BlueRemediomics

Project Number	**101082304**
Project Acronym	**BlueRemediomics**
Project Title	**BlueRemediomics: Harnessing the marine microbiome for novel sustainable biogenics and ecosystem services**
Funding Programme	**Horizon Europe**
Instrument	**RIA**
Project Start Date	**01/12/2022**
Duration of the Project	**48 months**
Deliverable Number and Name	**D4.2 Carbon fixation models**
Work Package	**WP4**
Lead	**CNRS**
Deliverable Due Date	**31/05/2024**
Submission date	**29/05/2024**
Author(s)	**EMBL – Anna Mankowski**
Dissemination Level	**Public**
Type	**R**

Funded by the European Union

BlueRemediomics has received funding from the European Union's Horizon Europe Programme under Grant Agreement No. 101082304. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

UK Research and Innovation

UK Partners on **BlueRemediomics** are supported by UK Research and Innovation (UKRI) under the UK Government's Horizon Europe funding guarantee Grant No. IFS 10061678 (University College London); IFS 10055633 (The Chancellors Masters and Scholars of the University of Cambridge); IFS 10057167 (University of Aberdeen).

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

The Swiss Partner (Eidgenoessische Technische Hochschule Zuerich) on **BlueRemediomics** has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) under Contract No. 22.00384.

CONTENTS

Introduction

Autotrophic CO_2 fixation is the cellular process by which inorganic carbon in the form of carbon dioxide (CO_2) or bicarbonate (HCO_3^-) is assimilated into organic compounds and biomass. Thus, it is a key part of the global carbon cycle and crucial for sustaining life on earth. In addition, autotrophic CO_2 fixation plays a crucial role in managing the global climate crisis. Increasing anthropogenic emissions of greenhouse gases are a main driver of climate change[1], and CO_2 is the most emitted greenhouse gas (34.9 Gt emitted in 2021) [2]. Autotrophic CO_2 fixation and subsequent sequestration in terrestrial and aquatic environments are involved in removing at least parts of CO_2 from the atmosphere[3].

So far, nine autotrophic CO_2 fixation pathways and pathway variants have been characterised or proposed. The most commonly known autotrophic CO_2 fixation pathway is the Calvin Benson Bassham cycle (CBB cycle)[4]. Besides the CBB cycle, six other carbon fixation pathways have been described: the reductive tricarboxylic acid cycle (rTCA)[5], the reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)[6], the 3-hydroxypropionate bicycle (3HPB)[5,7,8,9], the 3-hydroxypropionate/4-hydroxybutyrate cycle[10], the dicarboxylate/4-hydroxybutyrate cycle[11] and the reductive glycine pathway[12,13]. In addition to the description of novel pathways for carbon fixation, variants of known pathways are proposed in which the exchange of individual enzymes can alter e.g., the efficiency of the pathway. For example, an alternative, more energy efficient CBB cycle was proposed in which a PP_i dependent phosphofructokinase could replace the phosphoribulokinase, the sedoheptulose-1,2-bisphopshatase and the fructose-1,6-bisphosphatase[14,15]. Recently a completely reversed version of the TCA cycle, the reverse oxidative TCA cycle (roTCA cycle) was discovered, that can be used for carbon fixation when CO_2 partial pressure is high[16].

Most of the pathways for autotrophic CO_2 fixation occur in a variety of prokaryotic taxa (see e.g. Hügler & Sievert (2011) for an overview of the taxonomic distribution of autotrophic CO_2 fixation pathways in the marine realm)[17]. Since the publication of this review, more prokaryotic lineages capable of autotrophic CO_2 fixation keep being identified, including newly discovered, deep-branching prokaryotic lineages[18,19]. Interestingly, only the reductive acetyl-CoA pathway has been detected in both bacteria and archaea. The other pathways have been detected in either only archaea (the 3-hydroxypropionate/4-hydroxybutyrate

[1] Lacis, A. A. et al. Atmospheric CO2: Principal Control Knob Governing Earth's Temperature. Science (1979) 330, 356–359 (2010).
[2] Liu, Z. et al. Monitoring global carbon emissions in 2021. Nature Reviews Earth & Environment 3, 217–219 (2022).
[3] Friedlingstein, P. et al. Global Carbon Budget 2020. Earth System Science Data 12, 3269–3340 (2020).
[4] Bassham, J. A. & Calvin, M. The path of carbon in photosynthesis. in Die CO_2-Assimilation/The Assimilation of Carbon Dioxide 884–922 (Springer, 1960).
[5] Evans, M. C. et al. A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. PNAS 55, 928–934 (1966).
[6] Ljungdhal, L. G. The Autotrophic Pathway of Acetate Synthesis in Acetogenic Bacteria. Annual Review of Microbiology 40, 415–450 (1986).
[7] Holo, H. *Chloroflexus aurantiacus* secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO_2 and acetate. Archives of Microbiology 151, 252–256 (1989).
[8] Herter, S. et al. A Bicyclic Autotrophic CO_2 Fixation Pathway in *Chloroflexus aurantiacus* *. Journal of Biological Chemistry 277, 20277–20283 (2002).
[9] Zarzycki, J. et al. Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. Proceedings of the National Academy of Sciences 106, 21317–21322 (2009).
[10] Berg, I. A. et al. A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea. Science (1979) 318, 1782–1786 (2007).
[11] Huber, H. et al. A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. PNAS 105, 7851–7856 (2008).
[12] Sánchez-Andrea, I. et al. The reductive glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. Nature Comm. 11, 5090 (2020).
[13] Figueroa, I. A. et al. Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO_2 fixation pathway. PNAS 115, E92–E101 (2018).
[14] Reshetnikov, A. S. et al. Characterization of the pyrophosphate-dependent 6-phosphofructokinase from *Methylococcus capsulatus* Bath. FEMS Microbiol Lett 288, 202–210 (2008).
[15] Kleiner, M. et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. PNAS 109, E1173 LP-E1182 (2012).
[16] Steffens, L. et al. High CO_2 levels drive the TCA cycle backwards towards autotrophy. Nature 592, 784–788 (2021).
[17] Hügler M. & Sievert S. M. Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. Ann Rev Mar Sci 3, 261-289 (2011).
[18] Farnelid H. et al. Cell sorting reveals few novel prokaryote and photosynthetic picoeukaryote associations in the oligotrophic ocean. Environ Microbiol 23, 1469-1480 (2021).
[19] Lannes R. et al. Carbon Fixation by Marine Ultrasmall Prokaryotes. GBE 11 (4), 1166-1177 (2019).

cycle and the dicarboxylate/4-hydroxybutyrate cycle) or only in bacteria (all other pathways listed above). Besides being present in a broad taxonomic diversity of prokaryotes, most autotrophic CO_2 fixation pathways can be fueled by a variety of energy sources, including sunlight (i.e. photosynthesis) or the oxidation of reduced chemical compounds including ammonium, iron(II), hydrogen, methane, nitrite and sulfide (i.e. chemosynthesis)[18].

Photosynthetic organisms, i.e. vascular plants, (micro-) algae and other phytoplankton such as cyanobacteria, use light derived energy to fix ~100 Pg of carbon per year via the CBB cycle in terrestrial and surface aquatic ecosystems[20,21]. Thus, they perform the majority of annual NPP. However, research on the prevalence of pathways for non-photosynthetic carbon fixation in microbiomes from diverse terrestrial and aquatic ecosystems suggests that significant amounts of CO_2 are autotrophically fixed via chemosynthesis. For example, chemosynthetic autotrophic CO_2 fixation appears to be present in the microbiomes of desert soils, including soils from cold deserts[22] as well as sub-humid shrublands, semi-arid grasslands, arid deserts, and hyper-arid deserts[23]. In aquatic microbiomes, chemosynthetic autotrophic CO_2 fixation is not only present in the dark ocean[18,24] but also in coastal to open ocean sediments[25,26] and even in the euphotic water zone of oceans and lakes[26,27]. The ecological relevance of non-phototrophic CO_2 fixation (including chemosynthetic autotrophic CO_2 fixation) in the marine realm is underpinned by an estimated increase of the oceanic NPP by 2.5-11% or 1.2-11 Pg of carbon per year when considering non-phototrophic CO_2 fixation[28].

Despite the global ecological importance of and the ongoing research on autotrophic CO_2 fixation, we still lack a comprehensive understanding of the distribution, evolution and physiology of autotrophic prokaryotes, especially when considering chemosynthetic organisms that fix CO_2 via other pathways than the CBB cycle. In this project, I will fill this knowledge gap through in-depth analyses of the distribution, evolution and ecophysiology of prokaryotic autotrophic CO_2 fixation in globally distributed, environmental microbiomes.

In this task, we aim to characterise the oceans' CO_2 fixation potential of the ocean microbiome that goes beyond photosynthesis. This includes developing a framework to annotate the biochemical pathways that perform CO_2 fixation in global metagenomic datasets, identifying where (geographically and in which habitats) and in which organisms they occur, which energy sources they use in order to distinguish between chemo- and photosynthesis, and ultimately linking the prevalence and abundance of the pathways to known CO_2 fixation rates in order to generate estimates of the non-photosynthetic carbon capture potential of the oceans.

[20] Field C. B. et al. Primary Production of the Biosphere: Integrating Terrestrial and Ocean Components. Science 281 (5374), 237-249 (1998).
[21] Phillips R. & Milo R. A feeling for the numbers in biology. PNAS 106 (51), 21465-21471 (2009).
[22] Ray et al. Atmospheric chemosynthesis is phylogenetically and geographically widespread and contributes significantly to carbon fixation throughout cold deserts. ISME J (2022).
[23] Bay S. K. et al. Chemosynthetic and photosynthetic bacteria contribute differentially to primary production across a steep desert aridity gradient. ISME J 15, 3339-3356 (2021).
[24] Swan B. K. Atmospheric chemosynthesis is phylogenetically and geographically widespread and contributes significantly to carbon fixation throughout cold deserts. Science 333 (6047), 1296-1300 (2011).
[25] Middelburg J. J. Chemoautotrophy in the ocean. Geophys Res Lett 38, L24604 (2011).
[26] Dyksma S. et al. Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. ISME J 10, 1939-1953 (2016).
[27] Jaffe A. L. et al. Variable impact of geochemical gradients on the functional potential of bacteria, archaea, and phages from permanently stratified Lac Pavin. bioRxiv (2022).
[28] Baltar F. & Herndl G. J. Ideas and perspectives: Is dark carbon fixation relevant for oceanic primary production estimates? Biogeosciences 16, 3793-3799 (2019).

Progress

Done

- Development of a computational framework to annotate marker genes of pathways for autotrophic CO_2 fixation in metagenomic datasets
- Screening of global, metagenomic data with the developed models to generate an overview of the phylogeographic distribution of such pathways

Future work

- Validation of the preliminary data
- Annotation of energy conversation pathways
- Linking the prevalence of CO_2 fixation pathways and environmental metadata to published CO_2 fixation rates in order to model the CO_2 fixation potential across the global oceans

Results

Customised Hidden Markov Models for the identification of marker genes from autotrophic pathways

To reliably predict the presence of autotrophic CO_2 fixation, we developed customised reference profiles for Hidden Markov Model (HMM) annotations of marker genes of the respective pathways (profile HMMs, Table 1). The initial profile HMMs were optimised in order to reliably annotate only the marker genes of interest, carefully distinguishing between true positives (the actual gene of interest, TPs) and false positives (closely related and structurally similar, yet functionally different enzymes, FPs).

Table 1: Overview of the marker genes used for the annotation of the different CO_2 fixation pathways. Pathways highlighted in blue are known to occur in Bacteria, pathways highlighted in red are known to occur in Archaea and pathways highlighted in purple are known to occur in both, Bacteria and Archaea.

Pathway	Enzyme name	Gene names used in this report
Calvin Benson Bassham cycle	Ribulose-1,5-bisphosphate carboxylase/oxygenase	I II II.III III.a III.b III.c
3-hydroxypropionate bicycle	L-malyl-CoA/b-methylmalyl-CoA lyase malonyl-CoA reductase / 3-hydroxypropionate dehydrogenase (bacterial form)	mcl mcr_bac
Reductive tricarboxylic acid cycle	ATP citrate lyase	aclA aclB ccl ccsA ccsB
Reductive glycine cycle	Thioredoxin reductase	trx
Wood-Ljungdahl pathway	Carbon monoxide dehydrogenase/acetyl-CoA synthase	cdhD cdhE
3-hydroxypropionate/4-hydroxybutyrate cycle	4-hydroxybutyrate-CoA ligase malonyl-CoA reductase / 3-hydroxypropionate dehydrogenase (archaeal form)	4-hbl_a mcr_arc
Dicarboxylate/4-hydroxybutarate cycle	4-hydroxybutyrate-CoA ligase	4-hbl_c

The optimization was done with a combination of k-fold cross-validation as well as inspection of phylogenetic trees of the TPs, FPs and sequences identified in the proGenomes[29] databased with the original models. A first round of optimisation of the initial profile HMMs led to higher recoveries of TPs and lower recoveries of FPs (Figure 1). For the final search of the genes of interest on a global scale, we screened SPIRE[30] (a database of public, metagenomic data) with the optimised profile HMMs. We cleaned the results by also screening the same data with profile HMMs for the closely related FPs of each of the marker genes. If a predicted gene was picked up with both, the TP and the FP profile HMM, we only considered it truly the gene of interest if the threshold with which it was picked up by the TP profile HMM was higher than the threshold with which it was picked up by the FP profile HMM.

[29] Fullam A. *et al.* proGenomes3: approaching one million accurately and consistently annotated high-quality prokaryotic genomes. Nucleic Acids Res. 51, D760-D766 (2023)

[30] Schmidt T. S. B. *et al.* SPIRE: a Searchable, Planetary-scale mIcrobiome REsource, Nucleic Acids Res. 52 (D1), D777-D783 (2023)

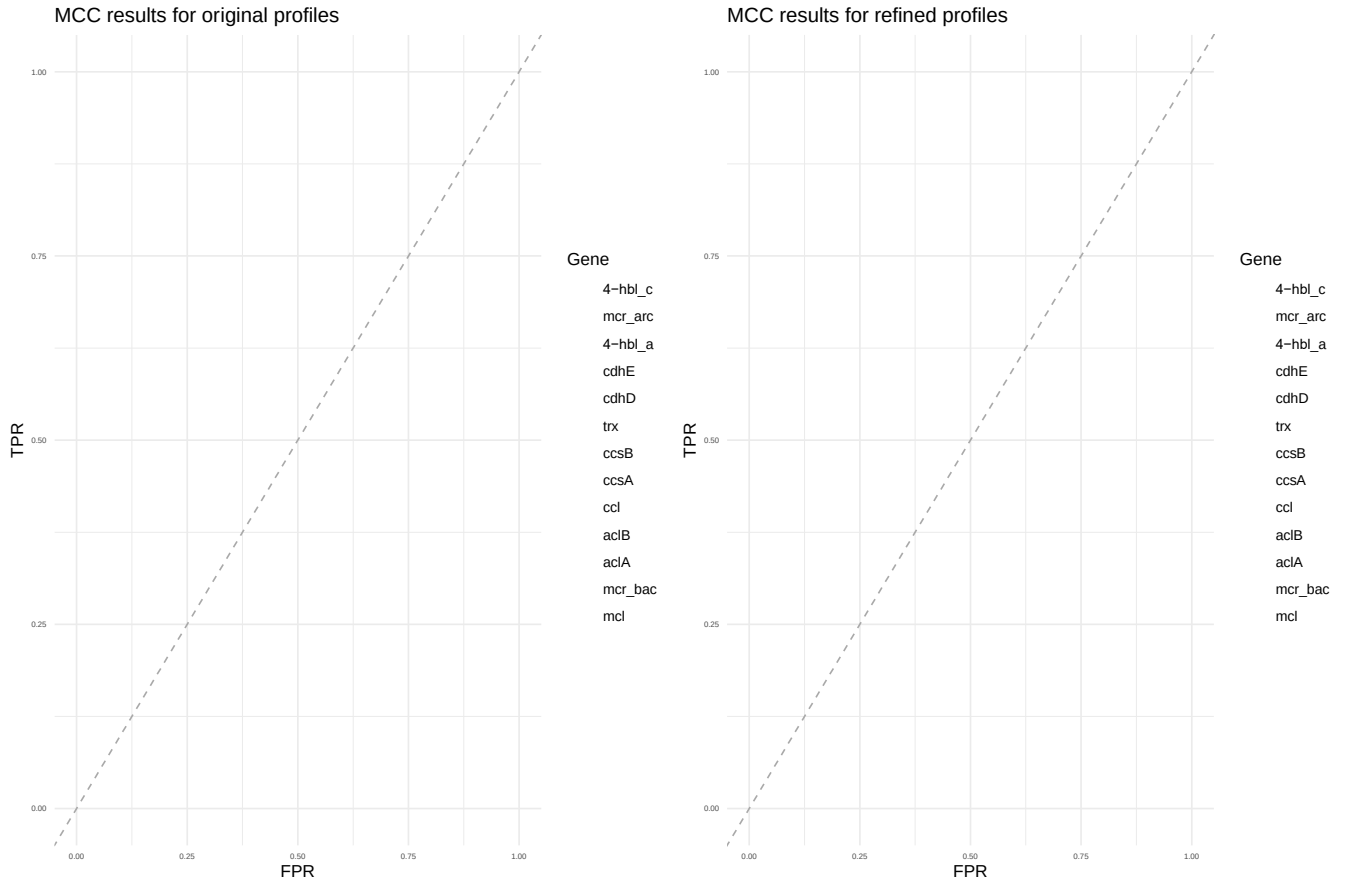

Figure 1: Results of the k-fold cross-validation of the original as well as the optimised profileHMMs for marker genes of autotrophic carbon fixation pathways. The aim are high recovery rates of true positives (TPR, y-axis) and low recovery rates of false positives (FPR, x-axis).

The marine ecosystem contains high numbers of genes involved in CO2 fixation

We initially screened a global collection of metagenomes from diverse habitats. The results were first filtered by only considering samples that were not representing the human microbiome or other animal gut microbiomes. We then checked for the prevalence of the genes in broader environmental categories by counting their occurrences (Figure 2, top left). To place these numbers into perspective, we normalised by the number of samples, resulting in an overview of how many genes per sample were found in each of the environments (Figure 2, top right). Habitats that were classified as “marine” had the third highest count of CO_2 fixation genes per sample, ~ 370, underpinning its importance in the global carbon cycle. To better compare the differences in abundance of the different pathways, we also computed the relative abundances of each of the genes (Figure 2, bottom).

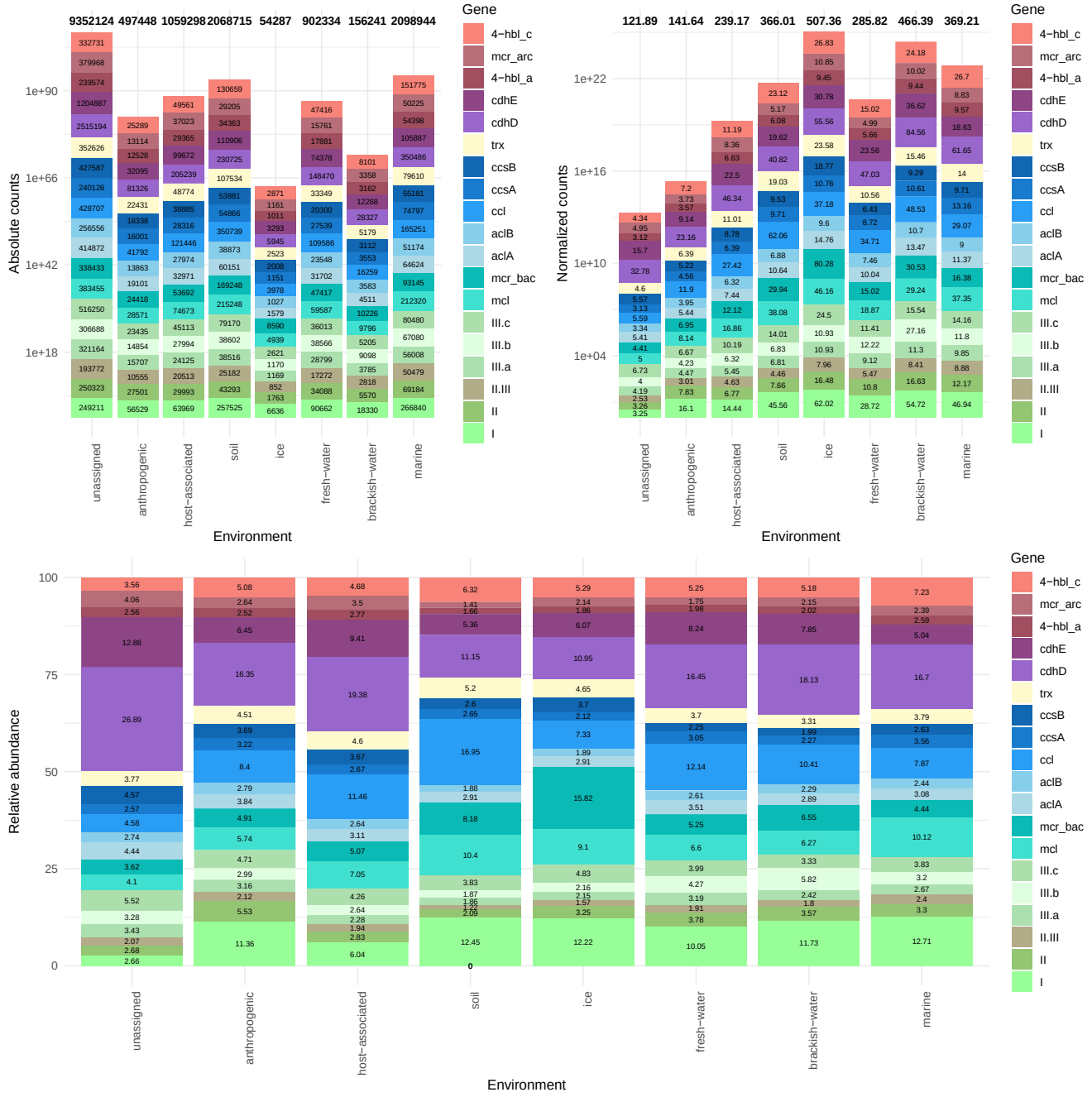

Figure 2: Prevalence of marker genes for autotrophic CO_2 fixation in different environmental habitats.

Autotrophic CO_2 fixation pathways are prevalent in ocean habitats around the planet but differ in their relative abundance

In a first step to analyse the distribution of autotrophic CO_2 fixation pathways in the global oceans, we simply mapped the present of marker genes onto a world map (Figure 3). Apparently, all pathways are present on a global scale and not limited to certain geographic regions. Yet, we would like to improve this analysis by not only considering presence of the genes but also abundance in order to be more sensitive in detecting regional differences.

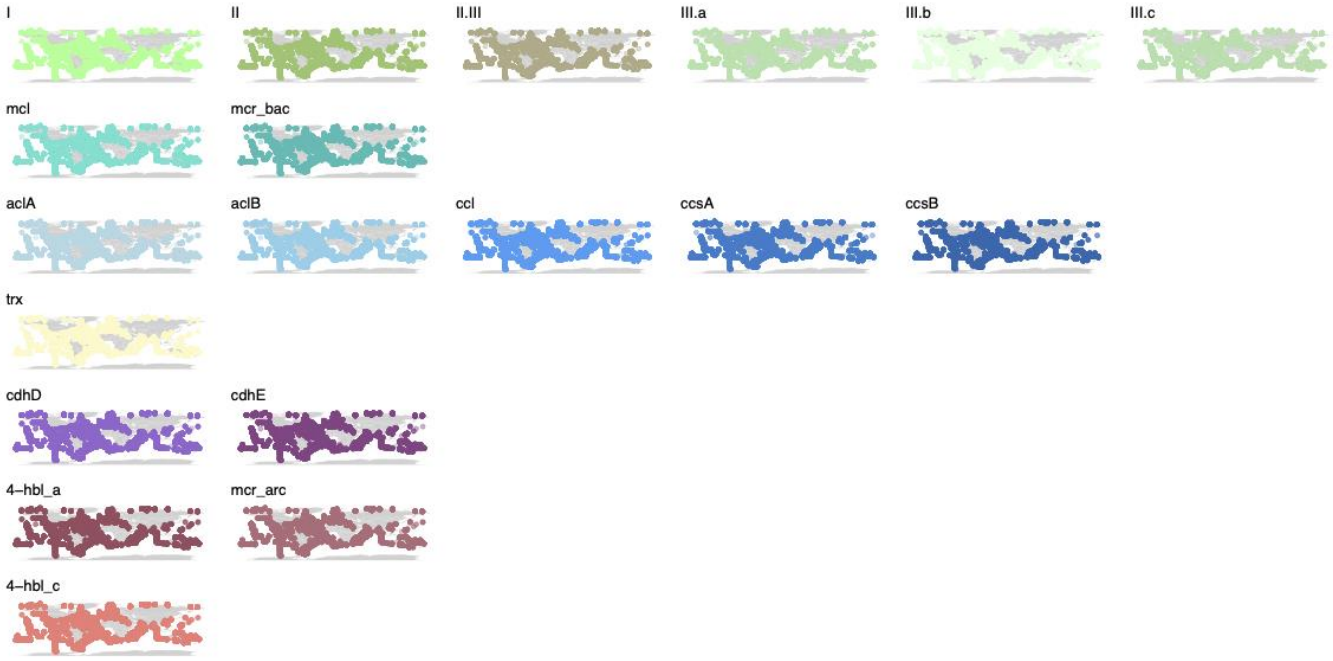

Figure 3: World maps highlight where the different marker genes of autotrophic CO_2 fixation pathways were detected.

For a more fine-scaled analysis of the prevalence of autotrophic CO_2 fixation in the different ocean ecosystems, we filtered our data for samples, that were assigned as “marine” and calculated the relative abundances of the marker genes for these samples (Figure 4). Although this is a very preliminary insight, this analysis illustrates that environmental properties could determine which pathways are most abundant under certain conditions (i.e., high abundance of the rTCA cycle in hydrothermal vent ecosystems or the Wood-Ljungdahl pathway in sediments) which would also impact the efficiency of carbon fixation and in consequence, the amount of CO_2 fixed in a certain environment and thus.

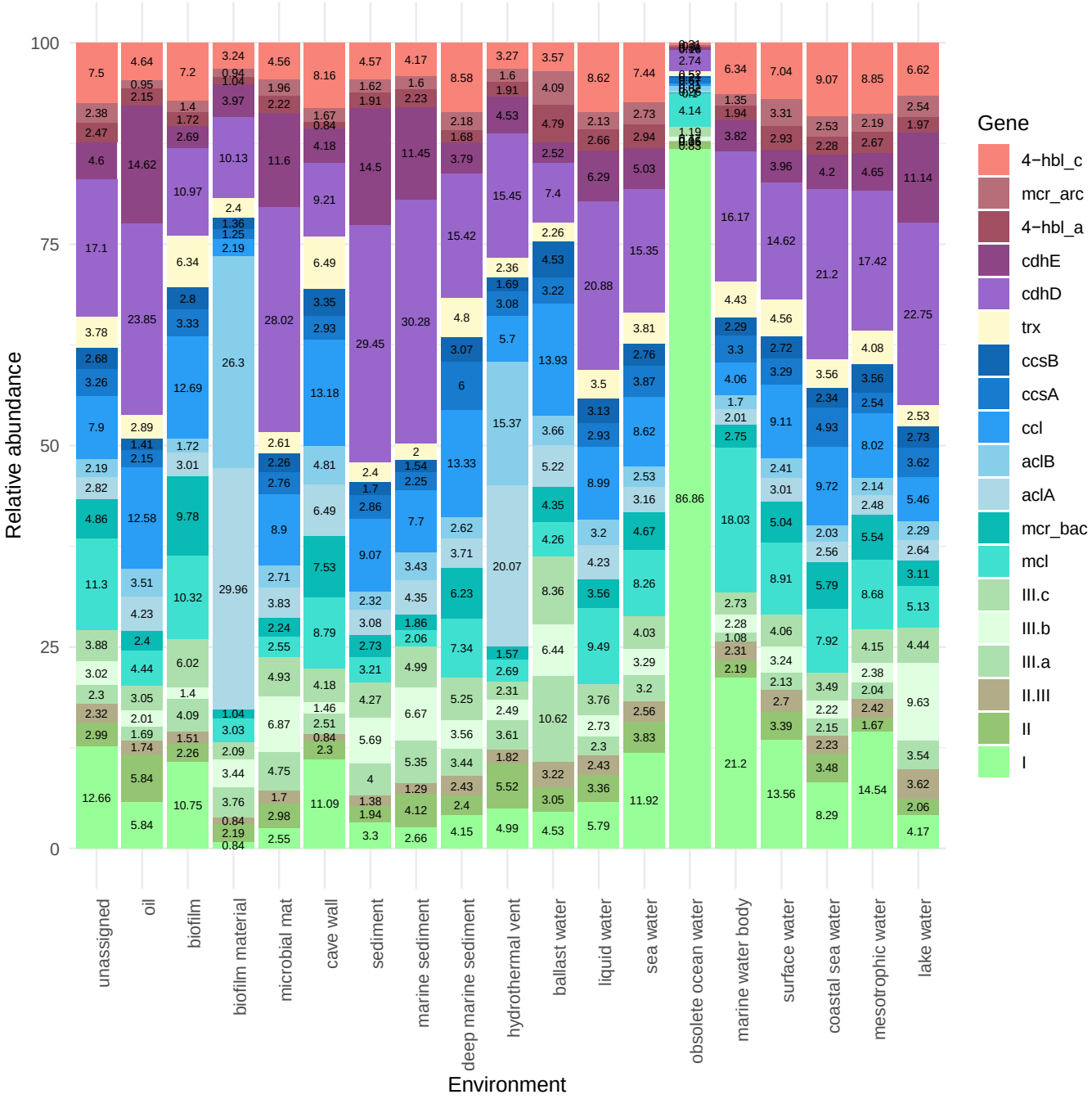

Figure 4: Relative abundance of the different marker genes for autotrophic CO_2 fixation pathways in different marine environments.

Conclusion

We successfully developed the computational framework and the data resources to understand the distribution of autotrophic CO_2 fixation beyond photosynthesis and already have indications that the different autotrophic pathways might play a substantial role in all ocean habitats around the planet. We still need future analysis to prepare a more fine-scaled analysis of the distribution of pathways and ultimately, the estimation of the non-photosynthetic carbon capture potential of the oceans.

Future directions

- Validation of the initial results
- Subset for sequences that are in metagenome assembled genomes (MAGs) that also have the other genes
- Annotate energy conversation pathways in MAGs of verified autotrophs in order to distinguish between photo- and chemosynthetic organisms
- Quantification based on read coverage rather than counts
- Proper modelling of potential rates
- Expand this analysis to non-autotrophic pathways, e.g., anaplerotic CO_2 fixation