Marine Proteobacteria metabolize glycolate via the β-hydroxyaspartate cycle

https://doi.org/10.1038/s41586-019-1748-4

Received: 29 March 2019

Accepted: 20 September 2019

Published online: 13 November 2019

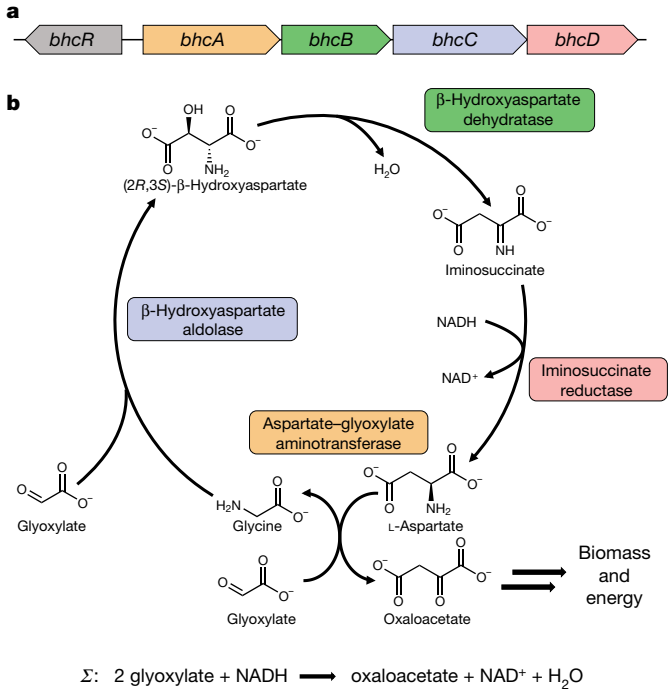
Lennart Schada von Borzyskowski[1]*, Francesca Severi[1], Karen Krüger[2], Lucas Hermann[3], Alexandre Gilardet[1], Felix Sippel[1], Bianca Pommerenke[1], Peter Claus[1], Niña Socorro Cortina[1], Timo Glatter[4], Stefan Zauner[5], Jan Zarzycki[1], Bernhard M. Fuchs[2], Erhard Bremer[3,6], Uwe G. Maier[5,6], Rudolf I. Amann[2] & Tobias J. Erb[1,6]*

One of the most abundant sources of organic carbon in the ocean is glycolate, the secretion of which by marine phytoplankton results in an estimated annual flux of one petagram of glycolate in marine environments[1]. Although it is generally accepted that glycolate is oxidized to glyoxylate by marine bacteria[2–4], the further fate of this C_2 metabolite is not well understood. Here we show that ubiquitous marine Proteobacteria are able to assimilate glyoxylate via the β-hydroxyaspartate cycle (BHAC) that was originally proposed 56 years ago[5]. We elucidate the biochemistry of the BHAC and describe the structure of its key enzymes, including a previously unknown primary imine reductase. Overall, the BHAC enables the direct production of oxaloacetate from glyoxylate through only four enzymatic steps, representing—to our knowledge—the most efficient glyoxylate assimilation route described to date. Analysis of marine metagenomes shows that the BHAC is globally distributed and on average 20-fold more abundant than the glycerate pathway, the only other known pathway for net glyoxylate assimilation. In a field study of a phytoplankton bloom, we show that glycolate is present in high nanomolar concentrations and taken up by prokaryotes at rates that allow a full turnover of the glycolate pool within one week. During the bloom, genes that encode BHAC key enzymes are present in up to 1.5% of the bacterial community and actively transcribed, supporting the role of the BHAC in glycolate assimilation and suggesting a previously undescribed trophic interaction between autotrophic phytoplankton and heterotrophic bacterioplankton.

Global net primary production has been estimated to be approximately 100 petagrams of carbon per year, equal parts of which are produced in terrestrial and marine habitats[6]. In the oceans, more than a third of primary production can be released into the water column by phytoplankton as dissolved organic carbon[7], generating a plethora of substrates for heterotrophic bacterioplankton. An abundant component of the pool of dissolved organic carbon is the carboxylic acid glycolate, which is released as a photorespiratory waste product of marine autotrophs[3,8,9]. Concentrations of glycolate in the nanomolar-to-low micromolar range have been measured in different marine habitats[1,2,10,11] (Extended Data Fig. 1), and the compound is readily taken up by bacterioplankton[12]. The first step in glycolate metabolism is its oxidation to glyoxylate, which is catalysed by the enzyme glycolate oxidase. The abundance and transcription of the *glcD* gene, which encodes a subunit of glycolate oxidase, has previously been used to investigate bacterial groups that are capable of glycolate utilization[4,13]. However, it has been assumed that glycolate is the subject of bacterial oxidation mainly to conserve energy[2–4]; the further fate of glyoxylate has not been described in detail. For SAR11 bacteria, it has been shown that glyoxylate can be used to replace the obligate glycine requirement[14]. In SAR11 and other bacteria, glyoxylate can be co-assimilated by malate synthase into the tricarboxylic acid cycle[14–16] or directly assimilated into central carbon metabolism through the well-studied glycerate pathway[17,18]. An alternative solution is the BHAC[5,19], which has been previously proposed to operate in the Alphaproteobacterium *Paracoccus denitrificans*[20,21]. However, the complete reaction sequence and the proteins comprising this pathway and their detailed biochemistry have remained unknown for the past 56 years.

On the basis of the sequence of a putative β-hydroxyaspartate aldolase gene (*dhaa*; GenBank accession number AB075600) from *P. denitrificans* IFO 13301[22], we identified a homologue in the genome of *P. denitrificans* DSM413 (BLT64_RS06500), annotated as a DSD1 family pyridoxal 5-phosphate (PLP)-dependent enzyme. This gene is part of a gene cluster, which consists of four structural genes and a putative transcriptional regulator that we termed *bhcABCD* and *bhcR* (Fig. 1a). In addition to the gene for the putative β-hydroxyaspartate aldolase (*bhcC*), the cluster comprises the open reading frames that encode a putative PLP-dependent aminotransferase (BLT64_RS06510, *bhcA*), a putative serine/threonine dehydratase (BLT64_RS06505, *bhcB*) and a putative ornithine cyclodeaminase (BLT64_RS06495, *bhcD*). The

[1]Department of Biochemistry & Synthetic Metabolism, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany. [2]Department of Microbial Ecology, Max Planck Institute for Marine Microbiology, Bremen, Germany. [3]Laboratory for Molecular Microbiology, Department of Biology, Philipps-University Marburg, Marburg, Germany. [4]Facility for Mass Spectrometry and Proteomics, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany. [5]Laboratory for Cell Biology, Department of Biology, Philipps-University Marburg, Marburg, Germany. [6]LOEWE-Center for Synthetic Microbiology, Philipps-University Marburg, Marburg, Germany. *e-mail: schada@mpi-marburg.mpg.de; toerb@mpi-marburg.mpg.de



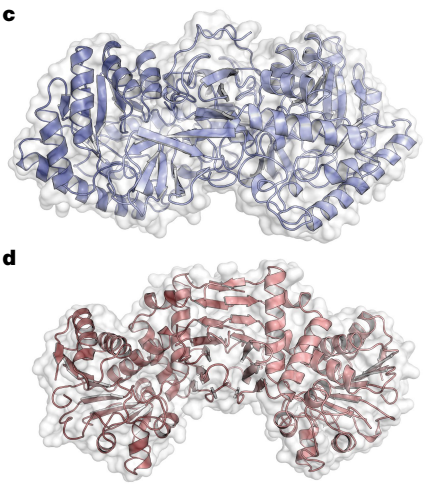


Fig. 1 | The BHAC. a, Genetic structure of the bhc gene cluster in *P. denitrificans* DSM 413. **b**, Reaction sequence and net balance of the BHAC. **c**, Cartoon representation of the β-hydroxyaspartate aldolase (BhcC) homodimer with superimposed protein surface (PDB 6QKB). **d**, Cartoon representation of the iminosuccinate reductase (BhcD) homodimer with superimposed protein surface (PDB 6RQA).

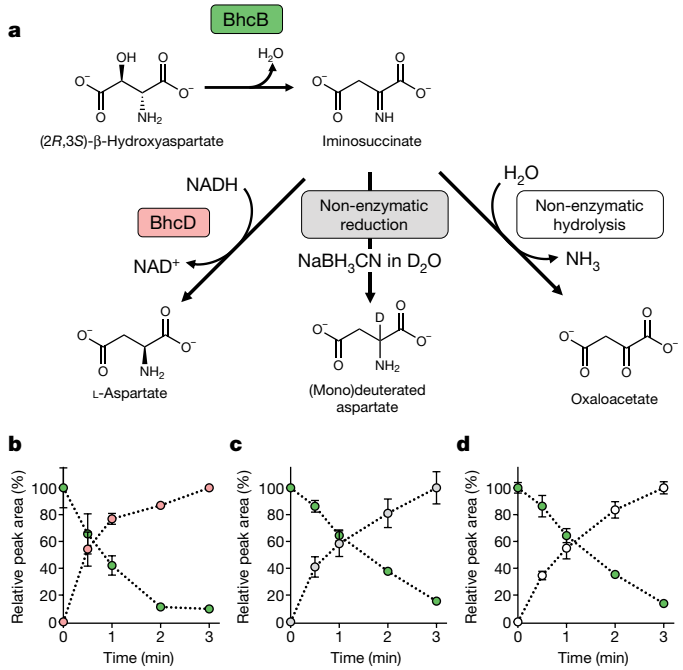


Fig. 2 | Reaction sequence catalysed by β-hydroxyaspartate dehydratase (BhcB) and iminosuccinate reductase (BhcD). a, Overview of the relevant reactions. **b**, Production of L-aspartate (red) from (2*R*, 3*S*)-β-hydroxyaspartate (green) by BhcB and reduction of iminosuccinate by BhcD. **c**, Production of (mono)deuterated aspartate (grey) from (2*R*, 3*S*)-β-hydroxyaspartate (green) by BhcB and reduction of iminosuccinate via $NaBH_3CN$ in D_2O. The data represent the formation of monodeuterated aspartate; owing to proton exchange, di- and trideuterated aspartate can also be formed in small quantities. **d**, Production of oxaloacetate (white) from (2*R*, 3*S*)-β-hydroxyaspartate (green) by BhcB and subsequent hydrolysis of iminosuccinate when neither BhcD nor $NaBH_3CN$ are added. **b**–**d**, Data are mean ± s.d.; $n = 3$ independent experiments.

putative transcriptional regulator (BLT64_RS06515), annotated as IclR-family regulator, is located in the opposite orientation to the four structural genes.

We expressed and characterized the four enzymes that are encoded in the gene cluster. BhcA is a PLP-dependent aminotransferase that transaminates glyoxylate into glycine using aspartate as the preferred amino group donor. BhcB functions as a β-hydroxyaspartate dehydratase. BhcC is a β-hydroxyaspartate aldolase, the key enzyme of the BHAC that catalyses the condensation of glyoxylate and glycine into β-hydroxyaspartate. This enzyme is closely related to D-threonine aldolases (Extended Data Fig. 2). The crystal structure of β-hydroxyaspartate aldolase that we solved at 1.7 Å (Protein Data Bank (PDB) 6QKB) shows that the three amino acids A160, A195 and S313 distinguish the active site of BhcC from that of D-threonine aldolases, providing a signature sequence for this enzyme family (Fig. 1c, Extended Data Fig. 2 and Extended Data Table 1). When combined, the BhcABC proteins were sufficient to reconstruct a reaction sequence from aspartate and two molecules of glyoxylate to two molecules of oxaloacetate and free ammonia. However, this left us puzzled about the function of the fourth open reading frame, the putative ornithine cyclodeaminase (*bhcD*).

When we tested BhcD in combination with BhcB, we discovered that it functions as an imine reductase (IRED) that accepts a labile iminosuccinate intermediate[23] formed by the latter enzyme (Fig. 2a, b). We used sodium cyanoborohydride trapping to demonstrate that BhcB produces iminosuccinate (Fig. 2c). Although this compound spontaneously decays into free ammonia and oxaloacetate in solution (Fig. 2d), iminosuccinate is reduced to L-aspartate in the presence of BhcD, thereby regenerating the amino group donor for the first step of the BHAC. IREDs are extensively investigated owing to their biotechnological potential[24]. Almost all IREDs described to date act on secondary imines, whereas the reduction of a free primary imine—as catalysed by BhcD—has not previously been described. The enzymatic reduction of primary imines is known only as part of the reaction sequence in glutamate dehydrogenase[25] and as part of a non-physiological side reaction of ketimine reductases[26]. The crystal structure of BhcD, which we solved to a resolution of 2.6 Å (PDB 6RQA), shows major differences in the active site compared to L-alanine dehydrogenase from *Archaeoglobus fulgidus* (PDB 1OMO), the closest structural homologue within the ornithine cyclodeaminase/μ-crystalline enzyme superfamily (Fig. 1d, Extended Data Fig. 3 and Extended Data Table 1). Phylogenetic analysis supports these active site differences and reveals that BhcD and its homologues constitute a novel family of primary IREDs within the ornithine cyclodeaminase/μ-crystalline superfamily (Extended Data Fig. 3).

Table 1 | Kinetic parameters of the four enzymes of the BHAC

Enzyme	Substrate	k_{cat} (s^{-1})	App. K_M (mM)	k_{cat}/K_M ($M^{-1} s^{-1}$)
Aspartate–glyoxylate aminotransferase (BhcA)	Glyoxylate	58 ± 1	0.43 ± 0.02	1.34×10^5
	L-Aspartate	56 ± 1	2.51 ± 0.10	2.25×10^4
	Glycine	0.76 ± 0.01	9.52 ± 0.40	7.97×10^1
	Oxaloacetate	0.76 ± 0.02	2.90 ± 0.27	2.62×10^2
	L-Serine	8.8 ± 0.3	2.10 ± 0.24	4.20×10^3
	L-Glutamate	5.0 ± 0.3	20.62 ± 2.33	2.44×10^2
β-Hydroxyaspartate dehydratase (BhcB)	(2*R*, 3*S*)-β-Hydroxyaspartate	35 ± 1	0.20 ± 0.02	1.75×10^5
β-Hydroxyaspartate aldolase (BhcC)	Glyoxylate	86 ± 4	0.23 ± 0.03	3.72×10^5
	Glycine	91 ± 2	4.31 ± 0.34	2.11×10^4
	(2*R*, 3*S*)-β-Hydroxyaspartate	33 ± 1	0.28 ± 0.03	1.18×10^5
	D-Threonine	76 ± 2	9.24 ± 0.86	8.25×10^3
Iminosuccinate reductase (BhcD)	Iminosuccinate	201 ± 10	0.09 ± 0.01	2.29×10^6
	NADH	–	0.02 ± 0.003	–
	NADPH	–	0.33 ± 0.05	–

Data are mean ± s.d., as determined from nonlinear fits of 18 data points with GraphPad Prism 8. Michaelis–Menten fits of enzyme kinetics and an SDS–PAGE gel showing purified proteins are provided in Extended Data Fig. 4 and Supplementary Fig. 1, respectively. For BhcA, kinetics for glyoxylate and L-aspartate were measured with 20 mM L-aspartate and 5 mM glyoxylate, respectively, and kinetics for glycine and oxaloacetate were measured with 20 mM oxaloacetate and 30 mM glycine, respectively. Kinetics for L-serine and L-glutamate were measured with 5 mM glyoxylate. For BhcC, kinetics for glycine and glyoxylate were measured with 5 mM glyoxylate and 20 mM glycine, respectively.

The kinetic parameters of all enzymes of the BHAC are reported in Table 1. The complete reaction sequence of the pathway is shown in Fig. 1b. The cycle extends the originally proposed reaction sequence[5] by the IRED reaction. Overall, the BHAC converts two molecules of glyoxylate (C_2) into oxaloacetate (C_4) without the loss of carbon as CO_2, under consumption of just one reducing equivalent and regeneration of the catalytic amino donor, which makes it one of the most efficient glyoxylate assimilation pathways described to date (Supplementary Table 1). Oxaloacetate formed in the BHAC can directly enter the tricarboxylic acid cycle or serve as substrate for anabolic reactions. The pathway is essential for the growth of *P. denitrificans* in the presence of glycolate and glyoxylate, and its enzymes are highly expressed and active in cells grown in the presence of glycolate (Extended Data Fig. 5). Glyoxylate negatively affected the interaction of the transcriptional regulator BhcR with the promoter region of the bhc gene cluster (Extended Data Fig. 5).

We next studied the phylogenetic distribution of the BHAC. The bhc gene cluster is widespread among the Rhizobiales and Rhodobacterales orders of the Alphaproteobacteria, and is also found in several gammaproteobacterial orders (Extended Data Fig. 6). Most of these bacteria were isolated from marine habitats, and the *Roseobacter* group within the Rhodobacterales is one of the three major bacterial groups responding to phytoplankton blooms[27]; *Roseobacter*-group bacteria can constitute up to 15% of the bacterial community in these blooms[28]. Notably, 94% of the isolates with the bhc gene cluster also encode glycolate oxidase in their genomes, enabling them to oxidize glycolate to glyoxylate for subsequent assimilation by the BHAC (Supplementary Data 1). BhcC is also ubiquitously present in marine metagenomes collected on the *Tara* Oceans expedition (Extended Data Fig. 7 and Supplementary Data 2), suggesting that the BHAC functions in glycolate assimilation in marine environments worldwide. Notably, the BHAC (represented by BhcC) is on average 20-fold more abundant than the glycerate pathway (represented by Gcl) in these datasets (Extended Data Fig. 7d).

To investigate the ecological importance of the BHAC, we focused our analyses on Helgoland (Extended Data Fig. 8a, b), an island in the North Sea that has already been used extensively as a study site to investigate the succession of bacterial populations during algal blooms[28,29]. We analysed metagenomes from seawater samples collected between 2010 and 2012 at Helgoland and detected the bhc gene cluster in all years at intermediate abundances (up to 3 reads per kilobase per million reads (RPKM), corresponding to roughly 1.5% of all cells[28]) (Extended Data Fig. 8c–e and Supplementary Data 3). To further investigate the role of the BHAC in situ, we monitored the spring phytoplankton bloom at Helgoland from March to May 2018. We determined chlorophyll *a* (Chl *a*) fluorescence as proxy for phytoplankton biomass and total microbial cell counts for each working day. Glycolate concentrations in the seawater were determined weekly.

The 2018 spring bloom was dominated by pennate diatoms and consisted of two peaks in phytoplankton growth in late April and late May (Fig. 3a). We determined a background concentration of glycolate in the seawater of 300 nM before the bloom, which is in line with previous measurements[1,2,4,9–11,30–39] (600 ± 340 nM) (Extended Data Fig. 1). During the bloom, from early March to late May, glycolate concentrations increased by approximately 350 nM (Fig. 3b), indicating the accumulation of phytoplankton-derived glycolate. At three time points in April and May, before and during the algal bloom, we determined bulk uptake rates of glycolate in the sea water. Glycolate uptake rates were in line with previously reported values[12] and increased more than threefold from 1.46 nM h^{-1} to 4.68 nM h^{-1} between the first and the last measurement (Fig. 3b), indicating that the capacity for glycolate uptake had multiplied at the same factor as the total microbial cell counts. Notably, these rates are comparable to uptake and consumption rates for dimethylsulfoniopropionate in the open ocean[40] and would enable a turnover of the total glycolate pool at our sampling site every 5–10 days. The bhc gene cluster was prevalent during the progression of the phytoplankton bloom. *bhcC* genes were detected at all of the time points, with the highest abundance per cell (around 1.5%) during the peaks of the phytoplankton bloom in April and May (Fig. 3c, Extended Data Fig. 9 and Supplementary Data 4). Transcription of *bhcC* was confirmed before and during the spring bloom (Fig. 3d, e), indicating that the BHAC is an active route for glycolate assimilation in the ocean.

In summary, our study provides the full reaction sequence and genetic basis of the BHAC. We demonstrate the biochemistry of the pathway, which involves a previously unknown family of IREDs, and provide support for its ecological importance in the assimilation of phytoplankton-derived dissolved organic carbon. The discovery of the BHAC as a ubiquitous pathway in marine environments adds a new

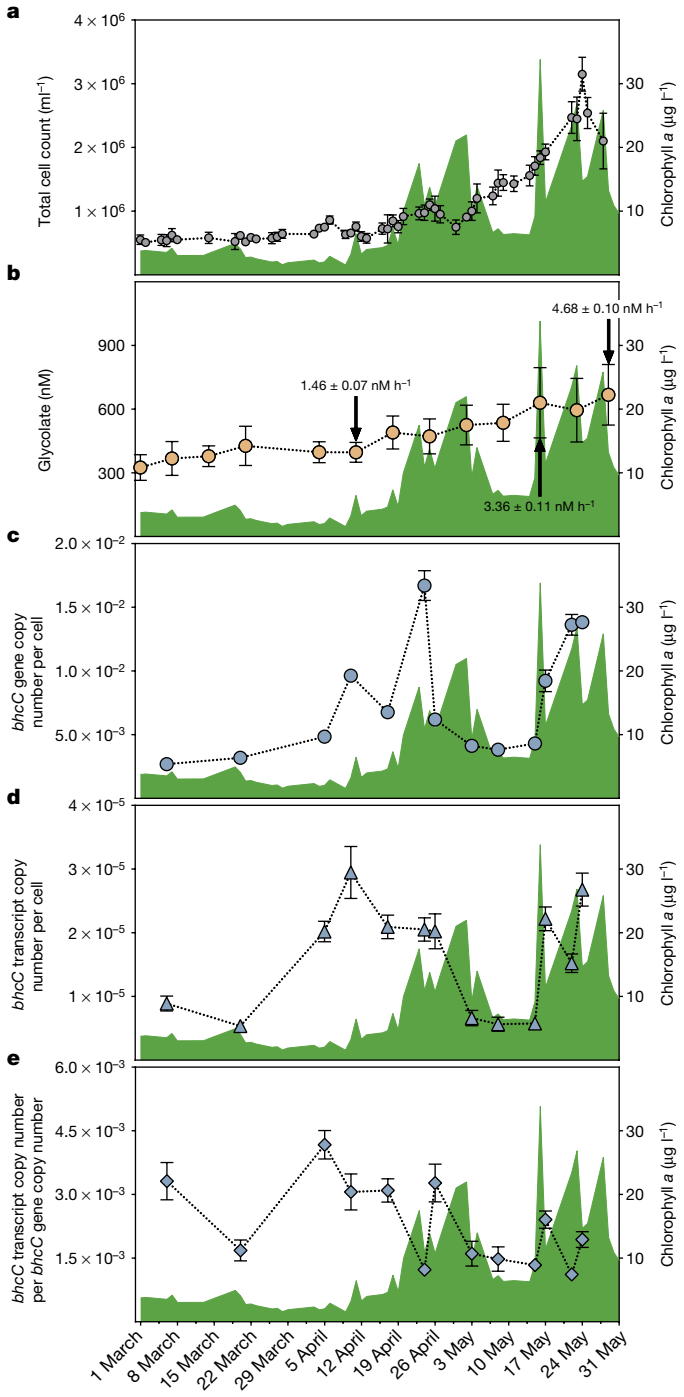


Fig. 3 | The BHAC during the spring phytoplankton bloom 2018 at Helgoland. a, From 1 March to 31 May, total microbial cell counts (grey) and Chl *a* concentrations (green) were determined each working day ($n = 1$). **b**, The concentration of glycolate (light brown) was determined once per week using liquid chromatography with mass spectrometry and increased from approximately 300 nM to around 650 nM. The uptake rate of bulk glycolate was determined at three time points through ^{14}C-glycolate incorporation and uptake rates are indicated. Data are the mean ± s.d. of $n = 5$ seawater samples for glycolate concentrations, and of $n = 4$ seawater samples for glycolate uptake rates. **c**, The *bhcC* gene copy number per cell (blue circles) was determined using qPCR. **d**, The *bhcC* transcript copy number per cell (blue triangles) was determined via cDNA synthesis followed by qPCR. **e**, *bhcC* transcript copy number divided by *bhcC* gene copy number (blue diamonds). **c**–**e**, Data are mean ± s.d.; $n = 3$ independent experiments.

dimension to the biochemical cycle of glycolate, an abundant organic acid in the global oceans. As the BHAC requires only one reducing equivalent and enables carbon-conserving glycolate assimilation, it may confer an advantage compared to the glycerate pathway, which releases CO_2. This may explain the high prevalence of the BHAC in marine Proteobacteria and could provide a starting point for future studies that investigate carbon fluxes from phytoplankton to heterotrophic bacterioplankton.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1748-4.

1. Wright, R. T. & Shah, N. M. Trophic role of glycolic acid in coastal seawater. II. Seasonal changes in concentration and heterotrophic use in Ipswich Bay, Massachusetts, USA. *Mar. Biol.* **43**, 257–263 (1977).
2. Wright, R. T. & Shah, N. M. Trophic role of glycolic acid in coastal seawater. I. Heterotrophic metabolism in seawater and bacterial cultures. *Mar. Biol.* **33**, 175–183 (1975).
3. Fogg, G. E. The ecological significance of extracellular products of phytoplankton photosynthesis. *Bot. Mar.* **26**, 3–14 (1983).
4. Lau, W. W., Keil, R. G. & Armbrust, E. V. Succession and diel transcriptional response of the glycolate-utilizing component of the bacterial community during a spring phytoplankton bloom. *Appl. Environ. Microbiol.* **73**, 2440–2450 (2007).
5. Kornberg, H. L. & Morris, J. G. β-Hydroxyaspartate pathway: a new route for biosyntheses from glyoxylate. *Nature* **197**, 456–457 (1963).
6. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
7. Duarte, C. M. & Cebrian, J. The fate of marine autotrophic production. *Limnol. Oceanogr.* **41**, 1758–1766 (1996).
8. Hellebust, J. A. Excretion of some organic compounds by marine phytoplankton. *Limnol. Oceanogr.* **10**, 192–206 (1965).
9. Tolbert, N. E. & Zill, L. P. Excretion of glycolic acid by algae during photosynthesis. *J. Biol. Chem.* **222**, 895–906 (1956).
10. Leboulanger, C., Descolasgros, C. & Jupin, H. HPLC determination of glycolic acid in seawater. An estimation of phytoplankton photorespiration in the Gulf of Lions, western Mediterranean Sea. *J. Plankton Res.* **16**, 897–903 (1994).
11. Leboulanger, C., Oriol, L., Jupin, H. & Descolas-Gros, C. Diel variability of glycolate in the eastern tropical Atlantic Ocean. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **44**, 2131–2139 (1997).
12. Casey, J. R., Ferrón, S. & Karl, D. M. Light-enhanced microbial organic carbon yield. *Front. Microbiol.* **8**, 2157 (2017).
13. Lau, W. W. & Armbrust, E. V. Detection of glycolate oxidase gene *glcD* diversity among cultured and environmental marine bacteria. *Environ. Microbiol.* **8**, 1688–1702 (2006).
14. Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. Nutrient requirements for growth of the extreme oligotroph 'Candidatus *Pelagibacter ubique*' HTCC1062 on a defined medium. *ISME J.* **7**, 592–602 (2013).
15. Eiler, A. et al. Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. *ISME J.* **10**, 1902–1914 (2016).
16. Tripp, H. J. et al. Unique glycine-activated riboswitch linked to glycine–serine auxotrophy in SAR11. *Environ. Microbiol.* **11**, 230–238 (2009).
17. Krakow, G. & Barkulis, S. S. Conversion of glyoxylate to hydroxypyruvate by extracts of *Escherichia coli*. *Biochim. Biophys. Acta* **21**, 593–594 (1956).
18. Hansen, R. W. & Hayashi, J. A. Glycolate metabolism in *Escherichia coli*. *J. Bacteriol.* **83**, 679–687 (1962).
19. Kornberg, H. L. & Morris, J. G. The utilization of glycollate by *Micrococcus denitrificans*: the β-hydroxyaspartate pathway. *Biochem. J.* **95**, 577–586 (1965).
20. Gibbs, R. G. & Morris, J. G. Assay and properties of β-hydroxyaspartate aldolase from *Micrococcus denitrificans*. *Biochim. Biophys. Acta* **85**, 501–503 (1964).
21. Gibbs, R. G. & Morris, J. G. Purification and properties of *erythro*-β-hydroxyaspartate dehydratase from *Micrococcus denitrificans*. *Biochem. J.* **97**, 547–554 (1965).
22. Liu, J. Q., Dairi, T., Itoh, N., Kataoka, M. & Shimizu, S. A novel enzyme, d-3-hydroxyaspartate aldolase from *Paracoccus denitrificans* IFO 13301: purification, characterization, and gene cloning. *Appl. Microbiol. Biotechnol.* **62**, 53–60 (2003).
23. Mortarino, M. et al. l-Aspartate oxidase from *Escherichia coli*. I. Characterization of coenzyme binding and product inhibition. *Eur. J. Biochem.* **239**, 418–426 (1996).
24. Lenz, M., Borlinghaus, N., Weinmann, L. & Nestl, B. M. Recent advances in imine reductase-catalyzed reactions. *World J. Microbiol. Biotechnol.* **33**, 199 (2017).
25. Hochreiter, M. C. & Schellenberg, K. A. α-Iminoglutarate formation by beef liver l-glutamate dehydrogenase. Detection by borohydride or dithionite reduction to glutamate. *J. Am. Chem. Soc.* **91**, 6530–6531 (1969).
26. Hallen, A., Cooper, A. J., Smith, J. R., Jamie, J. F. & Karuso, P. Ketimine reductase/CRYM catalyzes reductive alkylamination of α-keto acids, confirming its function as an imine reductase. *Amino Acids* **47**, 2457–2461 (2015).

27. Buchan, A., LeCleir, G. R., Gulvik, C. A. & González, J. M. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol*. **12**, 686–698 (2014).
28. Teeling, H. et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *eLife* **5**, e11888 (2016).
29. Teeling, H. et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**, 608–611 (2012).
30. Shah, N. M. & Wright, R. T. Occurrence of glycolic acid in coastal sea-water. *Mar. Biol*. **24**, 121–124 (1974).
31. Fogg, G. E., Burton, N. F. & Coughlan, S. J. The occurrence of glycollic acid in Antarctic waters. *Br. Antarct. Surv. Bull*. **41 & 42**, 193–195 (1975).
32. Hasan-Al, R. H., Coughlan, S. J., Pant, A. & Fogg, G. E. Seasonal variations in phytoplankton and glycollate concentrations in the Menai Straits, Anglesey. *J. Mar. Biol. Assoc. U.K*. **55**, 557–565 (1975).
33. Edenborn, H. M. & Litchfield, C. D. Glycolate turnover in the water column of the New York Bight apex. *Mar. Biol*. **95**, 459–467 (1987).
34. Leboulanger, C., Serve, L., Comellas, L. & Jupin, H. Determination of glycolic acid released from marine phytoplankton by post-derivatization gas chromatography mass spectrometry. *Phytochem. Anal*. **9**, 5–9 (1998).
35. Lord, J. M., Codd, G. A. & Merrett, M. J. The effect of light quality on glycolate formation and excretion in algae. *Plant Physiol*. **46**, 855–856 (1970).
36. Smith, W. O. Extracellular release of glycolic acid by a marine diatom. *J. Phycol*. **10**, 30–33 (1974).
37. Leboulanger, C., Martin-Jezequel, V., Descolas-Gros, C., Sciandra, A. & Jupin, H. J. Photorespiration in continuous culture of *Dunaliella tertiolecta* (Chlorophyta): relationships between serine, glycine, and extracellular glycolate. *J. Phycol*. **34**, 651–654 (1998).
38. Schnitzler Parker, M., Armbrust, E. V., Piovia-Scott, J. & Keil, R. G. Induction of photorespiration by light in the centric diatom *Thalassiosira weissflogii* (*Bacillariophyceae*): molecular characterization and physiological consequences. *J. Phycol*. **40**, 557–567 (2004).
39. Bertilsson, S., Berglund, O., Pullin, M. J. & Chisholm, S. W. Release of dissolved organic matter by *Prochlorococcus*. *Vie Milieu* **55**, 225–231 (2005).
40. Simó, R. & Pedrós-Alió, C. Short-term variability in the open ocean cycle of dimethylsulfide. *Glob. Biogeochem. Cycles* **13**, 1173–1181 (1999).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Chemicals and reagents

Unless otherwise stated, all chemicals and reagents were acquired from Sigma-Aldrich and were of the highest purity available.

Strains, medium and cultivation conditions

All strains used in this study are listed in Supplementary Table 2. *Escherichia coli* TOP10 (for genetic work), ST18 (for plasmid conjugation) and BL21 AI (for protein expression) were grown at 37 °C in lysogeny broth (LB)[41].

P. denitrificans DSM 413[42] and its derivatives were grown at 30 °C in LB or in mineral salt medium with TE3-Zn trace elements[43] supplemented with various carbon sources. To monitor growth, the optical density at 600 nm (OD_{600}) of culture samples was determined on a photospectrometer (Merck Chemicals).

Vector construction

The genes encoding the four enzymes of the BHAC (*bhcABCD*) as well as the *bhcR* gene encoding the transcriptional regulator were cloned into the standard expression vector pET16b (Merck Chemicals). To this end, the respective genes were amplified from genomic DNA of *P. denitrificans* DSM 413 with the primers provided in Supplementary Table 3. The resulting PCR products were digested with suitable restriction endonucleases (Thermo Fisher Scientific) as given in Supplementary Table 3 and ligated into the expression vector pET16b that had been digested with the same enzymes to create a vector for heterologous expression of the respective protein. Successful cloning of the desired open reading frames was verified by DNA sequencing (Eurofins Genomics). All plasmids used in this study are listed in Supplementary Table 2.

Expression and purification of recombinant proteins

For heterologous overexpression of the BhcA, BhcB, BhcC and BhcD enzymes, the corresponding plasmid encoding the respective enzyme was first transformed into chemically competent *E. coli* BL21 AI cells. The cells were then grown on LB agar plates containing 100 μg ml^{-1} ampicillin at 37 °C overnight. A starter culture in selective LB medium was inoculated from a single colony on the next day and left to grow overnight at 37 °C in a shaking incubator. The starter culture was used on the next day to inoculate an expression culture in selective terrific broth (TB) medium at a 1:100 dilution. The expression culture was grown at 37 °C in a shaking incubator to an OD_{600} of 0.5–0.7, induced with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) and 0.2% L-arabinose and subsequently grown overnight at 18 °C in a shaking incubator. Cells were collected at 6,000*g* for 15 min at 4 °C and cell pellets were stored at −20 °C until purification of enzymes. Cell pellets were resuspended in twice their volume in buffer A (300 mM NaCl, 25 mM Tris-HCl pH 8.0, 15 mM imidazole, 1 mM β-mercaptoethanol, 0.1 mM $MgCl_2$, 0.01 mM PLP and one tablet of SIGMAFAST protease inhibitor cocktail, EDTA-free per litre). The cell suspension was treated with a Sonopuls GM200 sonicator (BANDELIN Electronic) at an amplitude of 50% to lyse the cells and subsequently centrifuged at 50,000*g* and 4 °C for 1 h. The filtered supernatant (0.45-μm filter; Sarstedt) was loaded onto Protino Ni-NTA Agarose (Macherey-Nagel) in a gravity column, which had previously been equilibrated with 5 column volumes of buffer A. The column was washed with 20 column volumes of buffer A and 5 column volumes of 85% buffer A and 15% buffer B and the His-tagged protein was eluted with buffer B (buffer A with 500 mM imidazole). The eluate was desalted using PD-10 desalting columns (GE Healthcare) and buffer C (100 mM NaCl, 25 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 0.01 mM PLP, 0.1 mM dithiothreitol (DTT)). This was followed by purification on a size-exclusion column (Superdex 200 pg, HiLoad 16/600; GE Healthcare) connected to an ÄKTA Pure system (GE Healthcare) using buffer C. The concentrated protein solution (2 ml) was injected, and the flow was kept constant at 1 ml min^{-1}. Elution fractions containing pure protein were determined via SDS–PAGE analysis[44] on 12.5% gels. Purified enzymes in buffer C were used for crystallization or stored at −20 °C in buffer C containing 50% glycerol for later use in enzymatic assays.

BhcR was expressed and purified in the same way, except that buffer A contained 100 mM KCl, 20 mM HEPES-KOH pH 7.5, 10 mM $MgCl_2$, 4 mM β-mercaptoethanol, 5% glycerol and one tablet of SIGMAFAST protease inhibitor cocktail, EDTA-free per litre. Buffer C contained 100 mM KCl, 20 mM HEPES-KOH pH 7.5, 10 mM $MgCl_2$, 5% glycerol and 1 mM DTT.

NADH-dependent malate dehydrogenase (Mdh) and NADPH-dependent glyoxylate reductase (GhrA) from *E. coli* were overexpressed using the respective strains from the ASKA collection[45]. A starter culture in selective LB medium (34 μg ml^{-1} chloramphenicol) was inoculated from a single colony and left to grow overnight at 37 °C in a shaking incubator. The starter culture was used on the next day to inoculate an expression culture in selective TB medium at a 1:100 dilution. The expression culture was grown at 37 °C in a shaking incubator to an OD_{600} of 0.6, induced with 0.5 mM IPTG and grown another 4 h at 37 °C in a shaking incubator. The enzymes were affinity-purified in the same way as described above, except that buffer A contained 200 mM NaCl, 50 mM potassium phosphate pH 7.0, 15 mM imidazole, 1 mM β-mercaptoethanol and one tablet of SIGMAFAST protease inhibitor cocktail, EDTA-free per litre. Buffer C contained 100 mM NaCl, 50 mM potassium phosphate pH 7.0 and 0.1 mM DTT. The purified enzyme was stored at −20 °C in buffer C containing 50% glycerol.

Enzyme activity assays

For all enzyme assays, the oxidation of NADH or NADPH was followed at 340 nm or 360 nm on a Cary 60 UV-Vis photospectrometer (Agilent) in quartz cuvettes with a path length of 1 mm or 10 mm (Hellma Optik).

The enzyme assay to determine the kinetic parameters of BhcA with glyoxylate and L-aspartate as substrates was performed at 30 °C in a total volume of 300 μl. The reaction mixture contained 100 mM potassium phosphate buffer pH 7.5, 0.1 mM PLP, 0.2 mM NADH, different amounts of the respective substrates and 32 nM BhcA. Five hundred nanomolar Mdh was added as a coupling enzyme to convert oxaloacetate into malate. Kinetics for glyoxylate were measured with 20 mM L-aspartate; kinetics for L-aspartate were measured with 5 mM glyoxylate. To determine the kinetic parameters with oxaloacetate and glycine as substrates, the same assay mixture was used and 3 μM GhrA was added as a coupling enzyme to convert glyoxylate into glycolate. Kinetics for glycine were measured with 20 mM oxaloacetate; kinetics for oxaloacetate were measured with 30 mM glycine. To determine the kinetic parameters with L-serine or L-glutamate and glyoxylate as substrates, the same assay mixture was used and BhcB (3 μM), BhcC (1 μM) and Mdh (500 nM) were added as coupling enzymes. Kinetics for L-serine and L-glutamate were measured with 5 mM glyoxylate.

The enzyme assay to determine the kinetic parameters of BhcB was performed at 30 °C in a total volume of 300 μl. The reaction mixture contained 100 mM potassium phosphate buffer pH 7.5, 0.1 mM PLP, 0.2 mM NADH, different amounts of the substrate (2*R*, 3*S*)-β-hydroxyaspartate and 29 nM BhcB. Five hundred and eighty nanomolar BhcD was added as a coupling enzyme to convert iminosuccinate into L-aspartate. (2*R*, 3*S*)-β-Hydroxyaspartate was custom-synthesized by NewChem and was determined to be >95% pure by NMR analysis.

The enzyme assay to determine the kinetic parameters of BhcC with glyoxylate and glycine as substrates was performed at 30 °C in a total volume of 300 μl. The reaction mixture contained 100 mM potassium phosphate buffer pH 7.5, 0.1 mM PLP, 0.2 mM NADH, 0.5 mM $MgCl_2$, different amounts of the respective substrates and 4 nM BhcC. BhcB (200 nM) and BhcD (2 μM) were added as coupling enzymes. Kinetics

for glycine were measured with 5 mM glyoxylate; kinetics for glyoxylate were measured with 20 mM glycine. To determine the kinetic parameters with (2*R*, 3*S*)-β-hydroxyaspartate as substrate, the same assay mixture was used and 3 μM GhrA was added as a coupling enzyme to convert glyoxylate into glycolate. To determine the kinetic parameters with D-threonine as substrate, the same assay mixture was used and 3 μM alcohol dehydrogenase from *Saccharomyces cerevisiae* (Sigma-Aldrich) was added as coupling enzyme to convert acetaldehyde into ethanol.

The enzyme assay to determine the apparent kinetic parameters of BhcD was performed at 30 °C in a total volume of 250 μl. The reaction mixture contained 100 mM potassium phosphate buffer pH 7.5, 0.2 mM NADH, 0.1 mM PLP, different amounts of (2*R*, 3*S*)-β-hydroxyaspartate, and appropriate amounts of the enzymes BhcB and BhcD. Kinetics for iminosuccinate were measured with 15 nM BhcD, different amounts of (2*R*, 3*S*)-β-hydroxyaspartate and BhcB. To a given amount of (2*R*, 3*S*)-β-hydroxyaspartate, a tenfold molar excess of BhcB was added to start the reaction and completely and almost instantly convert the substrate pool into iminosuccinate. The initial reaction velocity of BhcD was determined after a mixing period of 3 s and the apparent concentration of iminosuccinate at this point in time was calculated on the basis of previously published values[23]. Kinetics for NADH and NADPH were measured with 2 mM (2*R*, 3*S*)-β-hydroxyaspartate, 214 nM BhcB, 28 nM BhcD and different amounts of the respective cofactor. No activity was measurable in a reaction mixture containing 100 mM potassium phosphate buffer pH 7.5, 0.2 mM NADH, 0.1 mM PLP and 3 mM oxaloacetate as well as 9 mM ammonium as putative substrates for BhcD.

The enzyme assay to generate iminosuccinate from (2*R*, 3*S*)-β-hydroxyaspartate (catalysed by BhcB) and further chemical reduction of iminosuccinate to L-aspartate with the reducing agent $NaBH_3CN$[46] was performed at 30 °C in a total volume of 1 ml. The reaction mixture contained 50 mM Tris pH 7.5, 1 mM (2*R*, 3*S*)-β-hydroxyaspartate, 0.1 mM PLP, 1 mM $MgCl_2$, 214 nM BhcB and 1 mM $NaBH_3CN$. The reaction was carried out in D_2O. Aliquots of 180 μl were taken after 0, 0.5, 1, 2 and 3 min and the reaction was immediately stopped by quenching with formic acid (4% final concentration). The samples were centrifuged at 17,000*g* and 4 °C for 15 min and the supernatant diluted 1:4 in double-distilled water for liquid chromatography–mass spectrometry (LC–MS) analysis. In negative control experiments, $NaBH_3CN$ was omitted from the reaction mixture. The same experiment was performed with added BhcD instead of $NaBH_3CN$ to enzymatically reduce iminosuccinate to L-aspartate. The reaction mixture contained 50 mM Tris pH 7.5, 1 mM (2*R*, 3*S*)-β-hydroxyaspartate, 2 mM NADH, 0.1 mM PLP, 1 mM $MgCl_2$, 214 nM BhcB and 28 nM BhcD.

LC–MS measurements were performed using an Agilent 6550 iFunnel Q-TOF LC–MS system equipped with an electrospray ionization (ESI) source set to negative ionization mode. LC was carried out as follows. The analytes were separated on an aminopropyl column (30 mm × 2 mm, particle size 3 μm, 100 Å; Luna NH2, Phenomenex) using a mobile phase system consisting of 95:5 20 mM ammonium acetate pH 9.3 (adjusted with ammonium hydroxide to a final concentration of approximately 10 mM): acetonitrile (A) and acetonitrile (B). Chromatographic separation was carried out using the following gradient condition at a flow rate of 250 μl min^{-1}: 0 min, 85% B; 3.5 min, 0% B, 7 min, 0% B; 7.5 min, 85% B; 8 min, 85% B. Column oven and autosampler temperature were maintained at 15 °C. The ESI source was set to the following parameters: capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 l min^{-1}, 225 °C) and sheath gas (12 l min^{-1}, 400 °C). The Q-TOF mass detector was calibrated before measurement using an ESI-L Low Concentration Tuning Mix (Agilent) with residuals and corrected residuals less than 2 ppm and 1 ppm, respectively. MS data were acquired with a scan range of 50–600 *m/z*. Autorecalibration was carried out using 113 *m/z* as reference mass. Subsequent peak integration of all analytes was performed using eMZed 2.29.4.0[47].

Enzyme activity assays in *P. denitrificans* cell extracts

P. denitrificans cultures were collected during mid-exponential phase (OD_{600} of 0.5–0.7), resuspended in ice-cold 100 mM potassium phosphate buffer (pH 7.2) and lysed by sonication. Cell debris was separated by centrifugation at 35,000*g* and 4 °C for 1 h. Total protein concentrations of the resulting cell-free extracts were determined by Bradford assay[48] using bovine serum albumin as standard. The assays for activity of BhcABCD were performed as described above, except that 100 mM potassium phosphate buffer pH 7.5 was replaced with 100 mM Tris pH 7.5. During BhcD assays, 90 μl samples were taken after 0.5, 1 and 2 min, and the reaction was immediately stopped by quenching with formic acid (4% final concentration). The samples were centrifuged at 17,000*g* and 4 °C for 15 min and the supernatant diluted 1:10 in double-distilled water for LC–MS analysis. L-Malate and L-aspartate in the samples were quantified using a standard curve of each compound ranging from 10 μM to 1,000 μM.

Genetic modification of *P. denitrificans*

The upstream and downstream flanking regions of the *bhcABCD* genes from *P. denitrificans* DSM 413 were cloned into the gene deletion vector pREDSIX[49]. To this end, the flanking regions were amplified from genomic DNA of *P. denitrificans* DSM 413 with the primers given in Supplementary Table 3. The resulting PCR products were used to perform Gibson assembly with the vector pREDSIX, which had been digested with MfeI. Subsequently, the resulting vector was digested with NdeI, and a kanamycin-resistance cassette, which had been cut out of the vector pRGD-Kan with NdeI, was ligated into the cut site to generate the final vectors for gene deletion. For gene deletion of each of the genes *bhcABCD* separately and of the complete bhc gene cluster, the corresponding plasmid was first transformed into chemically competent *E. coli* ST18[50] cells, which were then grown on LB agar plates containing 100 μg ml^{-1} ampicillin, 50 μg ml^{-1} kanamycin and 50 μg ml^{-1} aminolevulinic acid at 37 °C overnight. A culture in selective LB medium was inoculated the next day and left to grow overnight at 37 °C. The cultures were diluted the next morning to an OD_{600} of 0.1. A culture of wild-type *P. denitrificans* DSM 413 in LB medium was inoculated from a glycerol stock and grown at 30 °C. ST18 cultures were collected at an OD_{600} of around 0.7, and the *P. denitrificans* culture was collected at an OD_{600} of about 1.3. All cell pellets were washed once with sterile 10 mM $MgSO_4$ and resuspended to an OD_{600} of approximately 10 in sterile 10 mM $MgSO_4$. Suspensions of ST18 cells and *P. denitrificans* cells were mixed in a 2:1 ratio and spotted on minimal medium agar plates without any carbon source. Plates were incubated at 30 °C overnight. The next morning, spots were removed from the plates, resuspended in LB medium and plated on LB agar plates containing 25 μg ml^{-1} kanamycin. Plates were incubated at 30 °C for 3 days. The respective gene deletion was verified by colony PCR and DNA sequencing (Eurofins Genomics) and the deletion strain was propagated in selective LB medium.

High-throughput growth assays with *P. denitrificans* strains

Cultures of wild-type *P. denitrificans* DSM 413 and gene deletion strains were pre-grown at 30 °C in LB medium containing 25 μg ml^{-1} kanamycin, when necessary. Cells were collected, washed once with minimal medium containing no carbon source and used to inoculate growth cultures of 180 μl minimal medium containing an appropriate carbon source as well as 25 μg ml^{-1} kanamycin for gene deletion strains. Growth in 96-well plates (Thermo Fisher Scientific) was monitored at 30 °C at 600 nm in a Tecan Infinite M200Pro reader (Tecan). The resulting data were evaluated using GraphPad Prism 8.0.0.

Whole-cell shotgun proteomics

To acquire the proteome of *P. denitrificans* growing on different carbon sources, 30 ml cultures were grown to mid-exponential phase (OD_{600} of around 0.4) in minimal medium supplemented with 30 mM succinate

or 60 mM glycolate. Four replicate cultures were grown for each carbon source. Main cultures were inoculated from precultures grown in the same medium at a dilution of 1:1,000. Cultures were collected by centrifugation at 4,000*g* and 4 °C for 15 min. The supernatant was discarded and pellets were washed in 40 ml phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). After washing, cell pellets were resuspended in 1 ml PBS, transferred into Eppendorf tubes and centrifuged as described above. Cell pellets in Eppendorf tubes were snap-frozen in liquid nitrogen and stored at −80 °C until they were used for the preparation of samples for LC–MS analysis and label-free quantification.

For protein extraction, bacterial cell pellets were resuspended in 4% SDS and lysed by heating (95 °C, 15 min) and sonication (Hielscher Ultrasonics). Reduction was performed for 15 min at 90 °C in the presence of 5 mM tris(2-carboxyethyl)phosphine followed by alkylation using 10 mM iodoacetamide at 25 °C for 30 min. The protein concentration in each sample was determined using the BCA protein assay kit (Thermo Fisher Scientific) following the manufacturer's instructions. Protein clean-up and tryptic digestion were performed using the SP3 protocol as previously described[51] with minor modifications regarding protein digestion temperature and solid-phase extraction of peptides. SP3 beads were obtained from GE Healthcare. Trypsin (1 µg, Promega) was used to digest 50 µg of total solubilized protein from each sample. Tryptic digestion was performed overnight at 30 °C. Subsequently, all protein digestions were desalted using C18 microspin columns (Harvard Apparatus) according to the manufacturer's instructions.

LC–MS/MS analysis of protein digestions was performed on a Q-Exactive Plus mass spectrometer connected to an electrospray ion source (Thermo Fisher Scientific). Peptide separation was carried out using an Ultimate 3000 nanoLC-system (Thermo Fisher Scientific), equipped with an in-house-packed C18 resin column (Magic C18 AQ 2.4 µm; Dr. Maisch). The peptides were first loaded onto a C18 precolumn (preconcentration set-up) and then eluted in backflush mode with a gradient from 98% solvent A (0.15% formic acid) and 2% solvent B (99.85% acetonitrile and 0.15% formic acid) to 25% solvent B over 105 min, continued from 25% to 35% of solvent B up to 135 min. The flow rate was set to 300 nl min^{-1}. The data acquisition mode for the initial label-free quantification study was set to obtain one high-resolution MS scan at a resolution of 60,000 (*m/z* 200) with a scanning range from 375 to 1,500 *m/z* followed by MS/MS scans of the 10 most intense ions. To increase the efficiency of MS/MS shots, the charged-state screening modus was adjusted to exclude unassigned and singly charged ions. The dynamic exclusion duration was set to 30 s. The ion accumulation time was set to 50 ms (both MS and MS/MS). The automatic gain control was set to 3×10^6 for MS survey scans and 1×10^5 for MS/MS scans. Label-free quantification was performed using Progenesis QI (v.2.0). MS raw files were imported into Progenesis and the output data (MS/MS spectra) were exported in MGF format. MS/MS spectra were then searched using MASCOT (v.2.5) against a database of the predicted proteome from *P. denitrificans* downloaded from the UniProt database (https://www.uniprot.org/; download date 26 January 2017), containing 386 common contaminant and background proteins that were manually added. The following search parameters were used: full tryptic specificity required (cleavage after lysine or arginine residues); two missed cleavages allowed; carbamidomethylation (C) set as a fixed modification; and oxidation (M) set as a variable modification. The mass tolerance was set to 10 ppm for precursor ions and 0.02 Da for fragment ions for high-energy collision dissociation. Results from the database search were imported back into Progenesis, mapping peptide identifications to MS1 features. The peak heights of all MS1 features annotated with the same peptide sequence were summed, and protein abundance was calculated per LC–MS run. Next, the data obtained from Progenesis were evaluated using the SafeQuant R package v.2.2.2[52]. Then, the 1% false-discovery rate of identification and quantification as well as the intensity-based absolute quantification values were calculated.

Electrophoretic mobility shift assays

Fluorescently labelled DNA fragments for electrophoretic mobility shift assays were generated by PCR from genomic DNA of *P. denitrificans* DSM 413. For the *Pbhc* regulatory region, primers Pbhc_fw and Pbhc_rev-dye were used to generate a 238-bp fragment containing the putative *Pbhc* promoter. The primers bhcA_fw and bhcA_rev-dye were used to generate a 255-bp fragment containing a fragment of the *bhcA* gene as negative control. The primers *Pbhc*_rev-dye and *bhcA*_rev-dye were 5′-labelled with the Dyomics 781 fluorescent dye (Microsynth). Binding reactions between the DNA fragments (0.025 pmol), various amounts of the purified protein BhcR (400×, 2,000×, 4,000×, 10,000×, 20,000× and 40,000× molar excess), and various concentrations of glyoxylate (0.01, 0.05, 0.1, 0.2, 0.5 and 1 mM final concentration) were performed in buffer A (20 mM potassium phosphate pH 7.0, 1 mM DTT, 5 mM $MgCl_2$, 50 mM KCl, 15 µg ml^{-1} bovine serum albumin, 50 µg ml^{-1} herring sperm DNA, 5% v/v glycerol, 0.1% Tween-20) in a total volume of 20 µl. After the reaction mixtures were incubated at 37 °C for 20 min, the samples were loaded onto a native 5% polyacrylamide gel and electrophoretically separated at 110 V for 60 min. BhcR–DNA interactions were detected using an Odyssey FC Imaging System (LI-COR Biosciences).

Crystallization and structure determination of BhcC and BhcD

The sitting-drop vapour-diffusion method was used for crystallization at 16 °C. Purified BhcC (10 mg ml^{-1}) was mixed in a 1:1 ratio with solution A containing 20% PEG 3350, 0.2 M ammonium chloride, pH 6.3 (final drop volume 1.4 µl). Reservoirs were filled with 40 µl solution A. Crystals appeared within 14 days. Crystals were briefly soaked in mother liquor supplemented with 40% glycerol for cryoprotection before freezing in liquid nitrogen.

Purified BhcD (10 mg ml^{-1}) was mixed in a 1:1 ratio with solution B containing 20% PEG 3350, 0.2 M $Mg(NO_3)_2$, 5 mM NAD^+ and 5 mM Tb-Xo4, pH 6.4 (final drop volume 4 µl). Various additives were tested to improve crystal quality and size. The best results were achieved with the recently described nucleating and phasing agent Tb-Xo4[53]. Reservoirs were filled with 114 µl of solution B. Crystals appeared within a week. Crystals were briefly soaked in mother liquor supplemented with 40% ethylene glycol for cryoprotection before freezing in liquid nitrogen.

X-ray diffraction data were collected at the beamlines ID29 and ID30B of the ESRF (Grenoble, France) and at beamline P13 of DESY (Hamburg, Germany). The data were processed with the XDS[54] (build 20180126) and CCP4 v.7.0 software packages[55]. The structures were solved by molecular replacement. For BhcC, the structure of a D-threonine aldolase (PDB 4V15)[56] served as search model. For BhcD, a homology model was made based on the structure of L-alanine dehydrogenase (PDB 1OMO)[57] using Swiss-Model[58]. This homology model was then used as search model for the molecular replacement. The molecular replacement was carried out using Phaser of the Phenix software package[59] (v.1.14), built with Phenix.Autobuild and refined with Phenix.Refine. Additional modelling, manual refining and ligand fitting was done in Coot[60] (v.0.8.9). Final positional and B-factor refinements, as well as water picking, were performed using Phenix.Refine. The structure models for BhcC and BhcD were deposited at the Protein Data Bank in Europe (PDBe) under PDB accession numbers 6QKB and 6RQA, respectively. Figures were made using Pymol 1.8.

Analysis of North Sea metagenome data

Searches for the bhc gene cluster in 38 assembled surface seawater metagenomes sampled at the island of Helgoland between 2010 and 2012 were performed using the *Ruegeria pomeroyi* DSS-3 bhc gene cluster proteins as reference (NCBI protein IDs WP_011241924.1 (BhcR), WP_011241925.1 (BhcA), WP_011241926.1 (BhcB), WP_011241927.1 (BhcC), WP_044029519.1 (BhcD)). All identified proteins of the 38 metagenomes were searched against these proteins using DIAMOND[61] BLASTp and post-filtered to those hits for which the entire gene cluster

could be detected on a metagenome contig. These contigs were, if possible, linked to metagenome-assembled genomes (MAGs) binned from the same 38 metagenomes. MAGs were binned as previously described[62] and both the metagenome assemblies and MAGs are accessible under accession PRJEB28156 at the European Nucleotide Archive (ENA). MAG quality was assessed using CheckM v.1.0.7[63]. Abundance estimates of MAGs and the single unbinned contig were calculated based on read mapping as reads per kilobase per million reads (RPKM; 2 RPKM ≈ 1% relative abundance detected by fluorescence in situ hybridization[28]). Read mapping of all 38 metagenomes to MAGs and the single unbinned contig was performed as previously described[62] using BBMap v.35.14 (http://bbtools.jgi.doe.gov).

Phylogenetic analyses

A genome tree of bacterial strains and five MAGs with the bhc gene cluster was calculated using GTDBtk v.0.1.3 with GTDB v.86[64]. GTDBtk uses an alignment of 120 bacterial marker genes to infer taxonomic relationships. The GTDBtk calculated tree was subsampled to the 264 bhc gene cluster containing bacterial strains and the MAGs and visualized using iTOL[65].

Sequences of BhcABCD from 264 bacterial isolates and 6 metagenome contigs (five of which were linked to MAGs) were aligned using MUSCLE[66], manually curated to remove gaps and concatenated. A phylogenetic tree of concatenated sequences of BhcABCD was calculated using raxmlGUI[67] 1.5b2 using the PROTGAMMA model with Le-Gascuel substitution matrix[68], 100 bootstraps and 100 maximum-likelihood resamplings. The resulting tree was visualized using iTOL.

Sequences from the ornithine cyclodeaminase/μ-crystalline superfamily (Conserved Domain accession cl27428) and the type III PLP-dependent enzymes superfamily (Conserved Domain accession cl00261) were downloaded from the NCBI protein database and aligned using MUSCLE. Phylogenetic trees of the aligned sequences were calculated with raxmlGUI 1.5b2 using the PROTGAMMA model with Le-Gascuel substitution matrix, 100 bootstraps and 100 maximum-likelihood resamplings. The resulting trees were visualized using iTOL.

In total, 1,614 protein sequences from the ornithine cyclodeaminase/μ-crystalline superfamily were used for generation of a sequence similarity network (SSN) using the EFI-EST web tool[69] with a cut-off value of 1×10^{-50}. In this SSN, all connected sequences that shared 80% or more identity were grouped into a single node, resulting in 619 meta nodes. The SSN was visualized with Cytoscape 3.7.1 (https://cytoscape.org) and edges between nodes with less than 50% identity were removed.

Analysis of *Tara* Oceans metagenomes

BhcC from *P. denitrificans* DSM 413 (Uniprot A1B8Z1) and Gcl from *Starkeya novella* DSM 506 (Uniprot D7A6R1) were used as queries to search the OM-RGC_v1 database using the Ocean Gene Atlas[70] web tool (http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas) with a cut-off value of 1×10^{-100}. The resulting hits were inspected and sequences that were deemed to not belong to BhcC or Gcl were removed. The following criteria were used: at least 50% of the query sequence covered; at least two of the three residues A160, A195, S313 present for BhcC sequences; residues V25, V51, L421, L476, L478, I479[71] present for Gcl sequences. The coordinates of sampling sites with positive hits for BhcC in samples from surface water (0.22–3-μm size fraction) were downloaded and visualized using Ocean Data View 5.1.5 (Schlitzer, R., Ocean Data View, odv.awi.de, 2018). Taxonomic assignments of BhcC and Gcl sequences were downloaded and manually converted to GTDB taxonomy. Sequence IDs are listed in Supplementary Data 2.

Environmental sample collection and processing

Sampling was carried out on each working day (Monday–Friday) with the RV *Aade* (https://www.awi.de/en/expedition/ships/more-ships.html) at the research site 'Kabeltonne' (54° 11.3′ N, 7° 54.0′ E) from approximately 1 m water depth in 20 l carboys. The water samples for microbial biomass were subjected to fractionating filtration directly upon arrival in the Biologische Anstalt Helgoland laboratory (typically less than one hour after sampling). Three membrane 142-mm diameter filtration units were operated in parallel to keep filtration times to a minimum. First, samples were pre-filtered through 142-mm diameter 10-μm-pore-size polycarbonate filters (Merck Chemicals) by means of an air-pressure pump to remove large particles and eukaryotic plankton. Then, the water samples were filtered with air-pressure pumps onto 142-mm diameter 3-μm-pore-size polycarbonate filters (Merck Chemicals) to collect predominantly bacteria associated with smaller particles and algae. Afterwards, dedicated aliquots were filtered onto 142-mm diameter 0.2-μm pore-size-polyethersulfone filters (Merck Chemicals) for DNA and RNA extraction. Bacterioplankton dominated this 0.2-μm fraction. The entire filtration process for all fractions was usually finished within 3 h, that is, latest 4 h after the sampling. All filters were stored at −80 °C until further analyses.

Total cell counts

Samples were fixed with 1% formaldehyde and filtered onto polycarbonate membrane filters as described above. Total cell counts were determined from 10 ml fixed seawater samples. One filter section was cut and stained with 4′,6-diamidino-2-phenylindole (DAPI, 1 μg ml^{-1}). The stained filters were analysed manually; the total cell count includes heterotrophic bacteria as well as autofluorescent cyanobacteria, but not picoeukaryotic cells.

Concentration of Chl *a*

The concentration of Chl *a* was determined in subsurface water on each working day (Monday–Friday) as part of the Helgoland Roads LTER time series (https://www.awi.de/en/science/biosciences/shelf-sea-system-ecology/working-groups/long-term-observations-lto.html). The concentration of Chl *a* was assessed from fluorescence data using an algal group analyser (bbe moldaenke).

Determination of glycolate concentrations

Once per week, 5 aliquots of 2 ml each were taken from the filtrate after 0.2-μm filtration and stored at −80 °C until analysis. Glycolate concentrations were measured after derivatization of the samples with 3-nitrophenylhydrazine as previously described[72]. LC–MS analyses were performed on an Agilent 6495B Triple Quad LC–MS system equipped with an electrospray ionization source. The analytes were separated on a RP-18 column (50 mm × 2.1 mm, particle size 1.8 μm, ZORBAX RRHD Eclipse Plus C18; Agilent) kept at 40 °C using a mobile phase system that consisted of 0.1% formic acid in water (A) and acetonitrile (B). The gradient was as follows: 0 min, 5% B; 1 min, 5% B; 6 min, 95% B; 6.5 min, 95% B; 7 min, 5% B at a flow rate of 250 μl min^{-1}. Samples were held at 15 °C and injection volume was 5 μl. MS/MS data were acquired in negative MRM mode. Capillary voltage was set at 3 kV and nitrogen gas was used as nebulizing (25 psig), drying (11 l min^{-1}, 130 °C) and sheath gas (12 l min^{-1}, 400 °C). The dwell time and fragmentor voltage were 20 ms and 380 V, respectively. Optimized collision energy used for the derivatized glycolate (210 $m/z \rightarrow 137$ m/z) was 22 V. LC–MS data were analysed and quantified using MassHunter Qualitative Navigator and QQQ Quantitative Analysis software (Agilent).

Determination of glycolate uptake rates

Samples for glycolate uptake measurements were collected on 10 April, 15 May and 29 May 2018. All samples were used after filtration through a 3-μm filter and divided into 4 live 40 ml subsamples in sterile plastic tubes wrapped in aluminium foil and incubated with 165 nM calcium [1-^{14}C]glycolate (American Radiolabelled Chemicals; 55 mCi $mmol^{-1}$, 0.1 mCi ml^{-1} in sterile water) at 12 °C for 8 h. Controls consisted of four 40 ml subsamples killed in 10% formalin for 1 h before addition of 165 nM calcium [1-^{14}C]glycolate. Glycolate uptake was monitored over time by withdrawing 5 ml aliquots from each subsample, filtering each aliquot

onto a 0.2-µm pore size Nuclepore polycarbonate filter (GE Healthcare), rinsing the filter 3 times with 5 ml of filter-sterilized sea water and measuring the radioactivity with a Tri-Carb 4910 TR liquid scintillation analyser (PerkinElmer) using the Ultima Gold scintillation cocktail (PerkinElmer). Glycolate uptake rates were determined by linear fit of the counts per minute measured on the filters over time. Uptake rates were corrected to account for the presence of non-radioactive glycolate in the samples.

DNA and RNA extraction, cDNA synthesis and qPCR

DNA and RNA was extracted from filters using the AllPrep Bacterial DNA/RNA/Protein Kit (Qiagen) according to the manufacturer's instructions. The RNA samples were treated with the TURBO DNA-free Kit (Thermo Fisher Scientific) according to the manufacturer's instructions to exclude contamination with DNA. DNA and RNA concentrations were determined using the Qubit dsDNA/RNA HS Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. In total, 2 µg of RNA was used for cDNA synthesis with the GoScript Reverse Transcription System (Promega) and random hexamers according to the manufacturer's instructions.

Degenerate primers for the *bhcC* gene were designed using the j-CODEHOP software[73–75] and an alignment of 207 *bhcC* sequences from bacterial strains isolated from marine habitats. Sequences were aligned using MUSCLE. Extracted DNA and cDNA of RNA were quantified using a CFX Connect Real-Time System (Bio-Rad). SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich) was used for the PCR amplification mixture according to the manufacturer's instructions. Final $MgCl_2$ concentration was 3 mM, and the amplification protocol consisted of an initial enzyme activation step at 95 °C for 5 min, followed by 45 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s. Eight standard amounts ranging from 3×10^1 to 3×10^8 copies were run in triplicate for each set of analyses. Regression of all standard curves yielded an r^2 value of at least 0.998. All samples were run in triplicate. The starting copy numbers of *bhcC* in DNA and cDNA were calculated based on regression parameters of standard curves, and gene/transcript copy numbers per cell were calculated based on the volume of sea water filtered, the microbial cell count at the time of sampling, the amount of extracted DNA or RNA, and the volume of DNA or cDNA used per reaction. The degenerate primers were validated with genomic DNA of *P. denitrificans* DSM 413, *Rhodobacter sphaeroides* 2.4.1, and *E. coli* K-12 MG1655 as template using the same qPCR protocol as above. Standards for quantification were created by PCR using genomic DNA of *P. denitrificans* DSM 413 as template. Purified *bhcC* PCR product was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The coordinates and structure factors of the crystal structures generated from this research are available at the PDB under accession numbers 6QKB and 6RQA. Mass spectrometry proteomics data are available via ProteomeXchange with the identifier PXD013274. MAGs are available under accession PRJEB28156 at the European Nucleotide Archive (ENA). All other relevant data are available in the Article and the Supplementary Information. Source Data for Figs. 2, 3 and Extended Data Fig. 1, 4, 5, 7–9 are provided with the paper.

41. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293–300 (1951).
42. Beijerinck, M. W. & Minkman, D. C. J. Bildung und Verbrauch von Stickoxydul durch Bakterien. *Zentralbl. Bakteriol. Naturwiss.* **25**, 30–63 (1910).
43. Hahnke, S. M., Moosmann, P., Erb, T. J. & Strous, M. An improved medium for the anaerobic growth of *Paracoccus denitrificans* Pd1222. *Front. Microbiol.* **5**, 18 (2014).
44. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
45. Kitagawa, M. et al. 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 (2005).
46. Lane, C. F. Sodium cyanoborohydride — a highly selective reducing agent for organic functional groups. *Synthesis* **1975**, 135–146 (1975).
47. Kiefer, P., Schmitt, U. & Vorholt, J. A. eMZed: an open source framework in Python for rapid and interactive development of LC/MS data analysis workflows. *Bioinformatics* **29**, 963–964 (2013).
48. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
49. Ledermann, R., Strebel, S., Kampik, C. & Fischer, H. M. Versatile vectors for efficient mutagenesis of *Bradyrhizobium diazoefficiens* and other Alphaproteobacteria. *Appl. Environ. Microbiol.* **82**, 2791–2799 (2016).
50. Thoma, S. & Schobert, M. An improved *Escherichia coli* donor strain for diparental mating. *FEMS Microbiol. Lett.* **294**, 127–132 (2009).
51. Moggridge, S., Sorensen, P. H., Morin, G. B. & Hughes, C. S. Extending the compatibility of the SP3 paramagnetic bead processing approach for proteomics. *J. Proteome Res.* **17**, 1730–1740 (2018).
52. Glatter, T. et al. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* **11**, 5145–5156 (2012).
53. Engilberge, S. et al. Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence. *Chem. Sci.* **8**, 5909–5917 (2017).
54. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
55. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
56. Uhl, M. K. et al. The crystal structure of d-threonine aldolase from *Alcaligenes xylosoxidans* provides insight into a metal ion assisted PLP-dependent mechanism. *PLoS ONE* **10**, e0124056 (2015).
57. Gallagher, D. T. et al. Structure of alanine dehydrogenase from *Archaeoglobus*: active site analysis and relation to bacterial cyclodeaminases and mammalian mu crystallin. *J. Mol. Biol.* **342**, 119–130 (2004).
58. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
59. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
60. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
61. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
62. Francis, T. B., Krüger, K., Fuchs, B. M., Teeling, H. & Amann, R. I. *Candidatus* Prosiliicoccus vernus, a spring phytoplankton bloom associated member of the *Flavobacteriaceae*. *Syst. Appl. Microbiol.* **42**, 41–53 (2019).
63. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
64. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
65. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
66. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
67. Silvestro, D. & Michalak, I. raxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol.* **12**, 335–337 (2012).
68. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).
69. Gerlt, J. A. et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019–1037 (2015).
70. Villar, E. et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res.* **46**, W289–W295 (2018).
71. Kaplun, A. et al. Glyoxylate carboligase lacks the canonical active site glutamate of thiamine-dependent enzymes. *Nat. Chem. Biol.* **4**, 113–118 (2008).
72. Han, J., Gagnon, S., Eckle, T. & Borchers, C. H. Metabolomic analysis of key central carbon metabolism carboxylic acids as their 3-nitrophenylhydrazones by UPLC/ESI-MS. *Electrophoresis* **34**, 2891–2900 (2013).
73. Rose, T. M. et al. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* **26**, 1628–1635 (1998).
74. Rose, T. M., Henikoff, J. G. & Henikoff, S. CODEHOP (consensus-degenerate hybrid oligonucleotide primer) PCR primer design. *Nucleic Acids Res.* **31**, 3763–3766 (2003).
75. Boyce, R., Chilana, P. & Rose, T. M. iCODEHOP: a new interactive program for designing consensus-degenerate hybrid oligonucleotide primers from multiply aligned protein sequences. *Nucleic Acids Res.* **37**, W222–W228 (2009).

Acknowledgements We thank L. Franzmeyer and K.-P. Rücknagel as well as the crew of the *Aade* and the staff at the Biological Station at Helgoland (BAH) for sample collection and processing; K. H. Wiltshire for providing Chl *a* data collected at BAH; S. Vidal-Melgosa, A. Bolte and J.-H. Hehemann for sharing samples; T. Ferdelman for help with radioactive tracer work; B. Vögeli, T. Schwander, G. Stoffel and S. Burgener for helpful discussions. We acknowledge the support from the staff scientists at the European Synchrotron Radiation Facility Grenoble,

France (ESRF, beamlines ID29 & ID30B) as well as at the Deutsches Elektronen-Synchrotron Hamburg, Germany (DESY, beamline P13). Metagenome sequences were obtained within the COGITO (Coastal Microbe Genomic and Taxonomic Observatory) project granted to H. Teeling (MPI Bremen) as a community sequencing project by the Department of Energy's Joint Genome Institute in Walnut Creek, CA, USA (Proposal ID 998; https://doi.org/10.25585/1488076). This study was funded by the Max-Planck-Society (R.I.A. and T.J.E.), FET-Open Grant 686330 (Future Agriculture) and the German Research Foundation (SFB987 'Microbial diversity in environmental signal response' and FOR 2406 'Proteogenomics of marine polysaccharide utilization').

Author contributions L.S.v.B. identified the bhc gene cluster, purified proteins, performed enzyme kinetic analysis, qPCR, phylogenetic analysis and analysis of *Tara* Oceans metagenomes, generated and characterized mutant *P. denitrificans* strains and measured glycolate uptake rates. F. Severi performed enzyme kinetic analysis, crystallization of BhcD and enzyme assays in *P. denitrificans* cell-free extracts. K.K. performed phylogenetic analysis and analysis of Helgoland metagenomes. L.H. performed gel shift assays with BhcR. A.G. performed crystallization of BhcC. F. Sippel performed enzyme kinetic analysis. B.P. generated mutant *P. denitrificans* strains. P.C. and N.S.C. performed small-molecule mass spectrometry. T.G. performed mass spectrometry for proteomics. J.Z. collected X-ray datasets, solved, refined and analysed crystal structures. B.M.F. and R.I.A. planned and supervised fieldwork at Helgoland and provided reagents. L.S.v.B., E.B., S.Z., U.G.M., R.I.A. and T.J.E. planned experiments, analysed data and supervised the project. L.S.v.B. and T.J.E. wrote the manuscript, with contributions from all other authors.

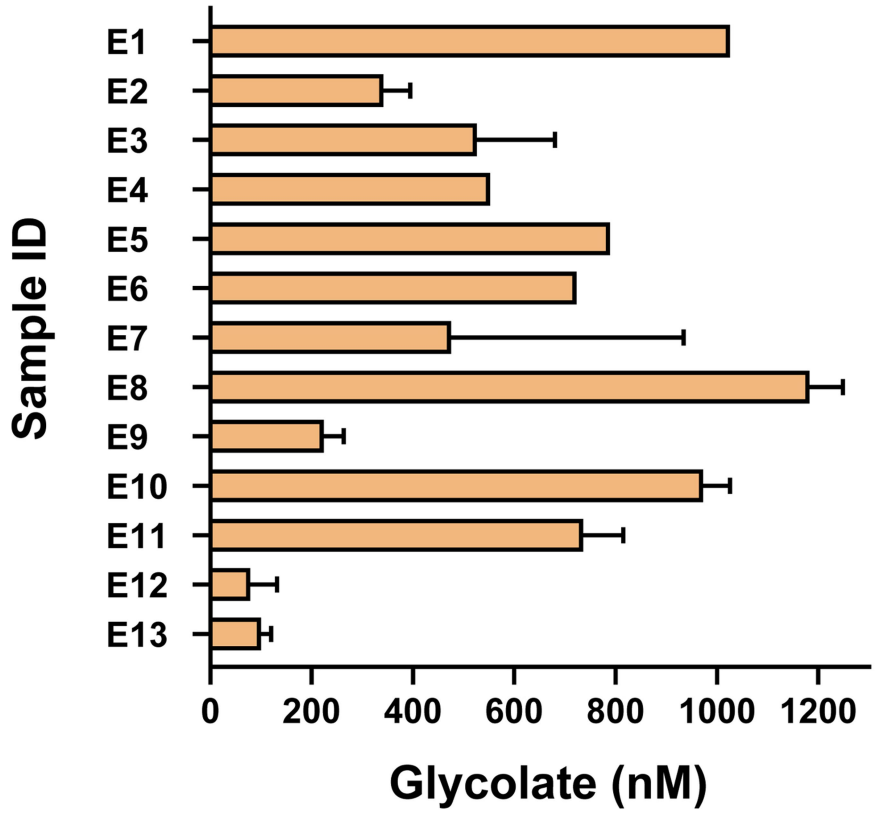
Competing interests The Max-Planck-Gesellschaft zur Förderung der Wissenschaften is the patent applicant for the following three patents. All patent applications are pending. L.S.v.B. and T.J.E. have filed European patent no. EP 19190404.4 for the production of plants with altered photorespiration due to implementation of the BHAC. L.S.v.B., J.Z. and T.J.E. have filed European patent no. EP 18167406.0 for the production of photoautotrophic organisms with altered photorespiration due to implementation of the BHAC. L.S.v.B. and T.J.E. have filed European patent no. 18211454.6 for the enantioselective preparation of primary amine compounds using the enzyme BhcD or its homologues.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1748-4.

Correspondence and requests for materials should be addressed to L.S.v.B. or T.J.E.

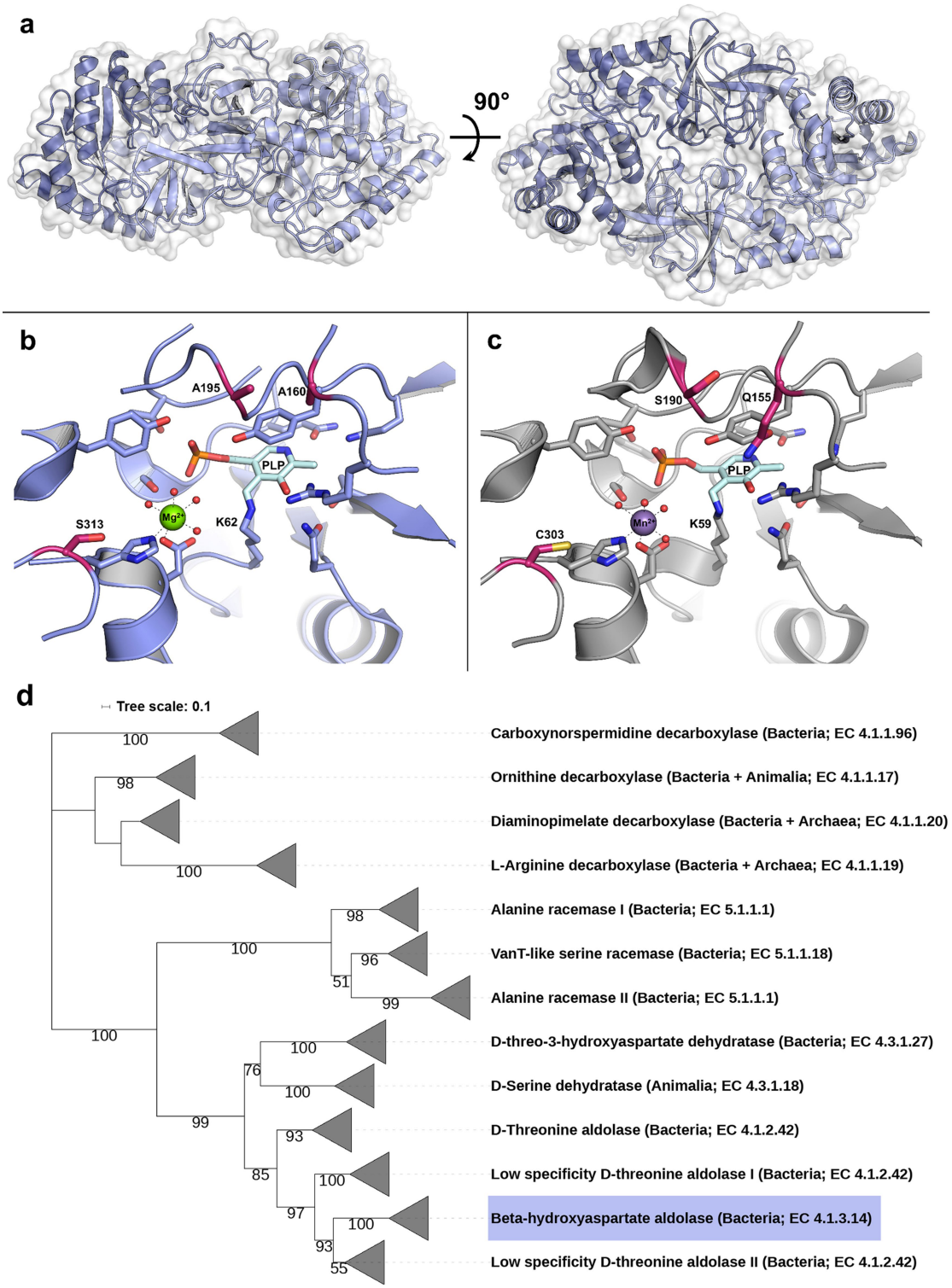
a



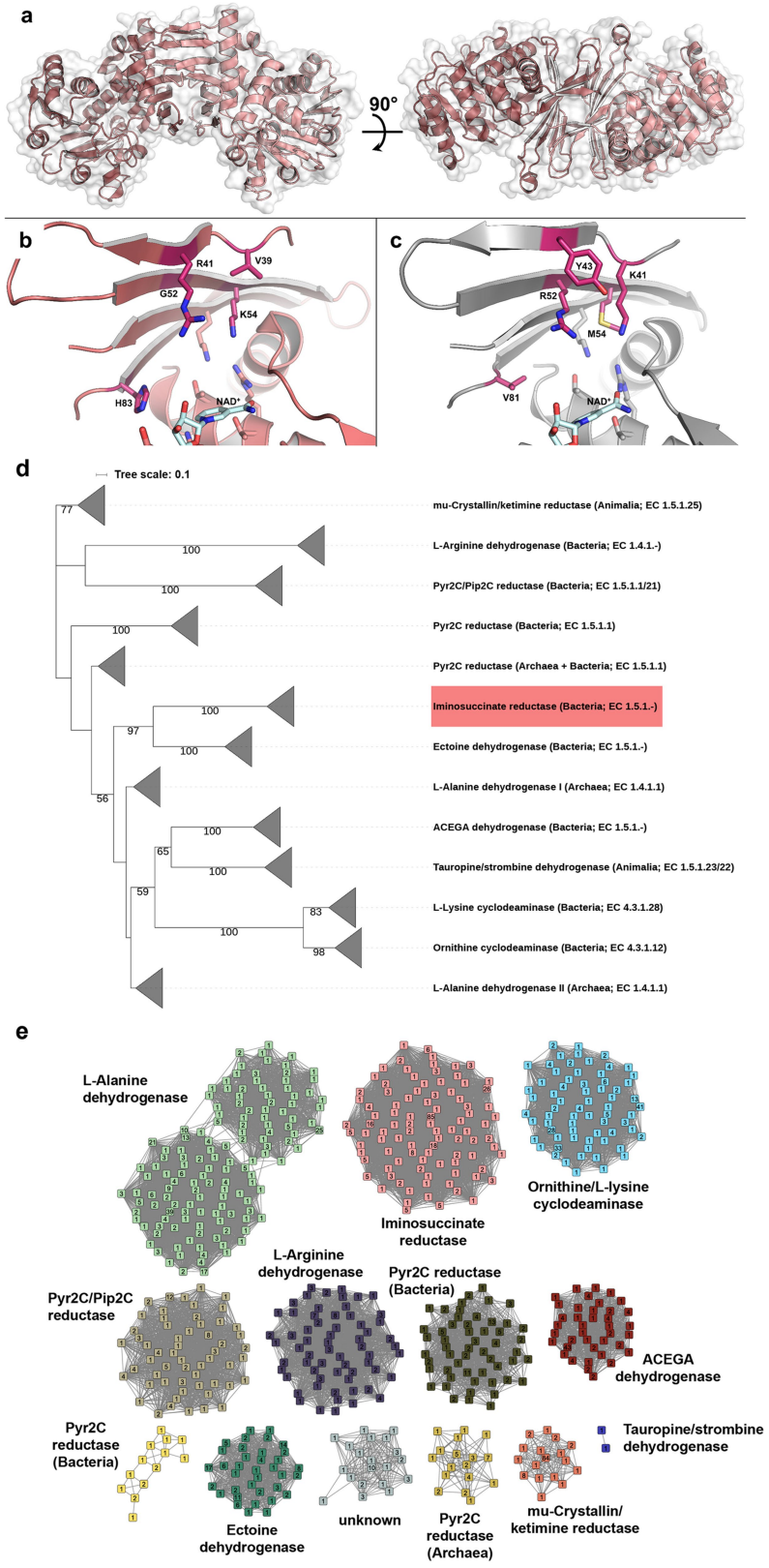
b

Sample ID	Sample	Maximum glycolate concentration [nM]	Analytical method	Reference
E1	Coastal seawater (Ipswich Bay, MA, USA)	1,026	Colorimetry	30
E2	Atlantic Ocean (51°40' S, 57°48' W)	342 ± 53	Colorimetry	31
E3	Antarctic lake (60°42' S, 45°37' W)	526 ± 154		
E4	Coastal seawater (Ipswich Bay, MA, USA)	552	Colorimetry	2
E5	Coastal seawater (Menai Straits, Anglesey, UK)	789	Colorimetry	32
E6	Coastal seawater (Ipswich Bay, MA, USA)	723	Colorimetry	1
E7	Coastal seawater (New York Bight, NY, USA)	475 ± 459	Colorimetry	33
E8	Mediterranean Sea (42°28' N, 30°16' E)	1,183 ± 66	HPLC	10
E9	Atlantic Ocean (oligotrophic waters, 21°01'54" N, 31°09'62" W)	224 ± 39	HPLC	11
E10	Atlantic Ocean (mesotrophic waters, 18°27'22" N, 21°10'18" W)	973 ± 53		
E11	Atlantic Ocean (eutrophic waters, 20°31'49" N, 18°34'39" W)	736 ± 79		
E12	Mediterranean Sea (43°25' N, 7°52' E)	79 ± 53	GC	34
E13	Coastal seawater (Dabob Bay, WA, USA)	100 ± 20	HPLC	4
C1	Culture of *Chlorella*	39,450 – 105,190	^{14}C-tracing	9
C2	Culture of *Euglena gracilis*	591,720	Colorimetry	35
C3	Culture of *Chaetoceros socialis*	240,630	Colorimetry	36
C4	Culture of *Dunaliella tertiolecta*	19,500 ± 1125	HPLC	37
C5	Culture of *Thalassiosira weissflogii*	799	HPLC	38
C6	Culture of *Prochlorococcus* MED4 (phosphorus-limited)	1873 ± 375	HPLC	39
C7	Culture of *Prochlorococcus* MED4 (phosphorus-replete)	2831 ± 458		
C8	Culture of *Prochlorococcus* MIT9312 (phosphorus-limited)	749 ± 916		
C9	Culture of *Prochlorococcus* MIT9312 (phosphorus-replete)	333 ± 125		

Extended Data Fig. 1 | Previously reported glycolate concentrations in environmental samples and cultures of photosynthetic organisms. a, Bar diagram of glycolate concentrations as previously reported in environmental samples. For details on samples, replicates, and analytics see **b** and the literature cited therein. **b**, Table of glycolate concentrations as previously reported in environmental samples (E1, E2 and so on) and cultures of photosynthetic organisms (C1, C2 and so on). When reported in the reference[1,2,4,9–11,30–39], the mean value ± error is given.

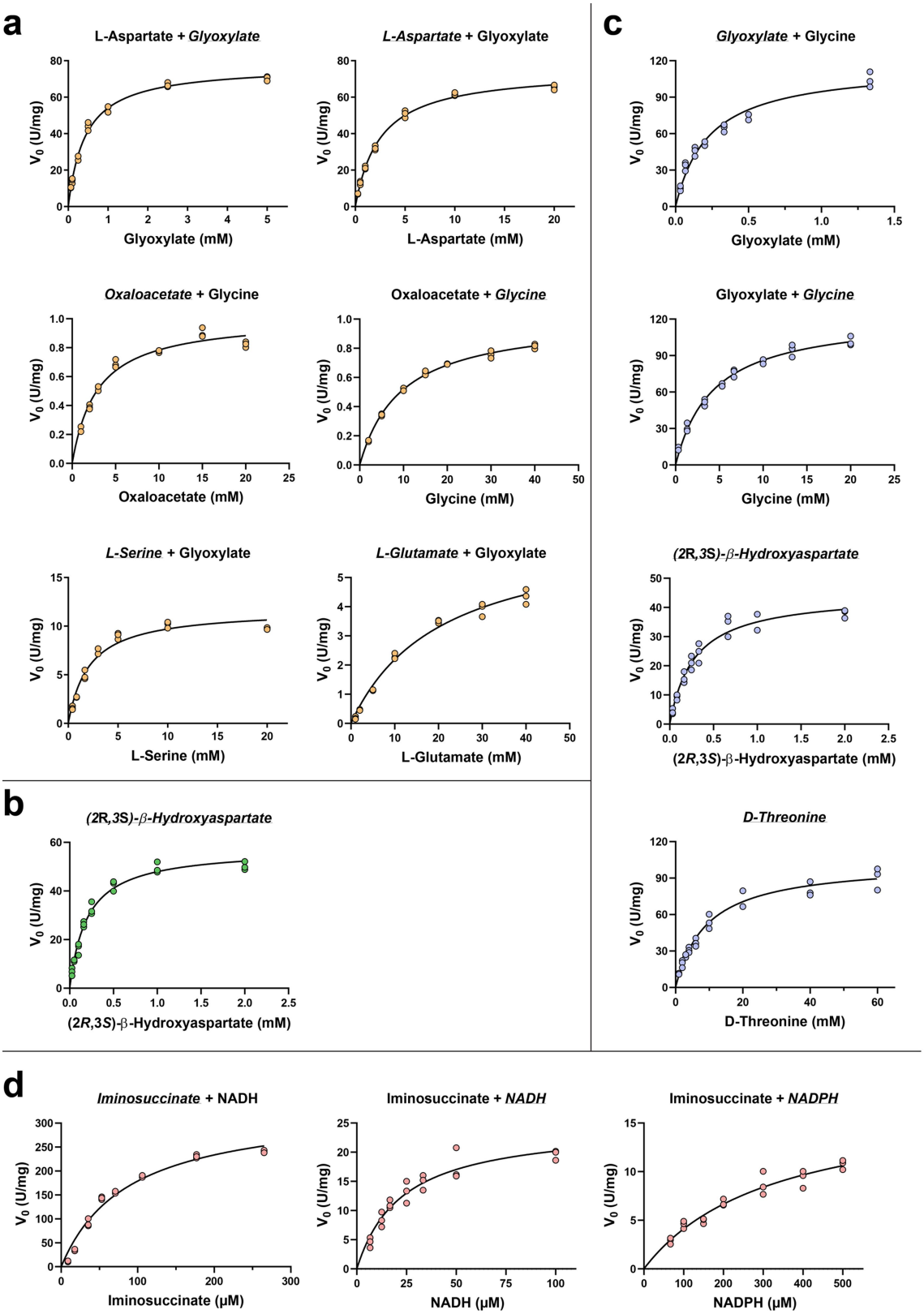


Extended Data Fig. 2 | Crystal structure and phylogenetic analysis of the β-hydroxyaspartate aldolase BhcC. a, Cartoon representation of the β-hydroxyaspartate aldolase homodimer (PDB 6QKB) with superimposed protein surface (left, side view; right, top view). **b**, Active site of β-hydroxyaspartate aldolase with covalently bound PLP (light cyan). Active site residues highlighted in pink (A160, A195 and S313) are completely conserved only among β-hydroxyaspartate aldolases, but differ in D-threonine aldolases. **c**, Active site of D-threonine aldolase (PDB 4V15). The corresponding conserved residues among D-threonine aldolases (Q155, S190 and C303) are highlighted as in **b**. **d**, Maximum likelihood phylogenetic tree of the type III PLP-dependent protein superfamily. Sequences of the β-hydroxyaspartate aldolase BhcC and its homologues form a distinct clade (blue) within the D-threonine aldolase branch of this superfamily. Bootstrap values of at least 50 are given on the respective nodes.

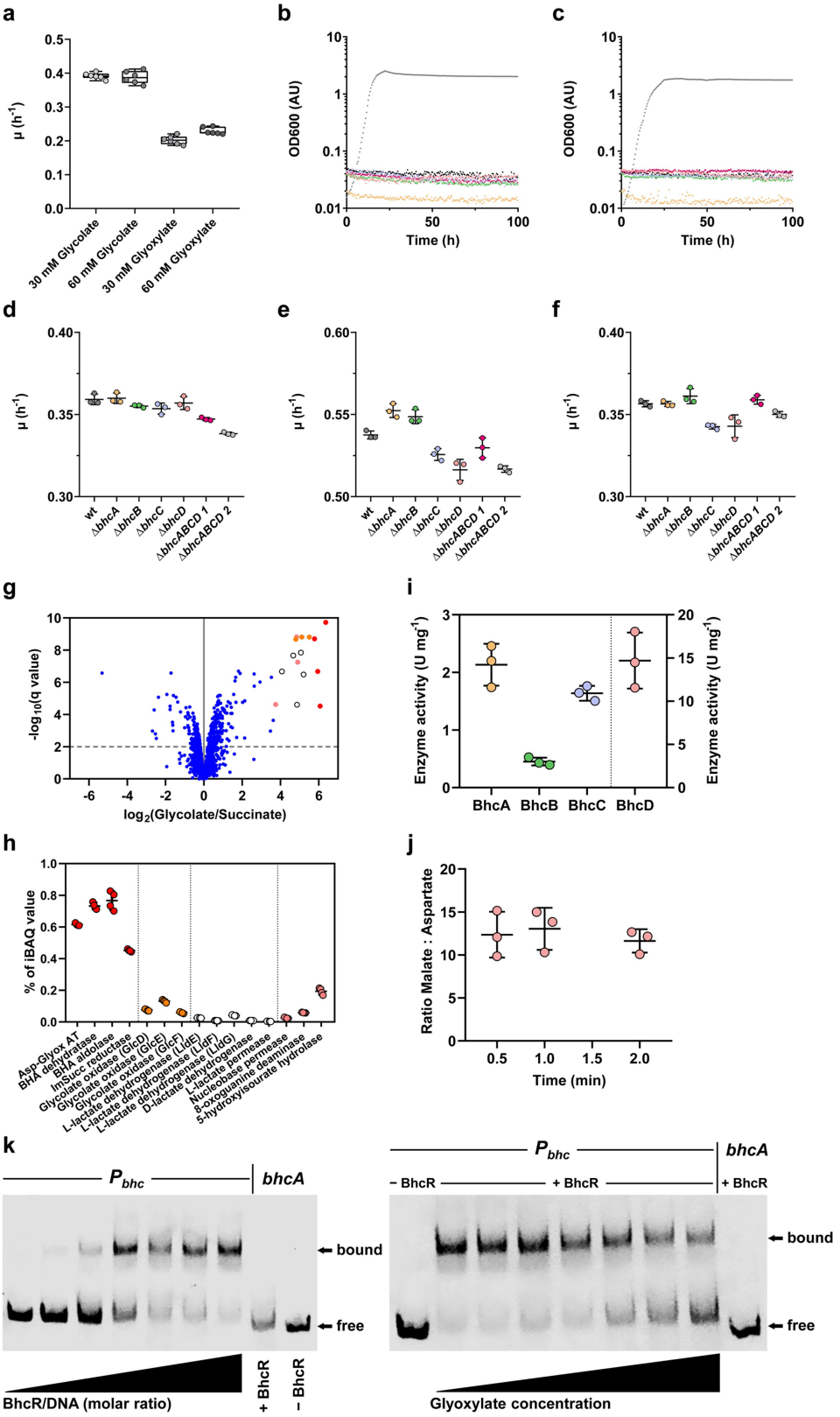


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Crystal structure and phylogenetic analysis of the iminosuccinate reductase BhcD. a, Cartoon representation of the iminosuccinate reductase homodimer (PDB 6RQA) with superimposed protein surface (left, side view; right, top view). **b**, Active site of BhcD with bound NAD^+ (light cyan). Residues highlighted in pink (V39, R41, G52, K54 and H83) may contribute to substrate binding and are conserved among iminosuccinate reductases, but differ in L-alanine dehydrogenases. **c**, Active site of L-alanine dehydrogenase (PDB 1OMO). The corresponding conserved residues among L-alanine dehydrogenases (K41, Y43, R52, M54 and V81) are highlighted as in **b**. **d**, Maximum likelihood phylogenetic tree of the ornithine cyclodeaminase/μ-crystalline protein superfamily. Sequences of the iminosuccinate reductase BhcD and its homologues form a distinct clade (red) within this superfamily. Bootstrap values of at least 50 are given on the respective nodes. **e**, Sequence similarity network of 1,614 sequences from the ornithine cyclodeaminase/μ-crystalline protein superfamily. Connected sequences with more than 80% identity are clustered into nodes. The number in each node gives the number of sequences contained within. Nodes with more than 50% identity are connected by edges. Similar to the phylogenetic analysis shown in **d**, sequences of the iminosuccinate reductase BhcD and its homologues form a distinct clade (red) within this superfamily.

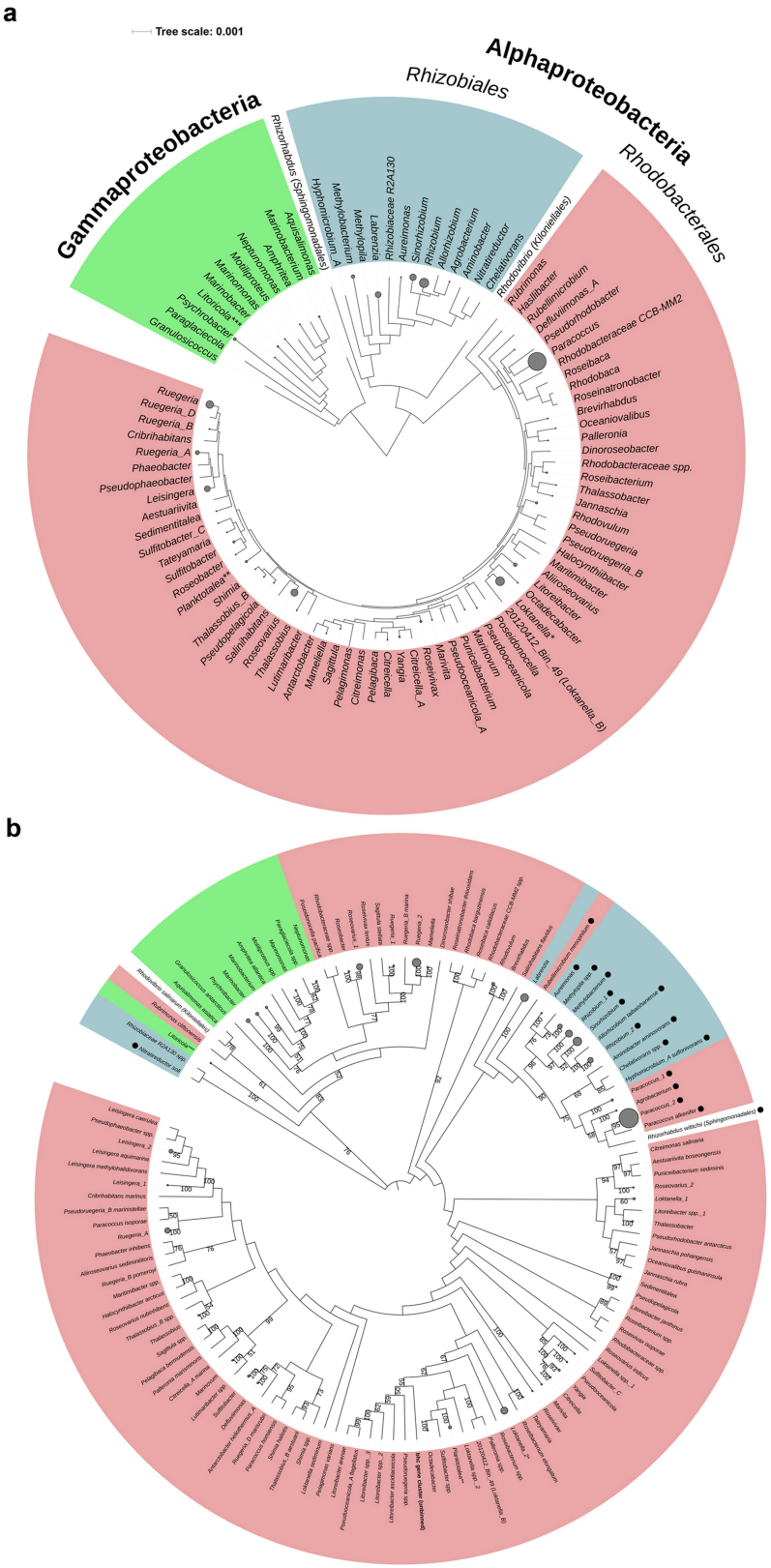


Extended Data Fig. 4 | Michaelis–Menten kinetics of all enzyme reactions characterized in this study. a, Michaelis–Menten kinetics for aspartate–glyoxylate aminotransferase (BhcA). **b**, Michaelis–Menten kinetics for β-hydroxyaspartate dehydratase (BhcB). **c**, Michaelis–Menten kinetics for β-hydroxyaspartate aldolase (BhcC). **d**, Michaelis–Menten kinetics for iminosuccinate reductase (BhcD). **a**–**d**, Data are shown from $n = 3$ independent experiments at different substrate concentrations. The data are summarized in Table 1.



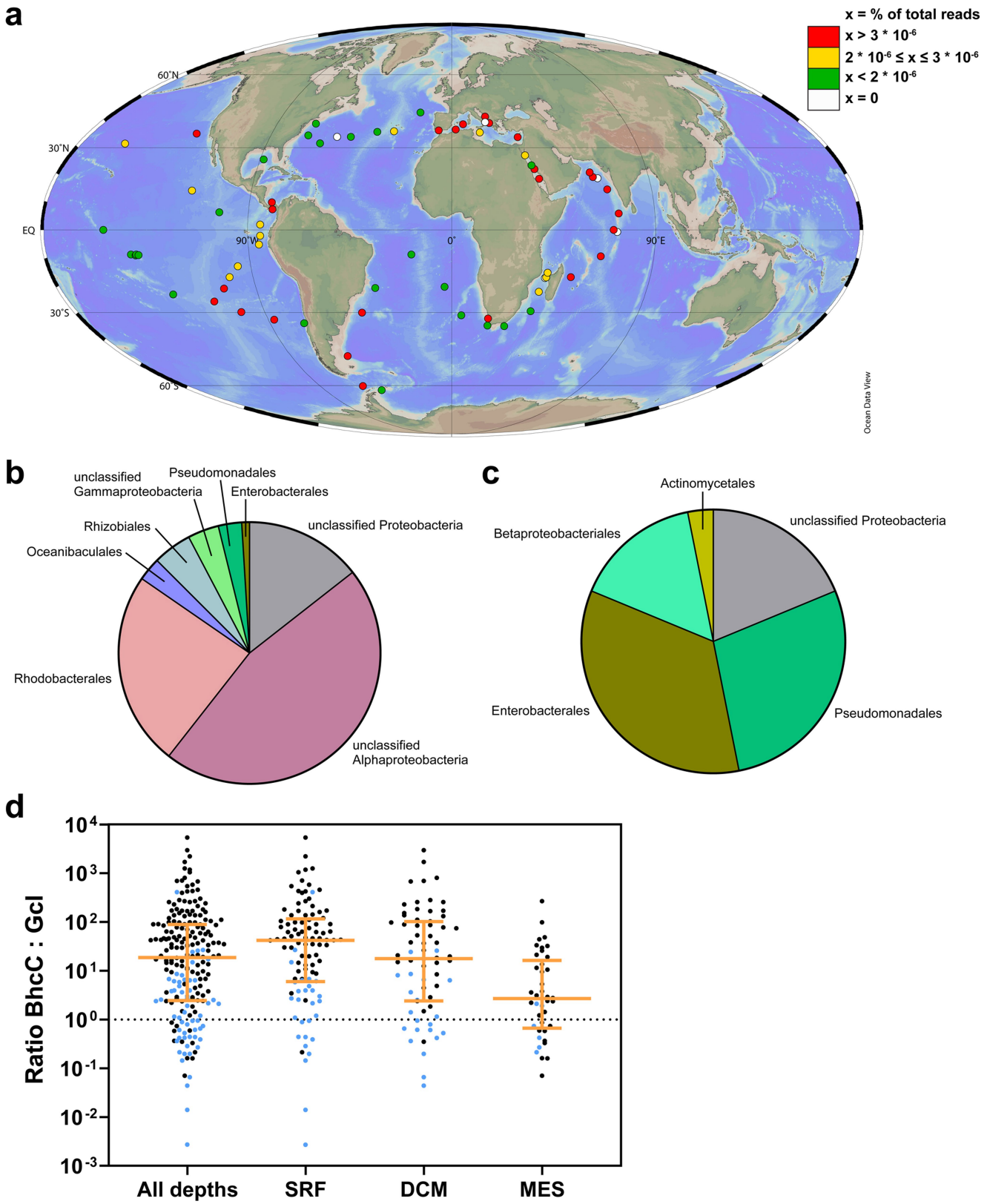
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Physiological role of the BHAC in *P. denitrificans* DSM 413. a, Growth rate of wild-type *P. denitrificans* DSM 413 on the BHAC substrates glycolate and glyoxylate. The middle line and box are the median and interquartile range of $n = 6$ independent experiments and the whiskers indicate the maximum range of the dataset. **b**, **c**, Representative growth curves of wild-type *P. denitrificans* DSM 413 (grey) and bhc deletion strains (coloured) grown in the presence of 60 mM glycolate (**b**) or 60 mM glyoxylate (**c**). Deletion of any single gene in the bhc gene cluster is sufficient to completely abolish growth in the presence of glycolate and glyoxylate. These experiments were repeated three times independently with similar results. **d**–**f**, Growth rates (μ) of wild-type *P. denitrificans* DSM 413 (grey) and BHAC deletion strains (coloured) grown in the presence of 60 mM acetate (**d**), 30 mM succinate (**e**) or 20 mM glucose (**f**). Deletion of any single gene in the bhc gene cluster, or of the whole bhc gene cluster, still permits growth on acetate, succinate or glucose with comparable growth rates as for the wild type. Data are the mean ± s.d. of $n = 3$ independently grown cultures. **g**, Analysis of the proteome of glycolate-grown compared to succinate-grown *P. denitrificans* DSM 413. All proteins that were quantified by at least three unique peptides are shown. The 15 proteins that showed the strongest increase in abundance are marked in the volcano plot. The four enzymes of the BHAC are marked in red, the three subunits of glycolate oxidase in orange, the proteins of a putative operon for lactate utilization in white and the proteins directly downstream of the bhc gene cluster in light red. **h**, The abundance of these proteins, given as the percentage of the intensity-based absolute quantification (iBAQ) value. Data are the mean ± s.d. of $n = 4$ independently grown cultures. **i**, Specific activities of BHAC enzymes in cell-free extracts of glycolate-grown *P. denitrificans* DSM 413, as measured spectrophotometrically. Note that the activity of BhcD is plotted on the right *y* axis and consists of the actual iminosuccinate reductase activity (iminosuccinate to L-aspartate) as well as endogenous malate dehydrogenase activity (oxaloacetate to L-malate). **j**, Ratio of malate to aspartate determined by LC–MS during the enzyme assay for BhcD activity. The ratio remains approximately constant at 12:1, indicating that only approximately 8% of the activity (around 1.3 U mg^{-1}) shown in **i** can be ascribed to iminosuccinate reductase. **i**, **j**, Data are the mean ± s.d. of $n = 3$ independently grown cultures; each data point represents the mean of $n = 3$ technical replicates. **k**, DNA-binding properties of BhcR. Left, a fluorescently labelled DNA fragment carrying the putative promoter region of the bhc gene cluster (P_{bhc}) was incubated with increasing amounts of purified BhcR protein and subsequently separated by electrophoresis to visualize DNA bound to BhcR and free DNA; a DNA fragment derived from the coding region of *bhcA* was used as a negative control. BhcR specifically forms a complex with the DNA fragment containing the putative promoter region of the bhc gene cluster. Right, the P_{bhc}–BhcR complex was incubated with increasing concentrations of glyoxylate and subsequently separated by electrophoresis to assess the effect of glyoxylate on complex formation; the *bhcA* DNA fragment together with BhcR was used as a negative control. Increasing concentrations of glyoxylate decrease the binding of BhcR to the P_{bhc} DNA fragment. For gel source data, see Supplementary Fig. 1.



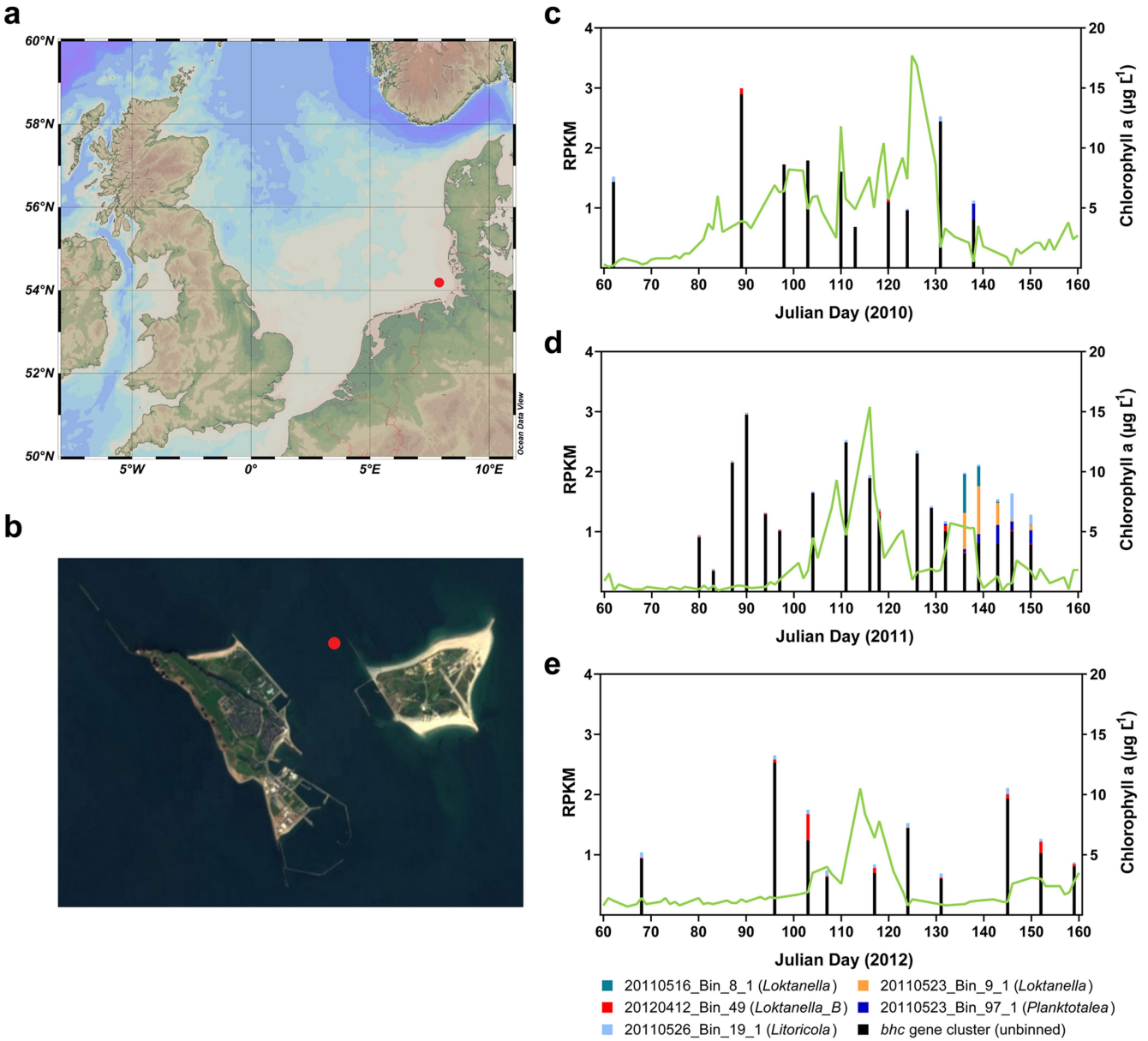
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Phylogenetic analysis of the bhc gene cluster. **a**, Genome-based maximum likelihood phylogenetic tree of bacterial strains with the bhc gene cluster. The bhc gene cluster is found in Gammaproteobacteria (green), and in the alphaproteobacterial orders Rhizobiales (blue) and Rhodobacterales (red), as well as in one member each of Sphingomonadales and Kiloniellales. The phylogenetic tree is based on an alignment of 120 bacterial marker genes from 264 publicly available bacterial genomes and 5 MAGs and was calculated using GTDB-Tk[64] (https://github.com/Ecogenomics/GtdbTk). If several strains from the same genus cluster together, nodes are collapsed at the genus level, and the size of the resulting circle corresponds to the respective number of strains. Loktanella*: collapsed node contains the MAGs 20110516_Bin_8_1 and 20110523_Bin_9_1; Planktotalea**: collapsed node contains the MAG 20110523_Bin_97_1; Litoricola***: collapsed node contains the MAG 20110526_Bin_19_1. **b**, Maximum likelihood phylogenetic tree of concatenated BHAC enzyme sequences. Colour code is the same as in **a**. Phylogenetic groups that were mostly isolated from terrestrial or freshwater habitats are marked with a black dot. Comparison with **a** reveals that the sequences of the BHAC enzymes are not phylogenetically representative, as, for example, alpha- and gammaproteobacterial sequences form a common branch and sequences from terrestrial or freshwater Rhizobiales and Rhodobacterales form another common branch. This suggests that the bhc gene cluster might have been subject to horizontal gene transfer between distantly related strains in shared habitats. The environmental bhc gene cluster sequence that could not be binned successfully is marked in bold and clusters together with isolated representatives of *Pseudoruegeria*, *Litoreibacter* and *Pseudooceanicola*. The phylogenetic tree is based on the concatenated alignments of the 4 enzymes (BhcA–BhcD) from 264 publicly available bacterial genomes and from 6 metagenome contigs. It was calculated using raxmlGUI[67]. Bootstrap values of at least 50 are given on the respective nodes; calculated branch lengths of the tree are ignored for the sake of better visualization. If several strains from the same genus cluster together, nodes are collapsed at the genus level, and the size of the resulting circle corresponds to the respective number of strains. If strains from the same genus cluster in more than one node, the respective branches are labelled as Genus_1, Genus_2, and so on, in a clockwise manner. Loktanella_2*: collapsed node contains the MAGs 20110516_Bin_8_1 and 20110523_Bin_9_1; Planktotalea**: collapsed node contains the MAG 20110523_Bin_97_1; Litoricola***: collapsed node contains the MAG 20110526_Bin_19_1. **a**, **b**, Taxonomy is based on GTDB (release 03-RS86; http://gtdb.ecogenomic.org/). All strains contained in the phylogenetic trees are listed in Supplementary Data 1.

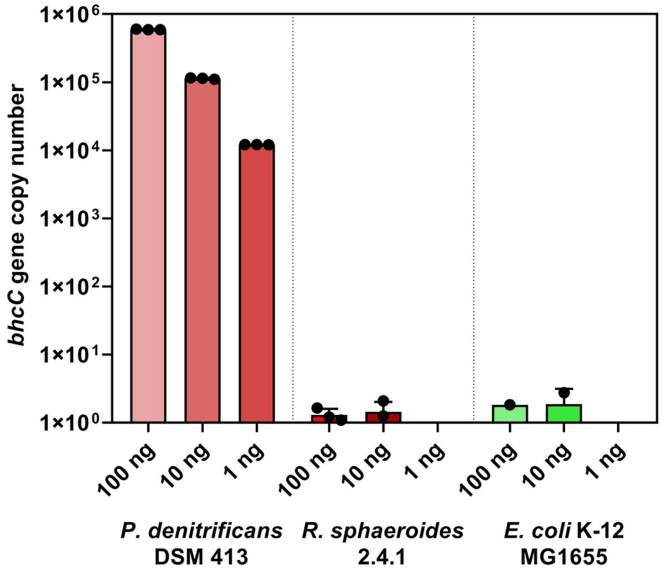


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Glyoxylate assimilation pathways in marine metagenomes. a, Metagenomes collected during the *Tara* Oceans expedition were searched for the presence of BhcC as representative enzyme of the BHAC. Dots on the map mark sampling locations of metagenomes containing BhcC sequences; the colour of the dot corresponds to BhcC abundance in samples from surface water (0.22–3-µm size fraction), as shown in the legend. The map was made with Ocean Data View 5.1.5 (Schlitzer, R., Ocean Data View, odv.awi.de, 2018). **b**, Phylogenetic distribution of 104 unique BhcC sequences found in *Tara* Oceans metagenomes. **c**, Phylogenetic distribution of 32 unique Gcl (as representative enzyme of the glycerate pathway) sequences found in *Tara* Oceans metagenomes. Whereas BhcC is mainly found in Alphaproteobacteria, Gcl is largely restricted to Gammaproteobacteria. **b**, **c**, Taxonomy is based on GTDB[64] (release 03-RS86; http://gtdb.ecogenomic.org/). **d**, Ratio of the abundances (in percentage of total reads) of BhcC to Gcl in *Tara* Oceans metagenomes. BhcC:Gcl ratios from $n = 210$ samples are plotted together (left) and clustered by sampling depth (SRF, upper layer zone ($n = 101$); DCM, deep chlorophyll maximum layer ($n = 68$); MES, mesopelagic zone ($n = 41$)). Samples from the 0.22–3-µm size fraction are denoted by a black dot; samples from the <0.22-µm size fraction are denoted by a blue dot. The median is shown in orange as centre value and error bars represent interquartile ranges. The median BhcC:Gcl ratio of all samples is 18.7. The highest BhcC:Gcl ratio is found in surface water samples (median = 41.8), with the ratio generally being higher in the 0.22–3-µm size fraction than in the <0.22-µm size fraction. Sequence IDs, abundances and BhcC:Gcl ratios are given in Supplementary Data 2.



Extended Data Fig. 8 | Abundance of the bhc gene cluster in Helgoland metagenomes. a, The location of Helgoland Island approximately 40 km offshore the northern German coastline in the North Sea is marked with a red dot. The map was made with Ocean Data View 5.1.5 (R. Schlitzer, Ocean Data View, odv.awi.de, 2018). **b**, The long-term ecological research site 'Kabeltonne' (red dot: 54° 11.3′ N, 7° 54.0′ E) is located between Helgoland Island (left) and the small island Düne (right). Satellite image from WorldWind Explorer (B. Schubert, worldwind.earth/explorer, 2016–2018); the image was adapted to indicate the sampling site. **c**–**e**, Abundance of the bhc gene cluster (in RPKM) was calculated in 38 metagenomes from samples collected during the algal spring blooms of 2010 to 2012 in the North Sea close to Helgoland[28]. Six different sequences were investigated, five of which could be assigned to metagenome bins (Extended Data Fig. 6 and Supplementary Data 3), whereas the remaining, most abundant sequence (black) could not be binned successfully.



Extended Data Fig. 9 | Validation of degenerate *bhcC* primers. Degenerate primers for *bhcC* were used for qPCR with different amounts of genomic DNA from *P. denitrificans* DSM 413, *Rhodobacter sphaeroides* 2.4.1, and *E. coli* K-12 MG1655 as template. While the *bhcC* gene from *P. denitrificans* DSM 413 is amplified, genomic DNA from organisms that lack the bhc gene cluster does not result in reliable amplification. Data are mean ± s.d.; $n = 3$ independent experiments.

Extended Data Table 1 | X-ray diffraction data collection and model refinement statistics

	β-hydroxyaspartate aldolase with bound pyridoxalphosphate (PDB ID 6QKB)	iminosuccinate reductase with bound NAD^+ (PDB ID 6RQA)
Data collection		
Space group	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$
Cell dimensions		
a, *b*, *c* (Å)	66.60, 75.25, 157.31	50.39, 72.41, 164.27
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	29.03 - 1.70 (1.79 - 1.70)	29.40 – 2.56 (2.70 – 2.56)
R_{merge}	0.134 (0.858)	0.097 (0.527)
$I / \sigma I$	10.4 (1.9)	12.6 (3.3)
$CC_{1/2}$ (%)	99.7 (70.8)	99.8 (90.7)
Completeness (%)	99.8 (99.0)	99.9 (100.0)
Redundancy	6.7 (6.5)	6.5 (6.5)
Refinement		
Resolution (Å)	29.03 - 1.70 (1.74 - 1.70)	29.40 – 2.56 (2.70 – 2.56)
No. unique reflections	87194 (5909)	20152 (2893)
R_{work} / R_{free}	0.158 / 0.177	0.176 / 0.225
No. atoms	6671	5027
Protein	5817	4780
Ligands	32	120
Water	822	127
B-factors		
Protein	17.05	52.02
Ligands	23.66	66.03
Water	31.58	47.34
R.m.s. deviations		
Bond lengths (Å)	0.006	0.004
Bond angles (°)	0.84	0.52

Numbers in parentheses indicate statistics for the highest resolution shell. The structures were determined from single crystals.

nature research

Corresponding author(s): Lennart Schada von Borzyskowski, Tobias J. Erb

Last updated by author(s): Sep 4, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data collection and analysis was performed using publicly available software as detailed in citations included in the manuscript and SI. Cary WinUV 5.0.0.999; GraphPad Prism 8.0.0; Excel 2013; BBMap 35.14; GTDBtk 0.1.3; raxmlGUI 1.5b2; XDS BUILT 20180126 ; CCP4 7.0; Phenix 1.14_3260; Coot 0.8.9; Pymol 1.8; eMZed 2.29.4.0; Progenesis QI 2.0; MASCOT 2.5; SafeQuant 2.2.2; iTOL 4.3.2; MUSCLE 3.8.31; Ocean Gene Atlas (tara-oceans.mio.osupytheas.fr/ocean-gene-atlas); Ocean Data View 5.1.5; MassHunter B.09.00; Cytoscape 3.7.1; SWISS-MODEL (swissmodel.expasy.org); CheckM 1.0.7; EFI-EST (efi.igb.illinois.edu/efi-est/index.php); QQQ Quantitative Analysis; j-CODEHOP (4virology.net/virology-ca-tools/j-codehop)
Data analysis	Data collection and analysis was performed using publicly available software as detailed in citations included in the manuscript and SI. Cary WinUV 5.0.0.999; GraphPad Prism 8.0.0; Excel 2013; BBMap 35.14; GTDBtk 0.1.3; raxmlGUI 1.5b2; XDS BUILT 20180126 ; CCP4 7.0; Phenix 1.14_3260; Coot 0.8.9; Pymol 1.8; eMZed 2.29.4.0; Progenesis QI 2.0; MASCOT 2.5; SafeQuant 2.2.2; iTOL 4.3.2; MUSCLE 3.8.31; Ocean Gene Atlas (tara-oceans.mio.osupytheas.fr/ocean-gene-atlas); Ocean Data View 5.1.5; MassHunter B.09.00; Cytoscape 3.7.1; SWISS-MODEL (swissmodel.expasy.org); CheckM 1.0.7; EFI-EST (efi.igb.illinois.edu/efi-est/index.php); QQQ Quantitative Analysis; j-CODEHOP (4virology.net/virology-ca-tools/j-codehop)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The coordinates and structure factors of the crystal structures generated from this research (Fig. 1, Extended Data Fig. 2 + 3) are available at the Protein Data Bank

under accession numbers 6QKB and 6RQA. Mass spectrometry proteomics data (Extended Data Fig. 5) are available via ProteomeXchange with the identifier PXD013274. Metagenome-assembled genomes (MAGs; Extended Data Fig. 6 + 8) are available under accession PRJEB28156 at the European Nucleotide Archive (ENA). All other relevant data are available in this article and its Supplementary Information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. The common rationale was to have a sample size that allows us to calculate a standard deviation. All Michaelis-Menten plots for the kinetic characterization of enzymes include 18 data points (three independent experiments at 6 different substrate concentrations). Controls verifying same levels of specific activities between different enzyme preparations were performed routinely. Three independent experiments (= assays in independent cuvettes) were conducted for the determination of enzyme activities in vitro, both with purified enzymes and in P. denitrificans cell extracts. Three biological replicates (= independent cultures) of P. denitrificans were used to generate cell-free extracts for determination of enzyme activities. Five biological replicates (= independent water samples) were used to determine glycolate concentrations in seawater. Three technical replicates (= repeated injections of the same sample) were measured for the determination of glycolate concentrations in each single seawater sample via LC-MS. Three independent experiments (= independent wells of 96-well qPCR plate) were conducted for all samples and standards in qPCR experiments. Three or six biological replicates (= independent cultures in different wells of 96 well-plate) were measured for the determination of P. denitrificans growth rates. Four biological replicates (= independent cultures) per condition were used to generate biomass for proteomic analysis. Four biological replicates (= independent water samples) were used to determine glycolate uptake rates.
Data exclusions	No data were excluded from the analyses.
Replication	Controls verifying same levels of specific activities between different enzyme preparations were performed routinely. When applicable, we performed our experiments using multiple independent samples (e.g., independent cultures, independent water samples). All attempts at replication of our findings were successful.
Randomization	The microorganisms used in this study were selected and divided randomly in the different conditions. No criteria of selection were applied.
Blinding	Blinding of samples was not applicable for the kind of experiments included in this study. Blinding was not possible because the person in charge of the analysis was the one responsible for taking the samples.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

nature research

Corresponding author(s): Lennart Schada von Borzyskowski, Tobias J. Erb

Last updated by author(s): Sep 4, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data collection and analysis was performed using publicly available software as detailed in citations included in the manuscript and SI. Cary WinUV 5.0.0.999; GraphPad Prism 8.0.0; Excel 2013; BBMap 35.14; GTDBtk 0.1.3; raxmlGUI 1.5b2; XDS BUILT 20180126 ; CCP4 7.0; Phenix 1.14_3260; Coot 0.8.9; Pymol 1.8; eMZed 2.29.4.0; Progenesis QI 2.0; MASCOT 2.5; SafeQuant 2.2.2; iTOL 4.3.2; MUSCLE 3.8.31; Ocean Gene Atlas (tara-oceans.mio.osupytheas.fr/ocean-gene-atlas); Ocean Data View 5.1.5; MassHunter B.09.00; Cytoscape 3.7.1; SWISS-MODEL (swissmodel.expasy.org); CheckM 1.0.7; EFI-EST (efi.igb.illinois.edu/efi-est/index.php); QQQ Quantitative Analysis; j-CODEHOP (4virology.net/virology-ca-tools/j-codehop)
Data analysis	Data collection and analysis was performed using publicly available software as detailed in citations included in the manuscript and SI. Cary WinUV 5.0.0.999; GraphPad Prism 8.0.0; Excel 2013; BBMap 35.14; GTDBtk 0.1.3; raxmlGUI 1.5b2; XDS BUILT 20180126 ; CCP4 7.0; Phenix 1.14_3260; Coot 0.8.9; Pymol 1.8; eMZed 2.29.4.0; Progenesis QI 2.0; MASCOT 2.5; SafeQuant 2.2.2; iTOL 4.3.2; MUSCLE 3.8.31; Ocean Gene Atlas (tara-oceans.mio.osupytheas.fr/ocean-gene-atlas); Ocean Data View 5.1.5; MassHunter B.09.00; Cytoscape 3.7.1; SWISS-MODEL (swissmodel.expasy.org); CheckM 1.0.7; EFI-EST (efi.igb.illinois.edu/efi-est/index.php); QQQ Quantitative Analysis; j-CODEHOP (4virology.net/virology-ca-tools/j-codehop)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The coordinates and structure factors of the crystal structures generated from this research (Fig. 1, Extended Data Fig. 2 + 3) are available at the Protein Data Bank

under accession numbers 6QKB and 6RQA. Mass spectrometry proteomics data (Extended Data Fig. 5) are available via ProteomeXchange with the identifier PXD013274. Metagenome-assembled genomes (MAGs; Extended Data Fig. 6 + 8) are available under accession PRJEB28156 at the European Nucleotide Archive (ENA). All other relevant data are available in this article and its Supplementary Information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Investigation of spring phytoplankton bloom 2018 at Helgoland island. Data on chlorophyll A concentrations, total cell count, glycolate concentrations and uptake rates, and abundance of bhcC genes and transcripts in microbial biomass from water samples are quantitative.
Research sample	Water samples were taken from a designated location (see below) close to Helgoland island; no statistical method was used to predetermine sample size. This sampling site was chosen because it had been used in previous sampling campaigns by the MPI Bremen and the Biologische Anstalt Helgoland. Therefore, it was relevant to take water samples in 2018 that could be compared to samples taken at the same site in previous years.
Sampling strategy	Five biological replicates (= independent water samples) were used to determine glycolate concentrations in seawater. Four biological replicates (= independent water samples) were used to determine glycolate uptake rates. One water sample was used to determine total cell count and chlorophyll A concentration. One water sample was used to obtain microbial biomass for DNA and RNA extraction. No sample size calculation was performed. The common rationale for replicate samples was to have a sample size that allows us to calculate a standard deviation.
Data collection	Sampling was carried out on each working day (Monday - Friday) by the crew of the RV Aade (www.awi.de/en/expedition/ships/more-ships.html) at the research site 'Kabeltonne' (54° 11.3' N, 7° 54.0' E) from approximately 1 m water depth in 20 L carboys. The water samples for microbial biomass were subjected to a fractionating filtration directly upon arrival in the Biologische Anstalt Helgoland laboratory (typically less than one hour after sampling). The entire filtration process for all fractions was usually finished within 3 h, i.e. latest 4 h after the sampling. Further data collection was carried out by the staff of Biologische Anstalt Helgoland and MPI Bremen (chlorophyll A concentrations, total cell count), or Lennart Schada von Borzyskowski (glycolate concentrations and uptake rates, abundance of bhcC genes and transcripts), together with Peter Claus and Nina Socorro Cortina (glycolate concentrations).
Timing and spatial scale	Water samples were collected once daily in the morning (between 8 and 10 am) on each working day between March 1 and May 31, 2018. Water samples were collected in the North Sea close to Helgoland island, at the research site 'Kabeltonne' (54° 11.3' N, 7° 54.0' E), from approximately 1 m water depth.
Data exclusions	No data were excluded from the analyses.
Reproducibility	Not applicable, since sampling was carried out once daily.
Randomization	The water samples used in this study were selected and divided randomly in the different conditions to determine glycolate uptake rates. No criteria of selection were applied.
Blinding	Blinding of samples was not applicable for the kind of experiments included in this study, since water samples (in replicates) were taken on different dates and labeled accordingly.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Water samples were collected once daily in the morning (between 8 and 10 am) on each working day between March 1 and May 31, independent of temperature or weather conditions.
Location	Water samples were collected in the North Sea close to Helgoland island, at the research site 'Kabeltonne' (54° 11.3' N, 7° 54.0' E), from approximately 1 m water depth.
Access and import/export	Water samples were collected without harming the habitat. No import or export of samples was conducted in the frame of this study.
Disturbance	No disturbance was caused by the study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging