# Carbon isotopic signatures of microbial trophic levels: insights from microbial mats

## Citation

Gonzalez-Nayeck, Ana. 2023. Carbon isotopic signatures of microbial trophic levels: insights from microbial mats. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

## Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37374611

## Terms of Use

# Share Your Story

## HARVARD UNIVERSITY
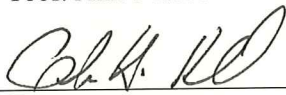### Graduate School of Arts and Sciences

## DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Earth and Planetary Sciences

have examined a dissertation entitled

"Carbon isotopic signatures of microbial trophic levels: insights from microbial mats"

presented by Ana Gonzalez-Nayeck

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature _____

*Typed name*:  Prof. Ann Pearson

Signature _____

*Typed name*:  Prof. Any Knoll

Signature _____

*Typed name*:  Prof. Dave Johnston

*Date*: December 7, 2022

Carbon isotopic signatures of microbial trophic levels: insights from microbial mats


A dissertation presented by

Ana Gonzalez-Nayeck

to

The Department of Earth and Planetary Sciences

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Earth and Planetary Sciences

Harvard University
Cambridge, Massachusetts

December, 2022

Dissertation Advisor: Ann Pearson                    Ana Gonzalez-Nayeck

**Carbon isotopic signatures of microbial trophic levels: insights from microbial mats**

**Abstract**

Microbial mat environments were likely widespread during the Proterozoic and early Paleozoic. As such, interpreting the carbon isotopic compositions ($\delta^{13}$C) of well-preserved organic matter from Proterozoic sediments requires understanding the isotopic consequences of carbon transfer within microbial mats specifically. In modern ecosystems, the $\delta^{13}$C ratios of consumers generally conform to the principle "you are what you eat, +1‰." However microbial mats, with complex diversity yet few taxa capable of phagocytosis, are not easily classified by canonical ecosystem methods. The primary goal of this thesis was to determine whether "you are what you eat, +1‰" applies to microbial heterotrophy. I use two modern microbial mat environments as analogues for ancient microbially-dominated ecosystems: a subaerial, oxygenated and highly photic environment (Chocolate Pots Hot Springs, Yellowstone National Park, USA; Chapter 2) and a submerged, low-oxygen and benthic environment (Middle Island Sinkhole, Lake Huron, USA; Chapter 3). In both instances, I used Protein Stable Isotope Fingerprinting (P-SIF) to measure the $\delta^{13}$C values of whole proteins separated from natural mat samples and classify the same proteins taxonomically via proteomics. In Chapter 2, we found that Cyanobacteria, obligate heterotrophs, and the monosaccharide moieties from exopolysaccharide (EPS) had indistinguishable $\delta^{13}$C signatures. From these data, we concluded that 1) producers and consumers in this system were sharing primary photosynthate as a common resource, and 2) Cyanobacteria were allocating most of their fixed carbon to exopolysaccharides. In Chapter 3, we found that Cyanobacteria (autotrophs), sulfate reducing bacteria (heterotrophs) and sulfur oxidizing bacteria (autotrophs or

mixotrophs), as well as the pentose and hexose moieties of EPS, were all isotopically heterogenous. We hypothesize that these isotopic patterns reflect cyanobacterial metabolic pathways that are relatively more active in low oxygen rather than oxygenated mat environments, resulting in isotopically more heterogeneous C sources in low oxygen mats. In Chapter 4, I use data compiled from the literature to evaluate a potential explanation for the isotopic patterns observed in chapters 2 and 3: that there is a kinetic isotope effect during exopolysaccharide synthesis in Cyanobacteria. Taken as a whole, this thesis cautions against applying "you are what you eat, +1‰" to microbial community food webs, as the $\delta^{13}C$ composition of microbial biomass is more closely tied to specific metabolites than to autotrophy versus heterotrophy. As such, interpretations of the $\delta^{13}C$ values in sediments derived from predominantly microbial ecosystems should be developed relative to the $\delta^{13}C$ values of specific molecular-level carbon sources.

**Table of Contents**

A mi familia, con la que nací y la que apareció después

## Acknowledgements

**Chapter 1**

**Introduction**

Unlike most ecosystems on Earth today, Precambrian and early Paleozoic ecosystems were mostly microbial. This is supported via evidence from ichnofossils suggesting that bioturbation did not reach modern intensity until the mid-Paleozoic (Tarhan et al., 2015), via the observation of microbial textures in Proterozoic rocks (Hagadorn and Bottjer, 1997; Gehling, 1999; Steiner and Reiter, 2001; Callow and Brasier, 2009), and most famously via the prevalence of stromatolites (lithifying microbial mats) in Archean rocks (Awramik, 1992; Hoffman, 2000; Schopf et al., 2007; Djokic et al., 2021). The closest modern analogues to these ancient ecosystems are environments that exclude animals due to their extreme temperature, salinity, or redox conditions, thereby allowing for microbial mats to grow (Canfield and des Marais, 1993; Burns et al., 2009). For example, researchers have used modern microbial mats to investigate the likelihood that Proterozoic mat environments hosted the first origin of eukaryotes (López-García and Moreira, 2020) and whether oxygenic cyanobacterial mats could have sustained animal life in otherwise low-oxygen Proterozoic ecosystems (Gingras et al., 2011). In these and other examples, interactions between microbes play an important role: the endosymbiosis that led to the first origin of eukaryotes was likely instigated by resource sharing between a heterotrophic bacterium and archaea, and cyanobacterial mats are only net sources of oxygen if oxygenic photosynthesis exceeds organic carbon respiration by adjacent heterotrophic bacteria. More broadly, understanding trophic interactions in microbial ecosystems is critical towards reconstructing the flow of carbon and energy resources in past environments.

There are numerous indications that the macrofaunal concept of "trophic levels" should be avoided in mat-dominated ecosystems. Some consumer taxa in mats are strictly heterotrophic, while others are mixotrophic or have flexible carbon metabolisms (Bennett et al., 2020; Hamilton et al., 2019; Klatt et al., 2013; van der Meer et al., 2007). One of the major organic carbon sources in mats, the binding matrix of extracellular polymeric substances (EPS, mostly exopolysaccharide), can be accessed by heterotrophic organisms via either extracellular digestion to monomers or by fermentation to smaller carbon units before assimilation (Anderson et al., 1987; Flemming & Wingender, 2010; Stuart et al., 2016). This complexity introduces challenges when using the classical methods of geobiology and isotope geochemistry.

Trophic levels in modern ecosystems are analyzed using both carbon ($\delta^{13}$C) and nitrogen ($\delta^{15}$N) isotope ratios (Fry and Sherr, 1984); (Cabana and Rasmussen, 1996). The colloquial phrase "you are what you eat, + 1‰," (De Niro, Michael, Epstein, 1978) reflects the $\delta^{13}$C values of small animals capable of holozoic feeding when grown on a complex diet. The analogous offset in $\delta^{15}$N values between trophic levels is *ca*. 3‰, making nitrogen isotopes the preferred tool to discern trophic structure in modern environments (Post, 2002). However, due to poor and non-specific nitrogen preservation, carbon – specifically the $\delta^{13}$C values of lipid biomarkers – remains more widely used to probe the ecology of ancient systems (Hayes et al., 1989; Freeman et al., 1994). While prior studies have assumed that "you are what you eat, + 1‰" applies to microbial heterotrophy (*e.g.,* Logan et al., 1995; Close et al., 2011), validation requires taxonomically resolved, natural abundance $\delta^{13}$C values at sub-1‰ resolution. A central goal of this thesis is to characterize the isotopic signatures of microbial heterotrophy, using modern microbial mat samples as analogues for microbially-dominated ancient ecosystems.

In **Chapter 2,** we use Protein Stable Isotope Fingerprinting (P-SIF), a method previously developed in the Pearson lab, to determine the trophic relationships in a microbial mat sample from Chocolate Pots Hot Springs, Yellowstone National Park (YNP), USA. P-SIF allows measurement of the $\delta^{13}C$ values of whole proteins, separated from environmental samples and identified taxonomically via proteomics. We find that in this highly photic and oxygenated mat, proteins from heterotrophic bacteria are indistinguishable from cyanobacterial proteins, indicating that "you are what you eat, +1‰" is not applicable. Cyanobacteria in similar mat-based ecosystems excrete large quantities of photosynthetic sugars into the surrounding environment. As such, we hypothesized that the measured isotopic similarity reflects an ecosystem where producers and consumers share primary photosynthate as a common resource. This idea was validated by confirming that glucose moieties in EPS were equal in $\delta^{13}C$ composition to both cyanobacterial and heterotrophic proteins. Overall, we confirm that the $\delta^{13}C$ composition of microbial biomass is tied to specific metabolites rather than to autotrophy versus heterotrophy or to individual trophic levels. In 1985, Neil Blair used laboratory cultures to demonstrate that isotopically, aerobic microbial heterotrophy is a case of "you are what you eat." To our knowledge, Chapter 2 provides the first evidence of this from an environmental sample.

In **Chapter 3,** we use P-SIF to determine the trophic relationships in a microbial mat sample from the submerged Middle Island Sinkhole (MIS) in Lake Huron, Michigan USA. This site has been previously identified as a potential analogue for benthic Proterozoic ecosystems in part due to its relatively low concentrations of dissolved oxygen (0-2 mg L$^{-1}$) amongst other geochemical parameters. We determined in Chapter 2 that the $\delta^{13}C$ composition of microbial biomass is tied to

specific metabolites. Relative to microbes in more oxygenated ecosystems, microbes at MIS are more likely to engage in carbon metabolisms such as fermentation that release multiple types of organic sources for heterotrophs. In this sample from the MIS, Cyanobacteria (autotrophs) were $^{13}$C-depleted relative to sulfate reducing bacteria (heterotrophs) and $^{13}$C-enriched relative to sulfur oxidizing bacteria (autotrophs or mixotrophs); furthermore, pentose moieties of EPS were systematically enriched in $^{13}$C relative to the hexose moieties of EPS. We hypothesize that these isotopic patterns reflect cyanobacterial metabolic pathways, particularly phosphoketolase, that are relatively more active in low oxygen rather than oxygenated mat environments, resulting in isotopically more heterogeneous C sources in low oxygen mats. Chapter 3 underscores the importance of flexible cyanobacterial metabolisms towards determining the $\delta^{13}$C composition of heterotrophic organic carbon sources.

In Chapters 2 and 3, we rely on the assumption that polyglucose synthesis carries an isotopic fractionation to explain our findings. In **<u>Chapter 4</u>**, I evaluate this assumption theoretically using published values for the $\delta^{13}$C compositions of hexose and pentose sugars, while assuming that the former represent primarily storage/excreted sugars and the latter represent primarily internal sugars. I find evidence for a systematic kinetic isotope effect during polyglucose synthesis of approximately 4‰ but note that controlled laboratory culture studies are a critical next step towards elucidating the exact mechanism responsible for this isotope effect.

In **<u>Chapter 5</u>**, I synthesize my findings from chapters 2-4 and emphasize a critical gap towards understanding the $\delta^{13}$C compositions of heterotrophic organic carbon sources: quantifying the isotopic consequences of cyanobacterial sugar synthesis, storage and excretion.

# References

Awramik S. M. (1992) The oldest records of photosynthesis. *Photosynth Res* **33**, 75–89.

Burns B. P., Anitori R., Butterworth P., Henneberger R., Goh F., Allen M. A., Ibañez-Peral R., Bergquist P. L., Walter M. R. and Neilan B. A. (2009) Modern analogues and the early history of microbial life. *Precambrian Res* **173**, 10–18.

Cabana G. and Rasmussen J. B. (1996) Comparison of aquatic food chains using nitrogen isotopes. *Proc Natl Acad Sci U S A* **93**, 10844–10847.

Callow R. H. T. and Brasier M. D. (2009) Remarkable preservation of microbial mats in Neoproterozoic siliciclastic settings: Implications for Ediacaran taphonomic models. *Earth Sci Rev* **96**, 207–219.

Canfield D. E. and des Marais D. J. (1993) Biogeochemical cycles of carbon, sulfur, and free oxygen in a microbial mat. *Geochim Cosmochim Acta* **57**, 3971–3984.

Djokic T., Kranendonk M. J. van, Campbell K. A., Havig J. R., Walter M. R. and Guido D. M. (2021) A Reconstructed Subaerial Hot Spring Field in the *3.5 Billion-Year-Old Dresser Formation, North Pole Dome, Pilbara Craton, Western Australia. *Astrobiology* **21**.

Freeman K. H., Wakeham S. G. and Hayes J. M. (1994) Predictive isotopic biogeochemistry of lipids from marine anoxic basins. *Org Geochem* **21**, 629–644.

Fry B. and Sherr E. B. (1984) delta super(13)C measurements as indicators of carbon flow in marine and freshwater ecosystems. *Contrib. Mar. Sci.* **27**, 13–47.

Gehling J. G. (1999) Microbial mats in terminal Proterozoic siliciclastics; Ediacaran death masks. *Palaios* **14**, 40–57.

Hagadorn J. W. and Bottjer D. J. (1997) Wrinkle structures: Microbially mediated sedimentary structures common in subtidal siliciclastic settings at the Proterozoic-Phanerozoic transition. *Geology* **25**, 1047–1050.

Hayes J. M., Freeman K. H., Popp B. N. and Hoham C. H. (1989) Compound specific isotope analysis, a novel tool for reconstruction of ancient biogeochemical processes. *Org Geochem* **16**, 1115–1128.

Hoffman H. J. (2000) Archean stromatolites as microbial archives. In *Microbial Sediments*

de Niro, Michael, Epstein S. (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochim Cosmochim Acta* **42**, 495–506.

Post D. M. (2002) Using stable isotopes to estimate trophic position: Models, methods, and assumptions. *Ecology* **83**, 703–718.

Schopf J. W., Kudryavtsev A. B., Czaja A. D. and Tripathi A. B. (2007) Evidence of Archean life: Stromatolites and microfossils. *Precambrian Res* **158**, 141–155.

Steiner M. and Reiter J. (2001) Evidence of organic structures in Ediacara-type fossils and associated microbial mats. *Geology* **29**, 1119–1122.

Tarhan L. G., Droser M. L., Planavsky N. J. and Johnston D. T. (2015) Protracted development of bioturbation through the early Palaeozoic Era. **8**.

# Chapter 2

## Absence of canonical trophic levels in a microbial mat

**Abstract**

In modern ecosystems, the carbon stable isotope ($\delta^{13}C$) ratios of consumers generally conform to the principle "you are what you eat, +1‰." However, this metric may not apply to microbial mat systems where diverse communities, using a variety of carbon substrates via multiple assimilation pathways, live in close physical association and phagocytosis is minimal or absent. To interpret the $\delta^{13}C$ record of the Proterozoic and early Paleozoic, when mat-based productivity likely was widespread, it is necessary to understand how a microbially driven producer-consumer structure affects the $\delta^{13}C$ compositions of biomass and preservable lipids. Protein Stable Isotope Fingerprinting (P-SIF) is a recently developed method that allows measurement of the $\delta^{13}C$ values of whole proteins, separated from environmental samples and identified taxonomically via proteomics. Here, we use P-SIF to determine the trophic relationships in a microbial mat sample from Chocolate Pots Hot Springs, Yellowstone National Park (YNP), USA. In this mat, proteins from heterotrophic bacteria are indistinguishable from cyanobacterial proteins, indicating that "you are what you eat, +1‰" is not applicable. To explain this finding, we hypothesize that sugar production and consumption dominate the net ecosystem metabolism, yielding a community in which producers and consumers share primary photosynthate as a common resource. This idea was validated by confirming that glucose moieties in exopolysaccharide were equal in $\delta^{13}C$ composition to both cyanobacterial and heterotrophic proteins, and by confirming that highly $^{13}C$-depleted fatty acids (FAs) of Cyanobacteria dominate the lipid pool, consistent with

flux-balance expectations for systems that overproduce primary photosynthate. Overall, the results confirm that the $\delta^{13}C$ composition of microbial biomass and lipids is tied to specific metabolites, rather than to autotrophy versus heterotrophy or to individual trophic levels. Therefore, we suggest that aerobic microbial heterotrophy is simply a case of "you are what you eat."

**Introduction**

Modern analogs are essential when contextualizing the carbon ($\delta^{13}C$) isotope ratios of organic matter from ancient ecosystems. Modern microbial mats may help illuminate Precambrian environments, given the prevalence of stromatolites in Archean rocks (Awramik, 1992; Hoffman, 2000; Schopf et al., 2007; Djokic et al., 2021), the observation of microbial textures in Proterozoic rocks (Callow & Brasier, 2009; Gehling, 1999; Hagadorn & Bottjer, 1997; Steiner & Reiter, 2001), and evidence that bioturbation did not reach modern intensity until the mid-Paleozoic (Tarhan, 2018). Furthermore, non-lithifying (i.e., non-stromatolite) microbial mats, which represent the majority of modern mat morphologies but are poorly preserved in the geologic record, might represent the majority of mat-based carbon fixation during the Precambrian (Schuler et al., 2017). Despite this importance, microbial mats with complex diversity yet few taxa capable of phagocytosis are not easily classified by canonical ecosystem methods (Anderson et al., 1987; van der Meer et al., 2007; Flemming & Wingender, 2010; Klatt et al., 2013; Stuart et al., 2016; Hamilton et al., 2019; Bennett et al., 2020). There are numerous indications that the macrofaunal concept of "trophic levels" should be avoided in mat-dominated ecosystems. Some consumer taxa in mats are strictly heterotrophic, while others are mixotrophic or have flexible carbon metabolisms (Bennett et al., 2020; Hamilton et al., 2019; Klatt et al., 2013; van der Meer et al., 2007). One of the major organic carbon sources in mats, the binding matrix of extracellular

polymeric substances (EPS, mostly exopolysaccharide), can be accessed by heterotrophic organisms via either extracellular digestion to monomers or by fermentation to smaller carbon units before assimilation (Anderson et al., 1987; Flemming & Wingender, 2010; Stuart et al., 2016). This complexity introduces challenges for reconstructing the flow of carbon and energy resources in past environments using the classical methods of geobiology and isotope geochemistry.

Trophic levels and the structure of modern, macrofaunal ecosystems traditionally are investigated through the lens of carbon ($\delta^{13}$C) and nitrogen ($\delta^{15}$N) isotope ratios (Cabana & Rasmussen, 1996; Fry & Sherr, 1984). The colloquial phrase "you are what you eat, +1‰," is grounded in a study of the $\delta^{13}$C values of small animals capable of holozoic feeding when grown on a complex diet, while simultaneously disguising the underpinning molecular-level heterogeneity that comprises both the food and the consumer (DeNiro & Epstein, 1978). For example, $\delta^{13}$C values for individual biochemical fractions from a fly fed horsemeat ranged from approximately 3‰ more negative (lipids) to 1‰ more positive (soluble protein) than the horsemeat substrate. This distribution is expected, given the fractionation associated with lipid synthesis (DeNiro & Epstein, 1977; Melzer & Schmidt, 1987; Monson & Hayes, 1982). It also reinforces that "you are what you eat, + 1‰" may only apply to organisms that consume whole prey, ingest herbaceous material, filter-feed on a diet of zooplankton, or can engulf whole microbes (Fry & Sherr, 1984; Fantle et al., 1999; Pinnegar & Polunin, 2000; van der Zanden & Rasmussen, 2001). By contrast, microbes incapable of phagocytosis selectively utilize specific substrates (Arnosti et al., 2011; Mahmoudi et al., 2017). Concomitantly, most bacteria repress the expression of catabolic enzymes for secondary carbon sources in the presence of their preferred substrate (Görke & Stülke, 2008).

Despite these arguments that microbial ecosystems should not have trophic organization, the "you are what you eat, +1‰" concept is used in the geobiological and geochemical literature as an estimate for the $\delta^{13}C$ signature of microbial heterotrophy, typically in the context of using lipid biomarkers to probe the ecology of ancient systems (e.g., Close et al., 2011; Logan et al., 1995; Luo et al., 2015; Pawlowska et al., 2013; van Maldegem et al., 2019). The rationale stems from generalizing the idea of an expressed fractionation during respiratory decarboxylation (Blair et al., 1985), resulting in the assumption that all heterotrophs would be systematically enriched in 13C relative to autotrophs. However, as previously noted by Breteler et al. (2002), even the data from DeNiro and Epstein (1978) show a range in respired carbon $\delta^{13}C$ values from +1.4‰ to −3.5‰ relative to the food, which is expected as this value will vary depending on the proportion of CO2 derived from the decarboxylation of pyruvate (relatively $^{13}C$-enriched) versus the decarboxylation of Krebs Cycle intermediates (relatively $^{13}C$-depleted) (Hayes, 2001). Nonetheless, the assumption of a 1‰ enrichment in heterotrophic biomass has been used to interpret the $\delta^{13}C$ patterns observed in both modern (e.g., Musilova et al., 2015; Pedrosa-Pàmies et al., 2018) and ancient (e.g., Osterhout et al., 2021; Williford et al., 2013) bulk organic matter, and in various biochemical fractions relative to one another (e.g., Close et al., 2011; Logan et al., 1995; Luo et al., 2015; Pawlowska et al., 2013; van Maldegem et al., 2019). Alternative hypotheses for the $^{13}C$ enrichments observed in microbial systems include diagenetic overprinting (Vinnichenko et al., 2021), a relatively greater assimilation of $^{13}C$-enriched substrates like acetate (Blair et al., 1985; Penning & Conrad, 2006), or increased fractional contributions from alternative metabolisms (House et al., 2003; van der Meer et al., 2001). It therefore remains critical to understand how, or if, a $\delta^{13}C$ signature of microbial heterotrophy would be preserved in the rock record.

To date, a major obstacle in this effort has been measuring taxonomically resolved, natural abundance $\delta^{13}C$ values for microbial communities at sub-1‰ resolution. Prior studies comparing the $\delta^{13}C$ values of phylum-specific biomarker lipids consistently report measurements at sub-1‰ resolution (e.g., Jahnke et al., 2004; Jahnke & des Marais, 2019; van der Meer et al., 2003; Werne et al., 2002). However, using these data to estimate the $\delta^{13}C$ values of autotrophic and heterotrophic microbial biomass is infeasible, given the variable offset in $\delta^{13}C$ values between lipids and biomass (Blair et al., 1985; DeNiro & Epstein, 1977; Tang et al., 2017). Protein Stable Isotope Fingerprinting (P-SIF) is a novel method for measuring the $\delta^{13}C$ values of whole proteins that have been separated from environmental samples and classified taxonomically via proteomics (Mohr et al., 2014). Because $\delta^{13}C$ values of proteins scale directly with biomass $\delta^{13}C$ values (Abelson & Hoering, 1961; Blair et al., 1985), this approach yields taxon-specific or group-specific $\delta^{13}C$ signatures of organisms. Although P-SIF has only modest taxonomic resolving power, it has several advantages that complement other recent stable isotope approaches (e.g., Dekas et al., 2019; Kleiner et al., 2018; Mayali et al., 2012; Radajewski et al., 2000): it does not require addition of isotope labels, incubations, or novel calibrations, and it has a mean analytical precision <1‰.

Here, we use P-SIF to assign $\delta^{13}C$ values for taxonomic groups in a previously characterized Cyanobacteria-Chloroflexi mat from Chocolate Pots Hot Springs (CP; Figure 2.1), Yellowstone National Park, USA (Klatt et al., 2013; Pierson et al., 1999; Pierson & Parenteau, 2000). We also report fatty acid (FA) $\delta^{13}C$ values to relate the protein data to lipid biomarkers that can be preserved over geologic time scales. Our results show that the canonically heterotrophic groups of organisms in the Chocolate Pots (CP) mat are isotopically indistinguishable from both the photoautotrophic Cyanobacteria and the sugar moieties of extracellular polymeric substances (EPS) produced by those Cyanobacteria. By contrast, the filamentous anoxygenic phototrophic

(FAP) Chloroflexi are moderately enriched in $^{13}C$ due to their mixotrophic lifestyle. These results indicate that high-density microbial communities do not conform to "you are what you eat +1‰", which we hypothesize is likely due to EPS-driven feeding and the complexities of small-molecule resource sharing.



**Figure 2.1** Chocolate Pots Hot Springs in Yellowstone National Park (YNP). Blue star indicates sampling location for this study. Photograph on the upper right is a close-up of the Synechococcus-Chloroflexi mat. Photographs by N. Parenteau

**Materials and Methods**

*Sample Collection*

Chocolate Pots Hot Springs (CP) contains four types of phototrophic microbial mats: Synechococcus-Chloroflexi (50-54°C), Pseudanabaena spp. (50-54°C), narrow Oscillatoria spp. (36-45°C), and Oscillatoria cf. princeps (37-47°C) (Pierson et al., 1999; Pierson & Parenteau, 2000). Synechococcus-Chloroflexi mat samples, the focus of this study, were collected in August

2014 under Yellowstone National Park Research Permit number 1549 and immediately placed on dry ice at the field site. A sample of approximately 15 grams was shipped to the laboratory on dry ice and subsequently frozen at −80°C until analysis. The sample was taken from the same location sampled for the YNP Metagenome Project (Inskeep et al., 2013; Klatt et al., 2013).

At collection, the pH of the vent water flowing over the mat was 6.1 and the temperature was 49.7°C. No further geochemical parameters were measured at the time of sampling; however, prior work on vent waters at Chocolate Pots Hot Springs (CP) has shown that the concentrations of various species have changed little over 80 years (see Table 2 from Parenteau & Cady 2010).

*Lipid extraction and identification*

Lipids were extracted from approximately 0.3 g (dry) of freezedried mat samples via a modified Bligh and Dyer procedure (Sturt et al., 2004). The total lipid extract was transesterified to generate fatty acid methyl esters (FAMEs; 5% HCl/methanol [v/v], 70°C, 4 h). The reaction was stopped by the addition of ultrapure $H_2O$, after which the organic phase was extracted into hexane/dichloromethane (4:1, v/v). FAME derivatives of n-C16:0, n-C19:0, and n-C24:0 FA standards with known δ13C compositions (−29.5‰, −31.7‰, and −30.8‰, respectively) were prepared in parallel to correct for the $^{13}C$ content of the derivatized carbon introduced during transesterification. FAMEs were further separated from the derivatized extract by elution over $SiO_2$ gel using the solvent program described in (Pearson et al., 2001).

The FAMEs were identified using gas chromatography-mass spectrometry (GC/MS; Agilent 6890N GC, 5973 MS equipped with a 30 m DB-5MS column) by comparison to known patterns of relative retention times (Pearson et al., 2001; Perry et al., 1979) and by comparison of fragment mass spectra to spectra from the National Institute of Standards and Technology Library

(Shen et al., 2017). The injection, oven temperature programs, and gas flow rates were adopted from Close et al., 2014.

*Protein stable isotope fingerprinting*

Protein Stable Isotope Fingerprinting (P-SIF) was performed as previously described (Mohr et al., 2014). Proteins were extracted from microbial mat samples by placing approximately 7 grams of wet mat material and up to 8 ml of bacterial protein extraction reagent (B-PER) protein extraction reagent (Thermo Scientific) in a 50 ml Teflon tube and sonicating using a 500-watt Qsonica ultrasonic processor equipped with a cup horn. The cup horn was filled with ice water and the sonicator was set to 25 s on and 35 s off for a total of 5 min sonication. Solids and cell material were removed by centrifugation at 16,000g. Proteins were precipitated from the supernatant in acetone and resuspended in 100 mM $NH_4HCO_3$, pH 9 to yield a total soluble protein extract. This extract was further separated into 960 fractions on an Agilent 1100 series HPLC with diode-array detector (DAD) and fraction collector using two orthogonal levels of chromatography: first by strong anion exchange (SAX; Agilent PL-SAX column; $4.6 \times 50$ mm, 8 μm) (20 fractions), then by reverse phase (RP; Agilent Poroshell 300SB-C3 column, $2.1 \times 75$ mm, 5 μm; 48 fractions), using the solvent gradients described in (Mohr et al., 2014). An aliquot of each final fraction is split into 96-well plates for isotope analysis (70%) and the remaining 30% is reserved for tryptic digestion followed by peptide sequencing.

*Protein taxonomic identification*

Plates for tryptic digestion were prepared as detailed in (Mohr et al., 2014). Peptides were sequenced by capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS) using an Agilent 1200 Series HPLC equipped with a Kinetex C18 column ($2.1$ mm $\times$ $100$ mm, $2.6$ μm particles) and an Agilent 6520 quadrupole time-of-flight mass spectrometer (QTOF-MS/MS).

Peptide LC-MS/ MS data were processed by searching all MS2 spectra against an in silico peptide library generated from protein-coding genes in a Chocolate Pots Metagenomic data set (Inskeep et al., 2013; Klatt et al., 2013). Relative phylogenetic abundances in each well were estimated by comparing the mean peptide intensities for proteins taxonomically assigned to a given phylogenetic group (defined in Table S2) to the sum of mean peptide intensity for all proteins in a given well (Mohr et al., 2014). This label-free protein quantification method is only semiquantitative, as are most label-free protein quantification methods (Bubis et al., 2017). Nonetheless, this quantification method was evaluated using a mixture of organisms of known quantity and found to have a ± 20% root-mean-square error (Figure S12 from Mohr et al., 2014).

*Sugar extraction and derivatization*

Extracellular polymeric substances (EPS) were extracted using a modified version of "method 8" as described in (see Table 1 from Klock et al., 2007). Briefly, 20 ml of 10% (w/v) NaCl was added to 10 grams of wet homogenized microbial mat sample and vortexed. This solution was incubated at 40°C for 15 min, followed by centrifugation at 8200g for 15 min. The supernatant was collected, and the precipitant was re-extracted with 20 ml 10% NaCl two more times. After cooling in an ice bath, 100% ethanol was added to the supernatant to a final concentration of 70%. EPS was precipitated at 4°C overnight and removed by centrifugation.

Extracted EPS were hydrolyzed into monomers using established methods (van Dongen et al., 2001). Briefly, the EPS were vortexed with 1 ml 12 M $H_2SO_4$ in a Teflon tube. A stir-bar was added, and the solution was stirred at ~400 rpm for 2 hr at room temperature, followed by dilution to 1 M and heating at 85°C for 4.5 h. After cooling to room temperature, the solution was neutralized to pH 7 using $BaCO_3$. Once neutralized, the solution was centrifuged at 4000 g for 5 min, after which the supernatant was collected, frozen, and lyophilized.

Lyophilized sugar monomers were derivatized immediately prior to isotope analysis, again using established protocols (van Dongen et al., 2001). Arabinose, xylose, glucose, and myo-inositol standards with known $\delta^{13}$C compositions (−11.7‰, −9.7 ‰, −11.1‰, and −14.4‰, respectively) were prepared in parallel to correct for the $^{13}$C content of the carbon introduced during derivitization. Briefly, 1 ml of a methylboronic acid/pyridine (10 mg/ml) mixture was added to 5 mg of lyophilized sample or to 1 mg of total glucose, arabinose, and xylose standards and then heated at 60°C for 30 min, followed by the addition of 100 μl N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and a further 5 min of heating. Because myo-inositol is insoluble in pyridine, 250-500 μg myo-inositol was instead dissolved in 1 ml dimethyl sulfoxide (DMSO) for derivatization. After the heating step, the myo-inositol standard mixture was cooled to room temperature before adding 1 ml of cyclohexane and 100 μl of BSTFA (Leblanc & Ball, 1978). All samples were dried under $N_2$ and quantitatively dissolved in ethyl acetate prior to isotope analysis.

*Isotope ratio mass spectrometry*

To measure the $\delta^{13}$C value of bulk total organic carbon (TOC), approximately 7.5 mg of triplicate freeze-dried microbial mat samples was placed in silver capsules (Costech) and acidified with 100 μl of 1 N HCl to remove dissolved inorganic carbon. Samples were dried at 50°C, enveloped in tin capsules (Costech), and analyzed on a Costech 4010 Elemental Analyzer connected to a Thermo Scientific Delta V IRMS (isotope ratio mass spectrometer).

Fatty acid methyl ester (FAME) and derivatized sugar monomer $\delta^{13}$C compositions were analyzed via gas chromatography-isotope ratio mass spectrometry (GC-IRMS; Thermo Scientific Delta V Advantage connected to a Trace GC Ultra via a GC Isolink interface). In both cases, 1 μl of sample was co-injected with 0.5 μl of internal standard (n-C32; 50 ng/μl). FAMEs were run on

a 30 m × 0.25 mm HP-5MS column as previously described (Close et al., 2014). Sugar monomers were run on a 30 m × 0.25 mm DB-1701 column; samples were transferred onto the column using a programmable temperature vaporizer (PTV) inlet at an injection temperature of 70°C followed by 330°C for 4 min. The GC-IRMS oven temperature gradient was adopted from van Dongen et al. (2001).

Stable carbon isotope analysis was conducted using spoolingwire microcombustion (SWiM)-IRMS (Caimi & Brenna, 1993; Brand & Dobberstein, 1996; Sessions et al., 2005; Thomas et al., 2011). The SWiM-IRMS configuration used here is adapted from Sessions et al. (2005) and is detailed in Mohr et al. (2014). Fractions (96-well plate aliquots) were measured in triplicate. Only data from wells containing >0.56 nmol C/µl (~350 mV peak amplitude, m/z 44) and with measurement standard deviations <2‰ were retained (Mohr et al., 2014).

*Data analysis*

Protein phylogenetic data were grouped into three taxonomic bins: Cyanobacteria + "other," Chloroflexi, and putative heterotrophs (Table S1). Estimates of the $\delta^{13}C$ compositions of proteins for each group ($\delta_{cyano}$, $\delta_{chloro}$, and $\delta_{het}$, respectively) were calculated via three different methods using the measured $\delta^{13}C$ compositions ($\delta_m$; Table S4) and the estimated relative taxonomic abundances in each well.

The first method (>95% Unique) calculates the values of $\delta_{cyano}$, $\delta_{chloro}$, and $\delta_{het}$ from the average $\delta_m$ for only those wells in which >95% of the peptide signal intensity was assigned to just one of the three taxonomic groups. Precision for this method is reported as ±1 SD from the mean. For the second (2-Component Mixture) and third (Linear Regression) methods, the $\delta_m$ values and estimated fractional abundances (from peptide signal intensity) were used in three overdetermined linear equations:

$$\delta_{m,i} = (\textstyle\sum Pep_{cyano,i} \times \delta_{cyano} + \sum Pep_{chloro,i} \times \delta_{chloro}\,)/\sum Pep_{total} \tag{1}$$

$$\delta_{m,i} = (\textstyle\sum Pep_{cyano,i} \times \delta_{cyano} + \sum Pep_{chloro,i} \times \delta_{chloro} + \sum Pep_{het,i} \times \delta_{het}\,)/\sum Pep_{total} \tag{2}$$

$$\delta_{m,i}/\sigma_{m,i} = (\textstyle\sum Pep_{cyano,i} \times \delta_{cyano} + \sum Pep_{chloro,i} \times \delta_{chloro} + \sum Pep_{het,i} \times \delta_{het}\,)\,/\,(\sum Pep_{total} \times \sigma_{m,i}\,) \tag{3}$$

where $\delta_{cyano}$, $\delta_{chloro}$ and $\delta_{het}$ are the unknowns, $\sum Pep$ is the summed QTOF-MS/MS ion counts for peptides, $\sigma_m$ is the precision for $\delta_m$ ($\pm$ 1 SD) and $i$ is each individual plate well for which both $\delta_m$ and peptide signal intensities were measured (Mohr et al., 2014).

The 2-Component Mixture method uses Equation 1 to estimate $\delta_{cyano}$ and $\delta_{chloro}$ using wells in which peptides were detected only for the Cyanobacteria and Chloroflexi taxonomic groups. Endmembers were estimated via a Deming least-squares regression as detailed previously (Mohr et al., 2014), Deming least-squares regression minimizes the variance in both x and y, accounting for errors in both the independent variable (x) and the dependent variable (y) (Deming, 1943). Precision for this method is reported as the square root of the error variance.

The Linear Regression method uses Equations 2 and 3 to estimate $\delta_{cyano}$, $\delta_{chloro}$, and $\delta_{het}$ for the full data set. Equation 2 represents an unweighted estimate. Equation 3 is weighted by the precision of the isotopic measurements for each well (Glover et al., 2011). Equations 2 and 3 were solved inversely for $\delta_{cyano}$, $\delta_{chloro}$, and $\delta_{het}$ by singular value decomposition (SVD) using the built-in Matlab SVD function (Glover et al., 2011). Precision for this method is reported as $\pm$ the square root of the error variance.

Individual amino acids (AAs) within organisms have different $\delta^{13}C$ values (Abelson & Hoering, 1961; Blair et al., 1985; Macko et al., 1987). However, the $\delta^{13}C$ compositions of average proteins are remarkably consistent (Hayes, 2001). To examine whether AA composition could be

responsible for the observed isotopic differences between wells, we used the full AA sequences of our named proteins to determine the relative proportion of individual AAs in each analyzed sample. The distributions of individual AAs were then compared to the $\delta^{13}C$ composition of the wells containing that AA.

**Results**

*Multidimensional protein chromatography*

The 20 initial chromatographic (SAX) fractions of the bulk protein extract had distinct colors (Figure 2.2, Figure S1) reflecting the presence of chlorophyll a, phycobilin, and bacteriochlorophyll (c and a) pigments associated with photosynthetic proteins (Pierson & Parenteau, 2000). Further separation by RP-HPLC resulted in distinct chromatograms for each SAX fraction (Figure 2.2), consistent with the theoretical average resolving power of ca. $10^3$ proteins for P-SIF chromatography (Mohr et al., 2014). SAX fractions 4-14 were chosen for further analysis using the integrated spectral absorbance of the RP-HPLC signal at 280 nm and criteria from prior work (Mohr et al., 2014).

*Protein taxonomic identifications*

Fifty percent (265/528) of the RP-HPLC fractions contained classifiable peptide sequences, yielding 277 unique proteins (Table S1): 170 proteins assigned to Cyanobacteria, 65 to Chloroflexi, 7 to Chlorobi, and all others (35 proteins) assigned to microbial groups containing four or fewer unique protein hits or to unclassified sources (Table S2). The mean and median number of unique peptides used to classify each protein were 4 and 3, respectively. Using the summed mean intensity of peptides assigned to proteins, we estimate the relative abundance of microbial groups to be 78.6% Cyanobacteria, 12.0% Chloroflexi and Chlorobi (11.6% Chloroflexi, 0.4% Chlorobi), and 9.4% other, which includes 5.1% putatively heterotrophic taxa and 4.3%

unclassified (Table S2). The ordering of these data differs from the phylogenetic distribution of best Basic Local Alignment Search Tool (BLAST) hits for protein-coding genes in the CP metagenome, of which 49.7% were assigned to Chloroflexi, 28.1% to Cyanobacteria, 22.0% to heterotrophic organisms, and 0.2% to Chlorobi (Inskeep et al., 2013; Klatt et al., 2013). In contrast, the distribution of nearly full-length 16S ribosomal RNA (rRNA) gene sequences in the CP metagenome (42% assigned to Cyanobacteria, 28% to Chloroflexi, 7% to Chlorobi, 13% to heterotrophic organisms, and 10% unclassified), and the relative abundances of single-copy marker genes from the metagenome assigned using AMPHORA (automated pipeline for phylogenomic analysis) both agree with the ordering of our relative abundance estimates (Inskeep et al., 2013; Klatt et al., 2013). The variability in relative abundances as estimated by these data is unsurprising, given that label-free protein quantification methods are only semiquantitative (Bubis et al., 2017) and copy numbers of 16 s rRNA genes in particular are variable among bacterial phyla (Větrovský & Baldrian, 2013). More generally, the samples were taken seven years apart in two different months (August vs. October). At CP, the measured pH, concentrations of inorganic constituents, and ventwater temperatures have been remarkably consistent over 80 years (see Table 2 from Parenteau & Cady 2010). (Ferris & Ward, 1997) report a relatively little change in microbial community composition at Octopus Springs in YNP over an annual cycle; nonetheless, the difference in solar flux at this latitude between August and October may contribute to temporal shifts in microbial community composition (Ruff-Roberts et al., 1994). Additionally, our data represent the relative contributions by different microbial groups to total protein biomass, which may differ from the relative abundances of individual organisms. For example, Finkel et al. (2016) estimate a median 43.1% dry weight protein content in Cyanobacteria relative to a 27.4% dry weight protein content in diatoms. Mohr et al. (2014) calculated a ±20% root-mean-square error

for P-SIF abundance estimates by analysis of known mixtures of cultured organisms (Figure S12 from Mohr et al., 2014). Since the difference in relative abundance between Cyanobacteria and all other groups is more than 3 times this ±20% root-mean-square error, we are confident that our method effectively distinguishes between the dominant autotrophic organisms and heterotrophic organisms in our sample. The relative abundance of Chloroflexi (Table S2) is supported by prior reports that the former are the predominant anoxygenic phototrophs in circumneutral to alkaline hot springs in YNP (Bennett et al., 2020; Hamilton et al., 2019).

*Protein carbon isotopic compositions*

Forty-two percent (224/528) of the RP-HPLC fractions contained enough carbon for isotopic measurement. Of these, 83% (186/224) had standard deviations <2.0‰ and were used in subsequent analyses. The resulting protein fraction $\delta^{13}C$ values (Figure 2.3a; Table S4) were not normally distributed (Shapiro-Wilk test, $p < 0.01$) with a mean of −25.4‰ (Table 2.1, Figure 2.3b). The data show a moderately positive skew (skewness = 1.0), indicating a small but statistically significant contribution of isotopically more positive proteins.

The average standard deviation of triplicate $\delta^{13}C$ measurements for protein fractions was 0.6‰ for the whole data set, and 0.4‰ for the most abundant 50% as determined by the IRMS peak area (Table S3). These values represent average measurement errors lower than the population standard deviation (1.0‰), indicating some degree of true variability among the protein $\delta^{13}C$ values (Table S3).

There were no statistically significant ($p = <0.05$, Student's t-test) correlations between the relative distribution of individual AAs and the $\delta^{13}C$ values in each well (Table S6).

*Estimates of protein δ13C values for microbial groups*

Four different methods (i) unique proteins, (ii) 2-component mixing, (iii) abundance-weighted multiple linear regression, and (iv) unweighted multiple linear regression were used to estimate the $\delta^{13}C$ values of proteins originating from different microbial groups (Table 2.2, Figure 2.4). The results from all three approaches agree within 1‰ and indicate that Cyanobacteria and putatively heterotrophic organisms are isotopically indistinguishable, with mean estimates ranging from −25.6‰ to −25.3‰. In contrast, the Chloroflexi are relatively $^{13}C$-enriched, at −24.0‰ to −22.7‰, depending on the estimation approach. Other taxonomic groups yielded too few assigned data points to resolve by mass-balance mixing approaches. However, in six wells, >45% of the detected protein was classified as deriving from Actinobacteria, including one well apparently composed entirely of actinobacterial protein and having a $\delta^{13}C$ value of −25.1‰. This value is indistinguishable within error from both the cyanobacterial and the heterotrophic mean values and is consistent with a relatively high abundance of Actinobacteria in the heterotrophic population.

We can compare these estimates for taxon-specific protein $\delta^{13}C$ compositions to biomarker $\delta^{13}C$ values reported by Parenteau (2007). Data from a CP sample collected from the same location as our 2014 sample in July 2005 include three wax ester compounds (average $\delta^{13}C$ composition of −30.2‰ ± 0.8‰), which are biomarkers for Chloroflexi and are typically approximately 2‰ depleted in $^{13}C$ relative to biomass (van der Meer et al., 2001). This suggests Chloroflexi biomass from the 2005 CP sample had a $\delta^{13}C$ composition of approximately −28.2‰. The 2005 CP sample also includes the cyanobacterial biomarkers *n*-heptadecane and phytol ($\delta^{13}C$ compositions of −39.4‰ ± 3.1‰ and −35.2‰ ± 0.7‰, respectively), which are typically approximately 8‰ and 6‰ depleted in $^{13}C$ relative to biomass, respectively (Sakata et al., 1997). These data suggest Cyanobacterial biomass from the 2005 CP sample had a $\delta^{13}C$ composition of approximately −31.4‰ ± 3.1‰ or −29.2‰ ± 0.7‰. These values should only be taken as back-of-the-envelope

estimates. Nonetheless, similarly to our protein $\delta^{13}C$ estimates, Chloroflexi biomass $\delta13C$ composition as estimated by these particular biomarker compounds is approximately 1‰-3‰ heavier than cyanobacterial biomass, in agreement with our protein data from the 2014 CP sample.

*Bulk, fatty acid, and sugar carbon isotope ratios*

Total organic carbon in this sample (TOC; $-27.0 \pm 0.1$‰; Table 2.1) was $^{13}C$-depleted relative to previous measurements of bulk biomass carbon reported previously for Chocolate Pots Synechococcus-Chloroflexi mats ($-23.2 \pm 0.8$‰ in 2004 and $-25.8 \pm 0.5$‰ in 2005; (Parenteau, 2007). This TOC value is 1.6‰ depleted in $^{13}C$ relative to average bulk protein, while the weighted average fatty acids (FAs) were 6.6‰ depleted in $^{13}C$ relative to TOC (Figure 2.5). The $\delta^{13}C$ of dissolved inorganic carbon (DIC) in the vent water above the mat was not measured at the time of sampling. However, given prior measurements in the same location ($-2.0$‰ $\pm$ 0.1‰ in 2004 and $-1.0$‰ $\pm$ 0.1‰ in 2005; Parenteau, 2007) and the narrow range in pH, bicarbonate concentrations, and temperatures at CP over 80 years (see Table 2 from Parenteau & Cady 2010), the TOC in our sample is approximately 25‰ to 26‰ depleted in 13C relative to the assumed DIC. This fractionation is within the expected range for organisms using the Calvin-Benson-Bassham Cycle (pentose phosphate cycle) for autotrophic carbon fixation in the 50-60°C range (Havig et al., 2011), which is consistent with the relatively high abundance of Cyanobacteria.

Individual FAs $n$-$C_{16:0}$, $n$-$C_{18:0}$, $n$-$C_{18:1}$, and $n$-$C_{18:2}$ were the only quantitatively significant FAs recovered (Table 2.3), in agreement with previous reports indicating these compounds comprise 93% of FAs in Synechococcus-Chloroflexi CP mats (Parenteau et al., 2014). The $\delta^{13}C$ values of these FAs ranged from 5.7 to 9.0‰ depleted in 13C relative to TOC, equivalent to 7.3 to 10.6‰ depleted relative to average protein (Figure 2.5). This relative isotopic ordering also agrees with prior reports, where $n$-$C_{16:0}$, $n$-$C_{18:0}$, and $n$-$C_{18:1}$ FAs (average of phospholipid, neutral,

and polar glycolipid FAs) had average $\delta^{13}C$ values that were 10.8‰, 12.1‰, and 12.8‰ offset from bulk biomass, respectively (Parenteau, 2007); their larger reported offsets are consistent with the more $^{13}C$-enriched value reported for TOC in that earlier sample.

Glucose was the only quantitatively important sugar monomer recovered from extracted EPS. It had a $\delta^{13}C$ value of $-25.1‰ \pm 0.8‰$ (Table 2.1), which is within error of the cyanobacterial and heterotrophic proteins as estimated by all methods (Table 2.2).

**Table 2.1** Summary of δ13C values for Chocolate Pots (CP) microbial mats

| Fraction | $\delta^{13}C$ (‰) |
|---|---|
| Total organic carbon | -27.0 ± 0.1 |
| Weighted average FA | -33.6 ± 0.3 |
| Average protein | -25.4 ± 1.0 |
| Glucose (EPS) | -25.1 ± 0.8 |

**Table 2.2** Estimates of the δ13C values (‰) of proteins for microbial groups from the Chocolate Pots mat as calculated from P-SIF data using four different methods.

| Microbial Group | >95% Unique | 2-Component mixing | Weighted linear regression | Unweighted linear regression |
|---|---|---|---|---|
| Cyanobacteria + Other | -25.3 ± 0.7 | -25.6 ± 0.6 | -25.4 ± 0.2 | -25.5 ± 0.3 |
| Chloroflexi | -23.5 ± 1.8 | -22.7 ± 0.6 | -23.1 ± 0.5 | -24.0 ± 0.5 |
| Heterotrophs | -25.3 ± 0.7 | | | |

**Table 2.3** $\delta^{13}C$ values for individual FAs from Chocolate Pots microbial mats.

| FA | $\delta^{13}C$ (‰)[a] | Relative abundance |
|---|---|---|
| $n$-C$_{16:0}$ | -32.7 ± 0.2 | 1 |
| $n$-C$_{18:0}$ | -35.1 ± 1.1 | 0.11 |
| $n$-C$_{18:1}$ | -35.4 ± 0.4 | 0.24 |
| $n$-C$_{18:2}$ | -35.8 ± 0.1 | 0.10 |

[a]Values are averaged from triplicate (n-C16:0) and duplicate (n-C18:x) GC-IRMS runs.
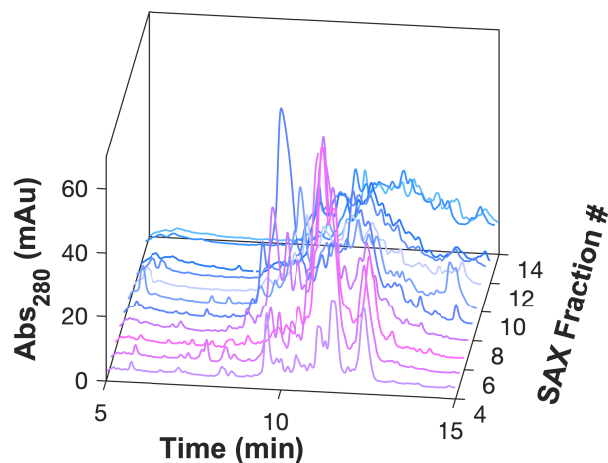
**Figure 2.2** RP-HPLC chromatograms of the spectral absorbance at 280 nm for individual SAX fractions S4-S14. Chromatogram color represents the visible color of each fraction (legend, top right; also see Figure S1).
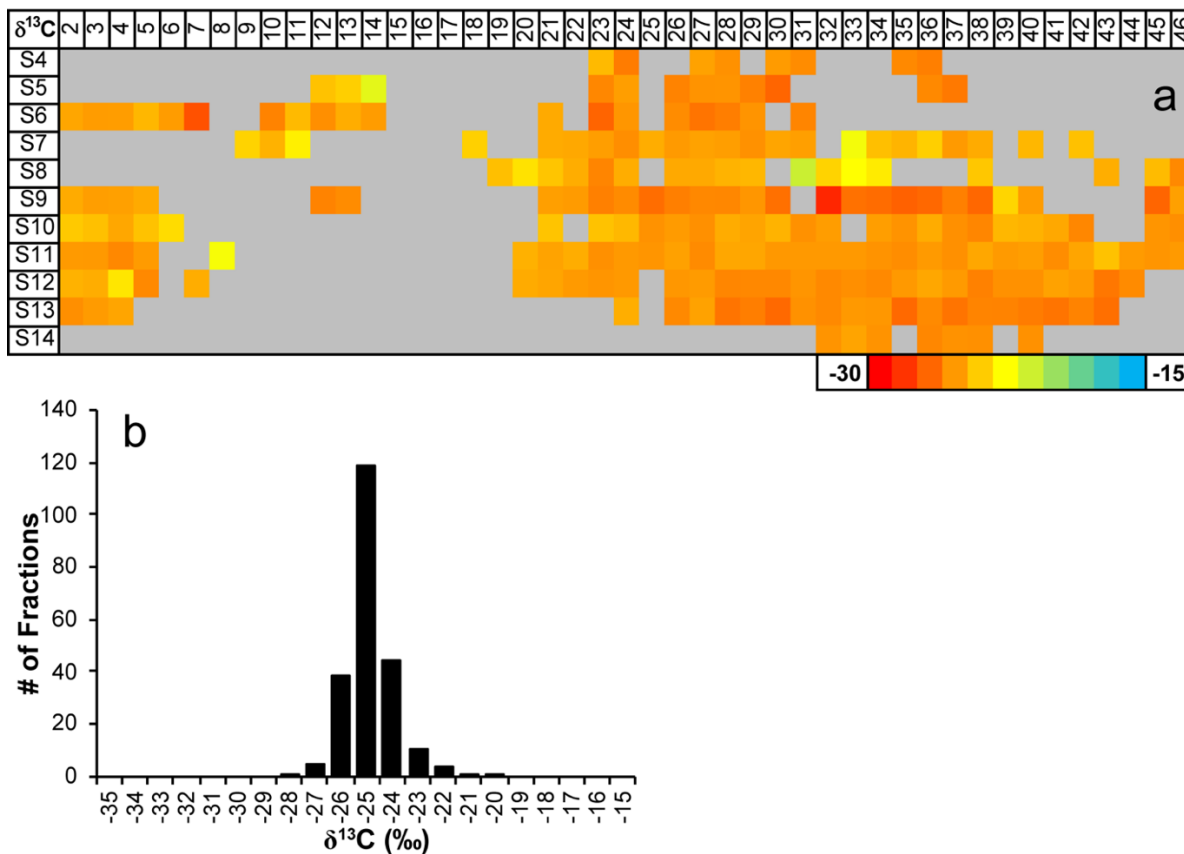


**Figure 2.3** (a) $\delta^{13}C$ values of individual RP chromatographic time slices (2-46) of SAX fractions S4-S14. Grey areas indicate no usable data (< 0.56 nmol C/μL and/or triplicate SD > 2‰). (b) Histogram of $\delta^{13}C$ values for SAX fractions shown in 3A. Values are normally distributed (Shapiro-Wilk test, $p < 0.01$).
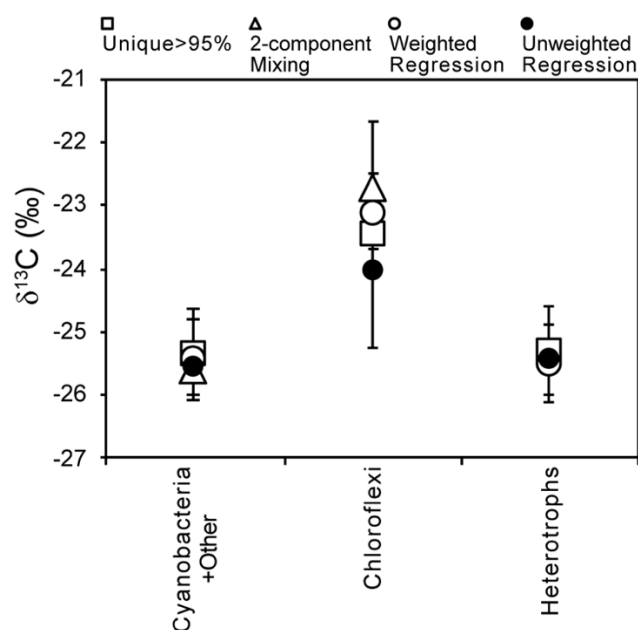
**Figure 2.4** Estimates of the $\delta^{13}C$ values of proteins from taxonomic groups as calculated by four independent methods: *i*) wells in which > 95% of detected proteins belonged only to the indicated group (squares); *ii*) 2-component end-member solution for wells in which > 95% of detected proteins belonged only to Cyanobacteria or Chloroflexi (triangles); *iii*) estimated from all taxonomic abundance and $\delta^{13}C$ data using an abundance-weighted multiple linear regression routine (unfilled circles); *iv*) estimated from all taxonomic abundance and $\delta^{13}C$ data using an unweighted multiple linear regression routine (filled circles); See Methods for details.
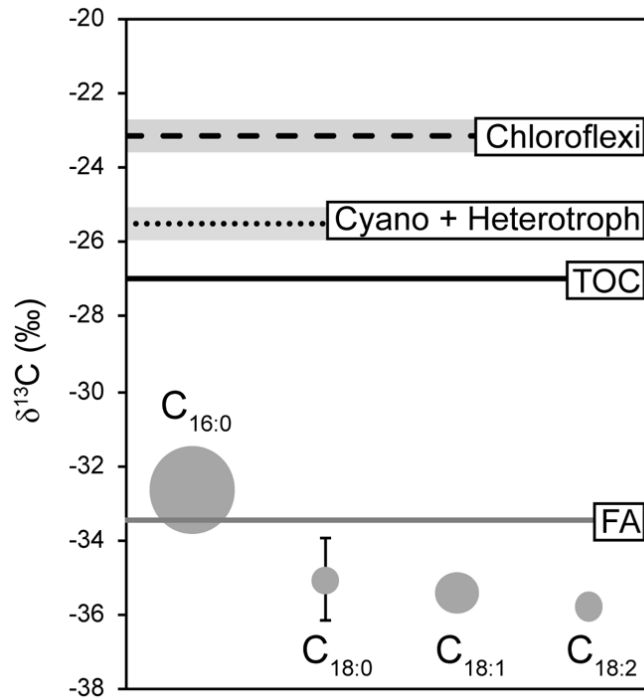
**Figure 2.5** A composite of Chocolate Pots carbon isotopic data. Chloroflexi, Cyanobacteria and heterotrophic protein $\delta^{13}C$ values are represented by dashed and dotted lines, respectively. The $\delta^{13}C$ value of TOC is shown in black. The weighted average FA pool is shown in gray. The $\delta^{13}C$ values of individual FA are indicated by gray circles; circle area corresponds to abundance relative to the $n$-$C_{16:0}$ FA. Shading and error bars represent $\pm$ 1 SD from the mean.

**Discussion**

*Cyanobacterial carbon production and excretion*

In phototrophic microbial mats, cyanobacterial organic carbon is available to heterotrophic organisms as excreted carbon storage molecules, photorespiration by-products, fermentation products, or via viral lysis (Bateson & Ward, 1988; Carreira et al., 2015; Nold & Ward, 1996; Stal & Moezelaar, 1997). Under the high light intensity typical of the top layer of microbial mats, the supply rate of photons is likely to be less limiting than the supply rate of nutrients. Under these conditions, cyanobacteria allocate excess photosynthate to both stored and excreted molecules, primarily polysaccharides (Braakman et al., 2017; Fogg, 1983). This allows cyanobacteria to maintain a high carbon fixation rate (Fogg, 1983), manage adenosine triphosphate (ATP) levels

(Cano et al., 2018), and store reduced carbon as an energy reserve for when light levels are low (Nold & Ward, 1996; Stal & Moezelaar, 1997). Cyanobacteria in a similar (48°C to 65°C) Synechococcus–Chloroflexi mat from Octopus Springs, YNP, allocated up to 85% of fixed CO2 to polysaccharide (Nold & Ward, 1996). This ratio is in great excess of the typical molecular composition of cells (> 30% protein for photosynthetic organisms; (Finkel et al., 2016), implying the balance of this sugar production was excreted extracellularly. Cyanobacteria-derived polysaccharide forms the majority of EPS in microbial mats, and this material, after digestion by extracellular enzymes, is assimilated by heterotrophic organisms or reused by Cyanobacteria (Bateson & Ward, 1988; Flemming & Wingender, 2010; Klock et al., 2007; Stuart et al., 2016). In mats similar to CP, glucose-rich EPS is the most likely primary source of organic carbon for heterotrophs.

Previous studies on carbon transfer in YNP microbial mats have focused primarily on the uptake of fermentation products by FAPs (e.g., Bateson & Ward, 1988; Nold & Ward, 1996; van der Meer et al., 2003). To our knowledge, there is no work directly demonstrating the preferential uptake of EPS-derived carbohydrates (as opposed to their fermentation products) by heterotrophs in YNP subaerial mats. Stuart et al. (2016) characterized the exoproteome of both natural and cultured marine cyanobacterial mats and found that the majority of proteins in EPS were related to carbohydrate and amino-acid metabolism. Furthermore, the two most abundant heterotrophic phyla in our CP mat (Actinobacteria and Bacteroidetes) are known to assimilate and ferment glucose in culture (de Vos et al., 2009; Whitman et al., 2010).

We thus hypothesize that the isotopic similarity between Cyanobacteria and heterotrophs is due to the latter organisms directly consuming glucose-rich EPS excreted by the former. In support of this idea, the glucose moieties of EPS extracted from our CP sample are isotopically

indistinguishable from the protein $\delta^{13}C$ values of both the Cyanobacteria and heterotrophs in the community. To our knowledge, there is only one prior report on the $\delta^{13}C$ composition of extracellular EPS in a microbial mat (rather than cellassociated sugar monomers), and in it, EPS again are isotopically equal to bulk organic carbon (Wieland et al., 2008).

Initially, both our findings here and the work of Wieland et al. (2008) appear to contradict prior work on the carbon isotopic composition of cell-associated glucose monomers (Teece & Fogel, 2007; van der Meer et al., 2003; van Dongen et al., 2002). In these reports, glucose monomers extracted from cells were enriched in $^{13}C$ relative to bulk biomass. However, this pattern excludes glucose from a cyanobacterial culture, which is depleted in $^{13}C$ relative to cyanobacterial biomass (Teece & Fogel, 2007). In either case, we can resolve this discrepancy by differentiating between internal (cell-associated) and external (EPS) sugars. Pereira et al. (2009) note that differences in the $\delta^{13}C$ compositions of internal hexose monomers in Cyanobacteria are likely due to isotopic fractionation during the polymerization of internal sugars to polysaccharide. Enzymatic products are typically depleted in $^{13}C$ relative to reactants. As such, EPS are likely depleted in $^{13}C$ relative to the internal pool of free sugars from which they are polymerized. Furthermore, because subaerial mat-forming Cyanobacteria allocate the majority of their fixed carbon to polysaccharides, we assume the expressed fractionation is minimal and the polysaccharide fraction (the dominant product) should isotopically resemble the initial photosynthate (Hayes, 2001).

In summary, cyanobacterial and heterotrophic proteins are isotopically indistinguishable because both organisms are utilizing simple sugars derived from cyanobacterial photosynthate as their anabolic carbon source. At the broader level, a hypothesis of near-zero $^{13}C$ fractionation by organisms growing on pure sugar substrates is consistent with some of the earliest works of stable

isotope biogeochemistry: Chlorella pyrenoidosa biomass was isotopically indistinguishable from the glucose substrate (Abelson & Hoering, 1961), while Escherichia coli biomass was depleted in $^{13}C$ relative to glucose substrate (Abelson & Hoering, 1961; Blair et al., 1985). The $^{13}C$-depletion in the latter organism may be understood as a reflection of total biomass (i.e., including its $^{13}C$-depleted lipid component), while lack of a significantly $^{13}C$-depleted bulk signal in Chlorella may reflect a smaller fractional contribution of lipids to the bulk cell.

*Chloroflexi carbon sources*

Chloroflexi can grow photoautotrophically (Chloroflexus spp.) via the anoxygenic 3-hydroxypropionate (3-HP) pathway, photoheterotrophically under light anoxic conditions (all FAPs) via the assimilation of low-molecular-weight organic compounds (Bauld & Brock, 1973; Giovannoni et al., 1987; Hanada et al., 2002; van der Meer et al., 2010; Zarzycki & Fuchs, 2011), or photomixotrophically by simultaneously incorporating inorganic and organic carbon sources (Chloroflexus, Roseiflexus) (Klatt et al., 2013). Chloroflexi can also grow heterotrophically under dark aerobic conditions. When grown purely photoautotrophically, the biomass of Chloroflexus aurantiacus is ~13‰ depleted in $^{13}C$ relative to the carbon source (van der Meer et al., 2001, House et al., 2003). Using our estimate for Chloroflexi biomass (−23.1 ± 0.5‰), a δ13C value of −1.0‰ ± 0.1‰ for CP dissolved inorganic carbon (Parenteau, 2007), an assumed photoautotrophic Chloroflexi biomass maximum value of −14.0‰ ± 0.1‰ (van der Meer et al., 2001), and our estimated heterotrophic endmember (−25.4‰ ± 0.2‰), isotope mass balance implies the FAPs obtain 20% ± 4% of their carbon autotrophically and the remaining ~80% by heterotrophic assimilation (if the latter substrate is sugar, or isotopically identical to sugar).

A recent report of in situ rates of anoxygenic photosynthesis across a broad range of mat systems in YNP mentions that anoxygenic photoautotrophy was not detected in sites below pH 6;

however, the same authors report the presence and transcripts of bchY genes (which are considered markers for anoxygenic photosynthesis, see Hamilton et al., 2012) at these sites, supporting active photoheterotrophy (Hamilton et al., 2019). While our CP sample was collected at pH 6.1, historical measurements have measured a range of pH 5.7–6.1, placing CP directly on this threshold and supporting primarily (but not necessarily exclusively) photoheterotrophic growth by Chloroflexi at CP (Parenteau & Cady 2010). In the present work, we do not distinguish between individual taxa of Chloroflexi. Bennett et al. (2020) report bulk biomass $\delta^{13}C$ values similar to autotrophic C. aurantiacus, and they find more abundant Chloroflexus operational taxonomic units (OTUs) and relatively heavier bulk $\delta^{13}C$ values at alkaline sites below 59.3°C than sites above 59.3°C. In contrast, Roseiflexus OTUs were most abundant between 58.3°C and 71.8°C (Bennett et al., 2020). It is thus possible that our sample represents a relatively larger contribution of Roseiflexus, albeit from a relatively acidic (pH 6.1) and cooler (49.7°C) site compared to those from Bennett et al. (2020).

Both laboratory culture and in situ studies support the idea that phototrophic Chloroflexi favor predominantly photoheterorophic growth. In laboratory culture, the fastest growth rates for C. aurantiacus and two strains of Roseiflexus are observed in media supplemented with organic carbon substrates (Bauld & Brock, 1973; Giovannoni et al., 1987; Hanada et al., 2002; van der Meer et al., 2010; Zarzycki & Fuchs, 2011). Attempts to grow Roseiflexus photoautotrophically in culture have been unsuccessful (Hanada et al., 2002; van der Meer et al., 2010), although it does contain the genes for the 3-HP pathway (Klatt et al., 2007), and has been shown to grow photoautotrophically and/or photomixotrophically in situ (Klatt et al., 2013). Chloroflexus can be cultured both autotrophically and heterotrophically (when photoautotrophy is inhibited) via the glyoxylate cycle, but these modes have slower growth rates (Giovannoni et al., 1987; Pierson &

Castenholz, 1974; Zarzycki & Fuchs, 2011). The pervasiveness of photoheterotrophic growth hypothesized by these culture studies is corroborated in situ via isotope-labeling experiments that show incorporation of labeled organic compounds by Chloroflexi in microbial mat environments; in particular, multiple studies point to the incorporation of fermentation products (e.g., acetate) in microbial mats (Anderson et al., 1987; Nold & Ward, 1996; van der Meer et al., 2005).

Therefore, an alternate explanation is that the phototrophic Chloroflexi in the CP mat may be growing nearly completely photoheterotrophically and are accessing a distinct carbon pool (with a different $\delta^{13}C$ value) than the other CP heterotrophs. While measurements of the $\delta^{13}C$ composition of bacterial fermentation products are scarce, acetate produced via fermentation by E. coli and Clostridium papyrosolvens is $^{13}C$-enriched relative to source glucose (Blair et al., 1985; Penning & Conrad, 2006). Given the metabolic flexibility of the phototrophic Chloroflexi, the CP community may grow photoheterotrophically on acetate and/or partially photoautotrophically, in addition to a generally high background rate of sugar-driven heterotrophy.

*Comparison to lipid $\delta^{13}C$*

Lipids and pigments are organic compounds that have a high preservation fidelity in the sedimentary record (Luo et al., 2019). Contextualizing $\delta^{13}C$ values of ancient lipid biomarkers requires estimating the difference in $\delta^{13}C$ values between lipids and biomass ($\varepsilon_{bio-lipid}$). Estimating $\varepsilon_{bio-lipid}$ accurately requires knowing the proportion of fixed carbon allocated to lipids, which can change depending on nutrient and light availability (Hayes, 2001; Mouginot et al., 2015; Tibocha-Bonilla et al., 2020). Furthermore, $\varepsilon_{bio-lipid}$ differs in heterotrophic organisms grown on glucose versus acetate (Blair et al., 1985; DeNiro & Epstein, 1977; Tang et al., 2017). Our phylum-specific protein $\delta^{13}C$ values allow calculation of the offset between FAs and specific microbial groups in our CP mats.

The 8.5–9.3‰ offset between TOC and the 18-carbon FAs is consistent with a dominantly cyanobacterial source, as the maximum $\varepsilon_{\text{bio-lipid}}$ reported (~8–11‰) is unique to Cyanobacteria (Parenteau et al., 2014; Sakata et al., 1997). In particular, the 11‰ offset between the $n$-$C_{18:2}$ FA and cyanobacterial protein is consistent with prior reports that this FA is primarily produced by Cyanobacteria (Kenyon et al., 1972; Parenteau et al., 2014).

The relatively large $\varepsilon_{\text{bio-lipid}}$ estimated for Cyanobacteria in CP is consistent with our hypothesis that Cyanobacteria are allocating the majority of fixed carbon to storage sugars. When only a small proportion of the initial C3 monomers of photosynthesis are decarboxylated to acetate, this enables greater expression of the isotope effect of pyruvate dehydrogenase (DeNiro & Epstein, 1977). Accordingly, $\varepsilon_{\text{bio-lipid}}$ increases both due to this greater expression of the isotopic fractionation for lipid synthesis and because the bulk cell is composed of less lipid overall (Hayes, 2001; Sakata et al., 1997).

*Implications for the geologic record*

Microbial mat environments were likely widespread during the Proterozoic and early Paleozoic (Callow & Brasier, 2009; Gehling, 1999; Hagadorn & Bottjer, 1997; Steiner & Reiter, 2001; Tarhan, 2018). As such, interpreting the $\delta^{13}C$ values of well-preserved organic matter from Proterozoic sediments requires understanding the isotopic consequences of carbon transfer within microbial mats specifically. Unlike water-column heterotrophs feeding on "marine snow," which is a diverse assemblage of organic compounds of various classes (Alldredge & Silver, 1988), our data indicate that heterotrophs in subaerial or shallow, highly photic mats are consuming photosynthetic sugars. This ecosystem represents both production and consumption dominated by aerobic metabolisms (oxygenic photosynthesis and aerobic respiration). We therefore would predict this pattern to be widespread in post-GOE (Great Oxidation Event) surface environments,

in which oxygenic Cyanobacteria came to dominate under conditions of high light intensity and/or low nutrient availability (Crockford et al., 2018; Havig et al., 2017; Reinhard et al., 2017). Any such system may allocate the majority of initial photosynthate to storage sugars (Fogg, 1983). If such mat environments were common during the Proterozoic and early Paleozoic, "you are what you eat", rather than "you are what you eat, +1‰" is more likely to apply to those environments. The corollary to such an assertion is that systematic $^{13}$C enrichment in bulk TOC, either on its own or relative to various biochemical fractions (e.g., Close et al., 2011; Logan et al., 1995; Pawlowska et al., 2013), would instead represent either a diagenetic signal (Cheng et al., 2015; Tang et al., 2005; Vinnichenko et al., 2021), a relatively greater assimilation of $^{13}$Cenriched substrates like acetate (Blair et al., 1985; Penning & Conrad, 2006), or a complex mixture of alternative metabolisms (Havig et al., 2017 and references therein) (including overprinting by other anaerobic degradation pathways in addition to acetate fermentation).

Contrary to biochemical expectations, straight-chain lipids extracted from Proterozoic sediments are consistently $^{13}$C-enriched relative to kerogen (Hayes, 2001; Logan et al., 1995). Initially, this inversion was attributed to the selective preservation of lipids from benthic heterotrophs assimilating organic carbon which, due to the slower sinking rate of Proterozoic organic matter, was subject to multiple rounds of remineralization in the water column (Logan et al., 1995). This hypothesis assumes that heterotrophic biomass becomes isotopically more positive per trophic level. It further requires that water-column remineralization is intense enough to enrich lipids in $^{13}$C while both (1) preserving sufficient heterotrophic lipid to form the majority of the preserved lipid pool and (2) degrading sufficient heterotrophic biomass such that primary biomass forms the majority of the kerogen pool (Close et al., 2011). While such conditions are predicted to occur only rarely, if ever, within the water column (Close et al., 2011), some have suggested that

these conditions are possible within microbial mats (Jahnke & des Marais, 2019; Pawlowska et al., 2013).

The absence of any distinguishable $\delta^{13}C$ signatures associated with heterotrophy in the subaerial CP *Synechococcus*–Chloroflexi mat suggests that the prevalence of microbial mat environments during the Proterozoic is not the sole explanation for the isotopic inversion between *n*-alkyl lipids and kerogen observed in rocks of Proterozoic age. Recent work favors a diagenetic source for the inversion. Samples from the Paleoproterozoic Barney Creek Formation have a relatively constant $\delta^{13}C$ composition of kerogen, while the $\delta^{13}C$ values of *n*-alkanes increase by 6.8‰ on average and are correlated with increasing thermal maturity (Vinnichenko et al., 2021). These data are supported by experiments on crude oil and Paleogene source rocks where the $\delta^{13}C$ compositions of nalkanes increased by approximately 3–6‰ after undergoing artificial maturation (Cheng et al., 2015; Tang et al., 2005; Tian et al., 2017), as well as by Phanerozoic field-based studies that show equivalent increases of approximately 2–4‰ (Clayton & Bjorøy 1994; Dawson et al., 2007; Odden et al., 2002; Cheng et al., 2015). Low-molecular weight (< C19) compounds can be enriched in $^{13}C$ by up to 4‰ relative to crude oil after biodegradation (Pedentchouk & Zhou, 2020; Sun et al., 2005). However, Vinnichenko et al. (2021) dismiss the effects of biodegradation on the $\delta^{13}C$ compositions of *n*-alkanes in their samples due to the constant relative abundances of *n*-alkanes in the samples. Furthermore, diagenesis on its own does not explain the relative lack of inverted isotope signal throughout the Phanerozoic (Logan et al., 1995). While this may partly reflect the greater availability of thermally immature sediments in the Phanerozoic combined with a researcher bias toward these sediments, it does not explain reemergence of unusual $^{13}C$ lipid ordering at the Permian– Triassic boundary coincident with extinction-associated ecosystem changes (Grice et al., 2005).

The only isotopically distinct group detected in the CP mat was the Chloroflexi, which can be attributed to either a bicarbonateutilizing autotrophic metabolism (3-HP), the assimilation of lowmolecular-weight compounds (e.g., acetate), or a combination of both signals. Molecular clock estimates suggest the 3-HP pathway did not evolve until the end of the Proterozoic (Shih et al., 2017). Nonetheless, these results suggest that the $\delta^{13}C$ composition of microbial biomass is more closely tied to specific metabolites than to autotrophy versus heterotrophy. As such, interpretations of the $\delta^{13}C$ values in sediments derived from predominantly microbial ecosystems should be developed relative to the $\delta^{13}C$ values of specific molecular-level carbon sources.

**Conclusions**

Here, the results for an oxygenic, photosynthetic microbial mat indicate that the protein subfractions of Cyanobacteria and obligate heterotrophs such as Actinobacteria have indistinguishable $\delta^{13}C$ signatures. Such ecosystems – specifically shallow or subaerial microbial mat environments – are dominated by oxygenic photosynthesis and aerobic, heterotrophic respiration. This suggests that sugar production and consumption dominate the net ecosystem metabolism, yielding a community in which producers and consumers share primary photosynthate as a common resource. Proteins assigned to the FAP bacteria from the phylum Chloroflexi (Roseiflexus sp. and Chloroflexus sp.) were approximately 2‰ enriched relative to both Cyanobacteria and other heterotrophs, indicating that they are growing partially photoautotrophically, or are consuming a carbon substrate (e.g., acetate) with a distinct isotopic composition. Our results caution against applying "you are what you eat, +1‰" to microbial community food webs, especially when interpreting $\delta^{13}C$ values of ancient sediments derived from predominantly microbial ecosystems such as microbial mats.

## Acknowledgements

## References

Abelson P. H. and Hoering T. C. (1961) Carbon isotope fractionation in formation of amino acids by photosynthetic organisms. *Proc Natl Acad Sci U S A* **47**, 623–632.

Alldredge A. L. and Silver M. W. (1988) Characteristics, dynamics and significance of marine snow. *Progress in Oceanography*.

Anderson K. L., Tayne T. a and Ward D. M. (1987) Formation and fate of fermentation products in hot spring cyanobacterial mats. *Appl Environ Microbiol* **53**, 2343–2352.

Arnosti C., Steen A. D., Ziervogel K., Ghobrial S. and Jeffrey W. H. (2011) Latitudinal gradients in degradation of marine dissolved organic carbon. *PLoS ONE* **6**, 8–13.

Awramik S. M. (1992) The oldest records of photosynthesis. *Photosynthesis Research* **33**, 75–89.

Bateson M. M. and Ward D. M. (1988) Photoexcretion and fate of glycolate in a hot spring cyanobacterial mat. *Appl Environ Microbiol* **54**, 1738–43.

Bauld J. and Brock T. D. (1973) Ecological studies of Chroflexis, a gliding photosynthetic bacterium. *Archiv für Mikrobiologie*.

Bennett A. C., Murugapiran S. K. and Hamilton T. L. (2020) Temperature impacts community structure and function of phototrophic Chloroflexi and Cyanobacteria in two alkaline hot springs in Yellowstone National Park. *Environmental Microbiology Reports* **12**, 503–513.

Blair N., Leu a, Muñoz E., Olsen J., Kwong E. and des Marais D. (1985a) Carbon isotopic fractionation in heterotrophic microbial metabolism. *Appl Environ Microbiol* **50**, 996–1001.

Braakman R., Follows M. J. and Chisholm S. W. (2017) Metabolic evolution and the self-organization of ecosystems. *Proceedings of the National Academy of Sciences* **114**, E3091–E3100.

Brand W. A. and Dobberstein P. (1996) Isotope-ratio-monitoring liquid chromatography mass spectrometry (IRM-LCMS): First results from a moving wire interface system. *Isotopes in Environmental and Health Studies* **32**, 275–283.

Breteler W., Grice K. and … S. S. (2002) Stable carbon isotope fractionation in the marine copepod Temora longicornis: unexpectedly low δ13C value of faecal pellets. *Marine ecolo* **240**, 195–204.

Bubis J. A., Levitsky L. I., Ivanov M. v., Tarasova I. A. and Gorshkov M. v. (2017) Comparative evaluation of label-free quantification methods for shotgun proteomics. *Rapid Communications in Mass Spectrometry* **31**, 606–612.

Cabana G. and Rasmussen J. B. (1996) Comparison of aquatic food chains using nitrogen isotopes. *Proc Natl Acad Sci U S A* **93**, 10844–10847.

Caimi, Richard and Brenna T. (1993) High-precision liquid chromatography-combustion isotope ratio mass spectrometry. *Analytical Chemistry* **65**, 3497–3500.

Callow R. H. T. and Brasier M. D. (2009) Remarkable preservation of microbial mats in Neoproterozoic siliciclastic settings: Implications for Ediacaran taphonomic models. *Earth-Science Reviews* **96**, 207–219.

Cano M., Holland S. C., Artier J., Burnap R. L., Ghirardi M., Morgan J. A. and Yu J. (2018) Glycogen Synthesis and Metabolite Overflow Contribute to Energy Balancing in Cyanobacteria. *Cell Reports*.

Carreira C., Staal M., Middelboe M. and Brussaard C. P. D. (2015) Counting viruses and bacteria in photosynthetic microbial mats. *Applied and Environmental Microbiology* **81**, 2149–2155.

Cheng P., Xiao X. M., Gai H. F., Li T. F., Zhang Y. Z., Huang B. J. and Wilkins R. W. T. (2015) Characteristics and origin of carbon isotopes of n-alkanes in crude oils from the western Pearl River Mouth Basin, South China sea. *Marine and Petroleum Geology* **67**, 217–229.

Clayton C. and Bjorøy M. (1994) Effect of maturity on 13C12C ratios of individual compounds in North Sea oils. *Organic Geochemistry* **21**, 737–750.

Close H. G., Bovee R. and Pearson A. (2011) Inverse carbon isotope patterns of lipids and kerogen record heterogeneous primary biomass. *Geobiology* **9**, 250–265.

Close H. G., Wakeham S. G. and Pearson A. (2014) Lipid and13C signatures of submicron and suspended particulate organic matter in the Eastern Tropical North Pacific: Implications for the contribution of Bacteria. *Deep-Sea Research Part I: Oceanographic Research Papers* **85**, 15–34.

Crockford P. W., Hayles J. A., Bao H., Planavsky N. J., Bekker A., Fralick P. W., Halverson G. P., Bui T. H., Peng Y. and Wing B. A. (2018) Triple oxygen isotope evidence for limited mid-Proterozoic primary productivity. *Nature 2018 559:7715* **559**, 613–616.

Dawson D., Grice K., and Alexander R. (2007) The effect of source and maturity on the stable isotopic compositions of individual hydrocarbons in sediments and crude oils from the Vulcan Sub-basin, Timor Sea. *Geochemistry.*

Dekas A. E., Parada A. E., Mayali X., Fuhrman J. A., Wollard J., Weber P. K. and Pett-Ridge J. (2019) Characterizing Chemoautotrophy and Heterotrophy in Marine Archaea and Bacteria With Single-Cell Multi-isotope NanoSIP. *Frontiers in Microbiology* **10**.

Deming W. E. (1943) Statistical adjustment of data.

DeNiro M. J. . and Epstein S. (1977) Mechanism of Carbon Isotope Fractionation Associated with Lipid Synthesis Author ( s ): Michael J . DeNiro and Samuel Epstein Published by : American Association for the Advancement of Science *Science (1979)* **197**, 261–263.

Djokic T., Kranendonk M. J. van, Campbell K. A., Havig J. R., Walter M. R. and Guido D. M. (2021) A Reconstructed Subaerial Hot Spring Field in the *3.5 Billion-Year-Old Dresser Formation, North Pole Dome, Pilbara Craton, Western Australia. *Astrobiology* **21**.

van Dongen B. E., Schouten S. and Sinninghe Damsté J. S. (2002) Carbon isotope variability in monosaccharides and lipids of aquatic algae and terrestrial plants. *Marine Ecology Progress Series* **232**, 83–92.

Fantle M. S., Dittel A. I., Schwalm S. M., Epifanio C. E. and Fogel M. L. (1999) A food web analysis of the juvenile blue crab, Callinectes sapidus, using stable isotopes in whole animals and individual amino acids. *Oecologia* **120**, 416–426.

Ferris M. J. and Ward D. M. (1997) Seasonal Distributions of Dominant 16S rRNA-Defined Populations in a Hot Spring Microbial Mat Examined by Denaturing Gradient Gel Electrophoresis. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* **63**, 1375–1381.

Finkel Z. v., Follows M. J., Liefer J. D., Brown C. M., Benner I. and Irwin A. J. (2016) Phylogenetic diversity in the macromolecular composition of microalgae. *PLoS ONE* **11**, 1–16.

Flemming H. C. and Wingender J. (2010) The biofilm matrix. *Nature Reviews Microbiology* **8**, 623–633.

Fogg G. E. (1983) The Ecological Significance of Extracellular Products of Phytoplankton Photosynthesis. *Botanica Marina.*

Fry B. and Sherr E. B. (1984) delta super(13)C measurements as indicators of carbon flow in marine and freshwater ecosystems. *Contrib. Mar. Sci.* **27**, 13–47.

Gehling J. G. (1999) Microbial mats in terminal Proterozoic siliciclastics; Ediacaran death masks. *Palaios* **14**, 40–57.

Giovannoni S. J., Revsbech N. P., Ward D. M. and Castenholz R. W. (1987) Obligately phototrophic Chloroflexus: primary production in anaerobic hot spring microbial mats. *Archives of Microbiology.*

Glover D. M., Jenkins W. J. and Doney S. C. (2011) *Modeling methods for marine science.*,

Görke B. and Stülke J. (2008) Carbon catabolite repression in bacteria: Many ways to make the most out of nutrients. *Nature Reviews Microbiology.*

Hagadorn J. W. and Bottjer D. J. (1997) Wrinkle structures: Microbially mediated sedimentary structures common in subtidal siliciclastic settings at the Proterozoic-Phanerozoic transition. *Geology* **25**, 1047–1050.

Hamilton T. L., Bennett A. C., Murugapiran S. K. and Havig J. R. (2019) Anoxygenic Phototrophs Span Geochemical Gradients and Diverse Morphologies in Terrestrial Geothermal Springs. *mSystems* **4**, 1–25.

Hamilton T. L., Vogl K. and Peters J. W. (2012) Environmental constraints defining the distribution, composition, and evolution of chlorophototrophs in thermal features of Yellowstone National Park.

Hanada S., Takaichi S., Matsuura K. and Nakamura K. (2002) Roseiflexus castenholzii gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. *International Journal of Systematic and Evolutionary Microbiology* **52**, 187–193.

Havig J., Hamilton T., Bachan A., Kump L. R., Havig J. R. and Hamilton T. L. (2017) Sulfur and carbon isotopic evidence for metabolic pathway evolution and a four-stepped Earth system progression across the Archean and Paleoproterozoic

Havig J. R., Raymond J., Meyer-Dombard D. R., Zolotova N. and Shock E. L. (2011) Merging isotopes and community genomics in a siliceous sinter-depositing hot spring. *Journal of Geophysical Research: Biogeosciences* **116**, 1–15.

Hayes J. M. (2001) Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes. *Reviews in Mineralogy and Geochemistry* **43**, 225–277.

Hoffman H. J. (2000) Archean stromatolites as microbial archives. In *Microbial Sediments*

House C. H., Schopf J. W. and Stetter K. O. Carbon isotopic fractionation by Archaeans and other thermophilic prokaryotes.

Inskeep W. P., Jay Z. J., Tringe S. G., Herrgård M. J. and Rusch D. B. (2013) The YNP metagenome project: Environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. *Frontiers in Microbiology* **4**.

Jahnke L., Embaye T., Hope J., Turk K. A., van Zuilen M., des Marais D. J., Farmer J. D. and Summons R. E. (2004) Lipid biomarker and carbon isotopic signatures for stromatolite-forming, microbial mat communities and Phormidium cultures from Yellowstone National Park. *Geobiology* **2**, 31–47.

Jahnke L. L. and des Marais D. J. (2019) Carbon isotopic composition of lipid biomarkers from an endoevaporitic gypsum crust microbial mat reveals cycling of mineralized organic carbon. *Geobiology* **17**, 643–659.

Kenyon C. N., Rippka R. and Stanier R. Y. (1972) Fatty acid composition and physiological properties of some filamentous blue-green algae. *Archiv für Mikrobiologie*.

Klatt C. G., Inskeep W. P., Herrgard M. J., Jay Z. J., Rusch D. B., Tringe S. G., Parenteau M. N., Ward D. M., Boomer S. M., Bryant D. A. and Miller S. R. (2013) Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Frontiers in Microbiology* **4**.

Kleiner M., Dong X., Hinzke T., Wippler J., Thorson E., Mayer B. and Strous M. (2018) Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc Natl Acad Sci U S A* **115**, E5576–E5584.

Klock J. H., Wieland A., Seifert R. and Michaelis W. (2007) Extracellular polymeric substances (EPS) from cyanobacterial mats: Characterisation and isolation method optimisation. *Marine Biology* **152**, 1077–1085.

Leblanc D. J. and Ball A. J. S. (1978) A fast one-step method for the silylation of sugars and sugar phosphates. *Analytical Biochemistry* **84**, 574–578.

Logan G. A., Hayes J. M., Hieshima G. B. and Summons R. E. (1995) Terminal Proterozoic reorganization of biogeochemical cycles. *Nature* **376**, 53–56.

Luo G., Hallmann C., Xie S., Ruan X. and Summons R. E. (2015) Comparative microbial diversity and redox environments of black shale and stromatolite facies in the Mesoproterozoic Xiamaling Formation. *Elsevier* **151**, 150–167.

Luo G., Yang H., Algeo T. J., Hallmann C. and Xie S. (2019) Lipid biomarkers for the reconstruction of deep-time environmental conditions. *Earth-Science Reviews*.

Macko S. A., Fogel M. L., Hare P. E. and Hoering T. C. (1987) Isotopic fractionation of nitrogen and carbon in the synthesis of amino acids by microorganisms. *Chemical Geology: Isotope Geoscience Section* **65**, 79–92.

Mahmoudi N., Beaupré S. R., Steen A. D. and Pearson A. (2017) Sequential bioavailability of sedimentary organic matter to heterotrophic bacteria. *Environmental Microbiology* **19**, 2629–2644.

van Maldegem L. M., Sansjofre P., Weijers J. W. H., Wolkenstein K., Strother P. K., Wörmer L., Hefter J., Nettersheim B. J., Hoshino Y., Schouten S., Sinninghe Damsté J. S., Nath N., Griesinger C., Kuznetsov N. B., Elie M., Elvert M., Tegelaar E., Gleixner G. and Hallmann C. (2019) Bisnorgammacerane traces predatory pressure and the persistent rise of algal ecosystems after Snowball Earth. *Nature Communications 2019 10:1* **10**, 1–11.

Mayali X., Weber P. K., Brodie E. L., Mabery S., Hoeprich P. D. and Pett-Ridge J. (2012) High-throughput isotopic analysis of RNA microarrays to quantify microbial resource use. *ISME Journal* **6**, 1210–1221.

van der Meer M. T. J., Klatt C. G., Wood J., Bryant D. A., Bateson M. M., Lammerts L., Schouten S., Sinninghe Damste J. S., Madigan M. T. and Ward D. M. (2010) Cultivation and genomic, nutritional, and lipid

biomarker characterization of Roseiflexus strains closely related to predominant in situ populations inhabiting yellowstone hot spring microbial mats. *Journal of Bacteriology* **192**, 3033–3042.

van der Meer M. T. J., Schouten S., Damsté J. S. S. and Ward D. M. (2007) Impact of carbon metabolism on 13C signatures of cyanobacteria and green non-sulfur-like bacteria inhabiting a microbial mat from an alkaline siliceous hot spring in Yellowstone National Park (USA). *Environmental Microbiology* **9**, 482–491.

van der Meer M. T. J., Schouten S., van Dongen B. E., Rijpstra W. I. C., Fuchs G., Sinninghe Damsté J. S., de Leeuw J. W. and Ward D. M. (2001a) Biosynthetic Controls on the 13C Contents of Organic Components in the Photoautotrophic Bacterium Chloroflexus aurantiacus. *Journal of Biological Chemistry* **276**, 10971–10976.

van der Meer M. T. J., Schouten S., van Dongen B. E., Rijpstra W. I. C., Fuchs G., Sinninghe Damsté J. S., de Leeuw J. W. and Ward D. M. (2001b) Biosynthetic Controls on the13C Contents of Organic Components in the Photoautotrophic Bacterium Chloroflexus aurantiacus. *Journal of Biological Chemistry* **276**, 10971–10976.

van der Meer M. T. J. van der, Schouten S., Jaap S., Damsté S., Leeuw J. W. de and Ward D. M. (2003a) Compound-Specific Isotopic Fractionation Patterns Suggest Different Carbon Metabolisms among Chloroflexus -Like Bacteria in Hot-Spring Microbial Mats Compound-Specific Isotopic Fractionation Patterns Suggest Different Carbon Metabolisms among Chloroflexus. *Applied & Environmental Microbiology* **69**, 6000–6006.

van der Meer M. T. J. van der, Schouten S., Mary M., Nübel U., Wieland A., Kühl M., Leeuw W. de, Damsté J. S. S., Ward D. M., Bateson M. M., Nu U., Ku M., Leeuw J. W. de and Damste J. S. S. (2005) Diel Variations in Carbon Metabolism by Green Nonsulfur-Like Bacteria in Alkaline Siliceous Hot Spring Microbial Mats from Yellowstone National Park. *American Society for Microbiology* **71**, 3978–3986.

Melzer E. and Schmidt H. L. (1987) Carbon Isotope Effects on the Pyruvate-Dehydrogenase Reaction and Their Importance for Relative C-13 Depletion in Lipids. *Journal of Biological Chemistry* **262**, 8159–8164.

Mohr W., Tang T., Sattin S. R., Bovee R. J. and Pearson A. (2014) Protein stable isotope fingerprinting: Multidimensional protein chromatography coupled to stable isotope-ratio mass spectrometry. *Analytical Chemistry* **86**, 8514–8520.

Monson K. D. and Hayes J. M. (1982) Carbon Isotopic Fractionation in the Biosynthesis of Bacterial Fatty-Acids - Ozonolysis of Unsaturated Fatty-Acids as a Means of Determining the Intramolecular Distribution of Carbon Isotopes. *Geochimica Et Cosmochimica Acta* **46**, 139–149.

Mouginot C., Zimmerman A. E., Bonachela J. A., Fredricks H., Allison S. D., van Mooy B. A. S. and Martiny A. C. (2015) Resource allocation by the marine cyanobacterium Synechococcus WH8102 in response to different nutrient supply ratios. *Limnology and Oceanography* **60**, 1634–1641.

Musilova M., Tranter M., Bennett S. A., Wadham J. and Anesio A. M. (2015) Stable microbial community composition on the Greenland Ice Sheet. *Frontiers in Microbiology* **6**.

de Niro, Michael, Epstein S. (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochimica et Cosmochimica Acta* **42**, 495–506.

Nold S. C. and Ward D. M. (1996) Photosynthate partitioning and fermentation in hot spring microbial mat communities. *Appl Environ Microbiol* **62**, 4598–4607.

Odden W., Barth T. and Talbot M. (2002) Compound-specific carbon isotope analysis of natural and artificially generated hydrocarbons in source rocks and petroleum fluids from offshore Mid-Norway. *Organic Geochemistry* **33**, 47–65.

Osterhout J., Schopf J. W., Williford K., McKeegan K., Kudryavtsev A. B. and Liu M.-C. (2021) Carbon isotopes of Proterozoic filamentous microfossils: SIMS analyses of ancient cyanobacteria from two disparate shallow-marine cherts. *Geomicrobiology Journal* **38**, 719–731.

Parenteau M. N. (2007) Microbial biosignatures in high-iron thermal springs. Portland State University.

Parenteau M. N., Jahnke L. L., Farmer J. D. and Cady S. L. (2014) Production and Early Preservation of Lipid Biomarkers in Iron Hot Springs. *Astrobiology* **14**, 502–521.

Pawlowska M. M., Butterfield N. J. and Brocks J. J. (2013) Lipid taphonomy in the Proterozoic and the effect of microbial mats on biomarker preservation. *Geology* **41**, 103–106.

Pearson A., McNichol A. P., Benitez-Nelson B. C., Hayes J. M. and Eglinton T. I. (2001) Origins of lipid biomarkers in Santa Monica Basin surface sediment: A case study using compound-specific Δ14C analysis. *Geochimica et Cosmochimica Acta* **65**, 3123–3137.

Pedrosa-Pàmies R., Conte M., … J. W.-P. in and 2018 undefined (2019) Carbon cycling in the Sargasso Sea water column: insights from lipid biomarkers in suspended particles. *Elsevier*.

Penning H. and Conrad R. (2006) Carbon isotope effects associated with mixed-acid fermentation of saccharides by Clostridium papyrosolvens. *Geochimica et Cosmochimica Acta* **70**, 2283–2297.

Perry G. J., Volkman J. K. and Johns R. B. (1979) Fatty acids of bacterial origin in contemporary marine sediments.

Pierson B. K. and Castenholz R. W. (1974) Studies of pigments and growth in Chloroflexus aurantiacus, a phototrophic filamentous bacterium. *Archives of Microbiology* **100**, 283–305.

Pierson B. K. and Parenteau M. N. (2000) Phototrophs in high iron microbial mats: Microstructure of mats in iron-depositing hot springs. *FEMS Microbiology Ecology*.

Pierson B. K., Parenteau M. N. and Griffin D. B. M. (1999) Phototrophs in high-iron-concentration microbial mats: Physiological ecology of phototrophs in an iron-depositing hot spring. *Applied and Environmental Microbiology* **65**, 5474–5483.

Pinnegar J. K. and Polunin N. V. C. (2000) Contributions of stable-isotope data to elucidating food webs of Mediterranean rocky littoral fishes. *Oecologia* **122**, 399–409.

Radajewski S., Ineson P., Parekh N. R. and Murrell J. C. (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* **403**, 646–649.

Reinhard C. T., Planavsky N. J., Gill B. C., Ozaki K., Robbins L. J., Lyons T. W., Fischer W. W., Wang C., Cole D. B. and Konhauser K. O. (2017) Evolution of the global phosphorus cycle. *nature.com* **541**.

Ruff-Roberts A. L., Gijs Kuenen J. and Ward' D. M. (1994) Distribution of Cultivated and Uncultivated Cyanobacteria and Chloroflexus-Like Bacteria in Hot Spring Microbial Mats. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 697–704.

Sakata S., Hayes J. M., McTaggart A. R., Evans R. A., Leckrone K. J. and Togasaki R. K. (1997) Carbon isotopic fractionation associated with lipid biosynthesis by a cyanobacterium: Relevance for interpretation of biomarker records. *Geochimica et Cosmochimica Acta* **61**, 5379–5389.

Schopf J. W., Kudryavtsev A. B., Czaja A. D. and Tripathi A. B. (2007) Evidence of Archean life: Stromatolites and microfossils. *Precambrian Research* **158**, 141–155.

Schuler C. G., Havig J. R. and Hamilton T. L. (2017) Hot spring microbial community composition, morphology, and carbon fixation: Implications for interpreting the ancient rock record. *Frontiers in Earth Science* **5**, 1–17.

Sessions A. L., Sylva S. P. and Hayes J. M. (2005) Moving-wire device for carbon isotopic analyses of nanogram quantities of nonvolatile organic carbon. *Analytical Chemistry* **77**, 6519–6527.

Shen, V.K., Siderius, D.W., Krekelberg, W.P., and Hatch H. W. ed. (2017) *NIST Standard Reference Simulation Website*. NIST Stand., National Institute of Standards and Technology, Gaithersburg MD.

Shih P. M., Ward L. M. and Fischer W. W. (2017) Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the Chloroflexi. *Proc Natl Acad Sci U S A* **114**, 10749–10754.

Stal L. J. and Moezelaar R. (1997) Fermentation in cyanobacteria. *FEMS Microbiology Reviews* **21**, 179–211.

Steiner M. and Reiter J. (2001) Evidence of organic structures in Ediacara-type fossils and associated microbial mats. *Geology* **29**, 1119–1122.

Stuart R. K., Mayali X., Lee J. Z., Craig Everroad R., Hwang M., Bebout B. M., Weber P. K., Pett-Ridge J. and Thelen M. P. (2016) Cyanobacterial reuse of extracellular organic carbon in microbial mats. *ISME Journal* **10**, 1240–1251.

Sturt H. F., Summons R. E., Smith K., Elvert M. and Hinrichs K. U. (2004) Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry - New biomarkers for biogeochemistry and microbial ecology. *Rapid Communications in Mass Spectrometry* **18**, 617–628.

Tang T., Mohr W., Sattin S. R., Rogers D. R., Girguis P. R. and Pearson A. (2017) Geochemically distinct carbon isotope distributions in Allochromatium vinosum DSM 180Tgrown photoautotrophically and photoheterotrophically. *Geobiology* **15**, 324–339.

Tang Y., Huang Y., Ellis G. S., Wang Y., Kralert P. G., Gillaizeau B., Ma Q. and Hwang R. (2005) A kinetic model for thermally induced hydrogen and carbon isotope fractionation of individual n-alkanes in crude oil. *Geochimica et Cosmochimica Acta* **69**, 4505–4520.

Tarhan L. G. (2018) The early Paleozoic development of bioturbation—Evolutionary and geobiological consequences. *Earth-Science Reviews* **178**, 177–207.

Teece M. A. and Fogel M. L. (2007) Stable carbon isotope biogeochemistry of monosaccharides in aquatic organisms and terrestrial plants. *Organic Geochemistry* **38**, 458–473.

Thomas A. T., Ognibene T., Daley P., Turteltaub K., Radousky H. and Bench G. (2011) Ultrahigh efficiency moving wire combustion interface for online coupling of high-performance liquid chromatography (HPLC). *Analytical Chemistry* **83**, 9413–9417.

Tian C., Xia Y., Song C., Ma S., Gao W. and Xing L. (2017) Changes in the carbon isotope composition of pristane and phytane with increasing maturity. *http://dx.doi.org/10.1080/10916466.2017.1324484* **35**, 1270–1276.

Tibocha-Bonilla J. D., Kumar M., Richelle A., Godoy-Silva R. D., Zengler K. and Zuñiga C. (2020) Dynamic resource allocation drives growth under nitrogen starvation in eukaryotes. *npj Systems Biology and Applications* **6**, 1–9.

Větrovský T. and Baldrian P. (2013) The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE* **8**, e57923.

Vinnichenko G., Jarrett A. J. M., van Maldegem L. M. and Brocks J. J. (2021) Substantial maturity influence on carbon and hydrogen isotopic composition of n-alkanes in sedimentary rocks. *Organic Geochemistry* **152**, 104171.

de Vos P., Garrity G. M., Jones D., Krieg N. R., Ludwig W., Rainey F. A., Schleifer K.-H. and Whitman W. B. (2009) *Bergey's manual of systematic bacteriology Volume Three The Firmicutes*.,

Werne J. P., Baas M. and Sinninghe Damsté J. S. (2002) Molecular isotopic tracing of carbon flow and trophic relationships in a methane-supported benthic microbial community. *Limnology and Oceanography* **47**, 1694–1701.

Whitman W. B., Parte A. C., Krieg N. R., Staley J. T., Brown D. R., Hedlund B. P., Paster B. J., Ward N. L. and Ludwig W. (2010) *Bergey's Manual of Systematic Bacteriology, Volume Four, The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*.,

Wieland A., Pape T., Möbius J., Klock J. H. and Michaelis W. (2008) Carbon pools and isotopic trends in a hypersaline cyanobacterial mat. *Geobiology* **6**, 171–186.

Williford K. H., Ushikubo T., Schopf J. W., Lepot K., Kitajima K. and Valley J. W. (2013) Preservation and detection of microstructural and taxonomic correlations in the carbon isotopic compositions of individual Precambrian microfossils. *Elsevier*.

van der Zanden M. J. and Rasmussen J. B. (2001) Variation in δ15N and δ13C trophic fractionation: Implications for aquatic food web studies. *Limnology and Oceanography* **46**, 2061–2066.

Zarzycki J. and Fuchs G. (2011) Coassimilation of organic substrates via the autotrophic 3-hydroxypropionate bi-cycle in Chloroflexus aurantiacus. *Applied and Environmental Microbiology*.

# Chapter 3

## Isotopic signatures of carbon transfer in a Proterozoic analogue microbial mat

This chapter is currently under review for publication in *Applied and Environmental Microbiology* in collaboration with co-authors Sharon L. Grim, Jacob Waldbauer, Gregory J. Dick and Ann Pearson.

**Abstract**

Modern microbial mats are potential analogues for Proterozoic ecosystems, yet only a handful of studies have characterized mats under low oxygen conditions relevant to Proterozoic environments. Here we use Protein-Stable Isotope Fingerprinting (P-SIF) to determine the protein carbon isotope ($\delta^{13}C$) values of autotrophic, heterotrophic, and mixotrophic organisms in a benthic microbial mat from the low oxygen Middle Island Sinkhole, Lake Huron, USA (MIS). We also measure the $\delta^{13}C$ values of the sugar moieties of exopolysaccharides (EPS) within the mat to explore the relationship between cyanobacterial exudates and heterotrophic anabolic carbon uptake. Our results show that Cyanobacteria (autotrophs) are $^{13}C$-depleted relative to sulfate reducing bacteria (heterotrophs) and $^{13}C$-enriched relative to sulfur oxidizing bacteria (autotrophs or mixotrophs). We also find that the pentose moieties of EPS are systematically enriched in $^{13}C$ relative to the hexose moieties of EPS. We hypothesize that these isotopic patterns reflect cyanobacterial metabolic pathways, particularly phosphoketolase, that are relatively more active in low oxygen rather than oxygenated mat environments, resulting in isotopically more heterogeneous C sources in low oxygen mats. While this might partially explain the isotopic variability observed in Proterozoic mat facies, further work is necessary to systematically characterize the isotopic fractionations associated with the synthesis of cyanobacterial exudates.

**Importance**

The $\delta^{13}C$ composition of heterotrophic microorganisms is dictated by the $\delta^{13}C$ composition of their organic carbon source. In both modern and ancient photosynthetic microbial mats, photosynthetic exudates are the most likely source of organic carbon for heterotrophs. We measured the $\delta^{13}C$ values of autotrophic, heterotrophic and mixotrophic bacteria, and the $\delta^{13}C$ value of the most abundant photosynthetic exudate (exopolysaccharide) in a modern analogue for a Proterozoic environment. Given these data, future studies will be better equipped to estimate the most likely carbon source for heterotrophs in both modern environments and Proterozoic environments preserved in the rock record.

**Introduction**

For the majority of life's history on Earth, ecosystems were entirely microbial (Knoll and Nowak, 2017, and references therein). Both ichnofossil evidence (Tarhan, 2018) and the presence of microbial textures in geologic features (Hagadorn and Bottjer, 1997; Gehling, 1999; Steiner and Reiter, 2001; Callow and Brasier, 2009) suggest microbial mats were widespread in the Proterozoic and early Paleozoic even after the oldest reliably eukaryotic fossils appear in the rock record at 1.65 Ga (Javaux, 2019). Using modern microbial mats as analogues, researchers have suggested Proterozoic mat environments hosted the first origin of eukaryotes (López-García and Moreira, 2020) and, considering oxygenic cyanobacterial mats specifically, sustained animal life in an otherwise low-oxygen environment (Gingras et al., 2011). However, few studies on modern microbial mats include phototrophic mats in persistently low-$O_2$ and/or sulfidic environments (Grim et al., 2021), conditions which were likely widespread in coastal Proterozoic habitats

43

(Hanson et al., 2013; Klatt et al., 2020). To test evolutionary hypotheses invoking Proterozoic microbial mats, more research is necessary in low-$O_2$ and/or sulfidic environments specifically.

The submerged Middle Island Sinkhole (MIS) in Lake Huron, the focus of this study, has been previously identified as a potential analogue for benthic Proterozoic ecosystems (Figure 3.1; Voorhies et al., 2012; Grim et al., 2021). Venting groundwater at the bottom of MIS (23m water depth) creates a stratified benthic layer with lower temperature (7-9°C), lower concentrations of dissolved oxygen (0-2 mg $L^{-1}$) and higher concentrations of dissolved sulfate (1250 mg $L^{-1}$) (Biddanda et al., 2009; Voorhies et al., 2012) than the overlying lake water. Within Cyanobacterial mats at MIS, organic carbon production is greater than oxygen production by oxygenic photosynthesis (OP), emphasizing the importance of anoxygenic photosynthesis (AP) and chemosynthesis and resulting in a net consumption of $O_2$ within the benthic layer (Voorhies et al. 2012). Metagenomic, metatranscriptomic and 16S rRNA profiles suggest the mats are dominated by versatile Cyanobacteria capable of OP, AP and fermentation (Nold et al., 2010; Voorhies et al., 2012). Additionally, the mats support an abundant and diverse set of Proteobacteria, including sulfate reducing bacteria (SRB) that are active during both night and day, including when the mat contains measurable dissolved oxygen from OP (Medina, 2017; Grim et al., 2021).

These Proterozoic analogue systems therefore have intriguing patterns of sequential redox cycling. In a mat sampled from a different system ("Main Spring" of the Frasassi cave outlet, Italy), Klatt et al. (2020) found that AP only occurred after sulfide was released via sulfate reduction by SRB, which itself only occurred after dissolved organic carbon was excreted via OP. In contrast to Lake Huron MIS, the dependence of both SRB and AP on the organic products excreted by OP made the Frasassi mat a net source of $O_2$ (Klatt et al., 2020). From these and similar studies, it appears that feedbacks between oxidizing and reducing chemical species may not have

been the only factors determining whether Cyanobacterial mats were net sources or sinks of $O_2$ during the Proterozoic, as has been previously suggested (e.g., Johnston et al., 2009), but also that the specific sequence of carbon transfer between organisms is important.

In modern ecosystems, networks of carbon transfer can be traced by adding $^{14}C$ and/or $^{13}C$-labeled carbon sources to either the surrounding environment (*e.g.*, van der Meer et al., 2005), cultures of microbial isolates (*e.g.,* Bateson and Ward, 1988; Nold and Ward, 1996) or incubations of environmental samples (*e.g.,* Nold and Ward 1996). For modern analogue work to yield information that can be applied to the rock record, natural abundance carbon stable isotope compositions ($\delta^{13}C$) of preservable biomolecules are more appropriate. Typically, the $\delta^{13}C$ values of ancient organic compounds are interpreted by comparison to culture studies of modern microbial metabolism. For example, microbial biomass produced via the Calvin-Benson-Bassham (reductive pentose phosphate) cycle using a Type I RuBisCO enzyme is depleted in $^{13}C$ by 12-26‰ relative to dissolved inorganic carbon (DIC), while biomass produced via the rTCA cycle is depleted by only 2-13‰ (House et al., 2003). In modern microbial mats, carbon transfer from autotrophs to heterotrophs occurs via the assimilation of metabolic intermediates excreted into the extracellular environment (*e.g.,* Bateson and Ward, 1988; Stuart et al., 2015). Since heterotrophic carbon assimilation does not fractionate organic carbon (DeNiro and Epstein, 1977; Blair et al., 1985), the $\delta^{13}C$ value of heterotrophic biomass reflects the $\delta^{13}C$ value of the specific organic compound assimilated. Depending on what is excreted by autotrophic metabolism, heterotrophic biomass may be isotopically indistinguishable (*e.g.,* exopolysaccharide, (Gonzalez-Nayeck et al., 2022), relatively $^{13}C$-depleted (*e.g.,* methane, Potter et al., 2009) or relatively $^{13}C$-enriched (*e.g.,* acetate, Blair et al., 1985; Penning and Conrad, 2006) relative to autotrophic biomass. In theory, the $\delta^{13}C$ composition of autotrophic and heterotrophic biomass, DIC, and excreted organic carbon

compounds in modern microbial mats might follow a reproducible pattern depending on the mode of carbon transfer in the mat. Characterizing these patterns in well-studied Proterozoic analogue systems may yield patterns that could be used to determine whether ancient microbial mat systems were likely sources or sinks of $O_2$.

Previously, we used Protein Stable Isotope Fingerprinting (P-SIF; Mohr et al., 2014) to measure the $\delta^{13}C$ values of whole proteins separated from an oxygenated, photosynthetic microbial mat in a terrestrial hydrothermal outflow channel in Yellowstone National Park (YNP), USA (GonzalezNayeck et al., 2022). The same proteins also were classified taxonomically via proteomics. Because $\delta^{13}C$ values of proteins are consistently offset from biomass $\delta^{13}C$ values (Abelson and Hoering, 1961; Blair et al., 1985), this approach yielded phylum-specific $\delta^{13}C$ signatures for the primary autotroph and heterotrophic populations present in the mat at the time of sampling. The results indicated that Cyanobacteria and obligate heterotrophs such as Actinobacteria in this system have indistinguishable $\delta^{13}C$ signatures. Concurrently, we measured the $\delta^{13}C$ values for *n*-alkyl lipids and the monosaccharide moieties from exopolysaccharide (EPS). Glucose moieties in exopolysaccharide were equal in $\delta^{13}C$ composition to both cyanobacterial and heterotrophic proteins, and the lipid pool was dominated by highly $^{13}C$-depleted fatty acids. From these data, we concluded that 1) producers and consumers in this system were sharing primary photosynthate as a common resource, and 2) Cyanobacteria were allocating most of their fixed carbon to exopolysaccharides. These results were consistent with prior literature on the fate of fixed carbon in microbial ecosystems within YNP which suggest that Cyanobacterial glycogen is a key source of organic carbon to other mat-based organisms (e.g., Nold and Ward 1996). However, it is unlikely that these results are applicable to benthic low oxygen ecosystems with

lower photon fluxes: under these conditions, Cyanobacteria are often light-limited and allocate a smaller portion of primary photosynthate to storage sugars (Cano et al., 2018).

For the present study, we hypothesized that Cyanobacteria in MIS microbial mats would allocate relatively less fixed carbon to exopolysaccharide than Cyanobacteria in surface microbial mats such as YNP. This may then imply that the resulting cyanobacterial exudates (i.e., heterotrophic anabolic carbon sources) should have variable $\delta^{13}C$ compositions and affect the signatures of heterotrophic consumers. To test this hypothesis, we measured the $\delta^{13}C$ composition of mat exopolysaccharide and fatty acids, and the protein $\delta^{13}C$ values of autotrophic, heterotrophic, and mixotrophic organisms in MIS.
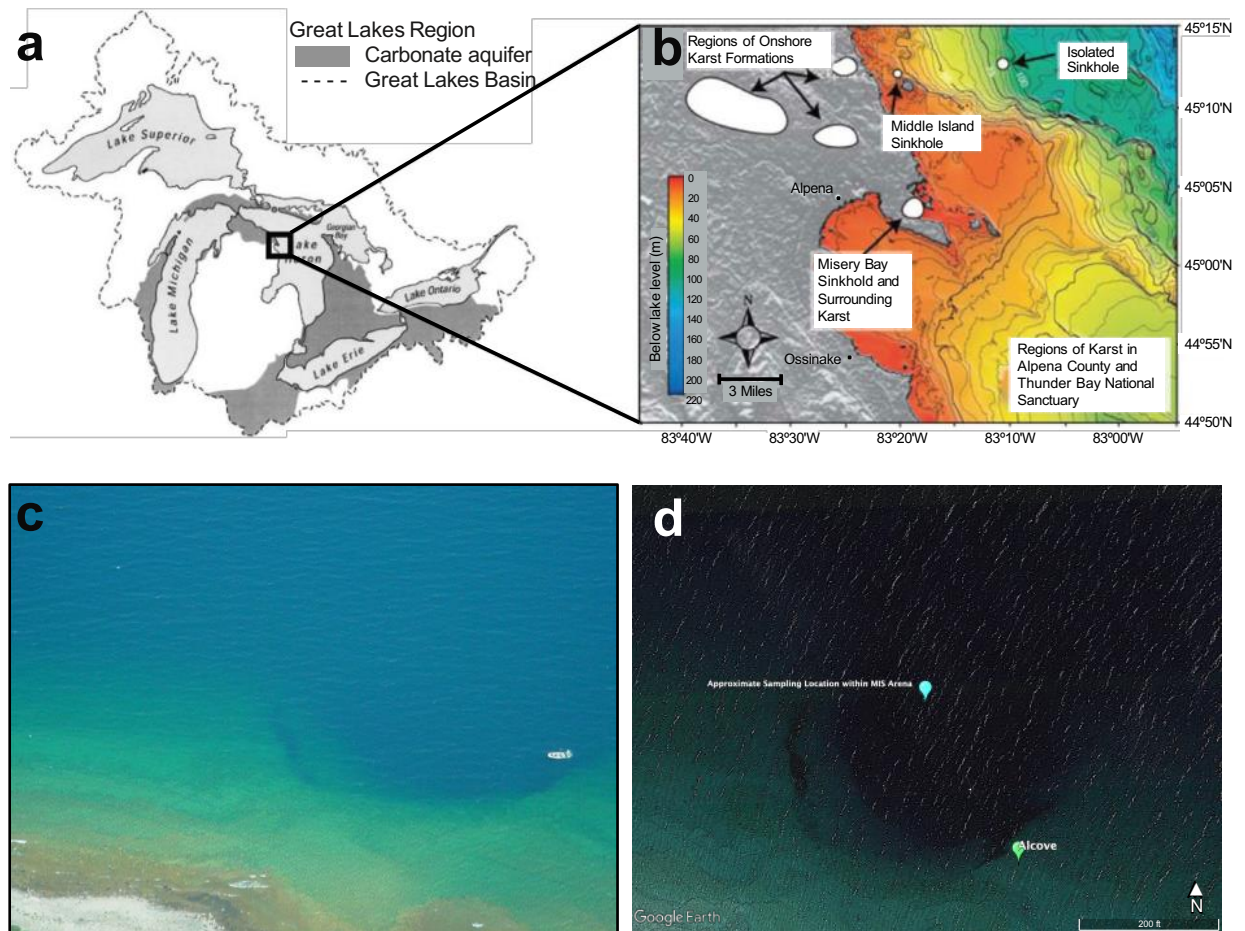
**Figure 3.1** An overview of our sampling location, modified from Biddanda et al., (2009). (a) Map of the North American Laurentian Great Lakes Basin showing regions of carbonate aquifers. (b) Regions of aboveground karst formations and submerged sinkholes within Lake Huron in the area of the black box in (a) (depth contours in meters, 0–150 meters; modified from Coleman [2002]). (c) Aerial photo of the Middle Island sinkhole (MIS), with a 9- meter Boston Whaler boat on the right for scale. (d) the specific sampling location within the MIS for our study. "Alcove" represents the lighter carbonate platform seen in (c).

**Materials and Methods**

*Sample collection*

Two flat-mat samples were collected by scuba divers from R/V *Storm* on May 31st, 2017 (hereafter

LH22) and August 7th, 2017 (hereafter LH47) from the same location within a 100-m area of the

Middle Island sinkhole arena (45.1984°N, 83.32721°W) by hand push core (Figure 3.1). Cores were rapidly transferred to the surface, and mats were immediately placed on dry ice before being shipped to the laboratory. The groundwater layer overlying the sediment had higher specific conductivity (1856 μS cm$^{-1}$), lower temperature (10-12˚C), and lower dissolved oxygen (2.38 mg/L or 23% saturation), than the surface of Lake Huron (205 μS cm$^{-1}$, 19˚C) (Grim et al. *in prep*).

*Lipid extraction and identification*

Lipids were extracted from approximately 0.3 grams (dry) of LH22 and 0.2 grams (dry) of LH47 freeze-dried mat samples via a modified Bligh and Dyer procedure (Sturt et al., 2004). The total lipid extracts were transesterified to generate fatty acid methyl esters (FAMEs; 5% HCl/methanol (v/v), 70°C, 4 hr). The reactions were stopped by the addition of ultrapure $H_2O$, after which the organic phases were extracted into hexane/dichloromethane (4:1, v/v). FAME derivatives of *n*-C$_{16:0}$, *n*-C$_{19:0}$, and *n*-C$_{24:0}$ FA standards with known $\delta^{13}C$ compositions (-29.5‰, -31.7‰, and -30.8‰, respectively) were prepared in parallel to correct for the $^{13}C$ content of the derivatized carbon introduced during transesterification. FAMEs were further separated from the derivatized extracts by elution over $SiO_2$ gel using the solvent program described in (Pearson et al., 2001).

FAMEs were identified using gas chromatography–mass spectrometry (GC/MS; Agilent 6890N GC, 5973 MS equipped with a 30m DB-5MS column) by comparison to known patterns of relative retention times (Perry et al., 1979; Pearson et al., 2001) and by comparison of fragment mass spectra to spectra from the National Institute of Standards and Technology Library (Shen, V.K., Siderius, D.W., Krekelberg, W.P., and Hatch, 2017). The injection, oven temperature programs and gas flow rates were adopted from Close et al., 2014.

*Protein stable isotope fingerprinting*

Protein Stable Isotope Fingerprinting was performed as previously described (Mohr et al., 2014). Due to the relatively large sample requirements (multiple grams) for P-SIF, P-SIF was only performed on LH47. Proteins were extracted by placing 21.5 grams of wet mat material and up to 8 mL of B-PER protein extraction reagent (Thermo Scientific) in two 50 mL Teflon tube and sonicating using a 500-watt Qsonica ultrasonic processor equipped with a cup horn. The cup horn was filled with ice water and the sonicator was set to 25 seconds on and 35 seconds off for a total of 5 minutes sonication. Solids and cell material were removed by centrifugation at 16,000 *g*. Proteins were precipitated from the supernatant in acetone and resuspended in 100 mM $NH_4HCO_3$, pH 9 to yield a total soluble protein extract. This extract was further separated into 960 fractions on an Agilent 1100 series HPLC with DAD detector and fraction collector using two orthogonal levels of chromatography: first by strong anion exchange (SAX; Agilent PL-SAX column; $4.6 \times 50$ mm, 8 μm) (20 fractions), then by reverse phase (RP; Agilent Poroshell 300SB-C3 column, $2.1 \times 75$ mm, 5 μm; 48 fractions), using the solvent gradients described in (Mohr et al., 2014). An aliquot of each final fraction is split into 96-well plates for isotope analysis (70%) and the remaining 30% is reserved for tryptic digestion followed by peptide sequencing.

*Protein taxonomic identification*

Plates for tryptic digestion were prepared as detailed in (Mohr et al., 2014). Peptide samples were separated by nanoflow liquid chromatography on a capillary C18 column (Thermo Acclaim PepMap 100 Å, 2 μm particles, 50 μm I.D. × 50 cm length) using a water/acetonitrile + 0.1% formic acid gradient (2–50% AcN over 180 min) at 90 nL/min using a Dionex Ultimate 3000 LC system with nanoelectrospray ionization (Proxeon Nanospray Flex source). Mass spectra were collected on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) operating in a

data-dependent acquisition (DDA) mode, with one high resolution (120,000 $m/\Delta m$) MS1 parent ion full scan triggering 15 Rapid-mode MS2 CID fragment ion scans of selected precursors.

Proteomic mass spectra were matched to peptide sequences using Sequest HT implemented in Proteome Discoverer 2.1. A combined metagenomic assembly of four metagenomes collected in 2016 from Middle Island Sinkhole microbial mat were used as the search database. Those data are available via IMG (taxon object ID: 3300028549), and SRA accession numbers (SRR21758066SRR21758069) under NCBI BioProject PRJNA72255. Proteomic mass spectral data are available via the MassIVE repository (massive.ucsd.edu) under accession MSV000090594.

Label-free protein abundances in each well were estimated by summing the integrated MS1 peptide peak areas for peptides assigned to specific proteins. Peptides mapping onto more than one protein were assigned to the protein with the most peptide evidence (Zhang et al., 2010). Relative abundances of phylogenetic groups in each well were determined by comparing the sum of all peptide peak areas for proteins taxonomically assigned to a given phylogenetic group to the sum of all peptide peak areas for proteins in a given well (Mohr et al., 2014).

*Sugar extraction and derivatization*

Due to sample limitation, EPS was only extracted from sample LH22. Briefly, 20 mL of 10% (w/v) NaCl was added to 10 grams of wet homogenized microbial mat sample and vortexed using established methods (Klock et al., 2007). This solution was incubated at 40ºC for 15 minutes, followed by centrifugation at 8,200 *g* for 15 minutes. The supernatant was collected, and the precipitant was re-extracted with 20 mL 10% NaCl two more times. After cooling in an ice bath, 100% ethanol was added to the supernatant to a final concentration of 70%. EPS was precipitated at 4ºC overnight and removed by centrifugation.

Extracted EPS was hydrolyzed into monomers using the method of van Dongen et al., (2001) by vortexing with 1mL 12M $H_2SO_4$ in a Teflon tube and then stirring at ~ 400 RPM for 2 hours at room temperature, followed by dilution to 1 M and heating (85ºC for 4.5hr). After cooling to room temperature, the solution was neutralized to pH 7 using $BaCO_3$. Once neutralized, the solution was centrifuged at 4,000 $g$ for 5 minutes, after which the supernatant was collected, frozen, and lyophilized.

Lyophilized sugar monomers were derivatized immediately prior to isotope analysis, again using established protocols (van Dongen et al., 2001). Arabinose, xylose, glucose and myo-inositol standards with known $\delta^{13}C$ compositions (-11.7‰, -9.7 ‰, -11.1‰, and -14.4‰, respectively) were prepared in parallel to correct for the $^{13}C$ content of the carbon introduced during derivatization. Briefly, 1 mL of a methylboronic acid/pyridine (10 mg/mL) mixture was added to 5 mg of sample or to 1 mg of total glucose, arabinose and xylose standards and then heated at 60ºC for 30 minutes, followed by addition of 100µL N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and a further 5 minutes of heating. Because myo-inositol is insoluble in pyridine, 250-500 µg myo-inositol was instead dissolved in 1 mL dimethyl sulfoxide (DMSO) for derivatization. After the heating step, the myo-inositol standard mixture was cooled to room temperature before adding 1 mL cyclohexane and 100 µL of BSTFA (Leblanc and Ball, 1978). All samples were dried under $N_2$ and quantitatively dissolved in ethyl acetate prior to isotope analysis.

*Isotope ratio mass spectrometry*

To measure the $\delta^{13}C$ value of bulk TOC, triplicate freeze-dried microbial mat samples were placed in silver capsules (Costech) and acidified with 100 µL of 1N HCl to remove dissolved inorganic

carbon. Samples were dried at 50ºC, enveloped in tin capsules (Costech), and analyzed on a Costech 4010 Elemental Analyzer connected to a Thermo Scientific Delta V IRMS.

FAME and derivatized sugar monomer $\delta^{13}C$ compositions were analyzed via gas chromatography– isotope ratio mass spectrometry (GC-IRMS; ThermoScientific Delta V Advantage connected to a Trace GC Ultra via a GC Isolink interface). In both cases, 1 µL of sample was co-injected with 0.5 µL of internal standard ($n$-C$_{32}$; 50 ng/µL). FAMEs were run on a 30 m × 0.25 mm HP- 5MS column as previously described (Close et al., 2014). Sugar monomers were run on a 30 m × 0.25 mm DB-1701 column; samples were transferred onto the column using a programmable temperature vaporizer (PTV) inlet at an injection temperature of 70ºC followed by 330ºC for 4 minutes. The GC-IRMS oven temperature gradient was adopted from van Dongen et al. (2001).

Stable carbon isotope analysis of P-SIF protein fractions was conducted using spooling-wire microcombustion (SWiM)-IRMS (Caimi, Richard and Brenna, 1993; Brand and Dobberstein, 1996). The SWiM-IRMS configuration used here is adapted from Sessions et al., (2005) and is detailed in Mohr et al., (2014). Fractions (96-well plate aliquots) were measured in triplicate. Only data from wells containing > 0.56 nmol C/µL (~350 mV peak amplitude, $m/z$ 44) and with measurement standard deviations < 2‰ were retained (Mohr et al., 2014).

*Data analysis*

Protein phylogenetic data were grouped in two different sets. First, by the most abundant phyla: Cyanobacteria, Proteobacteria and "remainder," which included all other annotated proteins. Given the relatively large proportion of Proteobacteria in LH47, we further subdivided the Proteobacteria into Deltaproteobacteria, Gammaproteobacteria, and remainder for the second set. The remaining data analysis is described in detail for the first set; methods were identical for

the second set but with two Proteobacterial variables instead of one. Estimates of the $\delta^{13}C$ compositions of proteins for each group ($\delta_{cyano}$, $\delta_{proteo}$, and $\delta_{rem}$, respectively) were calculated via multiple linear regression. Specifically, the $\delta^{13}C$ for a given well ($\delta_m$) and estimated fractional abundances (from summed protein peak areas) were used in two overdetermined linear equations:

$$\delta_{m,i} = (\textstyle\sum Prot_{,i} \times \delta_{cyano} + \sum Prot_{proteo,i} \times \delta_{proteo} + \sum prot_{rem,i} \times \delta_{rem}) / \sum Prot_{total} \tag{1}$$

$$\delta_{m,i}/\sigma_{m,i} = (\textstyle\sum Prot_{,i} \times \delta_{cyano} + \sum Prot_{proteo,i} \times \delta_{proteo} + \sum prot_{rem,i} \times \delta_{rem}) / (\sum Prot_{total} \times \sigma_{m,i}) \tag{2}$$

where $\delta_{cyano}$, $\delta_{proteo}$, and $\delta_{rem}$ are the unknowns, $\sum Prot$ is the sum of integrated peptide peak areas for proteins taxonomically assigned to the subscripted group, $\sigma_m$ is the precision for $\delta_m$ ($\pm$ 1 SD) and $i$ is each individual plate well for which both $\delta_m$ and peptide peak areas were measured (Mohr et al., 2014).

Equation 1 represents an unweighted estimate. Equation 2 is weighted by the precision of the isotopic measurements for each well (Glover et al., 2011). Equations 1 and 2 were solved inversely for $\delta_{cyano}$, $\delta_{proteo}$, and $\delta_{rem}$ by singular value decomposition using the built-in Matlab SVD function (Glover et al., 2011). Precision for this method is reported as $\pm$ the square root of the error variance.

**Results**

*Protein taxonomic identifications*

P-SIF analysis was conducted on sample LH47. SAX fractions 3-16 were chosen for further analysis using the integrated spectral absorbance of the RP-HPLC signal at 280 nm and criteria from prior work (Mohr et al., 2014). Six percent (43/672) of the RP-HPLC fractions contained identifiable peptide sequences, yielding 1188 unique microbial proteins (Table S1). Of these, 411

proteins were assigned to Cyanobacteria, 395 to Proteobacteria, 113 to Chloroflexi, 50 to Bacillariophyta, 16 to Bacteroidetes, 5 each to Spirochaetes and Thermotogae, and all others (193 proteins) to phyla containing 4 or fewer unique protein hits or to unclassified sources (Table S2). The mean and median number of unique peptides used to classify each protein was 2.7 and 2, respectively.

Table 3.1 contains our estimates for the relative abundance of microbial groups as determined via the summed integrated peak areas of peptides assigned to proteins. Mohr et al. (2014) calculated a $\pm$ 20% root-mean-square error for P-SIF abundance estimates by analysis of known mixtures of cultured organisms (Figure S12 from Mohr et al., 2014). Since our relative abundance estimates for Proteobacteria and Cyanobacteria overlap within error, we are unable to say definitively which of these two phyla dominated the microbial community at the time of sampling. However, we can compare our estimated relative abundances to the relative abundances of 16S rRNA gene sequences in a flat-mat sample collected from the sample location in August 2017 (the same month as our sampling): 16% Bacteroidetes, 29% Cyanobacteria, 42% Proteobacteria (28% Gammaproteobacteria, 9.9% Deltaproteobacteria, 1.3% Epsilonproteobacteria, and other Proteobacteria), 4.4% Chloroflexi, 2.0% Spirochaetes, 1.7% Verrucomicrobia, with the rest of the community composed of unclassifiable bacteria and other phyla each contributing <1% (Grim, 2019; Grim et al., *in prep*). As discussed for a previous P-SIF dataset (Gonzalez-Nayeck et al. 2022), label-free protein quantification methods are only semi-quantitative (Bubis et al., 2017). Nonetheless, these data support our estimate for an ecosystem dominated by both Proteobacteria and Cyanobacteria and suggest a relatively large contribution of Bacteriodetes to the "Remainder" category.

 *Protein carbon isotopic compositions*

Twenty percent (133/672) of the RP-HPLC fractions contained enough carbon for isotopic measurement. Of these, all but one had standard deviations < 2.0‰ and were used in subsequent analyses. The average standard deviation of triplicate $\delta^{13}$C measurements for protein fractions was 0.7‰ for the whole dataset, and 0.5‰ for the most abundant 50% as determined by IRMS peak area (Table S3). These values represent average measurement errors lower than the population standard deviation (1.6‰), indicating some degree of true variability among the protein $\delta^{13}$C values (Table S3). The resulting protein fraction $\delta^{13}$C values were not normally distributed (Shapiro-Wilk test, $p = 0.014$) with a mean of -24.6 ± 1.6‰ (Table 3.2; Figure S1). The data show a moderately negative skew (skewness = -0.63), indicating a small but statistically significant contribution of isotopically more negative proteins (Bulmer, 1979).

Six percent (43/672) of the RP-HPLC fractions contained sufficient carbon for both estimates of phylogenetic abundance (Table S1) and isotopic measurement (Table S4). These fractions were used in subsequent linear regression analyses.

*Estimates of protein $\delta^{13}$C values for microbial groups*

Microbial groups were defined in two different sets. The first set includes Cyanobacteria, Proteobacteria and all others grouped as "Remainder," and the second set further subdivides Proteobacteria into Deltaproteobacteria and Gammaproteobacteria, with other Proteobacteria included in "Remainder." Other taxonomic groups yielded too few assigned data points to resolve by mass-balance mixing approaches. The $\delta^{13}$C values of proteins originating from both sets of microbial groups were estimated via an abundance-weighted multiple linear regression and an unweighted multiple linear regression (Table 3.3, Figure 3.2). When Proteobacteria are grouped as one phylum, weighted protein estimates for Cyanobacteria and Proteobacteria are isotopically

indistinguishable at -24.3±0.4‰ and -24.4±0.3, respectively. When Proteobacteria are separated into the two dominant classes, weighted protein estimates for Cyanobacteria (-24.0±0.4‰) are isotopically more negative than Deltaproteobacteria (-23.3±0.6‰) but isotopically more positive than Gammaproteobacteria (-26.2±1.3‰). As noted above, we suspect that the estimated protein $\delta^{13}C$ values for "Remainder" contain a relatively large contribution from Bacteriodetes, the primary non-SRB heterotroph at MIS.

*Bulk, fatty acid and sugar carbon isotope ratios*

Total organic carbon (TOC) values in LH47 and LH22 were -26.9 ± 0.0‰; and -28.0 ± 0.2‰, respectively (Table 3.2). Both values are within error of the $\delta^{13}C$ composition (-28.1 ± 1.1‰) of a cyanobacterial mat sample taken in summer 2007 by Nold et al. (2013). The TOC value for is LH47 is 2.3 ± 1.6‰ depleted in $^{13}C$ relative to average protein.

The $\delta^{13}C$ value of DIC in the vent water above the mat was not measured at the time of sampling; furthermore, Nold et al. (2013) demonstrated that cyanobacterial mats at MIS utilize DIC from groundwater instead of lake water. The same authors measured groundwater DIC $\delta^{13}C$ composition at -6.0‰ ± 0.2‰ in summer 2007 (Nold et al., 2013). Assuming a substrate DIC $\delta^{13}C$ value of -6.0‰ for both samples, LH47 TOC is 20.9 ± 0.2‰ depleted in $^{13}C$ relative to DIC, and LH22 is 22.0 ± 0.2‰ depleted in $^{13}C$ relative to DIC.

Individual FAs *n*-C14:0, *n*-C15:0, *n*-C16:0, and *n*-C16:1, *n*-C17:0, *n*-C17:1, and *n*-C18x were the only measurable FAs recovered in both LH22 and LH47 (Table 3.4). The $\delta^{13}C$ values of these FAs ranged from 0.7 to 9.3‰ depleted in $^{13}C$ (LH47) and 1.2 to 6.3‰ depleted in $^{13}C$ (LH22) relative to TOC. Thus the weighted average fatty acids (FA, weighted by IRMS peak area) are 3.7 ± 0.7‰ depleted (LH47) and 3.6 ± 0.4‰ (LH22) in $^{13}C$ relative to TOC (Table 3.2).

Only LH22 had sufficient sample remaining after analyses to measure the $\delta^{13}C$ values of monosaccharide moieties extracted from EPS. Three monosaccharide sugars were present in sufficient abundance for isotopic measurement: glucose (-27.6 ± 1.5‰), xylose (-23.3 ± 1.2‰), and arabinose (-23.9 ± 0.3‰) (Table 3.2. Figure 3.2).

Throughout this work, the difference in $\delta^{13}C$ values (approximately 1.1‰; Table 3.2) between TOC, weighted average fatty acid, and individual fatty acids (Table 3.4) for both samples is consistent. This indicates that in both samples, relative isotopic differences between major carbon pools (e.g., protein, lipid and carbohydrate) are similar. As such, we can confidently assume that both the relative protein $\delta^{13}C$ values estimated for microbial groups (defined broadly, *i.e.*, at phylum-level or class-level instead of species level) and the relative sugar $\delta^{13}C$ values extracted from EPS are consistent for both samples, despite having each of these measurement types for only one sample. When comparing between samples (*e.g.,* LH22 sugars to LH47 proteins), we apply a ± 1.1‰ offset to the data.

**Table 3.1** Phylogenetic composition of sample LH47 as estimated by summed integrated peak area of peptides assigned to proteins.

| Phylum | Summed peak area of peptides | Abundance by peptide peak area (%) |
|---|---|---|
| Cyanobacteria | $1.10 \times 10^{10}$ | 29.1 |
| Proteobacteria | $1.75 \times 10^{10}$ | 46.3 |
| Chloroflexi | $1.79 \times 10^{9}$ | 4.7 |
| Bacillariophyta | $7.39 \times 10^{8}$ | 2.0 |
| Bacteroidetes | $1.23 \times 10^{9}$ | 3.2 |
| Spirochaetes | $8.35 \times 10^{7}$ | 0.2 |
| Thermotogae | $1.45 \times 10^{8}$ | 0.4 |
| Acidobacteria | $5.87 \times 10^{6}$ | 0.0 |
| Chlorobi | $5.61 \times 10^{6}$ | 0.0 |
| Unclassified | $5.33 \times 10^{9}$ | 14.1 |

Within Proteobacteria

| Class | Summed peak area of peptides | Abundance by peptide peak area (%) |
|---|---|---|
| Alpha | $1.93 \times 10^{8}$ | 1.1 |
| Beta | $6.39 \times 10^{7}$ | 0.4 |
| Delta | $1.15 \times 10^{10}$ | 65.4 |
| Epsilon | $2.22 \times 10^{7}$ | 0.1 |
| Gamma | $4.82 \times 10^{9}$ | 27.5 |
| Other | $9.68 \times 10^{8}$ | 5.5 |

**Table 3.2** Summary of $\delta^{13}C$ values for MIS microbial mat samples. Weighted average FA data were weighted by IRMS peak area.

| | $\delta^{13}C$ (‰)- LH47 | $\delta^{13}C$ (‰)- LH22 | Difference |
|---|---|---|---|
| Total Organic Carbon | $-26.9 \pm 0.0$ | $-28.0 \pm 0.2$ | $-1.1 \pm 0.2$ |
| Weighted Average Fatty Acid | $-30.6 \pm 0.7$ | $-31.6 \pm 0.4$ | $-1.1 \pm 0.8$ |
| Average Protein | $-24.6 \pm 1.6$ | n.m. | |
| EPS Glucose | n.m.; estimated $-26.5 \pm 1.7$ | $-27.6 \pm 1.5$ | |
| EPS Xylose | n.m.; estimated $-22.2 \pm 1.4$ | $-23.3 \pm 1.2$ | |
| EPS Arabinose | n.m.; estimated $-22.8 \pm 0.9$ | $-23.9 \pm 0.3$ | |

**Table 3.3** Estimates of the $\delta^{13}C$ values (‰) of proteins for microbial groups from MIS mat sample LH47 as calculated from P-SIF data using two different subsets of phylogenetic groups.

| Microbial Group (LH47 only) | Weighted Linear Regression | Unweighted Linear Regression |
|---|---|---|
| Proteobacteria Grouped by Phylum | | |
| Cyanobacteria | -24.3±0.4 | -25.0±1.2 |
| Proteobacteria | -24.4±0.3 | -24.4±0.6 |
| Other | -23.2±0.4 | -24.2±0.8 |
| Proteobacteria Grouped by Class | | |
| Cyanobacteria | -24.0±0.4 | -24.4±1.1 |
| Gammaproteobacteria | -26.2±1.3 | -27.5±1.7 |
| Deltaproteobacteria | -23.3±0.6 | -23.6±0.8 |
| Other | -24.0±0.3 | -24.1±0.6 |

**Table 3.4** $\delta^{13}C$ compositions of individual FAs extracted from LH47 and LH22, and the difference between the two sets of data.

| FA | $\delta^{13}C$ (‰)-LH47 | Abundance Relative to 16:0- LH47 | $\delta^{13}C$ (‰)-LH22 | Abundance Relative to 16:0- LH22 | Difference $\delta^{13}C$ |
|---|---|---|---|---|---|
| $n$-$C_{14:0}$ | -32.3 ± 1.6 | 0.24 | -34.0±0.0 | 0.28 | -1.7±1.6 |
| $n$-$C_{15:0}$ | -28.1 ± 1.9 | 0.46 | -29.3±0.6 | 0.82 | -1.3±2.5 |
| $n$-$C_{16:1}$ | -30.0 ± 0.2 | 1.10 | -31.9±1.0 | 1.22 | -1.9±1.2 |
| $n$-$C_{16:0}$ | -31.2 ± 0.2 | 1.00 | -32.2±0.1 | 1.00 | -0.9±0.3 |
| $n$-$C_{17:1}$ | -27.6 ± 1.2 | 0.53 | -29.2±0.2 | 0.64 | -1.5±1.4 |
| $n$-$C_{17:0}$ | -33.1 ± 1.3 | 0.27 | -32.7±0.0 | 0.39 | 0.4±1.4 |
| $n$-$C_{18:x}$ | -36.2 ± 0.6 | 0.32 | -34.3±0.2 | 0.45 | 1.9±0.8 |
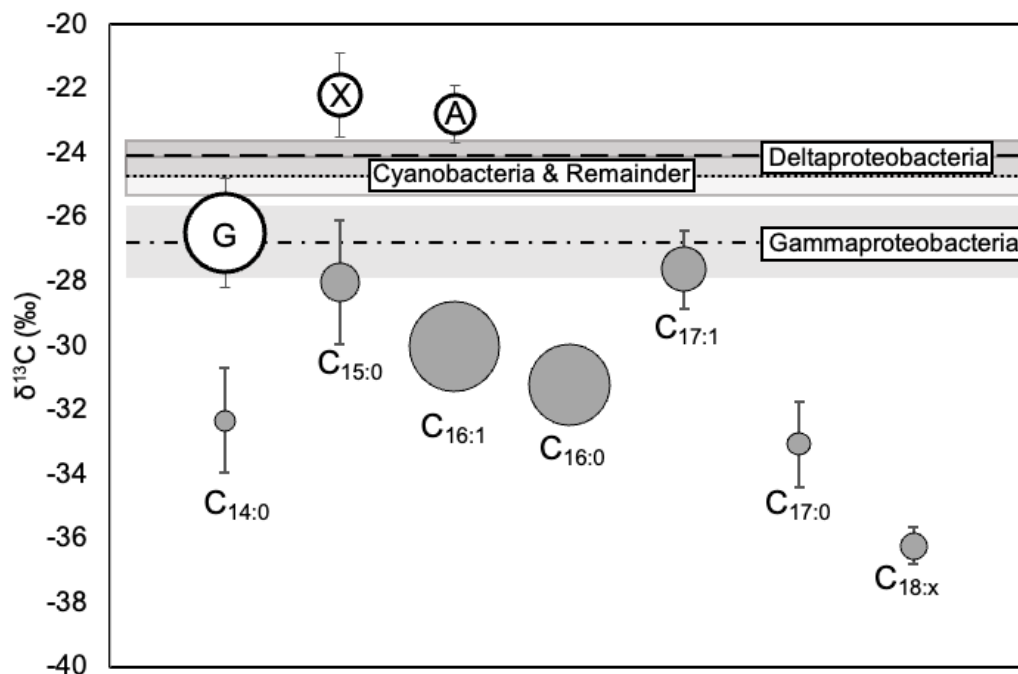
**Figure 3.2** A composite of MIS mat carbon isotopic data for sample LH47. Deltaproteobacteria, Cyanobacteria & Remainder, and Gammaproteobacteria protein $\delta^{13}C$ values are represented by dashed, dotted and dashed and dotted lines, respectively. The $\delta^{13}C$ values of individual FA are indicated by gray circles; circle area corresponds to abundance relative to the $n$-$C_{16:0}$ FA. The $\delta^{13}C$ estimates for individual glucose (G), xylose (X) and arabinose (A) moieties from extracted EPS are indicated by white circles; circle area corresponds to abundance relative to glucose moieties. Monosaccharide $\delta^{13}C$ estimates were generated by adding 1.1‰ to LH22 data, based on consistency in FA $\delta^{13}C$ patterns (Table 3.3). Shading and error bars represent $\pm$ 1 SD from the mean.
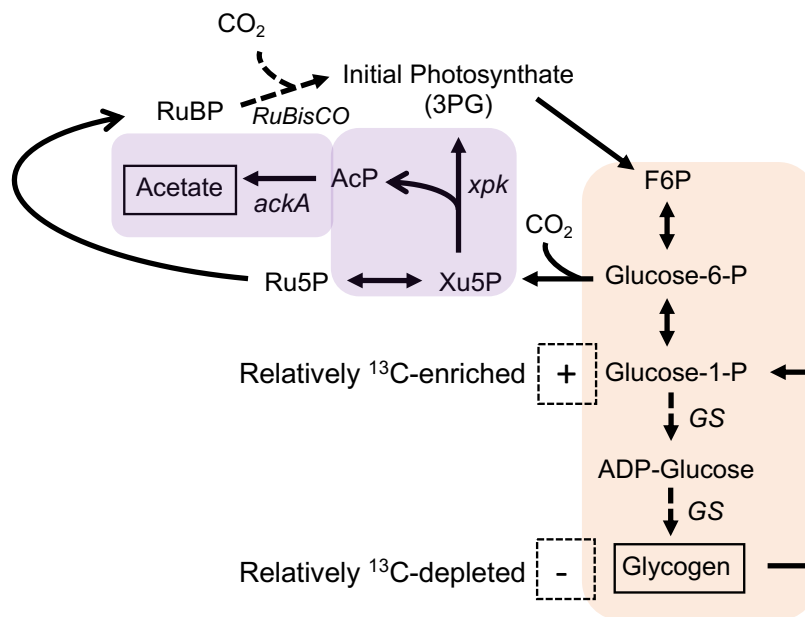
**Figure 3.3** Partial pathway for cyanobacterial glycogen synthesis and catabolism (orange shading) and the cyanobacterial phosphoketalase pathway (purple shading). Broken arrows represent reactions with confirmed or suspected carbon isotopic fractionation. Products excreted extracellularly are indicated by boxes. Enzymes are italicized. *ackA*, acetate kinase; *xpk*, phosphoketolase; *RuBisCO*, Ribulose1,5-bisphosphate carboxylase/oxygenase; *GS*, glycogen synthase. Glucose-1-P, glucose-1-phosphate; Glucose-6-P, glucose-6-phosphate; F6P, fructose-6-phosphate; Ru5P, ribulose-5-phosphate; RuBP, ribulose 1,5-bisphosphate; Xu5P, xylulose-5-phosphate; 3PG, 3-phosphoglycerate; AcP, acetylphosphate.

**Discussion**

Previous research at MIS has provided insight into the composition of the microbial community, its metabolic potential and activity, and the geochemical consequences of that activity (Biddanda et al., 2009; Nold et al., 2010; Voorhies et al., 2012; Medina, 2017; Klatt et al., 2020; Grim et al., 2021). Here, we further contextualize these observations by estimating the biomass $\delta^{13}C$ values of key microbial groups in the MIS ecosystem: Cyanobacteria, sulfur reducing Deltaproteobacteria (SRB), and sulfur oxidizing Gammaproteobacteria (SOB). Through P-SIF, we estimate that SRB are $^{13}C$-enriched relative to Cyanobacteria, while SOB are $^{13}C$-depleted. Below we discuss possible anabolic carbon sources for each group of organisms and their potential isotopic contributions to net microbial biomass.

*MIS Cyanobacteria and Gammaproteobacteria fix inorganic carbon via the Calvin-Benson-Bassham cycle*

Previously, MIS mats incubated with ^{14}C-labeled bicarbonate under low oxygen conditions were found to assimilate carbon via both anoxygenic photosynthesis and chemosynthesis in equal proportions, and via oxygenic photosynthesis when the sediments were suspended in oxygenated groundwater (Voorhies et al., 2012). Subsequently, metagenome-assembled genomes and metatranscriptomes for the dominant organisms at MIS suggested that photosynthetic Cyanobacteria and chemosynthetic Gammaproteobacteria are responsible for the bulk of inorganic carbon fixation via the Calvin-Benson-Bassham (CBB; reductive pentose phosphate) cycle (Grim et al., 2021). Cyanobacterial and Gammaproteobacterial proteins in the LH47 sample are approximately 18 and 20 ‰ depleted in ^{13}C relative to assumed DIC (-6‰), respectively. These values are within the reported range for both cultured Cyanobacteria fixing carbon via the CBB cycle and the *in vitro* fractionation for Proteobacterial RuBisCO enzyme (House et al., 200), and do not suggest DIC limitation as has been seen for other mat systems where the difference between biomass and DIC $\delta^{13}$C values is relatively small (Schouten et al., 2001).

*Cyanobacteria excrete ^{13}C-depleted EPS sugars during the day and ^{13}C-enriched acetate via the phosphoketalase pathway at night*

In microbial mats, heterotrophs and mixotrophs assimilate the metabolic by-products of autotrophic metabolism (Prieto-Barajas et al., 2018). Common sources of anabolic organic carbon in mats include storage sugars excreted extracellularly (EPS; *e.g.*, Nold and Ward, 1996), glycolate excreted during photorespiration (e.g., Bateson and Ward, 1988), and the products of overnight fermentation of storage sugars (e.g., van der Meer et al., 2005). Previously, we conducted P-SIF on a sub-aerial, highly oxygenated cyanobacterial mat and measured the $\delta^{13}$C of the

63

monosaccharide moieties of EPS extracted from the same mat (Gonzalez-Nayeck et al. 2022). We found that glucose, the only quantitatively important monosaccharide, was isotopically indistinguishable from both cyanobacterial and heterotrophic proteins. As such, we hypothesized that heterotrophic organisms in this oxygenated surface environment were assimilating cyanobacterial photosynthetic sugars, i.e., EPS.

In contrast, EPS extracted from MIS sample LH22 had three quantitatively important monosaccharides: glucose, xylose and arabinose. Extracted glucose was approximately 4‰ depleted in $^{13}C$ relative to both xylose and arabinose (Figure 3.3); furthermore, after normalizing the samples to each other (1.1‰ offset correction; see results section), the glucose is depleted and the xylose and arabinose are enriched in $^{13}C$ relative to cyanobacterial proteins. Both the offset between pentose and hexose sugars (~ 4 ‰) and relative ordering ( glucose < biomass < arabinose = xylose) agrees with prior $\delta^{13}C$ measurements of cell-associated monosaccharide sugars from a cultured freshwater Cyanobacterium (Teece and Fogel, 2007).

We propose that the isotopic ordering in monosaccharide sugars observed in MIS and by Teece and Fogel (2007) is due to carbon isotope fractionation during glycogen synthesis by Cyanobacteria (Figure 3.3. To our knowledge, there is no direct *in vitro* or *in vivo* evidence for an isotope effect during the polymerization of internal sugars to glycogen. However, indirect evidence supporting this idea includes differences in the $\delta^{13}C$ compositions of internal hexose monomers in Cyanobacteria (Teece and Fogel, 2007; Pereira et al., 2009), and an observation on a natural cyanobacterial mat that EPS is relatively $^{13}C$-depleted compared to TOC (Wieland et al., 2008). We assume that cyanobacterial xylose is isotopically similar to internal glucose, since pentose sugars are derived from the decarboxylation of internal hexose sugars (Figure 3.3; (White, 2000). If this assumption is correct, xylose and arabinose may be isotopically enriched relative to EPS

and external glucose because they are derived from a pool of residual internal glucose that is isotopically enriched due to fractionation during glycogen synthesis.

Glycolate excreted during cyanobacterial photorespiration is another potential source of organic carbon for heterotrophic organisms in microbial mats (Bateson and Ward, 1988). Since glycolate is derived from the RuBisCO substrate ribulose 1,5-bisphosphate, it should have the same $\delta^{13}C$ composition as initial photosynthate (White 2001). However, it is unlikely that glycolate is a quantitatively important source of organic carbon at MIS. Cyanobacteria possess the pathway to detoxify glycolate intracellularly and typically only excrete glycolate extracellularly under nitrogen limitation (Renström-Kellner and Bergman, 1989; Eisenhut et al., 2008). MIS sediment organic matter has a greater percentage of nitrogen than average Lake Huron sediments (Rico and Sheldon, 2019), and evidence for transcriptionally active nitrogen-fixation genes in the MIS mat suggests the microbial community can compensate during periods of low nitrogen availability (Grim et al. 2021). Furthermore, MIS is low oxygen, which makes it unlikely that the oxygenase reaction of RuBisCO (i.e., photorespiration) is quantitatively important (Warburg, 1928; Busch, 2020).

A key implication of this internal $^{13}C$ sorting is the likelihood that Cyanobacteria excrete $^{13}C$-enriched acetate produced by the phosphoketolase pathway. Cyanobacteria meet their maintenance energy requirements at night via either respiration (in oxic environments) or fermentation of storage sugars (in anoxic environments) (Stal 2012). Previously, it was believed that acetate-yielding cyanobacterial fermentation proceeded through the decarboxylation of pyruvate to acetyl-CoA (Stal and Moezelaar, 1997). However, recent work (Xiong et al., 2015; Chuang and Liao, 2021) has suggested the primary route instead may be via $C_5$ sugars.

Transformation of $^{13}C$-enriched xylose via the phosphoketolase pathway would result nighttime excretion of of $^{13}C$-enriched acetate (Figure 3.3), primarily by fermenting Cyanobacteria.

Phosphoketolases are glycolytic enzymes that convert either xylulose-5-phosphate and/or fructose-6-phosphate into acetyl-phosphate and glyceraldehyde-3-phosphate and/or acetyl-phosphate and erythrose-4-phosphate. In Cyanobacteria the genes encoding the enzymes are often found in genomic proximity with genes encoding acetate kinases (which convert acetyl-phosphate into acetate; (Sánchez et al., 2010). Experiments on the model Cyanobacteria *Synechococcus elongatus* PCC7942 (Chuang and Liao, 2021) and *Synechocystis sp.* PCC 6803 (Xiong et al., 2015) demonstrated that deleting the genes encoding the phosphoketolase enzyme reduced viability and prevented the excretion of acetate during both dark anaerobic conditions (*Synechococcus*) and dark aerobic conditions after being grown heterotrophically (*Synechocystis*). Phosphoketolase genes are found in most cyanobacterial genomes, and their expression peaks at sunset (Chuang and Liao, 2021). These lines of evidence suggest that some (if not most) Cyanobacteria meet their night-time maintenance energy requirements via phosphoketolase degradation of $C_5$ sugars rather than by decarboxylation of pyruvate. A metagenome-assembled genome from MIS for *Phormidium*, the dominant Cyanobacterial genus in sample LH22, contains both pyruvate:ferredoxin oxidoreductase and acetate kinase genes which previously were assumed to aid in night-time fermentation via pyruvate (Voorhies et al., 2012). Alternatively, here we propose that the pyruvate:ferredoxin oxidoreductase in *Phormidium* may primarily be associated with another function, either nitrogen fixation (Bothe et al., 2010) or aerobic photomixotrophic growth (Klatt et al., 2021; Wang et al., 2022). To investigate whether cyanobacterial phosphoketolases are present at MIS, we conducted a BLAST search against the same metagenome used for P-SIF (IMG taxon object ID: 3300028549) using protein sequence slr0453,

a putative phosphoketolase from Synechocystis sp. PCC 6803 (Xiong et al., 2015). This search yielded 77 putative proteins with at least 50% shared amino acid identities to the query sequence, including 12 proteins assigned to scaffolds within cyanobacterial bins. Furthermore, 11 proteins (including 3 within cyanobacterial bins) shared at least 70% shared amino acid identities to sequence slr0453, suggesting cyanobacterial phosphoketolase enzymes may indeed be present in MIS. In summary, we propose that cyanobacterial exudates at MIS are represented by two endmembers: relatively $^{13}$C-depleted glycogen from EPS, and relatively $^{13}$C-enriched acetate from overnight fermentation via the phosphoketolase pathway.

*Gammaproteobacteria at MIS likely conduct more inorganic carbon fixation than is expected for freshwater species*

The dominant Gammaproteobacteria identified at MIS via 16S rRNA abundances in August 2017 represented the sulfur-oxidizing bacterial (SOB) genus *Beggiatoa* (Grim et al., 2021), Grim et al. *in prep*). Freshwater *Beggiatoa* typically are heterotrophic or mixotrophic despite having a functional CBB cycle; they grow most readily in culture when supplied with acetate (Garrity et al., 2005). Most, but not all, strains described to date are incapable of growing in the laboratory with monosaccharide sugars as the sole carbon source (Dubinina et al., 2017). Even when growing mixotrophically, it is expected that inorganic carbon fixation accounts for a relatively small proportion (<10%) of total biomass production in freshwater *Beggiatoa* (Garrity et al., 2005). However, *Beggiatoa* from the Isolated Sinkhole in Lake Huron are phylogenetically nested between marine and freshwater strains and are capable of $H_2$-based lithoautotrophy, heterotrophy, and possibly mixotrophy (Sharrar et al., 2017). Based on this context, plus the detection of $HCO_3^-$ fixation (Voorhies et al. 2012), and the inconsistency between the measured $\delta^{13}$C composition of gammaproteobacterial protein and the expected $\delta^{13}$C value of acetate excreted by Cyanobacteria

(discussed above), we propose that SOB at MIS obtain a relatively larger proportion of their biomass carbon from inorganic carbon fixation than would be expected for freshwater Gammaproteobacteria growing with acetate as anabolic carbon source. An alternative or additional explanation for the difference between presumed acetate and SOB protein $\delta^{13}C$ values may be that they assimilate EPS sugars in preference to acetate, given the similarity between monosaccharide and gammaproteobacterial protein $\delta^{13}C$ values. If Gammaproteobacteria in MIS are indeed assimilating acetate as an anabolic carbon source, the acetate is either from a source other than cyanobacterial excretion, is assimilated in relatively small proportions, or our hypothesis of relatively $^{13}C$ enriched acetate (discussed above) is incorrect.

*Sulfate-reducing bacteria at MIS assimilate cyanobacterial acetate, while other heterotrophic bacteria assimilate EPS and/or the products of viral lysis*

Based on the similarity between the $\delta^{13}C$ composition of SRB proteins and xylose monomers, it appears likely that SRB are assimilating acetate generated by the cyanobacterial phosphoketolase pathway as their primary anabolic carbon source. Previously, it was reported that SRB from MIS expressed genes for the autotrophic Wood-Ljungdhal (acetyl-CoA) pathway that reduces $CO_2$ with $H_2$ as electron donor and produces acetyl-CoA, which is then the primary metabolite for biomass synthesis (Wood et al., 1986; Grim et al., 2021). However, SRB also are known to use the Wood-Ljungdahl pathway in reverse both to assimilate acetate for growth and for respiratory oxidation to $CO_2$ (Spormann and Thauer, 1988). This suggests the expression of Wood-Ljungdahl genes may have been detected because the pathway is being used for acetate uptake. Furthermore, prior research on carbon isotopic fractionation during lithoautotrophic growth by SRB using the Wood-Ljungdahl pathway (acetate-generating direction) suggests the resulting biomass should be $\sim 11$ ‰ depleted in $^{13}C$ relative to substrate $CO_2$ (Londry and des Marais, 2003). Here this would

yield biomass ~ -17 ‰, which is approximately 6 ‰ more positive than the measured SRB proteins (-23.3 ‰).

Although the proteins comprising the "Remainder" fraction could not be confidently assigned to any one taxonomic group, evidence from 16S rRNA analyses in the same location from August 2017 suggests that Bacteroidetes comprise a relatively large (~15%) proportion of the heterotrophic population (Grim et al. 2021, Grim et al. *in prep*). Proteins in the "Remainder" fraction were isotopically indistinguishable from Cyanobacterial proteins. Since heterotrophic metabolism does not impart significant $^{13}C$ fractionation in bacteria (DeNiro and Epstein, 1977; Blair et al., 1985), we can infer that the anabolic carbon source for Bacteroidetes is similar in $\delta^{13}C$ composition to the Cyanobacterial biomass. Bacteroidetes are broadly known for their ability to degrade a variety of complex polysaccharides (McKee et al., 2021), making EPS their most likely source of organic carbon. If we assume xylose, arabinose and glucose comprise the majority of EPS at MIS, the $\delta^{13}C$ of EPS as calculated by the abundance-weighted (IRMS peak areas) average of the three monomers is -24.5 ± 2.3 ‰.  This value is statistically indistinguishable from both Cyanobacterial and "Remainder" biomass, potentially indicating that Bacteroidetes at MIS are assimilating hydrolyzed EPS indiscriminately (i.e., all monosaccharide moieties), and/or the average products of total organic carbon from viral lysis of Cyanobacteria (Voorhies et al., 2016).

*Heterogeneous $\delta^{13}C$ of photosynthetic exudates contributes to isotopic variability in ancient microbial mat facies*

The difference in $\delta^{13}C$ composition between carbonate and syn-depositional organic matter ($\Delta\delta^{13}C$) has been remarkably consistent throughout the Phanerozoic (Krissansen-Totton et al., 2015). The relatively variable amplitudes in $\Delta\delta^{13}C$ from Precambrian facies has been interpreted

as reflecting local rather than global $\delta^{13}C$ signatures, specifically a greater contribution by microbial mats (Blumenberg et al., 2012; Schobben and van de Schootbrugge, 2019; Fox et al., 2020; Nelson et al., 2021). In contrast to our prior work on an oxygenated microbial mat (Gonzalez-Nayeck et al., 2022), suspected autotrophic and heterotrophic organisms at MIS had heterogeneous biomass $\delta^{13}C$ compositions, which we attribute to the heterogeneity in $\delta^{13}C$ composition between different cyanobacterial exudates. These studies encompass only two specific modern environments, and significantly more work is necessary before generalizing any observed trends. However, our work underscores the importance of considering local redox conditions when evaluating inter-species $\delta^{13}C$ signatures in the rock record, whether it be via comparing the isotopic compositions of lipid biomarkers for distinct groups of organisms (e.g., Blumenberg et al., 2012) or between individual microfossils (e.g., Peng et al., 2016; Schopf et al., 2018; Osterhout et al., 2021).

In a highly photic, oxygenated and relatively nutrient-limited surface microbial mat from Yellowstone National Park (YNP), we found that EPS sugars were indistinguishable from Cyanobacterial proteins, likely because Cyanobacteria in such environments allocate the majority of their initial photosynthate to glycogen excreted extracellularly (Nold and Ward, 1996; Gonzalez-Nayeck et al., 2022). If there is an isotopic fractionation during the polymerization of glycogen (the main component of cyanobacterial EPS), it is minimally expressed because the majority of the initial substrate has been converted to product (Hayes, 2001). Alternatively, Cyanobacteria in oxygenated environments might excrete large proportions of fixed carbon as glycolate due to the oxygenase reaction of RuBisCO (Bateson and Ward, 1988) and the nitrogen requirements of the glycolate detoxification pathway (Renström-Kellner and Bergman, 1989; Eisenhut et al., 2008). However, glycolate derives directly from cyanobacterial initial

photosynthate, so presumably glycolate-based heterotrophy would also result in biomass isotopically indistinguishable from this source.

In short, in an oxygenated environment with relatively high photon flux (e.g., YNP), we would predict that the biomass of Cyanobacteria, their net exudates, and therefore the total average composition of aerobic heterotrophic organisms, all will be isotopically homogenous. In contrast, MIS has relatively lower photon flux, is low oxygen and relatively higher in nutrients. As such, Cyanobacteria presumably can detoxify internal glycolate, and due to the demands of diel redox fluctuation may not allocate the majority of their initial photosynthate to EPS. Instead, Cyanobacteria at MIS may ferment a significant portion of their internal sugar, resulting in both (1) an expression of the suspected isotopic fractionation associated with glycogen synthesis, resulting in $^{13}$C-depleted EPS relative to intracellular sugars; and (2) $^{13}$C-enriched acetate derived from this relatively enriched internal pool. We suspect that at MIS, cyanobacterial exudates are isotopically heterogeneous as a direct consequence of night-time anoxia. Future work to directly measure the $\delta^{13}$C compositions of cyanobacterial metabolites, including acetate and glycolate, produced under different environmental conditions, would help test this hypothesis.

**Conclusions**

In this work we determined the protein $\delta^{13}$C values of autotrophic, heterotrophic, and mixotrophic organisms, and the $\delta^{13}$C values of the sugar moieties of EPS, all from flat mat samples from Middle Island Sinkhole in Lake Huron, a low oxygen Proterozoic analogue environment. Our results show that Cyanobacteria (autotrophs) are isotopically distinct from both sulfate reducing bacteria (heterotrophs) and sulfate oxidizing bacteria (autotrophs or mixotrophs), and that sugar moieties extracted from EPS are equally isotopically heterogeneous. We hypothesize that the observed $\delta^{13}$C

patterns in proteins from different bacterial groups are due to isotopic heterogeneity in cyanobacterial exudates, and that these patterns are common within benthic microbial mats that experience night-time anoxia. Proterozoic microbial mat facies often show greater variability in

$\delta^{13}C$ values internally than Phanerozoic facies, and this may partially be due to isotopically heterogeneous heterotrophic carbon sources, particularly cyanobacterial exudates.

**Acknowledgements**

**References**

Abelson P. H. and Hoering T. C. (1961) Carbon isotope fractionation in formation of amino acids by photosynthetic organisms. *Proc Natl Acad Sci U S A* **47**, 623–632.

Bateson M. M. and Ward D. M. (1988) Photoexcretion and fate of glycolate in a hot spring cyanobacterial mat. *Appl Environ Microbiol* **54**, 1738–43.

Biddanda B. A., Nold S. C., Ruberg S. A., Kendall S. T., Sanders T. G. and Gray J. J. (2009) Great Lakes Sinkholes: A Microbiogeochemical Frontier. *Eos, Transactions American Geophysical Union* **90**, 61–62.

Blair N., Leu a, Muñoz E., Olsen J., Kwong E. and des Marais D. (1985) Carbon isotopic fractionation in heterotrophic microbial metabolism. *Appl Environ Microbiol* **50**, 996–1001.

Blumenberg M., Thiel V., Riegel W., Kah L. C. and Reitner J. (2012) Biomarkers of black shales formed by microbial mats, Late Mesoproterozoic (1.1 Ga) Taoudeni Basin, Mauritania. *Precambrian Res* **196–197**, 113–127.

Bothe H., Schmitz O., Yates M. G. and Newton W. E. (2010) Nitrogen Fixation and Hydrogen Metabolism in Cyanobacteria. *Microbiology and Molecular Biology Reviews* **74**, 529–551.

Brand W. A. and Dobberstein P. (1996) Isotope-ratio-monitoring liquid chromatography mass spectrometry (IRM-LCMS): First results from a moving wire interface system. *Isotopes Environ Health Stud* **32**, 275–283.

Bubis J. A., Levitsky L. I., Ivanov M. v., Tarasova I. A. and Gorshkov M. v. (2017) Comparative evaluation of label-free quantification methods for shotgun proteomics. *Rapid Communications in Mass Spectrometry* **31**, 606–612.

Bulmer M. (1979) *Principles of statistics*.

Busch F. A. (2020) Photorespiration in the context of Rubisco biochemistry, $CO_2$ diffusion and metabolism. *The Plant Journal* **101**, 919–939.

Caimi, Richard and Brenna T. (1993) High-precision liquid chromatography-combustion isotope ratio mass spectrometry. *Anal Chem* **65**, 3497–3500.

Callow R. H. T. and Brasier M. D. (2009) Remarkable preservation of microbial mats in Neoproterozoic siliciclastic settings: Implications for Ediacaran taphonomic models. *Earth Sci Rev* **96**, 207–219.

Cano M., Holland S. C., Artier J., Burnap R. L., Ghirardi M., Morgan J. A. and Yu J. (2018) Glycogen Synthesis and Metabolite Overflow Contribute to Energy Balancing in Cyanobacteria. *Cell Rep*.

Chuang D. S. W. and Liao J. C. (2021) Role of cyanobacterial phosphoketolase in energy regulation and glucose secretion under dark anaerobic and osmotic stress conditions. *Metab Eng* **65**, 255–262.

Close H. G., Wakeham S. G. and Pearson A. (2014) Lipid and13C signatures of submicron and suspended particulate organic matter in the Eastern Tropical North Pacific: Implications for the contribution of Bacteria. *Deep Sea Res 1 Oceanogr Res Pap* **85**, 15–34.

DeNiro M. J. and Epstein S. (1977) Mechanism of Carbon Isotope Fractionation Associated with Lipid Synthesis. *Science,* **197**(4300)*,* 261-263.

van Dongen B. E., Schouten S. and Damsté J. S. S. (2001) Gas chromatography/combustion/isotope-ratio-monitoring mass spectrometric analysis of methylboronic derivatives of monosaccharides: A new method for determining natural 13C abundances of carbohydrates. *Rapid Communications in Mass Spectrometry* **15**, 496–500.

Dubinina G., Savvichev A., Orlova M., Gavrish E., Verbarg S. and Grabovich M. (2017) Beggiatoa leptomitoformis sp. Nov., the first freshwater member of the genus capable of chemolithoautotrophic growth. *Int J Syst Evol Microbiol* **67**, 197–204.

Eisenhut M., Ruth W., Haimovich M., Bauwe H., Kaplan A. and Hagemann M. (2008) The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proc Natl Acad Sci U S A* **105**, 17199–17204.

Fox C. P., Cui X., Whiteside J. H., Olsen P. E., Summons R. E. and Grice K. (2020) Molecular and isotopic evidence reveals the end-Triassic carbon isotope excursion is not from massive exogenous light carbon. *Proc Natl Acad Sci U S A* **117**, 30171–30178.

Garrity G., Bell J. A. and Lilburn T. (2005) *Bergey s manual of systematic bacteriology: the proteobacteria; part B: the gammaproteobacteria*.

Gehling J. G. (1999) Microbial mats in terminal Proterozoic siliciclastics; Ediacaran death masks. *Palaios* **14**, 40–57.

Gingras M., Hagadorn J. W., Seilacher A., Lalonde S. v., Pecoits E., Petrash D. and Konhauser K. O. (2011) Possible evolution of mobile animals in association with microbial mats. *Nat Geosci* **4**, 372–375.

Glover D. M., Jenkins W. J. and Doney S. C. (2011) *Modeling methods for marine science*.,

Gonzalez-Nayeck A. C., Mohr W., Tang T., Sattin S., Parenteau M. N., Jahnke L. L. and Pearson A. (2022) Absence of canonical trophic levels in a microbial mat. *Geobiology* **20**, 726–740.

Grim S. (2019) Genomic and Functional Investigations Into Seasonally-Impacted and Morphologically-Distinct Anoxygenic Photosynthetic Cyanobacterial Mats. University of Michigan.

Grim S. L., Voorhies A. A., Biddanda B. A., Jain S., Nold S. C., Green R. and Dick G. J. (2021) Omics-Inferred Partitioning and Expression of Diverse Biogeochemical Functions in a Low O2 Cyanobacterial Mat Community . *mSystems* **6**.

Hagadorn J. W. and Bottjer D. J. (1997) Wrinkle structures: Microbially mediated sedimentary structures common in subtidal siliciclastic settings at the Proterozoic-Phanerozoic transition. *Geology* **25**, 1047–1050.

Hanson T. E., Luther G. W., Findlay A. J., MacDonald D. J. and Hess D. (2013) Phototrophic sulfide oxidation: Environmental insights and a method for kinetic analysis. *Front Microbiol* **4**, 382.

Hayes J. M. (2001) Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes. *Rev Mineral Geochem* **43**, 225–277.

House C. H., Schopf J. W. and Stetter K. O. Carbon isotopic fractionation by Archaeans and other thermophilic prokaryotes.

Javaux E. J. (2019) Challenges in evidencing the earliest traces of life. *Nature 2019 572:7770* **572**, 451–460.

Johnston D. T., Wolfe-Simon F., Pearson A. and Knoll A. H. (2009) Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proceedings of the National Academy of Sciences* **106**, 16925–16929.

Klatt J. M., Chennu A., Arbic B. K., Biddanda B. A. and Dick G. J. (2021) Possible link between Earth's rotation rate and oxygenation. *Nature Geoscience 2021 14:8* **14**, 564–570.

Klatt J. M., Gomez-Saez G. v., Meyer S., Ristova P. P., Yilmaz P., Granitsiotis M. S., Macalady J. L., Lavik G., Polerecky L. and Bühring S. I. (2020) Versatile cyanobacteria control the timing and extent of sulfide production in a Proterozoic analog microbial mat. *ISME Journal* **14**, 3024–3037.

Klock J. H., Wieland A., Seifert R. and Michaelis W. (2007) Extracellular polymeric substances (EPS) from cyanobacterial mats: Characterisation and isolation method optimisation. *Mar Biol* **152**, 1077–1085.

Knoll A. H. and Nowak M. A. (2017) The timetable of evolution. Science advances, **3**(5), e1603076.

Krissansen-Totton J., Buick R. and Catling D. C. (2015) A statistical analysis of the carbon isotope record from the Archean to phanerozoic and implications for the rise of oxygen. *Am J Sci* **315**, 275–316.

Leblanc D. J. and Ball A. J. S. (1978) A fast one-step method for the silylation of sugars and sugar phosphates. *Anal Biochem* **84**, 574–578.

Londry K. L. and des Marais D. J. (2003) Stable carbon isotope fractionation by sulfate-reducing bacteria. *Appl Environ Microbiol* **69**, 2942–9.

López-García P. and Moreira D. (2020) The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nature Microbiology 2020 5:5* **5**, 655–667.

McKee L. S., la Rosa S. L., Westereng B., Eijsink V. G., Pope P. B. and Larsbrink J. (2021) Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature. *Environ Microbiol Rep* **13**, 559–581.

Medina M. J. (2017) Genomic and transcriptomic evidence for niche partitioning among sulfatereducing bacteria in redox-stratified cyanobacterial mats of the Middle Island Sinkhole.

Meer M. T. J. van der, Schouten S., Mary M., Nübel U., Wieland A., Kühl M., Leeuw W. de, Damsté J. S. S., Ward D. M., Bateson M. M., Nu U., Ku M., Leeuw J. W. de and Damste J. S. S. (2005) Diel Variations in Carbon Metabolism by Green Nonsulfur-Like Bacteria in Alkaline Siliceous Hot Spring Microbial Mats from Yellowstone National Park. *American Society for Microbiology* **71**, 3978–3986.

Mohr W., Tang T., Sattin S. R., Bovee R. J. and Pearson A. (2014) Protein stable isotope fingerprinting: Multidimensional protein chromatography coupled to stable isotope-ratio mass spectrometry. *Anal Chem* **86**, 8514–8520.

Nelson L. L., Ahm A. S. C., Macdonald F. A., Higgins J. A. and Smith E. F. (2021) Fingerprinting local controls on the Neoproterozoic carbon cycle with the isotopic record of Cryogenian carbonates in the Panamint Range, California. *Earth Planet Sci Lett* **566**, 116956.

Nold S. C., Bellecourt M. J., Kendall S. T., Ruberg S. A., Sanders T. G., Klump J. V. and Biddanda B. A. (2013) Underwater sinkhole sediments sequester Lake Huron's carbon. *Biogeochemistry* **115**, 235–250.

Nold S. C., Pangborn J. B., Zajack H. A., Kendall S. T., Rediske R. R. and Biddanda B. A. (2010) Benthic bacterial diversity in submerged sinkhole ecosystems. *Appl Environ Microbiol* **76**, 347–351.

Nold S. C. and Ward D. M. (1996) Photosynthate partitioning and fermentation in hot spring microbial mat communities. *Appl Environ Microbiol* **62**, 4598–4607.

Osterhout J., Schopf J. W., Williford K., McKeegan K., Kudryavtsev A. B. and Liu M.-C. (2021) Carbon isotopes of Proterozoic filamentous microfossils: SIMS analyses of ancient cyanobacteria from two disparate shallow-marine cherts. *Geomicrobiol J* **38**, 719–731.

Pearson A., McNichol A. P., Benitez-Nelson B. C., Hayes J. M. and Eglinton T. I. (2001) Origins of lipid biomarkers in Santa Monica Basin surface sediment: A case study using compoundspecific $\Delta$14C analysis. *Geochim Cosmochim Acta* **65**, 3123–3137.

Peng X., Guo Z., House C. H., Chen S. and Ta K. (2016) SIMS and NanoSIMS analyses of wellpreserved microfossils imply oxygen-producing photosynthesis in the Mesoproterozoic anoxic ocean. *Chem Geol* **441**, 24–34.

Penning H. and Conrad R. (2006) Carbon isotope effects associated with mixed-acid fermentation of saccharides by Clostridium papyrosolvens. *Geochim Cosmochim Acta* **70**, 2283–2297.

Pereira S., Zille A., Micheletti E., Moradas-Ferreira P., de Philippis R. and Tamagnini P. (2009) Complexity of cyanobacterial exopolysaccharides: Composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. *FEMS Microbiol Rev* **33**, 917–941.

Perry G. J., Volkman J. K. and Johns R. B. (1979) Fatty acids of bacterial origin in contemporary marine sediments.

Potter E. G., Bebout B. M. and Kelley C. A. (2009) Isotopic Composition of Methane and Inferred Methanogenic Substrates Along a Salinity Gradient in a Hypersaline Microbial Mat System. *https://home.liebertpub.com/ast* **9**, 383–390.

Prieto-Barajas C. M., Valencia-Cantero E. and Santoyo G. (2018) Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. *Electronic Journal of Biotechnology* **31**, 48–56.

Renström-Kellner E. and Bergman B. (1989) Glycolate metabolism in cyanobacteria. III. Nitrogen controls excretion and metabolism of glycolate in Anabaena cylindrica. *Physiol Plant* **77**, 46–51.

Rico K. I. and Sheldon N. D. (2019) Nutrient and iron cycling in a modern analogue for the redoxcline of a Proterozoic ocean shelf. *Chem Geol* **511**, 42–50.

Sánchez B., Zúñiga M., González-Candelas F., de Los Reyes-Gavilán C. G. and Margolles A. (2010) Bacterial and Eukaryotic Phosphoketolases: Phylogeny, Distribution and Evolution. *Microb Physiol* **18**, 37–51.

Schobben M. and van de Schootbrugge B. (2019) Increased stability in carbon isotope records reflects emerging complexity of the biosphere. *Front Earth Sci (Lausanne)* **7**, 87.

Schopf J. W., Kitajima K., Spicuzza M. J., Kudryavtsev A. B. and Valley J. W. (2018) SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proc Natl Acad Sci U S A* **115**, 53–58.

Schouten S., Hartgers W. A., Lòpez J. F., Grimalt J. O. and Sinninghe Damsté J. S. (2001) A molecular isotopic study of 13C-enriched organic matter in evaporitic deposits: recognition of CO2-limited ecosystems. *Org Geochem* **32**, 277–286.

Sessions A. L., Sylva S. P. and Hayes J. M. (2005) Moving-wire device for carbon isotopic analyses of nanogram quantities of nonvolatile organic carbon. *Anal Chem* **77**, 6519–6527.

Sharrar A. M., Flood B. E., Bailey J. v., Jones D. S., Biddanda B. A., Ruberg S. A., Marcus D. N. and Dick G. J. (2017) Novel large sulfur bacteria in the metagenomes of groundwater-fed chemosynthetic microbial mats in the Lake Huron basin. *Front Microbiol* **8**, 791.

Shen, V.K., Siderius, D.W., Krekelberg, W.P., and Hatch H. W. ed. (2017) *NIST Standard Reference Simulation Website*. NIST Stand., National Institute of Standards and Technology, Gaithersburg MD.

Spormann A. M. and Thauer R. K. (1988) Anaerobic acetate oxidation to CO2 by Desulfotomaculum acetoxidans. *Archives of Microbiology 1988 150:4* **150**, 374–380.

Stal L. J. and Moezelaar R. (1997) Fermentation in cyanobacteria. *FEMS Microbiol Rev* **21**, 179–211.

Steiner M. and Reiter J. (2001) Evidence of organic structures in Ediacara-type fossils and associated microbial mats. *Geology* **29**, 1119–1122.

Stuart R. K., Mayali X., Lee J. Z., Everroad R. C., Hwang M., Bebout B. M., Weber P. K., Pett-ridge J. and Thelen M. P. (2015) Cyanobacterial reuse of extracellular organic carbon in microbial mats. **10**, 1240–1251.

Sturt H. F., Summons R. E., Smith K., Elvert M. and Hinrichs K. U. (2004) Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry - New biomarkers for biogeochemistry and microbial ecology. *Rapid Communications in Mass Spectrometry* **18**, 617–628.

Tarhan L. G. (2018) The early Paleozoic development of bioturbation—Evolutionary and geobiological consequences. *Earth Sci Rev* **178**, 177–207.

Teece M. A. and Fogel M. L. (2007) Stable carbon isotope biogeochemistry of monosaccharides in aquatic organisms and terrestrial plants. *Org Geochem* **38**, 458–473.

Thomas A. T., Ognibene T., Daley P., Turteltaub K., Radousky H. and Bench G. (2011) Ultrahigh efficiency moving wire combustion interface for online coupling of high-performance liquid chromatography (HPLC). *Anal Chem* **83**, 9413–9417.

Voorhies A. A., Biddanda B. A., Kendall S. T., Jain S., Marcus D. N. and Nold S. C. (2012) Cyanobacterial life at low O 2 : community genomics and function reveal metabolic versatility and extremely low diversity in a Great Lakes sinkhole mat. , 250–267.

Voorhies A. A., Eisenlord S. D., Marcus D. N., Duhaime M. B., Biddanda B. A., Cavalcoli J. D. and Dick G. J. (2016) Ecological and genetic interactions between cyanobacteria and viruses in a low-oxygen mat community inferred through metagenomics and metatranscriptomics. *Environ Microbiol* **18**, 358–371.

Waldbauer J., Zhang L., Rizzo A. and Muratore D. (2017) DiDO-IPTL: A Peptide-Labeling Strategy for Precision Quantitative Proteomics. *Anal Chem* **89**, 11498–11504.

Wang Y., Chen X., Spengler K., Terberger K., Boehm M., Appel J., Barske T., Timm S., Battchikova N., Hagemann M. and Gutekunst K. (2022) Pyruvate:ferredoxin oxidoreductase and low abundant ferredoxins support aerobic photomixotrophic growth in cyanobacteria. *Elife* **11**.

Warburg O. (1928) Über die Geschwindigkeit der photochemischen Kohlensäurezersetzung in lebenden Zellen. II. *Über die Katalytischen Wirkungen der Lebendigen Substanz*, 341–366.

White D. (2000) *The Physiology and Biochemistry of Prokaryotes.*, Oxford University Press, New York.

Wieland A., Pape T., Möbius J., Klock J. H. and Michaelis W. (2008) Carbon pools and isotopic trends in a hypersaline cyanobacterial mat. *Geobiology* **6**, 171–186.

Wood H. G., Ragsdale S. W. and Pezacka E. (1986) The acetyl-CoA pathway of autotrophic growth. *FEMS Microbiol Rev* **2**, 345–362.

Xiong W., Lee T. C., Rommelfanger S., Gjersing E., Cano M., Maness P. C., Ghirardi M. and Yu J.
(2015) Phosphoketolase pathway contributes to carbon metabolism in cyanobacteria. *Nature Plants 2015 2:1* **2**, 1–8.

Zhang Y., Wen Z., Washburn M. P. and Florens L. (2010) Refinements to label free proteome quantitation: How to deal with peptides shared by multiple proteins. *Anal Chem* **82**, 2272– 2281.

# Chapter 4

**Exploring the magnitude of carbon isotope fractionation during polyglucose synthesis in Cyanobacteria**

[Draft in preparation for *Organic Geochemistry*]

Ana Gonzalez-Nayeck, Ann Pearson

## Abstract

Carbohydrates are rapidly recycled between autotrophic and heterotrophic organisms in a broad range of both modern and ancient environments. The carbon isotopic compositions of individual sugar compounds can be used to constrain the nature of this recycling, including the reactions of different monomeric units and the role of varying redox conditions in driving these reactions. Most compound-specific studies treat carbohydrates as a homogenous pool reflecting the $\delta^{13}C$ value of initial photosynthate. However, a few prominent exceptions have shown evidence for $^{13}C$-enrichment in pentose sugars relative to hexose sugars. Here, we hypothesize that fractionation during the synthesis of polyglucose storage polymers is partially responsible for this isotopic pattern. To test this hypothesis, we compiled reported $\delta^{13}C$ measurements of xylose (pentose) and glucose (hexose) from laboratory and natural environments and used these data along with the principles of open-system isotope mass balance to estimate the potential magnitude of such a fractionation. Our results show that the $^{13}C$-enrichment in xylose relative to glucose monomers is systematic, and that the maximum magnitude of the isotope effect for polyglucose synthesis is approximately 4‰. Polyglucose formation is a quantitatively important carbon-fixation process in the open ocean, soils and microbial mats, and its proportion relative to downstream processes such as internal recycling and fermentation of sugars to acetate is redox- and resource-sensitive. As such, our work has potential implications for understanding net community processes in complex biogeochemical systems.

**Introduction**

The systematic differences observed between the $\delta^{13}C$ values of carbohydrate, protein and lipids derived from a common biosynthetic or environmental source have long been known to hold significant interpretive value (DeNiro and Epstein, 1977; Macko et al., 1987; Hayes, 2001; Teece and Fogel, 2007). Compound-specific isotope analyses (CSIA) of organic matter preserved in the ancient sedimentary record are invaluable tools for examining how these differences illuminate biological processes (Hayes et al., 1989). For example, although the total organic carbon (TOC) of most primary producers using the Calvin-Benson-Bassham cycle has a $\delta^{13}C$ composition approximately 25‰ depleted in $^{13}C$ relative to $CO_2$ (Hayes et al., 1999), a $\geq$ 8‰ further depletion in the $\delta^{13}C$ value of associated *n*-alkyl lipids relative to TOC often indicates a large contribution specifically from Cyanobacteria (Sakata et al., 1997). As such, understanding the environmental and biosynthetic controls on the $\delta^{13}C$ compositions of different molecular classes is critical towards refining CSIA as a tool for organic geochemistry.

The $\delta^{13}C$ values of cellular lipids, proteins, carbohydrates and nucleic acids depend on the relative amount of carbon allocated to these fractions in tandem with the isotope effects of the biosynthetic steps (DeNiro and Epstein, 1977; Sakata et al., 1997; Hayes, 2001). Microbial organisms distribute carbon among these biochemical fractions differently under different nutrient, irradiance, and oxidative regimes (e.g., Halsey et al., 2014; Laws and McClellan, 2022), variables that likely have differed throughout Earth history (e.g., Meyer and Kump, 2008; Reinhard et al., 2020) and affect heterotrophs differently from photoautotrophs (Deniro and Epstein, 1978; Blair et al., 1985).

Carbohydrates comprise the majority of dissolved organic matter in the surface ocean (Benner et al., 1992; Aluwihare and Repeta, 1999). They also form the bulk of exopolymeric substances in microbial mats (Rossi and de Philippis, 2015) and organic matter in soils (Simpson et al., 2001), making exocellular polysaccharides (EPS) the most abundant organic carbon source for microbial heterotrophs globally. While higher animals ingest prey to meet their anabolic carbon requirements, microbial heterotrophs must assimilate low-molecular weight (LMW) substrates. Marine heterotrophs selectively utilize labile compounds such as phytoplankton-derived carbohydrates and amino acids before assimilating high molecular weight compounds such as polymeric plant material, which usually requires the help of extracellular enzymes (Harvey et al., 1995; Arnosti et al., 2011; Mahmoudi et al., 2017). Bacteria in microbial mats (Anderson et al., 1987; Flemming and Wingender, 2010; Stuart et al., 2015) and soils (Weijers et al., 2010) show a similar preference for LMW compounds. These LMW substrates are often derived from the lysis of organic material excreted by living photosynthetic bacteria (Fogg, 1983; Braakman et al., 2017) and generally are rich in carbohydrates. Microbial preference for these compounds results in a $\delta^{13}C$ composition of heterotrophic bacterial biomass equal to that of net community photosynthetic production (Blair et al., 1985; Gonzalez-Nayeck et al., 2022).

While the enzymatic mechanisms underpinning variation in carbon isotope ratios for lipids (e.g., Monson and Hayes, 1982; Melzer and Schmidt, 1987; Sakata et al., 1997) and proteins (e.g., Abelson and Hoering, 1961; Macko et al., 1987; Hayes, 2001) have received much attention, case studies of carbohydrates most often treat these compounds as a homogenous pool whose $\delta^{13}C$ values are dictated by the isotope effect of initial carbon fixation (e.g., Farquhar et al., 2001 and references therein). Nonetheless, a few studies on photoautotrophic bacteria, algae, higher plants, and modern microbial mats show evidence of $^{13}C$-enrichment in pentose sugars relative to hexose

sugars, with potential implications for understanding net community processes in these complex systems (van der Meer et al., 2001; van Dongen et al., 2002; Derrien et al., 2006; Teece and Fogel, 2007; Dungait et al., 2008; Gonzalez-Nayeck et al., in review).

Recent work suggests Cyanobacteria meet their night-time maintenance energy requirements via the phosphoketolase pathway, in which pentose sugars (e.g., xylose) are converted into xylulose-5-phosphate and subsequently broken down into acetyl-phosphate and glyceraldehyde-3-phosphate (Xiong et al., 2015; Chuang and Liao, 2021). The acetate excreted via this pathway is a likely anabolic carbon source for heterotrophic bacteria, while the glyceraldehyde-3-phosphate likely replenishes Calvin-cycle intermediates depleted during fast growth in Cyanobacteria (Makowka et al., 2020). As such, understanding the mechanisms underpinning the isotopic differences between pentose versus hexose sugars may further understanding both of the $\delta^{13}C$ compositions of heterotrophic bacteria and of the anaplerotic reactions of Cyanobacteria.

The formation of xylose from glucose proceeds via glucuronic acid and involves decarboxylation of the C-6 carbon from the precursor hexose (e.g., Smith and Bar-Peled, 2017). Because the C-6 carbon in the glucose moieties of sucrose isolated from beet syrup is relatively $^{13}C$-depleted compared to the rest of the molecule (Figure 4.4C; Rossmann et al., 1991; Gilbert et al., 2011), this decarboxylation reaction is often cited as the likely cause of the difference in $\delta^{13}C$ values between pentose and hexose sugars (van der Meer et al., 2001; van Dongen et al., 2002; Teece and Fogel, 2007; Dungait et al., 2008). The work of Rossman et al. (1991) and Gilbert et al. (2011) indicates the carbon contained in the C1-C5 carbons is ~1‰ enriched in $^{13}C$ compared to the average of all 6 carbon atoms, implying that while the individual C6 position is $^{13}C$-depleted by 5‰, the overall difference between C5 and C6 sugars should be small.

It therefore remains a mystery why data from multiple autotrophic organisms show larger differences (~3-5‰) between xylose and glucose than can be explained by this mechanism (Table 1, Figure 4.1). An additional source of fractionation is required to explain the full range of these observations. In the present study, we hypothesize that there is a significant kinetic isotope effect (KIE) that imparts fractionation during the irreversible formation of polyglucose (e.g., glycogen or other glucans) from glucose-1-phosphate, the precursor for xylose synthesis (Figure 4.2). Polyglucose in photoautotrophic organisms forms the building block for both intracellular (glycogen) and extracellular (exopolysaccharide) storage sugars. Because existing data on xylose and glucose $\delta^{13}C$ compositions span a range of growth conditions, and these physiological differences affect the amount of storage sugars that are synthesized, we propose that the magnitude of the isotope effect and the extent of its expression can be determined using these existing data.
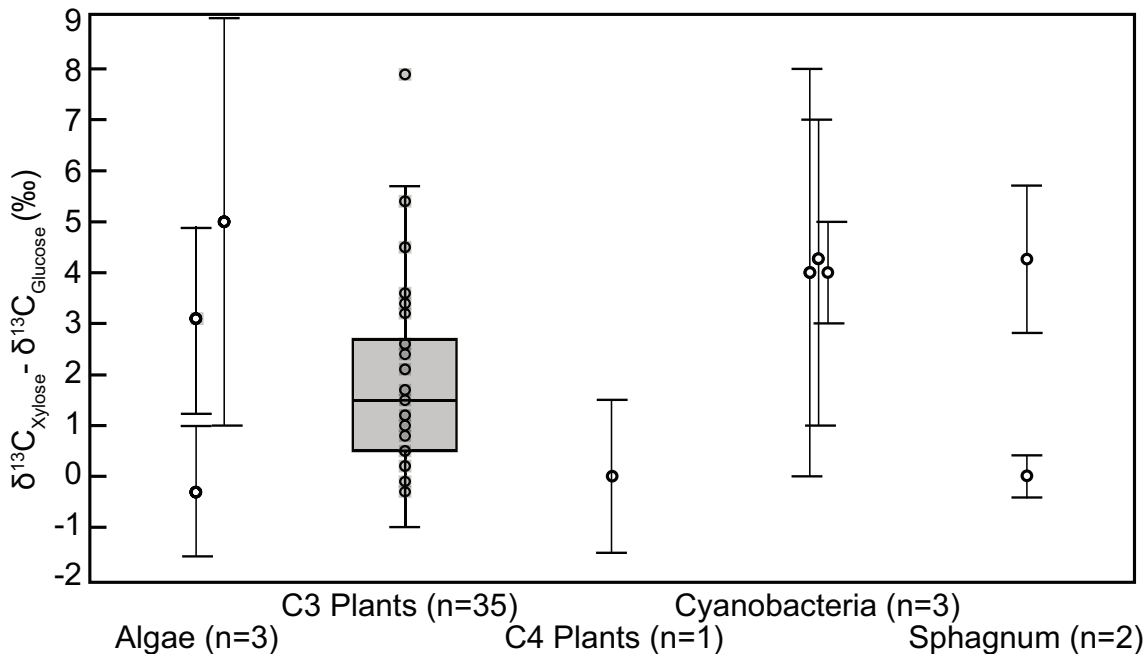


**Figure 4.1** Compilation of differences between $\delta^{13}C$ values for xylose and glucose. All data points represent individual datapoints with the exception of C3 plants which are represented by a box-and-whisker plot. Error bars represent the propagated error when accounting for subtraction; for most data individually measured error is approximately half of what is shown. Data from Macko et al., (1991), van Dongen et al., (2002), Teece and Fogel (2007), Derrien et al., (2006), Dungait et al., (2008) and Gonzalez-Nayeck et al., *in review*.

**Table 4.1** Results from literature search for $\delta^{13}$C composition of Glucose (Glu) and Xylose (Xyl) from autotrophic organisms. Measurement errors ("err") are as defined in each text; error values with an asterisk represent a single estimated error applied to all measurements.

| Reference | Derivatization Method | Type of Organism | *Genus* | Glu $\delta^{13}$C (‰) | Glu err | Xyl $\delta^{13}$C (‰) | Xyl err | Total Carb $\delta^{13}$C (‰) | TOC $\delta^{13}$C (‰) | Xyl-Glu $\delta^{13}$C (‰) |
|---|---|---|---|---|---|---|---|---|---|---|
| Van Dongen et al. 2002 | MBD | Algae | *I. galbana* | -6.8 | 0.3 | -7.1 | 0.9 | | -13.8 | -0.3 |
| Van Dongen et al. 2002 | MBD | Algae | *Phaeocystis* | -21.3 | 0.9 | -18.2 | 0.9 | | -17.5 | 3.1 |
| Teece and Fogel 2007 | Alditol acetate | Algae | *Cyanidium caldarium* | -13.0 | 2* | -8.0 | 2* | -10.0 | -13.0 | 5.0 |
| Derrien et al. 2006 | HMDS & TMCS | C3 plant | Wheat leaf | -26.0 | 2.0 | -27.0 | 1.0 | | -28.3 | -1.0 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Crepis sp.* | -30.9 | | -31.2 | | | -30.2 | -0.3 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Trifolium repens* | -27.7 | | -27.9 | | | -28.7 | -0.2 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Plantago lanceolata* | -29.4 | | -29.5 | | | -28.8 | -0.1 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Myosotis scorpiodes* | -29.8 | | -29.8 | | | -29.5 | 0.0 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Salix fragilis* | -27.4 | | -27.4 | | | -28.7 | 0.0 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Stellaria graminea* | -28.8 | | -28.6 | | | -27.7 | 0.2 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Sambucus nigra* | -28.0 | | -27.7 | | | -29.3 | 0.3 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Cerastium holosteoides* | -29.4 | | -28.9 | | | -28.2 | 0.5 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Centaurea nigra* | -29.2 | | -28.6 | | | -30.7 | 0.6 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Medicago lupulina* | -28.2 | | -27.6 | | | -29.0 | 0.6 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Taraxacum officinale* | -29.2 | | -28.6 | | | -30.4 | 0.6 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Holcus lanatus* | -29.8 | | -29.2 | | | -30.4 | 0.6 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Cirsium arvense* | -27.6 | | -26.8 | | | -28.2 | 0.8 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Prunella vulgaris* | -30.7 | | -29.8 | | | -30.2 | 0.9 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Ranunculus repens* | -29.2 | | -28.2 | | | -28.4 | 1.0 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Bellis perennis* | -32.2 | | -31.0 | | | -32.0 | 1.2 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Dactylis glomerata* | -28.9 | | -27.4 | | | -28.4 | 1.5 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Potentilla reptans* | -29.2 | | -27.7 | | | -29.2 | 1.5 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Cirsium vulgare* | -28.5 | | -26.8 | | | -28.7 | 1.7 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Acer campestre* | -25.9 | | -23.8 | | | -27.5 | 2.1 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Alisma plantago-aquatica* | -27.7 | | -25.3 | | | -27.8 | 2.4 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Malus sylvestris* | -25.8 | | -23.4 | | | -27.3 | 2.4 |

| Dungait et al. 2008 | Alditol acetate | C3 plant | *Veronica serpyllifolia* | -30.4 | | -27.9 | | | -31.0 | 2.5 |
|---|---|---|---|---|---|---|---|---|---|---|
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Achillea millefolium* | -31.0 | | -28.4 | | | -31.2 | 2.6 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Ranunculus acris* | -31.9 | | -29.2 | | | -30.7 | 2.7 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Rumex palustris* | -27.0 | | -24.3 | | | | 2.7 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Lotus tenuis* | -30.6 | | -27.4 | | | -30.5 | 3.2 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Crataegus monogyna* | -28.0 | | -24.6 | | | -29.7 | 3.4 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Leucanthemum vulgare* | -29.2 | | -25.6 | | | -29.5 | 3.6 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Potentilla anserine* | -30.4 | | -26.8 | | | -29.2 | 3.6 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Plantago media* | -28.0 | | -23.5 | | | -27.4 | 4.5 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Plantago major* | -31.0 | | -25.6 | | | -30.8 | 5.4 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Rumex acetosa* | -29.8 | | -24.1 | | | -28.5 | 5.7 |
| Dungait et al. 2008 | Alditol acetate | C3 plant | *Rumex conglomeratus* | -28.3 | | -20.4 | | | -28.4 | 7.9 |
| D. derrien et al. 2006 | HMDS & TMCS | C4 plant | Maize leaf | -12.0 | 2.0 | -12.0 | 1.0 | | -12.4 | 0.0 |
| Teece and Fogel 2007 | Alditol acetate | Cyanobacteria | *Anabaena* | -18.0 | 2.0 | -14.0 | 2.0 | -14.0 | -16.0 | 4.0 |
| Gonzalez-Nayeck et al. in prep | MBD | Cyanobacteria | *Cyano* | -26.5 | 1.7 | -22.2 | 1.3 | -24.5 | | 4.3 |
| Moers et al. 1993 | Alditol acetate | Cyanobacteria | *Lyngbia aestuarii and Microcoleus chtonoplastes* | -7.7 | 0.5 | -3.7 | 0.6 | | -9.2 | 4.0 |
| Macko et al. 1991 | Alditol acetate | Sphagnum | *Sphagnum* | -26.0 | 0.1 | -26.0 | 0.1 | | -26.0 | 0.0 |
| Van Dongen et al. 2002 | MBD | Sphagnum | *S. cuspidatum* | -23.8 | 0.7 | -19.5 | 0.7 | | -25.6 | 4.3 |

## Methods

*Data Compilation*

We conducted a literature search with the purpose of finding previous compound-specific (by gas chromatography-isotope ratio mass spectrometry, GC-IRMS) measurements of the $\delta^{13}C$ compositions of monosaccharide sugars extracted from autotrophic organisms or environments containing a large proportion of autotrophic organisms (*e.g.,* cyanobacterial microbial mats, or peat). Results of the search are shown in Table 1. Given that our analyses are focused on the

difference between xylose and glucose sugars, only data from organisms or environments with measurements for both sugars were included, except for a prior measurement by the authors (Gonzalez-Nayeck et al., 2022) from a system where glucose was the only quantifiable monosaccharide and the xylose value is estimated from first principles (see results and discussion).

These data represent three different derivatization methods to render monosaccharides GC-amenable prior to GC-IRMS measurement: the alditol acetate method (Klok et al., 1982), the methylboronic acid method (MBD; van Dongen et al., 2001), and trimethylsilylation via addition of hexamethyldisilazane (HMDS) and chlorotrimethylsilane (TMCS) (Derrien et al., 2006). The alditol acetate method derivatizes all monosaccharides completely but has relatively greater error than the MBD method, and the generated alditol acetates can represent multiple monosaccharides (e.g., glucitol can represent both glucose and fructose). This approach differs from the MBD method which preserves the original isomeric structure of the monosaccharide (van Dongen et al., 2001). However, the MBD method cannot be used to compare quantitative contributions from all monosaccharides because the derivatization of a few key monosaccharides, including mannose, ribose and galactose, is incomplete (van Dongen et al., 2001; Teece and Fogel, 2007). Nonetheless, both glucose and xylose are completely derivatized via the MBD method, indicating that the isotopic compositions of these two monosaccharides should be comparable between methods (van Dongen et al., 2001). Relatively little information is available on the derivatization yield of monosaccharides by trimethylsilylation via HMDS and TMCS, but this mechanism does not act on any carbon bonds and therefore carries no fractionation (Derrien et al., 2006). The derivatization method used in each study is noted in Table 1.

*Analytical Framework*

To explore the isotope effect responsible for the observed difference between pentose and hexose sugars in autotrophic organisms, we employ an analytical framework based on the principles of open-system isotope mass balance detailed in Hayes (2001). We use equation (1):

$$\delta_{P'} = \delta_{R'} - \varepsilon \tag{1}$$

where $\delta_{P'}$ and $\delta_{R'}$ are the instantaneous isotopic compositions of a product and reactant and $\varepsilon$ (expressed in ‰) is the isotope effect associated with the hypothetical reaction(s) causing the observed isotopic difference between the two (Hayes, 2001). In this work, we define $\delta_{P'}$ and $\delta_{R'}$ as the $\delta^{13}C$ compositions of glucose and xylose, respectively. This is counter to the traditional view where glucose is the reactant, which is converted to glucuronic acid and decarboxylated to form the product xylose (Figure 4.2; e.g., Smith and Bar-Peled, 2017). However, the products of an enzymatic reaction are almost always depleted relative to the reactant (Hayes, 2001). Given that glucose is systematically depleted in $^{13}C$ relative to xylose, we proceed with the assumption that there is an isotopic fractionation between a key intermediate that acts as a precursor to xylose synthesis and the production of glucose from that precursor. For our analysis, we rearrange equation (1) to

$$\varepsilon = \delta_{R'} - \delta_{P'} \tag{2}$$

and use the difference between xylose and glucose as an approximation of the magnitude of the isotope affect ($\varepsilon$) associated with this unidentified reaction.

We hypothesize that the transient pool of ADP-glucose in bacteria (or UDP-glucose in eukaryotes) is the $^{13}C$-enriched xylose precursor which is isotopically fractionated via the irreversible formation of polyglucose (red arrow, Figure 4.2). To test this hypothesis, we create a

plot (Figure 4.3) modeled after Figure 30 from Hayes (2001). In Figure 4.3, the $\delta^{13}C$ composition of xylose (reactant) and glucose (product) are plotted as a function of the presumed fraction of total carbohydrate (initial photosynthate) allocated to the final polyglucose. We only include data with reported measurements of total weighted-average monosaccharide $\delta^{13}C$ values, which should theoretically represent the $\delta^{13}C$ of initial photosynthate. By normalizing glucose and xylose $\delta^{13}C$ values to the $\delta^{13}C$ value of this initial photosynthate, we can compare across environments that are experiencing varying amounts of polyglucose synthesis. If the majority of initial photosynthate carbon is allocated to polyglucose, we expect that glucose $\delta^{13}C$ values would closely resemble the calculated $\delta^{13}C$ value of initial photosynthate, and the small proportion of residual glucose-1-phosphate (and by proxy xylose) would represent a maximum $^{13}C$ enrichment in a small yield of xylose. Conversely, if a relatively small proportion of initial photosynthate carbon is allocated to polyglucose, glucose $\delta^{13}C$ values would represent a maximum $^{13}C$ depletion and xylose production or recycling may be extensive. In all cases, we would expect the difference between the two endmembers to be relatively constant, reflecting the value of $\varepsilon$.

To create Figure 4.3, we used data from a nutrient-replete batch culture of the Cyanobacterium Dolichospermum *sp.* (formerly Anabaena *sp.*; Teece and Fogel, 2007) as well as data from two microbial mats dominated by Cyanobacteria: one from Chocolate Pots Hot Springs, Yellowstone National Park, representing a relatively nutrient-limited surface environment with high photon flux (Gonzalez-Nayeck et al., 2022) and one from the Middle Island Sinkhole, Lake Huron, USA representing a nutrient-limited environment with relatively lower photon flux (Gonzalez-Nayeck et al., *in review*). Given that polyglucose synthesis in Cyanobacteria is enhanced under conditions of nutrient limitation (Yoo et al., 2007) and high photon flux (Cano et

al., 2018), these environments should represent a gradient from near-minimum to near-maximum polyglucose synthesis.

*$\delta^{13}C$ measurement of monosaccharide moieties of EPS from modern cyanobacterial mats*

Methods for determining the $\delta^{13}C$ compositions of monosaccharide moieties of EPS extracted from modern microbial mat samples are detailed in Gonzalez-Nayeck et al. 2022 and are summarized here. Briefly, EPS were extracted from mat samples by incubation in 10% (w/v) NaCl followed by centrifugation to separate the supernatant. 100% ethanol was subsequently added to the supernatant to a final concentration of 70% and EPS were precipitated at 4ºC overnight ("Method 7" from Klock et al., 2007). EPS were hydrolyzed into monomers via the addition of 1 mL 12M $H_2SO_4$ and stirring for 2 hours at room temperature, followed by dilution to 1 M and heating at 85ºC for 4.5 hours (van Dongen et al., 2001). The solution was neutralized using $BaCO_3$, the resulting precipitant was removed via centrifugation, and the supernatant was collected, frozen and lyophilized (van Dongen et al., 2001). Lyophilized sugar monomers were derivatized via the MBD method as described by van Dongen et al., (2001).
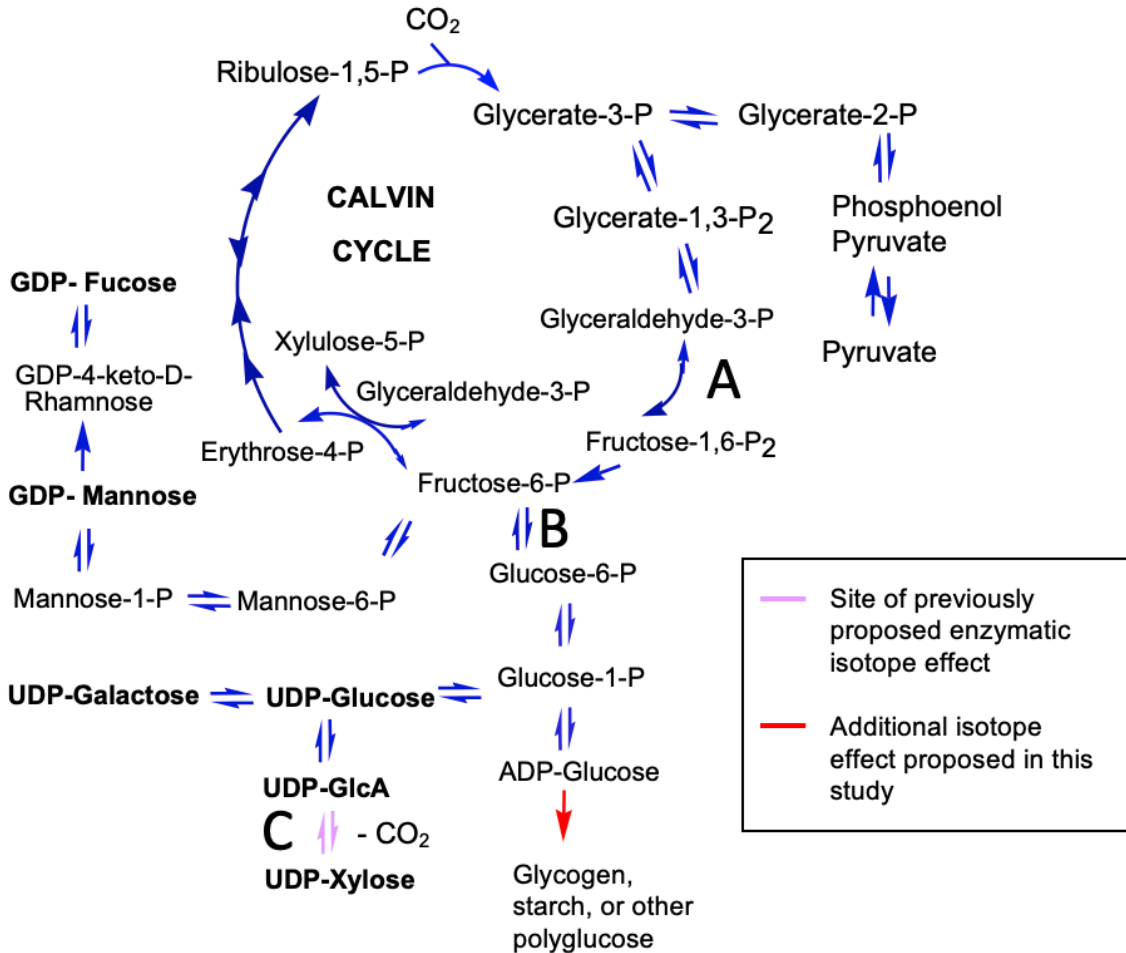
**Figure 4.2** Modified from Velmurugan and Incharoensakdi (2021). Simplified pathway for the synthesis of exopolysaccharide in Synechocystis sp. PCC 6803. The decarboxylation reaction previously suggested to account for the carbon isotopic difference between xylose and glucose is pictured in pink. The irreversible reaction proposed as an additional source of carbon isotope fractionation is pictured in red. A, B & C: Previously proposed mechanisms for generating xylose relatively depleted in $^{13}C$; details in Figure 4.4.

**Results and Discussion**

*Observed differences between xylose and glucose $\delta^{13}C$*

Figure 4.1 and Table 4.1 show the difference between xylose and glucose $\delta^{13}C$ values for algae (*n=3*), C3 plants (*n=35*), a C4 plant (*n=1*), Cyanobacteria (*n=3*) and Sphagnum (*n=2*). All values are positive within error, indicating that xylose is $^{13}C$-enriched relative to glucose in all observations from photosynthetic organisms. Because xylose is formed via the decarboxylation of glucuronic acid derived from glucose (pink arrows, Figure 4.2; Smith and Bar-Peled, 2017), this

ordering is counter to the expectation that the product of an enzymatic reaction is depleted in $^{13}$C

relative to reactants (Hayes, 2001). Previous explanations (van der Meer et al., 2001; van Dongen

et al., 2002; Teece and Fogel, 2007; Dungait et al., 2008) for this inverse isotope effect invoke

observed heterogeneity in the $\delta^{13}$C values of individual carbon atoms of glucose, where the C-6

carbon of glucose is approximately 5‰ depleted relative to the average of the whole molecule

(Rossmann et al., 1991). Because the relatively $^{13}$C-depleted C-6 carbon is cleaved from

glucuronic acid to form xylose (figure 4.4C), this reaction may yield xylose that is enriched in $^{13}$C

relative to the precursor glucose.

Since Rossmann et al. (1991) first measured the $^{13}$C depletion in the C-6 carbon of beet-

derived glucose, attempts to explain the mechanism behind this pattern have relied on

understanding the relative $^{13}$C enrichment of the C1-C5 positions (Rossmann et al., 1991; Gleixner

and Schmidt, 1997; Gilbert et al., 2012). Rossmann et al. (1991) originally proposed a KIE during

the bi-directional fructose-1,6 biphosphate aldolase reaction (Figure 4.4A), since the backwards

reaction (fructose-1,6 biphosphate to 2 triose phosphates) cleaves the bond linking the C3 and C4

carbons in linear fructose-1,6 biphosphate. This was subsequently tested *in-vitro* using a rabbit

muscle aldolase enzyme, resulting in a relative $^{13}$C-enrichment in the C3 and C4 carbons of

fructose-1,6 biphosphate and a $^{13}$C-depletion in the corresponding carbon atoms in the product

triose phosphates (Gleixner and Schmidt, 1997). Subsequent modeling exercises using *in-vivo*

isotope data have suggested that this isotope effect can explain the majority of the observed

patterns in the C3 and C4 carbons of fructose-1,6 biphosphate and its products (Tcherkez et al.,

2004).

However, the aldolase reaction by itself does not adequately explain the $^{13}$C depletion in

the C-6 carbon relative to the C-1 carbon in glucose. Gilbert et al. (2012) measured the kinetic and

equilibrium isotope fractionation *in-vitro* for the enzyme glucose isomerase, which facilitates the isomerization reaction between fructose-6-P and glucose-6-P (Figure 4.4B). In the fructose to glucose direction, the isomerization reaction involves (1) the opening of the fructose ring, (2) a hydride shift reaction in which the C-1 hydroxyl bond is converted into a carbonyl bond and vice-versa for the C-2 carbon, creating linear glucose (Figure 4.4B); and (3) the closing of the glucose ring (Nam, 2022). In the fructose to glucose direction, there is a kinetic fractionation of approximately 15‰ resulting in a relative $^{13}$C-depletion in the C-2 carbon of glucose. In the equilibrium fractionation, the C-2 and C-6 carbons in glucose are 7‰ and 4‰ depleted in $^{13}$C relative to fructose, and the C-1 carbon is approximately 13‰ enriched in $^{13}$C(Gilbert et al., 2012). Gilbert et al. (2012) suggest that it is the effect of this equilibrium fractionation on the C-1 carbon that is primarily responsible for observed depletion of C-6 relative to C-1 in glucose *in-vivo*. This agrees with the expectation that equilibrium will favor $^{13}$C in the relatively stronger carbonyl bond (glucose C-1) versus a hydroxyl bond (fructose C-1) (Bigeleisen, 1965; Cleland, 2005). Furthermore, the glucose isomerase reaction is infamous for reaching thermodynamic equilibrium quickly given difficulties in using the enzyme to create high-fructose corn syrup because the reaction will never fully isomerize glucose (Liu et al., 2019). Taking these observations in concert, it appears that the C-6 carbon in glucose is relatively depleted in $^{13}$C compared to the average of the molecule because fructose-1, 6 biphosphate aldolase and glucose isomerase result in $^{13}$C enrichments in the C-3 and C-4 and C-1 carbon atoms, respectively (Figure 4.4C). However, the net of these effects yields an average $\delta^{13}$C value of the total C-1 to C-5 carbons in glucose that is approximately 1‰ enriched relative to the average of the whole molecule (Rossmann et al., 1991; Gilbert et al., 2011), which is insufficient to account for the observed magnitude of $^{13}$C enrichment in xylose relative to glucose (approximately 4-5‰; Figures 4.1, 4.3).

The formation of polyglucose from ADP-Glucose in bacteria (or UDP-Glucose in eukaryotes) is one of the few unidirectional steps identified in the monosaccharide synthesis pathway (red arrow, Figure 4.2; Tymoczko et al., 2015; Velmurugan and Incharoensakdi, 2021). We propose that this reaction imparts an additional fractionation resulting in a pool of polyglucose depleted in $^{13}$C and a transient pool of ADP-glucose that is relatively enriched in $^{13}$C. Because ADP-glucose is interconverted with glucose-1-phosphate, a precursor for xylose synthesis, this process would also result in $^{13}$C-enriched xylose. In cyanobacteria, galactose is derived from the same precursor as xylose (Figure 4.2); accordingly, galactose from a cyanobacterial culture (Teece and Fogel, 2007) and from a cyanobacterial mat (Moers et al., 1993) was also determined to be enriched in $^{13}$C relative to glucose. An implication of this theory is that the pool of $^{13}$C-depleted glucose in autotrophs is derived almost entirely from polyglucose formation and subsequent recycling. This is supported by observations that the majority (>80%) of fixed carbon in Yellowstone cyanobacterial mats is allocated to the polyglucose compound glycogen (Nold and Ward, 1996), and the general principle that glycogen and starch represent the main glucose storage products in bacteria, eukaryotic algae, and plants (Tymoczko et al., 2015).

Polyglucose formation from ADP-glucose proceeds via the glycogenin (GN) and glycogen synthase (GS) enzyme complex (Figure 4.5; Marr et al., 2022). Since these enzymes act mostly on C-O bonds, the exact mechanism responsible for an isotope effect remains unclear; however, the activation energy for this process is notably high, and the intermediate states involved are currently unknown (Marr et al., 2022). We proceed with the assumption that there is a KIE during this process but note that further experimental work is necessary to understand the exact mechanism for such a KIE.

*The expected magnitude of carbon isotopic fractionation during polyglucose synthesis*

A consequence of the hypothesis that irreversible polyglucose synthesis imparts a $^{13}$C fractionation is that the magnitude of its expression would be directly proportional to the proportion of initial photosynthate that is allocated to storage or excreted polyglucose compounds. As detailed in the Methods section, we use this to test our hypothesis by comparing the relative offsets between glucose, xylose and initial photosynthate $\delta^{13}$C compositions in three cyanobacterial systems representing a range of polyglucose production. We draw the following observations (Figure 4.3):

1) When relatively more initial photosynthate is allocated to polyglucose compounds, the $\delta^{13}$C value of glucose resembles the $\delta^{13}$C value of initial photosynthate, and xylose is increasingly enriched in $^{13}$C.

2) When normalized to the pool of initial photosynthate, the $\delta^{13}$C offset between xylose and glucose in a nutrient-replete culture and a nutrient-limited microbial mat (MIS), are both approximately 4‰; this is similar to the maximum observed magnitude for xylose – glucose $\delta^{13}$C in Figure 4.1.

Observation 1 supports the existence of a kinetic isotope fractionation during polyglucose synthesis, at least in Cyanobacteria. Observation 2 suggests that the KIE of this process is likely to be approximately 4‰.

*Ranges of expressed $^{13}C$ fractionation in glucose relative to xylose in different organisms*

Observed ranges in $\varepsilon_{xyl\text{-}glu}$ are variable across different types of organisms, with C3 plants, for instance, showing a wider range in $\varepsilon_{xyl\text{-}glu}$ values than Cyanobacteria (Figure 4.1). This is certainly mostly an artifact of relative sample sizes ($n=35$ for C3 plants and $n=3$ for cyanobacteria). Nonetheless, there are a priori reasons to expect that this value would differ among algae, higher plants, Cyanobacteria and Sphagnum. Figure 4.6 shows a partial pathway for cyanobacterial glycogen synthesis and catabolism. The green arrow connecting polyglucose to glucose-1-P represents the recycling of storage molecules, which occurs in Cyanobacteria during the night-time respiration of glycogen as both an energy and anabolic carbon source to replenish metabolic intermediates (Shinde et al., 2020). Since glucose-1-P is the precursor for xylose synthesis, we can expect that a greater degree of recycling of sugar polymer back towards glucose-1-P will lead to a relatively smaller $\varepsilon_{xyl\text{-}glu}$ as $^{13}C$-depleted carbon is re-introduced into the precursor molecule.

In plant leaves, the major polyglucose molecule formed from ADP-glucose is starch stored in the chloroplast, which is accumulated during the daytime (Smith et al., 2012). While the cyanobacterial pathway is not directly analogous to the enzymatic pathway for starch degradation in eukaryotes, eukaryotic starch synthesis presumably evolved from cyanobacteria and the general mechanism for glycogen accumulation and breakdown is at least broadly comparable to starch synthesis (Ball et al., 2011). Plants subsequently degrade starch at night into glucose-1-P, which is exported out of the chloroplast and into the cytosol where it forms structural components including cellulose (Polko and Kieber, 2019). Algae, except for red algae, also generate and store starch in the plastid; however, the glucose-1-P generated and excreted into the cytosol via starch

94

degradation is primarily used for synthesizing storage glycogen instead of structural components (Ball et al., 2011). Red algae are unique in that they form a different type of starch ("Floridian starch") directly within the cytosol from UDP-glucose (Viola et al., 2001). Except for red algae, which we will not consider further since they do not form storage polyglucose (either glycogen or starch) via an ADP-glucose precursor, these generalizations appear to be widely applicable across photosynthetic organisms.

In higher plants, starches primarily are mobilized into the synthesis of structural sugars. In cyanobacteria, glycogen either is excreted as polysaccharide during the daytime or it is respired to generate cellular maintenance energy at night. Cyanobacterial excretion of polysaccharides is a quantitatively important process in many ecosystems, including microbial mats (Nold and Ward, 1996), the surface ocean (Fogg, 1983; Braakman et al., 2017), and soils (Weijers et al., 2010). We can therefore expect that plants retain relatively more of their polyglucose pool relative to Cyanobacteria. While current data are limited, especially for the Cyanobacteria, this idea is supported by the relative similarity in $\varepsilon_{xyl\text{-}glu}$ values for Cyanobacteria (Figure 4.1, Table 4.1) which all approximate the maximum value of 4‰. Mean C3 plant values for $\varepsilon_{xyl\text{-}glu}$ are relatively lower at approximately 1.5‰. Two haptophyte algae, *Isocrysis galbana* and *Phaocystis* have $\varepsilon_{xyl\text{-}glu}$ values of 0 and 3.1‰, respectively; the *Phaocystis* sample was collected from the natural environment during a bloom (van Dongen et al., 2002), during which large quantities of extracellular polysaccharide synthesis is expected to occur. If we consider algae to be polyglucose recyclers of intermediate efficiency, we might expect that $\varepsilon_{xyl\text{-}glu}$ for algae will be variable and depend on growth conditions.

*Implications for the geologic record*

Autotrophic carbon storage products, both retained and exuded, are increasingly recognized as a quantitatively important components of global fixed carbon. Recently, Wu et al. (2022) revisited existing models of modern ocean primary productivity and found that decoupling nutrient availability and carbon fixation (i.e., assuming that carbon fixation can occur in excess of nutrients required for growth and is limited solely by light-saturated photosynthetic capacity and $CO_2$ availability) results in a 30% increase in estimated global primary productivity, more closely matching observed oceanic C:P ratios. The role of photosynthetic exudates in microbial mats has been reviewed extensively (e.g., Rossi and de Philippis, 2015), and recent work (Gonzalez-Nayeck et al., 2022) underscores that it is the isotopic composition of individual exudates, not the photosynthetic organisms themselves, that determines the isotopic composition of heterotrophic mat organisms. Heterotrophic microorganisms in soils also assimilate autotrophic exudates as anabolic carbon sources (Weijers et al., 2010). If the formation of polyglucose indeed carries a carbon KIE, we might expect that there is a quantitatively important pool of global fixed carbon that is $^{13}C$-depleted relative to standing biomass or initial photosynthate. Since nutrient availability has varied throughout geologic history (e.g., Reinhard et al., 2017), and nutrient availability dictates the relative proportion of polyglucose formation in autotrophs (e.g., Laws and McClellan, 2022), we also may expect that the proportion of total fixed carbon that is directed toward storage carbon has differed over geologic history. As such, both the $\delta^{13}C$ composition of autotrophic exudates and the effects of exudates on total and compound-specific organic carbon $\delta^{13}C$ compositions represents a critical gap in our current understanding of the modern and geologic carbon cycle.
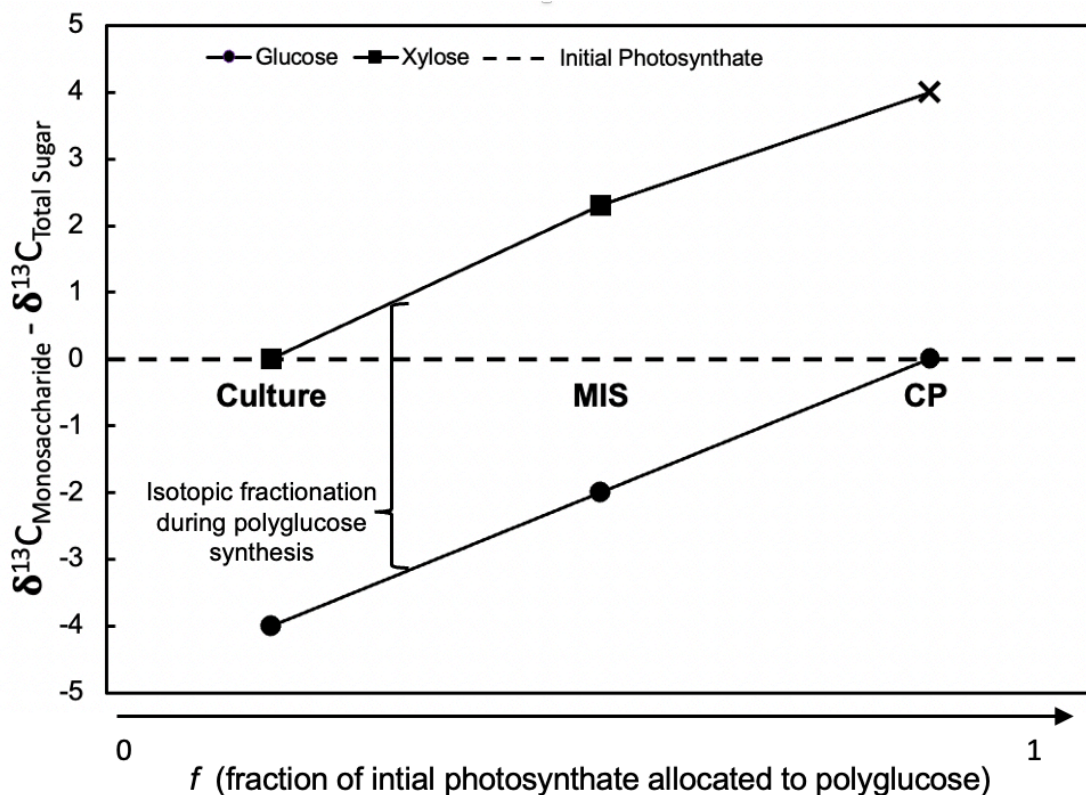
**Figure 4.3** Difference between measured individual monosaccharide $\delta^{13}C$ values and total sugar $\delta^{13}C$ values calculated from mass weighted averages of all individual sugars measured. Data shown are for a nutrient-replete cyanobacterial culture from Teece and Fogel (2007) ("Culture"), a cyanobacterial mat in a nutrient-limited and low-irradiance ecosystem ("MIS") and a cyanobacterial mat in a nutrient-limited and high-irradiance ecosystem ("CP"). The "x" datapoint for CP is estimated via the assumption that negligible precursor sugar is allocated to xylose.
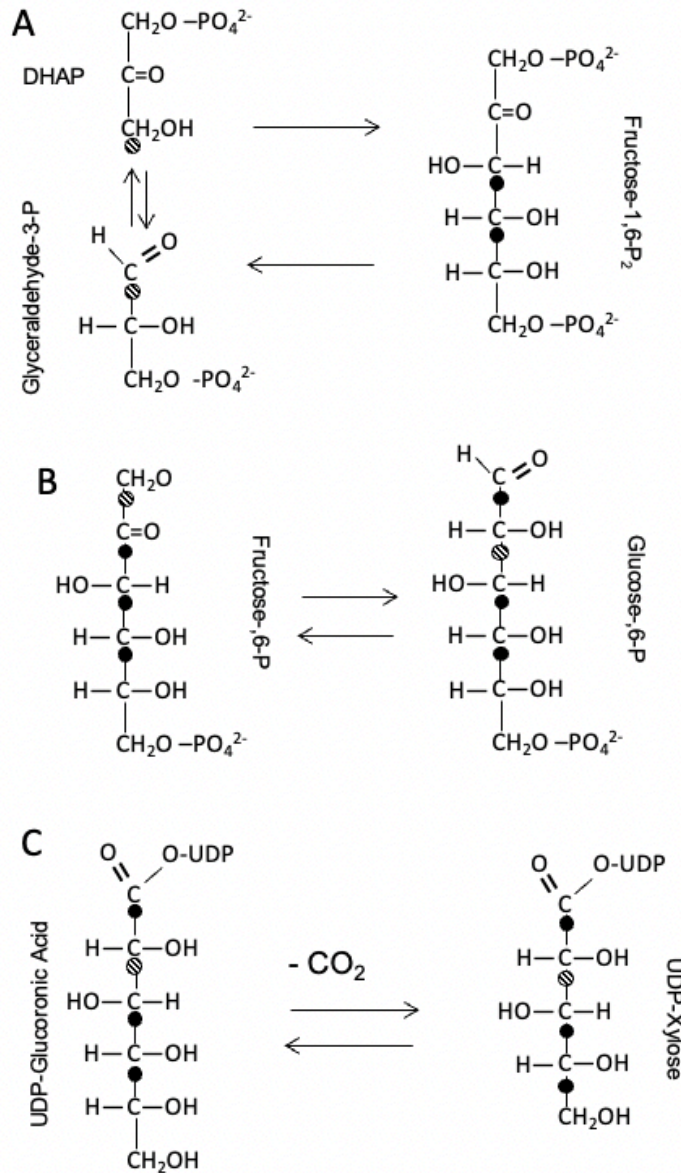
**Figure 4.4** Previously proposed mechanisms for generating xylose relatively depleted in $^{13}C$; letters correspond to enzymatic steps in figure 4.2. Black circles represent carbon atoms relatively $^{13}C$ enriched relative to the average of the molecule, striped circles represent carbon atoms relatively $^{13}C$ depleted. B: Modified from Gleixner and Schmidt 1997. The fructose biphosphate aldolase reaction cleaves the C-C bond between carbons 3 and 4 in the fructose to triose direction, resulting in a $^{13}C$ depletion in the corresponding carbons in the triose products and a relative $^{13}C$ enrichment in carbons 3 and 4 of fructose. C: Modified from Nam et al. 2022. At equilibrium, the glucose isomerase reaction results in glucose with a relative $^{13}C$ enrichment and depletion in carbons 1 and 2, respectively, due to the relative bond strengths of carbonyl and hydroxyl carbon bonds. D: Reactions B & C result in downstream products where carbons 1,3&4 are $^{13}C$ enriched compared to the average of the molecule. Carbon 6 is relatively $^{13}C$ depleted in comparison, and when this carbon is cleaved to form xylose it results in xylose that is relatively $^{13}C$-depleted compared to glucose.
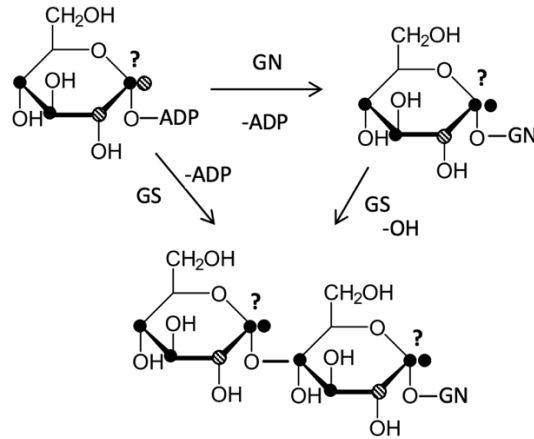
**Figure 4.5** Modified from Marr et al., (2022). Simplified pathway for the synthesis of glycogen (as a representative polyglucose) in bacteria. Black circles represent carbon atoms relatively $^{13}C$ enriched relative to the average of the molecule, striped circles represent carbon atoms relatively $^{13}C$ depleted, corresponding to the C atoms and mechanisms in figure 4.3. If there is a fractionation during polyglucose synthesis, the most likely position affected is the C-1 carbon, but the mechanism remains unclear.
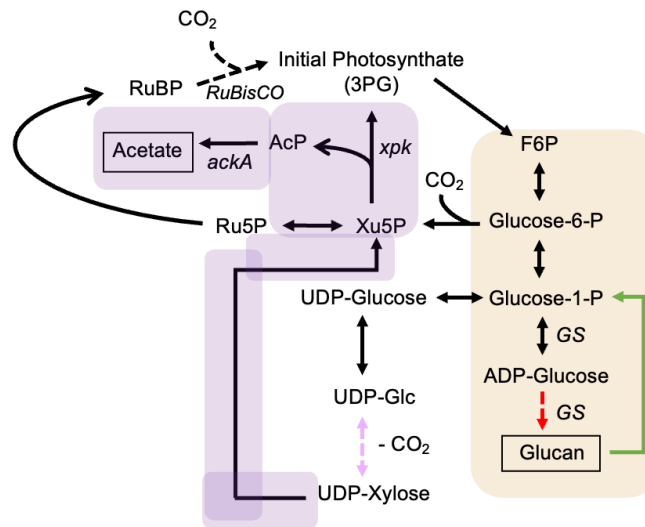


**Figure 4.6** Partial pathway for cyanobacterial glycogen synthesis and catabolism (orange shading) and the cyanobacterial phosphoketalase pathway (purple shading). Broken arrows represent reactions with confirmed or suspected carbon isotope fractionation. Pink and red colored arrows correspond to the pink and red colored arrows from Figure 4.1. The green arrow represents the recycling of glucans via hydrolysis and uptake. Products excreted extracellularly are indicated by boxes. Enzymes are italicized. *ackA*, acetate kinase; *xpk*, phosphoketolase; *RuBisCO*, Ribulose-1,5-bisphosphate carboxylase/oxygenase; *GS,* glycogen synthase. Glucose-1-P, glucose-1-phosphate; Glucose-6-P, glucose-6-phosphate; F6P, fructose-6-phosphate; Ru5P, ribulose-5-phosphate; RuBP, ribulose 1,5-bisphosphate; Xu5P, xylulose-5-phosphate; 3PG, 3-phosphoglycerate; AcP, acetylphosphate.

**Conclusions & Future directions**

Here we show that there is a systematic and quantifiable carbon isotopic difference between xylose (C5) and glucose (C6) monomers extracted from photosynthetic organisms. We propose that this fractionation is due to the irreversible formation of polyglucose compounds in both prokaryotic and eukaryotic organisms, and further hypothesize that the magnitude of this fractionation will depend on the degree of polyglucose recycling in different types of organisms and under different environmental conditions.

The main limitation in our analyses is the lack of taxonomically diverse data, especially for the Cyanobacteria. Subsequent work should isolate the primary variables of interest: polyglucose production and degree of recycling, as well as focus on obtaining data broadly across the Cyanobacteria and other photosynthetic microbial taxa. Both studies focused on one organism cultured under different conditions that promote variable polyglucose synthesis, and studies that sample wide physiological and taxonomic diversity would confirm whether a uniform KIE of 4‰ for this process ($\varepsilon_{xyl\text{-}glu}$, Figure 4.3) can be replicated under more strictly controlled experimental conditions.

**References**

Abelson P. H. and Hoering T. C. (1961) Carbon isotope fractionation in formation of amino acids by photosynthetic organisms. *Proc Natl Acad Sci U S A* **47**, 623–632.

Anderson K. L., Tayne T. a and Ward D. M. (1987) Formation and fate of fermentation products in hot spring cyanobacterial mats. *Appl Environ Microbiol* **53**, 2343–2352.

Arnosti C., Steen A. D., Ziervogel K., Ghobrial S. and Jeffrey W. H. (2011) Latitudinal gradients in degradation of marine dissolved organic carbon. *PLoS One* **6**, 8–13.

Ball S., Colleoni C., Cenci U., Raj J. N. and Tirtiaux C. (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J Exp Bot* **62**, 1775–1801.

Bigeleisen J. (1965) Chemistry of Isotopes. *Science (1979)* **147**, 463–471.

Blair N., Leu a, Muñoz E., Olsen J., Kwong E. and des Marais D. (1985) Carbon isotopic fractionation in heterotrophic microbial metabolism. *Appl Environ Microbiol* **50**, 996–1001.

Braakman R., Follows M. J. and Chisholm S. W. (2017) Metabolic evolution and the self-organization of ecosystems. *Proceedings of the National Academy of Sciences* **114**, E3091–E3100.

Cano M., Holland S. C., Artier J., Burnap R. L., Ghirardi M., Morgan J. A. and Yu J. (2018) Glycogen Synthesis and Metabolite Overflow Contribute to Energy Balancing in Cyanobacteria. *Cell Rep*.

Chuang D. S. W. and Liao J. C. (2021) Role of cyanobacterial phosphoketolase in energy regulation and glucose secretion under dark anaerobic and osmotic stress conditions. *Metab Eng* **65**, 255–262.

Cleland W. W. (2005) The use of isotope effects to determine enzyme mechanisms. *Arch Biochem Biophys* **433**, 2–12.

DeNiro M. J. . and Epstein S. (1977) Mechanism of Carbon Isotope Fractionation Associated with Lipid Synthesis Author ( s ): Michael J . DeNiro and Samuel Epstein Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/1744515 REFERENCES Li. *Science (1979)* **197**, 261–263.

Derrien D., Marol C., Balabane M. and Balesdent J. (2006) The turnover of carbohydrate carbon in a cultivated soil estimated by 13C natural abundances. *Eur J Soil Sci* **57**, 547–557.

van Dongen B. E., Schouten S. and Damsté J. S. S. (2001) Gas chromatography/combustion/isotope-ratio-monitoring mass spectrometric analysis of methylboronic derivatives of monosaccharides: A new method for determining natural 13C abundances of carbohydrates. *Rapid Communications in Mass Spectrometry* **15**, 496–500.

van Dongen B. E., Schouten S. and Sinninghe Damsté J. S. (2002) Carbon isotope variability in monosaccharides and lipids of aquatic algae and terrestrial plants. *Mar Ecol Prog Ser* **232**, 83–92.

Dungait J. A. J., Docherty G., Straker V. and Evershed R. P. (2008) Interspecific variation in bulk tissue, fatty acid and monosaccharide δ13C values of leaves from a mesotrophic grassland plant community. *Phytochemistry* **69**, 2041–2051.

Farquhar G. D., von Caemmerer S. and Berry J. A. (2001) Models of Photosynthesis. *Plant Physiol* **125**, 42–45.

Flemming H. C. and Wingender J. (2010) The biofilm matrix. *Nat Rev Microbiol* **8**, 623–633.

Fogg G. E. (1983) The Ecological Significance of Extracellular Products of Phytoplankton Photosynthesis. *Botanica Marina*.

Gilbert A., Robins R. J., Remaud G. S. and Tcherkez G. G. B. (2012) Intramolecular 13 C pattern in hexoses from autotrophic and heterotrophic C 3 plant tissues. **109**, 18204–18209.

Gilbert A., Silvestre V., Robins R. J., Tcherkez G. and Remaud G. S. (2011) A 13C NMR spectrometric method for the determination of intramolecular δ13C values in fructose from plant sucrose samples. *New Phytologist* **191**, 579–588.

Gleixner G. and Schmidt H. L. (1997) Carbon isotope effects on the fructose-1,6-bisphosphate aldolase reaction, origin for non-statistical 13C distributions in carbohydrates. *Journal of Biological Chemistry* **272**, 5382–5387.

Gonzalez-Nayeck A. C., Mohr W., Tang T., Sattin S., Parenteau M. N., Jahnke L. L. and Pearson A. (2022) Absence of canonical trophic levels in a microbial mat. *Geobiology* **20**, 726–740.

Halsey K. H., Jones B. and Jones B. M. (2014) Phytoplankton Strategies for Photosynthetic Energy Allocation.

Harvey H. R., Tuttle J. H. and Bell J. T. (1995) Article in Geochimica et Cosmochimica Acta. *Geochim Cosmochim Acta* **59**, 3367–3377.

Hayes J. M. (2001) Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes. *Rev Mineral Geochem* **43**, 225–277.

Hayes J. M., Freeman K. H., Popp B. N. and Hoham C. H. (1989) Compound specific isotope analysis, a novel tool for reconstruction of ancient biogeochemical processes. *Org Geochem* **16**, 1115–1128.

Hayes J. M., Strauss H. and Kaufman A. J. (1999) The abundance of 13C in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem Geol* **161**, 103–125.

Klock J. H., Wieland A., Seifert R. and Michaelis W. (2007) Extracellular polymeric substances (EPS) from cyanobacterial mats: Characterisation and isolation method optimisation. *Mar Biol* **152**, 1077–1085.

Klok J., Cox H. C., de Leeuw J. W. and Schenck P. A. (1982) Analysis of synthetic mixtures of partially methylated alditol acetates by capillary gas chromatography, gas chromatography-electron impact mass spectrometry and gas chromatography-chemical ionization mass spectrometry. *J Chromatogr A* **253**, 55–64.

Laws E. A. and McClellan S. A. (2022) Interactive effects of CO2, temperature, irradiance, and nutrient limitation on the growth and physiology of the marine cyanobacterium Synechococcus (Cyanophyceae). *J Phycol* **58**, 703–718.

Liu J. J., Zhang G. C., Kwak S., Oh E. J., Yun E. J., Chomvong K., Cate J. H. D. and Jin Y. S. (2019) Overcoming the thermodynamic equilibrium of an isomerization reaction through oxidoreductive reactions for biotransformation. *Nature Communications 2019 10:1* **10**, 1–8.

Macko S. A., Fogel M. L., Hare P. E. and Hoering T. C. (1987) Isotopic fractionation of nitrogen and carbon in the synthesis of amino acids by microorganisms. *Chemical Geology: Isotope Geoscience Section* **65**, 79–92.

Mahmoudi N., Beaupré S. R., Steen A. D. and Pearson A. (2017) Sequential bioavailability of sedimentary organic matter to heterotrophic bacteria. *Environ Microbiol* **19**, 2629–2644.

Makowka A., Nichelmann L., Schulze D., Spengler K., Wittmann C., Forchhammer K. and Gutekunst K. (2020) Glycolytic Shunts Replenish the Calvin-Benson-Bassham Cycle as Anaplerotic Reactions in Cyanobacteria. *Mol Plant* **13**, 471–482.

Marr L., Biswas D., Daly L. A., Browning C., Vial S. C. M., Maskell D. P., Hudson C., Bertrand J. A., Pollard J., Ranson N. A., Khatter H., Eyers C. E., Sakamoto K. and Zeqiraj E. (2022)

Mechanism of glycogen synthase inactivation and interaction with glycogenin. *Nature Communications 2022 13:1* **13**, 1–14.

van der Meer M. T. J., Schouten S., van Dongen B. E., Rijpstra W. I. C., Fuchs G., Sinninghe Damsté J. S., de Leeuw J. W. and Ward D. M. (2001) Biosynthetic Controls on the 13C Contents of Organic Components in the Photoautotrophic Bacterium Chloroflexus aurantiacus. *Journal of Biological Chemistry* **276**, 10971–10976.

Melzer E. and Schmidt H. L. (1987) Carbon Isotope Effects on the Pyruvate-Dehydrogenase Reaction and Their Importance for Relative C-13 Depletion in Lipids. *Journal of Biological Chemistry* **262**, 8159–8164.

Meyer K. M. and Kump L. R. (2008) Oceanic Euxinia in Earth History: Causes and Consequences.

Moers M. E. C., Jones D. M., Eakin P. A., Fallick A. E., Griffiths H. and Larter S. R. (1993) Carbohydrate diagenesis in hypersaline environments: application of GC-IRMS to the stable isotope analysis of derivatized saccharides from surficial and buried sediments. *Org Geochem* **20**, 927–933.

Monson K. D. and Hayes J. M. (1982) Carbon Isotopic Fractionation in the Biosynthesis of Bacterial Fatty-Acids - Ozonolysis of Unsaturated Fatty-Acids as a Means of Determining the Intramolecular Distribution of Carbon Isotopes. *Geochim Cosmochim Acta* **46**, 139–149.

Nam K. H. (2022) Glucose Isomerase: Functions, Structures, and Applications. *Applied Sciences 2022, Vol. 12, Page 428* **12**, 428.

de Niro, Michael, Epstein S. (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochim Cosmochim Acta* **42**, 495–506.

Nold S. C. and Ward D. M. (1996) Photosynthate partitioning and fermentation in hot spring microbial mat communities. *Appl Environ Microbiol* **62**, 4598–4607.

Polko J. K. and Kieber J. J. (2019) The Regulation of Cellulose Biosynthesis in Plants. *Plant Cell* **31**, 282.

Reinhard C. T., Planavsky N. J., Ward B. A., Love G. D., le Hir G. and Ridgwell A. (2020) The impact of marine nutrient abundance on early eukaryotic ecosystems. *Geobiology* **18**, 139–151.

Rossi F. and de Philippis R. (2015) Role of Cyanobacterial Exopolysaccharides in Phototrophic Biofilms and in Complex Microbial Mats. *Life* **5**, 1218.

Rossmann A., Butzenlechner M. and Schmidt H.-L. (1991) Evidence for a Nonstatistical Carbon Isotope Distribution in Natural Glucose. *Plant Physiol* **96**, 609–614.

Sakata S., Hayes J. M., McTaggart A. R., Evans R. A., Leckrone K. J. and Togasaki R. K. (1997) Carbon isotopic fractionation associated with lipid biosynthesis by a cyanobacterium: Relevance for interpretation of biomarker records. *Geochim Cosmochim Acta* **61**, 5379–5389.

Shinde S., Zhang X., Singapuri S. P., Kalra I., Liu X., Morgan-Kiss R. M. and Wang X. (2020) Glycogen metabolism supports photosynthesis start through the oxidative pentose phosphate pathway in cyanobacteria. *Plant Physiol* **182**, 507–517.

Simpson A. J., Kingery W. L., Spraul M., Dvortsak P. and Kerssebaum R. (2001) Research Separation of Structural Components in Soil Organic Matter by Diffusion Ordered Spectroscopy.

Smith A. M., Kruger N. J. and Lunn J. E. (2012) Source of sugar nucleotides for starch and cellulose synthesis. *Proceedings of the National Academy of Arts and Sciences* **109**, E776.

Smith J. A. and Bar-Peled M. (2017) Synthesis of UDP-apiose in Bacteria: The marine phototroph Geminicoccus roseus and the plant pathogen Xanthomonas pisi. *PLoS One* **12**, e0184953.

Stuart R. K., Mayali X., Lee J. Z., Everroad R. C., Hwang M., Bebout B. M., Weber P. K., Pettridge J. and Thelen M. P. (2015) Cyanobacterial reuse of extracellular organic carbon in microbial mats. **10**, 1240–1251.

Tcherkez G., Farquhar G., Badeck F., Ghashghaie J., Tcherkez G., Farquhar G., Badeck F. and Ghashghaie J. (2004) Theoretical considerations about carbon isotope distribution in glucose of C3 plants. *Functional Plant Biology* **31**, 857–877.

Teece M. A. and Fogel M. L. (2007) Stable carbon isotope biogeochemistry of monosaccharides in aquatic organisms and terrestrial plants. *Org Geochem* **38**, 458–473.

Tymoczko, John L, Berg, Jeremy M, Gatto, Gregory J, Stryer L. (2015) Biochemistry. , 1053.

Velmurugan R. and Incharoensakdi A. (2021) Overexpression of glucose-6-phosphate isomerase in Synechocystis sp. PCC 6803 with disrupted glycogen synthesis pathway improves exopolysaccharides synthesis. *Algal Res* **57**, 102357.

Viola R., Nyvall P. and Pederse¨n M. P. (2001) The unique features of starch metabolism in red algae.

Weijers J. W. H., Wiesenberg G. L. B., Bol R., Hopmans E. C. and Pancost R. D. (2010) Carbon isotopic composition of branched tetraether membrane lipids in soils suggest a rapid turnover and a heterotrophic life style of their source organism(s). *Biogeosciences* **7**, 2959–2973.

Xiong W., Lee T. C., Rommelfanger S., Gjersing E., Cano M., Maness P. C., Ghirardi M. and Yu J. (2015) Phosphoketolase pathway contributes to carbon metabolism in cyanobacteria. *Nature Plants 2015 2:1* **2**, 1–8.

Yoo S. H., Keppel C., Spalding M. and Jane J. lin (2007) Effects of growth condition on the structure of glycogen produced in cyanobacterium Synechocystis sp. PCC6803. *Int J Biol Macromol* **40**, 498–504.

## Chapter 5

## Conclusions

Within animal ecosystems, the $\delta^{13}C$ ratios of consumers generally conform to the principle "you are what you eat, +1‰." The initial goal of this thesis was to determine whether this metric applies to microbial systems where phagocytosis is minimal or absent and organisms assimilate a variety of carbon substrates via multiple different metabolic pathways. In chapter 2, we demonstrate that in a highly photic and oxygenated terrestrial microbial mat (Yellowstone National Park, USA), glucose moieties in exopolysaccharide are equal in $\delta^{13}C$ composition to both cyanobacterial and heterotrophic proteins. We hypothesized that this isotopic homogeneity reflects an ecosystem where producers (Cyanobacteria) and consumers (heterotrophic microorganisms) share primary photosynthate as a common resource.

Our conclusions in Chapter 2 led me to shift the goal of this thesis towards understanding the isotopic and environmental consequences of Cyanobacterial sugar synthesis, storage and excretion under different environmental conditions. In chapter 3, we characterize a benthic microbial mat from a suboxic sinkhole (Lake Huron, USA) with relative isotopic heterogeneity in the proteins of producers (Cyanobacteria), consumers (heterotrophic and mixotrophic microorganisms), and the glucose, arabinose and xylose moieties from exopolysaccharide. We hypothesize that in this system with a relatively lower photon flux and fluctuating oxidation states, internal pentose sugars act as substrate for Cyanobacterial fermentation, resulting in the excretion of organic compounds with heterogeneous isotopic compositions.

I propose a framework in which Cyanobacterial exopolysaccharide synthesis results in two pools of carbon: an external pool of exopolysaccharides with $\delta^{13}C$ values equal to or lower than initial photosynthetic carbon, and an internal pool of substrate sugars with relatively higher $\delta^{13}C$ values. This framework implies that greater exopolysaccharide production by Cyanobacteria results in lower overall organic carbon $\delta^{13}C$ compositions. In Chapter 4 I investigate the validity of this framework using the isotopic data from Chapters 2 and 3 as well as published $\delta^{13}C$ values for pentose and hexose sugars in autotrophic organisms. Using these data, I find that the likely magnitude of the proposed isotopic fractionation during polyglucose synthesis is approximately 4‰.

The most critical piece of information that would bolster the conclusions in this thesis is a mechanistic explanation for isotopic fractionation during polyglucose synthesis. A controlled laboratory culture study where substrate and polyglucose sugars are separated and isotopically characterized is an obvious next step towards obtaining this information.

More broadly, if this hypothesis is correct and the formation of polyglucose indeed carries a carbon kinetic isotope effect, we might expect that there is a quantitatively important pool of global fixed carbon that is $^{13}C$-depleted relative to standing biomass or initial photosynthate. This implies that environmental controls on cyanobacterial polysaccharide synthesis, such as nutrient availability, also affect local organic carbon $\delta^{13}C$ compositions.

# Appendix A

## Supporting information for Chapter 2



| | |
|---|---|
| 10 | 20 |
| 9 | 19 |
| 8 | 18 |
| 7 | 17 |
| 6 | 16 |
| 5 | 15 |
| 4 | 14 |
| 3 | 13 |
| 2 | 12 |
| 1 | 11 |

Figure S1.(Left) Photograph of SAX Fractions showing visible pigments of extracted proteins. (Right) Colors from image on left at full brightness.
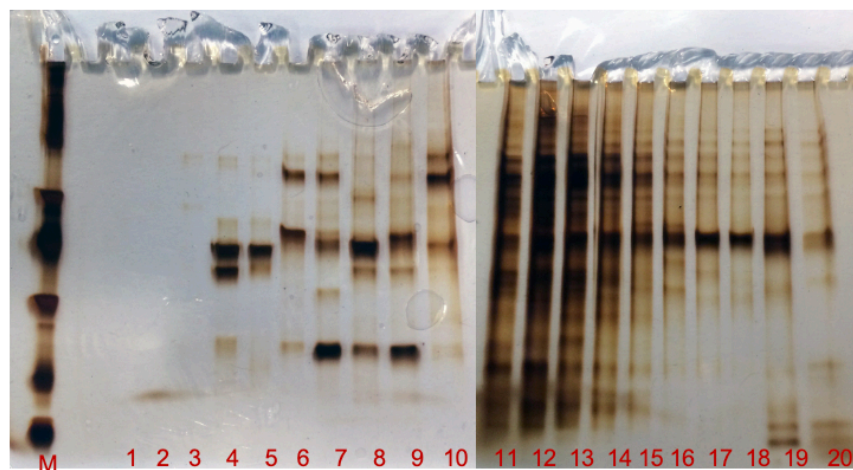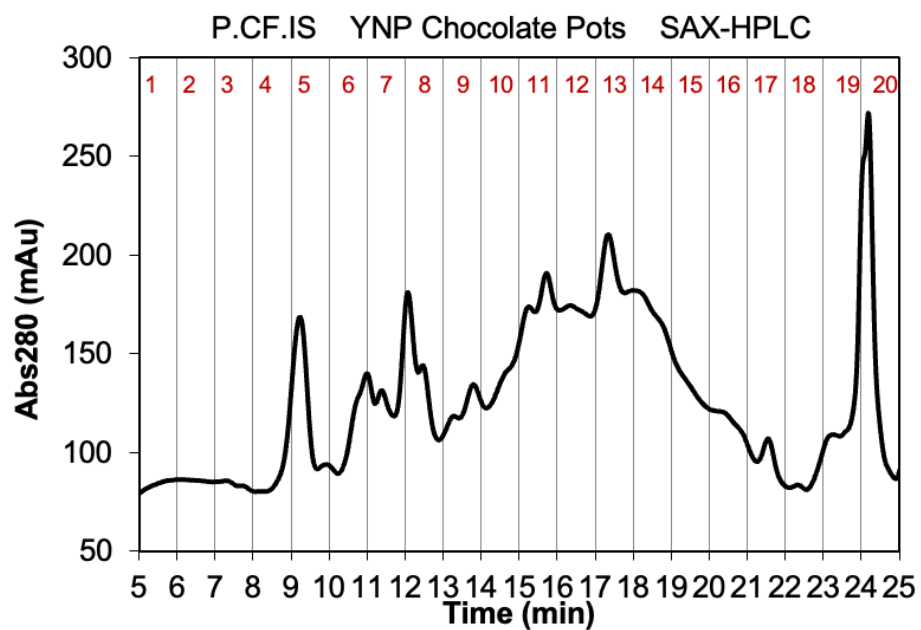
**Figure S2.** (Top) RP-HPLC chromatogram showing fractions collected for further separation by SAX chromatography. (Bottom) Aliquots of each SAX fraction analyzed by SDS-PAGE gel electrophoresis with silver stain.
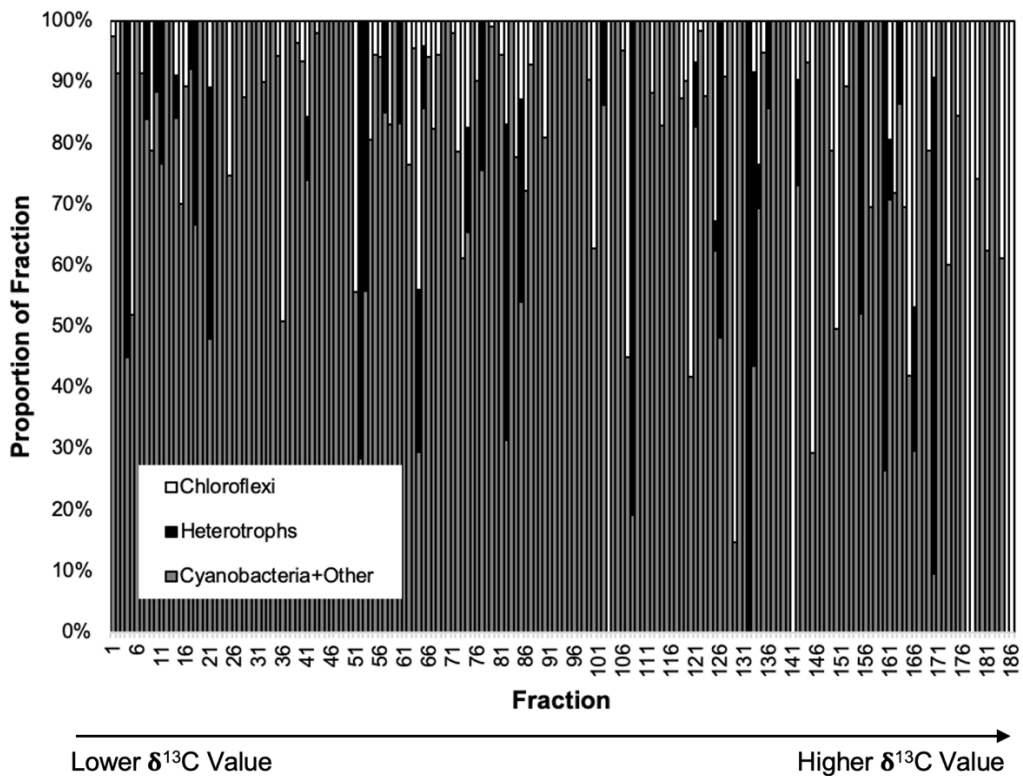
**Figure S3.** Proportions of taxonomic groups in all fractions that contained > 0.56 nmol C/μL and had triplicate SD < 2‰. Fractions are ordered from lowest $\delta^{13}C$ value (left) to highest $\delta^{13}C$ value (right).
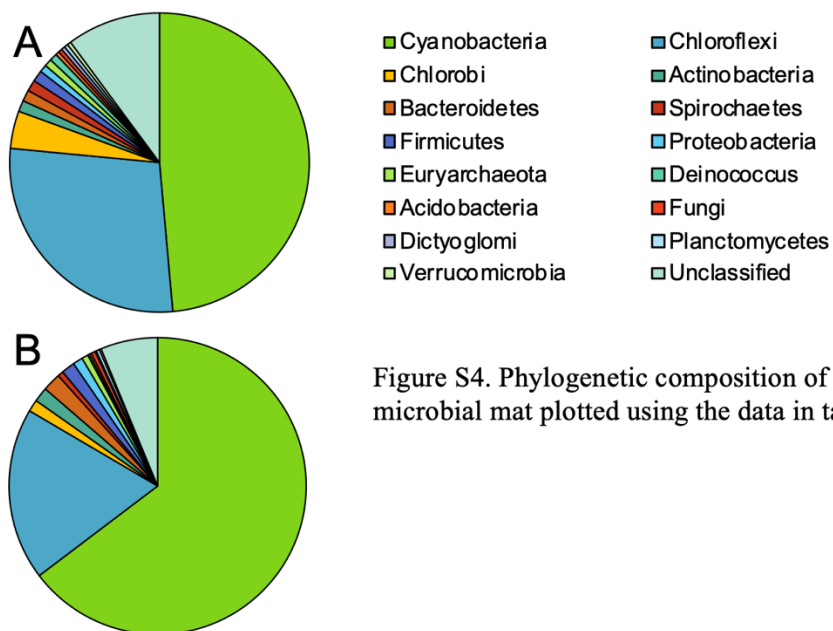


Figure S4. Phylogenetic composition of the CP microbial mat plotted using the data in table S2
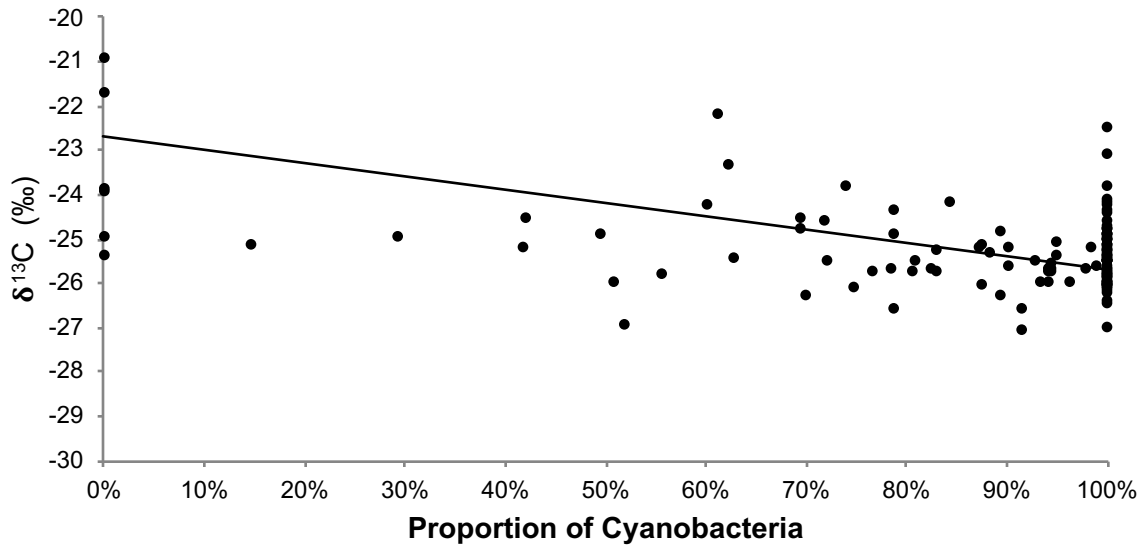
**Figure S5.** Result of two-component geometric (Deming) regression used to estimate the end-member $\delta^{13}C$ values for wells composed solely (> 95%) of mixtures of Cyanobacteria and Chloroflexi.

**Figure S6.** 16 plots below. Proportions of individual amino acids (AAs) in all fractions that contained > 0.56 nmol C/µL and had triplicate SD < 2‰, plotted against fraction $\delta^{13}C$ composition. See Table S6 for further statistics, and the Data analysis portion of the Methods section of the manuscript for additional information.

Proportion Ala vs. δ¹³C of Well

$y = 0.11x - 25.37$
$R^2 = 0.00$

Proportion His vs. δ¹³C of Well

$y = 10.53x - 25.54$
$R^2 = 0.02$

Proportion Arg vs. δ¹³C of Well

$y = -0.24x - 25.34$
$R^2 = 0.00$

Proportion Leu & Ile vs. δ¹³C of Well

$y = -0.65x - 25.28$
$R^2 = 0.00$

Proportion Tyr vs. δ¹³C of Well

$$y = -11.97x - 24.98$$
$$R^2 = 0.01$$

Proportion Thr vs. δ¹³C of Well

$$y = 0.40x - 25.38$$
$$R^2 = 0.00$$

Proportion Phe vs. δ¹³C of Well

$$y = -6.15x - 25.12$$
$$R^2 = 0.00$$

Proportion Ser vs. δ¹³C of Well

$$y = -3.92x - 25.17$$
$$R^2 = 0.01$$

Proportion Val vs. δ¹³C of Well

$y = 0.60x - 25.39$
$R^2 = 0.00$



Proportion Gly vs. δ¹³C of Well

$y = 3.77x - 25.64$
$R^2 = 0.01$



Proportion Trp vs. δ¹³C of Well

$y = -12.92x - 25.22$
$R^2 = 0.01$



Proportion Met vs. δ¹³C of Well

$y = -7.65x - 25.20$
$R^2 = 0.00$

Proportion Pro vs. δ¹³C of Well

y = -1.37x - 25.29
R² = 0.00

Proportion Cys vs. δ¹³C of Well

y = 13.36x - 25.47
R² = 0.00

# Appendix B

## Supporting Information for Chapter 3

**Figure S1.** Histogram of $\delta^{13}C$ values for all SAX fractions measured from sample LH47, separated into quartiles by decreasing IRMS peak area. Values are not normally distributed (Shapiro–Wilk test, $p < 0.01$)
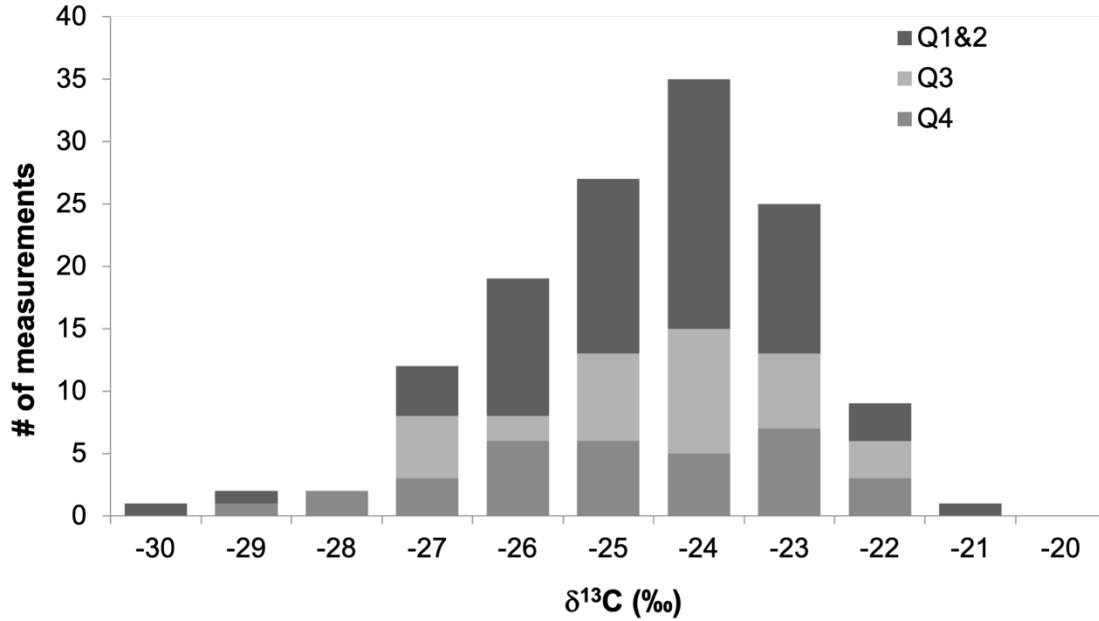


**Table S1.** Unique microbial proteins (n = 1188) detected by Orbitrap-MS/MS sequencing per RP fraction containing sufficient carbon for taxonomic identification.

| RP Fraction | unique peptides | Summed peptide peak area | Protein accession number | Gene product | Phylum |
|---|---|---|---|---|---|
| P.CW.IS03.1D | 5 | 9478386.469 | 3300002026_MIS_1000043442 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS03.1D | 1 | 30494990 | Ga0308317_144015415 | putative zinc ribbon protein | Chloroflexi |
| P.CW.IS03.1D | 2 | 12161883.5 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS03.1D | 1 | 6646202 | Ga0308317_10695267 | hypothetical protein | Chloroflexi |
| P.CW.IS03.1D | 1 | 5063847.5 | Ga0308317_129817518 | fumarate hydratase class II | Chloroflexi |
| P.CW.IS03.1D | 1 | 3917684.75 | 3300002026_MIS_100005419 | Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, alpha subunit | Chloroflexi |
| P.CW.IS03.1D | 1 | 3668450.688 | Ga0308317_106023111 | thiol peroxidase | Chloroflexi |
| P.CW.IS03.1D | 3 | 301852596.3 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1D | 2 | 96269350.91 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1D | 3 | 92307107.84 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1D | 1 | 68928208 | Ga0308317_122782710 | phycocyanin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 2 | 55221687 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS03.1D | 1 | 54410908 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS03.1D | 2 | 44313044.31 | Ga0308317_103184713 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1D | 1 | 33968028 | Ga0308317_122782739 | uncharacterized protein DUF3386 | Cyanobacteria |
| P.CW.IS03.1D | 3 | 30098812.31 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 2 | 25880644.88 | 3300002024_MIS_10023532 | Uncharacterized conserved protein - COG3937 | Cyanobacteria |
| P.CW.IS03.1D | 2 | 23068050 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS03.1D | 1 | 20496980 | Ga0308317_11304452 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 4 | 19689538.88 | Ga0308317_108417113 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS03.1D | 1 | 14584898 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1D | 2 | 10808367.44 | 3300002026_MIS_100192369 | Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family | Cyanobacteria |
| P.CW.IS03.1D | 2 | 10465668.5 | 3300002027_MIS_101479931 | Co-chaperonin GroES (HSP10) | Cyanobacteria |
| P.CW.IS03.1D | 3 | 8053181.75 | 3300002026_MIS_100054824 | NA | Cyanobacteria |
| P.CW.IS03.1D | 3 | 7464153.063 | Ga0308317_122782711 | phycocyanin alpha chain | Cyanobacteria |
| P.CW.IS03.1D | 1 | 5288434.5 | 3300002026_MIS_100321086 | F0F1-type ATP synthase, alpha subunit | Cyanobacteria |
| P.CW.IS03.1D | 1 | 4558649 | Ga0308317_113580413 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 2 | 4427941.688 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS03.1D | 1 | 3894237.75 | 3300002026_MIS_1001557614 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS03.1D | 1 | 3413473.5 | Ga0308317_14375973 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 8 | 249934170.9 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 1 | 66192434.25 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1D | 4 | 62962752.69 | 3300002027_MIS_100632912 | NA | Proteobacteria |
| P.CW.IS03.1D | 1 | 40047951 | Ga0308317_14352243 | acetyl-CoA synthetase | Proteobacteria |
| P.CW.IS03.1D | 3 | 23799291.34 | 3300002026_MIS_100072245 | Enolase | Proteobacteria |
| P.CW.IS03.1D | 6 | 15136087.5 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 4 | 14940344 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 4 | 14712202.5 | Ga0308317_14549566 | methyl-accepting chemotaxis protein/methyl-accepting chemotaxis protein | Proteobacteria |
| P.CW.IS03.1D | 1 | 11576324.69 | 3300002026_MIS_100319374 | NA | Proteobacteria |
| P.CW.IS03.1D | 3 | 9883926.5 | 3300002027_MIS_101963794 | NA | Proteobacteria |
| P.CW.IS03.1D | 2 | 9753531.938 | 3300002027_MIS_100755711 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS03.1D | 1 | 9734126 | Ga0308317_14797482 | rubredoxin-NAD+ reductase | Proteobacteria |
| P.CW.IS03.1D | 1 | 9315774 | 3300002026_MIS_100320112 | Fructose-1,6-bisphosphatase | Proteobacteria |
| P.CW.IS03.1D | 2 | 9041084.625 | Ga0308317_14129511 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1D | 1 | 9024177 | 3300002026_MIS_100005424 | NA | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1D | 3 | 8004448.5 | 3300002027_MIS_101006588 | NA | Proteobacteria |
| P.CW.IS03.1D | 2 | 6929809.5 | Ga0308317_15366783 | CspA family cold shock protein | Proteobacteria |
| P.CW.IS03.1D | 1 | 6867281.5 | Ga0308317_13112285 | EF hand domain-containing protein/EF hand domain-containing protein/EF hand domain-containing protein | Proteobacteria |
| P.CW.IS03.1D | 2 | 6396362 | Ga0308317_10164838 | trigger factor | Proteobacteria |
| P.CW.IS03.1D | 1 | 5919190 | Ga0308317_10281003 | signal transduction histidine kinase | Proteobacteria |
| P.CW.IS03.1D | 1 | 4539907.5 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1D | 1 | 4514909 | Ga0308317_11378883 | uncharacterized protein DUF4124 | Proteobacteria |
| P.CW.IS03.1D | 4 | 4395287.125 | Ga0308317_101705613 | chaperonin GroEL | Proteobacteria |
| P.CW.IS03.1D | 2 | 4089303.875 | Ga0308317_10016455 | CspA family cold shock protein | Proteobacteria |
| P.CW.IS03.1D | 1 | 3339703.5 | Ga0308317_12989787 | fructose-1,6-bisphosphatase I | Proteobacteria |
| P.CW.IS03.1D | 3 | 21811154.5 | 3300002027_MIS_100245793 | NA | Thermotogae |
| P.CW.IS03.1D | 1 | 228550880 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1D | 2 | 13225169.85 | 3300002026_MIS_100309382 | ABC-type branched-chain amino acid transport systems, periplasmic component | Unclassified |
| P.CW.IS03.1D | 3 | 12520994.19 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1D | 2 | 11176424 | 3300002026_MIS_100315161 | ATPases with chaperone activity, ATP-binding subunit | Unclassified |
| P.CW.IS03.1D | 1 | 6358878.5 | 3300002026_MIS_1000366010 | ABC-type sugar transport system, periplasmic component | Unclassified |
| P.CW.IS03.1D | 1 | 5990503.281 | Ga0308317_11002012 | elongation factor Tu | Unclassified |
| P.CW.IS03.1D | 1 | 5123254 | 3300002024_MIS_10944721 | Transketolase, thiamine diphosphate binding domain | Unclassified |
| P.CW.IS03.1D | 2 | 4277041 | Ga0308317_10604841 | peptidyl-prolyl cis-trans isomerase C | Unclassified |
| P.CW.IS03.1D | 3 | 4276620.813 | Ga0308317_10157129 | propionyl-CoA synthetase | Unclassified |
| P.CW.IS03.1D | 1 | 3453440.5 | Ga0308317_12144224 | nitrile hydratase alpha subunit | Unclassified |
| P.CW.IS03.1D | 1 | 1006402880 | Ga0308317_14757724 | hypothetical protein | Bacteroidetes |
| P.CW.IS03.1D | 2 | 122986346.8 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS03.1D | 15 | 4174922458 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 1 | 204426620.9 | Ga0308317_10589352 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 2 | 68831038.23 | Ga0308317_12399522 | sulfate adenylyltransferase | Proteobacteria |
| P.CW.IS03.1D | 1 | 43884580 | 3300002026_MIS_100199952 | Uncharacterized conserved protein - COG1262 | Spirochaetes |
| P.CW.IS03.1D | 1 | 149212816 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1D | 1 | 81209272 | 3300002027_MIS_100852841 | Protein kinase domain | Unclassified |
| P.CW.IS03.1D | 3 | 16783199.63 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS03.1D | 1 | 2496386 | Ga0308317_14023463 | transcriptional regulator with XRE-family HTH domain | Chloroflexi |
| P.CW.IS03.1D | 1 | 1805394.875 | Ga0308317_14165545 | phosphoglycerate kinase | Chloroflexi |
| P.CW.IS03.1D | 3 | 53572646.19 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1D | 3 | 16460200.19 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1D | 4 | 8230236.828 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1D | 1 | 6838420.5 | Ga0308317_14972441 | phosphoglycerate kinase | Cyanobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1D | 2 | 5454707.688 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS03.1D | 2 | 4365694.5 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS03.1D | 2 | 4105738.5 | Ga0308317_108417113 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS03.1D | 1 | 3396297 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS03.1D | 2 | 3309573.5 | 3300002026_MIS_100078659 | Predicted periplasmic or secreted lipoprotein | Cyanobacteria |
| P.CW.IS03.1D | 2 | 2124283.125 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 2 | 1849004.234 | 3300002026_MIS_1003608920 | NA | Cyanobacteria |
| P.CW.IS03.1D | 3 | 106629883.9 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 7 | 44457261.83 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 2 | 5837224.422 | Ga0308317_13495863 | tRNA 2-thiouridine synthesizing protein C | Proteobacteria |
| P.CW.IS03.1D | 1 | 5682499 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1D | 2 | 5278766.063 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1D | 1 | 4473249 | 3300002026_MIS_100307544 | NA | Proteobacteria |
| P.CW.IS03.1D | 1 | 2106484.75 | 3300002027_MIS_100094537 | NA | Proteobacteria |
| P.CW.IS03.1D | 2 | 1504023.734 | Ga0308317_10632712 | branched-chain amino acid transport system substrate-binding protein | Proteobacteria |
| P.CW.IS03.1D | 1 | 164287941.4 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1D | 3 | 4396303.719 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1D | 1 | 3852855.5 | Ga0308317_131278014 | GMP reductase | Unclassified |
| P.CW.IS03.1D | 1 | 2237217.75 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS03.1D | 1 | 1784967.125 | 3300002024_MIS_10666431 | NA | Unclassified |
| P.CW.IS03.1D | 2 | 1885016.609 | Ga0308317_10157129 | propionyl-CoA synthetase | Unclassified |
| P.CW.IS03.1D | 1 | 1488727.625 | Ga0308317_15263983 | cohesin domain-containing protein | Unclassified |
| P.CW.IS03.1D | 2 | 2782387.484 | 3300002027_MIS_101644101 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS03.1D | 3 | 4669827.063 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS03.1D | 2 | 1161989.781 | Ga0308317_129817518 | fumarate hydratase class II | Chloroflexi |
| P.CW.IS03.1D | 5 | 56036892.81 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1D | 3 | 39743572.55 | Ga0308317_12856282 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS03.1D | 4 | 9536782.781 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1D | 2 | 7687101.75 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1D | 2 | 3498045.375 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS03.1D | 2 | 1609484.625 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 1 | 1576991.75 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS03.1D | 1 | 1359786.172 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS03.1D | 7 | 76666178.95 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 1 | 3425519.75 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1D | 1 | 2634329.5 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1D | 1 | 144580482.9 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1D | 1 | 4300110 | 3300002024_MIS_11466371 | 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase | Unclassified |
| P.CW.IS03.1D | 3 | 2545980.75 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1D | 1 | 1271972 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS03.1D | 4 | 2165336.227 | 3300002026_MIS_1000043442 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS03.1D | 1 | 2063725.375 | 3300002024_MIS_11548661 | NA | Bacillariophyta |
| P.CW.IS03.1D | 2 | 1537327.656 | 3300002026_MIS_1000043414 | PsaD | Bacillariophyta |
| P.CW.IS03.1D | 1 | 45225864 | Ga0308317_127113013 | SusD-like starch-binding protein associating with outer membrane/putative outer membrane starch-binding protein | Bacteroidetes |
| P.CW.IS03.1D | 2 | 5687721 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS03.1D | 1 | 4403724.875 | Ga0308317_10340078 | Fe-Mn family superoxide dismutase | Chloroflexi |
| P.CW.IS03.1D | 3 | 2414955.781 | Ga0308317_129817518 | fumarate hydratase class II | Chloroflexi |
| P.CW.IS03.1D | 2 | 1885060.063 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS03.1D | 2 | 1624403.969 | Ga0308317_14165545 | phosphoglycerate kinase | Chloroflexi |
| P.CW.IS03.1D | 1 | 1282639.25 | Ga0308317_141014330 | hypothetical protein | Chloroflexi |
| P.CW.IS03.1D | 1 | 1225955 | Ga0308317_14401547 | acetyl-CoA C-acetyltransferase | Chloroflexi |
| P.CW.IS03.1D | 1 | 92578272 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1D | 6 | 59598608.47 | Ga0308317_123538692 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1D | 4 | 50197911.5 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1D | 3 | 44587020.5 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS03.1D | 4 | 41373715.83 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS03.1D | 2 | 40786326.2 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1D | 2 | 32848182.69 | Ga0308317_15402722 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1D | 4 | 19284731.41 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS03.1D | 2 | 11271310.5 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS03.1D | 1 | 6200991 | 3300002027_MIS_100251133 | NA | Cyanobacteria |
| P.CW.IS03.1D | 2 | 5564826.75 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS03.1D | 2 | 5372794.5 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 1 | 4138505.75 | 3300002027_MIS_100067606 | Rubredoxin | Cyanobacteria |
| P.CW.IS03.1D | 2 | 3622914.5 | Ga0308317_10126032 | hypothetical protein | Cyanobacteria |
| P.CW.IS03.1D | 2 | 3301129.563 | Ga0308317_11304452 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 1 | 3155880.5 | Ga0308317_13177228 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS03.1D | 4 | 2959230.313 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1D | 2 | 2597423.25 | Ga0308317_113580413 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1D | 2 | 2184811.797 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS03.1D | 1 | 2017492.625 | 3300002027_MIS_101906972 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS03.1D | 2 | 1974010.766 | Ga0308317_122782739 | uncharacterized protein DUF3386 | Cyanobacteria |
| P.CW.IS03.1D | 3 | 1687525.656 | Ga0308317_14361226 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS03.1D | 3 | 1386593.047 | Ga0308317_15235161 | SdrD B-like protein | Cyanobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1D | 3 | 1385423.344 | Ga0308317_122782711 | phycocyanin alpha chain | Cyanobacteria |
| P.CW.IS03.1D | 1 | 1348130 | Ga0308317_141694522 | hypothetical protein | Cyanobacteria |
| P.CW.IS03.1D | 8 | 73443006.23 | 3300002027_MIS_100632912 | NA | Proteobacteria |
| P.CW.IS03.1D | 4 | 61901368.5 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1D | 5 | 10387284.5 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 1 | 5984049.5 | 3300002026_MIS_100320112 | Fructose-1,6-bisphosphatase | Proteobacteria |
| P.CW.IS03.1D | 2 | 5347028.125 | Ga0308317_13169602 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1D | 1 | 3781292.5 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1D | 2 | 3081171.5 | Ga0308317_15288274 | uncharacterized protein (DUF697 family) | Proteobacteria |
| P.CW.IS03.1D | 2 | 2765066.125 | Ga0308317_10491392 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1D | 2 | 2360927.063 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 1 | 1703365.25 | 3300002026_MIS_100307544 | NA | Proteobacteria |
| P.CW.IS03.1D | 1 | 1366473.75 | Ga0308317_13996212 | phage shock protein A | Proteobacteria |
| P.CW.IS03.1D | 1 | 1297128 | 3300002026_MIS_100072245 | Enolase | Proteobacteria |
| P.CW.IS03.1D | 4 | 1294977.203 | 3300002027_MIS_100225507 | Adenosine-5'-phosphosulfate reductase beta subunit; | Proteobacteria |
| P.CW.IS03.1D | 2 | 2887442.719 | 3300002027_MIS_100245793 | NA | Thermotogae |
| P.CW.IS03.1D | 2 | 155173876.9 | Ga0308317_13886182 | actin beta/gamma 1 | Unclassified |
| P.CW.IS03.1D | 5 | 14016238.58 | 3300002024_MIS_10944721 | Transketolase, thiamine diphosphate binding domain | Unclassified |
| P.CW.IS03.1D | 3 | 10741222.25 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1D | 3 | 5402563.75 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1D | 2 | 2881927.25 | 3300002027_MIS_101184422 | Rubredoxin | Unclassified |
| P.CW.IS03.1D | 2 | 1536015.813 | 3300002027_MIS_101091071 | NA | Unclassified |
| P.CW.IS03.1D | 1 | 1454057.125 | Ga0308317_14432284 | hypothetical protein | Unclassified |
| P.CW.IS03.1D | 2 | 1242673.859 | Ga0308317_11156942 | hypothetical protein | Unclassified |
| P.CW.IS03.1D | 1 | 3144531.75 | Ga0308317_12453813 | hypothetical protein | Unclassified |
| P.CW.IS03.1D | 3 | 2331692 | Ga0308317_13508831 | peptidyl-prolyl cis-trans isomerase C | Unclassified |
| P.CW.IS03.1D | 2 | 36770226.56 | 3300002027_MIS_101946932 | CP12 domain | Cyanobacteria |
| P.CW.IS03.1D | 2 | 33254549.78 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1D | 2 | 6672166.5 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1D | 3 | 51659728.3 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1D | 1 | 22184640 | Ga0308317_130754517 | polyribonucleotide nucleotidyltransferase | Proteobacteria |
| P.CW.IS03.1D | 2 | 4711733.063 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1D | 1 | 53786548 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1D | 1 | 24761050 | 3300002026_MIS_10000009137 | Ribosomal protein L15 | Unclassified |
| P.CW.IS03.1E | 4 | 2521806.078 | 3300002027_MIS_101644102 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS03.1E | 2 | 3073847.25 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS03.1E | 1 | 2265003.5 | Ga0308317_129817518 | fumarate hydratase class II | Chloroflexi |
| P.CW.IS03.1E | 1 | 1668960.375 | Ga0308317_14165545 | phosphoglycerate kinase | Chloroflexi |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1E | 1 | 897186.25 | Ga0308317_10733153 | hypothetical protein | Chloroflexi |
| P.CW.IS03.1E | 2 | 20419256 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS03.1E | 4 | 9895270.375 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1E | 4 | 7586955.891 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1E | 1 | 2304112.75 | Ga0308317_141694522 | hypothetical protein | Cyanobacteria |
| P.CW.IS03.1E | 3 | 2208120.398 | Ga0308317_103184713 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1E | 2 | 1217377.625 | Ga0308317_10501821 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1E | 3 | 1180279.75 | Ga0308317_123538692 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1E | 3 | 1152303 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1E | 2 | 852919.6875 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS03.1E | 2 | 809934.3438 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1E | 2 | 796923.0469 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1E | 5 | 32134961.15 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 3 | 3824328.883 | 3300002027_MIS_100632912 | NA | Proteobacteria |
| P.CW.IS03.1E | 1 | 3737824.75 | 3300002027_MIS_101872221 | Superfamily II DNA/RNA helicases, SNF2 family | Proteobacteria |
| P.CW.IS03.1E | 2 | 3649950.023 | Ga0308317_13495863 | tRNA 2-thiouridine synthesizing protein C | Proteobacteria |
| P.CW.IS03.1E | 2 | 2774225 | Ga0308317_14797482 | rubredoxin-NAD+ reductase | Proteobacteria |
| P.CW.IS03.1E | 2 | 2401006.594 | Ga0308317_14352243 | acetyl-CoA synthetase | Proteobacteria |
| P.CW.IS03.1E | 3 | 1339260.328 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 1 | 1018892.563 | Ga0308317_11579021 | elongation factor G | Proteobacteria |
| P.CW.IS03.1E | 2 | 1002834.992 | 3300002027_MIS_100157287 | NA | Proteobacteria |
| P.CW.IS03.1E | 3 | 961565.6953 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 1 | 761709.875 | Ga0308317_11714423 | tRNA 2-thiouridine synthesizing protein C | Proteobacteria |
| P.CW.IS03.1E | 1 | 727584.4375 | 3300002026_MIS_100005424 | NA | Proteobacteria |
| P.CW.IS03.1E | 2 | 699894.9375 | 3300002026_MIS_100072245 | Enolase | Proteobacteria |
| P.CW.IS03.1E | 2 | 632210.0938 | Ga0308317_14549566 | methyl-accepting chemotaxis protein/methyl-accepting chemotaxis protein | Proteobacteria |
| P.CW.IS03.1E | 1 | 42552724 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1E | 1 | 16537650 | 3300002024_MIS_10666431 | NA | Unclassified |
| P.CW.IS03.1E | 4 | 3784343.75 | 3300002026_MIS_100315161 | ATPases with chaperone activity, ATP-binding subunit | Unclassified |
| P.CW.IS03.1E | 3 | 1690919.625 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1E | 1 | 1336426.625 | 3300002027_MIS_101311571 | NA | Unclassified |
| P.CW.IS03.1E | 3 | 8607453.895 | Ga0308317_14286762 | photosystem I subunit 7 | Unclassified |
| P.CW.IS03.1E | 3 | 2815654 | Ga0308317_10157129 | propionyl-CoA synthetase | Unclassified |
| P.CW.IS03.1E | 2 | 1036075.813 | Ga0308317_12144224 | nitrile hydratase alpha subunit | Unclassified |
| P.CW.IS03.1E | 1 | 43754316 | Ga0308317_12083728 | threonine/homoserine/homoserine lactone efflux protein | Bacteroidetes |
| P.CW.IS03.1E | 11 | 554493064.8 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 1 | 37239580 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1E | 1 | 3923325.75 | 3300002027_MIS_101644101 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS03.1E | 2 | 24119759.13 | Ga0308317_10358796 | CspA family cold shock protein | Bacteroidetes |
| P.CW.IS03.1E | 2 | 9195021 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS03.1E | 2 | 3684997.125 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS03.1E | 5 | 18103669.06 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS03.1E | 4 | 12513419.81 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1E | 2 | 11232128.19 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS03.1E | 1 | 10994375 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1E | 5 | 6578709.438 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS03.1E | 1 | 6068547.5 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1E | 2 | 4850955.258 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1E | 2 | 4788013 | 3300002026_MIS_1003608920 | NA | Cyanobacteria |
| P.CW.IS03.1E | 6 | 3732648.453 | Ga0308317_10084168 | ferredoxin | Cyanobacteria |
| P.CW.IS03.1E | 3 | 3192360.086 | 3300002027_MIS_100389562 | Uncharacterized protein conserved in bacteria | Cyanobacteria |
| P.CW.IS03.1E | 2 | 2278495.828 | 3300002027_MIS_100694095 | Chaperonin GroEL (HSP60 family) | Cyanobacteria |
| P.CW.IS03.1E | 8 | 379110788.1 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 5 | 29712209.88 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 1 | 27349875.22 | Ga0308317_12399522 | sulfate adenylyltransferase | Proteobacteria |
| P.CW.IS03.1E | 2 | 10683119 | 3300002027_MIS_101881042 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1E | 1 | 9133262 | 3300002026_MIS_100307544 | NA | Proteobacteria |
| P.CW.IS03.1E | 4 | 8952357.938 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1E | 1 | 23828012 | Ga0308317_12558412 | Tfp pilus assembly protein PilF | Spirochaetes |
| P.CW.IS03.1E | 1 | 160346960 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1E | 1 | 20990500 | 3300002024_MIS_10711273 | Aminotransferase class-III | Unclassified |
| P.CW.IS03.1E | 1 | 9159523 | 3300002027_MIS_100513031 | Ca2+-binding protein (EF-Hand superfamily) | Unclassified |
| P.CW.IS03.1E | 3 | 2996374.141 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1E | 1 | 2587109.5 | 3300002024_MIS_10472321 | NA | Unclassified |
| P.CW.IS03.1E | 2 | 18176578.56 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS03.1E | 3 | 9852782.82 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1E | 11 | 393494347.7 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 2 | 11334589.69 | Ga0308317_12399522 | sulfate adenylyltransferase | Proteobacteria |
| P.CW.IS03.1E | 1 | 4784255.5 | 3300002026_MIS_100307544 | NA | Proteobacteria |
| P.CW.IS03.1E | 2 | 3259289.125 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1E | 1 | 88922467.13 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1E | 1 | 4233057.5 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS03.1E | 1 | 1057193 | 3300002027_MIS_101644101 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS03.1E | 2 | 62716973.91 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1E | 2 | 6894270.063 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |

| P.CW.IS03.1E | 1 | 6259890.25 | Ga0308317_15402722 | CP12 domain-containing protein | Cyanobacteria |
|---|---|---|---|---|---|
| P.CW.IS03.1E | 2 | 3503142.5 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS03.1E | 3 | 2566545.125 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1E | 3 | 2484701.5 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1E | 1 | 2099888.25 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS03.1E | 1 | 1540790 | Ga0308317_11304452 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1E | 5 | 15281290.55 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 2 | 5512365.938 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1E | 2 | 4879670.688 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1E | 1 | 949618.4375 | Ga0308317_14976302 | hypothetical protein/hypothetical protein | Proteobacteria |
| P.CW.IS03.1E | 2 | 912559.875 | Ga0308317_13169602 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1E | 1 | 62108412 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1E | 1 | 1100895.656 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1E | 1 | 1327721.875 | 3300002026_MIS_1000043442 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS03.1E | 1 | 926590.375 | 3300002026_MIS_1000043428 | Ferredoxin | Bacillariophyta |
| P.CW.IS03.1E | 1 | 797223.25 | Ga0308317_10414141 | phenylalanyl-tRNA synthetase alpha chain | Bacteroidetes |
| P.CW.IS03.1E | 1 | 1056668.375 | 3300002026_MIS_100116427 | NA | Chloroflexi |
| P.CW.IS03.1E | 2 | 861552.0156 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS03.1E | 2 | 810844 | 3300002026_MIS_100048729 | NA | Chloroflexi |
| P.CW.IS03.1E | 3 | 810515.0938 | Ga0308317_14101438 | adenosylhomocysteinase | Chloroflexi |
| P.CW.IS03.1E | 5 | 16717935.29 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1E | 2 | 16483456.5 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1E | 9 | 6879422.227 | Ga0308317_122782739 | uncharacterized protein DUF3386 | Cyanobacteria |
| P.CW.IS03.1E | 3 | 5617630.563 | Ga0308317_10126032 | hypothetical protein | Cyanobacteria |
| P.CW.IS03.1E | 1 | 2782377.75 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1E | 1 | 2222536.25 | Ga0308317_11304452 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1E | 3 | 1754320.633 | Ga0308317_15402722 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1E | 1 | 1166341.625 | 3300002027_MIS_101906972 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS03.1E | 4 | 25997227.25 | Ga0308317_15042393 | rubredoxin | Proteobacteria |
| P.CW.IS03.1E | 2 | 16153602.55 | Ga0308317_12760225 | ferredoxin | Proteobacteria |
| P.CW.IS03.1E | 4 | 11428780.97 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1E | 3 | 10737516.31 | 3300002027_MIS_100632912 | NA | Proteobacteria |
| P.CW.IS03.1E | 2 | 4373195 | Ga0308317_14787195 | malate dehydrogenase | Proteobacteria |
| P.CW.IS03.1E | 2 | 2753269.078 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1E | 2 | 2712626.625 | 3300002027_MIS_100108445 | Nitrate reductase gamma subunit | Proteobacteria |
| P.CW.IS03.1E | 2 | 2179408.625 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1E | 3 | 805338.5313 | 3300002026_MIS_1000222523 | Rubrerythrin | Proteobacteria |
| P.CW.IS03.1E | 1 | 769551.625 | 3300002026_MIS_100060734 | Transcriptional regulatory protein, C terminal | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1E | 1 | 720023.5 | Ga0308317_10421892 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1E | 1 | 81326160 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1E | 1 | 19004422 | 3300002026_MIS_1000203919 | NA | Unclassified |
| P.CW.IS03.1E | 4 | 9551633.125 | 3300002027_MIS_100969412 | NA | Unclassified |
| P.CW.IS03.1E | 1 | 2199810 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS03.1E | 3 | 2093857.766 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1E | 4 | 1189966.336 | 3300002026_MIS_100076979 | Phycobilisome Linker polypeptide | Unclassified |
| P.CW.IS03.1F | 2 | 1623182.602 | 3300002027_MIS_101644102 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS03.1F | 3 | 774497.75 | 3300002026_MIS_1000043442 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS03.1F | 4 | 612061.1875 | 3300002026_MIS_1000043424 | NA | Bacillariophyta |
| P.CW.IS03.1F | 4 | 14567798.53 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS03.1F | 4 | 5299673.219 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS03.1F | 4 | 3651191.375 | Ga0308317_129817518 | fumarate hydratase class II | Chloroflexi |
| P.CW.IS03.1F | 4 | 18029562.13 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1F | 3 | 15916697.92 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS03.1F | 2 | 5060453.031 | Ga0308317_103184713 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1F | 2 | 4319001.781 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1F | 3 | 3433071.242 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1F | 2 | 3105327 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1F | 3 | 2941922.5 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS03.1F | 2 | 1183971.641 | Ga0308317_13177228 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS03.1F | 3 | 1156631.313 | 3300002027_MIS_100417313 | Manganese-stabilising protein / photosystem II polypeptide | Cyanobacteria |
| P.CW.IS03.1F | 2 | 1132512.656 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1F | 1 | 847693.6875 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS03.1F | 2 | 763170.625 | Ga0308317_10501821 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1F | 4 | 682598 | Ga0308317_13201486 | glyoxylase-like metal-dependent hydrolase (beta-lactamase superfamily II) | Cyanobacteria |
| P.CW.IS03.1F | 3 | 666993.1289 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1F | 1 | 658984.375 | 3300002026_MIS_1003608920 | NA | Cyanobacteria |
| P.CW.IS03.1F | 2 | 589678.9688 | Ga0308317_14361226 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS03.1F | 2 | 556786.875 | Ga0308317_108417113 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS03.1F | 1 | 531859.75 | 3300002027_MIS_100535401 | Hemolysin-type calcium-binding repeat (2 copies) | Cyanobacteria |
| P.CW.IS03.1F | 8 | 38560871.97 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 1 | 23144240.63 | Ga0308317_10589352 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 2 | 4178826.875 | Ga0308317_14352243 | acetyl-CoA synthetase | Proteobacteria |
| P.CW.IS03.1F | 2 | 3174545.734 | Ga0308317_10051721 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 1 | 2947441.5 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1F | 1 | 2451910.5 | 3300002026_MIS_100403762 | NA | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1F | 1 | 2245462.5 | Ga0308317_13390221 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 2 | 1991067.344 | Ga0308317_12147148 | ribonuclease E | Proteobacteria |
| P.CW.IS03.1F | 1 | 1078939.125 | Ga0308317_10421892 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 1 | 903507.375 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1F | 4 | 832109.1484 | 3300002027_MIS_100225507 | Adenosine-5'-phosphosulfate reductase beta subunit; | Proteobacteria |
| P.CW.IS03.1F | 3 | 771255.4219 | Ga0308317_14549566 | methyl-accepting chemotaxis protein/methyl-accepting chemotaxis protein | Proteobacteria |
| P.CW.IS03.1F | 2 | 678213.5313 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 2 | 652423.2813 | 3300002027_MIS_101006588 | NA | Proteobacteria |
| P.CW.IS03.1F | 1 | 576546.9375 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 1 | 50938192 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1F | 2 | 3529279.219 | 3300002027_MIS_100302972 | NA | Unclassified |
| P.CW.IS03.1F | 3 | 2654534.906 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1F | 3 | 1148410.273 | 3300002026_MIS_100315161 | ATPases with chaperone activity, ATP-binding subunit | Unclassified |
| P.CW.IS03.1F | 1 | 698198.4375 | 3300002024_MIS_10944721 | Transketolase, thiamine diphosphate binding domain | Unclassified |
| P.CW.IS03.1F | 2 | 550668.7969 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1F | 3 | 26439032.22 | Ga0308317_10157129 | propionyl-CoA synthetase | Unclassified |
| P.CW.IS03.1F | 1 | 451274496 | Ga0308317_10684161 | murein L,D-transpeptidase YcbB/YkuD | Proteobacteria |
| P.CW.IS03.1F | 4 | 2117069.375 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS03.1F | 9 | 167038311.6 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 2 | 3314059.461 | 3300002027_MIS_100642973 | NA | Proteobacteria |
| P.CW.IS03.1F | 1 | 56429088 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1F | 2 | 8060936.281 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS03.1F | 3 | 2641352.031 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1F | 1 | 2082693.75 | Ga0308317_10190262 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1F | 8 | 92141009.53 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 8 | 2808751.27 | Ga0308317_13827036 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 5 | 2287993.719 | Ga0308317_10627422 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 1 | 2045158.625 | 3300002026_MIS_100307544 | NA | Proteobacteria |
| P.CW.IS03.1F | 1 | 2011577 | 3300002027_MIS_100094537 | NA | Proteobacteria |
| P.CW.IS03.1F | 2 | 1881113.906 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1F | 1 | 1276056.5 | Ga0308317_13169602 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 1 | 63206176 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1F | 3 | 1231932.984 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1F | 1 | 462199.2188 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS03.1F | 2 | 5572417.5 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1F | 1 | 2315382 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1F | 1 | 1156734.5 | 3300002027_MIS_101362743 | NA | Cyanobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1F | 1 | 848941.1875 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS03.1F | 2 | 690162.3359 | Ga0308317_10646073 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1F | 1 | 620280.1875 | 3300002026_MIS_100078659 | Predicted periplasmic or secreted lipoprotein | Cyanobacteria |
| P.CW.IS03.1F | 3 | 10018372.5 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1F | 5 | 9772953.469 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 1 | 5105778.5 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1F | 3 | 1502414.766 | Ga0308317_10627422 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 2 | 732625.1328 | Ga0308317_13169602 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 3 | 541788.0313 | Ga0308317_13827036 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1F | 1 | 41987932 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1F | 1 | 3447670.969 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1F | 1 | 1043778.625 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS03.1F | 2 | 922019.125 | 3300002026_MIS_100355563 | Uncharacterized protein conserved in bacteria | Unclassified |
| P.CW.IS03.1F | 1 | 927012.0625 | 3300002026_MIS_1000043428 | Ferredoxin | Bacillariophyta |
| P.CW.IS03.1F | 2 | 7372746.469 | 3300002027_MIS_101946932 | CP12 domain | Cyanobacteria |
| P.CW.IS03.1F | 1 | 1394022.75 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS03.1F | 1 | 1341051.625 | Ga0308317_10646073 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1F | 1 | 849109.625 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1F | 1 | 3789175 | 3300002026_MIS_100307544 | NA | Proteobacteria |
| P.CW.IS03.1F | 1 | 1286281.5 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS03.1F | 2 | 1139048.664 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1F | 2 | 917113.1563 | Ga0308317_14715166 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 2 | 675506.5313 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1F | 1 | 38946396 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS03.1F | 2 | 825274.0859 | 3300002027_MIS_101184422 | Rubredoxin | Unclassified |
| P.CW.IS03.1G | 3 | 6782366.656 | 3300002024_MIS_10742481 | NA | Bacillariophyta |
| P.CW.IS03.1G | 1 | 4684538.5 | 3300002026_MIS_1000043416 | Ribosomal protein L7/L12 | Bacillariophyta |
| P.CW.IS03.1G | 5 | 3457337.836 | 3300002026_MIS_1000043424 | NA | Bacillariophyta |
| P.CW.IS03.1G | 1 | 1914830.969 | 3300002027_MIS_100584761 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS03.1G | 1 | 1109863.25 | 3300002026_MIS_1000043450 | Ribulose 1,5-bisphosphate carboxylase, large subunit | Bacillariophyta |
| P.CW.IS03.1G | 5 | 1030674.219 | 3300002026_MIS_1000043442 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS03.1G | 3 | 963718.2461 | 3300002026_MIS_100004349 | Mg-chelatase subunit ChlI | Bacillariophyta |
| P.CW.IS03.1G | 2 | 945400.6797 | 3300002027_MIS_101644102 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS03.1G | 1 | 937057.125 | 3300002024_MIS_11548661 | NA | Bacillariophyta |
| P.CW.IS03.1G | 3 | 3159978.016 | Ga0308317_10358796 | CspA family cold shock protein | Bacteroidetes |
| P.CW.IS03.1G | 3 | 18337268.75 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS03.1G | 1 | 6683097.5 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS03.1G | 3 | 3754937 | 3300002026_MIS_100008266 | Chaperonin GroEL (HSP60 family) | Chloroflexi |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1G | 1 | 3210813.5 | 3300002026_MIS_100048729 | NA | Chloroflexi |
| P.CW.IS03.1G | 3 | 3079031.609 | Ga0308317_14101438 | adenosylhomocysteinase | Chloroflexi |
| P.CW.IS03.1G | 6 | 83898927.13 | 3300002027_MIS_101479931 | Co-chaperonin GroES (HSP10) | Cyanobacteria |
| P.CW.IS03.1G | 3 | 14250031.35 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS03.1G | 4 | 11315523.53 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS03.1G | 4 | 11152339.55 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS03.1G | 5 | 10673740.09 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS03.1G | 2 | 9996809.75 | Ga0308317_103184713 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1G | 3 | 9618219.805 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS03.1G | 7 | 4698854.75 | Ga0308317_14591192 | chromosome segregation ATPase | Cyanobacteria |
| P.CW.IS03.1G | 1 | 4636928.5 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS03.1G | 4 | 4259677.156 | 3300002027_MIS_100523012 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS03.1G | 2 | 2372453.703 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS03.1G | 2 | 2321239.125 | 3300002026_MIS_100078659 | Predicted periplasmic or secreted lipoprotein | Cyanobacteria |
| P.CW.IS03.1G | 3 | 2127214.742 | 3300002027_MIS_100980214 | NA | Cyanobacteria |
| P.CW.IS03.1G | 2 | 2104995.703 | Ga0308317_12299312 | uncharacterized protein DUF4090 | Cyanobacteria |
| P.CW.IS03.1G | 3 | 1760832.469 | Ga0308317_108417113 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS03.1G | 2 | 1705053.766 | 3300002027_MIS_100417313 | Manganese-stabilising protein / photosystem II polypeptide | Cyanobacteria |
| P.CW.IS03.1G | 2 | 1553755.125 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS03.1G | 2 | 1492926.188 | 3300002026_MIS_1001557614 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS03.1G | 1 | 1250850.375 | 3300002026_MIS_1000143515 | Ribosomal protein L7/L12 | Cyanobacteria |
| P.CW.IS03.1G | 3 | 1213623.75 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS03.1G | 5 | 1199761.063 | Ga0308317_10084168 | ferredoxin | Cyanobacteria |
| P.CW.IS03.1G | 1 | 1193863.375 | Ga0308317_14110071 | Uma2 family endonuclease | Cyanobacteria |
| P.CW.IS03.1G | 4 | 1055473.148 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS03.1G | 2 | 1054869.281 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS03.1G | 2 | 984369.0938 | Ga0308317_103715913 | small subunit ribosomal protein S1 | Cyanobacteria |
| P.CW.IS03.1G | 3 | 967734.2969 | Ga0308317_13519002 | phage shock protein A | Cyanobacteria |
| P.CW.IS03.1G | 3 | 915653.3594 | 3300002026_MIS_1003608920 | NA | Cyanobacteria |
| P.CW.IS03.1G | 2 | 911102.1406 | 3300002027_MIS_100389562 | Uncharacterized protein conserved in bacteria | Cyanobacteria |
| P.CW.IS03.1G | 7 | 160913104.2 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1G | 2 | 141355990.6 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1G | 2 | 30211857.3 | Ga0308317_12399522 | sulfate adenylyltransferase | Proteobacteria |
| P.CW.IS03.1G | 2 | 19605012.75 | 3300002027_MIS_101881042 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1G | 6 | 10843096.63 | 3300002027_MIS_100632912 | NA | Proteobacteria |
| P.CW.IS03.1G | 1 | 4487350.438 | 3300002026_MIS_1002605312 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS03.1G | 2 | 3834163.25 | 3300002027_MIS_101283364 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS03.1G | 3 | 3763355.031 | Ga0308317_12147148 | ribonuclease E | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS03.1G | 7 | 3158837.922 | Ga0308317_12326721 | peptidyl-prolyl cis-trans isomerase C | Proteobacteria |
| P.CW.IS03.1G | 3 | 2702811.445 | Ga0308317_14352243 | acetyl-CoA synthetase | Proteobacteria |
| P.CW.IS03.1G | 1 | 2133388.703 | Ga0308317_11885202 | signal transduction histidine kinase | Proteobacteria |
| P.CW.IS03.1G | 3 | 2036871.875 | Ga0308317_14715166 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1G | 5 | 2021870.063 | 3300002026_MIS_100110857 | Cache domain | Proteobacteria |
| P.CW.IS03.1G | 1 | 1229557.25 | Ga0308317_10307526 | rubrerythrin | Proteobacteria |
| P.CW.IS03.1G | 1 | 1225710.375 | 3300002027_MIS_101006588 | NA | Proteobacteria |
| P.CW.IS03.1G | 2 | 1179220.641 | Ga0308317_11787262 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1G | 1 | 1061780.25 | 3300002026_MIS_1000681111 | FKBP-type peptidyl-prolyl cis-trans isomerases 2 | Proteobacteria |
| P.CW.IS03.1G | 1 | 953439.8125 | 3300002024_MIS_10299911 | NA | Proteobacteria |
| P.CW.IS03.1G | 4 | 835942.9766 | Ga0308317_12414362 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS03.1G | 4 | 819970.7813 | Ga0308317_11375573 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1G | 1 | 793053.0625 | Ga0308317_15131222 | hypothetical protein | Proteobacteria |
| P.CW.IS03.1G | 2 | 40128320 | Ga0308317_13886182 | actin beta/gamma 1 | Unclassified |
| P.CW.IS03.1G | 2 | 4115688.719 | 3300002024_MIS_10944721 | Transketolase, thiamine diphosphate binding domain | Unclassified |
| P.CW.IS03.1G | 3 | 3385557.781 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1G | 3 | 1765142.359 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS03.1G | 3 | 1364265.563 | 3300002026_MIS_100315161 | ATPases with chaperone activity, ATP-binding subunit | Unclassified |
| P.CW.IS03.1G | 1 | 858997.875 | 3300002027_MIS_100665091 | NA | Unclassified |
| P.CW.IS03.1G | 3 | 765816.125 | 3300002027_MIS_101567531 | Actin and related proteins | Unclassified |
| P.CW.IS03.1G | 2 | 7631231.5 | Ga0308317_10866993 | peptidyl-prolyl cis-trans isomerase B (cyclophilin B) | Unclassified |
| P.CW.IS03.1G | 3 | 6772414.969 | Ga0308317_10157129 | propionyl-CoA synthetase | Unclassified |
| P.CW.IS04.2B | 2 | 2040867.813 | 3300002027_MIS_101644102 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS04.2B | 2 | 1082292.742 | Ga0308317_10332984 | DNA repair exonuclease SbcCD ATPase subunit | Bacteroidetes |
| P.CW.IS04.2B | 1 | 6160411.953 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS04.2B | 4 | 4688381.859 | 3300002026_MIS_100089044 | Uncharacterized conserved protein - COG2835 | Chloroflexi |
| P.CW.IS04.2B | 1 | 3796393.5 | Ga0308317_10733153 | hypothetical protein | Chloroflexi |
| P.CW.IS04.2B | 3 | 11142292.74 | Ga0308317_137341848 | uncharacterized protein DUF4090 | Cyanobacteria |
| P.CW.IS04.2B | 3 | 6865597.063 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS04.2B | 2 | 1112155.875 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS04.2B | 5 | 169785717.9 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS04.2B | 2 | 16448050.5 | Ga0308317_12147148 | ribonuclease E | Proteobacteria |
| P.CW.IS04.2B | 3 | 3891444.109 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS04.2B | 4 | 985751.4375 | Ga0308317_13827036 | hypothetical protein | Proteobacteria |
| P.CW.IS04.2B | 1 | 2242242.5 | Ga0308317_10604841 | peptidyl-prolyl cis-trans isomerase C | Unclassified |
| P.CW.IS04.2B | 2 | 2240295.961 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS05.2A | 1 | 572439.25 | 3300002027_MIS_101644101 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2A | 1 | 1691976.5 | Ga0308317_12083728 | threonine/homoserine/homoserine lactone efflux protein | Bacteroidetes |
| P.CW.IS05.2A | 1 | 793909.125 | Ga0308317_14401547 | acetyl-CoA C-acetyltransferase | Chloroflexi |
| P.CW.IS05.2A | 1 | 789224.0625 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS05.2A | 4 | 17795575.3 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS05.2A | 3 | 5805610.406 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS05.2A | 3 | 2415188.391 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS05.2A | 1 | 1061876.617 | Ga0308317_12299312 | uncharacterized protein DUF4090 | Cyanobacteria |
| P.CW.IS05.2A | 1 | 853230.5625 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.2A | 1 | 815541.375 | Ga0308317_11304452 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.2A | 8 | 17062985.98 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2A | 1 | 2748736 | 3300002024_MIS_11415981 | NA | Proteobacteria |
| P.CW.IS05.2A | 3 | 2129853.938 | Ga0308317_13827036 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2A | 2 | 1979132.188 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2A | 3 | 1263865.313 | Ga0308317_10627422 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2A | 1 | 28674214 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.2A | 1 | 3521653.25 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS05.2A | 1 | 1122947.125 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS05.2A | 2 | 9113490.109 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS05.2A | 1 | 1601866.625 | Ga0308317_12299312 | uncharacterized protein DUF4090 | Cyanobacteria |
| P.CW.IS05.2A | 2 | 1188858.594 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS05.2A | 3 | 28334030.42 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2A | 8 | 5053635.75 | Ga0308317_13827036 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2A | 5 | 4030068.914 | Ga0308317_10627422 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2A | 1 | 18543698 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.2A | 6 | 113258384.3 | Ga0308317_136744817 | hypothetical protein | Chloroflexi |
| P.CW.IS05.2A | 2 | 9448498.594 | Ga0308317_12688292 | large subunit ribosomal protein L7/L12 | Chloroflexi |
| P.CW.IS05.2A | 1 | 126398656 | 3300002024_MIS_11232711 | FOG: WD40 repeat | Cyanobacteria |
| P.CW.IS05.2A | 2 | 18656298.7 | Ga0308317_112429825 | large subunit ribosomal protein L7/L12 | Cyanobacteria |
| P.CW.IS05.2A | 6 | 129791240 | 3300002027_MIS_101881042 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2A | 7 | 33753223.49 | Ga0308317_125581024 | uncharacterized protein DUF3365 | Proteobacteria |
| P.CW.IS05.2A | 3 | 23677820.38 | 3300002027_MIS_101283364 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2A | 8 | 11697806.55 | 3300002027_MIS_100342344 | NA | Proteobacteria |
| P.CW.IS05.2A | 2 | 8670297.5 | Ga0308317_12414363 | TolA-binding protein | Proteobacteria |
| P.CW.IS05.2A | 2 | 7742120.844 | 3300002027_MIS_100070992 | Cytochrome c553 | Proteobacteria |
| P.CW.IS05.2A | 1 | 236361472 | Ga0308317_13088821 | hypothetical protein | Unclassified |
| P.CW.IS05.2A | 1 | 8217875 | 3300002027_MIS_100331351 | DNA polymerase III, alpha subunit | Unclassified |
| P.CW.IS05.2A | 1 | 959930.1875 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS05.2A | 2 | 661626.2969 | 3300002026_MIS_100100153 | NA | Cyanobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2A | 1 | 377557.4375 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS05.2A | 1 | 1713372.875 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2A | 1 | 21643302 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.2A | 2 | 768606.625 | 3300002027_MIS_101184422 | Rubredoxin | Unclassified |
| P.CW.IS05.2A | 1 | 267591.5313 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS05.2A | 5 | 4792155 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS05.2A | 5 | 20766174.63 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS05.2A | 5 | 9984517.906 | 3300002027_MIS_101479931 | Co-chaperonin GroES (HSP10) | Cyanobacteria |
| P.CW.IS05.2A | 3 | 5423822.635 | 3300002026_MIS_1000222210 | NA | Cyanobacteria |
| P.CW.IS05.2A | 3 | 3792905.078 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS05.2A | 3 | 1986554.656 | 3300002027_MIS_100523012 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.2A | 12 | 540305958.8 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2A | 2 | 60065049.91 | Ga0308317_10589352 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2A | 2 | 3406526.25 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2A | 2 | 3393204.164 | Ga0308317_13474461 | chaperonin GroEL | Proteobacteria |
| P.CW.IS05.2A | 1 | 2273178.25 | Ga0308317_10487537 | alanine racemase | Proteobacteria |
| P.CW.IS05.2A | 2 | 2092653.875 | 3300002026_MIS_100383013 | Outer membrane protein | Proteobacteria |
| P.CW.IS05.2A | 1 | 1944271.625 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2A | 1 | 27253662 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.2A | 4 | 727907.6016 | Ga0308317_14702668 | Fe-S cluster assembly scaffold protein SufB/Fe-S cluster assembly scaffold protein SufB/intein/homing endonuclease | Chloroflexi |
| P.CW.IS05.2A | 5 | 19662889.95 | Ga0308317_112429825 | large subunit ribosomal protein L7/L12 | Cyanobacteria |
| P.CW.IS05.2A | 2 | 3278886.879 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS05.2A | 3 | 694667.625 | Ga0308317_102097225 | uncharacterized membrane protein YqiK | Cyanobacteria |
| P.CW.IS05.2A | 8 | 47290762.35 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2A | 5 | 26165663.5 | Ga0308317_152169619 | cytochrome c | Proteobacteria |
| P.CW.IS05.2A | 6 | 10730700.27 | 3300002027_MIS_101881042 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2A | 3 | 2175693.984 | Ga0308317_131768030 | cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2A | 2 | 1610668.766 | 3300002027_MIS_100070992 | Cytochrome c553 | Proteobacteria |
| P.CW.IS05.2A | 3 | 1364464.438 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS05.2A | 1 | 1277948.125 | Ga0308317_12708882 | catechol 2,3-dioxygenase-like lactoylglutathione lyase family enzyme | Proteobacteria |
| P.CW.IS05.2A | 2 | 1126785.621 | 3300002027_MIS_101283364 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2A | 2 | 1113188.25 | Ga0308317_10286231 | branched-chain amino acid transport system substrate-binding protein | Proteobacteria |
| P.CW.IS05.2A | 2 | 729371.875 | Ga0308317_13761634 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2A | 3 | 664278.9531 | 3300002026_MIS_100323342 | NA | Proteobacteria |
| P.CW.IS05.2A | 3 | 2756830.672 | Ga0308317_10446361 | hypothetical protein | Unclassified |
| P.CW.IS05.2A | 1 | 975539.375 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.2A | 1 | 665715.75 | 3300002027_MIS_100378483 | Cytochrome c553 | Unclassified |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2B | 2 | 2104702.266 | 3300002027_MIS_101644102 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS05.2B | 4 | 2037969 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS05.2B | 1 | 1467373.125 | Ga0308317_10733153 | hypothetical protein | Chloroflexi |
| P.CW.IS05.2B | 1 | 35841328.88 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS05.2B | 5 | 12195110.88 | Ga0308317_137341848 | uncharacterized protein DUF4090 | Cyanobacteria |
| P.CW.IS05.2B | 3 | 7221182.969 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS05.2B | 1 | 1218037.875 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS05.2B | 6 | 54394532.22 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2B | 7 | 5999193.617 | Ga0308317_10627422 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 7 | 5010560.531 | Ga0308317_13827036 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 1 | 2511966.5 | Ga0308317_12717772 | lysyl-tRNA synthetase class 2 | Proteobacteria |
| P.CW.IS05.2B | 1 | 14906691 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.2B | 3 | 3506996.57 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS05.2B | 1 | 2262228 | Ga0308317_13144313 | hypothetical protein | Unclassified |
| P.CW.IS05.2B | 1 | 1848393.75 | Ga0308317_12594022 | hypothetical protein | Unclassified |
| P.CW.IS05.2B | 1 | 2848229.484 | Ga0308317_13144313 | hypothetical protein | Unclassified |
| P.CW.IS05.2B | 4 | 24950873.06 | Ga0308317_12688292 | large subunit ribosomal protein L7/L12 | Chloroflexi |
| P.CW.IS05.2B | 4 | 19127467.42 | 3300002026_MIS_1000034633 | Ribosomal protein S1 | Chloroflexi |
| P.CW.IS05.2B | 1 | 13499316.53 | 3300002027_MIS_100755682 | Carbon dioxide concentrating mechanism/carboxysome shell protein | Cyanobacteria |
| P.CW.IS05.2B | 2 | 145405196 | 3300002027_MIS_101881042 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2B | 14 | 89428002.17 | 3300002027_MIS_100342344 | NA | Proteobacteria |
| P.CW.IS05.2B | 8 | 33957260.69 | Ga0308317_125581024 | uncharacterized protein DUF3365 | Proteobacteria |
| P.CW.IS05.2B | 3 | 26760575.5 | 3300002027_MIS_101283364 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2B | 1 | 9643787.25 | 3300002027_MIS_100070992 | Cytochrome c553 | Proteobacteria |
| P.CW.IS05.2B | 1 | 329458240 | 3300002024_MIS_11122321 | Acetate kinase | Unclassified |
| P.CW.IS05.2B | 1 | 27094142 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS05.2B | 1 | 526651.25 | 3300002026_MIS_1000043428 | Ferredoxin | Bacillariophyta |
| P.CW.IS05.2B | 1 | 1437658.125 | Ga0308317_10274931 | sigma-70-like protein | Chloroflexi |
| P.CW.IS05.2B | 2 | 16743780.84 | Ga0308317_15402722 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS05.2B | 7 | 5567554.469 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS05.2B | 2 | 4131844.5 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS05.2B | 2 | 2179981.484 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS05.2B | 1 | 599435.875 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS05.2B | 2 | 1513496.461 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS05.2B | 2 | 922053.0391 | Ga0308317_14863015 | rubrerythrin | Proteobacteria |
| P.CW.IS05.2B | 1 | 39971072 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.2B | 2 | 1814893.242 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS05.2B | 1 | 1503523.875 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2B | 1 | 1437255.5 | Ga0308317_11209342 | aminopeptidase N | Unclassified |
| P.CW.IS05.2B | 1 | 196864896 | 3300002024_MIS_10525911 | Chlorophyll A-B binding protein | Bacillariophyta |
| P.CW.IS05.2B | 7 | 87183764.63 | 3300002026_MIS_1000043442 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS05.2B | 3 | 50464618 | 3300002026_MIS_100004349 | Mg-chelatase subunit ChlI | Bacillariophyta |
| P.CW.IS05.2B | 3 | 36470605.25 | 3300002026_MIS_1000043414 | PsaD | Bacillariophyta |
| P.CW.IS05.2B | 4 | 35327220.25 | 3300002027_MIS_101644102 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS05.2B | 4 | 15882155.63 | 3300002026_MIS_1000043424 | NA | Bacillariophyta |
| P.CW.IS05.2B | 1 | 5431808.5 | 3300002026_MIS_100363859 | F0F1-type ATP synthase, beta subunit | Bacillariophyta |
| P.CW.IS05.2B | 1 | 5365466.438 | 3300002024_MIS_11548661 | NA | Bacillariophyta |
| P.CW.IS05.2B | 1 | 5614539 | Ga0308317_145685719 | small subunit ribosomal protein S3 | Chlorobi |
| P.CW.IS05.2B | 1 | 73662256 | Ga0308317_10340078 | Fe-Mn family superoxide dismutase | Chloroflexi |
| P.CW.IS05.2B | 2 | 51993525.25 | Ga0308317_14101438 | adenosylhomocysteinase | Chloroflexi |
| P.CW.IS05.2B | 8 | 44790069.88 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS05.2B | 2 | 43463538.25 | Ga0308317_14101437 | S-adenosylmethionine synthetase | Chloroflexi |
| P.CW.IS05.2B | 1 | 39076068 | Ga0308317_13224551 | multiple sugar transport system substrate-binding protein | Chloroflexi |
| P.CW.IS05.2B | 3 | 31445516.75 | Ga0308317_14023463 | transcriptional regulator with XRE-family HTH domain | Chloroflexi |
| P.CW.IS05.2B | 3 | 28555322.25 | Ga0308317_14165545 | phosphoglycerate kinase | Chloroflexi |
| P.CW.IS05.2B | 3 | 27518626 | Ga0308317_141014330 | hypothetical protein | Chloroflexi |
| P.CW.IS05.2B | 3 | 26465324 | Ga0308317_142163210 | glycine cleavage system H protein | Chloroflexi |
| P.CW.IS05.2B | 2 | 19961138.94 | Ga0308317_136328629 | SH3 domain-containing protein | Chloroflexi |
| P.CW.IS05.2B | 1 | 18116080 | Ga0308317_11685233 | succinyl-CoA synthetase beta subunit | Chloroflexi |
| P.CW.IS05.2B | 2 | 14358335.63 | 3300002026_MIS_100043273 | Translation elongation factors (GTPases) | Chloroflexi |
| P.CW.IS05.2B | 2 | 13283913.5 | Ga0308317_129817518 | fumarate hydratase class II | Chloroflexi |
| P.CW.IS05.2B | 2 | 12956052.25 | 3300002026_MIS_100044789 | NA | Chloroflexi |
| P.CW.IS05.2B | 3 | 12314574 | Ga0308317_14401547 | acetyl-CoA C-acetyltransferase | Chloroflexi |
| P.CW.IS05.2B | 1 | 11934388 | 3300002026_MIS_100003469 | ATP-dependent Lon protease, bacterial type | Chloroflexi |
| P.CW.IS05.2B | 2 | 11901291 | Ga0308317_13298891 | ABC-type xylose transport system substrate-binding protein/ABC-type xylose transport system substrate-binding | Chloroflexi |
| P.CW.IS05.2B | 1 | 11150432.25 | Ga0308317_144015415 | putative zinc ribbon protein | Chloroflexi |
| P.CW.IS05.2B | 2 | 8613537.25 | Ga0308317_11143075 | small subunit ribosomal protein S2 | Chloroflexi |
| P.CW.IS05.2B | 1 | 7446905.5 | 3300002026_MIS_100032683 | Glycine/serine hydroxymethyltransferase | Chloroflexi |
| P.CW.IS05.2B | 1 | 7374257 | 3300002026_MIS_1000139220 | Poly(3-hydroxyalkanoate) synthetase | Chloroflexi |
| P.CW.IS05.2B | 1 | 6071722 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS05.2B | 2 | 5715830.25 | Ga0308317_113407711 | ABC-type glycerol-3-phosphate transport system substrate-binding protein/ABC-type glycerol-3-phosphate transport system substrate-binding protein | Chloroflexi |
| P.CW.IS05.2B | 1 | 5506272 | 3300002026_MIS_100037676 | Methylmalonyl-CoA mutase, N-terminal domain/subunit | Chloroflexi |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2B | 7 | 882829219.4 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS05.2B | 3 | 522155620.1 | Ga0308317_103184713 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.2B | 4 | 286520379.5 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS05.2B | 5 | 209747024.1 | 3300002027_MIS_101659072 | Ycf66 protein N-terminus | Cyanobacteria |
| P.CW.IS05.2B | 5 | 204417103.5 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.2B | 4 | 142856612.3 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS05.2B | 1 | 140164049 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS05.2B | 1 | 117372656 | Ga0308317_14972441 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS05.2B | 3 | 107250252.3 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS05.2B | 4 | 103833112 | 3300002027_MIS_100980214 | NA | Cyanobacteria |
| P.CW.IS05.2B | 7 | 102612113.3 | Ga0308317_123538692 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS05.2B | 3 | 96069410.38 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS05.2B | 4 | 81943073.63 | Ga0308317_10126032 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.2B | 3 | 78262833.06 | Ga0308317_108840816 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.2B | 2 | 68641852.38 | 3300002024_MIS_11769591 | NA | Cyanobacteria |
| P.CW.IS05.2B | 4 | 67150221 | Ga0308317_108417113 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS05.2B | 1 | 62058222 | Ga0308317_105563459 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.2B | 3 | 59120336.19 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.2B | 3 | 54855427.88 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.2B | 4 | 52152326.75 | Ga0308317_13201486 | glyoxylase-like metal-dependent hydrolase (beta-lactamase superfamily II) | Cyanobacteria |
| P.CW.IS05.2B | 2 | 49491265 | 3300002026_MIS_1000143515 | Ribosomal protein L7/L12 | Cyanobacteria |
| P.CW.IS05.2B | 3 | 49393165.38 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS05.2B | 4 | 48867207.5 | Ga0308317_14361226 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.2B | 1 | 47493100 | Ga0308317_15402722 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS05.2B | 4 | 44606408 | Ga0308317_12286322 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.2B | 5 | 43215577.5 | Ga0308317_14591192 | chromosome segregation ATPase | Cyanobacteria |
| P.CW.IS05.2B | 3 | 40154624 | 3300002024_MIS_10023532 | Uncharacterized conserved protein - COG3937 | Cyanobacteria |
| P.CW.IS05.2B | 4 | 36967591.13 | Ga0308317_113580413 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.2B | 3 | 34235002.56 | Ga0308317_11304452 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.2B | 5 | 34128148.81 | 3300002024_MIS_10728201 | NA | Cyanobacteria |
| P.CW.IS05.2B | 3 | 32962907.38 | Ga0308317_122782711 | phycocyanin alpha chain | Cyanobacteria |
| P.CW.IS05.2B | 1 | 31165124 | 3300002027_MIS_101906972 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.2B | 2 | 28580529 | Ga0308317_15136002 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.2B | 2 | 26726597.5 | Ga0308317_122782739 | uncharacterized protein DUF3386 | Cyanobacteria |
| P.CW.IS05.2B | 2 | 24998310.5 | Ga0308317_15235161 | SdrD B-like protein | Cyanobacteria |
| P.CW.IS05.2B | 3 | 22566204.25 | Ga0308317_13377662 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.2B | 2 | 21521484 | 3300002026_MIS_100078659 | Predicted periplasmic or secreted lipoprotein | Cyanobacteria |

| P.CW.IS05.2B | 2 | 21387849.13 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
|---|---|---|---|---|---|
| P.CW.IS05.2B | 3 | 20781925.75 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS05.2B | 3 | 19821856 | 3300002026_MIS_1001557614 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.2B | 2 | 19247354 | 3300002026_MIS_100192369 | Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family | Cyanobacteria |
| P.CW.IS05.2B | 2 | 18592665.38 | Ga0308317_10379548 | enolase | Cyanobacteria |
| P.CW.IS05.2B | 2 | 17871274 | Ga0308317_14375973 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.2B | 1 | 17852276 | 3300002027_MIS_100251133 | NA | Cyanobacteria |
| P.CW.IS05.2B | 3 | 16389601.88 | Ga0308317_14310812 | molybdopterin converting factor small subunit | Cyanobacteria |
| P.CW.IS05.2B | 2 | 15656191 | Ga0308317_118065245 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS05.2B | 2 | 11495541.5 | 3300002027_MIS_100199341 | Transketolase | Cyanobacteria |
| P.CW.IS05.2B | 1 | 11287719.5 | Ga0308317_137341848 | uncharacterized protein DUF4090 | Cyanobacteria |
| P.CW.IS05.2B | 2 | 10812542 | Ga0308317_113275065 | aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit C | Cyanobacteria |
| P.CW.IS05.2B | 1 | 10539989 | 3300002027_MIS_101482251 | Oxidoreductase NAD-binding domain | Cyanobacteria |
| P.CW.IS05.2B | 2 | 10049002 | Ga0308317_122782710 | phycocyanin beta chain | Cyanobacteria |
| P.CW.IS05.2B | 2 | 9876044.25 | 3300002027_MIS_101998811 | FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor) | Cyanobacteria |
| P.CW.IS05.2B | 2 | 9873528.5 | 3300002026_MIS_100054824 | NA | Cyanobacteria |
| P.CW.IS05.2B | 1 | 9602371 | 3300002026_MIS_1000534011 | Chaperonin GroEL (HSP60 family) | Cyanobacteria |
| P.CW.IS05.2B | 2 | 8145256.125 | 3300002027_MIS_101074151 | Hemolysin-type calcium-binding repeat (2 copies) | Cyanobacteria |
| P.CW.IS05.2B | 2 | 7367799.75 | Ga0308317_11795261 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.2B | 1 | 7232943 | 3300002027_MIS_100173914 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.2B | 2 | 6400378.25 | Ga0308317_10420022 | photosystem II protein | Cyanobacteria |
| P.CW.IS05.2B | 1 | 6329210.5 | 3300002026_MIS_100318247 | Thioredoxin-like proteins and domains | Cyanobacteria |
| P.CW.IS05.2B | 2 | 6107492.5 | Ga0308317_101469348 | protein phosphatase | Cyanobacteria |
| P.CW.IS05.2B | 1 | 6096573.25 | Ga0308317_141694522 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.2B | 1 | 5523362.5 | Ga0308317_11713361 | diaminopimelate decarboxylase | Cyanobacteria |
| P.CW.IS05.2B | 3 | 5277677.25 | Ga0308317_13519002 | phage shock protein A | Cyanobacteria |
| P.CW.IS05.2B | 1 | 5193042 | 3300002027_MIS_100325692 | Aspartyl-tRNA synthetase | Cyanobacteria |
| P.CW.IS05.2B | 2 | 550989207.4 | Ga0308317_14352243 | acetyl-CoA synthetase | Proteobacteria |
| P.CW.IS05.2B | 5 | 186878642.8 | 3300002027_MIS_100632912 | NA | Proteobacteria |
| P.CW.IS05.2B | 5 | 181855533.8 | Ga0308317_10281003 | signal transduction histidine kinase | Proteobacteria |
| P.CW.IS05.2B | 8 | 177829318.6 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2B | 5 | 156120282.4 | 3300002027_MIS_101963794 | NA | Proteobacteria |
| P.CW.IS05.2B | 3 | 132818288.4 | Ga0308317_13112285 | EF hand domain-containing protein/EF hand domain-containing protein/EF hand domain-containing protein | Proteobacteria |
| P.CW.IS05.2B | 1 | 126885956 | Ga0308317_12399522 | sulfate adenylyltransferase | Proteobacteria |
| P.CW.IS05.2B | 2 | 96732700.13 | 3300002027_MIS_100755711 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS05.2B | 3 | 96466625.88 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2B | 4 | 93324208.75 | Ga0308317_101705613 | chaperonin GroEL | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2B | 5 | 86978950.88 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2B | 2 | 76601506.75 | 3300002026_MIS_100316452 | Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases | Proteobacteria |
| P.CW.IS05.2B | 5 | 75465152.75 | Ga0308317_101172414 | uncharacterized protein DUF4398 | Proteobacteria |
| P.CW.IS05.2B | 2 | 65097941 | 3300002026_MIS_100090083 | Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains | Proteobacteria |
| P.CW.IS05.2B | 3 | 59976019.5 | Ga0308317_14715166 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2B | 4 | 57972079.25 | Ga0308317_14787195 | malate dehydrogenase | Proteobacteria |
| P.CW.IS05.2B | 1 | 50198744 | Ga0308317_11579021 | elongation factor G | Proteobacteria |
| P.CW.IS05.2B | 2 | 49128778 | Ga0308317_11671024 | signal transduction histidine kinase | Proteobacteria |
| P.CW.IS05.2B | 1 | 48443623.25 | Ga0308317_13495863 | tRNA 2-thiouridine synthesizing protein C | Proteobacteria |
| P.CW.IS05.2B | 1 | 40741940.25 | 3300002026_MIS_100319374 | NA | Proteobacteria |
| P.CW.IS05.2B | 2 | 38457911.25 | Ga0308317_10164838 | trigger factor | Proteobacteria |
| P.CW.IS05.2B | 2 | 37956579.63 | Ga0308317_12147148 | ribonuclease E | Proteobacteria |
| P.CW.IS05.2B | 2 | 37736869 | Ga0308317_14549566 | methyl-accepting chemotaxis protein/methyl-accepting chemotaxis protein | Proteobacteria |
| P.CW.IS05.2B | 1 | 35321632 | 3300002026_MIS_1000679112 | Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains | Proteobacteria |
| P.CW.IS05.2B | 3 | 31863415.5 | 3300002027_MIS_100071524 | NA | Proteobacteria |
| P.CW.IS05.2B | 2 | 27840985.34 | 3300002027_MIS_101918742 | NA | Proteobacteria |
| P.CW.IS05.2B | 4 | 26730721.75 | 3300002027_MIS_100225507 | Adenosine-5'-phosphosulfate reductase beta subunit; | Proteobacteria |
| P.CW.IS05.2B | 2 | 25157987.75 | Ga0308317_13169602 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 3 | 24700892.5 | Ga0308317_12760225 | ferredoxin | Proteobacteria |
| P.CW.IS05.2B | 1 | 22573955 | 3300002026_MIS_100307544 | NA | Proteobacteria |
| P.CW.IS05.2B | 1 | 19981955.5 | 3300002026_MIS_100312755 | Formyltetrahydrofolate synthetase | Proteobacteria |
| P.CW.IS05.2B | 2 | 18652731 | Ga0308317_10047734 | rubrerythrin | Proteobacteria |
| P.CW.IS05.2B | 4 | 18309111 | Ga0308317_12326721 | peptidyl-prolyl cis-trans isomerase C | Proteobacteria |
| P.CW.IS05.2B | 1 | 17063532 | 3300002027_MIS_100761073 | NA | Proteobacteria |
| P.CW.IS05.2B | 2 | 16267981 | 3300002027_MIS_100094537 | NA | Proteobacteria |
| P.CW.IS05.2B | 1 | 16008510 | Ga0308317_11023992 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 2 | 15716707 | 3300002027_MIS_100956831 | NA | Proteobacteria |
| P.CW.IS05.2B | 4 | 15181984.16 | 3300002026_MIS_100005424 | NA | Proteobacteria |
| P.CW.IS05.2B | 2 | 14937198.25 | Ga0308317_14162442 | tellurite resistance protein TerA | Proteobacteria |
| P.CW.IS05.2B | 4 | 14406935.38 | Ga0308317_12079864 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 1 | 13962342 | 3300002026_MIS_100201834 | Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases | Proteobacteria |
| P.CW.IS05.2B | 3 | 13903714.69 | Ga0308317_11161885 | uncharacterized protein DUF4124 | Proteobacteria |
| P.CW.IS05.2B | 1 | 13270813 | Ga0308317_14108855 | patatin-like phospholipase/acyl hydrolase | Proteobacteria |
| P.CW.IS05.2B | 3 | 12995102.75 | Ga0308317_11997111 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 2 | 12368159.5 | Ga0308317_153522310 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 3 | 12366665.63 | 3300002027_MIS_100032986 | NA | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2B | 3 | 12229599.25 | Ga0308317_11167621 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 3 | 12000128.5 | 3300002027_MIS_100933923 | Enolase | Proteobacteria |
| P.CW.IS05.2B | 2 | 11706480.5 | 3300002027_MIS_100157287 | NA | Proteobacteria |
| P.CW.IS05.2B | 2 | 11278916.5 | 3300002026_MIS_1002605312 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS05.2B | 1 | 10103417.5 | Ga0308317_11714423 | tRNA 2-thiouridine synthesizing protein C | Proteobacteria |
| P.CW.IS05.2B | 1 | 9237834.875 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2B | 2 | 9096803.75 | Ga0308317_10307526 | rubrerythrin | Proteobacteria |
| P.CW.IS05.2B | 2 | 8986190 | 3300002027_MIS_101006588 | NA | Proteobacteria |
| P.CW.IS05.2B | 1 | 8873862 | Ga0308317_11079945 | outer membrane protein OmpA-like peptidoglycan-associated protein | Proteobacteria |
| P.CW.IS05.2B | 2 | 8612427.625 | Ga0308317_10016455 | CspA family cold shock protein | Proteobacteria |
| P.CW.IS05.2B | 4 | 8586612.5 | 3300002026_MIS_1000222523 | Rubrerythrin | Proteobacteria |
| P.CW.IS05.2B | 1 | 8561202 | Ga0308317_10765403 | F-type H+-transporting ATPase subunit b | Proteobacteria |
| P.CW.IS05.2B | 4 | 8365146 | 3300002026_MIS_100285303 | Uncharacterized protein conserved in bacteria | Proteobacteria |
| P.CW.IS05.2B | 1 | 8105000.5 | 3300002026_MIS_100272663 | Chemotaxis protein histidine kinase and related kinases | Proteobacteria |
| P.CW.IS05.2B | 1 | 7834700 | 3300002026_MIS_1000796610 | GTPases - translation elongation factors | Proteobacteria |
| P.CW.IS05.2B | 1 | 7508090 | 3300002027_MIS_101826251 | NA | Proteobacteria |
| P.CW.IS05.2B | 1 | 7346142 | Ga0308317_11932061 | formate--tetrahydrofolate ligase | Proteobacteria |
| P.CW.IS05.2B | 1 | 7313631.5 | Ga0308317_10302992 | tRNA 2-thiouridine synthesizing protein E | Proteobacteria |
| P.CW.IS05.2B | 2 | 7094580.875 | Ga0308317_132725521 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 1 | 6951650 | Ga0308317_10677553 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 2 | 6407834.5 | 3300002026_MIS_1001066610 | Enolase | Proteobacteria |
| P.CW.IS05.2B | 1 | 6254917.5 | Ga0308317_10672482 | ATP-dependent HslUV protease ATP-binding subunit HslU | Proteobacteria |
| P.CW.IS05.2B | 1 | 6142463 | Ga0308317_12631992 | TetR/AcrR family transcriptional regulator | Proteobacteria |
| P.CW.IS05.2B | 1 | 6105543.5 | 3300002027_MIS_100157284 | Desulfoferrodoxin | Proteobacteria |
| P.CW.IS05.2B | 2 | 5581243.125 | Ga0308317_14839643 | F-type H+-transporting ATPase subunit alpha | Proteobacteria |
| P.CW.IS05.2B | 1 | 5487254.5 | Ga0308317_10286231 | branched-chain amino acid transport system substrate-binding protein | Proteobacteria |
| P.CW.IS05.2B | 2 | 5362977.75 | 3300002027_MIS_101179512 | Membrane-fusion protein | Proteobacteria |
| P.CW.IS05.2B | 1 | 6033052.5 | 3300002024_MIS_11567531 | ABC-type xylose transport system, periplasmic component | Spirochaetes |
| P.CW.IS05.2B | 5 | 88783676 | 3300002027_MIS_100245793 | NA | Thermotogae |
| P.CW.IS05.2B | 1 | 239633408 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.2B | 4 | 63984528.75 | 3300002026_MIS_100076979 | Phycobilisome Linker polypeptide | Unclassified |
| P.CW.IS05.2B | 3 | 47849272.75 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS05.2B | 2 | 43471886 | 3300002027_MIS_101091071 | NA | Unclassified |
| P.CW.IS05.2B | 4 | 38565322.13 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS05.2B | 1 | 38125171 | Ga0308317_11025225 | elongation factor Tu | Unclassified |
| P.CW.IS05.2B | 3 | 27331635 | 3300002027_MIS_100130573 | Fructose-1,6-bisphosphatase | Unclassified |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2B | 2 | 17780134.75 | 3300002026_MIS_100309382 | ABC-type branched-chain amino acid transport systems, periplasmic component | Unclassified |
| P.CW.IS05.2B | 2 | 16864659.25 | 3300002024_MIS_10944721 | Transketolase, thiamine diphosphate binding domain | Unclassified |
| P.CW.IS05.2B | 4 | 13985096 | Ga0308317_14115661 | hypothetical protein | Unclassified |
| P.CW.IS05.2B | 4 | 11365294.5 | 3300002026_MIS_1000860718 | Chlorophyll A-B binding protein | Unclassified |
| P.CW.IS05.2B | 1 | 10053933 | 3300002026_MIS_100394952 | Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases | Unclassified |
| P.CW.IS05.2B | 2 | 9468024 | Ga0308317_15205092 | hypothetical protein | Unclassified |
| P.CW.IS05.2B | 1 | 8910530 | 3300002024_MIS_10441722 | NA | Unclassified |
| P.CW.IS05.2B | 1 | 8132352.5 | 3300002027_MIS_101217361 | Repulsive guidance molecule (RGM) C-terminus | Unclassified |
| P.CW.IS05.2B | 3 | 7968048.5 | 3300002027_MIS_101567531 | Actin and related proteins | Unclassified |
| P.CW.IS05.2B | 1 | 6819483 | Ga0308317_11002012 | elongation factor Tu | Unclassified |
| P.CW.IS05.2B | 2 | 6616811.5 | Ga0308317_12363301 | hypothetical protein | Unclassified |
| P.CW.IS05.2B | 1 | 6482437.5 | 3300002027_MIS_100977031 | NA | Unclassified |
| P.CW.IS05.2B | 4 | 244993882.8 | Ga0308317_10157129 | propionyl-CoA synthetase | Unclassified |
| P.CW.IS05.2B | 5 | 80554526.75 | Ga0308317_14286767 | ATP-dependent Clp protease ATP-binding subunit ClpC | Unclassified |
| P.CW.IS05.2B | 2 | 15252916 | Ga0308317_13399152 | nitrile hydratase alpha subunit | Unclassified |
| P.CW.IS05.2B | 1 | 6203458 | Ga0308317_142867611 | elongation factor Tu | Unclassified |
| P.CW.IS05.2B | 2 | 6272045.344 | 3300002024_MIS_10742481 | NA | Bacillariophyta |
| P.CW.IS05.2B | 1 | 5668131 | Ga0308317_14432443 | large subunit ribosomal protein L7/L12 | Bacteroidetes |
| P.CW.IS05.2B | 2 | 412123178.3 | 3300002026_MIS_1000216930 | Bacterial SH3 domain | Chloroflexi |
| P.CW.IS05.2B | 2 | 129136432 | Ga0308317_136328629 | SH3 domain-containing protein | Chloroflexi |
| P.CW.IS05.2B | 1 | 14064762 | Ga0308317_10733153 | hypothetical protein | Chloroflexi |
| P.CW.IS05.2B | 5 | 12271671.19 | Ga0308317_136744837 | hypothetical protein | Chloroflexi |
| P.CW.IS05.2B | 2 | 5513555.375 | 3300002026_MIS_1000216929 | NA | Chloroflexi |
| P.CW.IS05.2B | 5 | 73178246 | Ga0308317_14186367 | cytochrome b6-f complex iron-sulfur subunit | Cyanobacteria |
| P.CW.IS05.2B | 2 | 30779180.25 | 3300002027_MIS_101999964 | Thioredoxin domain-containing protein | Cyanobacteria |
| P.CW.IS05.2B | 3 | 20718935.5 | Ga0308317_120902322 | hemoglobin | Cyanobacteria |
| P.CW.IS05.2B | 2 | 18454713 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS05.2B | 4 | 14834856.03 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.2B | 1 | 14517032 | Ga0308317_10649065 | thioredoxin 1 | Cyanobacteria |
| P.CW.IS05.2B | 1 | 13019989.88 | Ga0308317_14718221 | uncharacterized protein DUF2488 | Cyanobacteria |
| P.CW.IS05.2B | 3 | 12958637.5 | 3300002026_MIS_100364606 | NA | Cyanobacteria |
| P.CW.IS05.2B | 1 | 8192585 | Ga0308317_13108231 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.2B | 2 | 5778810.625 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS05.2B | 27 | 802091908.4 | Ga0308317_103957321 | phosphate transport system substrate-binding protein | Proteobacteria |
| P.CW.IS05.2B | 3 | 390625993.8 | 3300002027_MIS_100119652 | NA | Proteobacteria |
| P.CW.IS05.2B | 13 | 268818156.3 | Ga0308317_11484977 | FKBP-type peptidyl-prolyl cis-trans isomerase FkpA/FKBP-type peptidyl-prolyl cis-trans isomerase FklB | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.2B | 11 | 198041122.2 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.2B | 17 | 164043429.4 | 3300002027_MIS_100342344 | NA | Proteobacteria |
| P.CW.IS05.2B | 4 | 79266146.44 | 3300002027_MIS_101881042 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.2B | 1 | 57313815.25 | Ga0308317_11787992 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 1 | 52985894.22 | 3300002026_MIS_100118007 | NA | Proteobacteria |
| P.CW.IS05.2B | 7 | 49272690.94 | Ga0308317_11787997 | FKBP-type peptidyl-prolyl cis-trans isomerase FkpA/FKBP-type peptidyl-prolyl cis-trans isomerase FklB | Proteobacteria |
| P.CW.IS05.2B | 6 | 39518421.88 | 3300002027_MIS_101563262 | Cache domain | Proteobacteria |
| P.CW.IS05.2B | 3 | 32772826.16 | Ga0308317_15120514 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 4 | 18177489.94 | Ga0308317_13761634 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 2 | 15656939.63 | 3300002026_MIS_100352862 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS05.2B | 2 | 14275406.09 | Ga0308317_13533542 | 2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase) | Proteobacteria |
| P.CW.IS05.2B | 2 | 9046376.375 | 3300002027_MIS_100070992 | Cytochrome c553 | Proteobacteria |
| P.CW.IS05.2B | 3 | 8554905.844 | Ga0308317_11242674 | molecular chaperone DnaK | Proteobacteria |
| P.CW.IS05.2B | 1 | 5589789.625 | Ga0308317_114891727 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 3 | 5403668.5 | Ga0308317_10421892 | hypothetical protein | Proteobacteria |
| P.CW.IS05.2B | 4 | 7982599.625 | 3300002024_MIS_11567531 | ABC-type xylose transport system, periplasmic component | Spirochaetes |
| P.CW.IS05.2B | 4 | 58848500.63 | Ga0308317_10604841 | peptidyl-prolyl cis-trans isomerase C | Unclassified |
| P.CW.IS05.2B | 3 | 10301920.48 | Ga0308317_10446361 | hypothetical protein | Unclassified |
| P.CW.IS05.2B | 3 | 8065664.813 | Ga0308317_10957565 | cytochrome c551/c552/cytochrome c551/c552 | Unclassified |
| P.CW.IS05.3B | 1 | 23578650.88 | Ga0308317_12193107 | basic membrane protein A | Chloroflexi |
| P.CW.IS05.3B | 2 | 7156985.75 | 3300002026_MIS_100216394 | NA | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3352016.75 | 3300002027_MIS_100830233 | Rieske Fe-S protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3312989.25 | Ga0308317_10371593 | uncharacterized protein YjbI with pentapeptide repeats | Cyanobacteria |
| P.CW.IS05.3B | 3 | 3065662 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 2811694.75 | Ga0308317_14361226 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.3B | 1 | 41719973.56 | Ga0308317_152169619 | cytochrome c | Proteobacteria |
| P.CW.IS05.3B | 4 | 13424309.25 | 3300002027_MIS_100342344 | NA | Proteobacteria |
| P.CW.IS05.3B | 1 | 13006011 | 3300002026_MIS_100323342 | NA | Proteobacteria |
| P.CW.IS05.3B | 1 | 5577616.5 | Ga0308317_10627423 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 2 | 3744314.75 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.3B | 1 | 3456684.5 | Ga0308317_10286231 | branched-chain amino acid transport system substrate-binding protein | Proteobacteria |
| P.CW.IS05.3B | 1 | 2035364.25 | 3300002026_MIS_100239869 | Malate/lactate dehydrogenases | Unclassified |
| P.CW.IS05.3B | 4 | 19567269 | 3300002026_MIS_1000043424 | NA | Bacillariophyta |
| P.CW.IS05.3B | 2 | 1922936.75 | 3300002024_MIS_10742481 | NA | Bacillariophyta |
| P.CW.IS05.3B | 1 | 4246352.5 | Ga0308317_10358796 | CspA family cold shock protein | Bacteroidetes |
| P.CW.IS05.3B | 1 | 4481609 | 3300002026_MIS_1000034621 | Negative regulator of beta-lactamase expression | Chloroflexi |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 1 | 3837125.5 | 3300002026_MIS_100048397 | NA | Chloroflexi |
| P.CW.IS05.3B | 2 | 2698725.5 | Ga0308317_14702668 | Fe-S cluster assembly scaffold protein SufB/Fe-S cluster assembly scaffold protein SufB/intein/homing endonuclease | Chloroflexi |
| P.CW.IS05.3B | 1 | 1927873.375 | Ga0308317_14023463 | transcriptional regulator with XRE-family HTH domain | Chloroflexi |
| P.CW.IS05.3B | 4 | 31163831.36 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 14922452.38 | Ga0308317_11304452 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 3 | 11734610.38 | 3300002026_MIS_100260982 | NA | Cyanobacteria |
| P.CW.IS05.3B | 3 | 9579257 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 3 | 9189638.125 | Ga0308317_15235161 | SdrD B-like protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 8692499 | Ga0308317_12001942 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 8584189.484 | 3300002027_MIS_101516003 | Thioredoxin-like proteins and domains | Cyanobacteria |
| P.CW.IS05.3B | 2 | 8477408.063 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS05.3B | 2 | 6797525.5 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS05.3B | 3 | 5378106.625 | Ga0308317_12299312 | uncharacterized protein DUF4090 | Cyanobacteria |
| P.CW.IS05.3B | 2 | 5148796.875 | Ga0308317_14361226 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.3B | 1 | 4224724.5 | 3300002026_MIS_100192369 | Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family | Cyanobacteria |
| P.CW.IS05.3B | 2 | 4074015.5 | 3300002027_MIS_100432273 | Molybdopterin converting factor, small subunit | Cyanobacteria |
| P.CW.IS05.3B | 3 | 4013711.875 | Ga0308317_13201486 | glyoxylase-like metal-dependent hydrolase (beta-lactamase superfamily II) | Cyanobacteria |
| P.CW.IS05.3B | 2 | 2014106.125 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 1800541.375 | Ga0308317_11058012 | SdrD B-like protein/SdrD B-like protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 1658396.25 | Ga0308317_108417113 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS05.3B | 1 | 1031094 | 3300002024_MIS_11493163 | Ribbon-helix-helix protein, copG family | Cyanobacteria |
| P.CW.IS05.3B | 2 | 1001869.313 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 7 | 51263958.88 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 3 | 43122283.56 | 3300002027_MIS_100484477 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.3B | 2 | 18010107.66 | Ga0308317_126373915 | signal transduction histidine kinase | Proteobacteria |
| P.CW.IS05.3B | 5 | 12709922 | 3300002027_MIS_100225507 | Adenosine-5'-phosphosulfate reductase beta subunit; | Proteobacteria |
| P.CW.IS05.3B | 2 | 8304356 | Ga0308317_13495863 | tRNA 2-thiouridine synthesizing protein C | Proteobacteria |
| P.CW.IS05.3B | 4 | 7380567.813 | Ga0308317_10627422 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 7 | 6014417.266 | Ga0308317_11484977 | FKBP-type peptidyl-prolyl cis-trans isomerase FkpA/FKBP-type peptidyl-prolyl cis-trans isomerase FklB | Proteobacteria |
| P.CW.IS05.3B | 2 | 4336961.5 | 3300002027_MIS_101881042 | Cytochrome c551/c552 | Proteobacteria |
| P.CW.IS05.3B | 2 | 3550799 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS05.3B | 1 | 2629818.25 | 3300002027_MIS_100070992 | Cytochrome c553 | Proteobacteria |
| P.CW.IS05.3B | 1 | 2297021.875 | Ga0308317_11787262 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 3 | 2178868.625 | 3300002026_MIS_100147282 | Adenosine-5'-phosphosulfate reductase beta subunit; | Proteobacteria |

| P.CW.IS05.3B | 1 | 1836164.25 | 3300002026_MIS_100392708 | Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain | Proteobacteria |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 2 | 1719012.656 | 3300002026_MIS_100131184 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS05.3B | 2 | 1576608.625 | Ga0308317_146689064 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 1 | 1316789.25 | 3300002026_MIS_100175136 | FOG: CheY-like receiver | Proteobacteria |
| P.CW.IS05.3B | 1 | 1056402.875 | Ga0308317_10225605 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 1 | 1790937.125 | 3300002026_MIS_100112239 | Rubrerythrin | Spirochaetes |
| P.CW.IS05.3B | 1 | 18799977.63 | 3300002024_MIS_10822431 | NA | Unclassified |
| P.CW.IS05.3B | 1 | 4685816.5 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS05.3B | 1 | 2810937 | Ga0308317_10957565 | cytochrome c551/c552/cytochrome c551/c552 | Unclassified |
| P.CW.IS05.3B | 1 | 1110597.75 | Ga0308317_11025225 | elongation factor Tu | Unclassified |
| P.CW.IS05.3B | 2 | 1040068.188 | 3300002026_MIS_100313508 | ABC-type branched-chain amino acid transport systems, periplasmic component | Unclassified |
| P.CW.IS05.3B | 2 | 6454922.563 | Ga0308317_11038726 | outer membrane protein | Bacteroidetes |
| P.CW.IS05.3B | 1 | 14005516 | 3300002026_MIS_1000034633 | Ribosomal protein S1 | Chloroflexi |
| P.CW.IS05.3B | 4 | 37907081.75 | Ga0308317_14361226 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.3B | 4 | 22692000 | Ga0308317_14591192 | chromosome segregation ATPase | Cyanobacteria |
| P.CW.IS05.3B | 1 | 19080730 | Ga0308317_11212645 | serine protease Do | Cyanobacteria |
| P.CW.IS05.3B | 3 | 10884856.13 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 8295521 | Ga0308317_10498013 | ferredoxin | Cyanobacteria |
| P.CW.IS05.3B | 1 | 6669786 | 3300002024_MIS_11769591 | NA | Cyanobacteria |
| P.CW.IS05.3B | 5 | 57238257.25 | Ga0308317_125581024 | uncharacterized protein DUF3365 | Proteobacteria |
| P.CW.IS05.3B | 2 | 56685073 | 3300002027_MIS_100070992 | Cytochrome c553 | Proteobacteria |
| P.CW.IS05.3B | 3 | 10037516.19 | Ga0308317_131768026 | nitrate reductase alpha subunit | Proteobacteria |
| P.CW.IS05.3B | 1 | 3097730 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 5871158.5 | Ga0308317_10662321 | methyl-accepting chemotaxis protein | Acidobacteria |
| P.CW.IS05.3B | 5 | 483605202.2 | 3300002027_MIS_101946932 | CP12 domain | Cyanobacteria |
| P.CW.IS05.3B | 4 | 11358198.25 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS05.3B | 1 | 8664563 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS05.3B | 3 | 6376623.375 | Ga0308317_15235161 | SdrD B-like protein | Cyanobacteria |
| P.CW.IS05.3B | 2 | 4099739.938 | 3300002026_MIS_100020563 | Photosynthetic reaction centre protein | Cyanobacteria |
| P.CW.IS05.3B | 3 | 29017468.25 | 3300002026_MIS_1000222523 | Rubrerythrin | Proteobacteria |
| P.CW.IS05.3B | 5 | 21504155.13 | Ga0308317_14715166 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 12111763 | 3300002024_MIS_10816301 | Aspartate-semialdehyde dehydrogenase | Proteobacteria |
| P.CW.IS05.3B | 5 | 11100343.22 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 47546492 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.3B | 1 | 7242833.5 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS05.3B | 2 | 4767746 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 1 | 14365075 | Ga0308317_10869475 | Skp family chaperone for outer membrane proteins | Unclassified |
| P.CW.IS05.3B | 1 | 77990918 | 3300002024_MIS_10525911 | Chlorophyll A-B binding protein | Bacillariophyta |
| P.CW.IS05.3B | 4 | 26476494.13 | 3300002026_MIS_100004349 | Mg-chelatase subunit ChlI | Bacillariophyta |
| P.CW.IS05.3B | 2 | 21736649 | 3300002026_MIS_1000043414 | PsaD | Bacillariophyta |
| P.CW.IS05.3B | 3 | 20304189 | 3300002026_MIS_1000043442 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS05.3B | 4 | 12804011.5 | 3300002026_MIS_1000043424 | NA | Bacillariophyta |
| P.CW.IS05.3B | 3 | 9447015.875 | 3300002027_MIS_101644102 | Cathepsin propeptide inhibitor domain (I29) | Bacillariophyta |
| P.CW.IS05.3B | 1 | 4734367.5 | 3300002026_MIS_1000043412 | NA | Bacillariophyta |
| P.CW.IS05.3B | 1 | 3751042.75 | 3300002026_MIS_100363859 | F0F1-type ATP synthase, beta subunit | Bacillariophyta |
| P.CW.IS05.3B | 1 | 3068158 | 3300002027_MIS_101318131 | NA | Bacillariophyta |
| P.CW.IS05.3B | 1 | 3407672.5 | 3300002026_MIS_1000019335 | GTPases - translation elongation factors | Bacteroidetes |
| P.CW.IS05.3B | 1 | 3241309.5 | Ga0308317_13041134 | small subunit ribosomal protein S1 | Bacteroidetes |
| P.CW.IS05.3B | 1 | 28126765.5 | Ga0308317_129817518 | fumarate hydratase class II | Chloroflexi |
| P.CW.IS05.3B | 1 | 17447200 | 3300002026_MIS_1000034621 | Negative regulator of beta-lactamase expression | Chloroflexi |
| P.CW.IS05.3B | 4 | 12865705 | Ga0308317_141014330 | hypothetical protein | Chloroflexi |
| P.CW.IS05.3B | 1 | 11634776 | Ga0308317_10623353 | RNA recognition motif-containing protein | Chloroflexi |
| P.CW.IS05.3B | 3 | 8882899 | Ga0308317_136328629 | SH3 domain-containing protein | Chloroflexi |
| P.CW.IS05.3B | 1 | 6653278 | 3300002026_MIS_100109615 | Uncharacterized conserved protein - COG1259 | Chloroflexi |
| P.CW.IS05.3B | 3 | 6111250.625 | Ga0308317_13298891 | ABC-type xylose transport system substrate-binding protein/ABC-type xylose transport system substrate-binding protein | Chloroflexi |
| P.CW.IS05.3B | 3 | 5464420 | 3300002026_MIS_100048729 | NA | Chloroflexi |
| P.CW.IS05.3B | 1 | 5241630.5 | Ga0308317_11685233 | succinyl-CoA synthetase beta subunit | Chloroflexi |
| P.CW.IS05.3B | 1 | 4851018 | Ga0308317_11869791 | ParB family chromosome partitioning protein | Chloroflexi |
| P.CW.IS05.3B | 2 | 4586304.938 | Ga0308317_14101438 | adenosylhomocysteinase | Chloroflexi |
| P.CW.IS05.3B | 2 | 3949747 | Ga0308317_10734327 | pyruvate,orthophosphate dikinase | Chloroflexi |
| P.CW.IS05.3B | 1 | 3605845.25 | Ga0308317_12193107 | basic membrane protein A | Chloroflexi |
| P.CW.IS05.3B | 1 | 3383263.5 | Ga0308317_14702668 | Fe-S cluster assembly scaffold protein SufB/Fe-S cluster assembly scaffold protein SufB/intein/homing endonuclease | Chloroflexi |
| P.CW.IS05.3B | 1 | 3308827.25 | 3300002026_MIS_100089044 | Uncharacterized conserved protein - COG2835 | Chloroflexi |
| P.CW.IS05.3B | 1 | 3052001.75 | 3300002026_MIS_100044789 | NA | Chloroflexi |
| P.CW.IS05.3B | 1 | 2711182.75 | 3300002026_MIS_100219953 | NA | Chloroflexi |
| P.CW.IS05.3B | 8 | 209433096.1 | Ga0308317_123538692 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS05.3B | 1 | 173003600 | Ga0308317_13900858 | CP12 domain-containing protein | Cyanobacteria |
| P.CW.IS05.3B | 3 | 144630109.5 | Ga0308317_103184713 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 5 | 135752677.8 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS05.3B | 7 | 119904049.1 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS05.3B | 3 | 99549606.75 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 4 | 85477738.5 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS05.3B | 5 | 45893490.5 | Ga0308317_15235161 | SdrD B-like protein | Cyanobacteria |
| P.CW.IS05.3B | 4 | 45427942.13 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 3 | 33982017.25 | Ga0308317_14361226 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.3B | 1 | 28921364 | Ga0308317_14972441 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS05.3B | 3 | 28471074.75 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS05.3B | 2 | 28377849.25 | 3300002026_MIS_1000143515 | Ribosomal protein L7/L12 | Cyanobacteria |
| P.CW.IS05.3B | 2 | 25894084 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS05.3B | 3 | 25649765.5 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS05.3B | 3 | 20917700.5 | Ga0308317_10501821 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 3 | 16438735 | 3300002027_MIS_100199341 | Transketolase | Cyanobacteria |
| P.CW.IS05.3B | 1 | 15886946.75 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS05.3B | 3 | 15404495.25 | Ga0308317_113275065 | aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit C | Cyanobacteria |
| P.CW.IS05.3B | 2 | 14409920.88 | 3300002026_MIS_100385418 | Cold shock proteins | Cyanobacteria |
| P.CW.IS05.3B | 1 | 14264313.69 | Ga0308317_108417113 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS05.3B | 2 | 14128791.44 | 3300002027_MIS_100980214 | NA | Cyanobacteria |
| P.CW.IS05.3B | 3 | 13425290.25 | Ga0308317_14591192 | chromosome segregation ATPase | Cyanobacteria |
| P.CW.IS05.3B | 1 | 12482374.75 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 1 | 11897059.75 | 3300002027_MIS_100564723 | NA | Cyanobacteria |
| P.CW.IS05.3B | 2 | 11313987.63 | Ga0308317_12286322 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.3B | 1 | 10609140 | 3300002027_MIS_101516003 | Thioredoxin-like proteins and domains | Cyanobacteria |
| P.CW.IS05.3B | 1 | 10426553.75 | Ga0308317_13377662 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.3B | 2 | 9706138 | 3300002024_MIS_10897352 | Molybdopterin converting factor, small subunit | Cyanobacteria |
| P.CW.IS05.3B | 2 | 9572595.25 | 3300002027_MIS_100523012 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 8781773 | Ga0308317_113580413 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 1 | 8186132.063 | Ga0308317_105289928 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS05.3B | 2 | 7217443.5 | Ga0308317_108840816 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 1 | 6962911 | 3300002026_MIS_100039505 | Protease subunit of ATP-dependent Clp proteases | Cyanobacteria |
| P.CW.IS05.3B | 2 | 6729413.875 | Ga0308317_13201486 | glyoxylase-like metal-dependent hydrolase (beta-lactamase superfamily II) | Cyanobacteria |
| P.CW.IS05.3B | 2 | 6611903.75 | 3300002027_MIS_100237813 | F0F1-type ATP synthase, beta subunit | Cyanobacteria |
| P.CW.IS05.3B | 1 | 5771291 | Ga0308317_105563459 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 5546993 | 3300002027_MIS_100417313 | Manganese-stabilising protein / photosystem II polypeptide | Cyanobacteria |
| P.CW.IS05.3B | 3 | 5371825.344 | Ga0308317_13141007 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 1 | 4757133 | Ga0308317_10156479 | fused signal recognition particle receptor | Cyanobacteria |
| P.CW.IS05.3B | 1 | 4702869.75 | 3300002027_MIS_100094373 | Ribulose 1,5-bisphosphate carboxylase, large subunit | Cyanobacteria |
| P.CW.IS05.3B | 2 | 4613854.938 | Ga0308317_100693036 | uncharacterized protein DUF2862 | Cyanobacteria |
| P.CW.IS05.3B | 1 | 4507269 | Ga0308317_11058012 | SdrD B-like protein/SdrD B-like protein | Cyanobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 2 | 4468493.5 | 3300002026_MIS_1001557614 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.3B | 2 | 4194988.313 | Ga0308317_13519002 | phage shock protein A | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3719991 | Ga0308317_101469348 | protein phosphatase | Cyanobacteria |
| P.CW.IS05.3B | 2 | 3653143.5 | Ga0308317_122782739 | uncharacterized protein DUF3386 | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3636759.75 | Ga0308317_13957862 | aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit C | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3611891.25 | Ga0308317_11132412 | ATP-dependent Clp protease protease subunit | Cyanobacteria |
| P.CW.IS05.3B | 2 | 3592397.5 | Ga0308317_13013517 | Fe-S cluster biogenesis protein NfuA | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3552194.25 | Ga0308317_137341848 | uncharacterized protein DUF4090 | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3371295.75 | Ga0308317_15294701 | fused signal recognition particle receptor | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3367340.75 | 3300002024_MIS_11607303 | NA | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3091521.5 | Ga0308317_10511763 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3066212.75 | 3300002026_MIS_100260982 | NA | Cyanobacteria |
| P.CW.IS05.3B | 3 | 2903060.375 | 3300002027_MIS_101412592 | Heme oxygenase | Cyanobacteria |
| P.CW.IS05.3B | 1 | 2770495 | 3300002027_MIS_100308741 | Phage shock protein A (IM30), suppresses sigma54-dependent transcription | Cyanobacteria |
| P.CW.IS05.3B | 1 | 2597369.75 | Ga0308317_100693045 | elongation factor Tu | Cyanobacteria |
| P.CW.IS05.3B | 1 | 2535603.25 | 3300002026_MIS_1000077318 | Ribosomal protein S13 | Cyanobacteria |
| P.CW.IS05.3B | 2 | 196044419 | 3300002026_MIS_100316452 | Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases | Proteobacteria |
| P.CW.IS05.3B | 3 | 181509974 | Ga0308317_14352243 | acetyl-CoA synthetase | Proteobacteria |
| P.CW.IS05.3B | 2 | 54898188.63 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 2 | 39770948.63 | 3300002026_MIS_100201834 | Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases | Proteobacteria |
| P.CW.IS05.3B | 1 | 22118398.75 | 3300002026_MIS_1002605312 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS05.3B | 1 | 21915028.25 | 3300002026_MIS_100072245 | Enolase | Proteobacteria |
| P.CW.IS05.3B | 1 | 17250124 | Ga0308317_14715166 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 12253165.88 | 3300002027_MIS_100157284 | Desulfoferrodoxin | Proteobacteria |
| P.CW.IS05.3B | 3 | 11312721.13 | 3300002026_MIS_100374701 | Ribulose 1,5-bisphosphate carboxylase, large subunit | Proteobacteria |
| P.CW.IS05.3B | 3 | 9302733.5 | Ga0308317_10307526 | rubrerythrin | Proteobacteria |
| P.CW.IS05.3B | 3 | 9247937.25 | Ga0308317_101705613 | chaperonin GroEL | Proteobacteria |
| P.CW.IS05.3B | 3 | 8643000.375 | 3300002027_MIS_100155414 | Methyl-accepting chemotaxis protein | Proteobacteria |
| P.CW.IS05.3B | 1 | 7111178.875 | Ga0308317_10632712 | branched-chain amino acid transport system substrate-binding protein | Proteobacteria |
| P.CW.IS05.3B | 2 | 7020862.313 | 3300002026_MIS_100110857 | Cache domain | Proteobacteria |
| P.CW.IS05.3B | 1 | 6754132.5 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 6725086.5 | Ga0308317_11117232 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 6638485.125 | 3300002026_MIS_100319374 | NA | Proteobacteria |
| P.CW.IS05.3B | 1 | 6351734.75 | Ga0308317_10672482 | ATP-dependent HslUV protease ATP-binding subunit HslU | Proteobacteria |
| P.CW.IS05.3B | 2 | 6064146 | Ga0308317_10047734 | rubrerythrin | Proteobacteria |
| P.CW.IS05.3B | 2 | 5645731.313 | 3300002026_MIS_100090083 | Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 2 | 5642314.5 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 5610561.5 | Ga0308317_10722186 | putative sensory transduction regulator | Proteobacteria |
| P.CW.IS05.3B | 3 | 5402743.75 | 3300002027_MIS_100225507 | Adenosine-5'-phosphosulfate reductase beta subunit; | Proteobacteria |
| P.CW.IS05.3B | 1 | 5337681.5 | Ga0308317_12414362 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 3 | 5113290.25 | 3300002026_MIS_100101317 | Putative phospholipid-binding domain | Proteobacteria |
| P.CW.IS05.3B | 1 | 4926087.5 | Ga0308317_11578061 | Na+-translocating ferredoxin:NAD+ oxidoreductase RNF subunit RnfB | Proteobacteria |
| P.CW.IS05.3B | 2 | 3876035 | Ga0308317_14839643 | F-type H+-transporting ATPase subunit alpha | Proteobacteria |
| P.CW.IS05.3B | 1 | 3472967 | Ga0308317_14162442 | tellurite resistance protein TerA | Proteobacteria |
| P.CW.IS05.3B | 1 | 3224610.5 | Ga0308317_11167621 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 1 | 3187374 | 3300002027_MIS_100894105 | 5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase | Proteobacteria |
| P.CW.IS05.3B | 1 | 2921625.5 | Ga0308317_13666641 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 2 | 2920965.625 | Ga0308317_12551502 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 2 | 2811898.813 | 3300002026_MIS_100131184 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS05.3B | 2 | 5676720 | 3300002027_MIS_100245793 | NA | Thermotogae |
| P.CW.IS05.3B | 1 | 94340752 | 3300002024_MIS_10538482 | Thiamine pyrophosphate enzyme, N-terminal TPP binding domain | Unclassified |
| P.CW.IS05.3B | 3 | 88142191.5 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS05.3B | 2 | 27553393.75 | 3300002026_MIS_100228095 | Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases | Unclassified |
| P.CW.IS05.3B | 3 | 22033932 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS05.3B | 3 | 11040670 | 3300002027_MIS_101802961 | LETM1-like protein | Unclassified |
| P.CW.IS05.3B | 4 | 7440054.625 | 3300002026_MIS_1000860718 | Chlorophyll A-B binding protein | Unclassified |
| P.CW.IS05.3B | 2 | 6924424 | 3300002027_MIS_101631721 | NA | Unclassified |
| P.CW.IS05.3B | 1 | 5939135 | 3300002026_MIS_100394952 | Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases | Unclassified |
| P.CW.IS05.3B | 2 | 5376517.531 | 3300002026_MIS_100315161 | ATPases with chaperone activity, ATP-binding subunit | Unclassified |
| P.CW.IS05.3B | 1 | 4729128 | 3300002027_MIS_101567531 | Actin and related proteins | Unclassified |
| P.CW.IS05.3B | 2 | 4471605.875 | 3300002024_MIS_11659621 | Ubiquitin family | Unclassified |
| P.CW.IS05.3B | 1 | 4103518.5 | 3300002024_MIS_10307593 | NA | Unclassified |
| P.CW.IS05.3B | 2 | 3867418.625 | 3300002026_MIS_100239869 | Malate/lactate dehydrogenases | Unclassified |
| P.CW.IS05.3B | 2 | 3492048.5 | 3300002027_MIS_100302972 | NA | Unclassified |
| P.CW.IS05.3B | 1 | 3392431.25 | 3300002024_MIS_10637324 | NA | Unclassified |
| P.CW.IS05.3B | 1 | 3371634.5 | Ga0308317_14115661 | hypothetical protein | Unclassified |
| P.CW.IS05.3B | 1 | 3334964.5 | Ga0308317_10604841 | peptidyl-prolyl cis-trans isomerase C | Unclassified |
| P.CW.IS05.3B | 3 | 83633961 | Ga0308317_10157129 | propionyl-CoA synthetase | Unclassified |
| P.CW.IS05.3B | 1 | 59236956 | Ga0308317_12184693 | hypothetical protein | Unclassified |
| P.CW.IS05.3B | 1 | 43521948 | Ga0308317_15263983 | cohesin domain-containing protein | Unclassified |
| P.CW.IS05.3B | 1 | 4075204.75 | Ga0308317_142867611 | elongation factor Tu | Unclassified |
| P.CW.IS05.3B | 3 | 24149482.5 | 3300002026_MIS_1000043442 | F0F1-type ATP synthase, alpha subunit | Bacillariophyta |
| P.CW.IS05.3B | 2 | 8659895.75 | 3300002026_MIS_1000043414 | PsaD | Bacillariophyta |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 2 | 6721679.5 | 3300002026_MIS_1000043428 | Ferredoxin | Bacillariophyta |
| P.CW.IS05.3B | 2 | 7099308.25 | Ga0308317_121057211 | hypothetical protein | Bacteroidetes |
| P.CW.IS05.3B | 5 | 13530804.31 | Ga0308317_112152823 | large subunit ribosomal protein L17 | Bacteroidetes |
| P.CW.IS05.3B | 3 | 52151534.63 | Ga0308317_14023463 | transcriptional regulator with XRE-family HTH domain | Chloroflexi |
| P.CW.IS05.3B | 2 | 15208773.25 | Ga0308317_142163210 | glycine cleavage system H protein | Chloroflexi |
| P.CW.IS05.3B | 1 | 11149664 | Ga0308317_11685233 | succinyl-CoA synthetase beta subunit | Chloroflexi |
| P.CW.IS05.3B | 1 | 10054182.25 | Ga0308317_14165545 | phosphoglycerate kinase | Chloroflexi |
| P.CW.IS05.3B | 2 | 7915061 | Ga0308317_10712513 | Uma2 family endonuclease | Chloroflexi |
| P.CW.IS05.3B | 1 | 5042610 | Ga0308317_136328629 | SH3 domain-containing protein | Chloroflexi |
| P.CW.IS05.3B | 2 | 4910150.25 | Ga0308317_14101438 | adenosylhomocysteinase | Chloroflexi |
| P.CW.IS05.3B | 2 | 4860981.375 | Ga0308317_144015415 | putative zinc ribbon protein | Chloroflexi |
| P.CW.IS05.3B | 1 | 4414184.5 | Ga0308317_141014330 | hypothetical protein | Chloroflexi |
| P.CW.IS05.3B | 1 | 3533739.5 | Ga0308317_13224551 | multiple sugar transport system substrate-binding protein | Chloroflexi |
| P.CW.IS05.3B | 8 | 331665145 | Ga0308317_122782772 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS05.3B | 3 | 107427300.8 | 3300002027_MIS_100217001 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.3B | 6 | 103755380.3 | Ga0308317_10126032 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 101574318.8 | Ga0308317_125818810 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS05.3B | 4 | 88887232.88 | Ga0308317_108981826 | allophycocyanin alpha subunit | Cyanobacteria |
| P.CW.IS05.3B | 1 | 79842306 | Ga0308317_123538665 | F-type H+-transporting ATPase subunit beta | Cyanobacteria |
| P.CW.IS05.3B | 8 | 68583995.13 | Ga0308317_122782739 | uncharacterized protein DUF3386 | Cyanobacteria |
| P.CW.IS05.3B | 6 | 47163736.5 | Ga0308317_123538692 | photosystem I subunit 4 | Cyanobacteria |
| P.CW.IS05.3B | 3 | 41755048.75 | Ga0308317_12056721 | phosphoglycerate kinase | Cyanobacteria |
| P.CW.IS05.3B | 2 | 37293940.5 | 3300002027_MIS_100980214 | NA | Cyanobacteria |
| P.CW.IS05.3B | 1 | 26569876 | Ga0308317_122782710 | phycocyanin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 2 | 25385306.5 | Ga0308317_108840816 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 4 | 23659740.38 | Ga0308317_108417113 | photosystem I subunit 7 | Cyanobacteria |
| P.CW.IS05.3B | 2 | 22456781 | 3300002027_MIS_101362743 | NA | Cyanobacteria |
| P.CW.IS05.3B | 2 | 19356314.75 | Ga0308317_14361225 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 3 | 16669348.13 | Ga0308317_113580413 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 4 | 16366631.38 | Ga0308317_15235161 | SdrD B-like protein | Cyanobacteria |
| P.CW.IS05.3B | 3 | 13391691.63 | Ga0308317_13377662 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.3B | 2 | 12428155.88 | Ga0308317_122782711 | phycocyanin alpha chain | Cyanobacteria |
| P.CW.IS05.3B | 2 | 9111967.5 | Ga0308317_106365312 | photosystem II P680 reaction center D2 protein | Cyanobacteria |
| P.CW.IS05.3B | 4 | 9005714.125 | Ga0308317_111576569 | chaperonin GroEL | Cyanobacteria |
| P.CW.IS05.3B | 1 | 8857891.5 | 3300002027_MIS_101516003 | Thioredoxin-like proteins and domains | Cyanobacteria |
| P.CW.IS05.3B | 2 | 8716596.25 | 3300002026_MIS_100156017 | NA | Cyanobacteria |
| P.CW.IS05.3B | 1 | 7824567 | Ga0308317_14972441 | phosphoglycerate kinase | Cyanobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 3 | 7423135.5 | Ga0308317_13201486 | glyoxylase-like metal-dependent hydrolase (beta-lactamase superfamily II) | Cyanobacteria |
| P.CW.IS05.3B | 1 | 7333149.5 | Ga0308317_11304452 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 2 | 7033560.25 | Ga0308317_12286322 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.3B | 1 | 6142287.625 | Ga0308317_14016542 | enolase | Cyanobacteria |
| P.CW.IS05.3B | 2 | 5971697.75 | 3300002026_MIS_1001557614 | Phycobilisome protein | Cyanobacteria |
| P.CW.IS05.3B | 3 | 5731843.75 | 3300002026_MIS_100100153 | NA | Cyanobacteria |
| P.CW.IS05.3B | 3 | 5616494.875 | 3300002027_MIS_101946932 | CP12 domain | Cyanobacteria |
| P.CW.IS05.3B | 3 | 4848581.313 | 3300002026_MIS_100192369 | Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family | Cyanobacteria |
| P.CW.IS05.3B | 3 | 4754573 | Ga0308317_14375973 | phycoerythrin beta chain | Cyanobacteria |
| P.CW.IS05.3B | 2 | 4560645 | Ga0308317_14361226 | phycoerythrin alpha chain | Cyanobacteria |
| P.CW.IS05.3B | 1 | 4019650.5 | Ga0308317_11765081 | hypothetical protein | Cyanobacteria |
| P.CW.IS05.3B | 2 | 3985674.75 | Ga0308317_10420022 | photosystem II protein | Cyanobacteria |
| P.CW.IS05.3B | 1 | 3409689 | 3300002027_MIS_100564723 | NA | Cyanobacteria |
| P.CW.IS05.3B | 6 | 225310238.1 | Ga0308317_12760225 | ferredoxin | Proteobacteria |
| P.CW.IS05.3B | 2 | 60321049.38 | 3300002027_MIS_100755711 | Outer membrane protein and related peptidoglycan-associated (lipo)proteins | Proteobacteria |
| P.CW.IS05.3B | 1 | 46611059 | 3300002026_MIS_100090083 | Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains | Proteobacteria |
| P.CW.IS05.3B | 7 | 45152742.06 | 3300002027_MIS_101963794 | NA | Proteobacteria |
| P.CW.IS05.3B | 6 | 39697117.34 | Ga0308317_13589363 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 39371623.75 | 3300002026_MIS_100319374 | NA | Proteobacteria |
| P.CW.IS05.3B | 1 | 36857242.94 | 3300002027_MIS_100877852 | Cache domain | Proteobacteria |
| P.CW.IS05.3B | 3 | 31566218.5 | 3300002027_MIS_100632912 | NA | Proteobacteria |
| P.CW.IS05.3B | 8 | 24890444.38 | Ga0308317_153992221 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 18628540 | 3300002026_MIS_100316452 | Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases | Proteobacteria |
| P.CW.IS05.3B | 1 | 18343983.97 | 3300002027_MIS_100094537 | NA | Proteobacteria |
| P.CW.IS05.3B | 3 | 18218902.88 | 3300002026_MIS_1000222523 | Rubrerythrin | Proteobacteria |
| P.CW.IS05.3B | 3 | 17836315.44 | Ga0308317_14715166 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 17490868 | Ga0308317_14549566 | methyl-accepting chemotaxis protein/methyl-accepting chemotaxis protein | Proteobacteria |
| P.CW.IS05.3B | 8 | 15770387.16 | 3300002027_MIS_101918742 | NA | Proteobacteria |
| P.CW.IS05.3B | 3 | 14255296.88 | Ga0308317_14787195 | malate dehydrogenase | Proteobacteria |
| P.CW.IS05.3B | 2 | 13209387.75 | 3300002026_MIS_100312755 | Formyltetrahydrofolate synthetase | Proteobacteria |
| P.CW.IS05.3B | 3 | 11822966.88 | Ga0308317_13100247 | peptidoglycan-associated lipoprotein | Proteobacteria |
| P.CW.IS05.3B | 1 | 9313223 | Ga0308317_10677553 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 4 | 7882102.25 | 3300002027_MIS_100032986 | NA | Proteobacteria |
| P.CW.IS05.3B | 1 | 7335922 | Ga0308317_11787262 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 1 | 6248888 | 3300002026_MIS_100072245 | Enolase | Proteobacteria |

| | | | | | |
|---|---|---|---|---|---|
| P.CW.IS05.3B | 1 | 6140068 | 3300002026_MIS_1000679112 | Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains | Proteobacteria |
| P.CW.IS05.3B | 3 | 5884253.813 | Ga0308317_10047734 | rubrerythrin | Proteobacteria |
| P.CW.IS05.3B | 2 | 5796966.625 | Ga0308317_10421892 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 1 | 4640062.5 | Ga0308317_14162442 | tellurite resistance protein TerA | Proteobacteria |
| P.CW.IS05.3B | 1 | 4569759.5 | Ga0308317_10281003 | signal transduction histidine kinase | Proteobacteria |
| P.CW.IS05.3B | 4 | 4489574.781 | Ga0308317_15209112 | uncharacterized protein DUF4124 | Proteobacteria |
| P.CW.IS05.3B | 2 | 4014346 | Ga0308317_10164838 | trigger factor | Proteobacteria |
| P.CW.IS05.3B | 5 | 3508437.375 | 3300002027_MIS_100225507 | Adenosine-5'-phosphosulfate reductase beta subunit; | Proteobacteria |
| P.CW.IS05.3B | 2 | 3506647 | Ga0308317_13495863 | tRNA 2-thiouridine synthesizing protein C | Proteobacteria |
| P.CW.IS05.3B | 3 | 3491569.188 | 3300002027_MIS_101676581 | ATP-dependent protease HslVU (ClpYQ), ATPase subunit | Proteobacteria |
| P.CW.IS05.3B | 1 | 3308430.75 | 3300002026_MIS_1000796610 | GTPases - translation elongation factors | Proteobacteria |
| P.CW.IS05.3B | 2 | 3248220.563 | Ga0308317_12080032 | NADPH-dependent glutamate synthase beta subunit-like oxidoreductase/protoporphyrinogen oxidase | Proteobacteria |
| P.CW.IS05.3B | 2 | 3207335.875 | 3300002027_MIS_101006588 | NA | Proteobacteria |
| P.CW.IS05.3B | 2 | 3018759.625 | Ga0308317_14393102 | hypothetical protein | Proteobacteria |
| P.CW.IS05.3B | 5 | 25974285.38 | 3300002027_MIS_100245793 | NA | Thermotogae |
| P.CW.IS05.3B | 1 | 248059501 | 3300002024_MIS_10574951 | Actin and related proteins | Unclassified |
| P.CW.IS05.3B | 3 | 97536917.75 | 3300002027_MIS_101091071 | NA | Unclassified |
| P.CW.IS05.3B | 1 | 79874962 | 3300002024_MIS_10666431 | NA | Unclassified |
| P.CW.IS05.3B | 3 | 31797536 | 3300002026_MIS_1000769710 | Phycobilisome protein | Unclassified |
| P.CW.IS05.3B | 4 | 19398328.38 | 3300002026_MIS_100076979 | Phycobilisome Linker polypeptide | Unclassified |
| P.CW.IS05.3B | 1 | 18731930 | 3300002027_MIS_101217361 | Repulsive guidance molecule (RGM) C-terminus | Unclassified |
| P.CW.IS05.3B | 1 | 17641650 | 3300002026_MIS_100315161 | ATPases with chaperone activity, ATP-binding subunit | Unclassified |
| P.CW.IS05.3B | 2 | 9054885.875 | Ga0308317_11002012 | elongation factor Tu | Unclassified |
| P.CW.IS05.3B | 2 | 6957478 | 3300002026_MIS_1000860718 | Chlorophyll A-B binding protein | Unclassified |
| P.CW.IS05.3B | 1 | 6621409 | 3300002024_MIS_10215871 | Starch binding domain | Unclassified |
| P.CW.IS05.3B | 2 | 6268601.125 | 3300002026_MIS_1000769711 | Phycobilisome protein | Unclassified |
| P.CW.IS05.3B | 3 | 5690603 | 3300002027_MIS_101658931 | NA | Unclassified |
| P.CW.IS05.3B | 2 | 5218701.5 | Ga0308317_11746661 | putative multiple sugar transport system substrate-binding protein | Unclassified |
| P.CW.IS05.3B | 2 | 4234287 | 3300002026_MIS_100309382 | ABC-type branched-chain amino acid transport systems, periplasmic component | Unclassified |
| P.CW.IS05.3B | 1 | 3998050 | 3300002027_MIS_100977031 | NA | Unclassified |
| P.CW.IS05.3B | 1 | 3436390.375 | 3300002027_MIS_100677291 | Chlorophyll A-B binding protein | Unclassified |
| P.CW.IS05.3B | 3 | 3383186.813 | 3300002027_MIS_101567531 | Actin and related proteins | Unclassified |
| P.CW.IS05.3B | 1 | 3301533.25 | 3300002027_MIS_100254473 | NA | Unclassified |
| P.CW.IS05.3B | 1 | 3157690.75 | 3300002026_MIS_100384576 | Predicted ATPase (AAA+ superfamily) - COG1373 | Unclassified |
| P.CW.IS05.3B | 2 | 86977001 | Ga0308317_10955155 | chaperonin GroEL | Unclassified |

| P.CW.IS05.3B | 2 | 2943507.875 | Ga0308317_10157129 | propionyl-CoA synthetase | Unclassified |
|---|---|---|---|---|---|
| P.CW.IS6.1D | 1 | 182272080 | Ga0308317_11095951 | cyclic beta-1,2-glucan synthetase | Proteobacteria |
| P.CW.IS6.1D | 1 | 58814700 | Ga0308317_14109403 | translation initiation factor IF-2 | Bacteroidetes |
| P.CW.IS6.1D | 1 | 1494433.25 | 3300002027_MIS_101148022 | NA | Proteobacteria |
| P.CW.IS6.1D | 1 | 48695628 | 3300002027_MIS_101372145 | NA | Unclassified |
| P.CW.IS10.1G | 1 | 58727252 | 3300002026_MIS_100377309 | NA | Proteobacteria |
| P.CW.IS10.1G | 1 | 13132601 | Ga0308317_11413212 | hypothetical protein | Unclassified |
| P.CW.IS16.1D | 1 | 69666448 | Ga0308317_111840611 | site-specific recombinase XerD | Unclassified |

**Table S2.** Phylogenetic composition of sample LH47 as estimated by summed integrated peak area of peptides assigned to proteins.

| Phylum | Summed peak area of peptides | Abundance by peptide peak area (%) |
|---|---|---|
| Cyanobacteria | $1.10 \times 10^{10}$ | 29.1 |
| Proteobacteria | $1.75 \times 10^{10}$ | 46.3 |
| Chloroflexi | $1.79 \times 10^{9}$ | 4.7 |
| Bacillariophyta | $7.39 \times 10^{8}$ | 2.0 |
| Bacteroidetes | $1.23 \times 10^{9}$ | 3.2 |
| Spirochaetes | $8.35 \times 10^{7}$ | 0.2 |
| Thermotogae | $1.45 \times 10^{8}$ | 0.4 |
| Acidobacteria | $5.87 \times 10^{6}$ | 0.0 |
| Chlorobi | $5.61 \times 10^{6}$ | 0.0 |
| Unclassified | $5.33 \times 10^{9}$ | 14.1 |
| Within Proteobacteria | | |
| Class | Summed peak area of peptides | Abundance by peptide peak area (%) |
| Alpha | $1.93 \times 10^{8}$ | 1.1 |
| Beta | $6.39 \times 10^{7}$ | 0.4 |
| Delta | $1.15 \times 10^{10}$ | 65.4 |
| Epsilon | $2.22 \times 10^{7}$ | 0.1 |
| Gamma | $4.82 \times 10^{9}$ | 27.5 |
| Other | $9.68 \times 10^{8}$ | 5.5 |

**Table S3.** Mean, standard deviation (SD) of full population and average measurement error (AME) of triplicate measurements for $\delta^{13}C$ values of individual wells ($n = 133$), separated into quartiles by decreasing IRMS peak area (Vs).

| | Q1-2 (‰) | Q3 (‰) | Q4 (‰) | All (‰) |
|---|---|---|---|---|
| Mean | -24.6 | -24.5 | -24.8 | -24.6 |
| SD | 1.5 | 1.5 | 1.8 | 1.6 |
| AME | 0.5 | 0.87 | 0.84 | 0.7 |

**Table S4.** Mean $\delta^{13}C$ composition and standard deviation (SD) of RP fractions used used for P-SIF linear regression analyses ($n = 43$). RP fraction labels correspond to those in Table S1.

| RP Fraction | $\delta^{13}C$ (‰) | SD (‰) |
|---|---|---|
| P.CW.IS03.1D | -24.7 | 1.3 |
| P.CW.IS03.1E | -24.7 | 0.2 |
| P.CW.IS03.1F | -24.6 | 0.4 |
| P.CW.IS03.1G | -24.6 | 0.7 |
| P.CW.IS03.1H | -24.4 | 0.3 |
| P.CW.IS04.1D | -23.8 | 0.1 |
| P.CW.IS04.1E | -24.2 | 0.0 |
| P.CW.IS04.1F | -24.3 | 0.3 |
| P.CW.IS04.1G | -24.0 | 0.3 |
| P.CW.IS04.1H | -23.8 | 0.2 |
| P.CW.IS04.2A | -23.4 | 0.2 |
| P.CW.IS04.2B | -23.3 | 0.4 |
| P.CW.IS04.3B | -29.1 | 1.0 |
| P.CW.IS04.3F | -27.0 | 0.3 |
| P.CW.IS05.1D | -25.0 | 0.3 |
| P.CW.IS05.1F | -23.8 | 0.4 |
| P.CW.IS05.1G | -23.3 | 0.3 |
| P.CW.IS05.1H | -22.8 | 0.7 |
| P.CW.IS05.2A | -22.9 | 0.2 |
| P.CW.IS05.2B | -23.9 | 0.3 |
| P.CW.IS05.3B | -30.2 | 0.4 |
| P.CW.IS6.1D | -24.9 | 0.1 |
| P.CW.IS06.1F | -23.8 | 0.0 |
| P.CW.IS06.1G | -23.1 | 0.2 |
| P.CW.IS06.1H | -23.0 | 0.4 |

| | | |
|---|---|---|
| P.CW.IS06.2A | -22.7 | 0.6 |
| P.CW.IS06.2B | -23.6 | 0.2 |
| P.CW.IS06.2G | -23.8 | 0.2 |
| P.CW.IS06.3B | -25.8 | 1.0 |
| P.CW.IS07.1D | -25.1 | 0.4 |
| P.CW.IS07.1E | -24.4 | 0.5 |
| P.CW.IS07.1F | -24.6 | 0.6 |
| P.CW.IS07.1G | -22.0 | 1.3 |
| P.CW.IS07.3B | -23.5 | 0.2 |
| P.CW.IS07.3C | -22.8 | 1.9 |
| P.CW.IS07.3D | -21.5 | 0.4 |
| P.CW.IS08.1F | -25.5 | 0.4 |
| P.CW.IS09.1D | -26.4 | 0.3 |
| P.CW.IS09.1F | -25.9 | 0.7 |
| P.CW.IS09.1H | -25.9 | 0.9 |
| P.CW.IS10.1E | -26.3 | 0.5 |
| P.CW.IS10.1F | -25.3 | 0.5 |
| P.CW.IS16.1D | -22.0 | 0.7 |