were designed to amplify 800–1,200 bp flanking regions of a selected \(oba\) protein coding sequences (PCSs) for cloning into pTS1 between XbaI and AvrII sites, and AvrII and BmtI sites. Flanking regions were designed to comprise 10–50 PCS codons at either end of the \(oba\) gene to be deleted to minimize polar effects, leaving a truncated chromosomal copy of the gene with an in-frame deletion and internal AvrII site cloning artefact following double homologous recombination. For genetic complementation, WT copies of \(oba\) PCSs were cloned from start to stop as BmtI-KpnI/XbaI fragments into pHK103 (Supplementary Note 2) for introduction and ectopic expression in the relevant mutant strains. For protein purification and expression, WT PCSs for \(obaG-I\) were cloned as NdeI–Xhol fragments into pET28a (+) for expression in \(E. coli\) NovBro13(DE3) (NEB) pLySs.
Mutagenesis and complementation experiments. pTSI knockout constructs were introduced into P. fluorescens ATCC 39502 via conjugation from E. coli S17-1 λpir and λpir+ conjugants were selected on LB plates containing 20 μg/ml of tetracycline and washed in 5 CV of their respective buffers containing 10, 20, 30 and 50 mM binding proteins. Eluted sample was loaded onto a HisTrap excel (GE Healthcare) chitin resin with gentle mixing for 30 min to remove any endogenous protein identity, using an autoflex Speed MALDI-TOF/TOF mass spectrometer bands were subsequently cut out, washed, reduced, alkylated and treated with 5% trypsin according to standard procedures adapted from Shevchenko et al. The tryptic peptide fragments were analysed by mass spectrometry to further confirm protein identity, using an autocut Speed MALDI-TOF/TOF mass spectrometer (Bruker Daltonics GmbH).

In vitro His<sub>6</sub>-ObaG discontinuous t-TTA activity assay. Reactions were performed in 100 μl reaction volumes comprising 50 μM His<sub>6</sub>-ObaG, 50 μM His<sub>6</sub>-ObaH, 10 mM MgCl<sub>2</sub> and 15 mM MgCl<sub>2</sub> using an Aminicon column (30kDa MWCO). Reactions were performed in 600 μl reaction volumes comprising 20 μM His<sub>6</sub>-ObaG and were initiated by addition of t-penicillamine (dissolved in reaction buffer) to a final concentration of 10 mM. Ultraviolet–visible spectra were recorded over 30 min using a Lambda 35 UV/Visible Spectrophotometer (PerkinElmer).

A second spectrophotometric assay was performed using His<sub>6</sub>-ObaG samples with excess PLP removed in which protein was treated with NaBH₄ to reduce the His<sub>6</sub>-ObaG-PLP aldime to form an amine adduct. Reactions were performed on ice for 15 min in 600 μl volumes comprising 20 μM His<sub>6</sub>-ObaG and were initiated by addition of NaBH₄ (dissolved in reaction buffer) to a final concentration of 1 mM. Samples were assayed using UV/Visible spectrophotometry.

Chemical synthesis and isolation of substrates. For the isolation of 1 and its methanolation product, the supernatant from 61 of culture broth was subjected to the liquid–liquid partition using equal volume of ethyl acetate. The organic fraction was concentrated under reduced pressure and the residue dissolved in MeCN. This was subjected to preparative reversed-phase HPLC (C18, 50 x 21.2 mm, 110 μm Phenomenex; A: water; B: MeCN, gradient 0-5 min 10% B (v/v), 5-35 min, 100-100% B (v/v), 35-40 min 100% B (v/v), 40-41 min 100% B (v/v), 41-45 min 10% B (v/v); flow rate was 20 min-1), monitored at 250 nm. The fractions containing 1 and its methanolation product were combined and further purified by preparative reversed phase HPLC (C18, 150 x 10 mm, 110, ΑΝΩΝ Phenomenex; A: water; B: MeCN; isocratic 50% B (v/v), 35-40 min 100% B (v/v), 40-41 min 100% B (v/v), 41-45 min 10% B (v/v); flow rate was 20 min-1), monitored at 250 nm. The fractions containing 1 and its methanolation product were combined and further purified by preparative reversed phase HPLC (C18, 50 x 21.2 mm, 110 μm Phenomenex; A: water; B: MeCN; isocratic 50% B (v/v)), 1: Pale yellow solid, (13 mg). 1H NMR (400 MHz, MeCN-d₃) δ 8.23 (1H, d, J = 8.8 Hz), 8.10 (2H, d, J = 8.8 Hz), 7.45 (2H, d, J = 8.8 Hz), 7.20 (1H, dd, J₁ = 8.2 Hz, J₂ = 1.27 Hz), 7.04 (1H, dd, J₁ = 7.9 Hz, J₂ = 1.2 Hz), 6.82 (1H, t, J = 7.96 Hz), 5.75 (1H, dd, J₁ = 8.5 Hz, J₂ = 6.21 Hz), 5.05 (1H, m, 3.38 (1H, dd, J₁ = 15.1 Hz, J₂ = 5.1 Hz), 3.21 (1H, dd, J₁ = 15.1 Hz, J₂ = 5.1 Hz); 13C NMR (100 MHz, MeCN-d₃) δ 171.47, 169.11, 150.43, 148.08, 147.02, 145.41, 131.20, 124.64, 120.50, 120.07, 118.68, 114.86, 78.50, 59.87, 36.16 p.p.m. HRMS (m/z): [M + H]⁺ calculated for C₉H₁₀NO₇, 183.0874, found, 183.0872.

PLP-dependence of His<sub>6</sub>-ObaG. Recombinant His<sub>6</sub>-ObaG was incubated with t-penicillamine. Excess PLP was removed from His<sub>6</sub>-ObaG samples by exchanging into a reaction buffer comprising 25 mM HEPES pH 7.8, 300 mM NaCl, and 15 mM MgCl₂ using an Aminicon column (30kDa MWCO). Reactions were performed in 600 μl reaction volumes comprising 20 μM His<sub>6</sub>-ObaG and were initiated by addition of t-penicillamine (dissolved in reaction buffer) to a final concentration of 10 mM. Ultraviolet–visible spectra were recorded over 30 min using a Lambda 35 UV/Visible Spectrophotometer (PerkinElmer).

Single-substrate kinetic analysis was carried out by performing the t-TTA activity assay with varying concentrations of t-threonine (1–200 mM). Reactions were performed using 25 mM enzyme and were incubated at 27°C, 700 r.p.m. for 4 min, before quenching, sample processing and HPLC analysis as described previously. Five replicates were performed for each concentration of t-threonine assayed, and data were fitted to the Michaelis–Menten equation using GraphPad Prism 5.04 (GraphPad Software, Inc, La Jolla, USA).

In vitro His<sub>6</sub>-ObaH decarboxylase activity assay. Reactions were performed in 100 μl reaction volumes comprising 50 μM His<sub>6</sub>-ObaH, 10 mM and 9 mM or phenylpyruvate, all in His<sub>6</sub>-ObaH buffer (25 mM HEPES pH 7.8, NaCl (300 mM), MgCl₂ (15 mM) and THDP (0.5 mM)). Reactions were initiated by introduction of the enzyme and were incubated at 27°C and 700 r.p.m. for 5 min. A boiled enzyme sample was used as a negative control. Reactions were terminated as described previously. HPLC and LCMS analysis was performed as described for the His<sub>6</sub>-ObaG t-TTA activity assay.

Coupling of His<sub>6</sub>-ObaH and His<sub>6</sub>-ObaG reactions. His<sub>6</sub>-ObaH was first exchanged into His<sub>6</sub>-ObaH buffer (see previous) to avoid reaction of 9 with unbound PLP. Reactions were performed in 100 μl reaction volumes comprising 50 μM His<sub>6</sub>-ObaG, 50 μM Hi<sub>6</sub>-ObaH, 10 mM 9 and 10 mM t-threonine, all in His<sub>6</sub>-ObaH buffer. Reactions were initiated by introduction of the enzyme and were incubated at 27°C and 700 r.p.m. for 2 h. Control reactions were also performed in which one or both of His<sub>6</sub>-ObaG and His<sub>6</sub>-ObaH were boiled before the reaction. Reactions were terminated by addition of MeOH (100 μl) and were incubated at 20°C for 1 h to ensure full enzyme precipitation. Precipitated enzyme was pelleted at 16,16g for 30 min before analysis by HPLC.

H<sub>6</sub>-Oba hydroxymethyl-trapping reaction. Reaction mixtures comprised 8.5 μM His<sub>6</sub>-Oba, 50 mM Tris-HCl pH 8, 4 ATP, (2.25 mM), hydroxylamine (150 mM), amino acid substrate (5 mM), MgCl₂ (15 mM) in a final volume of 300 μl and were allowed to proceed at 28°C for 5 h. Boiled enzyme and no substrate reactions were performed as negative controls. Following addition of Tris-HCl and centrifugation to pellet precipitated protein, samples were transferred to cuvettes and were measured at A₅₄₀ nm on a Spectronic Biomate 3 (Thermo Fisher Scientific).

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms15935 | www.nature.com/naturecommunications

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms15935

ARTICLE

8 | 15935 | DOI: 10.1038/ncomms15935 | www.nature.com/naturecommunications
Ethyl 3-(4-nitrophenyl)-4,5-dihydroxazol-4-carboxylate (11) was synthesized according to a literature procedure using p-nitrophenylacetaldehyde (990 mg, 6 mmol) and ethyl isocyanate (750 mg, 6.6 mmol) to yield 910 mg of the racemic trans-diastereoisomer. Yield: 55%. 1H NMR (400 MHz, CDCl3), δ 8.08 (2H, d, J = 8.8 Hz), 7.56 (2H, d, J = 8.8 Hz), 6.87 (1H, d, J = 9.4 Hz), 4.72 (1H, dd, J1 = 9.4 Hz, J2 = 2.9 Hz), 4.41-4.36 (1H, m), 4.22–4.13 (2H, m), 2.92-2.78 (2H, m), 1.23 (3H, t, J = 7.2 Hz). 13C NMR (100 MHz, CDCl3) δ 170.1, 161.8, 146.8, 145.3, 130.3, 123.6, 72.1, 62.2, 54.6, 40.1, 14.0 ppm. HRMS (m/z): [M + H]+ calcd. for C13H15N2O5, 279.0975; found, 279.0978.

1H NMR (400 MHz, D2O) δ 8.78 (2H, d, J = 8.8 Hz), 7.56 (2H, d, J = 8.8 Hz), 6.87 (1H, d, J = 9.4 Hz), 4.72 (1H, dd, J1 = 9.4 Hz, J2 = 2.9 Hz), 4.41-4.36 (1H, m), 4.22–4.13 (2H, m), 2.92-2.78 (2H, m), 1.23 (3H, t, J = 7.2 Hz). 13C NMR (100 MHz, CDCl3) δ 170.1, 161.8, 146.8, 145.3, 130.3, 123.6, 72.1, 62.2, 54.6, 40.1, 14.0 ppm. HRMS (m/z): [M + H]+ calcd. for C13H15N2O5, 279.0975; found, 279.0978.

(2S,R)-2-amino-3-hydroxy-4-(4-nitrophenyl)butanoic acid (2) was synthesized according to a literature procedure. Starting from 720 mg of 11 we obtained 770 mg of 2 (97.9%). 1H NMR (400 MHz, D2O) δ 8.31 (2H, d, J = 8.5 Hz), 7.63 (2H, d, J = 8.5 Hz), 4.44 (1H, m), 3.84 (1H, m), 3.27 (1H, dd, J1 = 14.0 Hz, J2 = 3.2 Hz), 3.05 (1H, dd, J1 = 14.0 Hz, J2 = Yield: 92%; 1.6 Hz); 13C NMR (100 MHz, CDCl3) δ 172.5, 146.6, 145.9, 130.5, 123.9, 70.3, 59.2, 39.8 ppm. HRMS (m/z): [M + H]+ calcd. for C10H15N2O3, 211.1077; found: 211.1081. NMR data were consistent with those of the synthetic reference.

Conditions for the ObaG discontinuous assay described above. The reaction was quenched after 2 h by the addition of MeOH (2 ml) and the solution was concentrated to yield 2 (1.8 mg, 38%). NMR data were consistent with those of the synthetic reference.

Phylogenetic analyses. Amino acid sequences of l-TAs and SHMTs were obtained from a previous phylogenetic study and were combined with additional l-TAs and SHMT sequences for enzymes which have been characterized or described in the literature and were obtained from the National Center for Biotechnology Information (NCBI) GenBank database and the Protein Data Bank (PDB).

Data availability. The authors declare that the data supporting the findings reported in this study are available within the article and the Supplementary Information, or are available from the authors on reasonable request. Nucleotide sequence data have been deposited in NCBI GenBank under the accession codes as follows: I BGC (KX931446); pTS1 (KX931445); pME3087 (KX931444).

References.
31. Choi, Y. S., Zhang, H., Brunzelle, J. S., Nair, S. K. & Zhao, H. In vitro reconstitution and crystal structure of p-aminobenzoate N-oxidase (AurF) involved in aurin acid biosynthesis. Proc. Natl. Acad. Sci. USA 105, 6858–6863 (2008).

32. Knood, C. J., Kovaléva, E. G. & Lipscomb, J. D. Crystal structure of CmlI, the aminolysis enzyme in the chloramphenicol biosynthetic pathway. J. Biol. Inorg. Chem. 21, 589–603 (2016).

33. Hu, L., Chanco, E. & Zhao, H. CmlI is an enzyme involved in the biosynthesis of the chloramphenicol. Tetrahedron 68, 7651–1654 (2012).

34. Rothig, M. et al. NISpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39, W362–W367 (2011).

35. Quadri, L. E. N., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem. Biol. 5, 631–643 (1998).

36. Pelletier, T. K., Nair, S. K., Cai, J. T. & Sattely, E. S. Enzymatic tailoring of ornithine in the biosynthesis of the aureothin biosynthesis. Biochemistry 49, 8815–8817 (2010).

37. Baltz, R. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. J. Ind. Microbiol. Biotechnol. 38, 1747–1760 (2011).

38. Heemstra, J. R., Walsh, C. T. & Sattely, E. S. Enzymatic tailoring of ornithine in the biosynthesis of the rhizobactin cyclic trihydroxamate siderophore vicibactin. J. Am. Chem. Soc. 131, 15317–15329 (2009).

39. Chen, H., Hubbard, B. K., O’Connor, S. E. & Walsh, C. T. Formation of β-hydroxy histidine in the formation of nikkomycin antibiotics. Chem. Biol. 9, 103–112 (2002).

40. Herbert, D. D., Boll, B., Zocher, G., Stehle, T. & Heide, L. Structural basis of the reconstituted biotransformation of 4-fluorothreonine from fluoride ion: In vitro reconstitution of biotransformation of 4-fluorothreonine from fluoride ion: application of the fluorinase. Chem. Biol. 15, 1268–1276 (2008).

41. Berman, H. M. et al. PDB ‒ the Protein Data Bank. Nucleic Acids Res. 44, D67–D72 (2016).

42. Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Anal. Chem. 68, 850–858 (1996).

43. Chen, D. & Frey, P. A. Identification of lysine 346 as a functionally important residue for pyridoxal-5'-phosphate binding and catalysis in lysine 2,3-aminomutase from Bacillus subtilis. Cell Chem. Biol. 56, 2192–2197 (2017).

44. Rao, M. N., Holkar, A. G. & Ayyangar, N. R. A novel transacylation method for reconstitution of N-acetyl-L-glutamate synthetase. Biochim. Biophys. Acta 970, 352–359 (1989).

45. Meyer, S. et al. Biochemical dissection of the natural diversification of fungal alanine racemase–a subgroup of strictly related enzymes specialized for nonribosomal peptide synthetase-independent siderophore synthetases. Biochemistry 56, 596–602 (2001).

46. Nair, S. K. et al. Biochemical and crystal structure of p-aminobenzoate N-oxidase AurF involved in aurin acid biosynthesis. Proc. Natl. Acad. Sci. USA 99, 13132–13137 (2002).

47. Miller, M. A., Pfeiffer, W. & Schwartz, T. in Proceedings of the Gateway Computing Environments Workshop (GCE), 1–8 (IEEE, New Orleans, LA, USA, 2010).

Acknowledgements
This work was supported by the BBSRC via Institute Strategic Programme Grant BB/I004561/1 to the John Innes Centre (JIC). The authors declare no conflicts of interest. We thank Dr Jacob Malone (JIC) for the supply of pME3807 and his invaluable assistance and expertise in Pseudomonas genetics. Plasmid pH110 was a kind gift of Professor Chris Thomas and Dr Joanne Hothersall (University of Birmingham, UK). We also thank Dr Lionel Hill and Dr Gerhard Saalbach (JIC) for their excellent metabolomics support. Genome sequencing was undertaken at the Earlham Institute (Norwich, UK).

Author contributions
B.W. and T.A.S designed the research. T.A.S., D.H. and B.W. wrote the manuscript. T.A.S., D.H. and B.W. performed detailed bioinformatics and phylogeny analysis, carried out the mutational analysis, purified proteins and ran biochemical experiments. D.H. carried out the ObaG NMR experiments and isolated enzymatically synthesized compounds. D.H. and B.W. synthesized chemical standards. Z.Q. isolated obafluorin and analogues and carried out NMR analysis.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Scott, T. A. et al. An 1-threonine transaldolase is required for L-threo-β-hydroxy-γ-oxo amino acid assembly during obafluorin biosynthesis. Nat. Commun. 8, 15935 doi: 10.1038/ncomms15935 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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