12-1-2015

# The impact of temperature, pH and environmental heterogeneity on prokaryotic diversity in Yellowstone National Park thermal springs

Xiaoben Jiang

Xiaoben Jiang

*Candidate*

Biology

*Department*

This dissertation is approved, and it is acceptable in quality and form for publication:

*Approved by the Dissertation Committee:*

Cristina Takacs-Vesbach, Chairperson

Robert Sinsabaugh

Diana Northup

Laura Crossey

**THE IMPACT OF TEMPERATURE, PH AND ENVIRONMENTAL HETEROGENEITY ON PROKARYOTIC DIVERSITY IN YELLOWSTONE NATIONAL PARK THERMAL SPRINGS**

by

**XIAOBEN JIANG**

B.S., Biotechnology, Shanghai Fisheries University, 2000

M.S., Biology, Brigham Young University, 2010

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Doctor of Philosophy**
**Biology**

The University of New Mexico
Albuquerque, New Mexico

**December 2015**

## DEDICATION

To my parents

## ACKNOWLEDGMENTS

With sincere gratitude, I thank my advisor, Dr. Cristina Takacs-Vesbach, for her guidance and support throughout the process of my dissertation work. Her broad knowledge, deep insight and great enthusiasm promoted the development and accomplishment of my dissertation projects. Not only has she enhanced my research experience, but my future career in science as well. The four years working with Dr. Takacs-Vesbach will benefit me all through my life.

I am indebted to my committee members, Dr. Robert Sinsabaugh, Dr. Diana Northup and Dr. Laura Crossey for their advice and support. I mostly owe thanks to Daniel Colman, who has provided his time, advice, enlightenment and encouragement and has been my best friend since we met. I would like to specially thank: David Van Horn, Jordan Okie, George Rosenberg and Heather Buelow.

I also want to extend thanks to my mother and father for their unending love, support and encouragement, throughout the pursuit of my PhD.

**The Impact of Temperature, pH and Environmental Heterogeneity on Prokaryotic Diversity in Yellowstone National Park Thermal Springs**

by

**Xiaoben Jiang**

## ABSTRACT

Yellowstone National Park (YNP) is one of the largest and most diverse hydrothermal areas on Earth. Extensive culture-independent studies in YNP thermal springs have shown dramatic taxonomic and metabolic diversity in microbial communities. We conducted a survey of bacterial communities along temperature gradients in three alkaline springs with similar geochemistries at the local scale. With these data, we investigated the influence of environmental variables on bacterial community diversity and assemblages along a broad temperature range using high throughput 454 pyrosequencing. Previous studies have suggested that pH is the driver of microbial diversity in thermal springs among geographical regions or at the global scale, whereas the temperature is an important factor controlling microbial diversity in springs within a similar pH range or within an individual geographical region. Our results revealed that temperature was the most important environmental variable to shape the structure of bacterial communities at the local scale. Similar community structure was observed in the sites with similar temperatures, irrespective of the origin of the thermal spring. The results of this study expand our current knowledge of the role of temperature in controlling community structure in YNP alkaline thermal springs.

In addition to temperature, pH is also a crucial factor in determining microbial assemblages in thermal springs. Although thermal springs in YNP are characterized by a broad range of pH (1-10), the pH distribution is bimodal. Most of springs are either vapor-dominated acidic springs or water-dominated neutral to slightly alkaline springs. Yet the intermediate pH (e.g., pH 4-5) habitats have been overlooked. The dearth of information on microbial communities in the intermediate pH sites has still hindered an understanding of the whole picture of microbial diversity in YNP thermal springs. We conducted the first metagenomic investigation of microbial phylogenetic and functional diversity in a pH 4 and low temperature site. We uncovered a high proportion of Chloroflexi, Bacteroidetes, Proteobacteria and Firmicutes in this site. Functional comparison indicated that the community was enriched in the COG functions related to energy production and conversion, transcription and carbohydrate transport, possibly to result from high microbial dynamics. This is the first study to examine a pH 4 and low temperature habitat, which is a key step towards an understanding of the whole picture of microbial diversity in YNP thermal springs.

Local environmental heterogeneity may also be responsible for the assemblages and distribution of some microbial groups. We examined microbial diversity and community composition in ten filamentous thermal springs with similar physiochemical properties in the Shoshone area by using the barcoded amplicon pyrosequencing approach. Despite all these samples from the same type of springs, statistical analyses suggest that the relatively small variation of environmental variables such as temperature, pH and conductivity among the springs can shape microbial composition. Additionally, we conducted a metagenomic analysis on a previously uninvestigated high temperature

and pH filamentous habitat. The results indicated that the site was exclusively dominated by the family Aquificaceae, and functions related to aerobic respiration and amino acid synthesis were enriched.

The research presented here is an in-depth investigation of thermal springs, enhanced by high-throughput community sequencing and a combination of environmental data. The work will fill the knowledge gap for microbial communities in uninvestigated habitats and offer a basis for understanding microbial ecology in global thermal springs.

## Table of Contents

# Chapter 1

## Microbes and Microbial Ecology in Thermal Springs of Yellowstone National Park

### Background

Microbial organisms are the most abundant and diverse organisms on Earth and can colonize every possible niche with an exploitable source of energy, accounting for most of the evolutionary diversity on Earth. Over the past century, the discovery of extreme environments and their microbial inhabitants have provided opportunities to study the extent to which physicochemical settings control the distribution, diversity and physiology of life. Microbial life in extreme environments attracts broad scientific interests. Owing to the analogous environments found on other celestial bodies such Mars, Europa and Titan (Cavicchioli 2002), extreme environments on Earth have made it more plausible to search extraterrestrial life and also to shed light on the possibility of panspermia (Rothschild and Mancinelli 2001).

Important members of extreme environments, terrestrial thermal springs are formed as the result of geothermally heated groundwater emerging from the Earth's crust. Terrestrial thermal springs are globally distributed. The earliest observation and description of the life in thermal springs of Yellowstone National Park (YNP) can date back to the late 1890s (Davis 1897). However, microbial life and microbial communities in thermal springs were not in the spotlight until the discovery and isolation of hyperthermophiles by Thomas D. Brock from thermal springs of YNP in the late 1960s (Brock and Brock 1978). Research on the extent of microbial diversity was sorely restricted by traditionally culture-dependent approaches, because only less than 1% of microbes in environments can be cultivated (Pace 1997; Rappé and Giovannoni 2003;

Schoenborn et al. 2004). During the past two decades, scientists have made great progress in understanding of the microbial life and communities in high temperature environments, due to the great leap in culture-independent molecular methods. For instance, recent advances in utilizing DNA sequencing and bioinformatics technologies are highlighted by our nascent understanding the diversity of microbial communities. Therefore, microbial communities in thermal springs have been studied throughout different continents, such as those in YNP, USA (Barns et al. 1996; Barns et al. 1994; De León et al. 2013; Hall et al. 2008; Hugenholtz et al. 1998; Inskeep et al. 2013a; Inskeep et al. 2013b; Inskeep et al. 2010; Klatt et al. 2013; Kozubal et al. 2012; Kozubal et al. 2013; Macur et al. 2013; Meyer-Dombard et al. 2005; Miller et al. 2009; Mitchell 2009; Pace 1997; Swingley et al. 2012; Takacs-Vesbach et al. 2013), Great Boiling Spring, USA (Cole et al. 2013), Lassen Volcanic National Park, USA (Siering et al. 2013; Wilson et al. 2008), El Tatio Geyser Field, Chile (Engel et al. 2013), Japan (Everroad et al. 2012; Nakagawa and Fukui 2002; Otaki et al. 2012), Tengchong and Tibet, China (Huang et al. 2011; Jiang et al. 2012; Lau 2007; Lau et al. 2009a; Pagaling et al. 2012; Song et al. 2010; Song et al. 2013; Valverde et al. 2012; Wang et al. 2013; Yim et al. 2006), Uttaranchal Himalaya, India (Kumar et al. 2004), Gedongsongo and Kamojang, Indonesia (Aditiawati et al. 2009; Aminin et al. 2008), Sungai Klah and Ulu Slim, Malaysia (Drammeh 2010), Philippines (Huang et al. 2013), Bor Khlueng, Thailand (Kanokratana et al. 2004), White Island, New Zealand (Donachie et al. 2002), Kamchatka, Russia (Bonch-Osmolovskaya et al. 1999; Burgess et al. 2012; Reigstad et al. 2009), Iceland (Kvist et al. 2007; Marteinsson et al. 2001; Mirete et al. 2011; Skirnisdottir et al. 2000; Takacs et al. 2001), Vulcano Island, Italy (Rogers and Amend

2005), Romania (Coman et al. 2013), Azores, Portugal (Hamamura et al. 2013), Tunisia (Sayeh et al. 2010), Limpopo, South Africa (Tekere et al. 2011), and Algerian (Amarouche-Yala et al. 2014). In conjunction with molecular techniques, these studies not only enable us to see more diverse microbial communities in extreme environments (Barns et al. 1994; Bryant et al. 2007; Hugenholtz et al. 1998), but also provide promising opportunities to gain insights into primitive life forms (Barns et al. 1996; Brochier-Armanet et al. 2012; Elkins et al. 2008; Kozubal et al. 2013).

Additionally, microorganisms living in extreme environments are also important sources for enzymes with unusual properties and desirable applications. For example, biotechnological applications of thermophiles ($T_{opt} \geq 50$ ºC) and hyperthermophiles ($T_{opt} \geq 65$ ºC) (Mesbah and Wiegel 2008) are drawing increasing attention from microbiologists because of their thermostabilization. A notable commercial application is thermostable DNA polymerases that were originally isolated from the YNP hyperthermal isolate, *Thermus aquaticus*. Extracellular enzymes including amylases, proteases, xylanases and pullulanases isolated from thermophiles and hyperthermophiles indicate highly thermal stability, making them attractive for broad biotechnological applications in detergent, paper bleaching, baking, brewing and biotechnology industries as well as bioremediation (Fujiwara 2002; Huber and Stetter 1998). Given these applications, we can anticipate that (hyper) thermophilic archaea and bacteria will play more important roles in future biotechnological and industrial processes.

Yellowstone National Park is one of the largest and most diverse hydrothermal areas in the world and it provides the most accessible and pristine geothermal fields for microbial ecological studies. Yellowstone National Park harbors more than 10,000

geothermal features covering a broad range of environmental gradients such as temperature (40-92 ℃) and pH (1-10) (Fournier 1989; Fournier 2005; Rye and Truesdell 2007). Previous studies (Barns et al. 1994; Boyd et al. 2007; Hall et al. 2008; Hugenholtz et al. 1998; Inskeep et al. 2013a; Inskeep et al. 2013b; Inskeep et al. 2010; Klatt et al. 2013; Kozubal et al. 2012; Kozubal et al. 2013; Macur et al. 2013; Meyer-Dombard et al. 2005; Mitchell 2009; Reysenbach et al. 2000b; Takacs-Vesbach et al. 2013; Ward et al. 1998a) have shown that YNP thermal springs harbor diverse chemotrophic and phototrophic thermophilic microbial communities, reflecting a wide range of geochemistries. While temperature has long been known to control microbial diversity, previous attempts to classify and measure the biodiversity across temperature gradients were based on either traditional molecular methods (Nakagawa and Fukui 2002; Sievert et al. 2000) or certain groups of bacteria (e.g. Cyanobacteria and Chloroflexi) (Miller et al. 2009). Thus, understanding the influence of temperature variation on the structures of microbial communities and interaction among different microbial communities is impeded by lack of available complete data generated from new techniques such as pyrosequencing.

In addition to temperature, pH is also a crucial factor in determining microbial assemblages in thermal springs. The bimodal distribution of pH values leads to the majority of the research on YNP thermophilic microbial communities focusing on two types of widely distributed springs (Brock et al. 1972; Jackson et al. 2001; Madigan 2003; Papke et al. 2003; Reysenbach et al. 2000b; Reysenbach et al. 1994; Shock et al. 2005; Ward et al. 1998a), which are vapor-dominated acidic springs and water-dominated circumneutral to slightly alkaline springs (Fournier 1989). Yet the intermediate pH (e.g.,

pH 4-5) habitats have been overlooked because of only few sites within a range of the intermediate pH. The lack of information on microbial communities in intermediate springs has still hindered an understanding of the whole picture of microbial diversity in YNP thermal springs.

Last, local environmental heterogeneity may also be responsible for the assemblages and distribution of some microbial populations. For example, *Synechococcus* spp., a member of thermophilic Cyanobacteria, from a single YNP thermal spring mat, has highly diverse ecotypes that adapt to different temperatures and photic micro-gradients (Ramsing et al. 2000; Ward et al. 1998a; Ward et al. 1990). Micro-gradients in light intensity, oxygen, and sulfide concentration also affect the distribution of some green non-sulfur bacteria such as *Chloroflexus* and *Roseiflexus* (van der Meer et al. 2005). Further studies on other microbial taxa are required to investigate the environmental heterogeneity controlling over microbial community assembly at the local scale and to unravel the forces that govern microbial diversity.

My dissertation research has concentrated on the study of microbial communities in YNP thermal springs, with the goal of determining how temperature, pH and environmental heterogeneity influence their structure, diversity and potential function *in situ*. My work addressed the following questions: i) In Chapter 2, how do temperature gradients affect bacterial community structure? ii) In Chapter 3, what are dominant microbial groups in pH 4 sites and what are their functional roles? iii) In Chapter 4, how does environmental heterogeneity impact on taxonomic and functional diversity in the same type of springs (neutral to slightly alkaline springs)? I summarize my important findings and suggest future research in Chapter 5.

## References


Aditiawati, P., Yohandini, H., Madayanti, F., and Akhmaloka (2009) Microbial diversity of acidic hot spring (kawah hujan B) in geothermal field of kamojang area, west java-indonesia. *Open Microbiol J* **3**: 58-66.

Amarouche-Yala, S., Benouadah, A., El Ouahab Bentabet, A., and López-García, P. (2014) Morphological and phylogenetic diversity of thermophilic cyanobacteria in Algerian hot springs. *Extremophiles* **18**: 1035-1047.

Aminin, A.L.N., Warganegara, F.M., Aditiawati, P., and Akhmaloka (2008) Simple enrichment and independent cultures to expand bacterial community analysis from gedongsongo hot spring. *Journal of Bioscience and Bioengineering* **106**: 211-214.

Barns, S.M., Fundyga, R.E., Jeffries, M.W., and Pace, N.R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 1609-1613.

Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences* **93**: 9188-9193.

Bonch-Osmolovskaya, E.A., Miroshnichenko, M.L., Slobodkin, A.I., Sokolova, T.G., Karpov, G.A., Kostrikina, N.A. et al. (1999) Biodiversity of anaerobic lithotrophic prokaryotes in terrestrial hot springs of Kamchatka. *Microbiology* **68**: 343-351.

Boyd, E.S., Jackson, R.A., Encarnacion, G., Zahn, J.A., Beard, T., Leavitt, W.D. et al. (2007) Isolation, characterization, and ecology of sulfur-respiring Crenarchaea inhabiting acid-sulfate-chloride-containing geothermal springs in Yellowstone National Park. *Applied and Environmental Microbiology* **73**: 6669-6677.

Brochier-Armanet, C., Gribaldo, S., and Forterre, P. (2012) Spotlight on the Thaumarchaeota. *ISME J* **6**: 227-230.

Brock, T., Brock, K., Belly, R., and Weiss, R. (1972) *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Archives of Microbiology* **84**: 54-68.

Brock, T.D., and Brock, T.D. (1978) *Thermophilic microorganisms and life at high temperatures*. New York: Springer-Verlag

Bryant, D.A., Costas, A.M.G., Maresca, J.A., Chew, A.G.M., Klatt, C.G., Bateson, M.M. et al. (2007) *Candidatus* Chloracidobacterium thermophilum: An Aerobic Phototrophic Acidobacterium. *Science* **317**: 523-526.

Burgess, E., Unrine, J., Mills, G., Romanek, C., and Wiegel, J. (2012) Comparative geochemical and microbiological characterization of two thermal pools in the Uzon caldera, Kamchatka, Russia. *Microbial Ecology* **63**: 471-489.

Cavicchioli, R. (2002) Extremophiles and the search for extraterrestrial life. *Astrobiology* **2**: 281-292.

Cole, J.K., Peacock, J.P., Dodsworth, J.A., Williams, A.J., Thompson, D.B., Dong, H. et al. (2013) Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J* **7**: 718-729.

Coman, C., Drugă, B., Hegedus, A., Sicora, C., and Dragoş, N. (2013) Archaeal and bacterial diversity in two hot spring microbial mats from a geothermal region in Romania. *Extremophiles* **17**: 523-534.

Davis, B.M. (1897) The vegetation of the hot springs of Yellowstone Park. *Science* **6**: 145.

De León, K.B., Gerlach, R., Peyton, B.M., and Fields, M.W. (2013) Archaeal and bacterial communities in three alkaline hot springs in Heart Lake Geyser Basin, Yellowstone National Park. *Frontiers in Microbiology* **4**.

Donachie, S., Christenson, B., Kunkel, D., Malahoff, A., and Alam, M. (2002) Microbial community in acidic hydrothermal waters of volcanically active White Island, New Zealand. *Extremophiles* **6**: 419-425.

Drammeh, I. (2010) Biodiversity of thermophiles in Sungai Klah and Ulu Slim hot springs. In *Faculty of Biosciences and Bioengineering* Universiti Teknologi Malaysia

Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L. et al. (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 8102-8107.

Engel, A.S., Johnson, L.R., and Porter, M.L. (2013) Arsenite oxidase gene diversity among Chloroflexi and Proteobacteria from El Tatio Geyser Field, Chile. *FEMS Microbiology Ecology* **83**: 745-756.

Everroad, R.C., Otaki, H., Matsuura, K., and Haruta, S. (2012) Diversification of bacterial community composition along a temperature gradient at a thermal spring. *Microbes and Environments* **27**: 374-381.

Fournier, R.O. (1989) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. *Annual Review of Earth and Planetary Sciences* **17**: 13-53.

Fournier, R.O. (2005) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. In *Geothermal biology and geochemistry in Yellowstone National Park.* Inskeep, W.P., and TR, M. (eds). Bozeman, MT: Thermal Biology Institute, Montana State University, pp. 4-30.

Fujiwara, S. (2002) Extremophiles: developments of their special functions and potential resources. *Journal of Bioscience and Bioengineering* **94**: 518-525.

Hall, J.R., Mitchell, K.R., Jackson-Weaver, O., Kooser, A.S., Cron, B.R., Crossey, L.J., and Takacs-Vesbach, C.D. (2008) Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes. *Appl Environ Microbiol* **74**: 4910-4922.

Hamamura, N., Meneghin, J., and Reysenbach, A.L. (2013) Comparative community gene expression analysis of Aquificales-dominated geothermal springs. *Environ Microbiol* **15**: 1226-1237.

Huang, Q., Jiang, H., Briggs, B.R., Wang, S., Hou, W., Li, G. et al. (2013) Archaeal and bacterial diversity in acidic to circumneutral hot springs in the Philippines. *FEMS Microbiology Ecology* **85**: n/a-n/a.

Huang, Q., Dong, C., Dong, R., Jiang, H., Wang, S., Wang, G. et al. (2011) Archaeal and bacterial diversity in hot springs on the Tibetan Plateau, China. *Extremophiles* **15**: 549-563.

Huber, H., and Stetter, K.O. (1998) Hyperthermophiles and their possible potential in biotechnology. *Journal of Biotechnology* **64**: 39-52.

Hugenholtz, P., Pitulle, C., Hershberger, K.L., and Pace, N.R. (1998) Novel division level bacterial diversity in a Yellowstone hot spring. *Journal of Bacteriology* **180**: 366-376.

Inskeep, W.P., Jay, Z.J., Tringe, S.G., Herrgard, M.J., and Rusch, D.B. (2013a) The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Front Microbiol* **4**: 67.

Inskeep, W.P., Jay, Z.J., Herrgard, M.J., Kozubal, M.A., Rusch, D.B., Tringe, S.G. et al. (2013b) Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate Linkages between metabolic potential and geochemistry. *Front Microbiol* **4**: 95.

Inskeep, W.P., Rusch, D.B., Jay, Z.J., Herrgard, M.J., Kozubal, M.A., Richardson, T.H. et al. (2010) Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS ONE* **5**: e9773.

Jackson, C.R., Langner, H.W., Donahoe-Christiansen, J., Inskeep, W.P., and McDermott, T.R. (2001) Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. *Environmental Microbiology* **3**: 532-542.

Jiang, H., Dong, C.Z., Huang, Q., Wang, G., Fang, B., Zhang, C., and Dong, H. (2012) Actinobacterial diversity in microbial mats of five hot springs in central and central-eastern Tibet, China. *Geomicrobiology Journal* **29**: 520-527.

Kanokratana, P., Chanapan, S., Pootanakit, K., and Eurwilaichitr, L. (2004) Diversity and abundance of Bacteria and Archaea in the Bor Khlueng Hot Spring in Thailand. *Journal of Basic Microbiology* **44**: 430-444.

Klatt, C.G., Inskeep, W.P., Herrgard, M.J., Jay, Z.J., Rusch, D.B., Tringe, S.G. et al. (2013) Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front Microbiol* **4**: 106.

Kozubal, M.A., Macur, R.E., Jay, Z.J., Beam, J.P., Malfatti, S.A., Tringe, S.G. et al. (2012) Microbial iron cycling in acidic geothermal springs of yellowstone national park:

integrating molecular surveys, geochemical processes, and isolation of novel fe-active microorganisms. *Front Microbiol* **3**: 109.

Kozubal, M.A., Romine, M., Jennings, R.d., Jay, Z.J., Tringe, S.G., Rusch, D.B. et al. (2013) Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J* **7**: 622-634.

Kumar, B., Trivedi, P., Kumar Mishra, A., Pandey, A., and Palni, L.M.S. (2004) Microbial diversity of soil from two hot springs in Uttaranchal Himalaya. *Microbiological Research* **159**: 141-146.

Kvist, T., Ahring, B.K., and Westermann, P. (2007) Archaeal diversity in Icelandic hot springs. *FEMS Microbiology Ecology* **59**: 71-80.

Lau, C.-y. (2007) Ecology of natural thermophilic communities in the Tibet autonomous region (China). In *Ecology and Biodiversity*. Hong Kong: University of Hong Kong.

Lau, M., Aitchison, J., and Pointing, S. (2009) Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles* **13**: 139-149.

Macur, R.E., Jay, Z.J., Taylor, W.P., Kozubal, M.A., Kocar, B.D., and Inskeep, W.P. (2013) Microbial community structure and sulfur biogeochemistry in mildly-acidic sulfidic geothermal springs in Yellowstone National Park. *Geobiology* **11**: 86-99.

Madigan, M.T. (2003) Anoxygenic phototrophic bacteria from extreme environments. *Photosynthesis research* **76**: 157-171.

Marteinsson, V.T., Hauksdóttir, S., Hobel, C.F.V., Kristmannsdóttir, H., Hreggvidsson, G.O., and Kristjánsson, J.K. (2001) Phylogenetic diversity analysis of subterranean hot springs in Iceland. *Applied and Environmental Microbiology* **67**: 4242-4248.

Mesbah, N., and Wiegel, J. (2008) Life at extreme limits. In *Incredible anaerobes: from physiology to genomics to fuels*, pp. 44-57.

Meyer-Dombard, D.R., Shock, E.L., and Amend, J.P. (2005) Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. *Geobiology* **3**: 211-227.

Miller, S.R., Strong, A.L., Jones, K.L., and Ungerer, M.C. (2009) Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Applied and Environmental Microbiology* **75**: 4565-4572.

Mirete, S., de Figueras, C.G., and Gonz áez-Pastor, J.E. (2011) Diversity of Archaea in Icelandic hot springs based on 16S rRNA and chaperonin genes. *FEMS Microbiology Ecology* **77**: 165-175.

Mitchell, K.R. (2009) Controls on microbial community structure in thermal environments : exploring bacterial diversity and the relative influence of geochemistry and geography. In *Biology*: University of New Mexico.

Nakagawa, T., and Fukui, M. (2002) Phylogenetic characterization of microbial mats and streamers from a Japanese alkaline hot spring with a thermal gradient. *The Journal of general and applied microbiology* **48**: 211-222.

Otaki, H., Everroad, R.C., Matsuura, K., and Haruta, S. (2012) Production and consumption of hydrogen in hot spring microbial mats dominated by a filamentous anoxygenic photosynthetic bacterium. *Microbes Environ* **27**: 293-299.

Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276**: 734-740.

Pagaling, E., Grant, W., Cowan, D., Jones, B., Ma, Y., Ventosa, A., and Heaphy, S. (2012) Bacterial and archaeal diversity in two hot spring microbial mats from the geothermal region of Tengchong, China. *Extremophiles* **16**: 607-618.

Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. (2003) Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology* **5**: 650-659.

Ramsing, N.B., Ferris, M.J., and Ward, D.M. (2000) Highly ordered vertical structure of *Synechococcus* populations within the one-millimeter-thick photic zone of a hot spring Cyanobacterial mat. *Applied and Environmental Microbiology* **66**: 1038-1049.

Rappé, M.S., and Giovannoni, S.J. (2003) The unclutured microbial majority. *Annual Review of Microbiology* **57**: 369-394.

Reigstad, L.J., Jorgensen, S.L., and Schleper, C. (2009) Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka. *ISME J* **4**: 346-356.

Reysenbach, A.L., Wickham, G.S., and Pace, N.R. (1994) Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Applied and Environmental Microbiology* **60**: 2113-2119.

Reysenbach, A.L., Ehringer, H., and Hershberger, K. (2000) Microbial diversity at 83°C in Calcite Springs, Yellowstone National Park: another environment where the *Aquificales* and "Korarchaeota" coexist. *Extremophiles* **4**: 61-67.

Rogers, K.L., and Amend, J.P. (2005) Archaeal diversity and geochemical energy yields in a geothermal well on Vulcano Island, Italy. *Geobiology* **3**: 319-332.

Rothschild, L.J., and Mancinelli, R.L. (2001) Life in extreme environments. *Nature* **409**: 1092-1101.

Rye, R.O., and Truesdell, A.H. (2007) The question of recharge to the deep thermal reservoir underlying the geysers and hot springs of Yellowstone National Park. In *Integrated Geoscience Studies in the Greater Yellowstone Area: Volcanic, Tectonic, and Hydrothermal Processes in the YellowstoneGeoecosystem*. al., L.A.M.e. (ed): U.S. Geol. Surv. Prof. Pap., pp. 1717, 1235-1270.

Sayeh, R., Birrien, J., Alain, K., Barbier, G., Hamdi, M., and Prieur, D. (2010) Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches. *Extremophiles* **14**: 501-514.

Schoenborn, L., Yates, P.S., Grinton, B.E., Hugenholtz, P., and Janssen, P.H. (2004) Liquid serial dilution is inferior to solid media for isolation of cultures representative of the phylum-level diversity of soil bacteria. *Appl Environ Microbiol* **70**: 4363-4366.

Shock, E.L., Holland, M., Meyer-Dombard, D.A.R., and Amend, J.P. (2005) Geochemical sources of energy for microbial metabolism in hydrothermal ecosystems: Obsidian Pool, Yellowstone National Park. In *Geothermal Biology and Geochemistry in Yellowstone National Park*. Inskeep, W.P., and McDermott, T.R. (eds). Bozeman, MT: Montana State University, pp. 95-109

Siering, P.L., Wolfe, G.V., Wilson, M.S., Yip, A.N., Carey, C.M., Wardman, C.D. et al. (2013) Microbial biogeochemistry of Boiling Springs Lake: a physically dynamic, oligotrophic, low-pH geothermal ecosystem. *Geobiology*.

Sievert, S.M., Ziebis, W., Kuever, J., and Sahm, K. (2000) Relative abundance of Archaea and Bacteria along a thermal gradient of a shallow-water hydrothermal vent quantified by rRNA slot-blot hybridization. *Microbiology* **146**: 1287-1293.

Skirnisdottir, S., Hreggvidsson, G.O., Hjörleifsdottir, S., Marteinsson, V.T., Petursdottir, S.K., Holst, O., and Kristjansson, J.K. (2000) Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Applied and Environmental Microbiology* **66**: 2835-2841.

Song, Z.-Q., Chen, J.-Q., Jiang, H.-C., Zhou, E.-M., Tang, S.-K., Zhi, X.-Y. et al. (2010) Diversity of Crenarchaeota in terrestrial hot springs in Tengchong, China. *Extremophiles* **14**: 287-296.

Song, Z.-Q., Wang, F.-P., Zhi, X.-Y., Chen, J.-Q., Zhou, E.-M., Liang, F. et al. (2013) Bacterial and archaeal diversities in Yunnan and Tibetan hot springs, China. *Environmental Microbiology* **15**: 1160-1175.

Swingley, W.D., Meyer-Dombard, D.A.R., Shock, E.L., Alsop, E.B., Falenski, H.D., Havig, J.R., and Raymond, J. (2012) Coordinating environmental genomics and geochemistry reveals metabolic transitions in a hot spring ecosystem. *PLoS ONE* **7**: e38108.

Takacs-Vesbach, C., Inskeep, W.P., Jay, Z.J., Herrgard, M.J., Rusch, D.B., Tringe, S.G. et al. (2013) Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three aquificales lineages. *Front Microbiol* **4**: 84.

Takacs, C.D., Ehringer, M., Favre, R., Cermola, M., Eggertsson, G., Palsdottir, A., and Reysenbach, A.-L. (2001) Phylogenetic characterization of the blue filamentous bacterial community from an Icelandic geothermal spring. *FEMS Microbiology Ecology* **35**: 123-128.

Tekere, M., Lotter, A., Olivier, J., Jonker, N., and Venter, S.N. (2011) Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa.

Valverde, A., Tuffin, M., and Cowan, D. (2012) Biogeography of bacterial communities in hot springs: a focus on the actinobacteria. *Extremophiles* **16**: 669-679.

van der Meer, M.T.J., Schouten, S., Bateson, M.M., Nübel, U., Wieland, A., Kühl, M. et al. (2005) Diel variations in carbon metabolism by green nonsulfur-like bacteria in alkaline siliceous hot spring microbial mats from Yellowstone National Park. *Applied and Environmental Microbiology* **71**: 3978-3986.

Wang, S., Hou, W., Dong, H., Jiang, H., Huang, L., Wu, G. et al. (2013) Control of temperature on microbial community structure in hot springs of the Tibetan Plateau. *PLoS ONE* **8**: e62901.

Ward, D.M., Weller, R., and Bateson, M.M. (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**: 63-65.

Ward, D.M., Ferris, M.J., Nold, S.C., and Bateson, M.M. (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiology and Molecular Biology Reviews* **62**: 1353-1370.

Wilson, M., Siering, P., White, C., Hauser, M., and Bartles, A. (2008) Novel Archaea and Bacteria Dominate Stable Microbial Communities in North America's Largest Hot Spring. *Microbial Ecology* **56**: 292-305.

Yim, L.C., Hongmei, J., Aitchison, J.C., and Pointing, S.B. (2006) Highly diverse community structure in a remote central Tibetan geothermal spring does not display monotonic variation to thermal stress. *FEMS Microbiology Ecology* **57**: 80-91.

## Chapter 2

# Influence of Temperature on Bacterial Community Structure in Three Alkaline Thermal Springs of Yellowstone National Park

**XIAOBEN JIANG[1†], DANIEL R. COLMAN[1], JORDAN G. OKIE[2] and CRISTINA TAKACS-VESBACH[1*]**

[1]Department of Biology, University of New Mexico, Albuquerque, New Mexico, 87131, United States of America

[2]School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287, United States of America

**Abstract**

One of the fundamental goals in ecology is to elucidate how environmental variables control biodiversity. In this study, we used 16S rRNA gene pyrosequencing to investigate the influence of environmental variables on bacterial community structure along the thermal gradients in three alkaline thermal springs of Yellowstone National Park. Our results indicated that temperature was the most important environmental variable to shape the structure of bacterial communities. Community dissimilarity between samples increased as temperature differences increased. UPGMA cluster analysis showed four distinct groups and the succession of predominant bacterial phyla in each group was systematically correlated with temperature. At high temperatures (81-87 ℃), *Aquificae* and EM3 predominated, whereas the temperature range of photosynthesis upper limit (73-75 ℃) was dominated by *Deinococcus-Thermus* and *Armatimonadetes*. Intermediate temperatures (63-68 ℃) and low temperatures (40-57 ℃) both favored *Cyanobacteria* and *Chloroflexi*, but bacterial communities at low temperatures were more diverse. In addition, Chao1 richness and phylogenetic diversity in phototrophic communities were significantly higher (Chao1, P=0.014 and phylogenetic diversity, P=0.009, respectively) than in non-phototrophic communities. We observed nonrandom distribution patterns of the *Cyanobacteria* and *Chloroflexi* populations. This pattern may be shaped by temperature and species interactions (*e.g.*, cross-feeding, competition for limited resources). The results of this study expand our current knowledge of the role of temperature in controlling community structure in YNP alkaline thermal springs and offer a basis for understanding microbial ecology in global thermal springs.

Microbes can colonize myriads of hostile environments on earth, such as terrestrial thermal springs (Barns et al. 1996; Hugenholtz et al. 1998; Inskeep et al. 2013b), hot deserts (Robinson et al. 2014), Antarctic cold deserts (Schwartz et al. 2014; Van Horn et al. 2014; Van Horn et al. 2013), hypersaline ponds (Ley et al. 2006), and even possibly other celestial bodies (Cavicchioli 2002; Rothschild and Mancinelli 2001). Diversification in microbial communities is determined by a variety of environmental variables, such as temperature, pH and salinity (Lozupone and Knight 2007). Depending on the environments and geographical scales, the extent that each environmental variable contributes to the diversification of microbial communities can vary with different habitats and geographical scales (Acosta-Mart ńez et al. 2008; Dequiedt et al. 2009b; Everroad et al. 2012; Lauber et al. 2009; Lozupone and Knight 2007; Nacke et al. 2011; Segawa et al. 2010; Signori et al. 2014; Van Horn et al. 2013; Wang et al. 2013).

Yellowstone National Park (YNP) is one of the largest and most diverse hydrothermal areas on Earth. Extensive culture-independent studies (Barns et al. 1994; Hall et al. 2008; Hugenholtz et al. 1998; Inskeep et al. 2013a; Inskeep et al. 2013b; Inskeep et al. 2010; Klatt et al. 2013; Kozubal et al. 2012; Kozubal et al. 2013; Macur et al. 2013; Meyer-Dombard et al. 2005; Mitchell 2009; Reysenbach et al. 2000b; Takacs-Vesbach et al. 2013; Ward et al. 1998a) have shown that YNP thermal springs harbor diverse chemotrophic and phototrophic microbial communities. Temperature and pH have the significant influence on a wide variety of physicochemical properties of water and on the kinetics and stability of biomolecules, resulting in the variation of microbial community assemblages. Previous studies have demonstrated that pH is a primary driver of microbial community composition in thermal springs among geographical regions or at the global scale (Boyd et al. 2013;

Boyd et al. 2010; Dequiedt et al. 2009a; Inskeep et al. 2013b; Song et al. 2013; Xie et al. 2014). In contrast, the influence of temperature on the structures of microbial communities has been normally observed in springs within a similar pH range or within an individual geographical region (Everroad et al. 2012; Wang et al. 2013; Xie et al. 2014). For example, studies on YNP alkaline springs have previously revealed that chemolithoautotrophic bacteria such as *Aquificae* are dominant primary producers at high temperature (>75 ℃) (Graber et al. 2001; Huber and Stetter 2001; Jahnke et al. 2001; Reysenbach et al. 2005). As temperature decreases below 75 ℃, thermophilic phototrophs such as *Cyanobacteria* and *Chloroflexi* become predominant at the moderate temperatures (*e.g.*, 60–75 ℃) (Brock and Brock 1978; Cox et al. 2011). Heterotrophic bacteria become important when temperature decreases below 60 ℃ (Zhang et al. 2004).

However, previous attempts to study the influence of environmental variables on microbial diversity in YNP thermal springs were either based on pilot surveys at large geographical scales (*i.e.*, sampling one or two sites from different springs) or focused on a narrow temperature range within one or two springs. Not many investigations have been performed for the same type of YNP thermal springs along a broad temperature range, and the microbial diversity and interactions along temperature gradients are still not well understood. The objectives of the present study were 1) to expand our current knowledge of the influence of environmental variables on bacterial community diversity and assemblages along a broad temperature range, and 2) to compare bacterial community composition within the same spring and among different springs. To accomplish these two objectives, samples from three thermal springs and a broad temperature range (40-87 ℃) were examined by using high throughput 454 pyrosequencing.

## Materials and Methods

*Site description and sample collection*

Microbial filament and mat samples were collected aseptically along the outflow channels of three alkaline thermal springs located in the Lower Geyser Basin of YNP: Red Terrace (RT, 44°33'54"N, -110°51'36"W), Bison Pool (BP, 44°34'11"N, -110°51'54"W) and Octopus Spring (OS, 44°32'03"N, -110°47'53"W) (Fig. S1). The sampling sites spanned distances of approximately 31 m at RT, 27 m at BP and 51 m at OS (Table 1). At each sampling location, water temperature and pH were measured using a Thermo Orion 290A+ meter and electrical conductivity was measured with a WTW meter with temperature correction. Hydrogen sulfide was measured in the field using the methylene blue method (Hach method 8131) and a hand-held colorimeter (Hach DR/850). After field measurements, biomass was collected into 2-mL microcentrifuge tubes using either a sterile syringe or sterilized forceps and preserved in sucrose lysis buffer (Giovannoni et al. 1990). The samples were stored on ice upon return from the field and stored in the laboratory at -80 ℃ until DNA extraction. Spring water was collected and analyzed as described previously (McCleskey 2005). Total dissolved ions were determined using inductively coupled plasma (ICP) spectrometry and ion chromatography in the laboratory. The geochemical data of BP and OS were obtained from published data by YNP Research Coordination Network (http://www.rcn.montana.edu) listed in Table S1. The geochemistries of YNP thermal springs are stable and lack annual and seasonal variations (Ball et al. 2010; Ball et al. 2002; Ball et al. 1998).

*DNA extraction and 454 pyrosequencing*

Total DNA was extracted from 200 μL of preserved samples following bead-beating disruption in a CTAB buffer and subsequent phenol-chloroform purification

steps as described previously (Mitchell and Takacs-Vesbach 2008). Briefly, 2 volumes of 1% CTAB buffer and proteinase K (final concentration 100 μg $mL^{-1}$) were added to the samples, which were then incubated for one hour at 60 ℃. SDS (final concentration 2%) and 0.1 mm diameter Zirconia/Silica beads (BioSpec products) were added. Samples were bead beaten for 45 s at 50 strokes per second. After incubating for one hour at 60 ℃, genomic DNA was extracted once with phenol-chloroform-isoamyl alcohol (25:24:1), and twice with chloroform. The DNA was then precipitated with 95% ethanol after the addition of 0.1 volume 3 M sodium acetate. Finally, the DNA was washed once with 70% ethanol, air-dried and resuspended in 50 μL 10mM Tris, pH 8.0.

Barcoded amplicon pyrosequencing of 16S rRNA genes was performed as described previously (Dowd et al. 2008a; Van Horn et al. 2013). The V1-V3 variable region of bacterial 16S rRNA gene was amplified with the universal bacterial primers 28F (5‘-GAGTTTGATCNTGGCTCAG-3’) and 519R (5‘-GTNTTACNGCGGCKGCTG-3’). PCR conditions for amplification were as follows: initial cycle of 95 ℃ for 5 min, 30 cycles of 95 ℃ for 30 s, 54 ℃ for 40 s and 72 ℃ for 1 min, and a final elongation for 10 min at 72 ℃. Amplification products were confirmed by agarose gel electrophoresis. Triplicate reaction mixtures per sample were combined and subsequently purified with an UltraClean® GelSpin® DNA Extraction Kit (MoBio Laboratories, Carlsbad, CA, USA). The purified DNA was quantified using a Nanodrop ND-2000c spectrophotometer and equimolar concentrations were pooled for 454 pyrosequencing using the GS FLX Titanium platform (454 Life Sciences, Branford, CT, USA), as previously described (Dowd et al. 2008a; Van Horn et al. 2013).

*Pyrosequencing data processing and analysis*

Raw 16S rRNA gene sequences were quality filtered, denoised and checked for chimeras using AmpliconNoise (Quince et al. 2011) integrated into QIIME (Ver. 1.7.0) (Caporaso et al. 2010b). Adapters, multiplex identifiers and primers were trimmed from denoised data. Denoised sequences were clustered into operational taxonomic units (OTUs) at the 97% DNA sequence similarity level using UCLUST (Edgar 2010) in QIIME. The most abundant sequence was picked from each OTU as a representative sequence (pick_rep_set.py) and aligned using the PyNAST aligner (Caporaso et al. 2010a) and Greengenes core set (GG 13_5) database (DeSantis et al. 2006) in QIIME. The aligned sequences were given taxonomic assignment using the Ribosomal Database Classifier program (Wang et al. 2007). To standardize for varying sequencing efforts across samples, all measurements of bacterial community structure were performed with randomly drawn subsets of 750 sequences from each sample.

All statistical analyses were performed using either QIIME or the Vegan package (Oksanen et al.) in R (Team 2011). Alpha diversity (Chao1 richness, Good's coverage, observed number of OTUs, Faith's phylogenetic diversity (Faith 1992), the Shannon and Simpson index) and beta diversity (Bray-Curtis dissimilarity) were calculated at the 97% DNA sequence similarity level. Linear regression was performed between the observed OTUs and temperature in each site. The unweighted pair group method with arithmetic mean (UPGMA) was used for sample clustering (based on Bray-Curtis dissimilarity) and jackknife confidence analysis. ANOSIM (Clarke 1993), Adonis (Anderson 2001), and MRPP (Mielke et al. 1981), which perform nonparametric multivariate statistical tests using distance matrices, were used to test the significance of cluster groups. Non-metric multidimensional scaling (NMDS) analysis was performed on Bray-Curtis dissimilarity matrix. The metaMDS

and envfit functions in ‘vegan’ package were used to test environmental variables correlated with bacterial community composition by fitting the four environmental variables (*i.e.*, temperature, pH, conductivity and hydrogen sulfide) measured in the field to the ordination. The BIO-ENV procedure (Clarke and Ainsworth 1993) was used to reveal any correlation between community data and major environmental variables. A Mantel test was also performed to check for any correlation between the community dissimilarity and the differences in temperature between samples. SIMPER (similarity percentage) analysis (Clarke 1993) was used to rank the top ten genera that accounted for the difference between any two adjacent temperature clusters from the results of UPGMA clustering.

*Null model analysis on phototrophic communities*

To better understand co-occurrence patterns between *Cyanobacteria* and *Chloroflexi*, we employed checkerboard score (C score), a presence/absence-based measurement of the co-occurrence of taxa, under a null model (Stone and Roberts 1990) using software EcoSim (Entsminger 2012) with the default settings. A standardized effect size (SES) for each matrix was calculated, which measures the difference between the observed C score and the mean of the simulated null model normalized by the standard deviation of the null distribution. Briefly, SES is calculated as the number of standard deviations. That is, the observed score is above or below the mean co-occurrence index for the simulated communities [$SES = (I_{obs} - I_{sim})/S_{sim}$]. Where $I_{obs}$ is the C score of the observed incidence matrix, $I_{sim}$ is the mean of 5,000 C scores generated from the simulated null model matrices, and $S_{sim}$ is the standard deviation of 5,000 simulated communities (Horner-Devine et al. 2007). If the observed C score is significantly higher than the C score generated from the simulated null distribution, this indicates a higher degree of nonrandom structure (*i.e.*, there is

segregation among species). By contrast, if the C score is significantly lower than the C score generated from the simulated null distribution, this indicates that a high degree of random structure exists (*i.e.*, there is aggregation among species).

*Nucleotide sequence accession numbers*

All raw sequencing data from this research are available through the NCBI Sequence Read Archive as SRP049962. The individual sff files from this study were assigned the accession numbers SRX761204, SRX761207, SRX761248 - SRX761254 and SRX761314 - SRX761319 under Bioproject PRJNA267597.

**Results**

*Geochemical characterization of sample locations*

The ranges of temperature and pH for each spring were from 54 ℃ to 84 ℃ and 8.9 to 9.1, 40 ℃ to 81 ℃ and 7.5 to 8.6, and 47 ℃ to 87 ℃ and 7.6 to 8.3 (RT, BP, OS, respectively) (Table 1). Other field measurements such as conductivity and hydrogen sulfide were also listed in Table 1. The major ion geochemistries of RT, BP and OS were compared in the Piper diagram (Fig. S2), indicating similarity in these geochemical variables in the three springs.

*Overview of 454 pyrosequencing data and bacterial diversity*

A total of 53,872 16S rRNA gene sequences (mean = 3,591 16S rRNA sequences/site) were obtained from 15 sites after denoising and removal of low quality sequences, spurious sequences and chimeras in AmpliconNoise. DNA sequences clustered into 374 OTUs at the level of 97% DNA sequence similarity. Good's coverages (Good 1953) of all samples were more than 90% (Table 2), indicating that a majority of OTU richness was detected in all samples. Bacterial 16S rRNA gene sequences retrieved from the three thermal springs represented the following phyla and candidate divisions: *Acidobacteria*, *Actinobacteria*, *Aquificae*,

*Armatimonadetes*, *Bacteroidetes*, *Chlorobi*, *Chloroflexi*, *Cyanobacteria*, *Deinococcus-Thermus*, *Elusimicrobia*, EM19, EM3, *Firmicutes*, *Nitrospirae*, OD1, OP1, OP9, OP11, *Planctomycetes*, *Proteobacteria*, *Spirochaetes*, *Thermodesulfobacteria*, *Thermotogae*, *Unknown Bacteria* (Fig. 1B and S4).

*UPGMA cluster analysis*

The UPGMA cluster analysis indicated that samples clustered into four distinct groups correlated with different temperature ranges (Fig. 1A): high temperature (HT, 81-87 ℃), photosynthesis upper limit (PUL, 73-75 ℃), intermediate temperature (IT, 63-68 ℃) and low temperature (LT, 40-57 ℃). Significant clustering of sample community composition by temperature was confirmed by ANOSIM ($R=0.7699$ and $P=0.001$), Adonis ($R^2=0.54134$ and $P=0.001$) and MRPP ($\delta=0.6234$ and $P=0.001$) tests. Microbial community composition among different springs was not significantly different based on ANOSIM, Adonis and MRPP tests ($R=-0.012$ and $P=0.481$, $R^2=0.14401$ and $P=0.416$, $\delta=0.8625$ and $P=0.495$, respectively) because of similar geochemical environments in these springs (Fig. S2). The group scheme detected in UPGMA cluster analysis was also verified by the NMDS ordination (Fig. 2B). Overall, HT and PUL groups were non-phototrophic communities, while IT and LT groups were phototrophic communities. Alpha diversity in phototrophic communities (<73 ℃) was significantly higher than in non-phototrophic communities (>73 ℃), based on the non-parametric two sample *t* test with 1000 Monte Carlo permutations (observed OTUs, $P=0.02$; Chao1, $P=0.014$; phylogenetic diversity, $P=0.009$; respectively)

Temperature was the strongest correlate among all measured environmental variables in structuring bacterial communities ($r^2=0.8760$, $P<0.001$). pH was also moderately correlated with the NMDS ordination ($r^2=0.4419$, $P=0.03$). Conductivity

and hydrogen sulfide were not significant at the 0.05 alpha level. Although both temperature and pH were significantly correlated with the NMDS ordination, the BIO-ENV analysis did not suggest that additional environmental factors, other than temperature, would improve the correlation between bacterial community structure and environmental factors (Table S2). A significant inverse linear relationship was found between the site temperature and the number of observed OTUs at the 97% OTU level ($R^2$=0.7, P<0.001, Fig. S3A). Additionally, the result of the mantel test showed that the dissimilarity between bacterial communities increased as the temperature differences between samples increased (r=0.6279, p=0.001, Fig. S3B), suggesting the importance of temperature in shaping the structure of microbial communities.

*Bacterial community composition*

Bacterial community composition varied dramatically among different temperature ranges within the same spring, but was relatively consistent at the same temperature range among different springs (Fig. 1B). Distinct bacterial groups dominated at different temperature ranges (Fig. S4). *Aquificae*, *EM3* and *Deinococcus-Thermus* were the predominant phyla in the HT group (52.1%, 24.3% and 14.9%, respectively), while *Deinococcus-Thermus*, *Armatimonadetes* and *Aquificae* were dominant phyla in the PUL group (50.0%, 29.6% and 7.9%, respectively). As temperature decreased below 73 ℃, *Cyanobacteria*, *Chloroflexi* and *Armatimonadetes* were abundant in the IT group (49.5%, 21.9% and 9.8%, respectively), while *Cyanobacteria*, *Chlorobi* and *Chloroflexi* were the predominant phyla within all LT samples (49.9%, 14.9% and 11.0%, respectively). The top ten genera that accounted for the dissimilarity among the temperature groups were identified by the SIMPER test (Table 3). Taxa such as *Aquificae* (unclassified genus

from Family *Aquificaceae* and *Hydrogenothermaceae*), *Thermotogae* (Genus *Fervidobacterium*) and *Thermodesulfobacteria* (Genus *Geothermobacterium*) were only found in the HT and PUL groups (>73 ℃), whereas other taxa such as *Cyanobacteria* (Genus *Gloeobacter*, *Pseudanabaena* and *Synechococcus*) and *Chloroflexi* (Genus *Chloroflexus*) were only present in the IT and LT groups (<73 ℃). Within the phototrophic temperature range of 73-75 ℃, the relative abundance of *Cyanobacteria* was negatively correlated with *Chloroflexi abundance* ($r=-0.9036$, $p<0.001$) (Fig. 3). Null model species co-occurrence analysis confirmed a nonrandom structure between these taxa (*Cyanobacteria*, $P_{obs>exp} = 0.001$; *Chloroflexi*, $P_{obs>exp} = 0.03$; *i.e.*, there was segregation of taxa).

**Discussion**

*The impact of temperature gradients on microbial communities in general*

Environmental variables have been considered to exert primary control on microbial community assemblages (Hanson et al. 2012). The primary objective of this project was to examine the influence of environmental variables on bacterial communities in alkaline thermal springs of YNP along a broad temperature range. Temperature was the strongest correlate to the bacterial community composition of all three springs. Microbial richness and diversity declined with increasing temperature, suggesting that temperature plays a significant role in shaping the distribution of microbial taxa. Similar patterns have been observed in a variety of terrestrial thermal springs on different continents, including Great Boiling Spring (Cole et al. 2013) and YNP springs (Miller et al. 2009) in North America, Nakabusa springs (Everroad et al. 2012) in Asia and Iceland springs (Tobler and Benning 2011) in Europe. Additionally, temperature has been found to play a more important role in determining microbial community diversity than geography, when communities are analyzed at several

hundred meter to kilometer scales (Miller et al. 2009). Our results strongly support this idea. We detected four temperature related groups (HT, PUL, IT and LT) at meter scales, where sites from the same temperature range showed a tendency to cluster together, irrespective of the origin of the thermal spring.

Bacterial community composition was significantly different among different temperature groups. In hotter temperatures (HT and PUL groups), samples contained a high abundance of non-photosynthetic bacteria, such as the phyla *Aquificae*, *Deinococcus-Thermus*, and *Thermodesulfobacteria*. In contrast, phototrophic bacteria such as *Cyanobacteria* and *Chloroflexi* were important in cooler temperature sites (IT and LT groups). The *in situ* temperature of these bacterial phyla is consistent with their distribution temperature from other global thermal spring surveys (Hou et al. 2013; Lau et al. 2009b; Otaki et al. 2011; Purcell et al. 2007; Reysenbach et al. 2005). The transition from non-phototrophic communities to phototrophic communities occurred at 73 ℃, which is in agreement with the previously established transition temperature (73 - 75 ℃) in alkaline thermal springs (Cox et al. 2011).

*Bacterial communities in the High Temperature group*

*Aquificae*, mostly *Aquificales* dominated in the HT group (Fig. S4). The *Aquificales* are widely found to inhabit both marine and terrestrial hydrothermal systems (Ferrera et al. 2007). Our data demonstrated that the *Aquificales* were only present above 66 ℃ (Fig. 1A), which is similar to previous reports on Nakabusa spring (Nakagawa and Fukui 2002) and Tibetan springs (Wang et al. 2013). However, previous study on Nakabusa spring revealed that some members of *Aquificales* (*e.g.*, *Sulfurihydrogenibium* sp.) occurred across all temperatures due to continuous availability of electronic donors for growth. The order of *Aquificales* has two important families, the *Hydrogenothermaceae* and *Aquificaceae*, which are both

widespread in various geothermal features of YNP (Reysenbach et al. 2005). Most of the OTUs associated with *Aquificales* in this study were classified into the family *Aquificaceae* (Table 3), which could explain the discrepancy in the distribution of the *Aquificales*. *Aquificales*, when present, are considered important primary producers in hydrothermal environments (Blank et al. 2002; Burgess et al. 2012; Eder and Huber 2002; Harmsen et al. 1997; Inagaki et al. 2003; Jahnke et al. 2001; Reysenbach et al. 2000b; Reysenbach et al. 1999; Spear et al. 2005; Takacs-Vesbach et al. 2013; Yamamoto et al. 1998; Zhang et al. 2004), because many representatives of the *Aquificales* are chemoautotrophs that can use a variety of electron donors and acceptors, and obtain carbon source from environmental $CO_2$ through the reductive tricarboxylic acid (rTCA) cycle rather than from organic carbon sources (Beh et al. 1993; Ferrera et al. 2007; Hugler et al. 2007; Reysenbach et al. 2009; Shiba et al. 1985; Takacs-Vesbach et al. 2013). The dominance of *Aquificae* at high temperatures throughout hydrothermal systems around the world collectively suggests that this group is of pivotal importance for the productivity of high temperature ecosystems. In addition to enhancing the ability of the whole community to fix carbon via the rTCA pathway, this alternative autotrophic pathway can produce a diverse array of metabolic intermediates and expand available niches in thermal springs, which may also contribute to *in situ* productivity and diversity (Everroad et al. 2012).

The candidate division EM3, the dominant component in Sample 10ymid57, is particularly intriguing because Sample 10ymid57 was the only non-*Aquificae* predominant sample within the HT group (Fig. 1B). The EM3 are often found in thermal springs (Meyer-Dombard et al. 2011; Zhang et al. 2013) and wastewaters (Hatamoto et al. 2007; Sekiguchi 2006) and has been described as a non-dominant component in some YNP sites. However, these sites are all dominated by *Aquificales*

and the temperatures are relatively lower (65 ~ 80 ℃) (Meyer-Dombard et al. 2011), whereas the sample temperature at which 10ymid57 was collected was 87 ℃. Molecular phylogenetic studies suggest that the candidate division EM3 is closely related to the phylum *Thermotogae* (Dunfield et al. 2012; Rinke et al. 2013), yet little is known about the metabolic potential and the ecology of the EM3 due to the lack of cultivated representatives.

*Bacterial communities in the Photosynthesis Upper Limit group*

The break between the HT group and the PUL group was at 73-75 ℃, below which is considered suitable for photosynthesis. The temperature range of photosynthesis upper limit (73-75 ℃) has long been acknowledged (Brock 1967). However, microbial communities at 73-75 ℃ were previously treated as a part of the phototrophic communities (Everroad et al. 2012; Wang et al. 2013). The microbial assemblages in the transition fringe between the chemosynthetic zone and photosynthetic zone have been neglected. Our cluster analysis indicated that microbial communities at 73-75 ℃ clustered into an individual group (Fig. 1 and 2).

The PUL group was dominated by the genus *Thermus* within the phylum *Deinococcus-Thermus*, while *Aquificae* became less important (Fig. 1B, S4 and Table 3). The occurrence of *Thermus* in the PUL group mirrors widespread distribution of this group in global thermal springs (Boomer et al. 2009; Cole et al. 2013; Costa et al. 2009; Gumerov et al. 2011; Lin et al. 2002; Purcell et al. 2007; Tobler and Benning 2011; Wang et al. 2013). However, the genus *Meiothermus*, another important genus within *Deinococcus-Thermus*, which commonly inhabits thermal springs, was completely absent at 73-75 ℃, suggesting distinct ecological niches between these two genera. Although both *Meiothermus* and *Thermus* can inhabit the temperature above 50 ℃ and circumneutral pH, most species within genus *Meiothermus* are aerobes with

an optimum growth temperature of 50-65 ℃. On the other hand, the *Thermus* are nitrate-respiring anaerobes with an optimum growth temperature of 65-75 ℃ (Nobre et al. 1996). Given that *Thermus* are normally chemoorganotrophs, the dominance of *Thermus* may coincide with the abundance of organic carbon as observed elsewhere (Wang et al. 2013), indicating organic carbon availability in this transition fringe.

In this study, *Armatimonadetes* were ubiquitous in all temperature groups and were large components of the PUL group, accounting for approximately 29.6% (Fig. S4). *Armatimonadetes* were originally discovered in Obsidian Pool of YNP and were named after candidate division OP10 (Hugenholtz et al. 1998). Despite only a few strains isolated to date (Boomer et al. 2009; Gumerov et al. 2011; Im et al. 2012; Lee et al. 2011), the phylum *Armatimonadetes* is estimated to comprise 12 groups (Dunfield et al. 2012), occurring in a wide diversity of environments, such as geothermal soils (Boomer et al. 2009), ginseng farm soils (Im et al. 2012) and a metal-rich lake (Stott et al. 2008). We detected the *Armatimonadetes* within a temperature range of 47-84 ℃. In contrast, other previous studies reached different conclusions that the *Armatimonadetes* were present within narrow temperature ranges. The *Armatimonadetes* were only found to inhabit below 60-66 ℃ in a Japanese thermal spring (Nakagawa and Fukui 2002), while the *Armatimonadetes* were detected at 72 ℃ in Alvord Hot Spring, Oregon (Connon et al. 2008). The disparate temperature distribution of the *Armatimonadetes* suggests that this phylum consists of diverse species with respective physiologies.

*Bacterial communities in the Intermediate Temperature group and the Low Temperature group*

Members of the ecologically diverse groups of *Cyanobacteria* and *Chloroflexi* are often detected below 73 ℃ in circumneutral to alkaline thermal springs (Inskeep et

al. 2013b; Klatt et al. 2011). The majority of *Cyanobacteria* sequences in this study were represented by two orders, the *Gloeobacterales* (*e.g., Gloeobacter*) and the *Pseudanabaenales* (*e.g., Pseudanabaena*) (Fig. 3B). Sequences affiliated with the *Gloeobacterales* (*e.g., Gloeobacter*) were widely distributed at temperatures below 73 ℃ in this study (Fig. 3B). The presence of the order *Pseudanabaenales* was also observed at low temperatures in Tibetan thermal springs (Wang et al. 2013). The phylum *Chloroflexi* detected here consisted of the class *Anaerolineae* (*e.g.*, uncultured group WCHB1-50) and the class *Chloroflexi* (*e.g.*, genus *Chloroflexus*) (Fig. 3B). The former is a chemoorganotroph obtaining energy from fermentation (Yamada et al. 2006), whereas the latter is an anoxygenic phototroph. The co-occurrence of these two classes in the IT and LT groups is in agreement with the previous study in Tibetan thermal springs (Wang et al. 2013). The successive additions of anoxygenic and oxygenic phototrophs in IT and LT group is expected to increase of photosynthesis and carbon fixation activities. The *Cyanobacteria* in thermal springs are considered to increase productivity, and availability of oxygen and extracellular organic compounds, which contribute to the association of various heterotrophs (Everroad et al. 2012). As was expected, bacterial communities at low temperatures were characterized by appearance and increase of members of heterotrophic bacterial groups including *Acidobacteria*, *Bacteroidetes*, *Chlorobi*, *Firmicutes*, *Planctomycetes* and *Proteobacteria* (Fig. 1B). This positive relationship between the productivity and biodiversity was observed at both mesophilic temperatures and thermophilic temperatures in freshwater ecosystems (Everroad et al. 2012; Fukami and Morin 2003).

The *Cyanobacteria* are important oxygenic phototrophs and play major roles in increasing productivity. Oxygen and organic compounds produced by the

*Cyanobacteria* are important nutrient and energy sources for a variety of heterotrophs in thermal springs. The *Chloroflexi* are major components of photosynthetic microbial mats widely distributed in different types of YNP thermal springs (Bauld and Brock 1973; Boomer et al. 2002; Giovannoni et al. 1987; Nubel et al. 2002; Ward et al. 1998b). Our analyses suggest that *Cyanobacteria* and *Chloroflexi* are not simple passive or transient colonizers in 'filamentous streamer' springs, but rather appear to be adapted to the optimal temperatures associated with phototrophs. For example, within the temperature range of 40-73 ℃, we observed that the relative abundance of *Cyanobacteria* was negatively correlated with that of Chloroflexi (Fig. 3). A previous trend between these two phyla was also observed within the temperature range of 38-73 ℃ elsewhere in YNP (Miller et al., 2009). Additionally, this pattern of taxonomical segregation was also corroborated by Null model analysis. According to the neutral theory (Hubbell 2001), stochasticity is the primary drive force shaping communities. If the role of stochastic processes is more important, the pattern of species co-occurrence is random. In contrast, if environmental variables and species interaction play major roles, community structures are not random (*i.e.*, there is segregation of taxa). Microbial community assemblages reflecting the transition of critical environmental variables such as moisture (Van Horn et al. 2014), salinity (Lozupone and Knight 2007), organic resources (Van Horn et al. 2014) and pH (Fierer and Jackson 2006) have been previously noted. Our null model analysis revealed this nonrandom distribution pattern of the *Cyanobacteria* and the *Chloroflexi*. The observed nonrandom structures may result from environmental filtering or species interactions. Temperature was a pronounced environmental factor in filtering the communities based on the BIO-ENV results (Table S2). If the nonrandom community structure is exclusively indicative of differences temperature tolerances among taxa,

we would expect to see a random structure for data pooled from the samples with similar temperatures, irrespective of the origin of thermal spring. To rule out alternative environmental factors, we pooled and sampled from similar temperatures from different springs. We observed random structure ($P_{obs>exp}$>0.05) in all cases: 10mid49 (OS, 66 ℃), 10ylow16 (RT, 67 ℃) and 10ylow28 (BP, 68 ℃); 03ylow18 (RT, 57 ℃), 03ylow17 (RT, 62 ℃) and 10ylow27 (BP, 63 ℃); 10ylow25 (BP, 40 ℃), 03ymid47 (OS, 47 ℃) and 03ylow19 (RT, 54 ℃). Thus, temperature is the most important environmental factor in shaping the structure of these phototrophic communities. The distribution of *Chloroflexus* tended to reflect the adaptation of temperature observed previously. For example, *Chloroflexus aurantiacus* Y-400 is only detected in high temperature mats from alkaline springs, while other uncultivated *Chloroflexus* have broader temperature tolerance (Ruff-Roberts et al. 1994). Additionally, species interactions such cross-feed and competition may also contribute to the species distribution. Stable carbon isotope analysis has suggested that some members in the *Chloroflexi* such as *Chloroflexus*-like organisms can grow photoheterotrophically and photoautotrophically (van der Meer et al. 2003). When *Chloroflexus* and *Cyanobacteria* occur in photosynthetic microbial mats, *Chloroflexus* normally switch to photoheterotrophic metabolism via cross-feeding the organic compounds produced by *Cyanobacteria* rather than compete with *Cyanobacteria* for inorganic carbon, which is particularly true of mats in alkaline springs, where $CO_2$ is limited (van der Meer et al. 2003). We observed that the relative abundance of *Cyanobacteria* was higher than that of the *Chloroflexi* in most of sites (Fig. 3), which may result from the photoautotrophic *Chloroflexi* outcompeted by the *Cyanobacteria*. Only *Chloroflexi* species that successfully switch to photoheterotrophic carbon metabolism can survive, which may explain the relatively low relative abundance in

most sites. When using the phototrophic metabolic strategy, the organic carbon sources largely depend on the primary producers, the *Cyanobacteria*. Given the organic carbon loss in springs or it consumed by other heterotrophic organisms, it is not surprising to observe a relatively low abundance of the *Chloroflexi*. Furthermore, species competition for limited resources or may also contribute to community organization. Light is considered the primary limiting resource and a master control for phototrophic communities such as plant communities (Gotelli and McCabe 2002). The isolates of the cyanobacterial genus *Synechococcus* from YNP thermal springs indicate broader temperature range distributions than environmental *Synechococcus* OTUs (Allewalt et al. 2006; Miller and Castenholz 2000; Miller et al. 2009). The difference between fundamental and realized niche width results from competition for limited resources such as light (Weltzer and Miller 2013).

**Conclusions**

The current study highlights that temperature plays a key role in controlling on the bacterial community structure and diversity in alkaline thermal springs of YNP. Despite previous efforts made in examining the microbial communities along the outflow channels in YNP thermal springs, the broader temperature range and a good coverage of the barcoded pyrosequencing adopted in this survey lent strong support to our results. Bacterial community structure and diversity sampled from three geochemically similar springs did not indicate any site-specific variations. Instead, we observed communities clustered into four distinct group related to temperature. Bacterial richness and diversity increased with decreasing temperature. In aggregate, our results demonstrate that temperature acts as a strong environmental stressor in alkaline springs at the local scale. In addition, temperature is of pivotal importance for determining community functions of geothermal ecosystems. In this study, as

temperature decreased along thermal gradients, transitions to downstream cooler communities were represented by the succession of chemolithotrophs, chemoorganotrophs, anoxygenic phototrophs and oxygenic phototrophs.

**References**


1. Acosta-Mart ńez, V., S. Dowd, Y. Sun, and V. Allen. 2008. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biol. Biochem. 40:2762-2770.
2. Allewalt, J.P., M.M. Bateson, N.P. Revsbech, K. Slack, and D.M. Ward. 2006. Effect of temperature and light on growth of and photosynthesis by *synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of yellowstone national park. Appl. Environ. Microbiol. 72:544-550.
3. Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26:32-46.
4. Ball, J.W., D.K. Nordstrom, K.M. Cunningham, M.A.A. Schoonen, Y. Xu, and J.M. DeMonge. 1998. Water-chemistry and on-site sulfur-speciation data for selected springs in yellowstone national park, wyoming, 1994-1995. U.S. Geological Survey Open-File Report:98-574.
5. Ball, J.W., R.B. McCleskey, D.K. Nordstrom, J.M. Holloway, and P.L. Verplanck. 2002. Water-chemistry data for selected springs, geysers, and streams in yellowstone national park, wyoming, 1999-2000. U.S. Geological Survey Open-File Report:02-382.
6. Ball, J.W., R.B. McCleskey, and D.K. Nordstrom. 2010. Water-chemistry data for selected springs, geysers, and streams in yellowstone national park, wyoming, 2006–2008. U.S. Geological Survey Open-File Report 2010–1192:109 p.
7. Barns, S.M., R.E. Fundyga, M.W. Jeffries, and N.R. Pace. 1994. Remarkable archaeal diversity detected in a yellowstone national park hot spring

environment. Proceedings of the National Academy of Sciences of the United States of America 91:1609-1613.

8. Barns, S.M., C.F. Delwiche, J.D. Palmer, and N.R. Pace. 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rrna sequences. Proceedings of the National Academy of Sciences 93:9188-9193.

9. Bauld, J., and T. Brock. 1973. Ecological studies of chloroflexis, a gliding photosynthetic bacterium. Archiv. Mikrobiol. 92:267-284.

10. Beh, M., G. Strauss, R. Huber, K.-O. Stetter, and G. Fuchs. 1993. Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *aquifex pyrophilus* and in the archaebacterium *thermoproteus neutrophilus*. Arch. Microbiol. 160:306-311.

11. Blank, C.E., S.L. Cady, and N.R. Pace. 2002. Microbial composition of near-boiling silica-depositing thermal springs throughout yellowstone national park. Appl. Environ. Microbiol. 68:5123-5135.

12. Boomer, S.M., D.P. Lodge, B.E. Dutton, and B. Pierson. 2002. Molecular characterization of novel red green nonsulfur bacteria from five distinct hot spring communities in yellowstone national park. Appl. Environ. Microbiol. 68:346-355.

13. Boomer, S.M., K.L. Noll, G.G. Geesey, and B.E. Dutton. 2009. Formation of multilayered photosynthetic biofilms in an alkaline thermal spring in yellowstone national park, wyoming. Appl. Environ. Microbiol. 75:2464-2475.

14. Boyd, E., T. Hamilton, J. Wang, L. He, and C. Zhang. 2013. The role of tetraether lipid composition in the adaptation of thermophilic archaea to acidity. Frontiers in Microbiology 4.

15. Boyd, E.S., T.L. Hamilton, J.R. Spear, M. Lavin, and J.W. Peters. 2010. [fefe]-hydrogenase in yellowstone national park: Evidence for dispersal limitation and phylogenetic niche conservatism. Isme j 4:1485-1495.

16. Brock, T.D. 1967. Life at high temperatures. Science 158:1012-1019.

17. Brock, T.D., and T.D. Brock. 1978. Thermophilic microorganisms and life at high temperatures. Springer-Verlag New York.

18. Burgess, E., J. Unrine, G. Mills, C. Romanek, and J. Wiegel. 2012. Comparative geochemical and microbiological characterization of two thermal pools in the uzon caldera, kamchatka, russia. Microb. Ecol. 63:471-489.

19. Caporaso, J.G., K. Bittinger, F.D. Bushman, T.Z. DeSantis, G.L. Andersen, and R. Knight. 2010. Pynast: A flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266-267.

20. Caporaso, J.G., J. Kuczynski, J. Stombaugh, *et al.* 2010. Qiime allows analysis of high-throughput community sequencing data. Nat Methods 7:335-336.

21. Cavicchioli, R. 2002. Extremophiles and the search for extraterrestrial life. Astrobiology 2:281-292.

22. Clarke, K., and M. Ainsworth. 1993. A method of linking multivariate community structure to environmental variables. Marine Ecology-Progress Series 92:205-205.

23. Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. Aust. J. Ecol. 18:117-143.

24. Cole, J.K., J.P. Peacock, J.A. Dodsworth, A.J. Williams, D.B. Thompson, H. Dong, G. Wu, and B.P. Hedlund. 2013. Sediment microbial communities in great boiling spring are controlled by temperature and distinct from water communities. ISME J 7:718-729.

25. Connon, S.A., A.K. Koski, A.L. Neal, S.A. Wood, and T.S. Magnuson. 2008. Ecophysiology and geochemistry of microbial arsenic oxidation within a high arsenic, circumneutral hot spring system of the alvord desert. FEMS Microbiol. Ecol. 64:117-128.

26. Costa, K., J. Navarro, E. Shock, C. Zhang, D. Soukup, and B. Hedlund. 2009. Microbiology and geochemistry of great boiling and mud hot springs in the united states great basin. Extremophiles 13:447-459.

27. Cox, A., E.L. Shock, and J.R. Havig. 2011. The transition to microbial photosynthesis in hot spring ecosystems. Chemical Geology 280:344-351.

28. Dequiedt, S., J. Thioulouse, C. Jolivet, *et al.* 2009. Biogeographical patterns of soil bacterial communities. Environmental microbiology reports 1:251-255.

29. Dequiedt, S., J. Thioulouse, C. Jolivet, *et al.* 2009. Biogeographical patterns of soil bacterial communities. Environmental Microbiology Reports 1:251-255.

30. DeSantis, T.Z., P. Hugenholtz, N. Larsen, *et al.* 2006. Greengenes, a chimera-checked 16s rrna gene database and workbench compatible with arb. Appl. Environ. Microbiol. 72:5069-5072.

31. Dowd, S., T. Callaway, R. Wolcott, Y. Sun, T. McKeehan, R. Hagevoort, and T. Edrington. 2008. Evaluation of the bacterial diversity in the feces of cattle using 16s rdna bacterial tag-encoded flx amplicon pyrosequencing (btefap). BMC Microbiol. 8:125.

32. Dunfield, P.F., I. Tamas, K.C. Lee, X.C. Morgan, I.R. McDonald, and M.B. Stott. 2012. Electing a candidate: A speculative history of the bacterial phylum op10. Environ Microbiol 14:3069-3080.

33. Eder, W., and R. Huber. 2002. New isolates and physiological properties of the aquificales and description of *thermocrinis albus* sp. Nov. Extremophiles 6:309-318.

34. Edgar, R.C. 2010. Search and clustering orders of magnitude faster than blast. Bioinformatics 26:2460-2461.

35. Entsminger, G. 2012. Ecosim professional: Null modeling software for ecologists, version 1. Acquired intelligence inc., kesey-bear, & pinyon publishing. Montrose, co 81403. http://www.garyentsminger.com/ecosim/index.htm.

36. Everroad, R.C., H. Otaki, K. Matsuura, and S. Haruta. 2012. Diversification of bacterial community composition along a temperature gradient at a thermal spring. Microbes Environ. 27:374-381.

37. Faith, D.P. 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61:1-10.

38. Ferrera, I., S. Longhorn, A. Banta, Y. Liu, D. Preston, and A.L. Reysenbach. 2007. Diversity of 16s rrna gene, its region and aclb gene of the *aquificales*. Extremophiles 11:57-64.

39. Fierer, N., and R.B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences of the United States of America 103:626-631.

40. Fukami, T., and P.J. Morin. 2003. Productivity-biodiversity relationships depend on the history of community assembly. Nature 424:423-426.

41. Giovannoni, S.J., N.P. Revsbech, D.M. Ward, and R.W. Castenholz. 1987. Obligately phototrophic chloroflexus: Primary production in anaerobic hot spring microbial mats. Arch. Microbiol. 147:80-87.

42. Giovannoni, S.J., E.F. DeLong, T.M. Schmidt, and N.R. Pace. 1990. Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. Appl. Environ. Microbiol. 56:2572-2575.

43. Good, I.J. 1953. The population frequencies of species and the estimation of population parameters. Biometrika 40:237-264.

44. Gotelli, N.J., and D.J. McCabe. 2002. Species co-occurrence: A meta-analysis of jm diamond's assembly rules model. Ecology 83:2091-2096.

45. Graber, J.R., J. Kirshtein, M. Speck, and A.-L. Reysenbach. 2001. Community structure along a thermal gradient in a stream near obsidian pool, yellowstone national park, p. 81-91, Thermophiles: Biodiversity, ecology, and evolution. Springer, New York.

46. Gumerov, V.M., A.V. Mardanov, A.V. Beletsky, E.A. Bonch-Osmolovskaya, and N.V. Ravin. 2011. Molecular analysis of microbial diversity in the zavarzin spring, uzon caldera, kamchatka. Microbiology 80:244-251.

47. Hall, J.R., K.R. Mitchell, O. Jackson-Weaver, A.S. Kooser, B.R. Cron, L.J. Crossey, and C.D. Takacs-Vesbach. 2008. Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rrna and metabolic genes. Appl Environ Microbiol 74:4910-4922.

48. Hanson, C.A., J.A. Fuhrman, M.C. Horner-Devine, and J.B.H. Martiny. 2012. Beyond biogeographic patterns: Processes shaping the microbial landscape. Nat Rev Micro 10:497-506.

49. Harmsen, H., D. Prieur, and C. Jeanthon. 1997. Group-specific 16s rrna-targeted oligonucleotide probes to identify thermophilic bacteria in marine hydrothermal vents. Appl. Environ. Microbiol. 63:4061-4068.

50. Hatamoto, M., H. Imachi, Y. Yashiro, A. Ohashi, and H. Harada. 2007. Diversity of anaerobic microorganisms involved in long-chain fatty acid degradation in methanogenic sludges as revealed by rna-based stable isotope probing. Appl. Environ. Microbiol. 73:4119-4127.

51. Horner-Devine, M.C., J.M. Silver, M.A. Leibold, *et al.* 2007. A comparison of taxon co-occurrence patterns for macro- and microorganisms. Ecology 88:1345-1353.

52. Hou, W., S. Wang, H. Dong, *et al.* 2013. A comprehensive census of microbial diversity in hot springs of tengchong, yunnan province china using 16s rrna gene pyrosequencing. PLoS ONE 8:e53350.

53. Hubbell, S.P. 2001. The unified neutral theory of biodiversity and biogeography vol. 32. Princeton University Press.

54. Huber, R., and K.O. Stetter. 2001. Aquificales, Els. John Wiley & Sons, Ltd.

55. Hugenholtz, P., C. Pitulle, K.L. Hershberger, and N.R. Pace. 1998. Novel division level bacterial diversity in a yellowstone hot spring. J. Bacteriol. 180:366-376.

56. Hugler, M., H. Huber, S.J. Molyneaux, C. Vetriani, and S.M. Sievert. 2007. Autotrophic $co_2$ fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum *aquificae*: Evidence for two ways of citrate cleavage. Environ Microbiol 9:81-92.

57. Im, W.-T., Z.-Y. Hu, K.-H. Kim, S.-K. Rhee, H. Meng, S.-T. Lee, and Z.-X. Quan. 2012. Description of *fimbriimonas ginsengisoli* gen. Nov., sp. Nov. Within the *fimbriimonadia* class nov., of the phylum *armatimonadetes*. Antonie Van Leeuwenhoek 102:307-317.

58. Inagaki, F., K. Takai, H. Hirayama, Y. Yamato, K.H. Nealson, and K. Horikoshi. 2003. Distribution and phylogenetic diversity of the subsurface microbial community in a japanese epithermal gold mine. Extremophiles 7:307-317.

59. Inskeep, W.P., D.B. Rusch, Z.J. Jay, *et al.* 2010. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS ONE 5:e9773.

60. Inskeep, W.P., Z.J. Jay, M.J. Herrgard, *et al.* 2013. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. Front Microbiol 4:95.

61. Inskeep, W.P., Z.J. Jay, S.G. Tringe, M.J. Herrgard, and D.B. Rusch. 2013. The ynp metagenome project: Environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. Front Microbiol 4:67.

62. Jahnke, L.L., W. Eder, R. Huber, J.M. Hope, K.-U. Hinrichs, J.M. Hayes, D.J. Des Marais, S.L. Cady, and R.E. Summons. 2001. Signature lipids and stable carbon isotope analyses of octopus spring hyperthermophilic communities compared with those of aquificales representatives. Appl. Environ. Microbiol. 67:5179-5189.

63. Klatt, C.G., J.M. Wood, D.B. Rusch, *et al.* 2011. Community ecology of hot spring cyanobacterial mats: Predominant populations and their functional potential. ISME J 5:1262-1278.

64. Klatt, C.G., W.P. Inskeep, M.J. Herrgard, *et al.* 2013. Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. Front Microbiol 4:106.

65. Kozubal, M.A., R.E. Macur, Z.J. Jay, J.P. Beam, S.A. Malfatti, S.G. Tringe, B.D. Kocar, T. Borch, and W.P. Inskeep. 2012. Microbial iron cycling in acidic geothermal springs of yellowstone national park: Integrating molecular surveys, geochemical processes, and isolation of novel fe-active microorganisms. Front Microbiol 3:109.

66. Kozubal, M.A., M. Romine, R.d. Jennings, Z.J. Jay, S.G. Tringe, D.B. Rusch, J.P. Beam, L.A. McCue, and W.P. Inskeep. 2013. Geoarchaeota: A new candidate phylum in the archaea from high-temperature acidic iron mats in yellowstone national park. ISME J 7:622-634.

67. Lau, M.C., J.C. Aitchison, and S.B. Pointing. 2009. Bacterial community composition in thermophilic microbial mats from five hot springs in central tibet. Extremophiles 13:139-149.

68. Lauber, C.L., M. Hamady, R. Knight, and N. Fierer. 2009. Pyrosequencing-based assessment of soil ph as a predictor of soil bacterial community structure at the continental scale. Appl. Environ. Microbiol. 75:5111-5120.

69. Lee, K.C.-Y., P.F. Dunfield, X.C. Morgan, *et al.* 2011. *Chthonomonas calidirosea* gen. Nov., sp. Nov., an aerobic, pigmented, thermophilic micro-organism of a novel bacterial class, *chthonomonadetes* classis nov., of the newly described phylum *armatimonadetes* originally designated candidate division op10. Int. J. Syst. Evol. Microbiol. 61:2482-2490.

70. Ley, R.E., J.K. Harris, J. Wilcox, *et al.* 2006. Unexpected diversity and complexity of the guerrero negro hypersaline microbial mat. Appl. Environ. Microbiol. 72:3685-3695.

71. Lin, L., C. Chen, Q. Peng, K. Ben, and Z. Zhou. 2002. *Thermus rehai* sp. Nov., isolated from rehai of tengchong, yunnan province, china. J Basic Microbiol 42:337-344.

72. Lozupone, C.A., and R. Knight. 2007. Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences 104:11436-11440.

73. Macur, R.E., Z.J. Jay, W.P. Taylor, M.A. Kozubal, B.D. Kocar, and W.P. Inskeep. 2013. Microbial community structure and sulfur biogeochemistry in mildly-acidic sulfidic geothermal springs in yellowstone national park. Geobiology 11:86-99.

74. McCleskey, R.B. 2005. Water-chemistry data for selected hot springs, geysers, and streams in yellowstone national park, wyoming, 2001-2002. US Geological Survey.

75. Meyer-Dombard, D.R., E.L. Shock, and J.P. Amend. 2005. Archaeal and bacterial communities in geochemically diverse hot springs of yellowstone national park, USA. Geobiology 3:211-227.

76. Meyer-Dombard, D.R., W. Swingley, J. Raymond, J. Havig, E.L. Shock, and R.E. Summons. 2011. Hydrothermal ecotones and streamer biofilm communities in the lower geyser basin, yellowstone national park. Environ Microbiol 13:2216-2231.

77. Mielke, P.W., K.J. Berry, P.J. Brockwell, and J.S. Williams. 1981. A class of nonparametric tests based on multiresponse permutation procedures. Biometrika 68:720-724.

78. Miller, S.R., and R.W. Castenholz. 2000. The evolution of thermotolerance in hot spring cyanobacteria of the genus *synechococcus*. J. Phycol. 36:48-48.

79. Miller, S.R., A.L. Strong, K.L. Jones, and M.C. Ungerer. 2009. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in yellowstone national park. Appl. Environ. Microbiol. 75:4565-4572.

80. Mitchell, K., and C. Takacs-Vesbach. 2008. A comparison of methods for total community DNA preservation and extraction from various thermal environments. J. Ind. Microbiol. Biotechnol. 35:1139-1147.

81. Mitchell, K.R. 2009. Controls on microbial community structure in thermal environments : Exploring bacterial diversity and the relative influence of geochemistry and geography. University of New Mexico.

82. Nacke, H., A. Thürmer, A. Wollherr, C. Will, L. Hodac, N. Herold, I. Schöning, M. Schrumpf, and R. Daniel. 2011. Pyrosequencing-based assessment of bacterial community structure along different management types in german forest and grassland soils. PloS one 6:e17000.

83. Nakagawa, T., and M. Fukui. 2002. Phylogenetic characterization of microbial mats and streamers from a japanese alkaline hot spring with a thermal gradient. The Journal of general and applied microbiology 48:211-222.

84. Nobre, M.F., H.G. Truper, and M.S.D. Costa. 1996. Transfer of *thermus ruber* (loginova et al. 1984), *thermus silvanus* (tenreiro et al. 1995), and *thermus chliarophilus* (tenreiro et al. 1995) to *meiothermus* gen. Nov. As *meiothermus ruber* comb. Nov., *meiothermus silvanus* comb. Nov., and *meiothermus chliarophilus* comb. Nov., respectively, and emendation of the genus t*hermus*. Int. J. Syst. Bacteriol. 46:604-606.

85. Nubel, U., M.M. Bateson, V. Vandieken, A. Wieland, M. Kuhl, and D.M. Ward. 2002. Microscopic examination of distribution and phenotypic properties of phylogenetically diverse chloroflexaceae-related bacteria in hot spring microbial mats. Appl Environ Microbiol 68:4593-4603.

86. Oksanen, J., F.G. Blanchet, R. Kindt, *et al.* Vegan: Community ecology package, 2012. R package version:2.0-5.

87. Otaki, H., R.C. Everroad, K. Matsuura, and S. Haruta. 2011. Production and consumption of hydrogen in hot spring microbial mats dominated by a filamentous anoxygenic photosynthetic bacterium. Microbes and environments/JSME 27:293-299.

88. Purcell, D., U. Sompong, L.C. Yim, T.G. Barraclough, Y. Peerapornpisal, and S.B. Pointing. 2007. The effects of temperature, ph and sulphide on the community structure of hyperthermophilic streamers in hot springs of northern thailand. FEMS Microbiol. Ecol. 60:456-466.

89. Quince, C., A. Lanzen, R.J. Davenport, and P.J. Turnbaugh. 2011. Removing noise from pyrosequenced amplicons. BMC Bioinformatics. 2011 Jan 28 12:38.

90. Reysenbach, A.-L., A. Banta, S.C. Daly, K. Mitchel, S. Lalonde, K. Konhauser, A. Rodman, K. Rusterholtz, and C. Takacs-Vesbach. 2005. Aquificales in yellowstone national park. Geothermal Biology and Geochemistry in YNP 1:129-142.

91. Reysenbach, A.-L., N. Hamamura, M. Podar, *et al.* 2009. Complete and draft genome sequences of six members of the *aquificales*. J. Bacteriol. 191:1992-1993.

92. Reysenbach, A.L., S. Seitzinger, J. Kirshtein, and E. McLaughlin. 1999. Molecular constraints on a high-temperature evolution of early life. Biological Bulletin 196:367-372.

93. Reysenbach, A.L., H. Ehringer, and K. Hershberger. 2000. Microbial diversity at 83°c in calcite springs, yellowstone national park: Another environment where the *aquificales* and "korarchaeota" coexist. Extremophiles 4:61-67.

94. Rinke, C., P. Schwientek, A. Sczyrba, *et al.* 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431-437.

95. Robinson, C.K., J. Wierzchos, C. Black, *et al.* 2014. Microbial diversity and the presence of algae in halite endolithic communities are correlated to atmospheric moisture in the hyper- arid zone of the atacama desert. Environ. Microbiol.

96. Rothschild, L.J., and R.L. Mancinelli. 2001. Life in extreme environments. Nature 409:1092-1101.

97. Ruff-Roberts, A.L., J.G. Kuenen, and D.M. Ward. 1994. Distribution of cultivated and uncultivated cyanobacteria and chloroflexus-like bacteria in hot spring microbial mats. Appl. Environ. Microbiol. 60:697-704.

98. Schwartz, E., D.J. Van Horn, H.N. Buelow, J.G. Okie, M.N. Gooseff, J.E. Barrett, and C.D. Takacs-Vesbach. 2014. Characterization of growing bacterial populations in mcmurdo dry valley soils through stable isotope probing with 18o-water, vol. 89.

99. Segawa, T., N. Takeuchi, K. Ushida, H. Kanda, and S. Kohshima. 2010. Altitudinal changes in a bacterial community on gulkana glacier in alaska. Microbes and environments / JSME 25:171-182.

100. Sekiguchi, Y. 2006. Yet-to-be cultured microorganisms relevant to methane fermentation processes. Microbes Environ. 21:1-15.

101. Shiba, H., T. Kawasumi, Y. Igarashi, T. Kodama, and Y. Minoda. 1985. The $co_2$ assimilation via the reductive tricarboxylic acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, *hydrogenobacter thermophilus*. Arch. Microbiol. 141:198-203.

102. Signori, C.N., F. Thomas, A. Enrich-Prast, R.C.G. Pollery, and S.M. Sievert. 2014. Microbial diversity and community structure across environmental gradients in bransfield strait, western antarctic peninsula. Frontiers in Microbiology 5:647.

103. Song, Z.-Q., F.-P. Wang, X.-Y. Zhi, *et al.* 2013. Bacterial and archaeal diversities in yunnan and tibetan hot springs, china. Environ. Microbiol. 15:1160-1175.

104. Spear, J.R., J.J. Walker, T.M. McCollom, and N.R. Pace. 2005. Hydrogen and bioenergetics in the yellowstone geothermal ecosystem. Proceedings of the National Academy of Sciences of the United States of America 102:2555-2560.

105. Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. Oecologia 85:74-79.

106. Stott, M.B., M.A. Crowe, B.W. Mountain, A.V. Smirnova, S. Hou, M. Alam, and P.F. Dunfield. 2008. Isolation of novel bacteria, including a candidate division, from geothermal soils in new zealand. Environ. Microbiol. 10:2030-2041.

107. Takacs-Vesbach, C., W.P. Inskeep, Z.J. Jay, *et al.* 2013. Metagenome sequence analysis of filamentous microbial communities obtained from

geochemically distinct geothermal channels reveals specialization of three aquificales lineages. Front Microbiol 4:84.

108. Team, R.D.C. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

109. Tobler, D.J., and L.G. Benning. 2011. Bacterial diversity in five icelandic geothermal waters: Temperature and sinter growth rate effects. Extremophiles 15:473-485.

110. van der Meer, M.T.J., S. Schouten, J.S. Sinninghe Damsté, J.W. de Leeuw, and D.M. Ward. 2003. Compound-specific isotopic fractionation patterns suggest different carbon metabolisms among chloroflexus-like bacteria in hot-spring microbial mats. Appl. Environ. Microbiol. 69:6000-6006.

111. Van Horn, D.J., M.L. Van Horn, J.E. Barrett, M.N. Gooseff, A.E. Altrichter, K.M. Geyer, L.H. Zeglin, and C.D. Takacs-Vesbach. 2013. Factors controlling soil microbial biomass and bacterial diversity and community composition in a cold desert ecosystem: Role of geographic scale. PLoS ONE 8:e66103.

112. Van Horn, D.J., J.G. Okie, H.N. Buelow, M.N. Gooseff, J.E. Barrett, and C.D. Takacs-Vesbach. 2014. Soil microbial responses to increased moisture and organic resources along a salinity gradient in a polar desert. Appl. Environ. Microbiol. 80:3034-3043.

113. Wang, Q., G.M. Garrity, J.M. Tiedje, and J.R. Cole. 2007. Naïve bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73:5261-5267.

114. Wang, S., W. Hou, H. Dong, *et al.* 2013. Control of temperature on microbial community structure in hot springs of the tibetan plateau. PLoS ONE 8:e62901.

115. Ward, D.M., M.J. Ferris, S.C. Nold, and M.M. Bateson. 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol. Mol. Biol. Rev. 62:1353-1370.

116. Ward, D.M., M.J. Ferris, S.C. Nold, and M.M. Bateson. 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62:1353-1370.

117. Weltzer, M., and S. Miller. 2013. Division-specific differences in bacterial community assembly of two alkaline hot spring ecosystems from yellowstone national park. Microb. Ecol. 65:537-540.

118. Xie, W., C.L. Zhang, J. Wang, *et al.* 2014. Distribution of ether lipids and composition of the archaeal community in terrestrial geothermal springs: Impact of environmental variables. Environ. Microbiol.:n/a-n/a.

119. Yamada, T., Y. Sekiguchi, S. Hanada, H. Imachi, A. Ohashi, H. Harada, and Y. Kamagata. 2006. *Anaerolinea thermolimosa* sp. Nov., *levilinea saccharolytica* gen. Nov., sp. Nov. And *leptolinea tardivitalis* gen. Nov., sp. Nov., novel filamentous anaerobes, and description of the new classes anaerolineae classis nov. And caldilineae classis nov. In the bacterial phylum chloroflexi. Int. J. Syst. Evol. Microbiol. 56:1331-1340.

120. Yamamoto, H., A. Hiraishi, K. Kato, H.X. Chiura, Y. Maki, and A. Shimizu. 1998. Phylogenetic evidence for the existence of novel thermophilic bacteria in hot spring sulfur-turf microbial mats in japan. Appl. Environ. Microbiol. 64:1680-1687.

121. Zhang, C., J. Wang, J. Dodsworth, A. Williams, C. Zhu, K.-U. Hinrichs, F. Zheng, and B.P. Hedlund. 2013. *In situ* production of branched glycerol

dialkyl glycerol tetraethers in a great basin hot spring (USA). Frontiers in Microbiology 4.

122. Zhang, C.L., B.W. Fouke, G.T. Bonheyo, A.D. Peacock, D.C. White, Y. Huang, and C.S. Romanek. 2004. Lipid biomarkers and carbon-isotopes of modern travertine deposits (yellowstone national park, USA): Implications for biogeochemical dynamics in hot-spring systems. Geochim. Cosmochim. Acta 68:3157-3169.

**Tables and Figures**

**Table 1.** Temperature and *in situ* water chemistry parameters for sampling sites.

**Table 2.** Alpha diversity calculations at the 97% similarity OTU level.

**Table 3.** Results of SIMPER analysis displaying top ten taxa (at the 97% similarity OTU level) accounting for the most dissimilarities of between adjacent temperature groups (*i.e.,* HT *vs.* PUL, PUL *vs.* IT and IT *vs.* LT)

**Fig. 1.** The similarity of samples based on the bacterial community composition at the phylum level. (A) Cluster analysis using Bray-Curtis dissimilarity, all nodes are supported by jackknife values >99.0%. (B) Bar charts showing the community composition of each sample. Only the phyla with >1% abundance are displayed. The phyla with <1% abundance are included as 'Others'. The numbers after bar chart represent the temperatures ( ℃) of sampling locations.

**Fig. 2.** The similarity of samples based on the bacterial community composition at the phylum level. (A) Cluster analysis using Bray-Curtis dissimilarity, all nodes with jackknife values >99.0%. (B) Non-metric multidimensional scaling (NMDS) ordination analysis constructed with Bray-Curtis dissimilarity. Color coding of samples from Red Terrace (diamonds), Bison Pool (squares) and Octopus Spring (triangles) indicates the temperatures of sampling locations.

**Fig. 3.** Distribution of *Cyanobacteria* and *Chloroflexi* within the temperature range (40-75 ℃). (A) The relative abundance of total *Cyanobacteria* and *Chloroflexi*. (B) The relative abundance of major genera within *Cyanobacteria* and *Chloroflexi* as a function of temperature. Each point represent an individual sample. For both phyla, only genera with relative abundance of >5% were included. The sum of these genera within *Cyanobacteria* and *Chloroflexi* accounted for mostly negative correlations between the two phyla.

**Supplementary figures**

**Table S1.** Detailed geochemical data and site information for all springs. The ion data of BP and OS are from YNP research coordination network (http://www.rcn.montana.edu)

**Fig. S1.** Geographic maps showing sampling locations

**Fig. S2.** Piper diagram of the major anions (right ternary plot) and cations (left ternary plot). Both anions and cations in the lower triangles are projected up onto the diamond.

**Fig. S3.** (A) Regression between temperature and the number of observed OTUs for each site collected at the 97% similarity OTU level. (B) Decay of bacterial community similarity with increased difference in temperature. X-axis represents the pairwise temperature difference between samples. Y-axis represents the community dissimilarity using Bray-Curtis distance.

**Fig. S4.** Pie charts show the relative proportions of bacterial phyla in each temperature group. Only the phyla with >1% abundance are displayed. The phyla with <1% abundance are included as 'Others'.

## Table 1.

| Sample ID | Spring name | Distance from source (m) | Temp (℃) | pH | Conductivity (μS $cm^{-1}$) | Hydrogen sulfide (mg $L^{-1}$) |
|---|---|---|---|---|---|---|
| 03ylow14 | Red Terrace | 7 | 84 | 8.9 | 1391 | 0.321 |
| 03ylow16 | Red Terrace | 18 | 67 | 9 | 1431 | 0.194 |
| 03ylow17 | Red Terrace | 23 | 62 | 9 | 1454 | 0.148 |
| 03ylow18 | Red Terrace | 28 | 57 | 9.1 | 1481 | 0.097 |
| 03ylow19 | Red Terrace | 31 | 54 | 9.1 | 1481 | 0.084 |
| 10ylow35 | Bison Pool | 2 | 81 | 7.5 | 1452 | 0.155 |
| 10ylow29 | Bison Pool | 11 | 73 | 7.9 | 1504 | 0.020 |
| 10ylow28 | Bison Pool | 14 | 68 | 7.6 | 1670 | 0 |
| 10ylow27 | Bison Pool | 18 | 63 | 7.7 | 1664 | 0.030 |
| 10ylow25 | Bison Pool | 27 | 40 | 8.6 | 1460 | 0.045 |
| 10ymid57 | Octopus Spring | 3 | 87 | 7.7 | 1515 | 0.010 |
| 10ymid54 | Octopus Spring | 9 | 82 | 7.6 | 1525 | 0.025 |
| 10ymid51 | Octopus Spring | 18 | 75 | 7.8 | 1534 | 0 |
| 10ymid49 | Octopus Spring | 25 | 66 | 8 | 1532 | 0.045 |
| 10ymid47 | Octopus Spring | 51 | 47 | 8.3 | 1530 | 0.000 |

**Table 2.**

| Sample ID | Good's coverage (%) | Richness (Chao1) | NO. observed OTUs | Shannon's index | Simpson index | Phylogenetic Diversity |
|---|---|---|---|---|---|---|
| 03ylow14 | 99% | 39 | 25 | 2.69 | 0.77 | 2.30 |
| 03ylow16 | 97% | 63 | 39 | 1.82 | 0.43 | 4.40 |
| 03ylow17 | 98% | 66 | 35 | 2.43 | 0.70 | 3.80 |
| 03ylow18 | 95% | 132 | 58 | 2.34 | 0.60 | 5.22 |
| 03ylow19 | 94% | 172 | 91 | 3.94 | 0.80 | 8.52 |
| 10ylow25 | 95% | 145 | 71 | 4.16 | 0.90 | 7.73 |
| 10ylow27 | 98% | 45 | 27 | 1.92 | 0.53 | 3.14 |
| 10ylow28 | 99% | 33 | 24 | 2.38 | 0.68 | 2.66 |
| 10ylow29 | 99% | 34 | 21 | 2.53 | 0.78 | 2.06 |
| 10ylow35 | 99% | 31 | 22 | 2.19 | 0.62 | 1.97 |
| 10ymid47 | 93% | 165 | 101 | 4.64 | 0.91 | 9.63 |
| 10ymid49 | 97% | 124 | 38 | 2.72 | 0.69 | 3.62 |
| 10ymid51 | 98% | 51 | 29 | 2.36 | 0.66 | 2.19 |
| 10ymid54 | 99% | 48 | 20 | 1.76 | 0.51 | 2.30 |
| 10ymid57 | 99% | 19 | 14 | 1.89 | 0.66 | 1.34 |

## Table 3.

| HT *vs.* PUL | | | | |
|---|---|---|---|---|
| Taxon[a] | Genus/Family[b] | Contr. (%)[c] | Avg.[d] HT (%) | Avg. PUL |
| *Aquificae* | *Aquificaceae* | 21.0 | 49.8 | 7.9 |
| *Deinococcus-Thermus* | *Thermus* | 17.6 | 14.8 | 49.9 |
| *Armatimonadetes* | *unidentified Armatimonadetes* | 13.3 | 2.9 | 29.6 |
| *EM3* | *unidentified EM3* | 10.9 | 24.3 | 3.9 |
| *Chlorobi* | *unidentified Chlorobi* | 3.3 | 0.0 | 6.7 |
| *Thermodesulfobacteria* | *Geothermobacterium* | 1.5 | 2.9 | 0.0 |
| *Aquificae* | *Hydrogenothermaceae* | 1.1 | 2.2 | 0.0 |
| *OP9* | *unidentified OP9* | 0.5 | 1.0 | 0.0 |
| *Thermotogae* | *Fervidobacterium* | 0.3 | 0.7 | 0.0 |
| *Acidobacteria* | *unidentified Acidobacteria* | 0.2 | 0.0 | 0.5 |
| PUL vs. IT | | | | |
| Taxon[a] | Genus/Family[b] | Contr. (%)[c] | Avg.[d] PUL (%) | Avg. IT |
| *Cyanobacteria* | *Gloeobacter* | 24.7 | 0.0 | 49.5 |
| *Deinococcus-Thermus* | *Thermus* | 21.5 | 49.9 | 6.9 |
| *Armatimonadetes* | *unidentified Armatimonadetes* | 10.1 | 29.6 | 9.7 |
| *Chloroflexi* | *unidentified Chloroflexi* | 8.3 | 0.0 | 16.7 |
| *Aquificae* | *Aquificaceae* | 3.3 | 7.9 | 1.7 |
| *Chlorobi* | *unidentified Chlorobi* | 3.0 | 6.7 | 1.5 |
| *Chlorobi* | *SM1B02* | 2.8 | 0.0 | 5.5 |
| *Chloroflexi* | *Chloroflexus* | 1.9 | 0.0 | 3.9 |
| *EM3* | *unidentified EM3* | 1.5 | 3.9 | 1.0 |
| *Chloroflexi* | *Roseiflexus* | 0.5 | 0.0 | 1.1 |
| IT *vs.* LT | | | | |
| Taxon[a] | Genus/Family[b] | Contr. (%)[c] | Avg.[d] IT (%) | Avg. LT |
| *Cyanobacteria* | *Gloeobacter* | 15.9 | 49.5 | 30.4 |
| *Chloroflexi* | *unidentified Chloroflexi* | 7.9 | 16.7 | 4.8 |
| *Chlorobi* | *SM1B02* | 6.0 | 5.5 | 13.4 |
| *Armatimonadetes* | *unidentified Armatimonadetes* | 4.6 | 9.7 | 0.6 |
| *Cyanobacteria* | *Pseudanabaena* | 4.3 | 0.0 | 8.5 |
| *Deinococcus-Thermus* | *Thermus* | 3.5 | 0.0 | 7.0 |
| *Acidobacteria* | *unidentified Acidobacteria* | 3.4 | 6.9 | 0.1 |
| *unknown Bacteria* | *unknown Bacteria* | 2.8 | 1.0 | 6.6 |
| *Cyanobacteria* | *Pseudanabaenaceae* | 2.3 | 0.0 | 4.7 |
| *Cyanobacteria* | *Synechococcus* | 2.1 | 0.0 | 4.2 |

[a]Phylum level.

[b]Genus was displayed; if genus was not available, family was displayed.

[c]Contribution of OTU to overall dissimilarity between groups.

[d]Average abudance of OTU in each group.

**Fig. 1.**

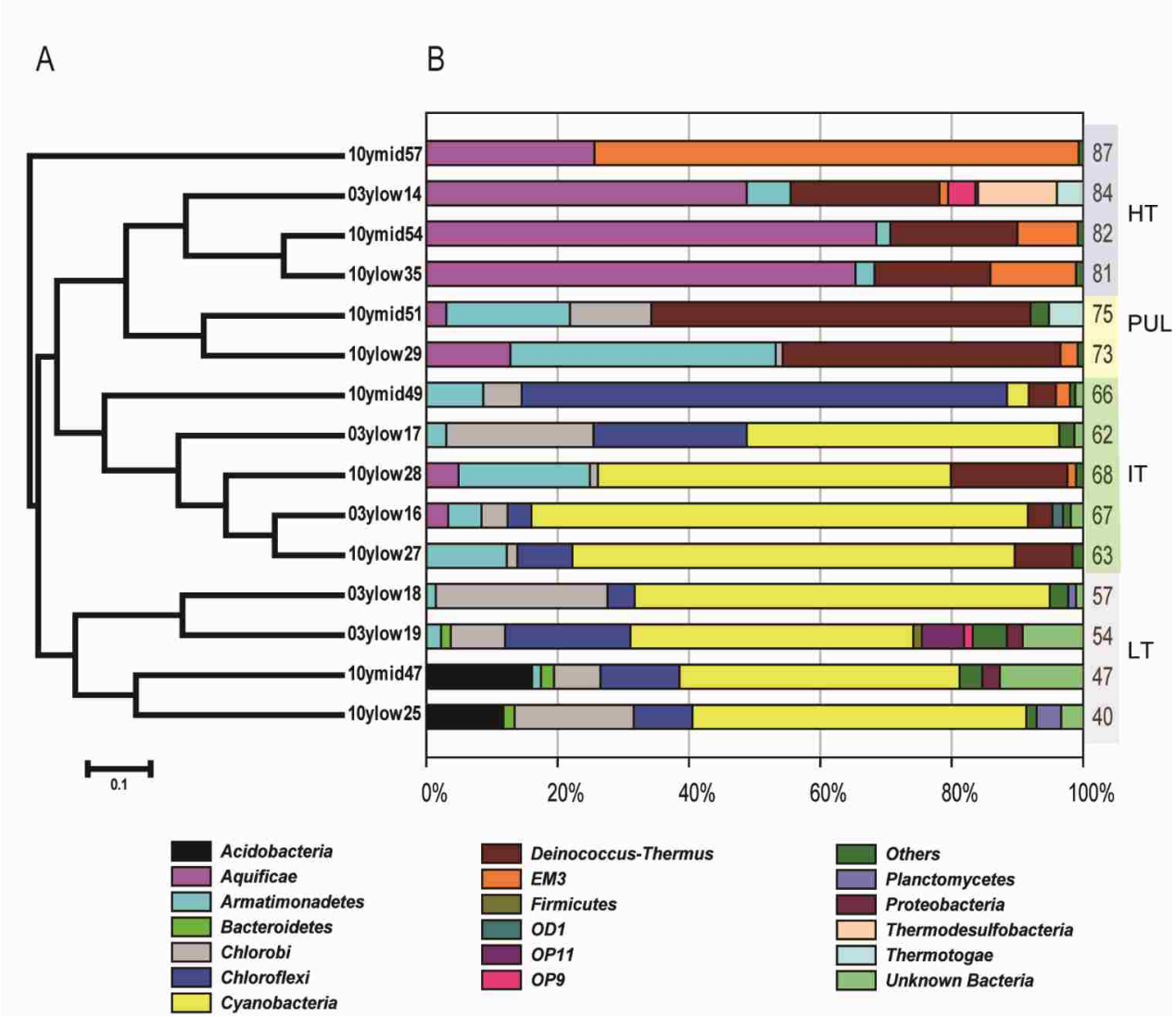

A
B
10ymid57
03ylow14
10ymid54
10ylow35
10ymid51
10ylow29
10ymid49
03ylow17
10ylow28
03ylow16
10ylow27
03ylow18
03ylow19
10ymid47
10ylow25
0.1
87
84
82
81
75
73
66
62
68
67
63
57
54
47
40
HT
PUL
IT
LT
0%
20%
40%
60%
80%
100%
Acidobacteria
Aquificae
Armatimonadetes
Bacteroidetes
Chlorobi
Chloroflexi
Cyanobacteria
Deinococcus-Thermus
EM3
Firmicutes
OD1
OP11
OP9
Others
Planctomycetes
Proteobacteria
Thermodesulfobacteria
Thermotogae
Unknown Bacteria

**Fig. 2.**

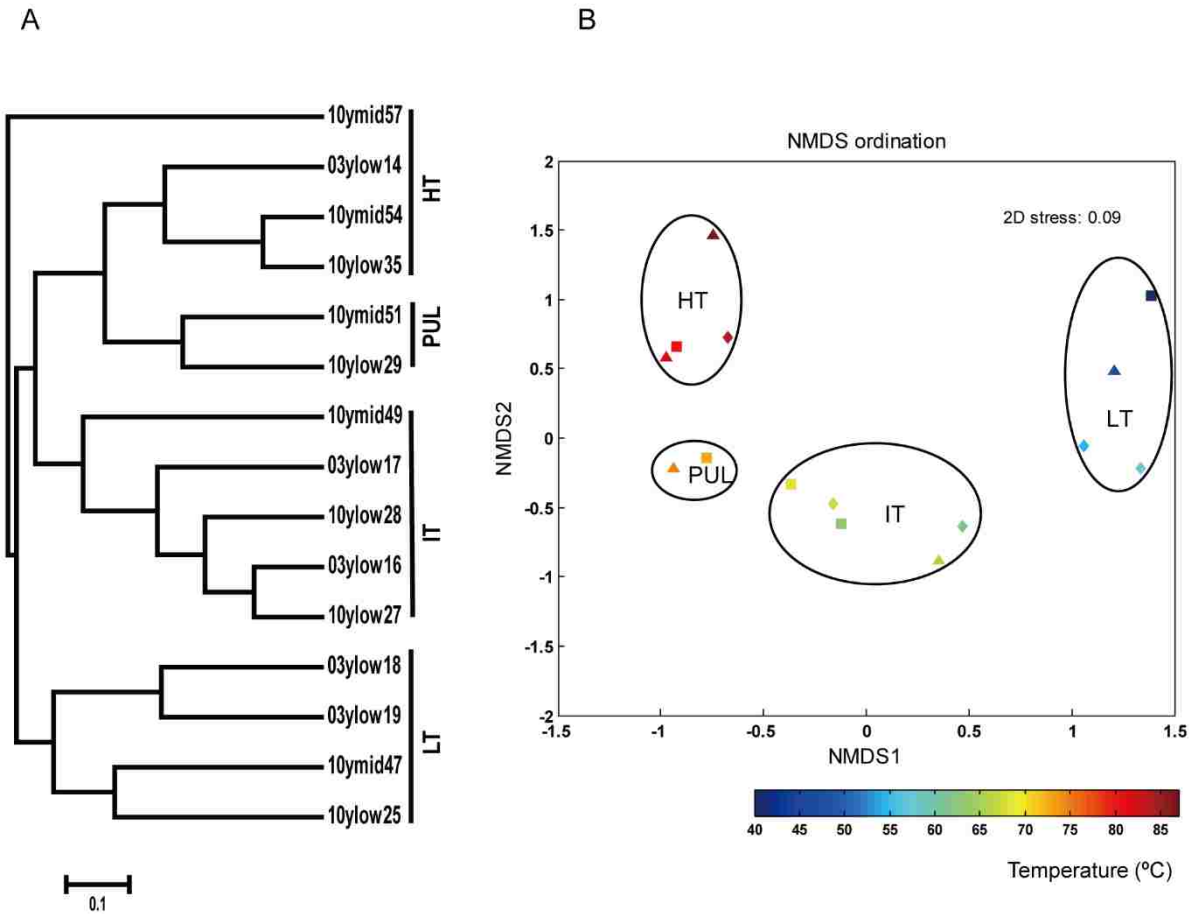

A
B
10ymid57
03ylow14
10ymid54
10ylow35
10ymid51
10ylow29
10ymid49
03ylow17
10ylow28
03ylow16
10ylow27
03ylow18
03ylow19
10ymid47
10ylow25
HT
PUL
IT
LT
0.1
NMDS ordination
2D stress: 0.09
NMDS1
NMDS2
Temperature (ºC)

**Fig. 3.**

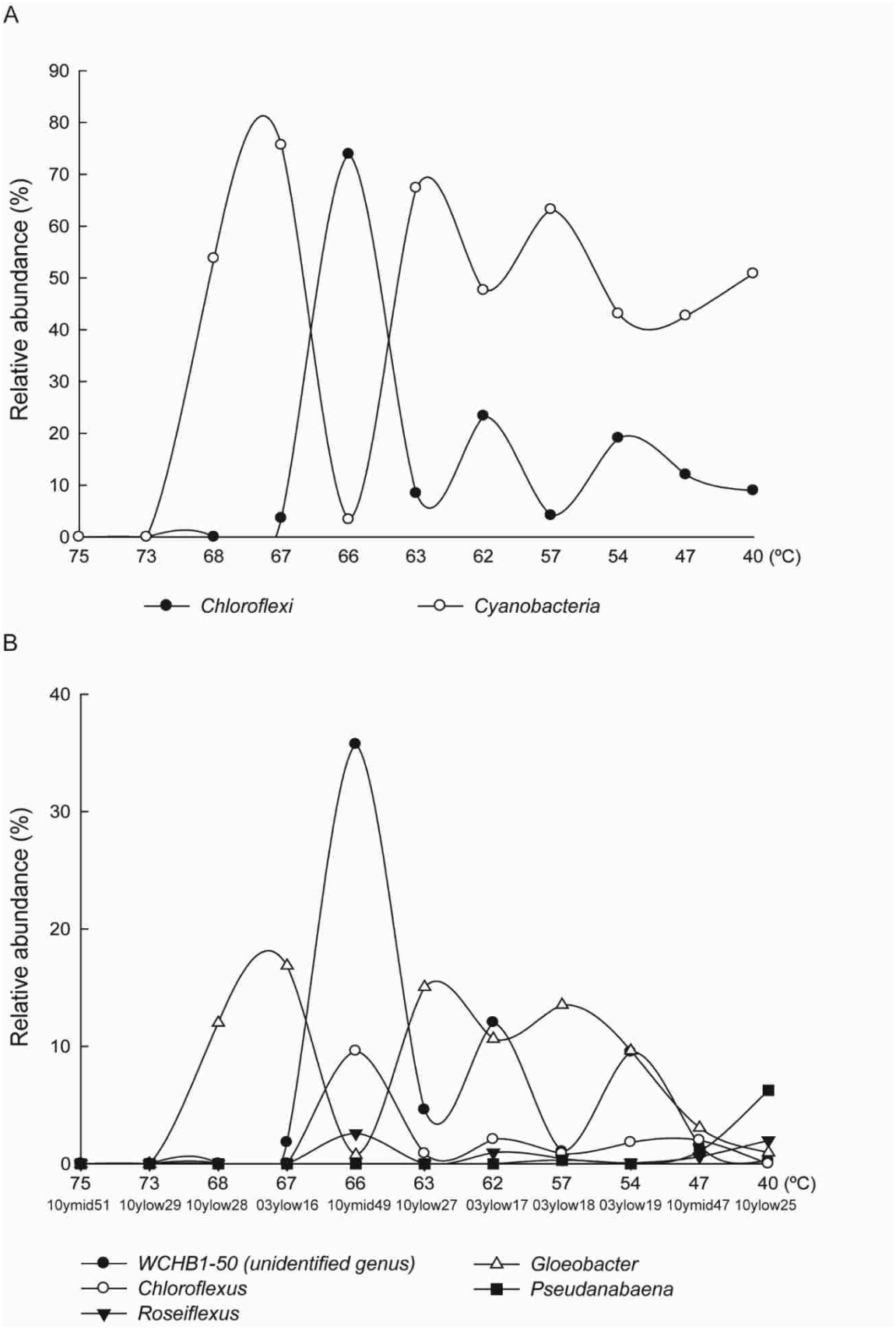

A
Relative abundance (%)
90
80
70
60
50
40
30
20
10
0
75 73 68 67 66 63 62 57 54 47 40 (ºC)
Chloroflexi
Cyanobacteria
B
Relative abundance (%)
40
30
20
10
0
75 73 68 67 66 63 62 57 54 47 40 (ºC)
10ymid51 10ylow29 10ylow28 03ylow16 10ymid49 10ylow27 03ylow17 03ylow18 03ylow19 10ymid47 10ylow25
WCHB1-50 (unidentified genus)
Chloroflexus
Roseiflexus
Gloeobacter
Pseudanabaena

**Table S1.** Temperature and in situ water chemistry parameters for sampling sites

| Sample ID | 03ylow14 | 03ylow16 | 03ylow17 | 03ylow18 | 03ylow19 | 10ylow25 | 10ylow27 | 10ylow28 |
|---|---|---|---|---|---|---|---|---|
| Spring name | Red Terrace | Red Terrace | Red Terrace | Red Terrace | Red Terrace | Bison Pool | Bison Pool | Bison Pool |
| Distance from source (m) | 7 | 18 | 23 | 28 | 31 | 27 | 18 | 14 |
| Temp (℃) | 84 | 67 | 62 | 57 | 54 | 40 | 63 | 68 |
| pH | 8.9 | 9 | 9 | 9.1 | 9.1 | 8.6 | 7.7 | 7.6 |
| Conductivity (μS $cm^{-1}$) | 1391 | 1431 | 1454 | 1481 | 1481 | 1460 | 1664 | 1670 |
| Hydrogen sulfide (mg $L^{-1}$) | 0.321 | 0.194 | 0.148 | 0.097 | 0.084 | 0.045 | 0.030 | 0 |
| Dissolved oxygen (mg $L^{-1}$) | NA | NA | NA | NA | NA | 4.8 | 1.3 | 1.1 |
| Al (ppm) | 0 | 0 | 0 | 0 | 0 | 0 | NA | NA |
| Alkalinity as $HCO_3^-$ (ppm) | 307 | 313 | 315 | 319 | 318 | 337 | NA | NA |
| As(III) (ppm) | 1 | 1 | 1 | 1 | 1 | NA | NA | NA |
| As(T) (ppm) | 1 | 1 | 1 | 1 | 1 | 1 | NA | NA |
| B (ppm) | 3 | 4 | 4 | 4 | 4 | 4 | NA | NA |
| Ba (ppm) | 0 | <0.0008 | 0 | <0.0008 | <0.0008 | BD | NA | NA |
| Be (ppm) | 0 | 0 | 0 | 0 | 0 | BD | NA | NA |
| Br (ppm) | 1 | 1 | 1 | 1 | 1 | 1 | NA | NA |
| Ca (ppm) | <0.4 | <0.4 | <0.4 | <0.4 | <0.4 | 0 | NA | NA |
| Cd (ppm) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | BD | NA | NA |
| Cl (ppm) | 266 | 262 | 256 | 266 | 263 | 189 | NA | NA |
| Co (ppm) | <0.002 | <0.002 | <0.002 | <0.002 | <0.002 | BD | NA | NA |
| Cr (ppm) | <0.002 | <0.002 | <0.002 | <0.002 | <0.002 | BD | NA | NA |
| Cu (ppm) | <0.003 | <0.003 | <0.003 | <0.003 | <0.003 | BD | NA | NA |
| F (ppm) | 26 | 28 | 28 | 29 | 29 | 23 | NA | NA |
| Fe(II) (ppm) | 0 | 0 | 0 | 0 | 0 | BD | NA | NA |
| Fe(T) (ppm) | 0 | 0 | 0 | 0 | 0 | BD | NA | NA |
| K (ppm) | 10 | 10 | 10 | 10 | 11 | 12 | NA | NA |
| Li (ppm) | 2 | 2 | 2 | 2 | 2 | NA | NA | NA |
| Mg (ppm) | <0.04 | <0.04 | <0.04 | <0.04 | <0.04 | BD | NA | NA |
| Mn (ppm) | 0 | <0.001 | <0.001 | <0.001 | <0.001 | BD | NA | NA |
| Mo (ppm) | 0 | 0 | 0 | 0 | 0 | BD | NA | NA |
| $NO_2^-$ (ppm) | 0 | 0 | 0 | 0 | 0 | BD | NA | NA |
| Na (ppm) | 305 | 302 | 281 | 302 | 309 | 301 | NA | NA |
| $NH_4^+$ (ppm) | 0 | 1 | 0 | 0 | 0 | 0 | NA | NA |
| Ni (ppm) | 0 | 0 | 0 | 0 | 0 | BD | NA | NA |
| $NO_3^-$ (ppm) | <0.1 | <0.1 | <0.1 | <0.1 | <0.1 | BD | NA | NA |
| Pb (ppm) | <0.008 | <0.008 | <0.008 | <0.008 | <0.008 | BD | NA | NA |
| $PO_4^{3-}$ (ppm) | <0.3 | <0.3 | <0.3 | <0.3 | <0.3 | BD | NA | NA |
| $S_2O_3^{2-}$ (ppm) | <0.1 | 0 | 0 | 0 | 0 | NA | NA | NA |
| Se (ppm) | <0.04 | <0.04 | <0.04 | <0.04 | <0.04 | BD | NA | NA |
| $SiO_4^{2-}$ (ppm) | 276 | 249 | 244 | 294 | 308 | 141 | NA | NA |
| $SO_4^{2-}$ (ppm) | 15 | 16 | 16 | 17 | 17 | 17 | NA | NA |
| Sr (ppm) | <0.0003 | <0.0003 | <0.0003 | <0.0003 | <0.0003 | BD | NA | NA |
| V (ppm) | <0.005 | <0.005 | <0.005 | <0.005 | <0.005 | BD | NA | NA |
| Zn (ppm) | <0.004 | <0.004 | 0 | <0.004 | <0.004 | BD | NA | NA |

| Sample ID | 10ylow29 | 10ylow35 | 10ymid47 | 10ymid49 | 10ymid51 | 10ymid54 | 10ymid57 |
|---|---|---|---|---|---|---|---|
| Spring name | Bison Pool | Bison Pool | Octopus Spring | Octopus Spring | Octopus Spring | Octopus Spring | Octopus Spring |
| Distance from source (m) | 11 | 2 | 51 | 25 | 18 | 9 | 3 |
| Temp (℃) | 73 | 81 | 47 | 66 | 75 | 82 | 87 |
| pH | 7.9 | 7.5 | 8.3 | 8 | 7.8 | 7.6 | 7.7 |
| Conductivity (μS $cm^{-1}$) | 1504 | 1452 | 1530 | 1532 | 1534 | 1525 | 1515 |
| Hydrogen sulfide (mg $L^{-1}$) | 0.020 | 0.155 | 0.000 | 0.045 | 0 | 0.025 | 0.010 |
| Dissolved oxygen (mg $L^{-1}$) | 1.0 | 0.3 | 3.3 | 0.9 | 0.9 | 0.6 | 0.3 |
| Al (ppm) | NA | NA | 0 | NA | NA | NA | NA |
| Alkalinity as $HCO_3^-$ (ppm) | NA | NA | 344 | NA | NA | NA | NA |
| As(III) (ppm) | NA | NA | NA | NA | NA | NA | NA |
| As(T) (ppm) | NA | NA | 1 | NA | NA | NA | NA |
| B (ppm) | NA | NA | 3 | NA | NA | NA | NA |
| Ba (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Be (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Br (ppm) | NA | NA | 1 | NA | NA | NA | NA |
| Ca (ppm) | NA | NA | 0 | NA | NA | NA | NA |
| Cd (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Cl (ppm) | NA | NA | 202 | NA | NA | NA | NA |
| Co (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Cr (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Cu (ppm) | NA | NA | BD | NA | NA | NA | NA |
| F (ppm) | NA | NA | 21 | NA | NA | NA | NA |
| Fe(II) (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Fe(T) (ppm) | NA | NA | BD | NA | NA | NA | NA |
| K (ppm) | NA | NA | 15 | NA | NA | NA | NA |
| Li (ppm) | NA | NA | NA | NA | NA | NA | NA |
| Mg (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Mn (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Mo (ppm) | NA | NA | BD | NA | NA | NA | NA |
| $NO_2^-$ (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Na (ppm) | NA | NA | 310 | NA | NA | NA | NA |
| $NH_4^+$ (ppm) | NA | NA | 0 | NA | NA | NA | NA |
| Ni (ppm) | NA | NA | BD | NA | NA | NA | NA |
| $NO_3^-$ (ppm) | NA | NA | 0 | NA | NA | NA | NA |
| Pb (ppm) | NA | NA | BD | NA | NA | NA | NA |
| $PO_4^{3-}$ (ppm) | NA | NA | BD | NA | NA | NA | NA |
| $S_2O_3^{2-}$ (ppm) | NA | NA | NA | NA | NA | NA | NA |
| Se (ppm) | NA | NA | BD | NA | NA | NA | NA |
| $SiO_4^{2-}$ (ppm) | NA | NA | 117 | NA | NA | NA | NA |
| $SO_4^{2-}$ (ppm) | NA | NA | 18 | NA | NA | NA | NA |
| Sr (ppm) | NA | NA | BD | NA | NA | NA | NA |
| V (ppm) | NA | NA | BD | NA | NA | NA | NA |
| Zn (ppm) | NA | NA | BD | NA | NA | NA | NA |

**Table S2. BIO-ENV results showing correlation between microbial community composition and each subset of environmental variables**

| | Size | Correlation coefficient |
|---|---|---|
| Temp | 1 | 0.8562 |
| Temp pH | 2 | 0.6446 |
| Temp pH $H_2S$ | 3 | 0.4308 |
| Temp pH $H_2S$ Conductivity | 4 | 0.2785 |

**Fig. S1.**

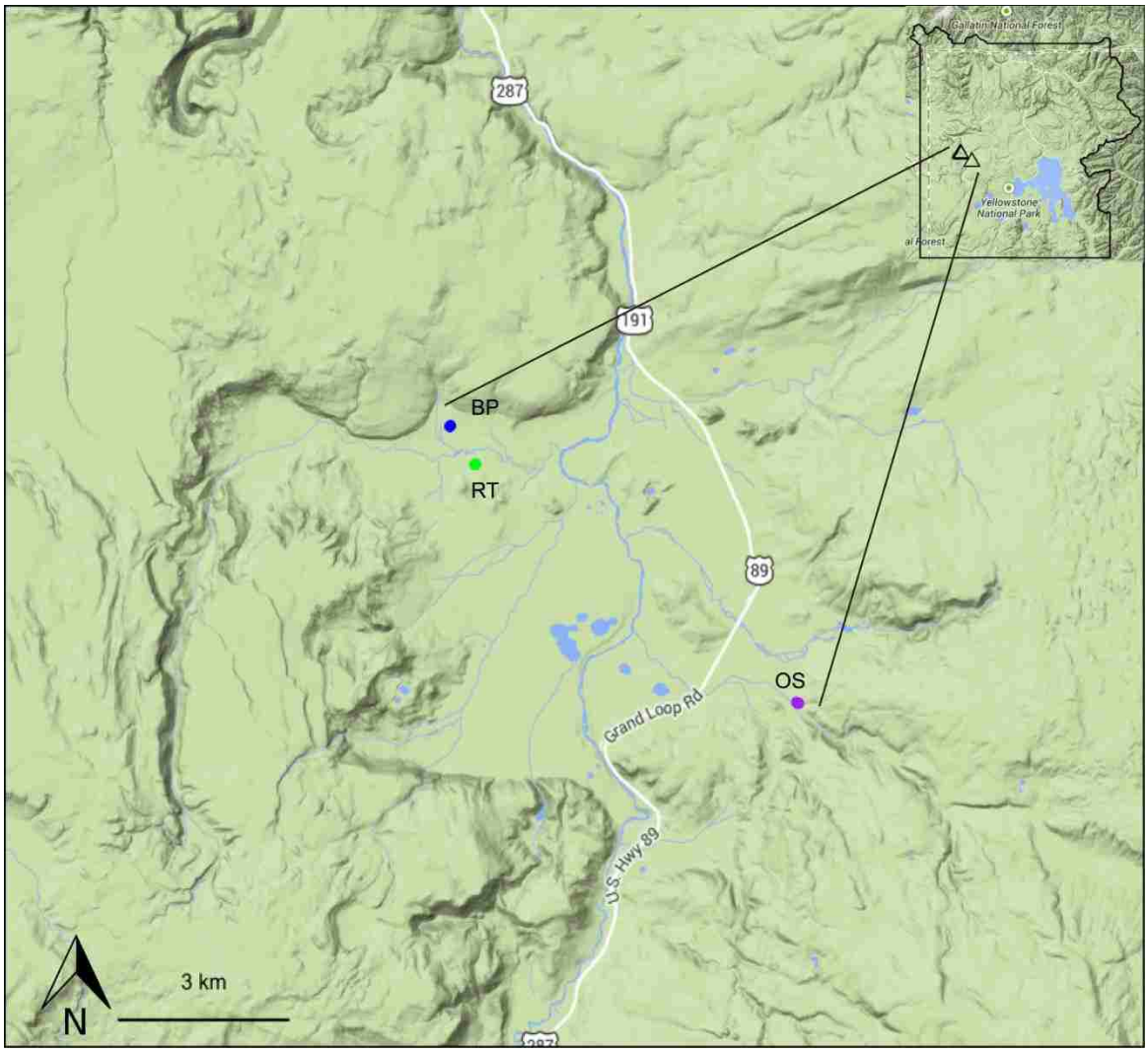

Gallatin National Forest
Yellowstone National Park
287
191
89
BP
RT
OS
Grand Loop Rd
U.S. Hwy 89
3 km
N

**Fig. S2.**

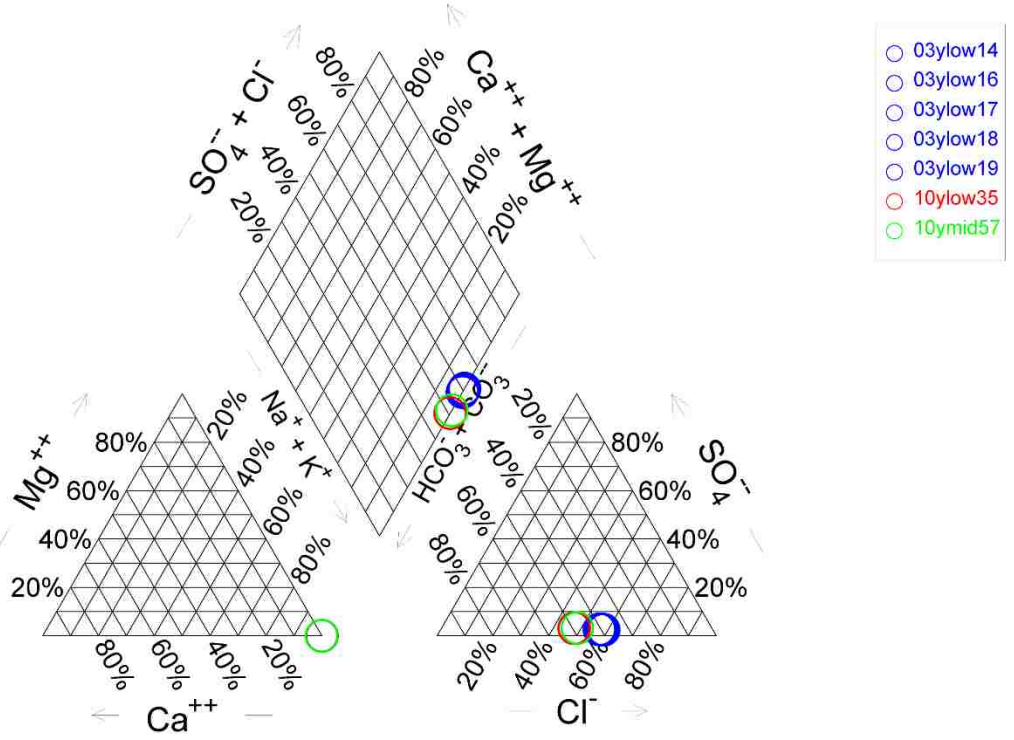

03ylow14
03ylow16
03ylow17
03ylow18
03ylow19
10ylow35
10ymid57
$SO_4^{--} + Cl^-$
$Ca^{++} + Mg^{++}$
$Na^+ + K^+$
$CO_3^{--} + HCO_3^-$
$Mg^{++}$
$SO_4^{--}$
$Ca^{++}$
$Cl^-$
20%
40%
60%
80%

**Fig. S3.**

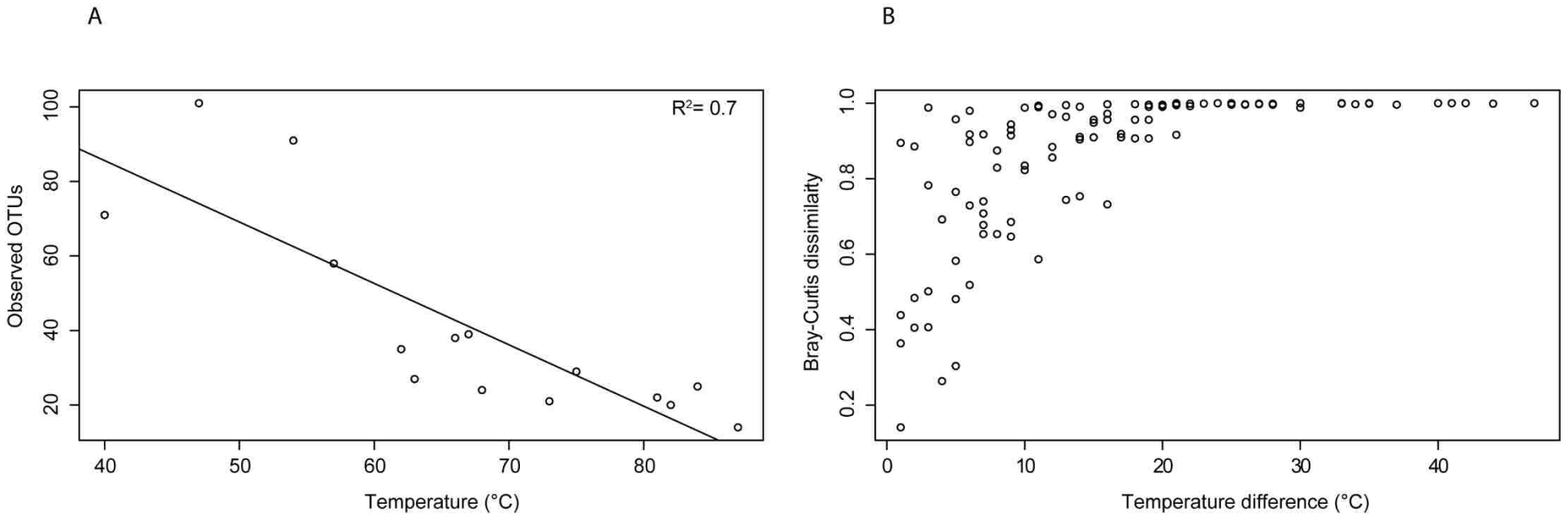
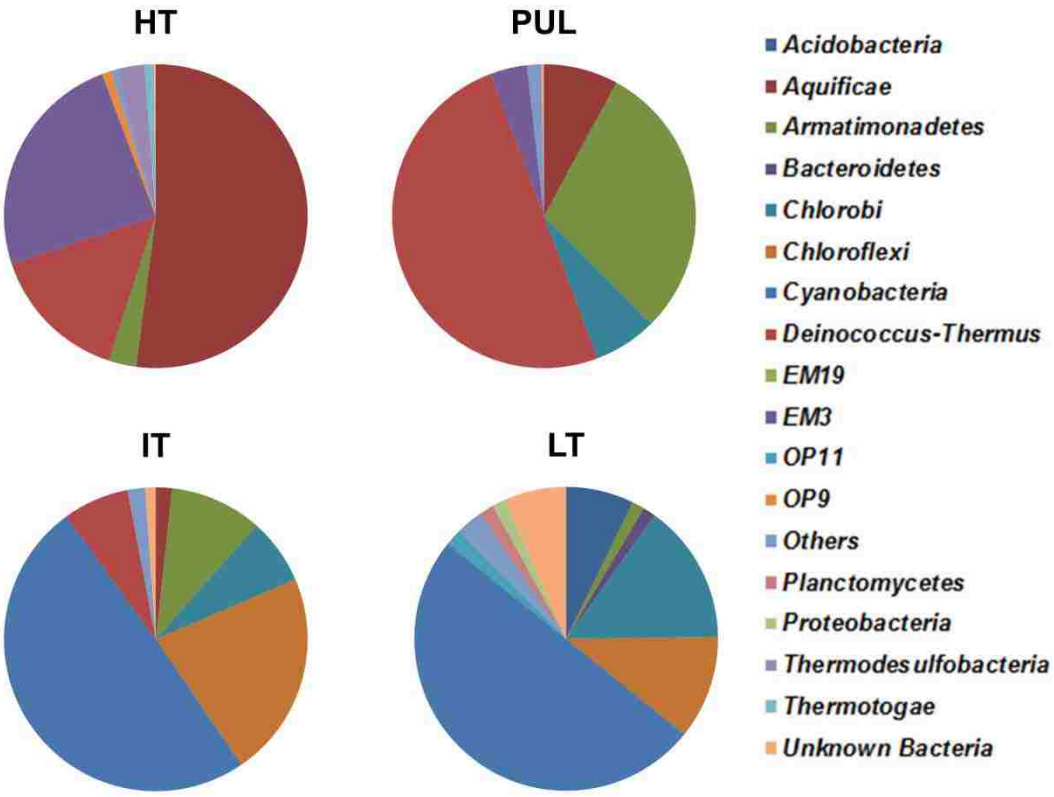

A
B
Observed OTUs
Temperature (°C)
R²= 0.7
Bray-Curtis dissimilarity
Temperature difference (°C)

**Fig. S4.**


HT
PUL
IT
LT
Acidobacteria
Aquificae
Armatimonadetes
Bacteroidetes
Chlorobi
Chloroflexi
Cyanobacteria
Deinococcus-Thermus
EM19
EM3
OP11
OP9
Others
Planctomycetes
Proteobacteria
Thermodesulfobacteria
Thermotogae
Unknown Bacteria

# Chapter 3

## Taxonomic and Functional Insights of a Low Temperature pH 4 Thermal Spring Microbial Community in Yellowstone National Park

Xiaoben Jiang ·Cristina Takacs-Vesbach

**Abstract**

In this study, we used 16S rRNA gene pyrosequencing to investigate microbial community structure sampled from four different temperature pH 4 thermal springs in Yellowstone National Park (YNP), USA. Our results suggest that temperature exerts a strong control on taxonomic assemblages. The taxonomic composition and metabolic functional potential of the microbial community from a low temperature (55 ℃) and pH 4 site in the Seven Mile Hole region was investigated using shotgun metagenome sequencing. The taxonomic classification, based on 372 Mbp of unassembled metagenomic reads and 16S rRNA barcoded amplicon sequencing, indicated that this community included a high proportion of Chloroflexi, Bacteroidetes, Proteobacteria and Firmicutes. Functional comparison with other YNP metagenomic datasets indicated that this community was enriched in the COG functions related to energy production and conversion, transcription, and carbohydrate transport. Analysis of genes involved in the nitrogen metabolism revealed the presence of assimilatory and dissimilatory nitrate reduction. We uncovered that genes involved in sulfur metabolism were mostly related to the reduction of sulfate to adenylylsulfate, sulfite and $H_2S$.

## Introduction

Yellowstone National Park (YNP) is one of the largest and most diverse hydrothermal areas on Earth and it harbors more than 10,000 thermal springs that are characterized by a broad range of tempeature (40-92 ℃), pH (1-10) and geochemical properties (Fournier 1989; Rye and Truesdell 2007). YNP thermal springs often have abundant and diverse electron donors (e.g., $H_2$, sulfide, $S^0$, thiosulfate and $Fe^{2+}$) and electron acceptors (e.g., dissolved $O_2$, $S^0$ and $Fe^{3+}$), which may serve to increase potential niches. As a consequence, thermal springs support microbial communities that comprise a diverse array of metabolisms including photoautotrophs, photoheterotrophs, chemolithotrophs and chemoorganotrophs (Amend and Shock 2001).

pH is a primary environmental factor that directly influences microbial community composition in thermal springs at the regional and global scale (Boyd et al. 2013; Boyd et al. 2010; Dequiedt et al. 2009a; Inskeep et al. 2013b; Song et al. 2013; Xie et al. 2014). While the range of pH in YNP thermal springs is broad (1-10), the majority of thermal springs in YNP can be classified into two categories by pH – the vapor-dominated acid system and the water-dominated circumneutral to alkaline system (Fournier 1989). The vapor-dominated springs, often with little discharge of liquid water, contain $H_2S$ that oxidizes to $H_2SO_4$ when it contacts air in perched pools of ground water. In contrast, the water-dominated systems discharge significant amounts of circumneutral or alkaline water enriched in chloride (Fournier 1989). As a consequence, there are few thermal features in Yellowstone with intermediate pH in the range of 4 to 5. For example, of the more than 7000 thermal featuers inventoried by the National Park Service (data available online at http://www.rcn.montana.edu/Default.aspx), only ~ 5 % of the entries

have a pH between 4 and 5. Microbial communities in the two pH systems common in YNP (acidic and circumneutral to alkaline) have been extensively studied and are known to each harboring distinct communities. For example, circumneutral to alkaline springs often include microbial communities that are dominated by members of the Aquificales, Chlorobi, Chloroflexi and Cyanobacteria (Inskeep et al. 2013b; Inskeep et al. 2010; Madigan 2003; Meyer-Dombard et al. 2005; Meyer-Dombard et al. 2011; Reysenbach et al. 2000b; Reysenbach et al. 1994; Ward et al. 1998a). Although the Archaea are present in circumneutral springs, they are estimated to be a lesser fraction of the total biomass (Inskeep et al. 2013b). Conversely, Archaea dominate the microbial communities in acidic vapor dominated springs (Brock et al. 1972; Jackson et al. 2001; Meyer-Dombard et al. 2005; Whitaker et al. 2003). A bimodal pH distribution among terrestrial themal springs has been noted for thermal areas worldwide (Brock 1971). For example, previous surveys of microbial communities from other geothermal hotspots around the world including El Tatio, Chile (Engel et al. 2013), Kamojang, Indonesia (Aditiawati et al. 2009), Nakabusa, Japan (Everroad et al. 2012), Odisha, India (Sen and Maiti 2014) and Tibet, China (Song et al. 2013; Wang et al. 2013) mainly forcused on acidic springs with pH below 3 or slight-neutural to alkaline springs with pH above 6. Thus, given the limited number of intermediate pH springs in YNP, little is known about their microbial ecology.

The goals of this study were to (1) investigate taxonomic profiles of four YNP springs with intermediate pH (pH 4.05 to 4.35) using 16S rRNA amplicon pryosequencing, and (2) characterize the metabolic potential of one of these sites, a low temperature spring that contained a novel microbial community. To our knowledge, the

present study is the first survey on microbial taxonomic and functionl diversity of pH 4 springs by applying a combined 16 rRNA amplicon and metagnomic sequencing approaches.

**Materials and Methods**

*Site description and Sample Collection*

Four geothermal areas (Norris, Mary Bay Area, Mud Kettles and Seven Mile Hole) in YNP were selected for field measurements and sample collection (Table S1 and Fig. 1). In the field, water for geochemical analysis was filtered through a 0.02 μm Sterivex filter using sterile 50-mL syringes and preserved as appropriate for the analysis to be performed (McCleskey et al. 2005). Macroscopically visible sediment, mat or filament samples were collected aseptically into 2-mL microcentrifuge tubes and preserved in sucrose lysis buffer (SLB; 20 mM EDTA, 200 mM NaCl, 0.75 M sucrose, 50 mM Tris-HCl, pH 9.0). Samples were stored at ambient temperature (~10 to 26 ℃) for up to 1 days before they were stored at -80 ℃. Previous experiments indicated that storage of samples in SLB without freezing did not lead to a loss of DNA or microbial diversity in the samples relative to samples immediately frozen in liquid nitrogen (Mitchell and Takacs-Vesbach 2008). Once in the laboratory, samples were stored at -80 ℃ until DNA extraction.

*Geochemical Analyses*

At each sampling location, water temperature and pH were measured using a Thermo Orion 290A+ meter and electrical conductivity was measured with a WTW meter with temperature correction. Hydrogen sulfide was measured using a hand held colorimeter (Hach DR/850) in the field. Geochemical analyses, including anions and

cations, were conducted using standard USGS methods (McCleskey et al. 2005).

*DNA extraction*

Total DNA was extracted from preserved samples following bead-beating disruption in a CTAB buffer (1 % CTAB, 0.75 M NaCl, 50 mM Tris pH 8, 10 mM EDTA) and subsequent phenol-chloroform purification steps as described in (Mitchell and Takacs-Vesbach 2008). Briefly, 2 volumes of 1 % CTAB buffer and proteinase K (final concentration 100 μg $mL^{-1}$) were added to the samples, which were then incubated for one hour at 60 ℃. SDS (final concentration 2 %) and 0.1 mm diameter Zirconia/Silica beads were added. Samples were bead beated for 45 s at 50 strokes per second. After incubating for one hour at 60 ℃, DNA was extracted once with an equal volume phenol:chloroform:isoamyl alcohol (25:24:1), followed by two extractions with an equal volume chloroform. Finally, the DNA was precipitated in 95 % ethanol, washed with 70 % ethanol, dried by speed-vac and reconstituted with 50 μl of filter-sterilized, autoclaved 10 mM Tris pH 8.0. DNA extracts were quantified using a Nanodrop ND-2000c spectrophotometer.

*16S rRNA pyrosequencing*

Barcoded amplicon pyrosequencing of 16S rRNA genes was performed as described previously (Van Horn et al. 2013). Briefly, DNA isolated from each sample was amplified using the universal bacterial primers 28F (5‘-GAGTTTGATCNTGGCTCAG-3’) and 519R (5‘-GTNTTACNGCGGCKGCTG-3’), and archaeal primers Arch349F (5’-GYGCASCAGKCGMGAAW-3’) and Arch806R (5’-GGACTACVSGGGTATCTAAT-3’) targeting the 16S rRNA genes as described previously (Colman et al. 2015; Rhoads et al. 2012). PCR was performed as follows:

initial cycle of 95 ℃ for 5 min, 30 cycles of 95 ℃ for 30 s, 54 ℃ for 40 s and 72 ℃ for 1 min, and a final elongation for 10 min at 72 ℃. Successful amplification was confirmed by agarose gel electrophoresis. Triplicate reaction mixtures per sample were combined and subsequently purified with an UltraClean™ GelSpin™ DNA Extraction Kit (MoBio Laboratories, Carlsbad, CA, USA). The purified DNA was quantified using a Nanodrop ND-2000c spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Amplicons from all samples were pooled at equimolar concentrations for pyrosequencing on a 454 GS FLX (454 Life Sciences, Branford, CT, USA) using Titanium reagents according to the manufacturer's protocol.

*16S rRNA pyrosequencing data processing*

Raw sequences obtained from pyrosequencing were denoised to correct for sequencing errors and remove low quality sequences and potential sequencing chimeras using AmpliconNoise (Quince et al. 2011) integrated into QIIME (Ver. 1.8.0, Caporaso et al. 2010b). Adapters, multiplex identifiers and primers were trimmed from denoised data. Operational taxonomic units (OTUs) were identified at the 97 % DNA similarity level using UCLUST (Edgar 2010) in QIIME. The most abundant sequence from each OTU was picked as a representative sequence and aligned using the PyNAST aligner (Caporaso et al. 2010a) and the Greengenes database (GG 13_5, DeSantis et al. 2006). Taxonomic assignments were made using the Ribosomal Databse Classifier program (Wang et al. 2007). All alpha and beta diversity estimates (e.g., rarefaction curves, Good's coverage) were performed with randomly drawn subsets of 800 sequences per sample to standardize for varying sequencing efforts across samples.

*Shotgun metagenome sequencing*

The Seven Mile Hole (04YSMH020, Table S1) sample was selected for further characterization by metagenomic sequencing because it contained a relatively unique microbial community compared to the other samples described here. Approximately 500 ng of DNA was used for library construction. Metagenome library preparation and sequencing was performed on one-half of a picotiter plate according to manufacturer's protocol on a 454 GS FLX Titanium platform (454 Life Sciences, Branford, CT, USA).

*Metagenomic assembly, comparison and analysis*

Metagenomic sequencing reads were quality-filtered and assembled using Newbler 2.6 (Margulies et al. 2005) using default settings. Contigs and singleton reads were submitted to the JGI IMG/M annotation pipeline (Markowitz et al. 2012). In order to determine the type of geobiological ecosystem in the low temperature pH 4 site, Clusters of Orthologous Group (COG, Tatusov et al. 2000) functions of the 04YSMH020 metagenome assembly were compared to those of other YNP metagenomes (Table S2, Inskeep et al. 2013b). For each metagenome, data were normalized by total number of COG functions detected, weighted by contig depth if assembly information was available. For unassembled singleton reads, a contig depth of one was assumed. A Bray-Curtis dissimilarity matrix was built based on the COG function abundance table. COG functions were subsequently classified into COG categories on IMG/M and a Bray-Curtis dissimilarity matrix based on the COG category abundance table was also constructed. The COG category and function dissimilarity matrices were subjected to hierarchical agglomerative cluster analysis. Principal Coordinates Analysis (PCoA) was performed with both COG categories and functions to confirm that observed patterns were stable regardless of hierarchical level. A two-way hierarchical clustering was done on COG

category abundance. All multivariate comparisons and ordinations were performed using R (Team 2011) statistical software with 'vegan' (Oksanen et al.) and 'cluster' (Maechler et al. 2013) packages. In addition, unassembled raw reads were also submitted for annotation on the metagenomics analysis server, MG-RAST (Meta Genome Rapid Annotation using Subsystem Technology, v3.3, Glass et al. 2010), using the default quality control pipeline. Microbial composition analyses were conducted via the MG-RAST best hit classification tool against the GenBank (NCBI-nr), M5NR (M5 non-redundant protein) and RefSeq databases using the minimum identity of 60 %, e-value cutoff of $10^{-5}$ and minimum alignment length of 50 bp.

*COG function enrichment analysis of the site 04YSMH020*

For an indication of what functions might be unusually prominent in a low temperature pH 4 site, we used the "Function Comparison" tool on IMG/M to compare the metagenome of 04SMYH020 against seven of the most similar metagenomic datasets detected through cluster analyses. The relative abundance of COG function was calculated based on normalized gene counts and expressed as D-scores, which were used to calculate the standard variation from the null hypothesis (i.e., relative gene counts in metagenome A = relative gene counts in metagenome B). For each comparison, the *P* value cutoff for significant D-scores was determined using a false discovery rate of 0.05.

*Metabolic mapping of energy metabolism*

After metagenome library adapters and low quality ends were trimmed from the unassembled raw reads, individual reads were annotated by BLASTX against to NCBI non-redundant (NR) protein database (Altschul et al. 1997) using the e-value cutoff of $10^{-5}$. The results of BLASTX were imported into MEtaGenome ANalyzer software

(MEGAN v4.70.4, Huson et al. 2007) and were taxonomically classified using the least common ancestor (LCA) algorithm based on the top 10 BLAST alignments for each read. Energy metabolism pathways were annotated using KEGG database (Ogata et al. 1999). The sequences in each pathway (oxidative phosphorylation, methane metabolism, nitrogen metabolism, carbon fixation pathways in prokaryotes, carbon fixation in photosynthetic organisms, sulfur metabolism and photosynthesis) were given taxonomical assignments at the phylum level. The pathways related to nitrogen and sulfur metabolism were mapped and reconstructed using KEGG identifiers.

*Sequence Data Submission*

The metagenome is publicly available on IMG/M (04YSMH020: IMG submission ID 13526) and MG-RAST (04YSMH020: ID 4523620.3). All raw sequencing data from this research are available through the NCBI Sequence Read Archive as SRP058441. The individual sff files from this study were assigned the accession numbers SRX1031281 - SRX1031284 under Bioproject PRJNA284196.

**Results**

*Taxonomic profiles* of 16S rRNA amplicon sequencing

A total of 6,432 bacterial 16S rRNA gene sequences were obtained from the four pH 4 sites after denoising and removing low quality or chimeric sequences. We solely focused on diversity of the Bacteria, because no archaeal 16S rRNA gene sequences were amplified with the Arch349F and Arch806R primers. The Bacterial 16S rRNA gene sequences clustered into 226 OTUs at 97 % DNA similarity and a majority of rarefaction curves approached saturation (Fig. S2). Good's coverages (Good 1953), which provide the estimate of sampling completeness, ranged from 91.1 % to 98.5 %, with an average of

95.9 %. The majority of bacterial phyla in the two high temperature sites (Table S1), 03YNOR021 and 03YMRY047, were dominated by phototrophic bacteria. For example, Cyanobacteria, Proteobacteria and Chloroflexi were the predominant bacterial phyla within 03YNOR021 (38.5 %, 23.4 % and 7.9 % of total 16S rRNA gene sequences, respectively, Fig. 5A), whereas Cyanobacteria, Chlorobi and Chloroflexi were the three most abundant bacterial groups within 03YMRY047 (82.5 %, 7.9 % and 5 % of total 16S rRNA gene sequences, respectively, Fig. 5A). In contrast, 03YMKL049 was comprised mostly of Aquificae (93.5 % of total 16S rRNA gene sequences), while 04YSMH020 was dominated by Armatimonadetes, Chloroflexi and Bacteroidetes (52.8 %, 18.6 % and 18 % of total 16S rRNA gene sequences, respectively, Fig. 5A).

*Metagenome sequencing, coverage and overview of microbial community groups*

Metagenome sequencing generated 848,583 reads, with 438 bp mean length, totaling 372 Mbp, for 04YSMH020 (Table 1). Assembly of this metagenomic sequence dataset yielded 19,346 contigs. An N50 contig length of 4,303 bp was obtained from 04YSMH020.

Hierarchical agglomerative cluster analyses based on COG functional categories and COG functions indicated that the 21 YNP metagenomes compared here were clustered into three distinct groups that could be characterized by their dominant members as: (1) archaeal communities; (2) Aquificales communities; and (3) phototrophic communities (Figs. 3, S1). The grouping detected through hierarchical cluster analyses was also confirmed by PCoA ordination based on COG functional categories (Fig. 4A) and on COG functions (Fig. 4B).

*Taxonomic profiles of metagenomics*

In this study, rarefaction curves were performed based on the taxonomic information retrieved from annotation results on MG-RAST, suggesting a majority of individual genomes were sampled and covered in the metagenome (Fig. 2). Overall community structure analysis performed with the M5NR (M5 non-redundant protein) database on MG-RAST indicated that 04YSMH020 was dominated by Bacteria (81.04 %, Table S3). The remaining sequences from 04YSMH020 matched with Archaea (6.36 %), Eukaryota (0.21 %), and unclassified sequences, or were unassigned (12.38 %). A closer look at the metagenome of 04YSMH020 revealed the taxonomic distribution of numerically abundant microbial phyla derived from the metagenome of 04YSMH020, indicating that Chloroflexi (~17.8 %), Bacteroidetes (~17.7 %), Proteobacteria (~13.5 %) and Firmicutes (~12.0 %) were the top four most abundant phyla, according to GenBank, M5NR and RefSeq database (Fig. 6). The class Ktedonobacteria was the dominant class (~58.0 %) in phylum Chloroflexi based on all three databases. All of the sequences affiliated with the order Ktedonobacterales were assigned to *Ktedonobacter racemifer* (Fig. 7A). Sphingobacteria was the dominant class in the phylum Bacteroidetes (~52.4 %, Fig. 6), with all sequences related to the order Sphingobacteriales (Fig. 7B), that included *Chitinophaga pinensis* (~73.2 %, Fig. 7B). Sequences assigned to Deltaproteobacteria were dominated by two orders, Myxococcales (~38.6 %) and Desulfuromonadales (~32.0 %, Fig. 7C). At the class level, Clostridia accounted for ~ 66.8% of all Firmicutes reads, followed by Bacilli (~28.8 %), Negativicutes (~4.2 %) and Erysipelotrichi (~0.3 %). Within the class Clostridia, a large proportion of sequences were assigned to Clostridiales (~58.7 %) represented by a variety of genera (Fig. 7D).

*Gene functions enriched in the metagenome of 04YSMH020.*

The metagenome of 04YSMH020 provided the first information on the functional capabilities of microbial community in a low temperature pH 4 site. Cluster analyses showed that the functional profile of the 04YSMH020 metagenome was similar to metagenomes of phototrophic communities (Figs. 3, 4, S1). A total of 27 COG functions were significantly overrepresented in the 04YSMH020 dataset in at least six of the seven comparisons (Table S4). "Energy production and conversion", "Transcription" and "Carbohydrate metabolism and transport" were the top three most abundant COG categories among these functions (25.9 %, 18.5 % and 11.1 %, respectively) and were dominant among those with highest enrichment D-scores (Fig. 8).

*Energy metabolism mapping*

Clearly, the functional assignment of the unassembled 04YSMH020 metagenomic dataset provided information about possible functions in this community. A total of 14,957 reads were assigned to energy metabolism using BLASTX against the NCBI-nr database, and the majority of reads were related to the domain Bacteria (~ 92 %). Among the bacterial reads, most were mapped to Bacteroidetes, Chloroflexi, Firmicutes, Proteobacteria and Planctomycetes involved in diverse pathways, such as oxidative phosphorylation, methane metabolism, nitrogen metabolism, carbon fixation pathways in prokaryotes, carbon fixation in photosynthetic organisms, sulfur metabolism and photosynthesis (Fig. 9). A total of 2,239 and 638 reads were mapped to nitrogen and sulfur metabolism, respectively. Major KEGG function categories and unique hits assigned to each category are listed in Tables S5 and S6. Genes involved in nitrate reduction were among the abundant categories associated nitrogen metabolism (Table

S5). Genes encoding sulfate adenylyltransferase, cysteine synthase and sulfite reductase were also highly enriched in the metagenome of 04YSMH020 (Table S6).

**Discussion**

In this study, we employed several approaches and databases to examine microbial diversity. First, 16S rRNA sequencing results suggest that temperature exerts a strong control on taxonomic assemblages. Diverse community assemblages were recovered at sites with different temperatures (Fig. 5A), although the pH values were close (Table S1). For example, the Aquificae were dominant in the site 03YMKL049 (Fig. 5A), one of the high temperature sites of pH 4 (Table S1), which is consistent with previous studies in YNP. The Aquificae are considered thermophilic microorganisms and are one of the basal lineages of Bacteria domain (Takacs-Vesbach et al. 2013). The phylum Aquificae normally predominates in high temperature springs with temperature above 70 ℃ (Inskeep et al. 2013b), or high temperature reaches within a spring where photosynthesis is limited (Cole et al. 2013; Everroad et al. 2012; Hall et al. 2008; Huber and Stetter 2001). In contrast, diverse bacterial groups (e.g., Armatimonadetes, Bacteroidetes and Chloroflexi) were uncovered from the low temperature site 04YSMH020 (Fig. 5A). Second, the metagenomic analysis based on the M5NR database indicated that the relative abundance of Bacteroidetes and Chloroflexi in 04YSMH020 was consistent with their proportions retrieved in 16S rRNA gene sequencing (Fig. 5B). The annotations based on M5NR database indicated that Archaea were present in very low abundance in 04YSMH020 (Figs. 5C), which could explain why we were unable to amplify archaeal 16S rRNA gene using archaeal specific primers. Interestingly, compared to a large number of sequences retrieved from 16S rRNA gene sequencing,

Armatimonadetes sequences were not detected from the metagenomic annotation, although similar relative abundance of Chloroflexi and Bacteroidetes was found in both metagenomic and 16S rRNA sequencing data (18.6% and 18.1% of total of bacterial sequences according to M5NR, respectively; 18.6% and 18% of total 16S rRNA gene sequences, respectively, Fig. 5B). Instead, additional prevalent bacterial groups detected from 04YSMH020 were affiliated with Acidobacteria, Actinobacteria, Firmicutes, Planctomycetes and Proteobacteria (Fig. 5B). The phylum Armatimonadetes is a new described phylum and is estimated to comprise 12 groups, occurring in a variety of environments (Dunfield et al. 2012). However, only several strains of Armatimonadetes have been isolated to date (Dunfield et al. 2012). The lack of isolates and classification can lead to inaccurate annotations and incongruent results. Unexpectedly, our 16S rRNA based results showed that phototrophic organisms (e.g., Cyanobacteria) were present in 03YNOR021 and 03YMRY047 (Fig. 5A), where temperatures were both above 75 ℃ (Table. S1). Generally, the temperature limit for photosynthesis is considered ~75 ℃ (Rothschild and Mancinelli 2001). Samples collected from site 03YNOR021 and 03YMRY047 belong to sediment layers, where dead cells from allochthonous phototrophic microorganisms might deposit and accumulate.

We compared the metagenome of a pH 4 site, 04YSMH020, to 20 YNP metagenomes that were publicly available through IMG/M (Inskeep et al. 2013b). The observation of three main geobiological ecosystem types is in agreement with previously metagenomic surveys on YNP thermal springs (Inskeep et al. 2013a; Inskeep et al. 2013b; Klatt et al. 2013; Takacs-Vesbach et al. 2013). Branch lengths appeared to be short in the clustering using the highest COG categories, suggesting overall

compositional similarity among the metagenomes (Fig. 3A). In contrast, branch lengths in the clustering using lowest COG functions were longer, revealing the specificities among metagenomes (Fig. 3B). Most interestingly, site 04YSMH020 and the archaeal sites did not fall within the same group, although pH values across these sites were close (Tables S1, S2). Instead, site 04YSMH020, representing the low temperature pH 4 habitat, appeared to cluster together with phototrophic sites at both COG hierarchical levels (Fig. 3), despite large differences in pH (Tables S1, S2).

Of significance is that, unlike the other phototrophic communities that were dominated by one or two phyla such as Cyanobacteria and Chloroflexi (Inskeep et al. 2013b), site 04YSMH020 was dominated by more diverse bacterial groups at the phylum level (Fig. 6), which may lead to *in situ* functional distinction and metabolic diversification. A close look into the deep taxonomic distribution of 04YSMH020 can provide the linkage between the taxonomic diversity and metabolic diversity. *Ktedonobacter racemifer* was the most frequently encountered species in the phylum Chloroflexi (3,429 assigned reads using GenBank database, Fig. 7A). *K. racemifer* is a mesophilic aerobic heterotroph often found in soils (Cavaletti et al. 2006; Weber and King 2010) and the only isolated strain DSM 44963 produces $H_2S$ but cannot reduce nitrate *in vitro* (Cavaletti et al. 2006; Chang et al. 2011). Despite being assigned to the strain DSM 44963, the *K. racemifer* uncovered from our environmental sample displayed differences in functional gene content. Apart from genes (e.g., *cys*C) related to sulfur metabolism (Table S6), the assimilatory nitrate reduction gene (e.g., *nas*A) was detected in association with the *K. racemifer* from site 04YSMH020, suggesting the *in situ* metabolic versatility of *K. racemifer*. Not surprisingly, several common genera within

class Chloroflexi were also detected in site 04YSMH020, which were previously implicated in the primary production of phototrophic communities, including the following bacteria previously isolated from YNP thermal springs or detected through culture-independent approaches: *Roseiflexus* spp. and *Chloroflexus* spp. (Pierson and Castenholz 1974; Ward et al. 1990). The presence of order Sphingobacteriales within class Sphingobacteriia (phylum Bacteroidetes, Figs. 6, 7B) is not surprising, as this order has also been uncovered from alkaline thermal springs in Africa using 16S rRNA amplicon pyrosequencing (Tekere et al. 2011). Of the order Sphingobacteriales, the most abundant species found in site 04YSMH020, *Chitinophaga pinensis*, can degrade the chitin and grow at pH between 4 to 10 (Del Rio et al. 2010; Sangkhobol and Skerman 1981). Temperature and pH in site 04YSMH020 are suitable for *C. pinensis* growth. Additionally, the thick clumpy mat structure in site 04YSMH020 (Table S1) can provide additional niches and available organic compounds for the growth of *C. pinensis*. Within Deltaproteobacteria, the order Myxococcales was the predominant order (Fig. 7C) with a few of sequences related to *Anaeromyxobacter*. Members of genus *Anaeromyxobacter* are found to utilize a wide spectrum of electron donors and acceptors (Sanford et al. 2002; Wu et al. 2006). For example, *Anaeromyxobacter dehalogenans* Strain 2CP-C detected in this metagenome possesses two Ni-Fe-type hydrogenases using inorganic (e.g., hydrogen) or organic (e.g., acetate) as electronic donors and it can also utilize Fe (III), U (VI), nitrate, nitrite, fumarate or oxygen as electron acceptors. Bacterial respiratory versatility may be fueled by the continuous supply of electron donors from thermal spring water and diverse electron acceptors from the spring sediments. Sequences assigned to the order Clostridiales within class Clostridia (Fig. 7D) are commonly found

in global thermal springs (Almarsdottir et al. 2010; Fardeau et al. 2010; Meyer-Dombard et al. 2005; Sayeh et al. 2010; Tekere et al. 2011). Genus *Thermaerobacter*, one of the dominant genera within the Clostridiales (Fig. 7D), represents aerobic and thermophilic heterotrophs previously detected in both acidic and alkaline terrestrial thermal springs through 16S rRNA analyses (Burgess et al. 2012; Inagaki et al. 2001), growing in the presence of organic substrates such as cellulose, chitin, sugars and amino acids (Spanevello et al. 2002). Overall, despite the short length of reads for the unassembled metagenome and the limitation of public databases that are biased toward readily cultured representatives, the results of the unassembled 04YSMH020 metagenome reveal a general pattern of the taxonomic profile, which is consistent with part of our 16S rRNA amplicon sequencing results (Figs. 5A, 6). The microbial community of 04YSMH020 supports a diverse array of phototrophic, heterotrophic and chemotrophic microorganisms, which can contribute to the metabolic potential and diversity in a low temperature pH 4 environment.

The metagenome of 04YSMH020 provides the first picture of the functional profile in a low temperature pH 4 site of YNP. Additionally, comparison of 04YSMH020 metagenomic data to those from other phototrophic sites offers information on prominent functional capabilities of a low temperature pH 4 microbial community. Among these overrepresented functions, many were responsible for various energy production and conversion such as redox reactions and electron transports. For example, some enriched COG functions (Table S4), such as coenzyme F420 (COG 2141), heme Cu oxidase (COG1622) and carbon monoxide dehydrogenase (COG1529, COG2080 and COG1319) and proteins (COG3794, COG0723) involved in electron transportation can be related

directly to the microbial dynamics and metabolic potentials. For example, COG 2141 associated with coenzyme F420-dependent 5,10-methylenetetrahydromethanopterin reductase belongs to the family of oxidoreductases responsible for the redox reaction in many Actinobacteria and methanogenic Archaea (Deppenmeier 2002). Heme-copper-type oxidases (COG 1622) representing the terminal energy-transfer enzymes of respiratory chains play a significant role in aerobic metabolism (Garc á-Horsman et al. 1994a). In addition, housekeeping functions associated with transcription were well represented in 04YSMH020 (COG1595, COG1522, COG1846, COG0789 and COG5662), suggesting that the microbial community has the potential to reproduce quickly. Among the overrepresented functions, we observed many relevant to carbohydrate metabolism and transport. For instance, COG2814 belongs to the family of "arabinose efflux permease". Proteins of this COG function belong to the major facilitator superfamily (MFS) that can transport a diverse array of substrates, such as, amino acids, drugs, ions and sugars across the membrane (Law et al. 2008). Another overrepresented function COG2271, which is responsible for transport sugar derived from the environment, is also affiliated with MFS. The enrichment of functions involved in transportation in site 04YSMH020 suggests higher microbial activities and metabolic dynamics than the other phototrophic sites. More intermediates produced by higher microbial activities may expand available niches in the community and increase the matter and energy flux in the whole system, thus facilitating the observed increase of microbial diversity and competition. For example, the functions (e.g., COG 1131 and COG 0841) associated with multidrug resistance with high abundance are known to export antibiotics and toxic molecules (Piddock 2006). Bacteria bearing these functions

can defend against toxic compounds produced by competitors. The functions (COG1629 and COG1914) related to inorganic ion transport and metabolism, were significantly abundant in 04YSMH020. Bacteria in this site are expected to encounter heavy metals and possess the genes involved in heavy metal transport (e.g., iron transporters), because site 04YSMH020 has considerable concentration of iron (Table S1).

Pathways involved in nitrogen and sulfur metabolism are known to be important in thermal springs, where non $O_2$ electronic acceptors such as arsenate, $CO_2$, element sulfur, ferric iron, nitrate, sulfate and thiosulfate normally play more important roles (Hall et al. 2008; Inskeep et al. 2010; Jiménez et al. 2012). Approximately 81.1% of genes detected in association with nitrogen metabolism were related to Bacteroidetes, Chloroflexi, Firmicutes, Proteobacteria and Planctomycetes. Genes coding for dissimilatory nitrate reductases (e.g., *nar*G, *nar*H, *nar*I, *nar*J) and nitric oxide reductases (e.g., *nor*B, *nor*C, *nor*D, *nor*Q) were prevalent in site 04YSMH020 (Table S5, Fig.S3). The presence of genes coding for these two types of enzymes have been previously identified in a high temperature and acidic system of YNP (e.g., Joseph's Coat hot spring, Inskeep et al. 2010). Additionally, the gene coding for nitrite reductase (*nir*K) was also uncovered (Table S5, Fig. S3), which is an indispensable step of the dissimilatory nitrate reduction pathway. According to the models of dissimilatory nitrate reduction in bacteria (Gonzalez et al. 2006; Moreno-Vivián et al. 1999; Richardson et al. 2001), a nitrite reductase (*nir*K or *nir*S) is the requisite for producing NO, which serves as a substrate for nitric oxide reductase to produce $N_2O$. The sequences related to gene *nos*Z coding for nitrous oxide reductase (associated with class Aquificae, Bacteroidetes, Flavobacteriia, Ignavibacteriae, Sphingobacteriia and Thermomicrobia) are important for the last step of

denitrification, which converts $N_2O$ to $N_2$ (Table S5, Fig. S3). It has been suggested that the *in situ* process of denitrification largely depends on the environmental factors such as temperature, pH, oxygen, nitrate availability and organic matter content (Jiménez et al. 2012). The discovery of all the important genes associated with denitrification pathways has important implications for the N cycle in low temperature pH 4 sites. Compared to high denitrification activities, the absence of *nif*K, a gene involved in the synthesis of molybdenum dependent nitrogenase, suggests that nitrogen fixation is not an important metabolic pathway in 04YSMH20 (Dos Santos et al. 2012). The lack of detection of genes (e.g., *amo*A) coding for ammonium monooxygenase is consistent with the previous report in YNP thermal springs (Inskeep et al. 2010), suggesting that this may not be a predominant metabolic process in this site. Based on the taxonomic composition of the assigned reads, assimilatory reduction of nitrate was performed mostly by bacterial classes associated with Acidobacteria, Actinobacteria and Chloroflexi, while dissimilatory reduction of nitrate was carried out by microorganisms related to Actinobacteria, Betaproteobacteria, Deltaproteobacteria, Firmicutes, Nitrospirae and Archaea (Table S5, Fig S3).

Most of the genes involved in the sulfur metabolism were related to the conversion of sulfate into adenylylsulfate and to the subsequent production of sulfite and $H_2S$ (Table S6, Fig. S4), similar to what has been reported previously in other thermal springs (Jiménez et al. 2012). Genes responsible for cysteine synthase A (*cys*K) and B (*cys*M) are implicated in formation of adenylylsulfate (Table S6, Fig. S4). The environmental *apr*A and *apr*B sequences coding for adenosine-5'-phosphosulfate (APS) reductase (Apr) exhibited closest matches to members of Betaproteobacteria and

Thermoprotei (Table S6, Fig. S4). Based on the current model of dissimilatory sulfate reduction and sulfur oxidation in prokaryotes, Apr is a pivotal enzyme. During the process of sulfate reduction, the function of Apr is to convert APS to sulfite, once sulfate is activated to APS by ATP-sulfurylase at the expense of ATP. Sulfite is subsequently reduced to sulfide by dissimilatory sulfite reductases (DSRs, Meyer and Kuever 2008). The alpha subunits of Apr enzymes is considered to be ubiquitous in all known sulfate reducing and most of sulfur oxidizing prokaryotes (Meyer and Kuever 2008). For example, environmental *apr*A reads found in site 04YSMH020 showed excellent identity to those annotated in the *Thiobacillus plumbophilus* and *Caldivirga maquilingensis* genomes (e-value < $10^{-35}$, Table S6, Fig. S4). *Thiobacillus plumbophilus* requires $H_2S$ as an electron donor (Drobner et al. 1992), whereas *Caldivirga maquilingensis* respires sulfur, thiosulfate or sulfate (Itoh et al. 1999). Genes coding for the reduction of adenylylsulfate to sulfite (e.g., *apr*A, *apr*B, *cys*H) and the subsequent reduction of sulfite to $H_2S$ (e.g., *cys*I, *sir*) observed in site 04YSMH020 suggest that sulfate and sulfite reduction pathways are dominant processes in the environment studied here. However, enzymes such as sulfite oxidase and sulfide quinone oxidoreductase were not detected in our dataset, either due to not enough sequencing depth was accomplished or because they may be absent or be present in very low abundance.

**Conclusions**

In the present study, microbial community structure and diversity in pH 4 springs were investigated using 16S rRNA pyrosequencing. In doing so, we provided the first information on the microbial communities of pH 4 springs with different temperatures, which indicated that temperature was an important factor in controlling microbial

assemblages. Additionally, we assessed the functional profiles of the microbial community in a low temperature pH 4 spring that was previously unexplored, using shotgun metagenome sequencing. Functional cluster analyses revealed that this unexplored geobiological ecosystem type belonged to phototrophic communities. However, taxonomically, this spring community exhibited the diverse and atypical community composition, which was very different from other phototrophic communities. Apart from Chloroflexi that were commonly found in phototrophic communities, Bacteroidetes, Proteobacteria and Firmicutes were abundant in this spring. The taxonomic diversity resulted in metabolic diversity (e.g., chemotrophs, heterotrophs), as observed here. Compared to other YNP phototrophic metagenomes, the metagenome of 04YSMH020 indicated that the enrichment functions were involved in energy production and conversion, transcription and carbohydrate transport. The identification of genes coding for nitrogen and sulfur cycling revealed that the microbial population was involved in assimilatory and dissimilatory reduction of nitrate, and conversion of sulfate into adenylylsulfate, sulfite and $H_2S$. Microbial communities in pH 4 springs deserve more attention in YNP, and we expect that the results generated from this research will provide a foundation for understanding microbial communities in these less common springs.

**References**

Almarsdottir AR, Taraceviz A, Gunnarsson I, Orlygsson J (2010) Hydrogen production from sugars and complex biomass by *Clostridium* species, $AK_{14}$, isolated from Icelandic hot spring. Icelandic Agricultural Sciences:61-71

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Amend JP, Shock EL (2001) Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. FEMS Microbiol Rev 25:175-243

Ball JW, McCleskey RB, Nordstrom DK (2010) Water-chemistry data for selected springs, geysers, and streams in Yellowstone National Park, Wyoming, 2006–2008. US Geological Survey Open-File Report 2010–1192:109 p

Ball JW, McCleskey RB, Nordstrom DK, Holloway JM, Verplanck PL (2002) Water-chemistry data for selected springs, geysers, and streams in Yellowstone National Park, Wyoming, 1999-2000. US Geological Survey Open-File Report:02-382

Ball JW, Nordstrom DK, Cunningham KM, Schoonen MAA, Xu Y, DeMonge JM (1998) Water-chemistry and on-site sulfur-speciation data for selected springs in Yellowstone National Park, Wyoming, 1994-1995. US Geological Survey Open-File Report:98-574

Brock T, Brock K, Belly R, Weiss R (1972) *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Microbiol 84:54-68

Burgess E, Unrine J, Mills G, Romanek C, Wiegel J (2012) Comparative geochemical and microbiological characterization of two thermal pools in the Uzon caldera, Kamchatka, Russia. Microb Ecol 63:471-489

Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, Krause DJ, Whitaker RJ (2012) Patterns of gene flow define species of thermophilic Archaea. PLoS Biol 10:e1001265

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266-267

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010b) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-336

Cavaletti L, Monciardini P, Bamonte R, Schumann P, Rohde M, Sosio M, Donadio S (2006) New lineage of filamentous, spore-forming, gram-positive bacteria from soil. Appl Environ Microbiol 72:4360-4369

Cavicchioli R, Amils R, Wagner D, McGenity T (2011) Life and applications of extremophiles. Environ Microbiol 13:1903-1907

Chang Y-J, Land M, Hauser L, Chertkov O, Del Rio TG, Nolan M, Copeland A, Tice H, Cheng J-F, Lucas S (2011) Non-contiguous finished genome sequence and

contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21$^T$). Standards in genomic sciences 5:97

Cole JK, Peacock JP, Dodsworth JA, Williams AJ, Thompson DB, Dong H, Wu G, Hedlund BP (2013) Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. ISME J 7:718-729

Del Rio TG, Abt B, Spring S, Lapidus A, Nolan M, Tice H, Copeland A, Cheng J-F, Chen F, Bruce D (2010) Complete genome sequence of *Chitinophaga pinensis* type strain (UQM 2034$^T$). Standards in genomic sciences 2:87

Deppenmeier U (2002) Redox-driven proton translocation in methanogenic Archaea. CMLS, Cell Mol Life Sci 59:1513-1533

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069-5072

Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R (2012) Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genomics 13:162

Drobner E, Huber H, Rachel R, Stetter KO (1992) *Thiobacillus plumbophilus* spec. nov., a novel galena and hydrogen oxidizer. Arch Microbiol 157:213-217

Dunfield PF, Tamas I, Lee KC, Morgan XC, McDonald IR, Stott MB (2012) Electing a candidate: a speculative history of the bacterial phylum OP10. Environ Microbiol 14:3069-3080

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461

Everroad RC, Otaki H, Matsuura K, Haruta S (2012) Diversification of bacterial community composition along a temperature gradient at a thermal spring. Microbes Environ 27:374-381

Fardeau M-L, Barsotti V, Cayol J-L, Guasco S, Michotey V, Joseph M, Bonin P, Ollivier B (2010) *Caldinitratiruptor microaerophilus*, gen. nov., sp. nov. isolated from a French hot spring (Chaudes-Aigues, Massif Central): a novel cultivated facultative microaerophilic anaerobic thermophile pertaining to the *Symbiobacterium* branch within the Firmicutes. Extremophiles 14:241-247

Fournier RO (1989) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. Annual Review of Earth and Planetary Sciences 17:13-53

Garc á-Horsman JA, Barquera B, Rumbley J, Ma J, Gennis RB (1994) The superfamily of heme-copper respiratory oxidases. Journal of Bacteriology 176:5587-5600

Giovannoni SJ, DeLong EF, Schmidt TM, Pace NR (1990) Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. Appl Environ Microbiol 56:2572-2575

Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F (2010) Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harbor Protocols 2010:pdb.prot5368

Gonzalez P, Correia C, Moura I, Brondino C, Moura J (2006) Bacterial nitrate reductases: molecular and biological aspects of nitrate reduction. Journal of inorganic biochemistry 100:1015-1023

Good IJ (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40:237-264

Hall JR, Mitchell KR, Jackson-Weaver O, Kooser AS, Cron BR, Crossey LJ, Takacs-Vesbach CD (2008) Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes. Appl Environ Microbiol 74:4910-4922

Huber R, Stetter KO (2001) Aquificales. In: eLS. John Wiley & Sons, Ltd,

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377-386

Inagaki F, Motomura Y, Doi K, Taguchi S, Izawa E, Lowe DR, Ogata S (2001) Silicified microbial community at Steep Cone hot spring, Yellowstone National Park. Microbes Environ 16:125-130

Inskeep WP, Jay ZJ, Herrgard MJ, Kozubal MA, Rusch DB, Tringe SG, Macur RE, Jennings R, Boyd ES, Spear JR, Roberto FF (2013a) Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate Linkages between metabolic potential and geochemistry. Front Microbiol 4:95

Inskeep WP, Jay ZJ, Tringe SG, Herrgard MJ, Rusch DB (2013b) The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front Microbiol 4:67

Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, Macur RE, Hamamura N, Jennings Rd, Fouke BW, Reysenbach A-L, Roberto F, Young M, Schwartz A, Boyd ES, Badger JH, Mathur EJ, Ortmann AC, Bateson M, Geesey

G, Frazier M (2010) Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS ONE 5:e9773

Itoh T, Suzuki K-i, Sanchez PC, Nakase T (1999) *Caldivirga maquilingensis* gen. nov., sp. nov., a new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. Int J Syst Bacteriol 49:1157-1163

Jackson CR, Langner HW, Donahoe-Christiansen J, Inskeep WP, McDermott TR (2001) Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. Environ Microbiol 3:532-542

Jiménez DJ, Andreote FD, Chaves D, Montaña JS, Osorio-Forero C, Junca H, Zambrano MM, Baena S (2012) Structural and functional insights from the metagenome of an acidic hot spring microbial planktonic community in the Colombian Andes. PLoS ONE 7:e52069

Klatt CG, Inskeep WP, Herrgard MJ, Jay ZJ, Rusch DB, Tringe SG, Niki Parenteau M, Ward DM, Boomer SM, Bryant DA, Miller SR (2013) Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. Front Microbiol 4:106

Konhauser KO, Jones B, Reysenbach A-L, Renaut RW (2003) Hot spring sinters: keys to understanding Earth's earliest life forms. Canadian Journal of Earth Sciences 40:1713-1724

Law CJ, Maloney PC, Wang D-N (2008) Ins and outs of major facilitator superfamily antiporters. Annu Rev Microbiol 62:289

Madigan MT (2003) Anoxygenic phototrophic bacteria from extreme environments. Photosynthesis Res 76:157-171

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2013) Cluster: cluster analysis basics and extensions.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380

Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Grechkin Y, Ratner A, Jacob B, Pati A, Huntemann M, Liolios K, Pagani I, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Res 40:D123-D129

McCleskey RB (2005) Water-chemistry data for selected hot springs, geysers, and streams in Yellowstone National Park, Wyoming, 2001-2002. US Geological Survey,

McCleskey RB, Ball JW, Nordstrom DK, Holloway JM, Taylor HE (2005) Water-chemistry data for selected hot springs, geysers, and streams in Yellowstone

National Park, Wyoming 2001-2002. US Geological Survey Open-File Report 2004-1316

Meyer-Dombard DR, Shock EL, Amend JP (2005) Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. Geobiology 3:211-227

Meyer-Dombard DR, Swingley W, Raymond J, Havig J, Shock EL, Summons RE (2011) Hydrothermal ecotones and streamer biofilm communities in the Lower Geyser Basin, Yellowstone National Park. Environ Microbiol 13:2216-2231

Meyer B, Kuever J (2008) Homology modeling of dissimilatory APS reductases (AprBA) of sulfur-oxidizing and sulfate-reducing prokaryotes. PloS one 3:e1514

Mitchell K, Takacs-Vesbach C (2008) A comparison of methods for total community DNA preservation and extraction from various thermal environments. J Ind Microbiol Biotechnol 35:1139-1147

Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F (1999) Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. J Bacteriol 181:6573-6584

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H Vegan: community ecology package, 2012. R package version:2.0-5

Piddock LJ (2006) Multidrug-resistance efflux pumps - not just for resistance. Nat Rev Microbiol 4:629-636

Pierson B, Castenholz R (1974) A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. Arch Microbiol 100:5-24

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 2011 Jan 28 12:38

Reysenbach AL, Ehringer H, Hershberger K (2000) Microbial diversity at 83℃ in Calcite Springs, Yellowstone National Park: another environment where the *Aquificales* and "Korarchaeota" coexist. Extremophiles 4:61-67

Reysenbach AL, Seitzinger S, Kirshtein J, McLaughlin E (1999) Molecular constraints on a high-temperature evolution of early life. Biological Bulletin 196:367-372

Reysenbach AL, Wickham GS, Pace NR (1994) Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. Appl Environ Microbiol 60:2113-2119

Rhoads DD, Wolcott RD, Sun Y, Dowd SE (2012) Comparison of culture and molecular identification of bacteria in chronic wounds. International Journal of Molecular Sciences 13:2535-2550

Richardson D, Berks B, Russell D, Spiro S, Taylor C (2001) Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. CMLS, Cell Mol Life Sci 58:165-178

Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. Nature 409:1092-1101

Rye RO, Truesdell AH (2007) The question of recharge to the deep thermal reservoir underlying the geysers and hot springs of Yellowstone National Park. In: al. LAMe (ed) Integrated Geoscience Studies in the Greater Yellowstone Area: Volcanic, Tectonic, and Hydrothermal Processes in the YellowstoneGeoecosystem. U.S. Geol. Surv. Prof. Pap., pp 1717, 1235-1270

Sanford RA, Cole JR, Tiedje JM (2002) Characterization and Description of Anaeromyxobacter dehalogenans gen. nov., sp. nov., an Aryl-Halorespiring Facultative Anaerobic Myxobacterium. Appl Environ Microbiol 68:893-900

Sangkhobol V, Skerman VBD (1981) *Chitinophaga*, a new genus of chitinolytic myxobacteria. Int J Syst Bacteriol 31:285-293

Sayeh R, Birrien J, Alain K, Barbier G, Hamdi M, Prieur D (2010) Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches. Extremophiles 14:501-514

Spanevello MD, Yamamoto H, Patel BKC (2002) *Thermaerobacter subterraneus* sp. nov., a novel aerobic bacterium from the Great Artesian Basin of Australia, and emendation of the genus Thermaerobacter. Int J Syst Evol Microbiol 52:795-800

Takacs-Vesbach C, Inskeep WP, Jay ZJ, Herrgard MJ, Rusch DB, Tringe SG, Kozubal MA, Hamamura N, Macur RE, Fouke BW, Reysenbach AL, McDermott TR, Jennings R, Hengartner NW, Xie G (2013) Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three aquificales lineages. Front Microbiol 4:84

Takacs-Vesbach C, Mitchell K, Jackson-Weaver O, Reysenbach A-L (2008) Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. Environ Microbiol 10:1681-1689

Team RDC (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Tekere M, Lotter A, Olivier J, Jonker N, Venter SN (2011) Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261-5267

Ward DM, Ferris MJ, Nold SC, Bateson MM (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62:1353-1370

Ward DM, Weller R, Bateson MM (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345:63-65

Weber CF, King GM (2010) Distribution and diversity of carbon monoxide-oxidizing bacteria and bulk bacterial communities across a succession gradient on a Hawaiian volcanic deposit. Environ Microbiol 12:1855-1867

Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 301:976-978

Wu Q, Sanford RA, Löffler FE (2006) Uranium(VI) reduction by *Anaeromyxobacter dehalogenans* strain 2CP-C. Appl Environ Microbiol 72:3608-3614

Xie W, Zhang CL, Wang J, Chen Y, Zhu Y, de la Torre JR, Dong H, Hartnett HE, Hedlund BP, Klotz MG (2014) Distribution of ether lipids and composition of the archaeal community in terrestrial geothermal springs: impact of environmental variables. Environ Microbiol:n/a-n/a

**Tables and Figures**

**Table 1.** 454 pyrosequencing and Newbler assembly metrics of the metagenomic DNA sample from site 04YSMH020

| Parameter | 04YSMH020 |
|---|---|
| Total number of reads | 848,583 |
| Mean read length | 438 bp |
| Metagenome size (unassembled reads) | 372 Mbp |
| Metagenome size (assembled reads) | 38 Mbp (10.2%) |
| Number of reads in contigs | 641,401 |
| Number of contigs | 19,346 |
| Reads/contig | 31.15 |
| Largest contig (bp) | 187,560 |
| Mean contig length (bp) | 2,554 |
| N50 contig length (bp) | 4,303 |
| Number of singletons | 129,660 |

**Table 2.** Features of the thermal spring metagenomes based on MG-RAST[a] and IMG/M[b] annotations

| Annotation Platform | MG-RAST | IMG/M |
|---|---|---|
| Metagenome/Features | 04YSMH020 | 04YSMH020 |
| Total number of reads post MG-RAST quality control | 730,387 | - |
| Total DNA scaffolds post IMG/M data processing | - | 19,293 |
| Average GC content | 52 ±10 % | - |
| Protein coding sequences | 426,466 | 53,332 |
| Protein coding sequences with function prediction | 185,985 (43.6%) | 32,441 (60.16%) |
| rRNA genes | 402 | 63 |

[a]Features from unassembled reads that passed MG-RAST quality control.

[b]Features from Newbler assembled reads post IMG/M data processing.

-, Not applicable or not determined.

**Table 3.** Top 30 (by sequence count) COG functions represented in 04YSMH020 metagenomic assembled sequences

| COG category | COG ID | COG Name | sequence count | ranking |
|---|---|---|---|---|
| M | COG0438 | Glycosyltransferase | 247 | 1 |
| G | COG2814 | Arabinose efflux permease | 173 | 2 |
| K | COG1595 | DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog | 169 | 3 |
| TK | COG0745 | Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain | 161 | 4 |
| R | COG0673 | Predicted dehydrogenases and related proteins | 146 | 5 |
| IQR | COG1028 | Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) | 145 | 6 |
| RTKL | COG0515 | Serine/threonine protein kinase | 142 | 7 |
| H | COG2226 | Methylase involved in ubiquinone/menaquinone biosynthesis | 141 | 8 |
| V | COG1131 | ABC-type multidrug transport system, ATPase component | 138 | 9 |
| NU | COG2165 | Type II secretory pathway, pseudopilin PulG | 124 | 10 |
| R | COG4783 | Putative Zn-dependent protease, contains TPR repeats | 122 | 11 |
| NU | COG3063 | Tfp pilus assembly protein PilF | 119 | 12 |
| M | COG0463 | Glycosyltransferases involved in cell wall biogenesis | 117 | 13 |
| T | COG2204 | Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains | 109 | 14 |
| C | COG1529 | Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL homologs | 108 | 15 |
| MG | COG0451 | Nucleoside-diphosphate-sugar epimerases | 107 | 16 |
| E | COG0747 | ABC-type dipeptide transport system, periplasmic component | 107 | 17 |
| I | COG1960 | Acyl-CoA dehydrogenases | 105 | 18 |
| P | COG1629 | Outer membrane receptor proteins, mostly Fe transport | 93 | 19 |
| E | COG0531 | Amino acid transporters | 92 | 20 |
| T | COG0642 | Signal transduction histidine kinase | 91 | 21 |
| I | COG1024 | Enoyl-CoA hydratase/carnithine racemase | 90 | 22 |
| O | COG1225 | Peroxiredoxin | 89 | 23 |
| R | COG0596 | Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) | 88 | 24 |
| TK | COG2197 | Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain | 88 | 25 |
| C | COG1012 | NAD-dependent aldehyde dehydrogenases | 86 | 26 |
| IQ | COG0318 | Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II | 84 | 27 |
| V | COG0841 | Cation/multidrug efflux pump | 82 | 28 |
| T | COG2205 | Osmosensitive K+ channel histidine kinase | 78 | 29 |
| C | COG2141 | Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases | 75 | 30 |

**Fig. 1.** Geographic maps showing sampling locations.

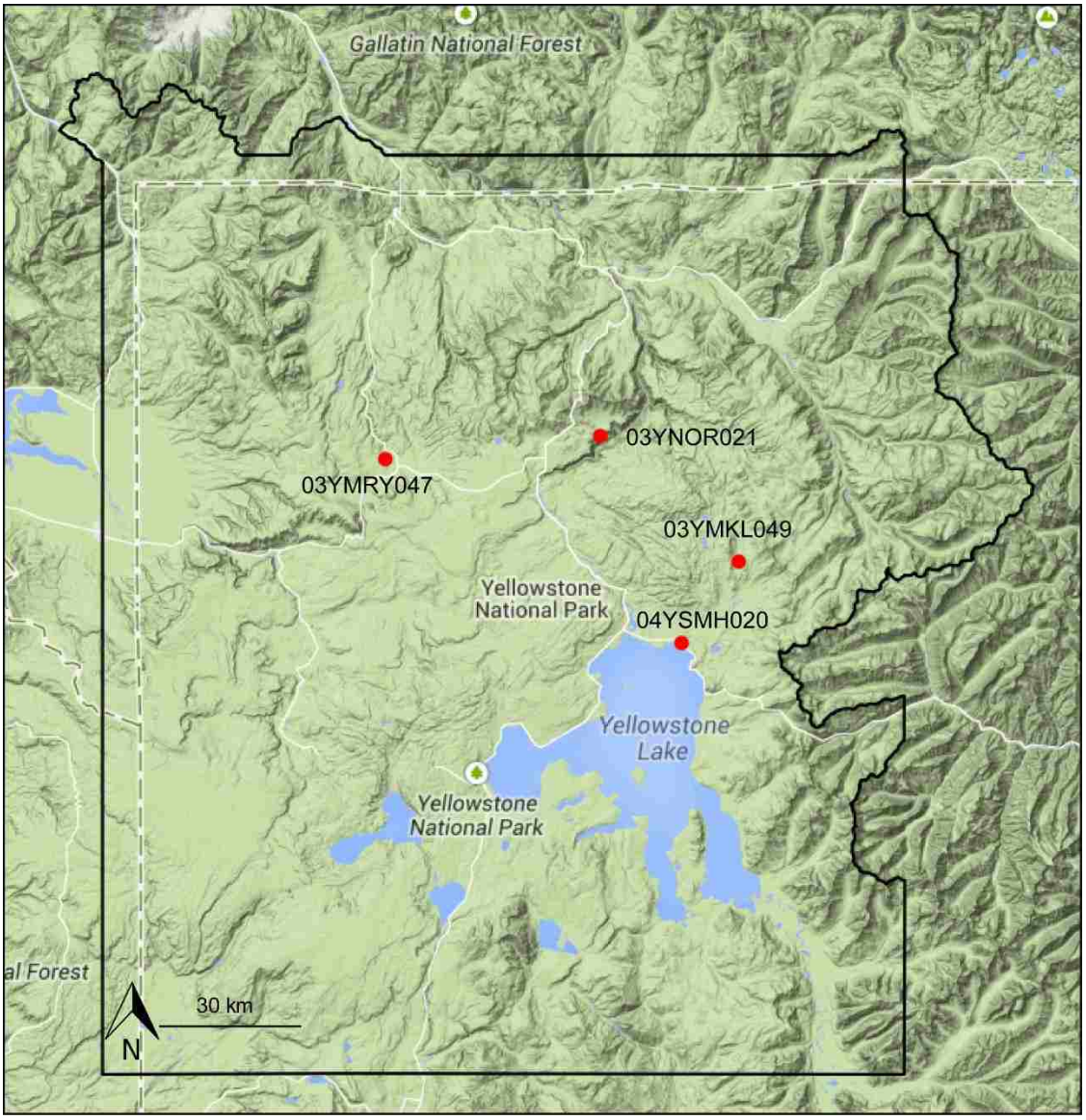


**Fig. 2.** Rarefaction curves of the total number of species annotations (MG-RAST M5NR database) as a function of the number sequences.

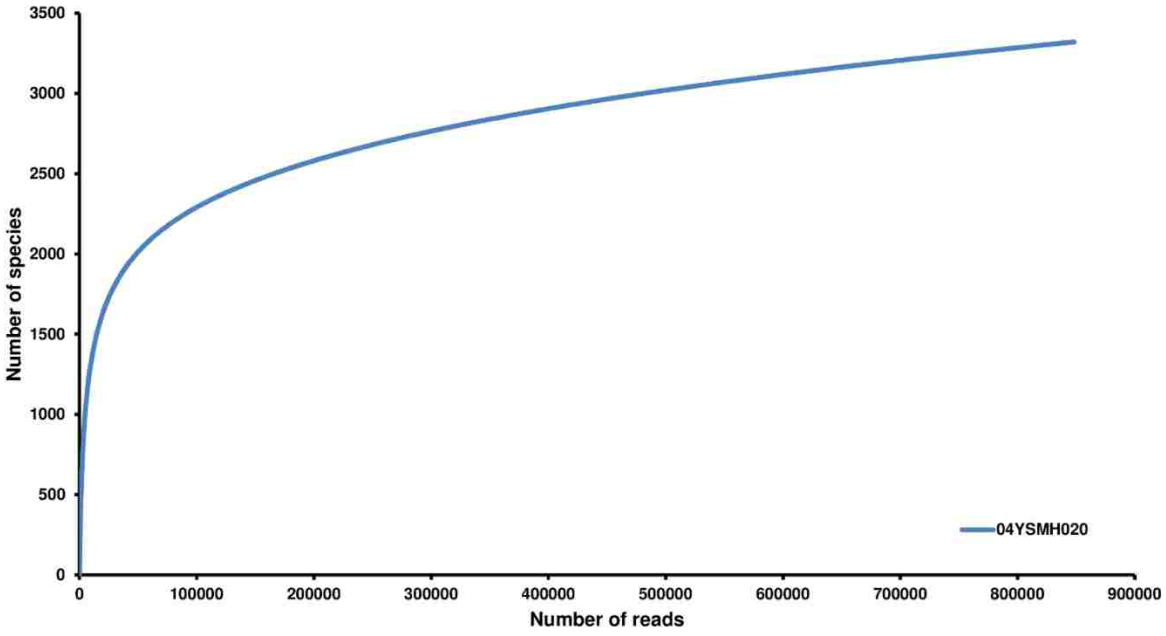


**Fig. 3.** Hierarchical clustering of functional gene groups of 21 YNP metagenomes. (A) Cluster analysis based on COG categories; (B) cluster analysis based on COG functions.

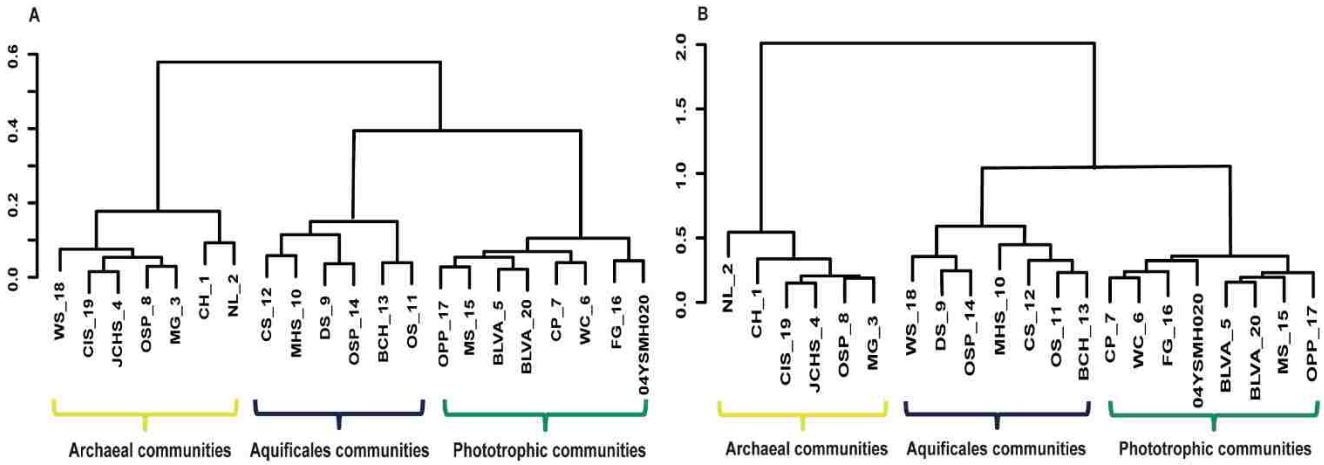


**Fig. 4.** Principal coordinates analysis (PCoA) of 21 YNP metagenomes. (A) Cluster analysis based on COG categories; (B) cluster analysis based on COG functions.

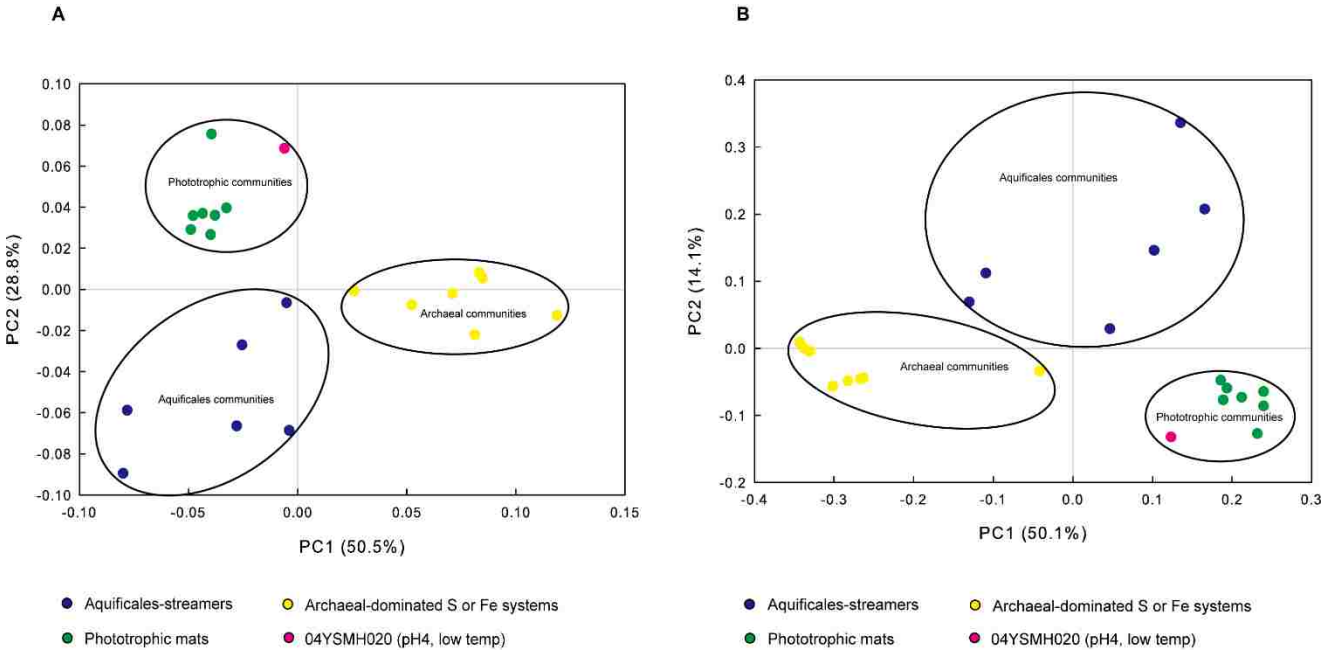


**Fig. 5.** Taxonomic classification and comparison of 16S rRNA and metagenomic reads. (A) Taxonomic classification based on 16S rRNA amplicon pyrosequencing using the Greengenes database; (B) Comparison of 04YSMH020 taxonomic profiles between 16S rRNA amplicon pyrosequencing and metagenomics, based on the Green genes database and MG-RAST M5NR database, respectively.

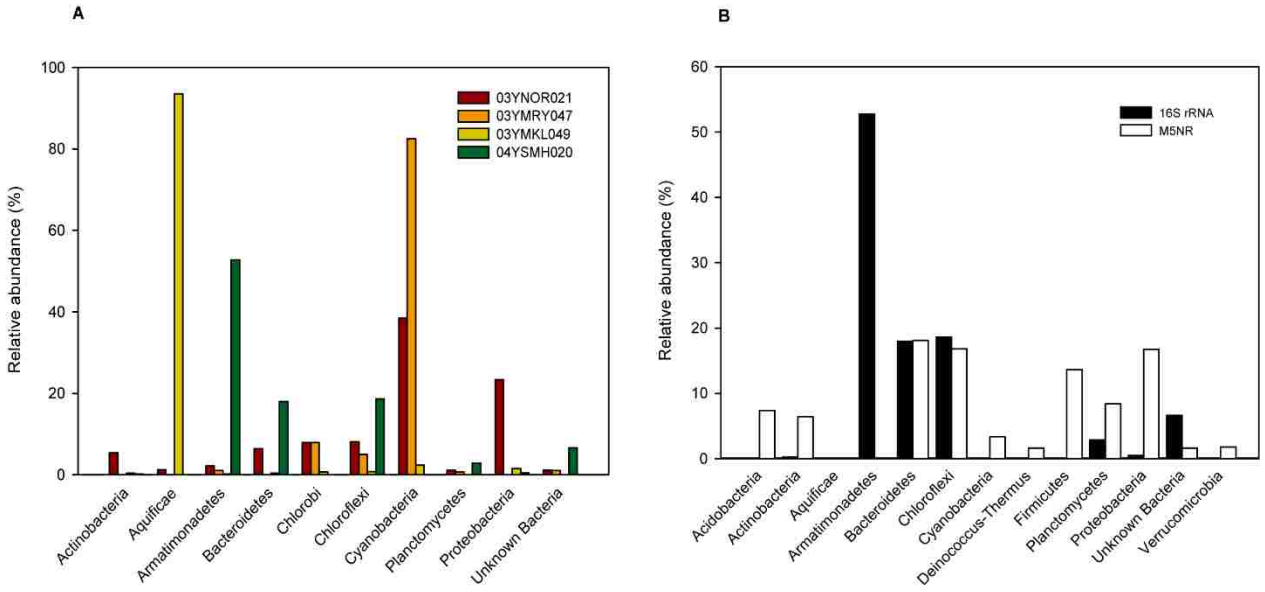


**Fig. 6.** Comparison of the taxonomic assignment of unassembled 04YSMH020 metagenomic sequences based on GenBank (NCBI-nr), M5NR and RefSeq databases

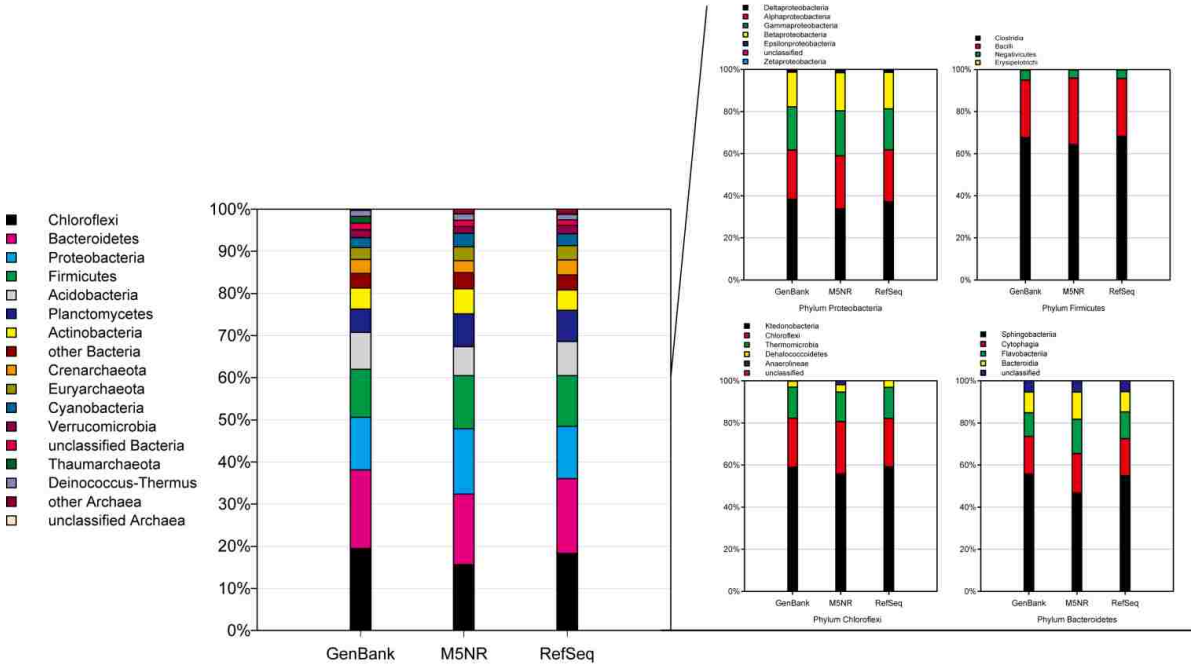


**Fig. 7.** Taxonomic comparison within most abundant classes based on GenBank (NCBI-nr), M5NR and RefSeq databases. A) Ktedonobacteria, B) Sphingobacteria, C) Deltaproteobacteria and D) Clostridia.

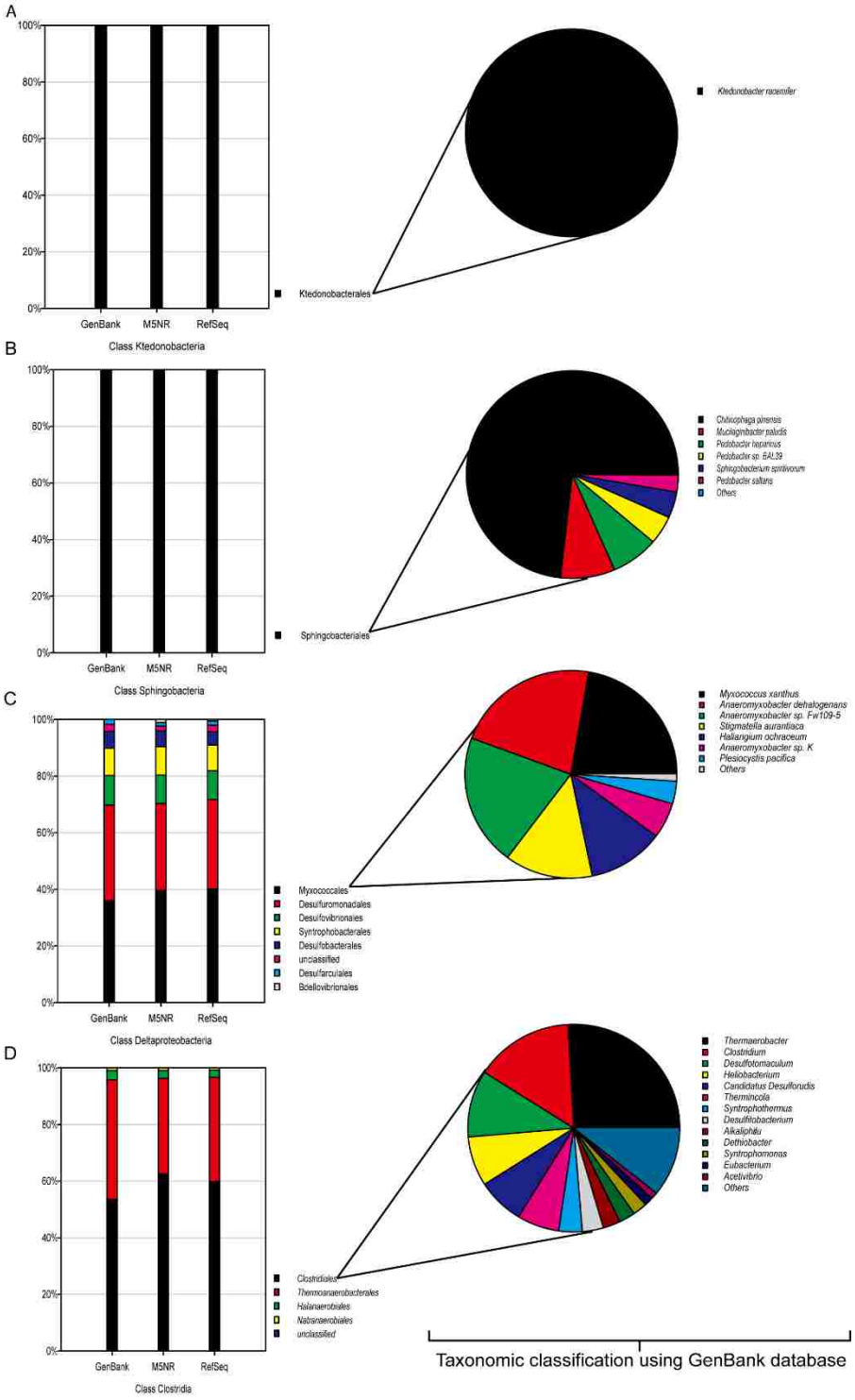


**Fig. 8.** Overrepresented COG functions in the metagenome of 04YSMH020 relative to seven other phototrophic metagenomes. These represent the 20 most enriched functions from a total of 27 COG functions that received significant enrichment scores in six of seven comparisons (Table S2). Average D-score represents the mean enrichment score over all seven comparisons. Letters above graphs represent COG category

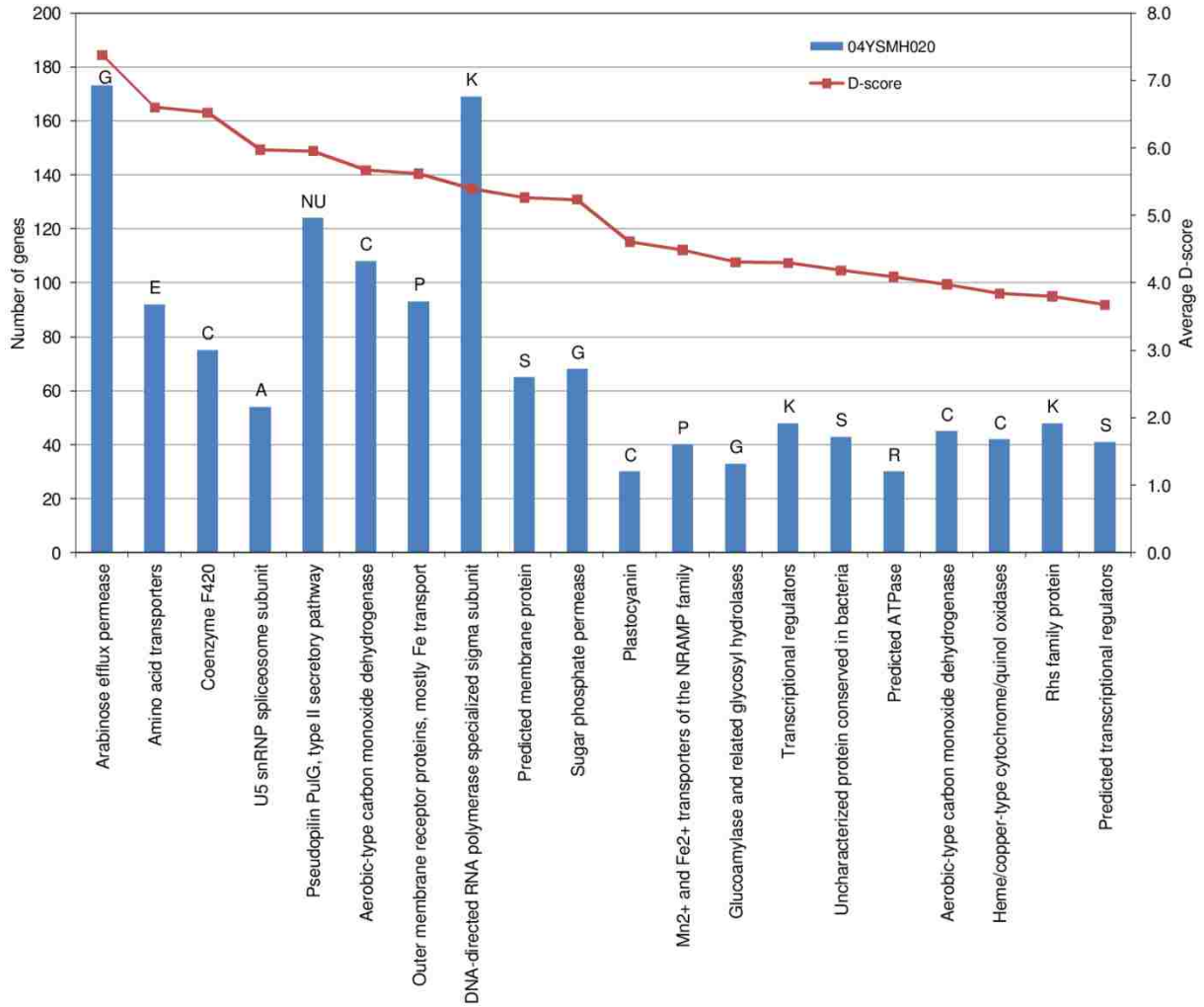


**Fig. 9.** Taxonomic assignment of metagenomic reads from the site 04YSMH020 related to energy metabolism (KEGG identifiers)

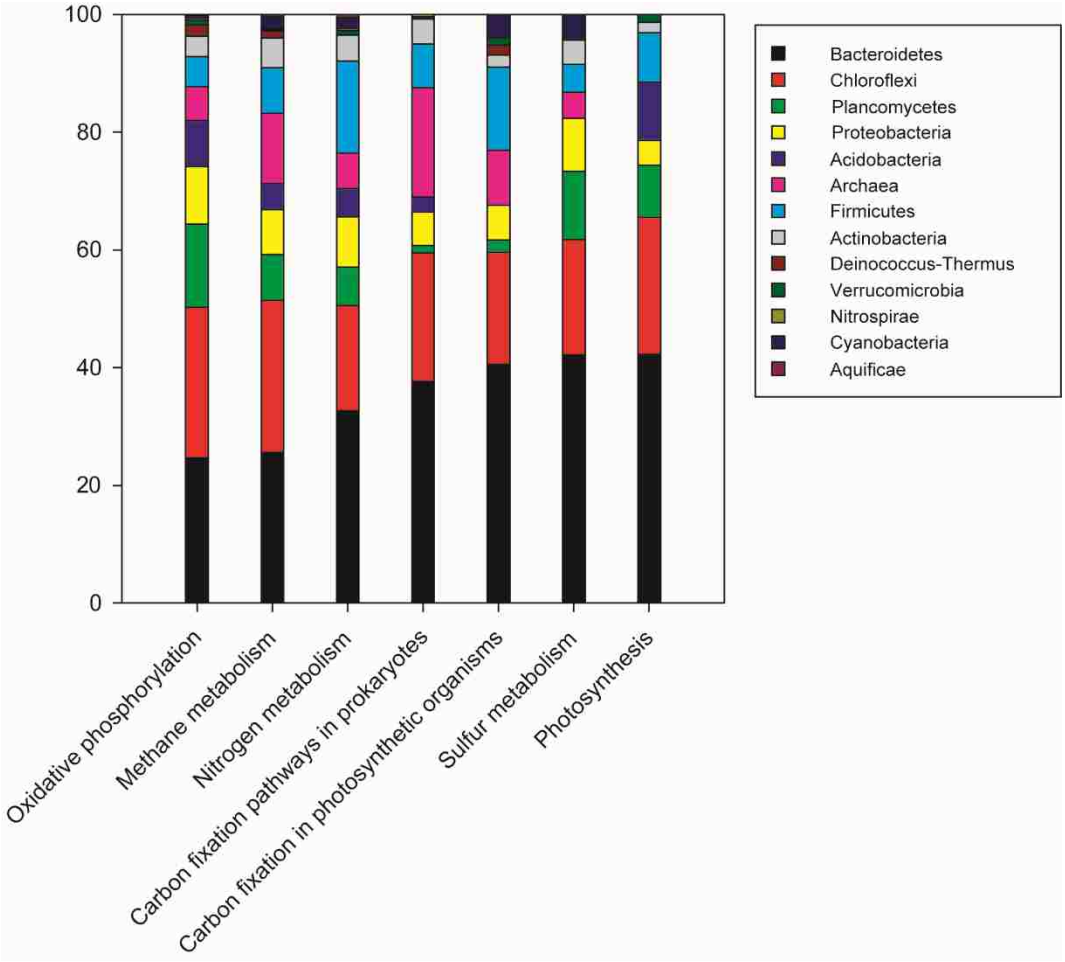


**Table S1.** Geographic and geochemical parameters for the four sampling sites in YNP

| Sample area | Norris | Mary Bay Area | Mud Kettles | Seven Mile Hole |
|---|---|---|---|---|
| Sample ID | 03YNOR021 | 03YMRY047 | 03YMKL049 | 04YSMH020 |
| GPS Location (N/W) | 44.732423/ -110.709777 | 44.553361/ -110.3045805 | 44.634664/ -110.6045546 | 44.754916/ -110.4158659 |
| Physical Context | Sediment, pool edge | Sediment, pool edge | Sediment, gas turbulent pool bottom | Clumpy mat, thermal creek |
| Temperature (℃) | 84 | 80 | 72 | 55 |
| pH | 4.34 | 4.32 | 4.35 | 4.05 |
| Conductivity (μS/cm) | 2130 | 1260 | 1170 | 1538 |
| DOC (mg $L^{-1}$) | 0.3 | 0 | 1.7 | 0 |
| Ca (mg $L^{-1}$) | 4.3 | 6 | 46 | 88.8 |
| K (mg $L^{-1}$) | 44 | 4.2 | 23 | 30 |
| Na (mg $L^{-1}$) | 382.2 | 4.6 | 137.6 | 152.5 |
| Mg (mg $L^{-1}$) | 0.02 | 4 | 19 | 29.1 |
| Alkalinity as $HCO_3^-$ (mg $L^{-1}$) | 0 | 0 | 0 | 0 |
| $SO_4^{2-}$ (mg $L^{-1}$) | 73.3 | 542 | 571 | 306 |
| $Cl^-$ (mg $L^{-1}$) | 599 | 3.67 | 1.82 | 272 |
| Al (mg $L^{-1}$) | 0.91 | 0.047 | 0.43 | 0.263 |
| As(Total) (mg $L^{-1}$) | 2.16 | 0.038 | 0.002 | 0.2 |
| As(III) (mg $L^{-1}$) | 2.16 | 0.025 | 0.001 | 0.005 |
| Ba (mg $L^{-1}$) | 0.019 | 0.04 | 0.083 | 0.073 |
| Be (mg $L^{-1}$) | 0.0038 | 0.00022 | 0.00025 | 0.0009 |
| Cd (mg $L^{-1}$) | 0 | 0.00006 | 0 | 0.00008 |
| Co (mg $L^{-1}$) | 0 | 0 | 0.000027 | 0.00293 |
| Cu (mg $L^{-1}$) | 0.0014 | 0.00098 | 0.00063 | 0.0009 |
| Fe(Total) (mg $L^{-1}$) | 0.033 | 1.4 | 0.264 | 0.009 |
| Fe(II) (mg $L^{-1}$) | 0.031 | 0.359 | 0.252 | 0.003 |
| Li (mg $L^{-1}$) | 3.8 | 0.0047 | 0.072 | 0.611 |
| Mn (mg $L^{-1}$) | 0.0031 | 0.13 | 0.29 | 1.07 |
| Ni (mg $L^{-1}$) | 0 | 0.0005 | 0 | 0.01 |
| P (mg $L^{-1}$) | 0.002 | 0.017 | 0.13 | 0.00002 |
| Pb (mg $L^{-1}$) | 0 | 0.0005 | 0.000083 | 0 |
| Sb (mg $L^{-1}$) | 0.06 | 0.0017 | 0.000049 | 0.00064 |
| Se (mg $L^{-1}$) | 0.0072 | 0 | 0 | 0.0035 |
| $SiO_2$ (mg $L^{-1}$) | 396 | 237 | 282 | 190 |
| Sr (mg $L^{-1}$) | 0.014 | 0.019 | 1.1 | 1.01 |
| V (mg $L^{-1}$) | 0 | 0.0009 | 0.00067 | 0 |
| Zn (mg $L^{-1}$) | 0.016 | 0.067 | 0.078 | 0.0308 |
| $NH_4^-$ (mg $L^{-1}$) | 7.19 | 0 | 33.06 | 16.9 |
| $NO_2^-$ (mg $L^{-1}$) | 0.0029 | 0 | 0.0013 | 0.053 |
| $NO_3^-$ (mg $L^{-1}$) | 0 | 0.23 | 0 | 37 |
| Br (mg $L^{-1}$) | 2.01 | 0 | 0 | 0.907 |
| Cr (mg $L^{-1}$) | 0 | 0 | 0 | 0 |
| $F^-$ (mg $L^{-1}$) | 8.7 | 0.25 | 0.5 | 0.518 |
| $H_2S$ (mg $L^{-1}$) | 0.327 | 1.175 | 1.625 | 0 |

**Table S2.** General features of 20 selected YNP metagenomes for functional comparison. The table adapted from Inskeep et al. 2013b

| Site | Site abbreviation | T (ºC) | pH | DS (µM) | $S^0$ | Geobiological ecosystem type | IMG submission ID |
|---|---|---|---|---|---|---|---|
| White Creek | WC_6 | 48-50 | 8.2 | <1 | N | Phototrophic mats | 341 |
| Chocolate Pots | CP_7 | 52 | 6.2 | <1 | N | Phototrophic mats | 396 |
| Mushroom Spring | MS_15 | 60 | 8.2 | <1 | N | Phototrophic mats | 485 |
| Fairy Geyser | FG_16 | 36-38 | 9.1 | <1 | N | Phototrophic mats | 891 |
| Bath Lake Vista | BLVA_5 | 56-57 | 6.2 | 117 | Y | Phototrophic mats | 340 |
| Bath Lake Vista | BLVA_20 | 54-56 | 6.2 | 120 | Y | Phototrophic mats | 605 |
| Obsidian Pool Prime | OPP_17 | 56 | 5.7 | <1 | N | Bacterial OP Divisions | 572 |
| Dragon Spring | DS_9 | 70-72 | 3.1 | 80 | Y | Aquificales-rich 'filamentous-streamer' communities | 394 |
| 100 Spring Plain | OSP_14 | 72-74 | 3.5 | 10 | N | Aquificales-rich 'filamentous-streamer' communities | 344 |
| Mammoth Hot Spring | MHS_10 | 70-72 | 6.5 | 70 | Y | Aquificales-rich 'filamentous-streamer' communities | 888 |
| Calcite Springs | CS_12 | 76 | 7.8 | 105 | Y | Aquificales-rich 'filamentous-streamer' communities | 395 |
| Octopus Spring | OS_11 | 82 | 7.9 | <1 | N | Aquificales-rich 'filamentous-streamer' communities | 397 |
| Bechler Spring | BCH_13 | 82 | 7.8 | <1 | N | Aquificales-rich 'filamentous-streamer' communities | 343 |
| 100 Spring Plain | OSP_8 | 72 | 3.4 | <1 | N | Crenarchaeota,Mixed, Novel Archaea | 342 |
| Crater Hills | CH_1 | 76 | 2.6 | 2 | Y | Archaeal-dominated sediments | 392 |
| Nymph Lake | NL_2 | 88 | ~4 | 3 | Y | Archaeal-dominated sediments | 484 |
| Monarch Geyser | MG_3 | 78-80 | 4 | 25 | Y | Archaeal-dominated sediments | 875 |
| Cistern Spring | CIS_19 | 78-80 | 4.4 | 20 | Y | Archaeal-dominated sediments | 483 |
| Joseph's Coat Spring | JCHS_4 | 80 | 6.1 | 20 | Y | Archaeal-dominated sediments | 378 |
| Washburn Spring | WS_18 | 76 | 6.4 | 160 | Y | Archaeal-dominated sediments | 571 |

**Table S3.** Domain distribution on the metagenomic sample based on M5NR database

| Database | M5NR |
|---|---|
| Domain | 04YSMH020 % |
| Archaea | 6.36 |
| Bacteria | 81.04 |
| Eukaryota | 0.21 |
| Viruses | - |
| Other sequences | 0.01 |
| Unassigned | 12.32 |
| Unclassified sequences | 0.06 |

**Table S4.** Overrepresented COG functions in the metagenome 04YSMH020 compared to other YNP metagenomes from phototrophic communities (MS_15, FG_16, OPP_17, BLVA_20, BLVA_5, WC_6 and CP_7). Grey shading indicates significant D-scores

| | | | Gene counts | D-scores[a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| COG | COG category | COG description | 04YSMH020 | Mean | MS_15 | FG_16 | OPP_17 | BLVA_20 | BLVA_5 | WC_6 | CP_7 |
| COG2814 | G | Arabinose efflux permease | 173 | 7.4 | 8.73 | 7.55 | 5.52 | 7.43 | 6.91 | 9.42 | 6.07 |
| COG0531 | E | Amino acid transporters | 92 | 6.6 | 8.78 | 6.31 | 3.62 | 7.54 | 6.81 | 6.66 | 6.49 |
| COG2141 | C | Coenzyme F420 | 75 | 6.5 | 8.46 | 4.28 | 6.28 | 7.17 | 6.45 | 6.43 | 6.6 |
| COG5178 | A | U5 snRNP spliceosome subunit | 54 | 6.0 | 6.72 | 5.57 | 6.02 | 6.35 | 5.82 | 5.27 | 6.06 |
| COG2165 | NU | Pseudopilin PulG, type II secretory pathway | 124 | 6.0 | 3.93 | 5.15 | 5.18 | 6.12 | 5.64 | 8.66 | 6.99 |
| COG1529 | C | Aerobic-type carbon monoxide dehydrogenase | 108 | 5.7 | 5.22 | 5.17 | 4.55 | 6.17 | 5.31 | 7.21 | 6.05 |
| COG1629 | P | Outer membrane receptor proteins, mostly Fe transport | 93 | 5.6 | 7.35 | 4.84 | 4.29 | 6.83 | 7.55 | 4.01 | 4.45 |
| COG1595 | K | DNA-directed RNA polymerase specialized sigma subunit | 169 | 5.4 | 7.44 | 5.32 | 5.16 | 5.96 | 5.57 | 4.24 | 4.11 |
| COG5373 | S | Predicted membrane protein | 65 | 5.3 | 5.4 | 4.29 | 5.35 | 5.7 | 6.42 | 5.29 | 4.36 |
| COG2271 | G | Sugar phosphate permease | 68 | 5.2 | 5.09 | 4.79 | 2.12 | 7.01 | 6.62 | 5.81 | 5.19 |
| COG3794 | C | Plastocyanin | 30 | 4.6 | 5.58 | 4.5 | 4.36 | 5.19 | 4.94 | 4.41 | 3.28 |
| COG1914 | P | $Mn^{2+}$ and $Fe^{2+}$ transporters of the NRAMP family | 40 | 4.5 | 6.05 | 5.42 | 3.33 | 3.85 | 4.09 | 4.88 | 3.78 |
| COG3387 | G | Glucoamylase and related glycosyl hydrolases | 33 | 4.3 | 5.7 | 3.68 | 2.68 | 4.98 | 3.59 | 4.7 | 4.8 |
| COG1522 | K | Transcriptional regulators | 48 | 4.3 | 4.63 | 4.9 | 3.07 | 4.29 | 4.44 | 4.08 | 4.64 |
| COG4102 | S | Uncharacterized protein conserved in bacteria | 43 | 4.2 | 3.94 | 2.62 | 5.32 | 5.66 | 5.15 | 2.96 | 3.65 |
| COG1672 | R | Predicted ATPase | 30 | 4.1 | 3.75 | 4.5 | 3.3 | 4.68 | 3.94 | 4.41 | 4.01 |
| COG1319 | C | Aerobic-type carbon monoxide dehydrogenase | 45 | 4.0 | 3.89 | 3.77 | 3 | 4.69 | 3.17 | 5.1 | 4.21 |
| COG1622 | C | Heme/copper-type cytochrome/quinol oxidases | 42 | 3.8 | 5.19 | 3.48 | 2.92 | 4.59 | 4.66 | 3.03 | 3.01 |
| COG3209 | M | Rhs family protein | 48 | 3.8 | 2.76 | 4.38 | 4.78 | 2.79 | 2.84 | 4.6 | 4.45 |
| COG0789 | K | Predicted transcriptional regulators | 41 | 3.7 | 3.99 | 3.73 | 3.22 | 3.95 | 3.81 | 2.93 | 4.08 |
| COG1961 | L | Site-specific recombinases, DNA invertase Pin homologs | 26 | 3.6 | 4.39 | 2.46 | 3.44 | 3.06 | 4.28 | 4 | 3.36 |
| COG5662 | K | Predicted transmembrane transcriptional regulator (anti-sigma factor) | 35 | 3.5 | 3.42 | 4.31 | 2.87 | 4.72 | 3.59 | 3.01 | 2.91 |
| COG0517 | R | FOG: CBS domain | 36 | 3.4 | 3.25 | 2.69 | 4.32 | 3.44 | 3.7 | 2.93 | 3.42 |
| COG0723 | C | Rieske Fe-S protein | 24 | 3.4 | 4.59 | 2.2 | 3.23 | 4.02 | 2.78 | 3.51 | 3.42 |
| COG4099 | R | Predicted peptidase | 20 | 3.4 | 3.76 | 3.38 | 2.5 | 3.52 | 3.92 | 3.02 | 3.6 |
| COG2080 | C | Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homologs | 38 | 3.3 | 3.79 | 3.08 | 1.99 | 3.65 | 3.15 | 3.32 | 4.25 |
| COG1846 | K | Transcriptional regulators | 44 | 3.3 | 5.24 | 3.03 | 2.04 | 2.82 | 3.56 | 2.9 | 3.37 |

[a]P-value cutoffs used to identify significant D-scores (False Discovery Rate of 0.05): MS_15, ≤ 3.15e-03; FG_16, ≤ 1.61e-03; OPP_17, ≤ 1.51e-03; BLVA_20, ≤ 2.82e-03; BLVA_5, ≤ 3.02e-03; WC_6, ≤ 2.28e-03; CP_7, ≤ 1.95e-03.

**Table S5.** Sequences related to specific functions and taxa within nitrogen cycle using KEGG pathways

| Number of reads assigned | EC number | Orthology KEGG | Enzyme name | Gene | Metagenome best hit organism | Class | E-value |
|---|---|---|---|---|---|---|---|
| 234 | 1.7.99.4 | K00373 | nitrate reductase delta subunit | *nar*J | *Kyrpidia tusciae DSM 2912* | Bacilli | 2e-16 |
| | | K00372 | assimilatory nitrate reductase catalytic subunit | *nas*A | *Streptomyces sp.* | Actinobacteria | 3e-46 |
| | | K08345 | | | *Thermobaculum terrenum* | n.a. | 3e-55 |
| | | | | | *Candidatus Solibacter usitatus Ellin6076* | Solibacteres | 1e-53 |
| | | | | | *Ktedonobacter racemifer DSM 44963* | Ktedonobacteria | 6e-64 |
| | | K00374 | nitrate reductase gamma subunit | *nar*I | *Vulcanisaeta distributa DSM 14429* | Thermoprotei | 6e-24 |
| | | | | | *Kyrpidia tusciae DSM 2912* | Bacilli | 6e-34 |
| | | | | | *Salinispora tropica CNB-440* | Actinobacteria | 9e-17 |
| | | | | | *Sulfobacillus acidophilus TPY* | Clostridia | 4e-21 |
| | | | | | *Metallosphaera yellowstonensis MK1* | Thermoprotei | 9e-23 |
| | | | | | *Rubrobacter xylanophilus DSM 9941* | Actinobacteria | 3e-26 |
| | | | | | *Salinispora arenicola CNS-205* | Actinobacteria | 1e-22 |
| | | | | | *Geobacillus kaustophilus HTA426* | Bacilli | 2e-32 |
| | | | | | *Geobacillus sp. Y412MC61* | Bacilli | 1e-32 |
| | | | | | *Bacillus coagulans 36D1* | Bacilli | 2e-20 |
| | | | | | *Rubrobacter xylanophilus DSM 9941* | Actinobacteria | 4e-26 |
| | | | | | *Burkholderia pseudomallei 576* | Betaproteobacteria | 5e-16 |
| | | | | | *Catenulispora acidiphila DSM 44928* | Actinobacteria | 9e-20 |
| | | | | | *Paenibacillus peoriae KCTC 3763* | Bacilli | 4e-18 |
| | | | | | *Geobacillus thermodenitrificans NG80-2* | Bacilli | 1e-29 |
| | | K00371 | nitrate reductase beta subunit | *nar*H | *Bacillus cytotoxicus NVH 391-98* | Bacilli | 3e-63 |
| | | | | | *Kyrpidia tusciae DSM 2912* | Bacilli | 7e-83 |
| | | | | | *Bacillus thuringiensis serovar finitimus YBT-020* | Bacilli | 3e-82 |
| | | | | | *Bacillus sp. 10403023* | Bacilli | 2e-55 |
| | | K00370 | nitrate reductase alpha subunit | *nar*G | *Geobacillus sp. WCH70* | Bacilli | 7e-26 |
| | | | | | *Vulcanisaeta distributa DSM 14429* | Thermoprotei | 8e-52 |
| | | | | | *Geobacillus thermodenitrificans NG80-2* | Bacilli | 4e-65 |
| | | | | | *Kyrpidia tusciae DSM 2912* | Bacilli | 7e-74 |
| | | | | | *Bacillus selenitireducens MLS10* | Bacilli | 6e-54 |
| | | | | | *Streptomyces griseoaurantiacus M045* | Actinobacteria | 3e-46 |
| | | | | | *Paenibacillus terrae HPL-003* | Bacilli | 2e-60 |
| | | | | | *Geobacillus sp. G11MC16* | Bacilli | 1e-49 |
| | | | | | *Rubrobacter xylanophilus DSM 9941* | Actinobacteria | 2e-34 |
| | | | | | *Streptomyces griseoaurantiacus M045* | Actinobacteria | 4e-29 |
| | | | | | *Acidothermus cellulolyticus 11B* | Actinobacteria | 5e-34 |
| | | | | | *Paenibacillus sp. Aloe-11* | Bacilli | 1e-58 |
| | | | | | *Desulfitobacterium hafniense DP7* | Clostridia | 4e-46 |
| | | | | | *Bacillus coagulans 2-6* | Bacilli | 8e-76 |
| | | | | | *Bacillus azotoformans LMG 9581* | Bacilli | 3e-59 |
| | | | | | *Desulfobacterium autotrophicum HRM2* | Deltaproteobacteria | 2e-32 |
| | | | | | *Candidatus Nitrospira defluvii* | Nitrospira | 6e-54 |
| | | | | | *Desulfomonile tiedjei DSM 6799* | Deltaproteobacteria | 4e-37 |
| | | | | | *Paenibacillus terrae HPL-003* | Bacilli | 7e-21 |
| | | | | | *Lentibacillus sp. Grbi* | Bacilli | 1e-20 |
| | | | | | *Streptomyces sp.* | Actinobacteria | 4e-26 |
| 25 | 1.7.2.1 | K00368 | nitrite reductase | *nir*K | *Psychrobacter sp. 1501(2011)* | Gammaproteobacteria | 2e-24 |
| | | | | | *Actinobacillus pleuropneumoniae serovar 1 str. 4074* | Gammaproteobacteria | 7e-34 |
| | | | | | *Rhodothermus marinus SG0.5JP17-172* | Bacteroidetes | 4e-35 |
| | | | | | *Chryseobacterium gleum ATCC 35910* | Flavobacteriia | 2e-59 |
| | | | | | *Turneriella parva DSM 21527* | Spirochaetia | 8e-32 |
| | | | | | *Kingella oralis ATCC 51147* | Betaproteobacteria | 3e-43 |
| | | | | | *Flavobacteriaceae bacterium 3519-10* | Flavobacteriia | 8e-40 |
| | | | | | *Flavobacterium johnsoniae* | Flavobacteriia | 2e-47 |
| | | | | | *Aequorivita sublithincola DSM 14238* | Flavobacteriia | 3e-40 |
| 104 | 1.7.99.7 | K04748 | nitric oxide reductase NorQ protein | *nor*Q | *Desulfococcus oleovorans Hxd3* | Deltaproteobacteria | 3e-24 |
| | | | | | *Methanosarcina mazei Go1* | Methanomicrobia | 5e-59 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | K02305 | nitric oxide reductase subunit C | *nor*C | *Hydrogenobacter thermophilus TK-6* | Aquificae | 6e-48 |
| | | K02448 | nitric oxide reductase NorD protein | *nor*D | *Desulfococcus oleovorans Hxd3* | Deltaproteobacteria | 9e-37 |
| | | K04561 | nitric oxide reductase subunit B | *nor*B | *Geobacter sp. FRC-32* | Deltaproteobacteria | 8e-48 |
| | | | | | *Anaeromyxobacter sp. K* | Deltaproteobacteria | 1e-47 |
| | | | | | *Microlunatus phosphovorus NM-1* | Actinobacteria | 3e-60 |
| | | | | | *Anaeromyxobacter dehalogenans 2CP-C* | Deltaproteobacteria | 6e-61 |
| | | | | | *Sorangium cellulosum 'So ce 56'* | Deltaproteobacteria | 2e-45 |
| | | | | | *Capnocytophaga sp. CM59* | Flavobacteriia | 5e-38 |
| | | | | | *Rhodococcus equi 103S* | Actinobacteria | 1e-40 |
| | | | | | *Anaeromyxobacter sp. Fw109-5* | Deltaproteobacteria | 9e-67 |
| | | | | | *Mycobacterium avium subsp. avium ATCC 25291* | Actinobacteria | 2e-50 |
| | | | | | *Bacillus sp. 1NLA3E* | Bacilli | 3e-61 |
| | | | | | *Allochromatium vinosum DSM 180* | Gammaproteobacteria | 8e-30 |
| | | | | | *Intrasporangium calvum DSM 43043* | Actinobacteria | 7e-20 |
| | | | | | *Brachybacterium faecium DSM 4810* | Actinobacteria | 1e-66 |
| 38 | 1.7.99.6 | K00376 | nitrous-oxide reductase | *nos*Z | *Melioribacter roseus P3M* | Ignavibacteriae | 3e-62 |
| | | | | | *Persephonella marina EX-H1* | Aquificae | 2e-32 |
| | | | | | *Imtechella halotolerans K1* | Flavobacteriia | 2e-59 |
| | | | | | *Rhodothermus marinus SG0.5JP17-172* | Bacteroidetes | 7e-41 |
| | | | | | *Thermomicrobium roseum DSM 5159* | Thermomicrobia | 1e-50 |
| | | | | | *Niastella koreensis GR20-10* | Sphingobacteriia | 4e-45 |
| 58 | 1.7.7.1 | K00366 | ferredoxin-nitrite reductase | nirA | *Chthoniobacter flavus Ellin428* | Verrucomicrobia | 4e-51 |
| | | | | | *Paenibacillus sp.* | Bacilli | 5e-37 |
| | | | | | *Acidobacterium capsulatum ATCC 51196* | Acidobacteriia | 1e-57 |
| | | | | | *Geobacillus sp. WCH70* | Bacilli | 6e-31 |
| | | | | | *Rhodothermus marinus SG0.5JP17-172* | Bacteroidetes | 6e-33 |
| | | | | | *Candidatus Koribacter versatilis Ellin345* | Solibacteres | 3e-58 |
| | | | | | *Corynebacterium aurimucosum ATCC 700975* | Actinobacteria | 2e-29 |
| | | | | | *Acidothermus cellulolyticus 11B* | Actinobacteria | 8e-35 |
| | | | | | *Nitrolancetus hollandicus* | Thermomicrobia | 2e-61 |
| | | | | | *Brevibacillus sp. BC25* | Bacilli | 4e-23 |
| | | | | | *Bacillus sp. NRRL B-14911* | Bacilli | 1e-28 |
| | | | | | *Thermosinus carboxydivorans Nor1* | Negativicutes | 6e-31 |
| | | | | | *Candidatus Nitrospira defluvii* | Nitrospira | 2e-46 |
| | | | | | *Gloeobacter violaceus PCC 7421* | Gloeobacteria | 1e-35 |
| 31 | 1.7.7.4 | K00362 | nitrite reductase large subunit | *nir*B | *Planctomyces maris DSM 8797* | Planctomycetia | 2e-49 |
| | | | | | *Schlesneria paludicola DSM 18645* | Planctomycetia | 1e-64 |
| | | | | | *Ktedonobacter racemifer DSM 44963* | Ktedonobacteria | 1e-34 |
| | | | | | *Niastella koreensis GR20-10* | Sphingobacteriia | 8e-29 |
| | | | | | *Thermobacillus composti KWC4* | Bacilli | 1e-29 |
| | | | | | *Schlesneria paludicola DSM 18645* | Planctomycetia | 8e-69 |

**Table S6.** Sequences related to specific functions and taxa within sulfur cycle using KEGG pathways

| Number of reads assigned | EC number | Orthology KEGG | Enzyme name | Gene | Metagenome best hit organism | Class |
|---|---|---|---|---|---|---|
| 60 | 2.3.1.30 | K00640 | serine O-acetyltransferase | *cys*E | *Mucilaginibacter paludis DSM 18603* | Sphingobacteriia |
| | | | | | *Emticicia oligotrophica DSM 17448* | Cytophagia |
| | | | | | *Ktedonobacter racemifer DSM 44963* | Ktedonobacteria |
| | | | | | *Sphingobacterium sp. 21* | Sphingobacteriia |
| | | | | | *Roseiflexus sp. RS-1* | Chloroflexi |
| | | | | | *Thermovibrio ammonificans HB-1* | Aquificae |
| | | | | | *Gemmata obscuriglobus UQM 2246* | Planctomycetia |
| | | | | | *Cyanothece sp. ATCC 51472* | Oscillatoriophycideae |
| | | | | | *Blastopirellula marina DSM 3645* | Planctomycetia |
| | | | | | *Roseiflexus castenholzii DSM 13941* | Chloroflexi |
| | | | | | *gamma proteobacterium HIMB30* | Gammaproteobacteria |
| | | | | | *Pedobacter sp. BAL39* | Sphingobacteriia |
| | | | | | *Rhodopirellula baltica SH 1* | Planctomycetia |
| | | | | | *Pedosphaera parvula Ellin514* | Verrucomicrobia |
| | | | | | *Desulfurobacterium thermolithotrophum DSM 11699* | Aquificae |
| 146 | 2.7.7.4 | K00956 | sulfate adenylyltransferase subunit 1 | *cys*N | *Runella slithyformis DSM 19594* | Cytophagia |
| | | | | | *Nitrosococcus oceani AFC27* | Gammaproteobacteria |
| | | | | | *Ornithobacterium rhinotracheale DSM 15997* | Flavobacteriia |
| | | | | | *Dyadobacter fermentans DSM 18053* | Cytophagia |
| | | | | | *Niastella koreensis GR20-10* | Sphingobacteriia |
| | | | | | *Chryseobacterium gleum ATCC 35910* | Flavobacteriia |
| | | | | | *Spirosoma linguale DSM 74* | Cytophagia |
| | | | | | *Verrucomicrobiae bacterium DG1235* | Verrucomicrobia |
| | | K00957 | sulfate adenylyltransferase subunit 2 | *cys*D | *Fluviicola taffensis DSM 16823* | Flavobacteriia |
| | | | | | *Niastella koreensis GR20-10* | Sphingobacteriia |
| | | | | | *Pedobacter sp. BAL39* | Sphingobacteriia |
| | | | | | *Niabella soli DSM 19437* | Sphingobacteriia |
| | | | | | *Capnocytophaga sp. CM59* | Flavobacteriia |
| | | | | | *Capnocytophaga canimorsus Cc5* | Flavobacteriia |
| | | | | | *Capnocytophaga ochracea DSM 7271* | Flavobacteriia |
| | | | | | *Sphingobacterium sp. 21* | Sphingobacteriia |
| | | | | | *Niastella koreensis GR20-10* | Sphingobacteriia |
| | | | | | *Spirosoma linguale DSM 74* | Cytophagia |
| | | | | | *Clostridium cellulovorans 743B* | Clostridia |
| | | | | | *Flexibacter litoralis DSM 6794* | Cytophagia |
| | | K00958 | sulfate adenylyltransferase | *sat* | *Rhodothermus marinus SG0.5JP17-172* | Bacteroidetes |
| | | | | | *Planctomyces brasiliensis DSM 5305* | Planctomycetia |
| | | | | | *Ktedonobacter racemifer DSM 44963* | Ktedonobacteria |
| | | | | | *Brevibacillus brevis NBRC 100599* | Bacilli |
| | | | | | *Ferroglobus placidus DSM 10642* | Archaeoglobi |
| | | | | | *Paenibacillus polymyxa E681* | Bacilli |
| | | | | | *Pyrococcus yayanosii CH1* | Thermococci |
| | | | | | *Planctomyces maris DSM 8797* | Planctomycetia |
| | | | | | *Rubrobacter xylanophilus DSM 9941* | Actinobacteria |
| | | | | | *Desulfatibacillum alkenivorans AK-01* | Deltaproteobacteria |
| | | | | | *Thermaerobacter marianensis DSM 12885* | Clostridia |
| | | | | | *Thermobaculum terrenum ATCC BAA-798* | unclassified Bacteria |
| | | | | | *Thiorhodococcus drewsii AZ1* | Gammaproteobacteria |
| | | | | | *Nitrolancetus hollandicus* | Thermomicrobia |
| | | | | | *Aciduliprofundum boonei T469* | unclassified Euryarchaeota |
| | | | | | *Nitrococcus mobilis Nb-231* | Gammaproteobacteria |
| | | | | | *Rhodobacter sphaeroides ATCC 17025* | Alphaproteobacteria |
| | | | | | *Oscillatoria sp. PCC 6506* | Oscillatoriophycideae |
| | | | | | *Planctomyces limnophilus DSM 3776* | Planctomycetia |
| | | | | | *Cyanobacterium stanieri PCC 7202* | Oscillatoriophycideae |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | *Ferroglobus placidus DSM 10642* | Archaeoglobi |
| | | K00955 | bifunctional enzyme CysN/CysC | *cys*NC | *Kribbella flavida DSM 17836* | Actinobacteria |
| 20 | 1.8.99.2 | K00394 | adenylylsulfate reductase, subunit A | *apr*A | *Thiobacillus plumbophilus* | Betaproteobacteria |
| | | | | | *Caldivirga maquilingensis IC-167* | Thermoprotei |
| | | K00395 | adenylylsulfate reductase, subunit B | *apr*B | *Sideroxydans lithotrophicus ES-1* | Betaproteobacteria |
| 60 | 2.7.1.25 | K00955 | bifunctional enzyme CysN/CysC | *cys*NC | *Candidatus Nitrospira defluvii* | Nitrospira |
| | | | | | *Opitutus terrae PB90-1* | Verrucomicrobia |
| | | | | | *Kribbella flavida DSM 17836* | Actinobacteria |
| | | K00860 | adenylylsulfate kinase | *cys*C | *Thermobaculum terrenum ATCC BAA-798* | unclassified Bacteria |
| | | | | | *Oscillatoria sp. PCC 6506* | Oscillatoriophycideae |
| | | | | | *Haliangium ochraceum DSM 14365* | Deltaproteobacteria |
| | | | | | *Ktedonobacter racemifer DSM 44963* | Ktedonobacteria |
| | | | | | *Candidatus Solibacter usitatus Ellin6076* | Solibacteres |
| | | | | | *Methanohalobium evestigatum Z-7303* | Methanomicrobia |
| | | | | | *Gemmata obscuriglobus UQM 2246* | Planctomycetia |
| | | | | | *Rubrobacter xylanophilus DSM 9941* | Actinobacteria |
| | | | | | *planctomycete KSU-1* | Planctomycetia |
| 153 | 2.5.1.47 | K12339 | cysteine synthase B | *cys*M | *Rubrobacter xylanophilus DSM 9941* | Actinobacteria |
| | | | | | *Ktedonobacter racemifer DSM 44963* | Ktedonobacteria |
| | | | | | *Emticicia oligotrophica DSM 17448* | Cytophagia |
| | | | | | *Candidatus Methylomirabilis oxyfera* | unclassified Bacteria |
| | | | | | *Gloeobacter violaceus PCC 7421* | Gloeobacteria |
| | | | | | *Synechococcus sp. JA-3-3Ab* | Oscillatoriophycideae |
| | | | | | *Thermobaculum terrenum ATCC BAA-798* | unclassified Bacteria |
| | | | | | *Mucilaginibacter paludis DSM 18603* | Sphingobacteriia |
| | | | | | *Mycobacterium vaccae RIVM* | Actinobacteria |
| | | | | | *Hydrogenobaculum sp. Y04AAS1* | Aquificae |
| | | K01738 | cysteine synthase A | *cys*K | *Symbiobacterium thermophilum IAM 14863* | Clostridia |
| | | | | | *Bacteroides uniformis ATCC 8492* | Bacteroidia |
| | | | | | *Desulfomicrobium baculatum DSM 4028* | Deltaproteobacteria |
| | | | | | *Singulisphaera acidiphila DSM 18658* | Planctomycetia |
| | | | | | *Cyanobium sp. PCC 7001* | Oscillatoriophycideae |
| | | | | | *Neisseria cinerea ATCC 14685* | Betaproteobacteria |
| | | | | | *Methylovorus glucosetrophus SIP3-4* | Betaproteobacteria |
| | | | | | *Eikenella corrodens ATCC 23834* | Betaproteobacteria |
| | | | | | *Gemmata obscuriglobus UQM 2246* | Planctomycetia |
| | | | | | *Anaerolinea thermophila UNI-1* | Anaerolineae |
| | | | | | *Synechococcus sp. PCC 7335* | Oscillatoriophycideae |
| | | | | | *Cyanobacterium stanieri PCC 7202* | Oscillatoriophycideae |
| | | | | | *Chroococcidiopsis thermalis PCC 7203* | unclassified Cyanobacteria |
| | | | | | *Isosphaera pallida ATCC 43644* | Planctomycetia |
| | | | | | *Mycobacterium parascrofulaceum ATCC BAA-614* | Actinobacteria |
| | | | | | *Lysinibacillus fusiformis ZC1* | Bacilli |
| | | | | | *Acidobacterium capsulatum ATCC 51196* | Acidobacteriia |
| | | | | | *Ktedonobacter racemifer DSM 44963* | Ktedonobacteria |
| | | | | | *Methylacidiphilum infernorum V4* | unclassified Verrucomicrobia |
| | | | | | *Desulfovibrio alaskensis G20* | Deltaproteobacteria |
| | | | | | *Desulfosporosinus acidiphilus SJ4* | Clostridia |
| | | | | | *Acaryochloris sp. HICR111A* | Oscillatoriophycideae |
| | | | | | *Nostoc sp. PCC 7524* | Nostocales |
| | | | | | *Cyanothece sp. PCC 7425* | Oscillatoriophycideae |
| 30 | 1.8.4.8 | K00390 | phosphoadenosine phosphosulfate reductase | *cys*H | *Gemmata obscuriglobus UQM 2246* | Planctomycetia |
| | | | | | *Sulfolobus islandicus L.S.2.15* | Thermoprotei |
| | | | | | *Vulcanisaeta moutnovskia 768-28* | Thermoprotei |
| | | | | | *Symbiobacterium thermophilum IAM 14863* | Clostridia |
| | | | | | *Pedobacter saltans DSM 12145* | Sphingobacteriia |
| | | | | | *Chitinophaga pinensis DSM 2588* | Sphingobacteriia |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | *Rhodothermus marinus DSM 4252* | Bacteroidetes |
| | | | | | *Acidothermus cellulolyticus 11B* | Actinobacteria |
| | | | | | *Flavobacterium frigoris PS1* | Flavobacteriia |
| 57 | 1.8.1.2 | K00381 | sulfite reductase (NADPH) hemoprotein beta-component | *cys*I | *Sphingobacterium sp. 21* | Sphingobacteriia |
| | | | | | *Chitinophaga pinensis DSM 2588* | Sphingobacteriia |
| | | | | | *Emticicia oligotrophica DSM 17448* | Cytophagia |
| | | | | | *Schlesneria paludicola DSM 18645* | Planctomycetia |
| | | | | | *Paenibacillus vortex V453* | Bacilli |
| | | | | | *Robiginitalea biformata HTCC2501* | Flavobacteriia |
| | | | | | *Emticicia oligotrophica DSM 17448* | Cytophagia |
| | | | | | *Gemmata obscuriglobus UQM 2246* | Planctomycetia |
| | | | | | *Mucilaginibacter paludis DSM 18603* | Sphingobacteriia |
| | | | | | *Sphingobacterium spiritivorum ATCC 33300* | Sphingobacteriia |
| 4 | 1.8.7.1 | K00392 | sulfite reductase (ferredoxin) | *sir* | *Gemmata obscuriglobus UQM 2246* | Planctomycetia |

**Fig. S1.** Two way-hierarchical clustering of normalized COG protein family abundance data averaged across the level of COG categories. The data was standardized (subtract mean and divide by standard deviation) across sites before clustering so that the color scale units represent standard deviations from the mean across sites. Red colors correspond to values that are higher than the site mean and green colors to values that are lower than the mean

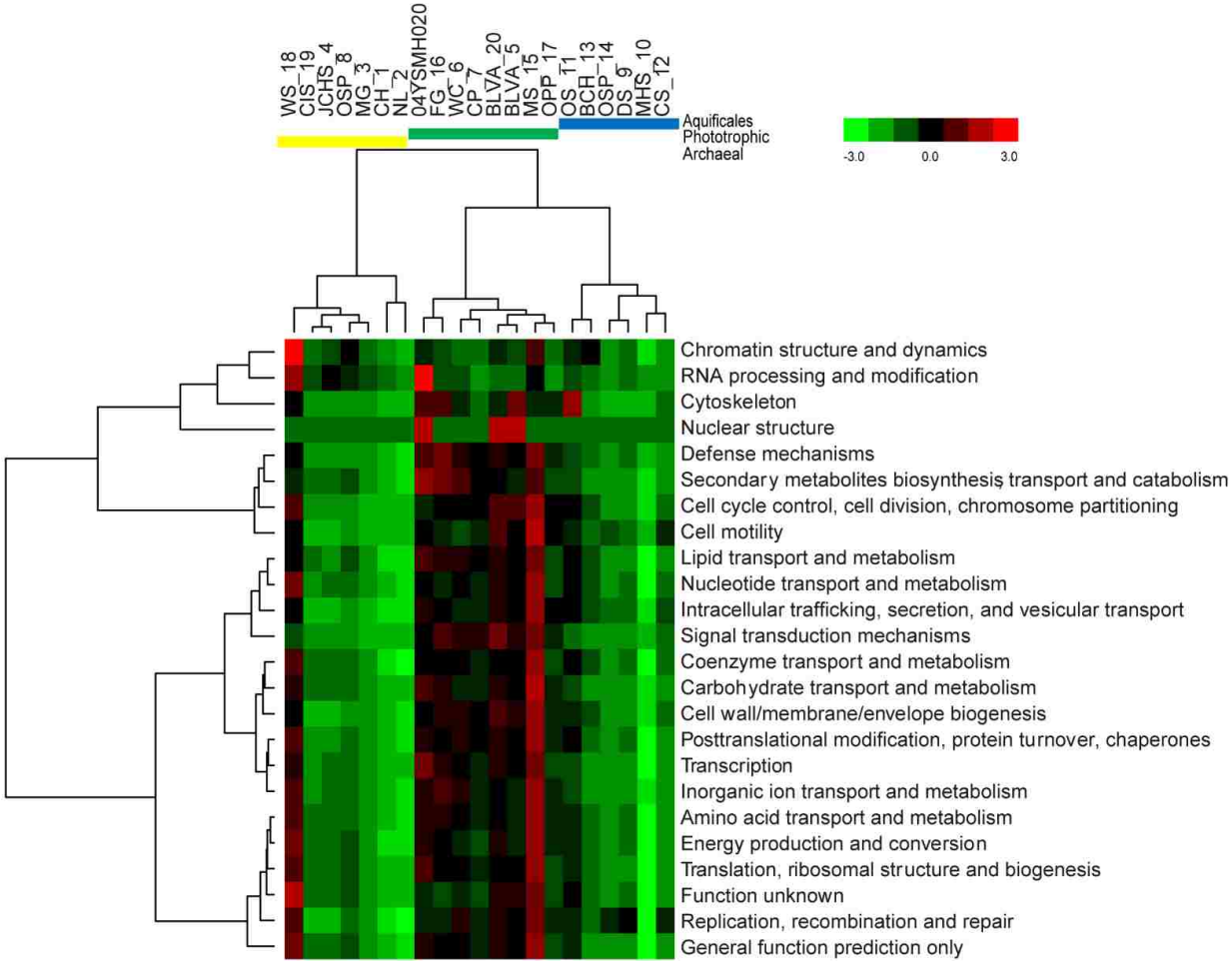


**Fig. S2.** 16S rRNA gene rarefaction curves for four pH 4 samples depicting the observed OTUs (97% similarity) for randomly drawn subsets of 800 sequences per sample

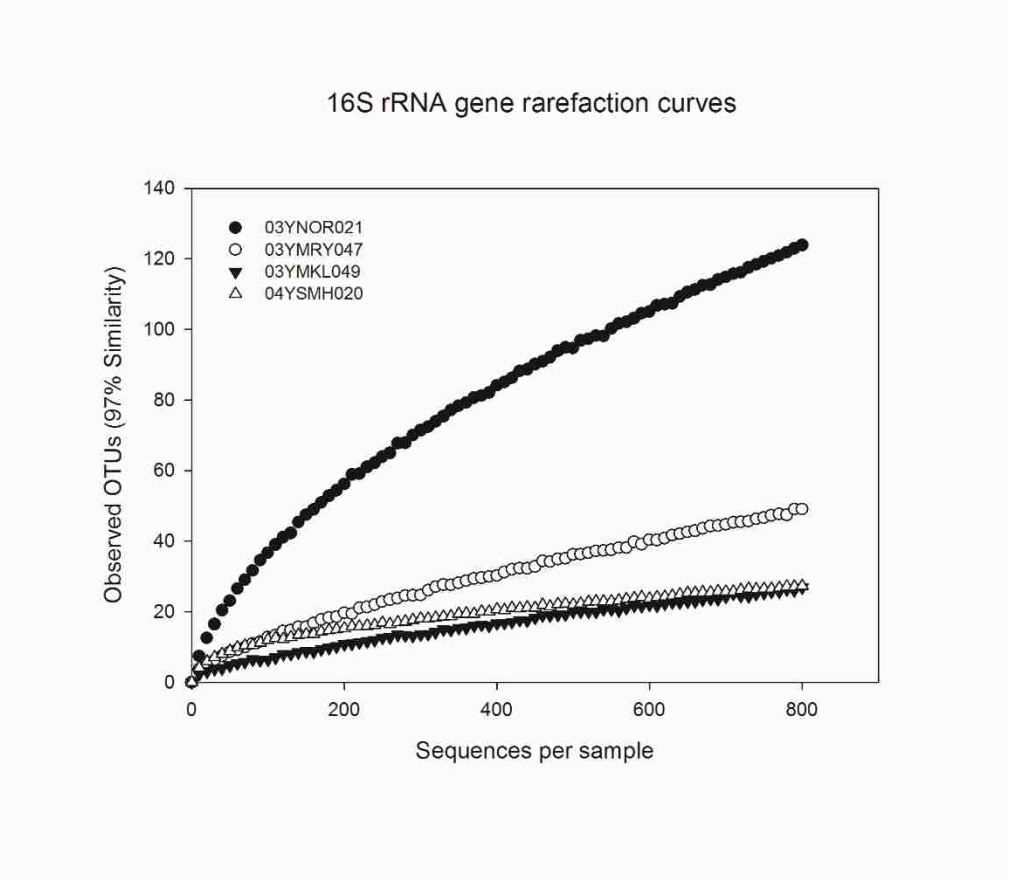


**Fig. S3.** Partial nitrogen pathways identified by KEGG affiliation of the sequences from the site 04YSHM020. Boxes indicate the KEGG identifiers and numbers in gray circles indicate the number of sequences assigned to the KEGG function

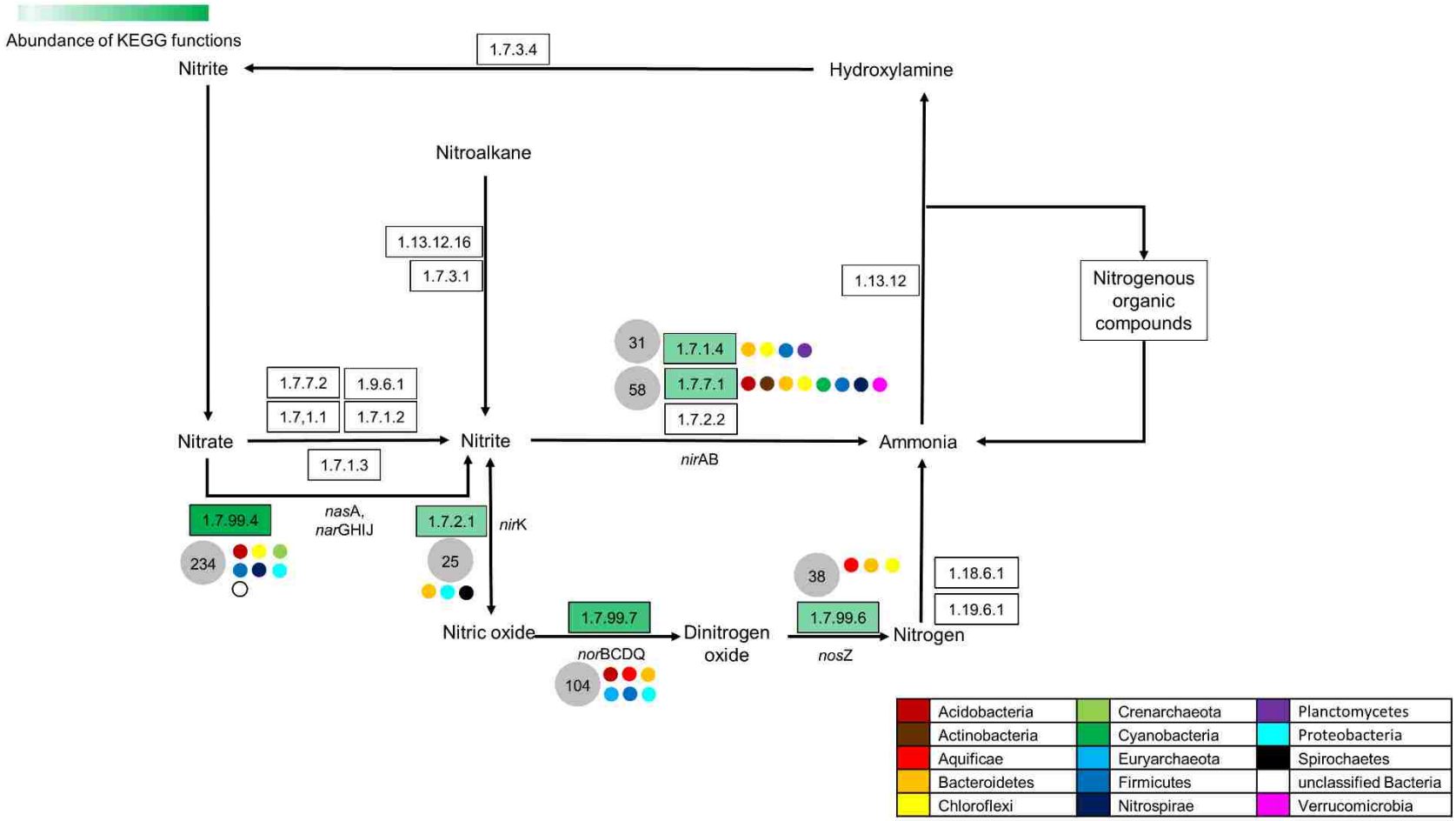
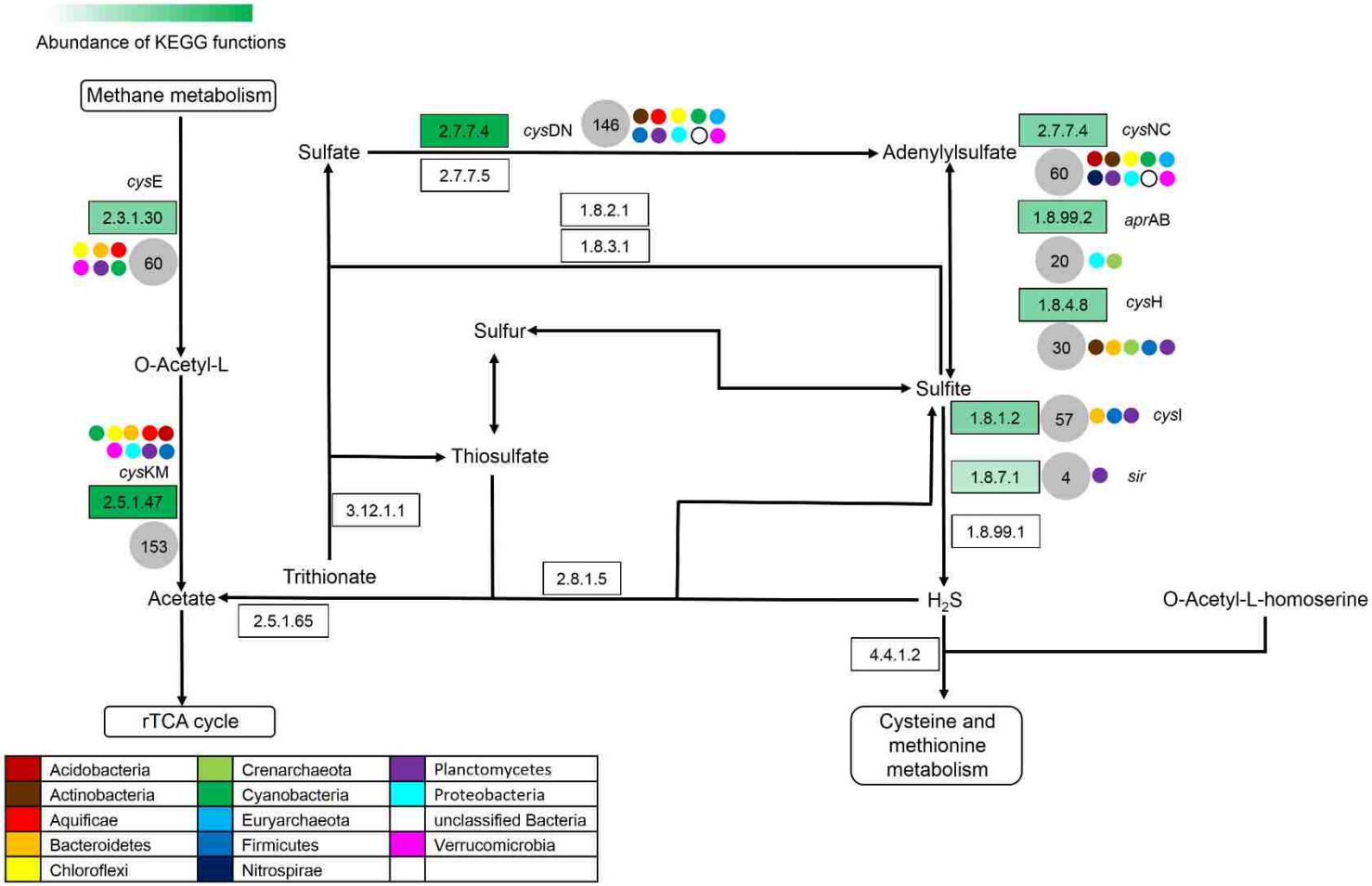


**Fig. S4.** Partial sulfur pathways identified by KEGG affiliation of the sequences from the site 04YSHM020. Boxes indicate the KEGG identifiers and numbers in gray circles indicate the number of sequences assigned to the KEGG function



# Chapter 4

## Changes in the Microbial Community Structure in Response to Environmental Heterogeneity in Filamentous Thermal Springs, Yellowstone National Park

Xiaoben Jiang[a], Kendra Maas[b] and Cristina Takacs-Vesbach[a]

[a]Department of Biology University of New Mexico Albuquerque, NM 87131, USA

[b]Department of Biology, University of British Columbia, Vancouver, V6T1Z4, BC, Canada

**Abstract**

In this study, we investigated microbial diversity and community composition in 10 filamentous thermal springs with similar physiochemical properties in the Shoshone area by using the barcoded amplicon pyrosequencing approach. Non-metric multidimensional scaling ordination based on 16S rRNA genes indicated that the 20 samples fell into three groups. Temperature and pH were the two strongest correlates environmental variables ($R^2$ = 0.7315, P = 0.001; $R^2$ = 0.7979, P = 0.001; respectively), followed by conductivity ($R^2$ = 0.6141, P = 0.007). Even all these samples from the same type of springs, our results revealed that the relatively small variation of environmental factors could shape microbial community composition. Indicator species analysis indicated that each group was dominated by different phyla. Functional analysis of the unique Aquificaceae dominant site indicated that functions related to aerobic respiration and amino acid synthesis were enriched (P < 0.05), suggesting highly microbial activities *in situ*. Energy metabolic pathway analysis showed that rTCA cycle played a dominant role in energy metabolism. The identification of genes coding for nitrogen and sulfur cycling indicated that the microbial population was involved in assimilatory and dissimilatory nitrate reduction, and conversion of sulfate into adenylylsulfate and sulfite.

## Introduction

One of the fundamental goals of ecology is to understand the distribution of organisms, the range of their habitats and the factors influencing on their distributions. Biogeography is the study of the distribution of biodiversity over varying spatial and temporal scales; research on the biogeography of microorganisms can offer insights into microbial speciation, extinction and dispersal. A century ago, Baas-Becking proposed the most famous hypothesis on microbial biogeography: "[E]verything is everywhere, but the environment selects" (Becking 1934). According to Bass-Becking's hypothesis, microorganisms appear to have cosmopolitan distributions due to their vast population size and enormous capacity for long-distance dispersal (Griffin et al., 2002). This view of microbial biogeography has been the paradigm for many years and is supported by some studies, especially for the patterns of the protist distributions (Fenchel and Finlay 2004; Finlay 2002). However, this paradigm of microbial biogeography has been challenged by recent studies using molecular techniques. Geographic distances as a barrier to dispersal do exist at the global scale. For example, globally biogeographic patterns have been detected in the distribution of some thermophiles such as *Sulfolobus* and *Synechococcus* (Martiny et al. 2006). In addition to geographic barriers, like macro-organisms, a variety of environmental factors such as temperature (Everroad et al. 2012; Wang et al. 2013), pH (Fierer and Jackson 2006; Lauber et al. 2009; Van Horn et al. 2013) and salinity (Kunin et al. 2008; Ley et al. 2006) controlling the distribution microorganisms and communities have been observed at different habitats and geographical scales.

Yellowstone National Park (YNP) is one of the largest and most diverse hydrothermal areas in the world, containing more than 14,000 geothermal features that

possess a broad range of multiple environmental gradients such as temperature (40-92 ℃) and pH (1-10) (Fournier 1989; Fournier 2005; Rye and Truesdell 2007). It is generally held that YNP thermal springs have three major community types: (1) phototrophic mats; (2) Aquificales-rich 'filamentous-streamer' communities; and (3) archaeal-dominated sediments (Inskeep et al. 2013b); the classification is based on dominant species and geochemical properties. It has been suggested that environmental variables of local habitats are important in controlling over local distribution microbial taxa and communities in phototrophic mats (Ramsing et al. 2000; van der Meer et al. 2005; Ward et al. 1998a; Ward et al. 1990). Filamentous-streamer communities are characterized by the dominance of Aquificales. However, how environmental heterogeneity influences the diversification of Aquificales-rich 'filamentous-streamer' communities at a local scale is not clear. Previous studies of 16S rRNA gene-based sequencing (Reysenbach et al. 2005), together with the YNP metagenome survey (Inskeep et al. 2013b), have shed light on the distribution of the Aquificales. Despite current breakthroughs in understanding the relationship between environmental variables and metabolic potential in some YNP geothermal features, several important microbial habitat types (e.g., pH, 5-9; temperature, >70 ℃) are yet to be examined (Inskeep et al. 2013b). Moreover, of the previous pilots studies that attempted to link environmental variables with microbial community composition, most concentrated on the vent of each spring and therefore were hindered by not enough number of representative samples or sequencing depth (Coman et al. 2013; Huang et al. 2013; Meyer-Dombard et al. 2005; Purcell et al. 2007; Reigstad et al. 2009; Tobler and Benning 2011). The influence of local environmental heterogeneity

(especially, fine temperature and pH variation) on the structure of Aquificales-rich 'filamentous-streamer' communities requires further investigation.

The goal of the present study was to expand the current understanding of how the environmental heterogeneity structures microbial (in particular, bacterial) communities in filamentous springs of YNP. To address this question, a survey of molecular diversity of bacteria using 16S rRNA gene-based pyrosequencing on 20 sites from 10 circumneutral to alkaline filamentous springs was undertaken. In addition to the 16S rRNA gene marker, we also explored the phylogenetic diversity of the Aquificales using the internal transcribed spacer (ITS) region between the 16S rRNA and 23S rRNA genes, in order to assess the diversity of the Aquificales in these springs. Finally, we provided a functional analysis of metagenomic sequences obtained from a high temperature and pH filamentous community to fill the knowledge gap for this uninvestigated habitat type.

**Materials and Methods**

*Sampling sites*

Ten thermal springs within the Shoshone region of YNP were sampled during 2007 (Figure S1 and Table 1). Within each spring, two visible filamentous samples were collected. One sample was collected from the highest temperature locality where filaments existed, whereas the other sample was collected from the lowest temperature locality where filaments existed. Prior to collecting filaments, spring water for geochemical analysis was filtered through a 0.02 μm filter (Millipore) using sterile 50-mL syringes and preserved as appropriate for the analysis to be performed. Geochemical analyses were carried out using standard USGS methods (McCleskey 2005). Macroscopically visible filamentous samples were collected aseptically into 2-mL

microcentrifuge tubes and preserved in sucrose lysis buffer (Giovannoni et al. 1990). Samples were stored at ambient temperature (~10 to 26 ℃) for up to one days. Once in the laboratory, they were stored at -80 ℃ until DNA extraction.

*DNA extraction, barcoded pyrosequencing and shotgun metagenome sequencing*

Total DNA was extracted from preserved samples following bead-beating disruption in a cetyltrimethylammonium bromide (CTAB) buffer (1% CTAB, 0.75 M NaCl, 50 mM Tris pH 8, 10 mM EDTA) and subsequent phenol-chloroform purification steps as described before (Mitchell and Takacs-Vesbach 2008). Barcoded amplicon pyrosequencing of 16S rRNA genes was performed as described previously (Andreotti et al. 2011; Dowd et al. 2008b; Van Horn et al. 2013) using universal bacterial primers 939F (5‘- TTG ACG GGG GCC CGC AG -3’) and 1492R (5‘-GGT TAC CTT GTT ACG ACT T -3’). For each sample, barcoded amplicon pyrosequencing was also carried out using specific ITS primers 1492F (5’-AAG TCG TAA CAA GGT AAC C-3’) and 115R (5’-GGG TTB CCC CAT TCR G-3’) designed for the Aquificales (Ferrera et al. 2007). We focused solely on the bacterial composition and diversity, because Archaea and Eukarya were not dominant groups in these circumneutral to alkaline thermal systems of YNP (Miller et al. 2009; Ward et al. 1998a). Triplicate reaction mixtures per sample were combined and subsequently purified with an UltraClean™ GelSpin™ DNA Extraction Kit (MoBio Laboratories, Carlsbad, CA, USA). The purified amplicons were quantified using a Nanodrop ND-2000c spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Amplicons from all samples were pooled with equimolar concentrations for pyrosequencing, using Roche titanium reagents and titanium procedures on a Roche 454 GS FLX instrument (454 Life Sciences, Branford, CT, USA).

In addition, we selected one sample, 07SHO09A, collected from the high temperature and pH site for metagenomic sequencing to gain a better understanding of the metabolic potential of the communities. Approximately 500 ng of quality-checked DNA was used for library construction. Metagenome sequencing was performed according to manufacturer's protocol on a 454 GS FLX Titanium platform (454 Life Sciences, Branford, CT, USA).

*16S rRNA and ITS pyrosequencing data processing*

All 16S rRNA gene raw sequences were quality-filtered, denoised and checked for chimeras using the AmpliconNoise tool (Quince et al. 2011) integrated in the Quantitative Insights into Microbial Ecology (QIIME, Ver. 1.8.0) (Caporaso et al. 2010b). Adapters, multiplex identifiers and primers were trimmed from denoised data. The QIIME pipeline was employed to analyze the denoised sequences. Species level operational taxonomic units (OTUs) were identified at 97% similarity using UCLUST (Edgar 2010). The most abundant sequence was picked from each OTU as a representative sequence and aligned using the PyNAST aligner (Caporaso et al. 2010a) and Greengenes database (DeSantis et al. 2006). The phylogenetic affiliations were assigned by the Ribosomal Database Classifier program (Wang et al. 2007). The same settings in QIIME were used for ITS gene sequence data analysis, except representative sequences were aligned using MUSCLE (Edgar 2004) and given taxonomic assignments using BLASTN (Altschul et al. 1997) against the NCBI nucleotide (NT) database. In order to standardize for varying sequencing efforts across samples, all measurements of community structure were performed with randomly drawn subsets of 867 sequences from each sample for the 16S rRNA gene sequencing and 824 sequences for ITS gene

sequencing from each sample, respectively. Alpha diversity indices, including observed OTUs, Good's coverage, Shannon index and phylogenetic diversity, were calculated at the 97% similarity OTU level for both 16S rRNA gene and ITS gene sequences using tools implemented in QIIME (Caporaso et al. 2010b).

*Statistical analysis of bacterial community*

All statistical analyses were conducted using the R (Team 2011) statistical package with the 'vegan' library (Oksanen et al. 2013), unless otherwise noted. For 16S rRNA sequences, the metaMDS and envfit functions in the 'vegan' package were used to test environmental variables correlated with bacterial community composition by fitting all environmental variables measured to the ordination. To assess the similarity of the microbial community structures among the sites, a hierarchical cluster analysis was performed. The significance of the cluster groups was determined using a permutational MANOVA test (Anderson 2001). An indicator species analysis was conducted to obtain taxonomic signatures of each group. Differences in environmental variables among cluster groups were tested using the one-way ANOVA test. Box plots were also built to show the significance of environmental variables among cluster groups.

*Metagenomic assembly and analysis*

Metagenome sequencing reads were quality-filtered and assembled using Newbler 2.6 (Margulies et al. 2005) under default settings. All contigs and remaining singleton reads were submitted to the JGI IMG/M annotation pipeline (Markowitz et al. 2012). Unassembled raw reads were also submitted for annotation on the metagenomics analysis server, MG-RAST (Meta Genome Rapid Annotation using Subsystem Technology, v3.3.9) (Glass et al. 2010), using its default quality control pipeline. Microbial taxa

analysis was conducted *via* the MG-RAST best-hit classification tool against the GenBank (NCBI-NR), M5NR (M5 non-redundant protein) and RefSeq databases using the minimum identity of 60%, e-value cutoff of $10^{-5}$ and minimum alignment length of 50 bp.

*Metabolic mapping of energy metabolism*

Unassembled 07YSHO09A raw reads were annotated by BLASTX (Altschul et al. 1997) against the NCBI non-redundant (NR) protein database using the e-value cutoff of $10^{-5}$. The results of BLASTX were imported into MEtaGenome ANalyzer software (MEGAN v5.5.3) (Huson et al. 2011) and were taxonomically classified using the least common ancestor (LCA) algorithm based on the top 10 BLAST alignments for each read. The pathways of energy metabolism were recognized using KEGG identifiers (Ogata et al. 1999). The sequences in each pathway (oxidative phosphorylation, methane metabolism, nitrogen metabolism, carbon fixation pathways in prokaryotes, carbon fixation in photosynthetic organisms, sulfur metabolism and photosynthesis) were given taxonomical assignments at phylum level. In order to detect important electron transfer processes likely linked to thermal springs, the metabolic pathways involved in carbon fixation pathways in prokaryotes, nitrogen and sulfur metabolism were identified using KEGG identifiers.

*Comparisons to other YNP metagenomes*

A cluster of Orthologous Group (COG) functions of the 07YSHO09A metagenome assembly was compared to available YNP metagenomes (Table S1) (Inskeep et al. 2013b) to better characterize the potential community functions of this sample. Annotated COG functions of these metagenomes were obtained from the IMG/M

server. Prior to quantitative characterization, metagenomic data were normalized by the total number of COG functions detected, weighted by contig depth if assembly information was available. For unassembled singleton reads, a contig depth of one was assumed. COG functions were subsequently classified into COG categories on the IMG/M and a Bray-Curtis dissimilarity matrix was constructed, based on the COG category abundance table. The COG category dissimilarity matrix was subject to hierarchical agglomerative cluster analysis. Moreover, the standardized data were analyzed using a principal coordinates analysis (PCoA). All multivariate comparisons and ordinations were performed using the R (Team 2011) statistical package with 'vegan' (Oksanen et al. 2013) and 'cluster' (Maechler et al. 2013) libraries.

We used the "Function Comparison" tool on the IMG/M to compare the metagenome of 07YSHO09A against metagenomic datasets from other communities, in purpose to determine whether unusually prominent functions existed in this unexplored site. The relative abundance of COG function was calculated based on normalized gene counts and was expressed as D-scores, which were used to calculate the standard variation from the null hypothesis (*i.e.*, relative gene counts in metagenome A = relative gene counts in metagenome B). For each comparison, the P value cutoff for significant D-scores was determined using a false discovery rate of 0.05.

*Sequence Data Submission*

The metagenome of 07YSHO09A is publicly available on the IMG/M (IMG submission ID 13525) and the MG-RAST (ID 4523621.3). The individual sff files of 16S rRNA gene amplicon sequencing were assigned the accession numbers SRX1093556 - SRX1093565, SRX1093570 - SRX1093571 and SRX1093579 - SRX1093586 under

Bioproject PRJNA289700. The individual sff files of ITS rRNA gene amplicon sequencing were assigned the accession numbers SRX1100239 - SRX1100258 and SRX1100241 - SRX1100250 under Bioproject PRJNA290194.

**Results**

*Water chemistry of the studied sites*

A total of 20 filamentous samples were collected from 10 different thermal springs. Water temperature of the sampled thermal springs ranged from 62 to 87 ℃. The pH was from circumneutral to alkaline (6.21-8.88). The conductivity was from 1060 to 1475 (μS/cm). The concentration of hydrogen sulfide in these springs ranged from 0.035 to 4.45 mg $L^{-1}$. The concentration of major anions detailed in Table 1.

*Microbial community composition of 16S rRNA and ITS amplicons*

Collectively, 120,662 16S rRNA gene sequences (mean = 6,033 sequences/site and 80,332 ITS gene sequences (mean = 4016 sequences/site) were obtained after denoising, respectively. Good's coverage calculation indicated that the analyzed sequences covered the diversity of bacteria (Table 3). The Shannon diversity index of 16S rRNA gene sequences ranged from 0.062 to 3.815, with an average of 2.099 (Table 3). The Shannon diversity index of ITS gene sequences ranged from 0.158 to 1.746, with an average of 0.857 (Table 3). The Shannon diversity index, based on 16S rRNA gene sequences, clearly demonstrated that samples collected from low temperature sites had about 79.5% higher species richness than those collected from high temperature sites (P = 0.005, one-tailed *t*-test). However, there was no significant differences in the Shannon diversity between high and low temperature sites (P = 0.308, one-tailed t-test), based on ITS gene sequences.

Bacterial communities of all samples was composed of sequences related to seven major phyla and candidate divisions (Figure 1 and Figure 3A). Candidate division EM3, Aquificae, Proteobacteria and Deinococcus-Thermus were the four most dominant phyla, contributing 33.67%, 24.17%, 19.08% and 14.22% of the total 16S RNA sequences, respectively. The cluster analysis revealed three distinct clusters (groups A, B and C; Figure 3A and Figure 3B). The cluster groups were confirmed by the permutational MANOVA test ($R^2 = 0.78728$, $P < 0.001$). In contrast, microbial communities among springs were not significantly different, according to the permutational MANOVA test ($R^2 = 0.5318$, $P = 0.264$). Each cluster significantly affiliated with one or several phyla, according to the indicator species analysis ($P < 0.05$). For instance, Aquificae and EM3 were associated with group A and group C, respectively, whereas Armatimonadetes, Cyanobacteria, Deinococcus-Thermus were closely related to group B.

*Correlation between bacterial communities and environmental variables*

Non-metric multidimensional scaling (nMDS) ordination based on 16S rRNA gene sequences was shown in Figure 4 (2D stress = 0.1). Several environmental variables were significantly correlated to the first two axes of the ordination (Figure 4). Temperature and pH were the two strongest correlates among all environmental variables ($R^2 = 0.7315$, $P = 0.001$; $R^2 = 0.7979$, $P = 0.001$; respectively), followed by conductivity ($R^2 = 0.6141$, $P = 0.007$). Significant correlations were not detected between ordination and other environmental variables ($P > 0.05$), including $H_2S$, $F^-$, $Cl^-$, $NO_2^-$, $NO_3^-$ $SO_4^{2-}$ and $PO_4^{3-}$. Significant differences in temperature and pH among groups were found, according to the ANOVA test (Both $P<0.05$), but the difference in conductivity among the three groups was close to the marginally significant level ($P = 0.06$). Generally, group

A represented the low temperature and pH habitat, and group B represented the low temperature and high pH habitat, in contrast to group C exhibiting high temperature and pH (Figure 5).

*Metagenome sequencing summary, coverage and classification of the microbial community*

Metagenomic pyrosequencing generated 765,145 reads, with 435 bp by mean length, totaling 333 Mbp from site 07YSHO09A (Table 2). Assembly of these reads yielded 26,440 contigs, with the N50 contig length of 1,162 (Table 2). The result of hierarchical clustering analysis mirrored the previous characterization of YNP thermal springs (Inskeep et al. 2013b), indicating three distinct community groups: (1) archaeal communities; (2) Aquificales communities; and (3) phototrophic communities (Figure 6A). Site 07YSHO09A affiliated with Aquificales communities. Additionally, the group scheme of PCoA ordination confirmed the same group pattern (Figure 6B).

*The taxonomic profile of metagenomics*

As shown in Table S2, taxonomic classification with M5 non-redundant protein (M5NR) database on MG-RAST detected that Bacteria domain (87.71%) was predominant in site 07YSHO09A. The remaining sequences from 07YSHO09A were associated with Archaea (6.36%), Eukaryota (0.01%), Virus (0.05%) and unassigned sequences (8.71%). Among the Bacteria, the most abundant sequences were associated with phylum Aquificae (91.45%). This result was similar to that from 16S rRNA gene sequencing, in which the Aquificae sequences were predominantly retrieved, accounting for 93.61% of total 16S rRNA gene sequences (Figure S2). BLAST analyses using different databases revealed detailed taxonomic information with respect to this

unexplored site. All our results unequivocally demonstrated that the Aquificae predominated over other phyla in site 07YSHO09A, according to GenBank (93.64%), M5NR (90%) and RefSeq (93.31%) database (Figure 7). The family Aquificaceae was the dominant family (~93%) within order Aquificales based on all three databases (Figure 8). Hydrogenothermaceae (~7%) was also detected in this site (Figure 8). With all sequences related to the family Aquificaceae, *Aquifex aeolicus* (36.70%), *Thermocrinis albus* (30.79%) and *Hydrogenobacter thermophilus* (27.89%) were the three most abundant species (Figure 8). Within the family Hydrogenothermaceae, a large proportion of sequences were assigned to *Sulfurihydrogenibium sp. YO3AOP1* (57.50%), followed by *Sulfurihydrogenibium yellowstonense* (29.13%) and *Sulfurihydrogenibium azorense* (8.84%) (Figure 8).

*Functional comparison with other YNP metagenomes*

Both cluster and PCoA analysis unequivocally demonstrated that the functional profile of 07YSHO09A metagenome was similar to metagenomes from the Aquificales communities. Therefore, we compared the functional profile of 07YSHO09A to those of metagenomes from the archaeal group and phototrophic group, respectively. First, we uncovered that 21 COG functions were overrepresented in the 07YSHO09A dataset in at least five of the seven comparisons (Table S3). “Energy production and conversion”, “amino acid transport and metabolism” and “translation, ribosomal structure and biogenesis” were three most abundant COG categories among these functions (19%, 14.3% and 14.3%, respectively) and were dominant among those with the highest enrichment D-scores (Table S3 and Figure 9). In addition, 140 COG functions were enriched within the metagenome of 07YSHO09A in at least five of the six comparisons

(Table S4). “Amino acid transport and metabolism”, “translation, ribosomal structure and biogenesis” and “general function prediction only” were the three most abundant COG categories (15.7%, 10% and 10%, respectively; Table S4 and Figure 10).

*Energy metabolism, rTCA, nitrogen and sulfur metabolism*

Functional assignment of the unassembled 07YSHO09A metagenome provided detailed information about possible functions in this habitat. A total of 59,235 reads were assigned to functions related to energy metabolism using BLASTX against the NCBI-nr database. Among the reads of energy metabolism, most were mapped to the phylum Aquificae and involved in a variety of pathways, such as oxidative phosphorylation, methane metabolism, nitrogen metabolism and carbon fixation pathways in prokaryotes, carbon fixation in photosynthetic organisms, sulfur metabolism and photosynthesis (Figure S3). A total of 14,281, 14,636 and 277 reads were assigned to carbon fixation pathways in prokaryotes (rTCA), nitrogen metabolism and sulfur metabolism, respectively. Major KEGG functional categories and the unique hits mapped to each categories were listed (Table S5, Table S6 and Table S7).

**Discussion**

*Microbial community environmental heterogeneity*

The first objective of this study was to examine the effects of local environmental heterogeneity on microbial communities from filamentous springs in the Shoshone area. Aquificales have been found to be dominant in high temperature (> 65 ℃), neutral-alkaline, filamentous springs in YNP (Hamamura et al. 2013; Meyer-Dombard et al. 2005; Meyer-Dombard et al. 2011; Reysenbach et al. 2000b; Takacs-Vesbach et al. 2013); as well as in other similar terrestrial hydrothermal systems (China, Iceland and

Japan) (Kato et al. 2004; Nakagawa and Fukui 2002; Takacs et al. 2001; Wang et al. 2013). The 10 filamentous springs in this project are considered ideal habitats for Aquificales communities (Inskeep et al. 2013b). However, the order Aquificales is not the exclusively dominant taxonomic group in these springs, despite the fact that considerable sequences related to the Aquificales have been previously recovered from springs in Shoshone (Takacs-Vesbach et al. 2008). Previous reports correlating thermal springs around the world have shown that pronounced temperature and pH variation can affect the structure and distribution of microbial communities among springs or within a spring (Allewalt et al. 2006; Everroad et al. 2012; Inskeep et al. 2010; Miller et al. 2009; Purcell et al. 2007; Wang et al. 2013). Here, we observed that environmental heterogeneity contributed to the considerable variation of microbial taxonomic assemblages and richness, despite all these sites belonging to the same type of springs within the Shoshone area (Figure 3).

Based on the correlations between the phylogenetic composition and physicochemical conditions, the 20 sites can be classified into three distinct groups, designated here as group A, group B and group C (Figure 3A and Figure 3B), which suggest significant differences in temperature and pH (Figure 5). In group A, microbial communities were mainly composed of sequences related to the Aquificae (Figure 3A and Figure 3B). Within the phylum Aquificae, all sequences were related to the family Hydrogenothermaceae (77.5% of all Aquificae sequences) and Aquificaceae within the order Aquificales (22.5% of all Aquificae sequences). The sites dominated by the Aquificales represent relatively low temperature and low pH sites among the 20 sites (Figure 5). Six of seven sites in group A were dominated by members of the family

Hydrogenothermaceae (Figure 3B and Figure 4), with a temperature range of 62 - 78 ℃ and a pH range of 6.21-7.10. By contrast, the only site dominated by the family Aquificaceae was from the 83 ℃ and pH 8.83 site (Table 1, Figure 3B and Figure 4), accounting for 87.4% of all the Aquificaceae sequences. The further classification using the ITS gene indicated that genus *Sulfurihydrogenibium* was the major taxonomic group in the family Hydrogenothermaceae (Figure 2). Additionally, metagenomic sequencing revealed that genus *Aquifex*, *Thermocrinis* and *Hydrogenobacter* were three most abundant bacterial groups in the family Aquificaceae (Figure 8). The genus *Sulfurihydrogenibium* associated with the family Hydrogenothermaceae normally inhabits circumneutral sulfidic springs of YNP (Hugenholtz et al. 1998; Reysenbach et al. 2005; Reysenbach et al. 2000a). The two well-studied groups in the family Aquificaceae, the *Hydrogenobaculum* spp. and *Thermocrinis*-like group, appear to be divergent. These two groups normally colonize two distinct niches in thermal springs. The *Hydrogenobaculum* spp. generally dominate in pH springs (pH<4), whereas the *Thermocrinis*-like organisms dominate in higher pH systems (pH 6-9) (Takacs-Vesbach et al. 2013). Given that the concentration of hydrogen sulfide is not significantly different among cluster groups by ANOVA test, the differences in temperature and pH among sites seem to play a more important role in controlling the distribution of the Hydrogenothermaceae and Aquificaceae in the Shoshone. Furthermore, the distribution of members from these two families is consistent with previous observed temperature and pH range elsewhere (Blank et al. 2002; Deckert et al. 1998; Hugenholtz et al. 1998; Pitulle et al. 1994; Reysenbach et al. 2005; Reysenbach et al. 2000a; Takacs-Vesbach et al. 2013).

In contrast to group A, group B was detected at sites with higher pH. The bacterial communities in group B contained a high proportion of sequences associated with the Deinococcus-Thermus (35.6% of total Bacteria in group B, mainly genus *Thermus*. The presence of *Thermus* in group B coincides with the widespread distribution of this group in global thermal springs with temperature of 50 - 99 ℃ and circumneutral to alkaline pH (Boomer et al. 2009; Cole et al. 2013; Costa et al. 2009; Gumerov et al. 2011; Lin et al. 2002; Meyer-Dombard et al. 2011; Purcell et al. 2007; Tobler and Benning 2011; Wang et al. 2013). Despite the widespread distribution of the *Thermus*, the dominance of this genus in terrestrial thermal springs is rarely observed, except for a few springs in Iceland (Tobler and Benning 2011) and Tibet (Wang et al. 2013). The dominance of the genus *Thermus* usually is associated with higher available organic carbon content (Wang et al. 2013). Given the abundance of *Thermus*, coupled with the evidence that the highest abundance of autotrophic bacteria, Cyanobacteria (mainly genus *Gloeobacter*) found in group B (Figure 3A), we can reason that sites in group B correlate with more available organic carbon.

The candidate division EM3 was the indicator of group C (Figure 3A and Figure 4), exhibiting high temperature and pH among these three clustering groups. In group C, the average temperature is 82 ℃ and the average pH is 8.01, which mirror the temperature and pH at Octopus Spring where the EM3 were originally reported (Reysenbach et al. 1994). However, the EM3 seem to be well adapted to different environments. Aside from thermal springs (Meyer-Dombard et al. 2011; Zhang et al. 2013), the EM3 are also present in wastewaters (Hatamoto et al. 2007; Sekiguchi 2006). Previously, molecular phylogenetic studies suggested that the candidate division EM3

was closely related to the phylum Thermotogae (Dunfield et al. 2012; Rinke et al. 2013). The recent study implies that this candidate division is the deepest lineage of domain Bacteria (Colman et al. in prep). Despite this phylogenetic information, little is known about the metabolic pathway and the ecology of the EM3 due to the lack of cultivated representatives. Although one-way ANOVA analysis revealed that the difference in conductivity among three groups was at the marginally significant level ($P = 0.06$), the conductivity was significantly correlated with the first two axes of the ordination (Figure 4). In particular, group C showed a strong linkage to high conductivity (Figure 4). Conductivity is an important environmental variable to control the global distribution of bacterial community composition (Lozupone and Knight 2007). Abrupt salinity gradients often occur in habitats such as estuaries, wetlands, salt marshes and coastal lagoons (Jeffries et al. 2012). We observed that the higher salinity seemed to favor the EM3, suggesting that the fine salinity heterogeneity has an important influence on the local distribution of bacteria.

*Functional profile of 07YSHO09A metagenome*

The second goal of this research was to characterize the functional profile of the microbial community in 07YSHO09A. Although our cluster analysis based on 16S rRNA gene sequencing indicated that 07YSHO09A clustered with sites dominated by the Aquificales (Figure 3B), it separated itself from the other Aquificales-dominated sites on nMDS plot (Figure 4) due to the dominance of the family Aquificaceae rather than the family Hydrogenothermaceae. It is noteworthy that the predominance of the Aquificaceae is also supported by metagenomic analysis (Figure 8).

The metagenome of 07YSHO09A provides the functional profile in a relatively low elemental sulfur, high temperature and pH site of YNP. In addition, the comparison of 07YSHO09A metagenomic data to those from sites associated with other communities can reveal the information concerning prominent functional capabilities of the 07YSHO09A (Figure 9, Figure 10, Table S3 and Table S4). Among all the overrepresented functions, many were responsible for a variety of energy production and conversion such as carbohydrate metabolism. Particularly, some functions found overrepresented in both comparisons (Table S3 and Table S4), such as Succinyl-CoA synthetase (COG 0045) and NADH:ubiquinone oxidoreductase (COG1007, COG0649, COG1009), can be related directly to aerobic respiration. Succinyl-CoA synthetase is an enzyme that catalyzes the reversible reaction of succinyl-CoA to succinate, which plays a key role as one of the catalysts involved in the citric acid cycle (TCA) (Voet and Voet 2011) and the rTCA cycle (Williams et al. 2006). Moreover, NADH:ubiquinone oxidoreductase is one of the main entry enzymes for the respiratory chains of myriad organisms (Friedrich et al. 1998).

Heme copper oxidases (subunit 1 of terminal oxidase complexes, COG0843) were enriched, when compared to metagenomes from phototrophic communities (Table S4). Heme copper oxidases that can catalyze the reduction of $O_2$ to $H_2O$ are good indicators of the potential for aerobic respiration (Garc á-Horsman et al. 1994b; Pereira et al. 2004). This result, together with the predominance of the Aquificaceae populations implies that the majority of these organisms respire oxygen. The sample 07YSHO09A was collected from visible filaments. Here, the overabundance of functions contributing to amino acid synthesis (e.g., COG0019, COG0436, COG0620 and COG0498) and concomitant with

the function implicated in cell division (e.g., COG0445) suggest high productivity existing in the filaments (Table S3 and Table S4). Higher microbial productivity produces more intermediates, which may expand the available niches and increase the microbial diversity and competition in the whole spring (Everroad et al. 2012). For instance, the presence of the function involved in multidrug resistance (COG0841), which is to export antibiotics and toxic molecules (Piddock 2006) (Table 3), suggests that microorganisms can defend against toxic compounds produced by competitors. The over overrepresented COG functions (Table S4) also included functions related to DNA repair (COG0419 and COG0497). In contrast to phototrophic communities, the high proportion of these COGs can be indicative of the need to respond to the high temperature and pH environment of site 07YSHO09A. Among the enriched functions present in 07YSHO09A, we observed the protein involved with trace element detoxification (COG1055, Na+/H+ antiporter NhaD and related arsenite permeases, Table S4). Arsenic and mercury are two common toxic constituents originating from YNP geothermal features (Ball et al. 2002; Boyd et al. 2009; Stauffer and Thompson 1984). The COG1055 coded by gene *ars*B is responsible for transporting arsenite out of the cell under toxic conditions (Mukhopadhyay et al. 2002; Stolz et al. 2006). Previous reports have indicated the widespread distribution of *ars*B gene in all three main communities of YNP (Inskeep et al. 2013b; Inskeep et al. 2010). Compared to phototrophic communities, the significant abundance of COG1055 has implications for higher activities in arsenic efflux *in situ*.

Potential primary production strategies in thermal springs are of primary interest because the species richness and primary production in thermophilic environments are considered to be low compared to mesophilic environments (Brock 1994). Aquificae,

mainly Aquificales dominant in site 07YSHO09A, are known for their roles as primary producers in hydrothermal systems (Blank et al. 2002; Eder and Huber 2002; Harmsen et al. 1997; Reysenbach et al. 2000a; Reysenbach et al. 1999; Spear et al. 2005; Takacs-Vesbach et al. 2013; Yamamoto et al. 1998). The rTCA cycle was well represented compared to the nitrogen cycle and the sulfur cycle (Figure S4, Figure S5 and Figure S6), according to the KEGG database. The rTCA cycle is the reversed process of the oxidative TCA cycle, providing an alternative strategy to the Calvin cycle for fixing $CO_2$. During the rTCA cycle, three molecules of $CO_2$ are fixed, forming acetyl-CoA and subsequently pyruvate, the precursors of all other central metabolites. Three key enzymes are involved in the TCA cycle, including ATP citrate lyase, pyruvate ferredoxin oxidoreductase and 2-oxoglutarate ferredoxin oxidoreductase (He et al. 2013; Hugler et al. 2007). Although ATP citrate lyase was not detected by KEGG pathway analysis, pyruvate ferredoxin oxidoreductase and 2-oxoglutarate ferredoxin oxidoreductase were prevalent in 07YSHO09A. In addition, despite the low abundance, ATP citrate lyase was detected using the COG database (see above discussion). Apart from $CO_2$ fixation *via* rTCA cycle, the microbial community in site 07YSHO09A contained the evidence of the reductive acetyl-CoA pathway for $CO_2$ (marked by the gene *ACSS*, acetyl-CoA synthetase, Table S5 and Figure S4). These *ACSS* sequences were phylogenetically related to Aquificae, Betaproteobacteria, Chloroflexi, Crenarchaeota, Firmicutes and Thermodesulfobacteria. The genes coding for rTCA and reductive acetyl-CoA pathway have been previously reported to be prevalent in other YNP springs (*e.g.*, Crater Hills, Mammoth Hot Springs) (Inskeep et al. 2010), where the Aquificales predominate.

Pathways related to nitrogen and sulfur metabolism are of importance in thermal springs, where terminal electronic acceptors other than $O_2$ (*e.g.*, $CO_2$, element sulfur, ferric iron, nitrate, sulfate and thiosulfate) normally play more important roles (Hall et al. 2008; Inskeep et al. 2010; Jiménez et al. 2012). Genes coding for dissimilatory nitrate reductases (*e.g.*, *nar*I) and nitric oxide reductases (*e.g.*, *nor*B, *nor*C, *nor*Q) occurred in 07YSHO09A (Table S6 and Figure S5). The presence of genes coding for these two enzymes has been previously reported in an Archaeal-dominated spring of YNP (Joseph's Coat Hot Spring) (Inskeep et al. 2010). However, genes coding for nitrite reductase (*nir*K or *nir*S) were not present. During the process of dissimilatory nitrate reduction, a nitrite reductase (*nir*K or *nir*S) is the requisite for producing NO, which serves as a substrate for nitric oxide reductase to produce $N_2O$ (Gonzalez et al. 2006; Moreno-Vivián et al. 1999; Richardson et al. 2001). The lack of detection of nitrite reductase in 07YSHO09A, coupled with evidence of the absence of nitrous oxide reductase (*nos*Z), which converts $N_2O$ to $N_2$, suggests that denitrification might be not an important metabolic pathway in this site. Most microorganisms, with the exception of a few members in the Euryarchaeota, can only fix nitrogen up to 64 ℃ (Mehta and Baross 2006). Given the *in situ* temperature of site 07YSHO09A (Table 1), not surprisingly, genes (*e.g.*, *nif*H) contributing to nitrogen fixation were not detected (Table S6, Figure S5). The absence of genes (*e.g.*, *amo*A) coding for ammonium monooxygenase is in agreement with the previous finding in other YNP thermal springs (Inskeep et al. 2010), suggesting that biological ammonia oxidation may not be predominant in YNP thermal springs (Table S6, Figure S5).

In the sulfur cycle, genes related to the conversion of sulfate into adenylylsulfate and to the further generation of sulfite were detected (Table S7 and Figure S6). The environmental *apr*A and *apr*B sequences coding for adenosine-5'-phosphosulfate (APS) reductase (Apr) exhibited closest matches to members of Thermodesulfobacteria and Thermoprotei. Based on the current model of dissimilatory sulfate reduction and sulfur oxidation in prokaryotes, APS reductase (Apr) is a key enzyme. During the process of sulfate reduction, the function of Apr is to convert APS to sulfite, once sulfate is activated to APS by ATP-sulfurylase at the expense of ATP (Meyer and Kuever 2008). The alpha subunits of Apr enzymes are present in all known sulfate reducing and most of sulfur oxidizing prokaryotes (Meyer and Kuever 2008). For instance, the environmental *apr*A reads found in site 07YSHO09A indicate high identity to those in the genome of hyperthermophilic sulfate reducer (e-value < $10^{-80}$), *Thermodesulfobacterium geofontis* OPF15 (Table S7), an isolate from the YNP thermal spring (Elkins et al. 2013; Hamilton-Brehm et al. 2013). *Thermodesulfobacterium geofontis* can reduce sulfate with an optimal temperature of 83 ℃. Genes coding for the subsequent reduction of sulfite to $H_2S$ were not detected (Table S7 and Figure S6), either because not enough sequencing depth was accomplished or due to the fact that they were absent or present in very low abundance.

**Conclusions**

Phylogenetic analysis of amplicon sequencing data from 10 circumneutral to alkaline thermal springs suggests that the variation of microbial taxonomic assemblages within the same type of community (*i.e.*, 'filamentous streamer' community) can result from the fine physicochemical heterogeneity of local habitats. Here, we observed that the bacterial community composition changed in response to the relatively small variation of

temperature, pH and conductivity ($R^2$ = 0.7315, P = 0.001; $R^2$ = 0.7979, P = 0.001; $R^2$ = 0.6141, P = 0.007; respectively). Phylogenetic and functional analyses of a high temperature and pH, Aquificales-dominated site (07YSHO09A) revealed that the family Aquificaceae was exclusively abundant. Compared to Archaeal-dominated communities and phototrophic communities in YNP, the metagenome of 07YSHO09A indicated that functions such as aerobic respiration and amino acid synthesis were enriched, implying active microbial activities *in situ*. The identification of genes coding for the rTCA cycle suggests that rTCA cycle plays a dominant role in *in situ* energy metabolism. The identification of genes coding for nitrogen and sulfur cycling indicated that the microbial population was involved in assimilatory and dissimilatory nitrate reduction, and conversion of sulfate into adenylylsulfate and sulfite. The results from this study suggest that the taxonomic diversity in Aquificales-rich 'filamentous-streamer' communities may not be homogenous. The dominant bacterial groups can change in response to the small variation of environmental factors.

**References**

Acosta-Martínez V, Dowd S, Sun Y, Allen V (2008) Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biol Biochem 40:2762-2770

Aditiawati P, Yohandini H, Madayanti F, Akhmaloka (2009) Microbial diversity of acidic hot spring (kawah hujan B) in geothermal field of kamojang area, west java-indonesia. Open Microbiol J 3:58-66

Allewalt JP, Bateson MM, Revsbech NP, Slack K, Ward DM (2006) Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the Octopus spring microbial mat community of Yellowstone National Park. Appl Environ Microbiol 72:544-550

Almarsdottir AR, Taraceviz A, Gunnarsson I, Orlygsson J (2010) Hydrogen production from sugars and complex biomass by *Clostridium* species, $AK_{14}$, isolated from Icelandic hot spring. Icelandic Agricultural Sciences:61-71

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Amarouche-Yala S, Benouadah A, El Ouahab Bentabet A, López-García P (2014) Morphological and phylogenetic diversity of thermophilic cyanobacteria in Algerian hot springs. Extremophiles 18:1035-1047

Amend JP, Shock EL (2001) Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. FEMS Microbiol Rev 25:175-243

Aminin ALN, Warganegara FM, Aditiawati P, Akhmaloka (2008) Simple enrichment and independent cultures to expand bacterial community analysis from gedongsongo hot spring. J Biosci Bioeng 106:211-214

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecol 26:32-46

Andreotti R, Perez de Leon A, Dowd S, Guerrero F, Bendele K, Scoles G (2011) Assessment of bacterial diversity in the cattle tick Rhipicephalus (Boophilus) microplus through tag-encoded pyrosequencing. BMC Microbiol 11:6

Ball JW, McCleskey RB, Nordstrom DK (2010) Water-chemistry data for selected springs, geysers, and streams in Yellowstone National Park, Wyoming, 2006–2008. US Geological Survey Open-File Report 2010–1192:109 p

Ball JW, McCleskey RB, Nordstrom DK, Holloway JM, Verplanck PL (2002) Water-chemistry data for selected springs, geysers, and streams in Yellowstone National Park, Wyoming, 1999-2000. US Geological Survey Open-File Report:02-382

Ball JW, Nordstrom DK, Cunningham KM, Schoonen MAA, Xu Y, DeMonge JM (1998) Water-chemistry and on-site sulfur-speciation data for selected springs in Yellowstone National Park, Wyoming, 1994-1995. US Geological Survey Open-File Report:98-574

Barns SM, Delwiche CF, Palmer JD, Pace NR (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proceedings of the National Academy of Sciences 93:9188-9193

Barns SM, Fundyga RE, Jeffries MW, Pace NR (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proceedings of the National Academy of Sciences of the United States of America 91:1609-1613

Bauld J, Brock T (1973) Ecological studies of Chloroflexis, a gliding photosynthetic bacterium. Archiv Mikrobiol 92:267-284

Becking LGMB (1934) Geobiologie, of inleiding tot de milieukunde: met literatuurlijst en ind. W.P. Van Stockum & Zoon, Hague, Netherlands

Beh M, Strauss G, Huber R, Stetter K-O, Fuchs G (1993) Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaebacterium *Thermoproteus neutrophilus*. Arch Microbiol 160:306-311

Blank CE, Cady SL, Pace NR (2002) Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. Appl Environ Microbiol 68:5123-5135

Bonch-Osmolovskaya EA, Miroshnichenko ML, Slobodkin AI, Sokolova TG, Karpov GA, Kostrikina NA, Zavarzina DG, Prokof'eva MI, Rusanov, II, Pimenov NV (1999) Biodiversity of anaerobic lithotrophic prokaryotes in terrestrial hot springs of Kamchatka. Microbiology 68:343-351

Boomer SM, Lodge DP, Dutton BE, Pierson B (2002) Molecular Characterization of Novel Red Green Nonsulfur Bacteria from Five Distinct Hot Spring Communities in Yellowstone National Park. Appl Environ Microbiol 68:346-355

Boomer SM, Noll KL, Geesey GG, Dutton BE (2009) Formation of multilayered photosynthetic biofilms in an alkaline thermal spring in Yellowstone National Park, Wyoming. Appl Environ Microbiol 75:2464-2475

Boyd E, Hamilton T, Wang J, He L, Zhang C (2013) The role of tetraether lipid composition in the adaptation of thermophilic archaea to acidity. Frontiers in Microbiology 4

Boyd ES, Hamilton TL, Spear JR, Lavin M, Peters JW (2010) [FeFe]-hydrogenase in Yellowstone National Park: evidence for dispersal limitation and phylogenetic niche conservatism. Isme j 4:1485-1495

Boyd ES, Jackson RA, Encarnacion G, Zahn JA, Beard T, Leavitt WD, Pi Y, Zhang CL, Pearson A, Geesey GG (2007) Isolation, characterization, and ecology of sulfur-respiring Crenarchaea inhabiting acid-sulfate-chloride-containing geothermal springs in Yellowstone National Park. Appl Environ Microbiol 73:6669-6677

Boyd ES, King S, Tomberlin JK, Nordstrom DK, Krabbenhoft DP, Barkay T, Geesey GG (2009) Methylmercury enters an aquatic food web through acidophilic microbial mats in Yellowstone National Park, Wyoming. Environ Microbiol 11:950-959

Brochier-Armanet C, Gribaldo S, Forterre P (2012) Spotlight on the Thaumarchaeota. ISME J 6:227-230

Brock T, Brock K, Belly R, Weiss R (1972) *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Microbiol 84:54-68

Brock TD (1967) Life at high temperatures. Science 158:1012-1019

Brock TD (1971) Bimodal distribution of pH values of thermal springs of the world. Geological Society of America Bulletin 82:1393-1394

Brock TD (1994) Life at high temperatures. In. Yellowstone National Park, WY: Yellowstone Association for Natural Science, History & Education, Inc.,

Brock TD, Brock TD (1978) Thermophilic microorganisms and life at high temperatures. Springer-Verlag New York

Bryant DA, Costas AMG, Maresca JA, Chew AGM, Klatt CG, Bateson MM, Tallon LJ, Hostetler J, Nelson WC, Heidelberg JF, Ward DM (2007) *Candidatus* Chloracidobacterium thermophilum: An Aerobic Phototrophic Acidobacterium. Science 317:523-526

Burgess E, Unrine J, Mills G, Romanek C, Wiegel J (2012) Comparative geochemical and microbiological characterization of two thermal pools in the Uzon caldera, Kamchatka, Russia. Microb Ecol 63:471-489

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266-267

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010b) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-336

Cavaletti L, Monciardini P, Bamonte R, Schumann P, Rohde M, Sosio M, Donadio S (2006) New lineage of filamentous, spore-forming, gram-positive bacteria from soil. Appl Environ Microbiol 72:4360-4369

Cavicchioli R (2002) Extremophiles and the search for extraterrestrial life. Astrobiology 2:281-292

Chang Y-J, Land M, Hauser L, Chertkov O, Del Rio TG, Nolan M, Copeland A, Tice H, Cheng J-F, Lucas S (2011) Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21$^T$). Standards in genomic sciences 5:97

Clarke K, Ainsworth M (1993) A method of linking multivariate community structure to environmental variables. Marine Ecology-Progress Series 92:205-205

Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. Aust J Ecol 18:117-143

Cole JK, Peacock JP, Dodsworth JA, Williams AJ, Thompson DB, Dong H, Wu G, Hedlund BP (2013) Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. ISME J 7:718-729

Colman DR, Thomas R, Maas KR, Takacs-Vesbach CD (2015) Detection and analysis of elusive members of a novel and diverse archaeal community within a thermal spring streamer consortium. Extremophiles 19:307-313

Coman C, Drugă B, Hegedus A, Sicora C, Dragoş N (2013) Archaeal and bacterial diversity in two hot spring microbial mats from a geothermal region in Romania. Extremophiles 17:523-534

Connon SA, Koski AK, Neal AL, Wood SA, Magnuson TS (2008) Ecophysiology and geochemistry of microbial arsenic oxidation within a high arsenic, circumneutral hot spring system of the Alvord Desert. FEMS Microbiol Ecol 64:117-128

Costa K, Navarro J, Shock E, Zhang C, Soukup D, Hedlund B (2009) Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. Extremophiles 13:447-459

Cox A, Shock EL, Havig JR (2011) The transition to microbial photosynthesis in hot spring ecosystems. Chemical Geology 280:344-351

Davis BM (1897) The vegetation of the hot springs of Yellowstone Park. Science 6:145

De León KB, Gerlach R, Peyton BM, Fields MW (2013) Archaeal and bacterial communities in three alkaline hot springs in Heart Lake Geyser Basin, Yellowstone National Park. Frontiers in Microbiology 4

Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olsen GJ, Swanson RV (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature 392:353-358

Del Rio TG, Abt B, Spring S, Lapidus A, Nolan M, Tice H, Copeland A, Cheng J-F, Chen F, Bruce D (2010) Complete genome sequence of *Chitinophaga pinensis* type strain (UQM $2034^{T}$). Standards in genomic sciences 2:87

Deppenmeier U (2002) Redox-driven proton translocation in methanogenic Archaea. CMLS, Cell Mol Life Sci 59:1513-1533

Dequiedt S, Thioulouse J, Jolivet C, Saby N, Lelievre M, Maron PA, Martin MP, Prévost-Bouré NC, Toutain B, Arrouays D (2009a) Biogeographical patterns of soil bacterial communities. Environmental microbiology reports 1:251-255

Dequiedt S, Thioulouse J, Jolivet C, Saby NPA, Lelievre M, Maron P-A, Martin MP, Prévost-Bouré NC, Toutain B, Arrouays D, Lemanceau P, Ranjard L (2009b)

Biogeographical patterns of soil bacterial communities. Environmental Microbiology Reports 1:251-255

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069-5072

Donachie S, Christenson B, Kunkel D, Malahoff A, Alam M (2002) Microbial community in acidic hydrothermal waters of volcanically active White Island, New Zealand. Extremophiles 6:419-425

Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R (2012) Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genomics 13:162

Dowd S, Callaway T, Wolcott R, Sun Y, McKeehan T, Hagevoort R, Edrington T (2008a) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol 8:125

Dowd SE, Sun Y, Wolcott RD, Domingo A, Carroll JA (2008b) Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) for microbiome studies: bacterial diversity in the ileum of newly weaned Salmonella-infected pigs. Foodborne pathogens and disease 5:459-472

Drammeh I (2010) Biodiversity of thermophiles in Sungai Klah and Ulu Slim hot springs. Universiti Teknologi Malaysia

Drobner E, Huber H, Rachel R, Stetter KO (1992) *Thiobacillus plumbophilus* spec. nov., a novel galena and hydrogen oxidizer. Arch Microbiol 157:213-217

Dunfield PF, Tamas I, Lee KC, Morgan XC, McDonald IR, Stott MB (2012) Electing a candidate: a speculative history of the bacterial phylum OP10. Environ Microbiol 14:3069-3080

Eder W, Huber R (2002) New isolates and physiological properties of the Aquificales and description of *Thermocrinis albus* sp. nov. Extremophiles 6:309-318

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461

Elkins JG, Hamilton-Brehm SD, Lucas S, Han J, Lapidus A, Cheng JF, Goodwin LA, Pitluck S, Peters L, Mikhailova N, Davenport KW, Detter JC, Han CS, Tapia R, Land ML, Hauser L, Kyrpides NC, Ivanova NN, Pagani I, Bruce D, Woyke T, Cottingham RW (2013) Complete genome sequence of the hyperthermophilic sulfate-reducing bacterium *Thermodesulfobacterium geofontis* OPF15T. Genome Announc 1:e0016213

Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, Lapidus A, Goltsman E, Barry K, Koonin EV, Hugenholtz P, Kyrpides N, Wanner G, Richardson P, Keller M, Stetter KO (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. Proceedings of the National Academy of Sciences of the United States of America 105:8102-8107

Engel AS, Johnson LR, Porter ML (2013) Arsenite oxidase gene diversity among Chloroflexi and Proteobacteria from El Tatio Geyser Field, Chile. FEMS Microbiol Ecol 83:745-756

Entsminger G (2012) EcoSim Professional: Null modeling software for ecologists, Version 1. Acquired Intelligence Inc., Kesey-Bear, & Pinyon Publishing. Montrose, CO 81403. http://wwwgaryentsmingercom/ecosim/indexhtm

Everroad RC, Otaki H, Matsuura K, Haruta S (2012) Diversification of bacterial community composition along a temperature gradient at a thermal spring. Microbes Environ 27:374-381

Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1-10

Fardeau M-L, Barsotti V, Cayol J-L, Guasco S, Michotey V, Joseph M, Bonin P, Ollivier B (2010) *Caldinitratiruptor microaerophilus*, gen. nov., sp. nov. isolated from a French hot spring (Chaudes-Aigues, Massif Central): a novel cultivated facultative microaerophilic anaerobic thermophile pertaining to the *Symbiobacterium* branch within the Firmicutes. Extremophiles 14:241-247

Fenchel TOM, Finlay BJ (2004) The ubiquity of small species: patterns of local and global diversity. Bioscience 54:777-784

Ferrera I, Longhorn S, Banta A, Liu Y, Preston D, Reysenbach AL (2007) Diversity of 16S rRNA gene, ITS region and aclB gene of the *Aquificales*. Extremophiles 11:57-64

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences of the United States of America 103:626-631

Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. Science 296:1061-1063

Fournier RO (1989) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. Annual Review of Earth and Planetary Sciences 17:13-53

Fournier RO (2005) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. In: Inskeep WP, TR M (eds) Geothermal biology and geochemistry in Yellowstone National Park. Thermal Biology Institute, Montana State University, Bozeman, MT, pp 4-30

Friedrich T, Abelmann A, Brors B, Gu énebaut V, Kintscher L, Leonard K, Rasmussen T, Scheide D, Schlitt A, Schulte U (1998) Redox components and structure of the respiratory NADH: ubiquinone oxidoreductase (complex I). Biochimica et Biophysica Acta (BBA)-Bioenergetics 1365:215-219

Fujiwara S (2002) Extremophiles: developments of their special functions and potential resources. J Biosci Bioeng 94:518-525

Fukami T, Morin PJ (2003) Productivity-biodiversity relationships depend on the history of community assembly. Nature 424:423-426

Garc á-Horsman JA, Barquera B, Rumbley J, Ma J, Gennis RB (1994a) The superfamily of heme-copper respiratory oxidases. Journal of Bacteriology 176:5587-5600

Garc á-Horsman JA, Barquera B, Rumbley J, Ma J, Gennis RB (1994b) The superfamily of heme-copper respiratory oxidases. J Bacteriol 176:5587

Giovannoni SJ, DeLong EF, Schmidt TM, Pace NR (1990) Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. Appl Environ Microbiol 56:2572-2575

Giovannoni SJ, Revsbech NP, Ward DM, Castenholz RW (1987) Obligately phototrophic Chloroflexus: primary production in anaerobic hot spring microbial mats. Arch Microbiol 147:80-87

Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F (2010) Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harbor Protocols 2010:pdb.prot5368

Gonzalez P, Correia C, Moura I, Brondino C, Moura J (2006) Bacterial nitrate reductases: molecular and biological aspects of nitrate reduction. Journal of inorganic biochemistry 100:1015-1023

Good IJ (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40:237-264

Gotelli NJ, McCabe DJ (2002) Species co-occurrence: a meta-analysis of JM Diamond's assembly rules model. Ecology 83:2091-2096

Graber JR, Kirshtein J, Speck M, Reysenbach A-L (2001) Community structure along a thermal gradient in a stream near Obsidian Pool, Yellowstone National Park. In: Thermophiles: Biodiversity, Ecology, and Evolution. Springer, New York, pp 81-91

Gumerov VM, Mardanov AV, Beletsky AV, Bonch-Osmolovskaya EA, Ravin NV (2011) Molecular analysis of microbial diversity in the Zavarzin Spring, Uzon Caldera, Kamchatka. Microbiology 80:244-251

Hall JR, Mitchell KR, Jackson-Weaver O, Kooser AS, Cron BR, Crossey LJ, Takacs-Vesbach CD (2008) Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes. Appl Environ Microbiol 74:4910-4922

Hamamura N, Meneghin J, Reysenbach AL (2013) Comparative community gene expression analysis of Aquificales-dominated geothermal springs. Environ Microbiol 15:1226-1237

Hamilton-Brehm S, Gibson R, Green S, Hopmans E, Schouten S, Meer MJ, Shields J, Damsté JS, Elkins J (2013) *Thermodesulfobacterium geofontis* sp. nov., a hyperthermophilic, sulfate-reducing bacterium isolated from Obsidian Pool, Yellowstone National Park. Extremophiles 17:251-263

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Micro 10:497-506

Harmsen H, Prieur D, Jeanthon C (1997) Group-specific 16S rRNA-targeted oligonucleotide probes to identify thermophilic bacteria in marine hydrothermal vents. Appl Environ Microbiol 63:4061-4068

Hatamoto M, Imachi H, Yashiro Y, Ohashi A, Harada H (2007) Diversity of anaerobic microorganisms involved in long-chain fatty acid degradation in methanogenic sludges as revealed by RNA-based stable isotope probing. Appl Environ Microbiol 73:4119-4127

He P, Liu Y, Yue W, Huang X (2013) Key genes expression of reductive tricarboxylic acid cycle from deep-sea hydrothermal chemolithoautotrophic *Caminibacter profundus* in response to salinity, pH and $O_2$. Acta Oceanol Sin 32:35-41

Horner-Devine MC, Silver JM, Leibold MA, Bohannan BJ, Colwell RK, Fuhrman JA, Green JL, Kuske CR, Martiny JB, Muyzer G, Ovreas L, Reysenbach AL, Smith VH (2007) A comparison of taxon co-occurrence patterns for macro- and microorganisms. Ecology 88:1345-1353

Hou W, Wang S, Dong H, Jiang H, Briggs BR, Peacock JP, Huang Q, Huang L, Wu G, Zhi X, Li W, Dodsworth JA, Hedlund BP, Zhang C, Hartnett HE, Dijkstra P, Hungate BA (2013) A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. PLoS ONE 8:e53350

Huang Q, Dong C, Dong R, Jiang H, Wang S, Wang G, Fang B, Ding X, Niu L, Li X, Zhang C, Dong H (2011) Archaeal and bacterial diversity in hot springs on the Tibetan Plateau, China. Extremophiles 15:549-563

Huang Q, Jiang H, Briggs BR, Wang S, Hou W, Li G, Wu G, Solis R, Arcilla CA, Abrajano T, Dong H (2013) Archaeal and bacterial diversity in acidic to circumneutral hot springs in the Philippines. FEMS Microbiol Ecol 85:n/a-n/a

Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography vol 32. Princeton University Press,

Huber H, Stetter KO (1998) Hyperthermophiles and their possible potential in biotechnology. J Biotechnol 64:39-52

Huber R, Stetter KO (2001) Aquificales. In: eLS. John Wiley & Sons, Ltd,

Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998) Novel division level bacterial diversity in a Yellowstone hot spring. J Bacteriol 180:366-376

Hugler M, Huber H, Molyneaux SJ, Vetriani C, Sievert SM (2007) Autotrophic $CO_2$ fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum *Aquificae*: evidence for two ways of citrate cleavage. Environ Microbiol 9:81-92

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377-386

Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Res 21:1552-1560

Im W-T, Hu Z-Y, Kim K-H, Rhee S-K, Meng H, Lee S-T, Quan Z-X (2012) Description of *Fimbriimonas ginsengisoli* gen. nov., sp. nov. within the *Fimbriimonadia* class nov., of the phylum *Armatimonadetes*. Antonie Van Leeuwenhoek 102:307-317

Inagaki F, Motomura Y, Doi K, Taguchi S, Izawa E, Lowe DR, Ogata S (2001) Silicified microbial community at Steep Cone hot spring, Yellowstone National Park. Microbes Environ 16:125-130

Inagaki F, Takai K, Hirayama H, Yamato Y, Nealson KH, Horikoshi K (2003) Distribution and phylogenetic diversity of the subsurface microbial community in a Japanese epithermal gold mine. Extremophiles 7:307-317

Inskeep WP, Jay ZJ, Herrgard MJ, Kozubal MA, Rusch DB, Tringe SG, Macur RE, Jennings R, Boyd ES, Spear JR, Roberto FF (2013a) Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats

demonstrate Linkages between metabolic potential and geochemistry. Front Microbiol 4:95

Inskeep WP, Jay ZJ, Tringe SG, Herrgard MJ, Rusch DB (2013b) The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front Microbiol 4:67

Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, Macur RE, Hamamura N, Jennings Rd, Fouke BW, Reysenbach A-L, Roberto F, Young M, Schwartz A, Boyd ES, Badger JH, Mathur EJ, Ortmann AC, Bateson M, Geesey G, Frazier M (2010) Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS ONE 5:e9773

Itoh T, Suzuki K-i, Sanchez PC, Nakase T (1999) *Caldivirga maquilingensis* gen. nov., sp. nov., a new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. Int J Syst Bacteriol 49:1157-1163

Jackson CR, Langner HW, Donahoe-Christiansen J, Inskeep WP, McDermott TR (2001) Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. Environ Microbiol 3:532-542

Jahnke LL, Eder W, Huber R, Hope JM, Hinrichs K-U, Hayes JM, Des Marais DJ, Cady SL, Summons RE (2001) Signature lipids and stable carbon isotope analyses of Octopus Spring hyperthermophilic communities compared with those of Aquificales representatives. Appl Environ Microbiol 67:5179-5189

Jeffries T, Seymour J, Newton K, Smith R, Seuront L, Mitchell J (2012) Increases in the abundance of microbial genes encoding halotolerance and photosynthesis along a sediment salinity gradient. Biogeosciences 9:815-825

Jiang H, Dong CZ, Huang Q, Wang G, Fang B, Zhang C, Dong H (2012) Actinobacterial diversity in microbial mats of five hot springs in central and central-eastern Tibet, China. Geomicrobiol J 29:520-527

Jiménez DJ, Andreote FD, Chaves D, Montaña JS, Osorio-Forero C, Junca H, Zambrano MM, Baena S (2012) Structural and functional insights from the metagenome of an acidic hot spring microbial planktonic community in the Colombian Andes. PLoS ONE 7:e52069

Kanokratana P, Chanapan S, Pootanakit K, Eurwilaichitr L (2004) Diversity and abundance of Bacteria and Archaea in the Bor Khlueng Hot Spring in Thailand. J Basic Microbiol 44:430-444

Kato K, Kobayashi T, Yamamoto H, Nakagawa T, Maki Y, Hoaki T (2004) Microbial mat boundaries between chemolithotrophs and phototrophs in geothermal hot spring effluents. Geomicrobiol J 21:91-98

Klatt CG, Inskeep WP, Herrgard MJ, Jay ZJ, Rusch DB, Tringe SG, Niki Parenteau M, Ward DM, Boomer SM, Bryant DA, Miller SR (2013) Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. Front Microbiol 4:106

Klatt CG, Wood JM, Rusch DB, Bateson MM, Hamamura N, Heidelberg JF, Grossman AR, Bhaya D, Cohan FM, Kuhl M, Bryant DA, Ward DM (2011) Community

ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. ISME J 5:1262-1278

Kozubal MA, Macur RE, Jay ZJ, Beam JP, Malfatti SA, Tringe SG, Kocar BD, Borch T, Inskeep WP (2012) Microbial iron cycling in acidic geothermal springs of yellowstone national park: integrating molecular surveys, geochemical processes, and isolation of novel fe-active microorganisms. Front Microbiol 3:109

Kozubal MA, Romine M, Jennings Rd, Jay ZJ, Tringe SG, Rusch DB, Beam JP, McCue LA, Inskeep WP (2013) Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. ISME J 7:622-634

Kumar B, Trivedi P, Kumar Mishra A, Pandey A, Palni LMS (2004) Microbial diversity of soil from two hot springs in Uttaranchal Himalaya. Microbiol Res 159:141-146

Kunin V, Raes J, Harris JK, Spear JR, Walker JJ, Ivanova N, von Mering C, Bebout BM, Pace NR, Bork P, Hugenholtz P (2008) Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. Mol Syst Biol 4

Kvist T, Ahring BK, Westermann P (2007) Archaeal diversity in Icelandic hot springs. FEMS Microbiol Ecol 59:71-80

Lau C-y (2007) Ecology of natural thermophilic communities in the Tibet autonomous region (China). University of Hong Kong

Lau M, Aitchison J, Pointing S (2009a) Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. Extremophiles 13:139-149

Lau MC, Aitchison JC, Pointing SB (2009b) Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. Extremophiles 13:139-149

Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75:5111-5120

Law CJ, Maloney PC, Wang D-N (2008) Ins and outs of major facilitator superfamily antiporters. Annu Rev Microbiol 62:289

Lee KC-Y, Dunfield PF, Morgan XC, Crowe MA, Houghton KM, Vyssotski M, Ryan JLJ, Lagutin K, McDonald IR, Stott MB (2011) *Chthonomonas calidirosea* gen. nov., sp. nov., an aerobic, pigmented, thermophilic micro-organism of a novel bacterial class, *Chthonomonadetes* classis nov., of the newly described phylum *Armatimonadetes* originally designated candidate division OP10. Int J Syst Evol Microbiol 61:2482-2490

Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM, Maresca JA, Bryant DA, Sogin ML, Pace NR (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. Appl Environ Microbiol 72:3685-3695

Lin L, Chen C, Peng Q, Ben K, Zhou Z (2002) *Thermus rehai* sp. nov., isolated from Rehai of Tengchong, Yunnan Province, China. J Basic Microbiol 42:337-344

Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences 104:11436-11440

Macur RE, Jay ZJ, Taylor WP, Kozubal MA, Kocar BD, Inskeep WP (2013) Microbial community structure and sulfur biogeochemistry in mildly-acidic sulfidic geothermal springs in Yellowstone National Park. Geobiology 11:86-99

Madigan MT (2003) Anoxygenic phototrophic bacteria from extreme environments. Photosynthesis Res 76:157-171

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2013) Cluster: cluster analysis basics and extensions.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380

Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Grechkin Y, Ratner A, Jacob B, Pati A, Huntemann M, Liolios K, Pagani I, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Res 40:D123-D129

Marteinsson VT, Hauksdóttir S, Hobel CFV, Kristmannsdóttir H, Hreggvidsson GO, Kristjánsson JK (2001) Phylogenetic diversity analysis of subterranean hot springs in Iceland. Appl Environ Microbiol 67:4242-4248

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, Reysenbach A-L, Smith VH, Staley JT (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Micro 4:102-112

McCleskey RB (2005) Water-chemistry data for selected hot springs, geysers, and streams in Yellowstone National Park, Wyoming, 2001-2002. US Geological Survey,

McCleskey RB, Ball JW, Nordstrom DK, Holloway JM, Taylor HE (2005) Water-chemistry data for selected hot springs, geysers, and streams in Yellowstone National Park, Wyoming 2001-2002. US Geological Survey Open-File Report 2004-1316

Mehta MP, Baross JA (2006) Nitrogen fixation at 92 degrees C by a hydrothermal vent archaeon. Science 314:1783-1786

Mesbah N, Wiegel J (2008) Life at extreme limits. In: Incredible anaerobes: from physiology to genomics to fuels. pp 44-57

Meyer-Dombard DR, Shock EL, Amend JP (2005) Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. Geobiology 3:211-227

Meyer-Dombard DR, Swingley W, Raymond J, Havig J, Shock EL, Summons RE (2011) Hydrothermal ecotones and streamer biofilm communities in the Lower Geyser Basin, Yellowstone National Park. Environ Microbiol 13:2216-2231

Meyer B, Kuever J (2008) Homology modeling of dissimilatory APS reductases (AprBA) of sulfur-oxidizing and sulfate-reducing prokaryotes. PloS one 3:e1514

Mielke PW, Berry KJ, Brockwell PJ, Williams JS (1981) A class of nonparametric tests based on multiresponse permutation procedures. Biometrika 68:720-724

Miller SR, Castenholz RW (2000) The evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*. J Phycol 36:48-48

Miller SR, Strong AL, Jones KL, Ungerer MC (2009) Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. Appl Environ Microbiol 75:4565-4572

Mirete S, de Figueras CG, González-Pastor JE (2011) Diversity of Archaea in Icelandic hot springs based on 16S rRNA and chaperonin genes. FEMS Microbiol Ecol 77:165-175

Mitchell K, Takacs-Vesbach C (2008) A comparison of methods for total community DNA preservation and extraction from various thermal environments. J Ind Microbiol Biotechnol 35:1139-1147

Mitchell KR (2009) Controls on microbial community structure in thermal environments : exploring bacterial diversity and the relative influence of geochemistry and geography. University of New Mexico

Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F (1999) Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. J Bacteriol 181:6573-6584

Mukhopadhyay R, Rosen BP, Phung LT, Silver S (2002) Microbial arsenic: from geocycles to genes and enzymes. FEMS Microbiol Rev 26:311-325

Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M, Daniel R (2011) Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. PloS one 6:e17000

Nakagawa T, Fukui M (2002) Phylogenetic characterization of microbial mats and streamers from a Japanese alkaline hot spring with a thermal gradient. The Journal of general and applied microbiology 48:211-222

Nobre MF, Truper HG, Costa MSD (1996) Transfer of *Thermus ruber* (Loginova et al. 1984), *Thermus silvanus* (Tenreiro et al. 1995), and *Thermus chliarophilus* (Tenreiro et al. 1995) to *Meiothermus* gen. nov. as *Meiothermus ruber* comb. nov., *Meiothermus silvanus* comb. nov., and *Meiothermus chliarophilus* comb. nov., respectively, and emendation of the genus T*hermus*. Int J Syst Bacteriol 46:604-606

Nubel U, Bateson MM, Vandieken V, Wieland A, Kuhl M, Ward DM (2002) Microscopic examination of distribution and phenotypic properties of phylogenetically diverse Chloroflexaceae-related bacteria in hot spring microbial mats. Appl Environ Microbiol 68:4593-4603

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27:29-34

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H Vegan: community ecology package, 2012. R package version:2.0-5

Oksanen J, Blanchet FG, Kindt R, Oksanen MJ, Suggests M (2013) Package 'vegan'. Community ecology package Version 2:0-0

Otaki H, Everroad RC, Matsuura K, Haruta S (2011) Production and consumption of hydrogen in hot spring microbial mats dominated by a filamentous anoxygenic photosynthetic bacterium. Microbes and environments/JSME 27:293-299

Otaki H, Everroad RC, Matsuura K, Haruta S (2012) Production and consumption of hydrogen in hot spring microbial mats dominated by a filamentous anoxygenic photosynthetic bacterium. Microbes and environments / JSME 27:293-299

Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276:734-740

Pagaling E, Grant W, Cowan D, Jones B, Ma Y, Ventosa A, Heaphy S (2012) Bacterial and archaeal diversity in two hot spring microbial mats from the geothermal region of Tengchong, China. Extremophiles 16:607-618

Papke RT, Ramsing NB, Bateson MM, Ward DM (2003) Geographical isolation in hot spring cyanobacteria. Environ Microbiol 5:650-659

Pereira M, Bandeiras T, Fernandes A, Lemos R, Melo AP, Teixeira M (2004) Respiratory chains from aerobic thermophilic prokaryotes. J Bioenerg Biomembr 36:93-105

Piddock LJ (2006) Multidrug-resistance efflux pumps - not just for resistance. Nat Rev Microbiol 4:629-636

Pierson B, Castenholz R (1974) A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. Arch Microbiol 100:5-24

Pitulle C, Yang Y, Marchiani M, Moore ER, Siefert JL, Aragno M, Jurtshuk P, Jr., Fox GE (1994) Phylogenetic position of the genus Hydrogenobacter. Int J Syst Bacteriol 44:620-626

Purcell D, Sompong U, Yim LC, Barraclough TG, Peerapornpisal Y, Pointing SB (2007) The effects of temperature, pH and sulphide on the community structure of hyperthermophilic streamers in hot springs of northern Thailand. FEMS Microbiol Ecol 60:456-466

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 2011 Jan 28 12:38

Ramsing NB, Ferris MJ, Ward DM (2000) Highly ordered vertical structure of *Synechococcus* populations within the one-millimeter-thick photic zone of a hot spring Cyanobacterial mat. Appl Environ Microbiol 66:1038-1049

Rappé MS, Giovannoni SJ (2003) The unclutured microbial majority. Annu Rev Microbiol 57:369-394

Reigstad LJ, Jorgensen SL, Schleper C (2009) Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka. ISME J 4:346-356

Reysenbach A-L, Banta A, Daly SC, Mitchel K, Lalonde S, Konhauser K, Rodman A, Rusterholtz K, Takacs-Vesbach C (2005) Aquificales in Yellowstone National Park. Geothermal Biology and Geochemistry in YNP 1:129-142

Reysenbach A-L, Hamamura N, Podar M, Griffiths E, Ferreira S, Hochstein R, Heidelberg J, Johnson J, Mead D, Pohorille A, Sarmiento M, Schweighofer K, Seshadri R, Voytek MA (2009) Complete and draft genome sequences of six members of the *Aquificales*. J Bacteriol 191:1992-1993

Reysenbach AL, Banta AB, Boone DR, Cary SC, Luther GW (2000a) Microbial essentials at hydrothermal vents. Nature 404:835

Reysenbach AL, Ehringer H, Hershberger K (2000b) Microbial diversity at 83˚C in Calcite Springs, Yellowstone National Park: another environment where the *Aquificales* and "Korarchaeota" coexist. Extremophiles 4:61-67

Reysenbach AL, Seitzinger S, Kirshtein J, McLaughlin E (1999) Molecular constraints on a high-temperature evolution of early life. Biological Bulletin 196:367-372

Reysenbach AL, Wickham GS, Pace NR (1994) Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. Appl Environ Microbiol 60:2113-2119

Rhoads DD, Wolcott RD, Sun Y, Dowd SE (2012) Comparison of culture and molecular identification of bacteria in chronic wounds. International Journal of Molecular Sciences 13:2535-2550

Richardson D, Berks B, Russell D, Spiro S, Taylor C (2001) Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. CMLS, Cell Mol Life Sci 58:165-178

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM,

Hugenholtz P, Woyke T (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431-437

Robinson CK, Wierzchos J, Black C, Crits-Christoph A, Ma B, Ravel J, Ascaso C, Artieda O, Valea S, Roldán M (2014) Microbial diversity and the presence of algae in halite endolithic communities are correlated to atmospheric moisture in the hyper-arid zone of the Atacama Desert. Environ Microbiol

Rogers KL, Amend JP (2005) Archaeal diversity and geochemical energy yields in a geothermal well on Vulcano Island, Italy. Geobiology 3:319-332

Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. Nature 409:1092-1101

Ruff-Roberts AL, Kuenen JG, Ward DM (1994) Distribution of cultivated and uncultivated cyanobacteria and Chloroflexus-like bacteria in hot spring microbial mats. Appl Environ Microbiol 60:697-704

Rye RO, Truesdell AH (2007) The question of recharge to the deep thermal reservoir underlying the geysers and hot springs of Yellowstone National Park. In: al. LAMe (ed) Integrated Geoscience Studies in the Greater Yellowstone Area: Volcanic, Tectonic, and Hydrothermal Processes in the YellowstoneGeoecosystem. U.S. Geol. Surv. Prof. Pap., pp 1717, 1235-1270

Sanford RA, Cole JR, Tiedje JM (2002) Characterization and Description of Anaeromyxobacter dehalogenans gen. nov., sp. nov., an Aryl-Halorespiring Facultative Anaerobic Myxobacterium. Appl Environ Microbiol 68:893-900

Sangkhobol V, Skerman VBD (1981) *Chitinophaga*, a new genus of chitinolytic myxobacteria. Int J Syst Bacteriol 31:285-293

Sayeh R, Birrien J, Alain K, Barbier G, Hamdi M, Prieur D (2010) Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches. Extremophiles 14:501-514

Schoenborn L, Yates PS, Grinton BE, Hugenholtz P, Janssen PH (2004) Liquid serial dilution is inferior to solid media for isolation of cultures representative of the phylum-level diversity of soil bacteria. Appl Environ Microbiol 70:4363-4366

Schwartz E, Van Horn DJ, Buelow HN, Okie JG, Gooseff MN, Barrett JE, Takacs-Vesbach CD (2014) Characterization of growing bacterial populations in McMurdo Dry Valley soils through stable isotope probing with 18O-water vol 89. vol 2.

Segawa T, Takeuchi N, Ushida K, Kanda H, Kohshima S (2010) Altitudinal changes in a bacterial community on Gulkana Glacier in Alaska. Microbes and environments / JSME 25:171-182

Sekiguchi Y (2006) Yet-to-be cultured microorganisms relevant to methane fermentation processes. Microbes Environ 21:1-15

Sen R, Maiti N (2014) Genomic and functional diversity of bacteria isolated from hot springs in Odisha, India. Geomicrobiol J 31:541-550

Shiba H, Kawasumi T, Igarashi Y, Kodama T, Minoda Y (1985) The $CO_2$ assimilation via the reductive tricarboxylic acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. Arch Microbiol 141:198-203

Shock EL, Holland M, Meyer-Dombard DAR, Amend JP (2005) Geochemical sources of energy for microbial metabolism in hydrothermal ecosystems: Obsidian Pool,

Yellowstone National Park. In: Inskeep WP, McDermott TR (eds) Geothermal Biology and Geochemistry in Yellowstone National Park. Montana State University, Bozeman, MT, pp 95-109

Siering PL, Wolfe GV, Wilson MS, Yip AN, Carey CM, Wardman CD, Shapiro RS, Stedman KM, Kyle J, Yuan T, Van Nostrand JD, He Z, Zhou J (2013) Microbial biogeochemistry of Boiling Springs Lake: a physically dynamic, oligotrophic, low-pH geothermal ecosystem. Geobiology

Sievert SM, Ziebis W, Kuever J, Sahm K (2000) Relative abundance of Archaea and Bacteria along a thermal gradient of a shallow-water hydrothermal vent quantified by rRNA slot-blot hybridization. Microbiology 146:1287-1293

Signori CN, Thomas F, Enrich-Prast A, Pollery RCG, Sievert SM (2014) Microbial diversity and community structure across environmental gradients in Bransfield Strait, Western Antarctic Peninsula. Frontiers in Microbiology 5:647

Skirnisdottir S, Hreggvidsson GO, Hjörleifsdottir S, Marteinsson VT, Petursdottir SK, Holst O, Kristjansson JK (2000) Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. Appl Environ Microbiol 66:2835-2841

Song Z-Q, Chen J-Q, Jiang H-C, Zhou E-M, Tang S-K, Zhi X-Y, Zhang L-X, Zhang C-L, Li W-J (2010) Diversity of Crenarchaeota in terrestrial hot springs in Tengchong, China. Extremophiles 14:287-296

Song Z-Q, Wang F-P, Zhi X-Y, Chen J-Q, Zhou E-M, Liang F, Xiao X, Tang S-K, Jiang H-C, Zhang CL, Dong H, Li W-J (2013) Bacterial and archaeal diversities in Yunnan and Tibetan hot springs, China. Environ Microbiol 15:1160-1175

Spanevello MD, Yamamoto H, Patel BKC (2002) *Thermaerobacter subterraneus* sp. nov., a novel aerobic bacterium from the Great Artesian Basin of Australia, and emendation of the genus Thermaerobacter. Int J Syst Evol Microbiol 52:795-800

Spear JR, Walker JJ, McCollom TM, Pace NR (2005) Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. Proceedings of the National Academy of Sciences of the United States of America 102:2555-2560

Stauffer RE, Thompson JM (1984) Arsenic and antimony in geothermal waters of Yellowstone National Park, Wyoming, USA. Geochim Cosmochim Acta 48:2547-2561

Stolz JF, Basu P, Santini JM, Oremland RS (2006) Arsenic and selenium in microbial metabolism. Annu Rev Microbiol 60:107-130

Stone L, Roberts A (1990) The checkerboard score and species distributions. Oecologia 85:74-79

Stott MB, Crowe MA, Mountain BW, Smirnova AV, Hou S, Alam M, Dunfield PF (2008) Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. Environ Microbiol 10:2030-2041

Swingley WD, Meyer-Dombard DAR, Shock EL, Alsop EB, Falenski HD, Havig JR, Raymond J (2012) Coordinating environmental genomics and geochemistry reveals metabolic transitions in a hot spring ecosystem. PLoS ONE 7:e38108

Takacs-Vesbach C, Inskeep WP, Jay ZJ, Herrgard MJ, Rusch DB, Tringe SG, Kozubal MA, Hamamura N, Macur RE, Fouke BW, Reysenbach AL, McDermott TR, Jennings R, Hengartner NW, Xie G (2013) Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct

geothermal channels reveals specialization of three aquificales lineages. Front Microbiol 4:84

Takacs-Vesbach C, Mitchell K, Jackson-Weaver O, Reysenbach A-L (2008) Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. Environ Microbiol 10:1681-1689

Takacs CD, Ehringer M, Favre R, Cermola M, Eggertsson G, Palsdottir A, Reysenbach A-L (2001) Phylogenetic characterization of the blue filamentous bacterial community from an Icelandic geothermal spring. FEMS Microbiol Ecol 35:123-128

Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33-36

Team RDC (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Tekere M, Lotter A, Olivier J, Jonker N, Venter SN (2011) Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa.

Tobler DJ, Benning LG (2011) Bacterial diversity in five Icelandic geothermal waters: temperature and sinter growth rate effects. Extremophiles 15:473-485

Valverde A, Tuffin M, Cowan D (2012) Biogeography of bacterial communities in hot springs: a focus on the actinobacteria. Extremophiles 16:669-679

van der Meer MTJ, Schouten S, Bateson MM, Nübel U, Wieland A, Kühl M, de Leeuw JW, Sinninghe Damsté JS, Ward DM (2005) Diel variations in carbon metabolism

by green nonsulfur-like bacteria in alkaline siliceous hot spring microbial mats from Yellowstone National Park. Appl Environ Microbiol 71:3978-3986

van der Meer MTJ, Schouten S, Sinninghe Damsté JS, de Leeuw JW, Ward DM (2003) Compound-specific isotopic fractionation patterns suggest different carbon metabolisms among chloroflexus-like bacteria in hot-spring microbial mats. Appl Environ Microbiol 69:6000-6006

Van Horn DJ, Okie JG, Buelow HN, Gooseff MN, Barrett JE, Takacs-Vesbach CD (2014) Soil microbial responses to increased moisture and organic resources along a salinity gradient in a polar desert. Appl Environ Microbiol 80:3034-3043

Van Horn DJ, Van Horn ML, Barrett JE, Gooseff MN, Altrichter AE, Geyer KM, Zeglin LH, Takacs-Vesbach CD (2013) Factors controlling soil microbial biomass and bacterial diversity and community composition in a cold desert ecosystem: role of geographic scale. PLoS ONE 8:e66103

Voet D, Voet JG (2011) Biochemistry, 4-th Edition. Wiley & Sons, New York,

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261-5267

Wang S, Hou W, Dong H, Jiang H, Huang L, Wu G, Zhang C, Song Z, Zhang Y, Ren H, Zhang J, Zhang L (2013) Control of temperature on microbial community structure in hot springs of the Tibetan Plateau. PLoS ONE 8:e62901

Ward DM, Ferris MJ, Nold SC, Bateson MM (1998a) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62:1353-1370

Ward DM, Ferris MJ, Nold SC, Bateson MM (1998b) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62:1353-1370

Ward DM, Weller R, Bateson MM (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345:63-65

Weber CF, King GM (2010) Distribution and diversity of carbon monoxide-oxidizing bacteria and bulk bacterial communities across a succession gradient on a Hawaiian volcanic deposit. Environ Microbiol 12:1855-1867

Weltzer M, Miller S (2013) Division-Specific Differences in Bacterial Community Assembly of Two Alkaline Hot Spring Ecosystems from Yellowstone National Park. Microb Ecol 65:537-540

Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 301:976-978

Williams TJ, Zhang CL, Scott JH, Bazylinski DA (2006) Evidence for autotrophy via the reverse tricarboxylic acid cycle in the marine magnetotactic coccus strain MC-1. Appl Environ Microbiol 72:1322-1329

Wilson M, Siering P, White C, Hauser M, Bartles A (2008) Novel Archaea and Bacteria Dominate Stable Microbial Communities in North America's Largest Hot Spring. Microb Ecol 56:292-305

Wu Q, Sanford RA, Löffler FE (2006) Uranium(VI) reduction by *Anaeromyxobacter dehalogenans* strain 2CP-C. Appl Environ Microbiol 72:3608-3614

Xie W, Zhang CL, Wang J, Chen Y, Zhu Y, de la Torre JR, Dong H, Hartnett HE, Hedlund BP, Klotz MG (2014) Distribution of ether lipids and composition of the

archaeal community in terrestrial geothermal springs: impact of environmental variables. Environ Microbiol:n/a-n/a

Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H, Kamagata Y (2006) *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. Int J Syst Evol Microbiol 56:1331-1340

Yamamoto H, Hiraishi A, Kato K, Chiura HX, Maki Y, Shimizu A (1998) Phylogenetic evidence for the existence of novel thermophilic bacteria in hot spring sulfur-turf microbial mats in Japan. Appl Environ Microbiol 64:1680-1687

Yim LC, Hongmei J, Aitchison JC, Pointing SB (2006) Highly diverse community structure in a remote central Tibetan geothermal spring does not display monotonic variation to thermal stress. FEMS Microbiol Ecol 57:80-91

Zhang C, Wang J, Dodsworth J, Williams A, Zhu C, Hinrichs K-U, Zheng F, Hedlund BP (2013) *In situ* production of branched glycerol dialkyl glycerol tetraethers in a Great Basin hot spring (USA). Frontiers in Microbiology 4

Zhang CL, Fouke BW, Bonheyo GT, Peacock AD, White DC, Huang Y, Romanek CS (2004) Lipid biomarkers and carbon-isotopes of modern travertine deposits (Yellowstone National Park, USA): Implications for biogeochemical dynamics in hot-spring systems. Geochim Cosmochim Acta 68:3157-3169

## Tables and Figures

**Table 1.** Geographic and geochemical parameters for the 20 sampling sites in YNP.

| Thermal Inventory ID[a] | GPS Location (N/W) | Sample ID | Distance from source pool to A (cm) | Distance from A to B (cm) | Temperature (°C) | pH | Conductivity (µS/cm) | $H_2S$ (mg $L^{-1}$) | $F^-$ (mg $L^{-1}$) | $Cl^-$ (mg $L^{-1}$) | $NO_2^-$ (mg $L^{-1}$) | $NO_3^-$ (mg $L^{-1}$) | $SO_4^{2-}$ (mg $L^{-1}$) | $PO_4^{3-}$ (mg $L^{-1}$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SMMG045 | 44°21'18"/-110°47'55" | 07YSHO01A | 52 | 284 | 83[a] | 7.67[a] | 1400[a] | 0.53 | 18.726 | 167.7 | 0 | 0 | 44.777 | 0 |
| | | 07YSHO01B | | | 77 | 8.31 | 1348 | 0.1 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SMMG028 | 44°21'19"/-110°47'53" | 07YSHO02A | 1 | 741 | 80 | 8.69 | 1334 | 2.65 | 18.177 | 161.2 | 0 | 0.54 | 51.208 | 0 |
| | | 07YSHO02B | | | 70 | 8.88 | 1367 | 0.53 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SOG003 | 44°21'11"/-110°47'57" | 07YSHO03A | 13 | 15 | 78 | 6.76 | 1195 | 0.125 | 17.847 | 160.7 | 0 | 0 | 46.607 | 0 |
| | | 07YSHO03B | | | 72 | 7.04 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SOG002 | 44°21'11"/-110°47'56" | 07YSHO04A | 173 | 259 | 86 | 8.06 | 1475 | 0.303 | 22.167 | 209.5 | 0 | 0 | 48.929 | 0 |
| | | 07YSHO04B | | | 63 | 8.39 | n.d. | 0.057 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SSG016 | 44°21'14"/-110°48'04" | 07YSHO05A | 64 | 530 | 87 | 8.49 | 1367 | 0.15 | 19.564 | 181.1 | 0 | 0 | 44.309 | 0 |
| | | 07YSHO05B | | | 72 | 8.41 | 1435 | 0.043 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SLGG043 | 44°21'22"/-110°47'54" | 07YSHO06A | 192 | 193 | 73 | 8.21 | 1260 | 0.197 | 17.172 | 150.4 | 0 | 0 | 40.466 | 0 |
| | | 07YSHO06B | | | 67 | 8.15 | 1288 | 0.107 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SNG036 | 44°21'16"/-110°47'58" | 07YSHO07A | 3 | 343 | 80 | 6.65 | 1350 | 0.555 | 19.62 | 173.8 | 0 | 0 | 45.451 | 0 |
| | | 07YSHO07B | | | 71 | 7.3 | 1378 | 0.047 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SNG040 | 44°21'15"/-110°47'57" | 07YSHO08A | 2 | 71 | 64 | 6.47 | n.d. | 0.483 | 19.525 | 169.2 | 0 | 0 | 47.144 | 0 |
| | | 07YSHO08B | | | 62 | 6.79 | n.d. | 0.053 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SLGG034 | 44°21'06"/-110°47'42" | 07YSHO09A | 216 | 638 | 83 | 8.83 | 1303 | n.d. | 19.091 | 160.5 | 0 | 0 | 42.004 | 0 |
| | | 07YSHO09B | | | 74 | 8.51 | 1330 | 0.035 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |
| SLG017 | 44°21'05"/-110°47'45" | 07YSHO10A | 0 | 645 | 76 | 6.21 | 1060 | 4.45 | 18.931 | 202.968 | 0 | 0 | 50.79 | 0 |
| | | 07YSHO10B | | | 67 | 7.1 | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. |

[a] Site ID given by Yellowstone Center for Resources in their parkwide survey of Yellowstone thermal features, more info available at network (http://www.rcn.montana.edu)

[b] n.d. = not determined

**Table 2.** 454 GS FLX Titanium pyrosequencing and Newbler assembly metrics of the metagenomic DNA sample from 07YSHO09A

| Parameter | 07YSHO09A |
|---|---|
| Total number of reads | 765,145 |
| Mean read length | 435 bp |
| Metagenome size (unassembled reads) | 333 Mbp |
| Metagenome size (assembled reads) | 14 Mbp (4.2%) |
| Number of reads in contigs | 552,192 |
| Number of contigs | 26,440 |
| Reads/contig | 20.88 |
| Largest contig (bp) | 17,623 |
| Mean contig length (bp) | 1,095 |
| N50 contig length (bp) | 1,162 |
| Number of singletons | 66,251 |

**Table 3.** Alpha-diversity indices of 16S rRNA gene and ITS gene; this table was calculated by randomly drawing subsets of 867 sequences per sample from 16S rRNA sequences and subsets of 824 sequences per sample from ITS sequences, respectively.

| | 16S rRNA gene | | | | ITS gene | | | |
|---|---|---|---|---|---|---|---|---|
| Sample ID | NO. observed OTUs | Good's coverage | Shannon's index | Phylogenetic Diversity | NO. observed OTUs | Good's coverage | Shannon's index | Phylogenetic Diversity |
| 07YSHO01A | 26 | 0.975 | 0.570 | 2.232 | 4 | 0.999 | 1.255 | 0.675 |
| 07YSHO01B | 43 | 0.971 | 2.781 | 4.728 | 3 | 1.000 | 1.232 | 0.673 |
| 07YSHO02A | 20 | 0.984 | 0.960 | 1.908 | 4 | 0.999 | 1.053 | 0.733 |
| 07YSHO02B | 43 | 0.973 | 2.898 | 3.652 | 5 | 0.998 | 0.320 | 0.954 |
| 07YSHO03A | 19 | 0.993 | 2.080 | 2.203 | 6 | 0.996 | 1.062 | 1.405 |
| 07YSHO03B | 31 | 0.980 | 1.818 | 2.509 | 6 | 0.999 | 1.727 | 1.266 |
| 07YSHO04A | 12 | 0.990 | 0.657 | 1.269 | 2 | 1.000 | 0.280 | 0.653 |
| 07YSHO04B | 33 | 0.990 | 3.363 | 4.055 | 2 | 1.000 | 0.158 | 0.653 |
| 07YSHO05A | 5 | 0.997 | 0.062 | 0.728 | 3 | 0.999 | 0.178 | 0.681 |
| 07YSHO05B | 36 | 0.980 | 3.012 | 4.168 | 4 | 1.000 | 0.515 | 1.216 |
| 07YSHO06A | 43 | 0.970 | 2.693 | 4.748 | 4 | 0.999 | 0.882 | 0.679 |
| 07YSHO06B | 54 | 0.975 | 3.815 | 5.793 | 9 | 0.998 | 0.595 | 1.377 |
| 07YSHO07A | 21 | 0.991 | 1.784 | 2.442 | 12 | 0.996 | 1.746 | 1.641 |
| 07YSHO07B | 43 | 0.975 | 3.472 | 3.815 | 7 | 0.999 | 1.081 | 1.478 |
| 07YSHO08A | 90 | 0.932 | 2.980 | 8.184 | 9 | 0.995 | 0.843 | 1.526 |
| 07YSHO08B | 39 | 0.975 | 2.079 | 4.627 | 3 | 0.999 | 0.529 | 1.024 |
| 07YSHO09A | 16 | 0.987 | 0.663 | 1.966 | 5 | 0.996 | 0.294 | 0.767 |
| 07YSHO09B | 8 | 0.998 | 1.189 | 1.455 | 3 | 1.000 | 0.556 | 0.751 |
| 07YSHO10A | 34 | 0.985 | 2.569 | 3.733 | 10 | 0.998 | 1.572 | 1.525 |
| 07YSHO10B | 46 | 0.975 | 2.532 | 3.633 | 5 | 0.998 | 1.268 | 0.873 |

**Figure 1.** Bacterial diversity in different sites based on 16S rRNA sequences. Stacked bars represent the relative distribution of major phyla. Only phyla with relative abundance of more than 1% are shown. Phyla with relative abundance of less than 1% are shown as other Bacteria.

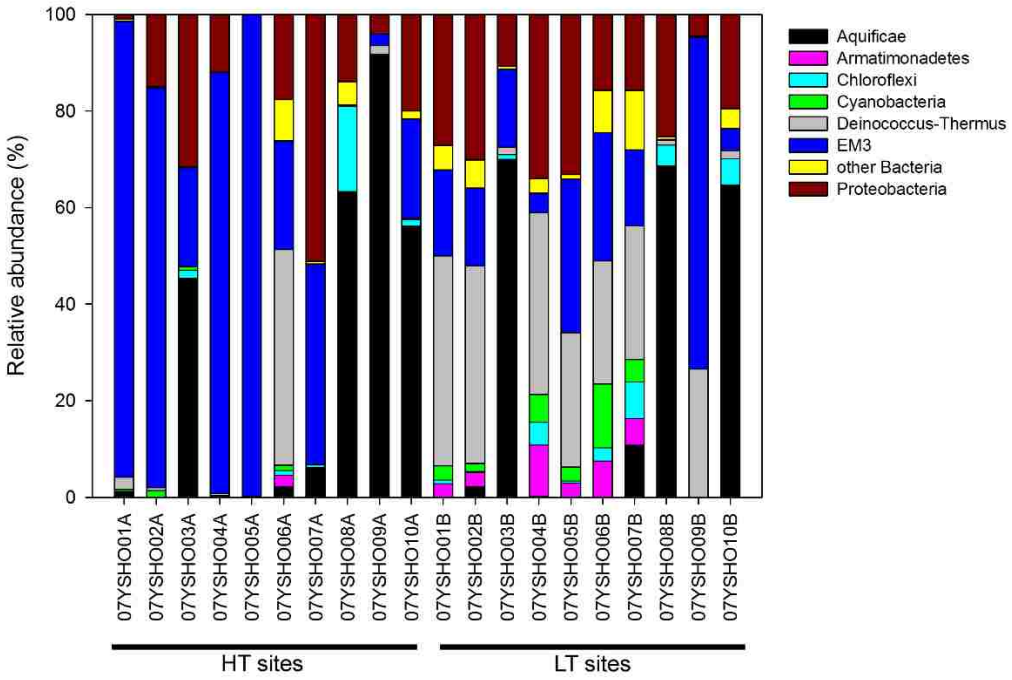


**Figure 2.** The Aquificales diversity in different sites based on ITS sequences. Stacked bars represent the relative distribution of species level (97% similarity of ITS gene identity).

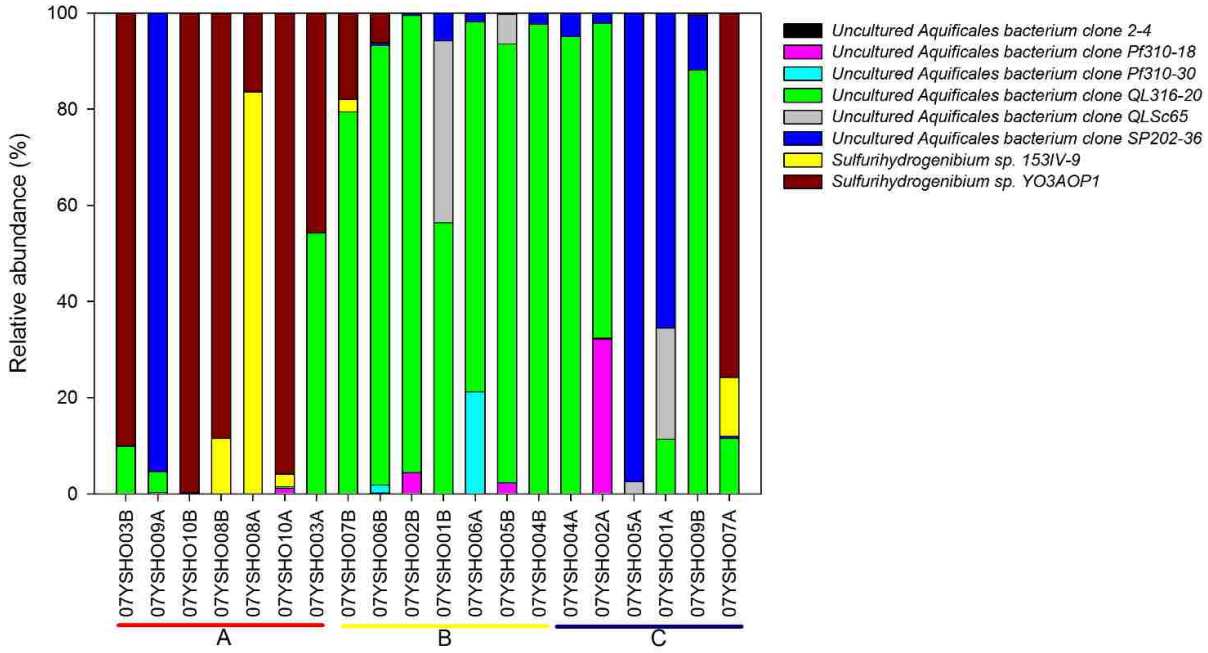


**Figure 3A.** The similarity of samples based on the bacterial community composition at the phylum level. The right dendrogram showing sample clustering at phylum-level using Bray-Curtis dissimilarity. The right bar charts showing the community composition of each sample. Only phyla with relative abundance of more than 1% are shown. Phyla with relative abundance of less than 1% are shown as other Bacteria.

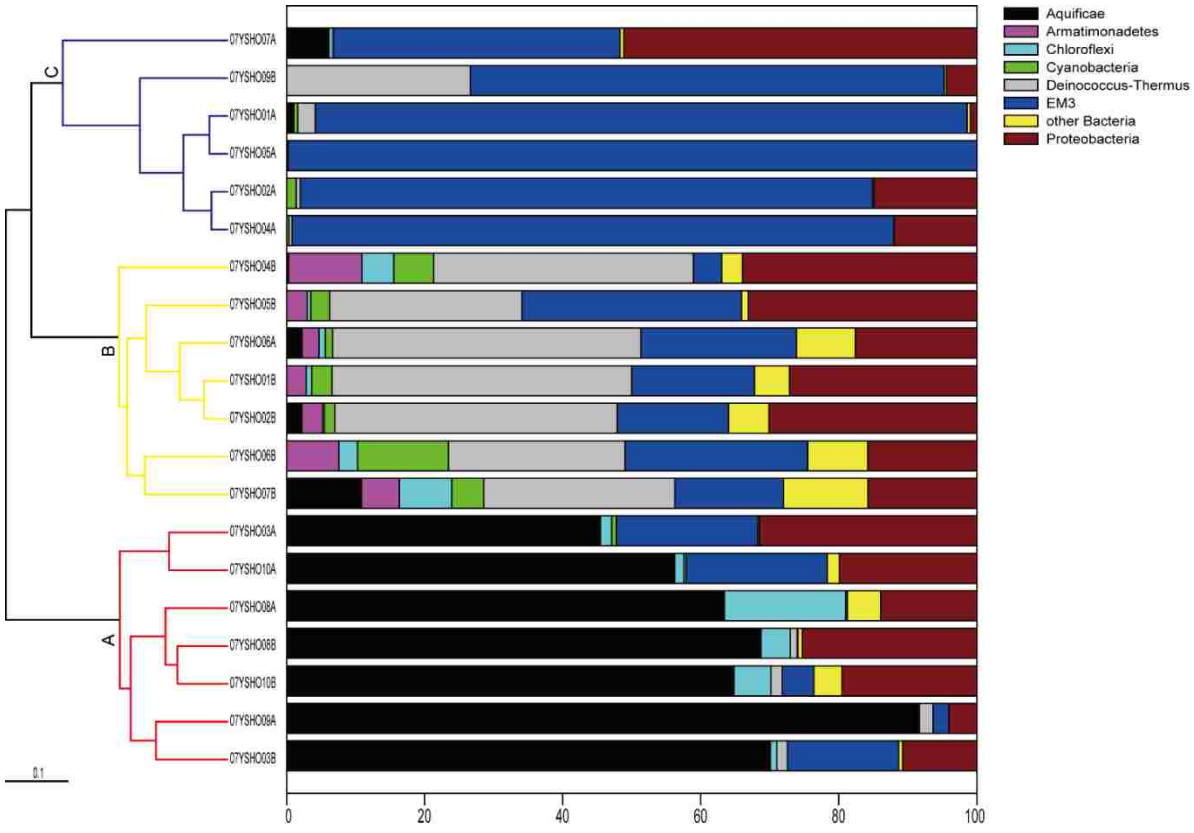


**Figure 3B.** Heatmap of 16S rRNA gene abundance in each sample ordered by taxonomic classification vertically and sample clustering horizontally. Dendrogram of sample clustering at phylum-level using Bray-Curtis dissimilarity is shown above the heatmap with each cluster colored according to the group showed at the node of each cluster. Rows in the heatmap represent genus-level OTUs (at the 95% similarity OTU level).

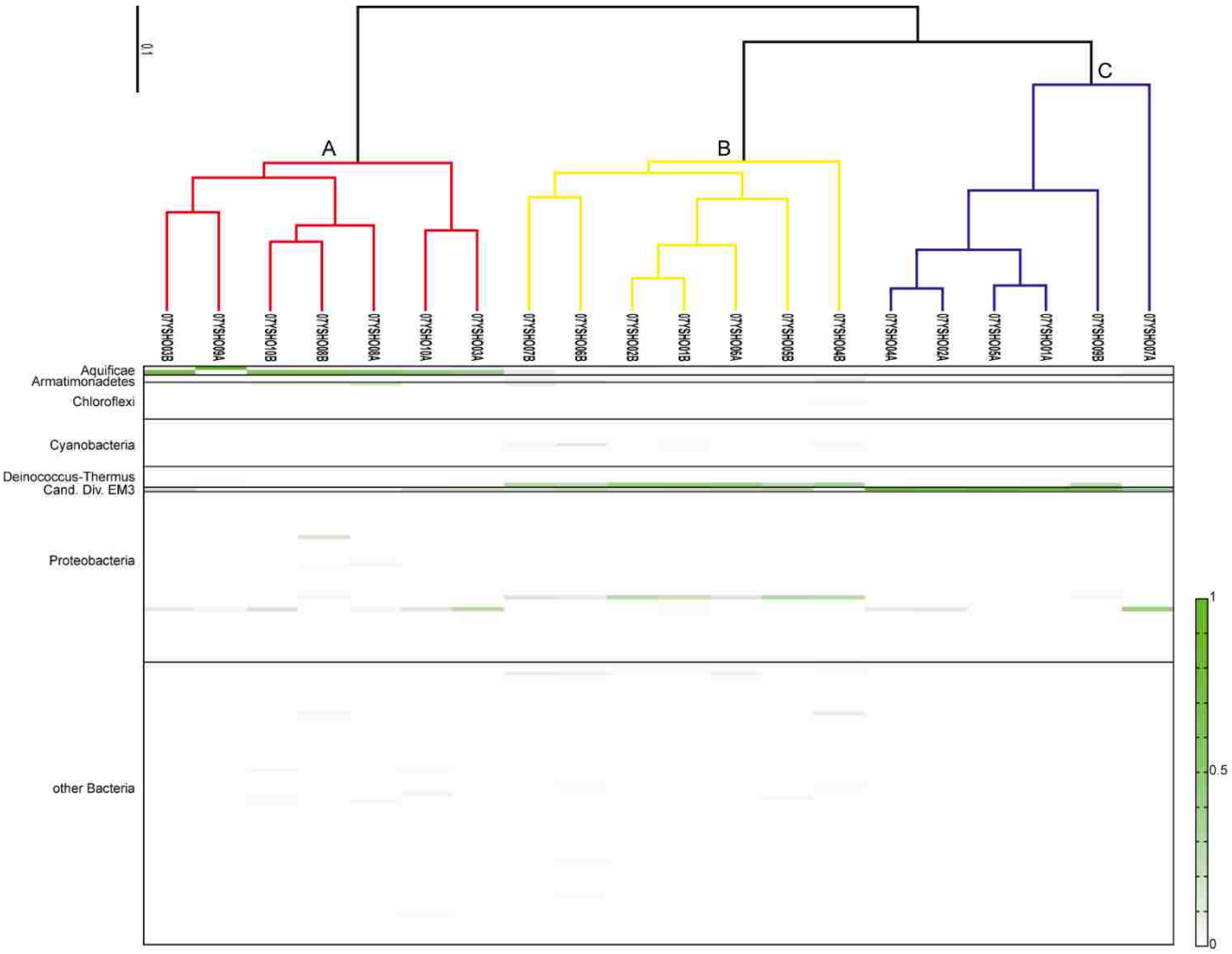


**Figure 4.** nMDS ordination depicting the distributions of twenty sites and major representative genus (solid circles). Environmental variables significantly ($P < 0.05$) correlating to ordination positions are given by arrows, indicting the directionality and relative magnitude of the correlations to the plot. The names of samples are colored by the groups they fall into based on the cluster analysis, as shown in Figure 3B.

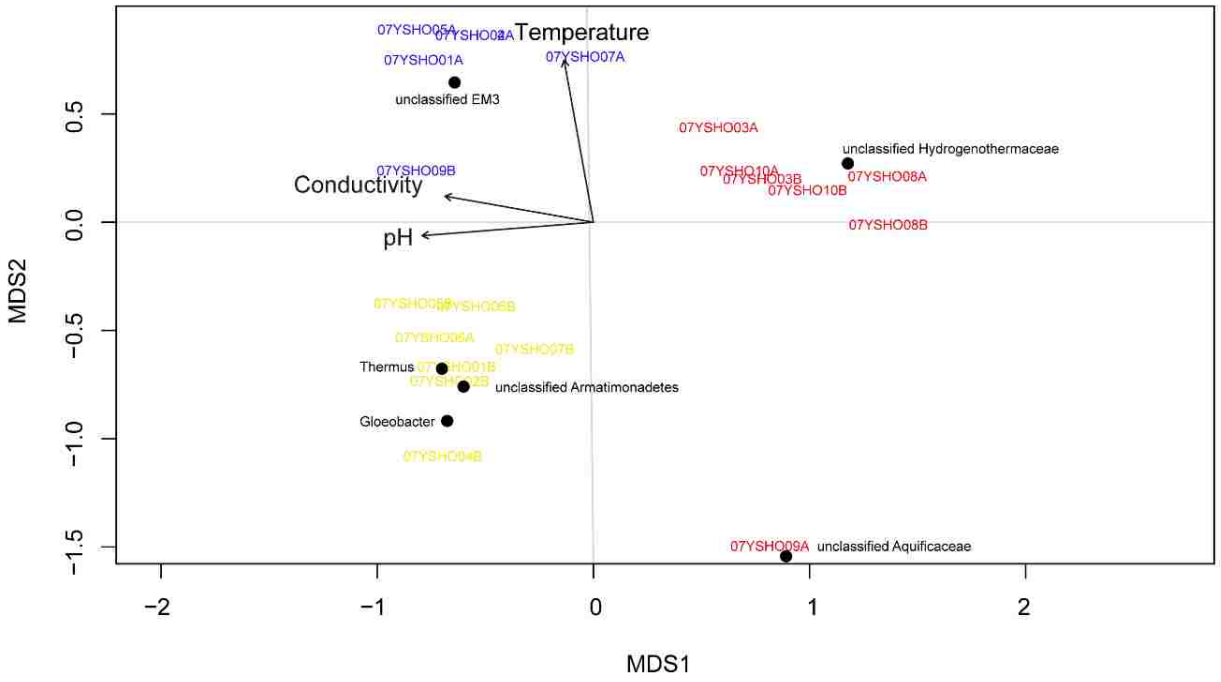


**Figure 5.** Box plot showing the two significant environmental variables (temperature and pH) detected in ANOVA analyses. Ranges of maximum and minimum are shown as whiskers, with outliners as open circles beyond these values. The median of each group is shown as a dark line.

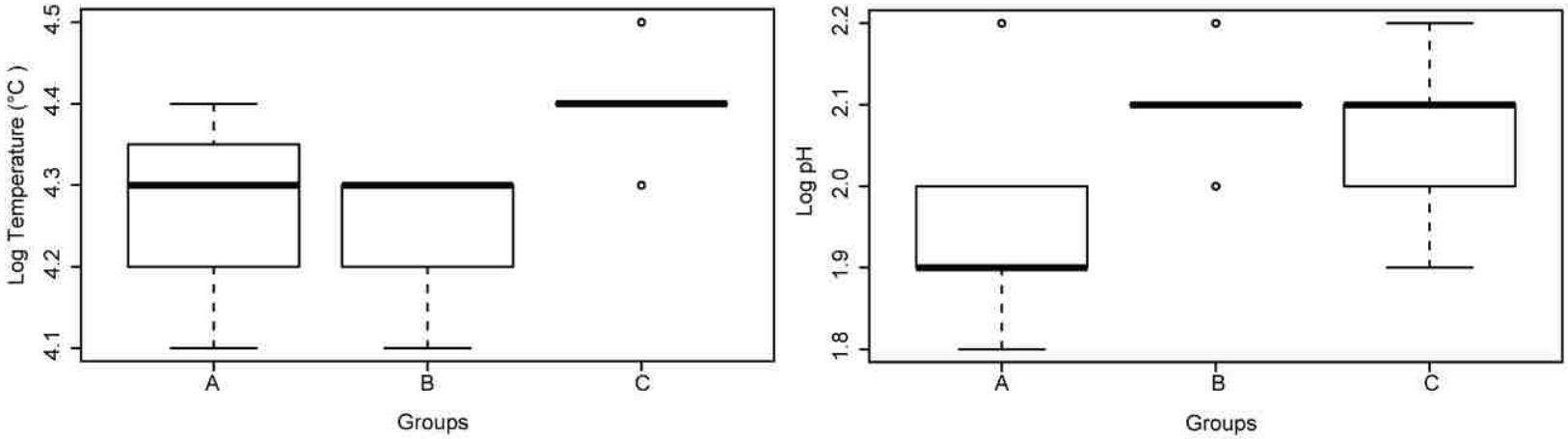


**Figure 6.** (A) Hierarchical clustering of COG categories of 21 YNP metagenomes. (B) Principal coordinates analysis (PCoA) of 21 YNP metagenomes based on COG categories.

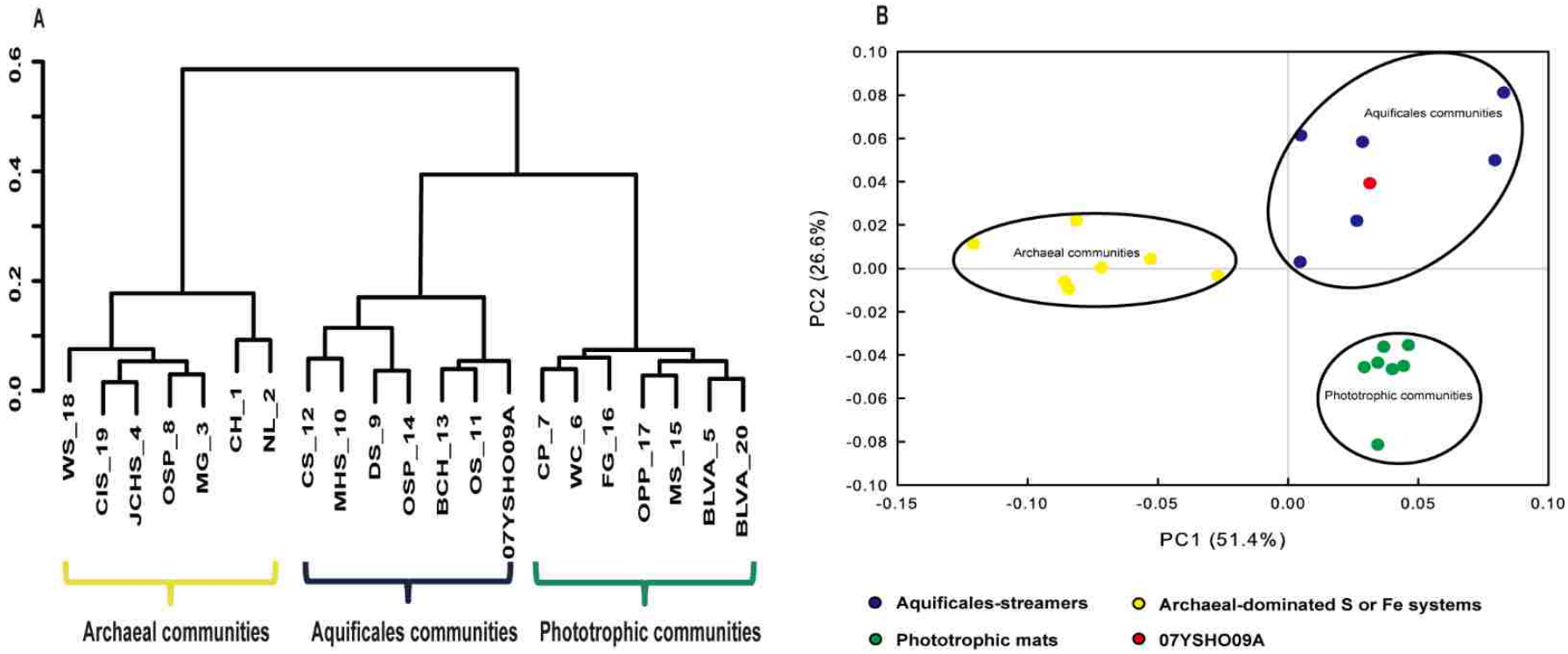


**Figure 7.** Comparison of the taxonomic assignment of unassembled metagenomic sequences based on GenBank (NCBI-NR), M5NR and RefSeq databases.

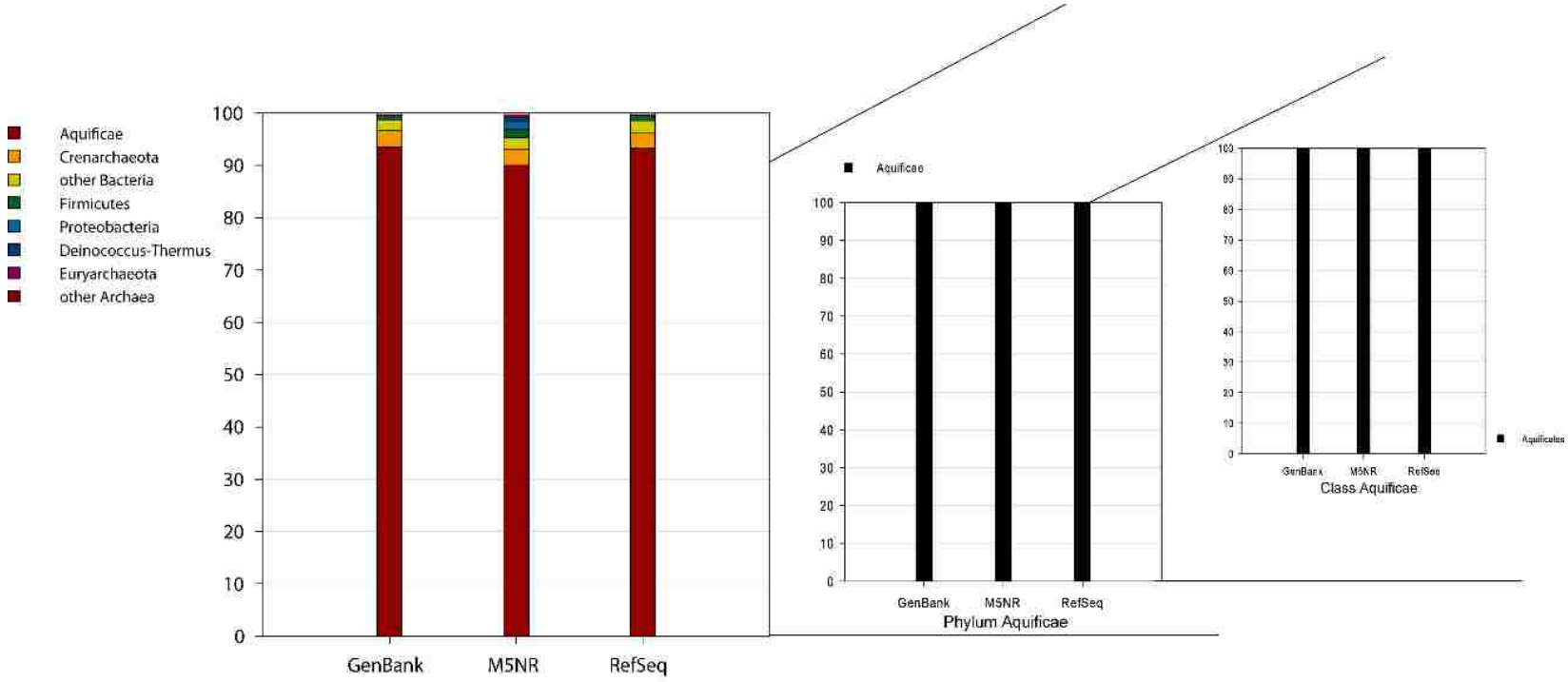


**Figure 8.** Taxonomic comparison within the most abundant class (Aquificales) based on GenBank (NCBI-NR), M5NR and RefSeq databases.

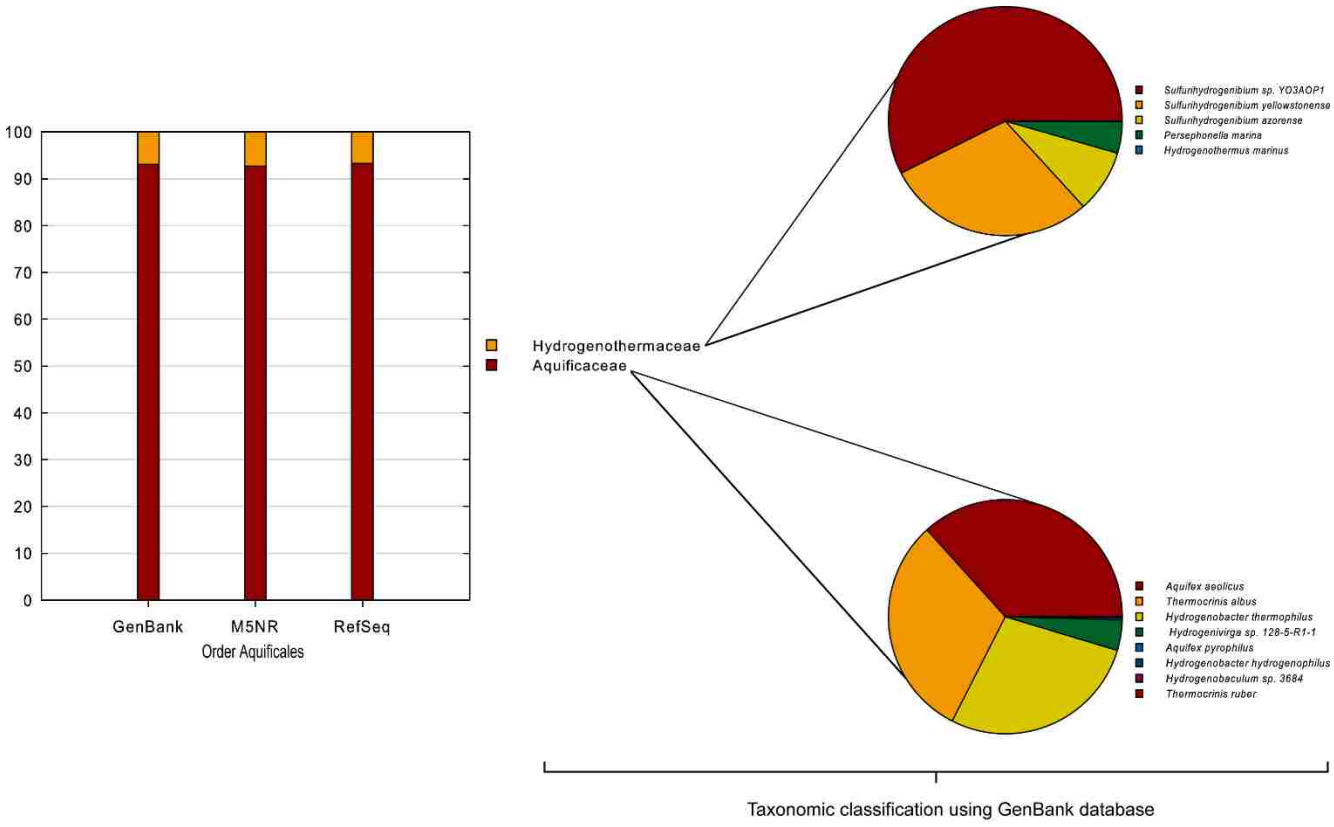


**Figure 9.** Overrepresented COG functions in the metagenome of 07YSHO09A relative to seven other archaeal metagenomes. These represent the 20 most enriched functions from 21 COG functions that received significant enrichment scores in five of seven comparisons (Table S3). Average D-score represents the mean enrichment score over all seven comparisons. Letters above graphs denote COG category.

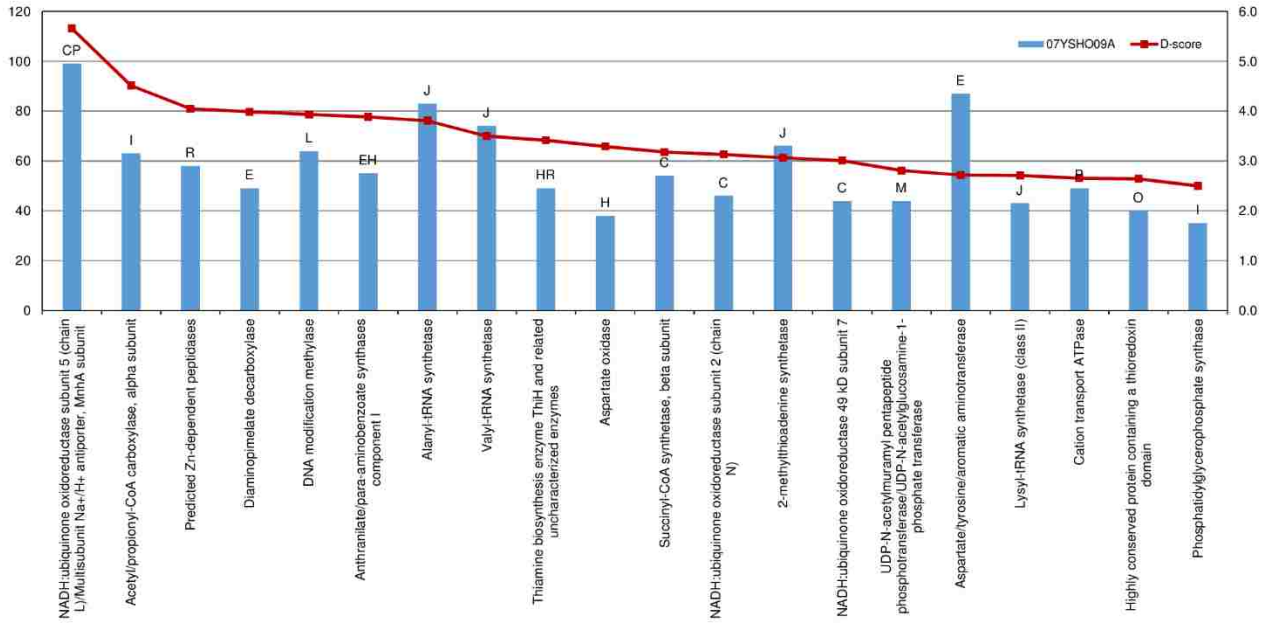


**Figure 10.** Overrepresented COG functions in the metagenome of 07YSHO09A relative to six other phototrophic metagenomes. These represent the 20 most enriched functions from 140 COG functions that received significant enrichment scores in five of six comparisons (Table S4). Average D-score represents the mean enrichment score over all seven comparisons. Letters above graphs denote COG category.

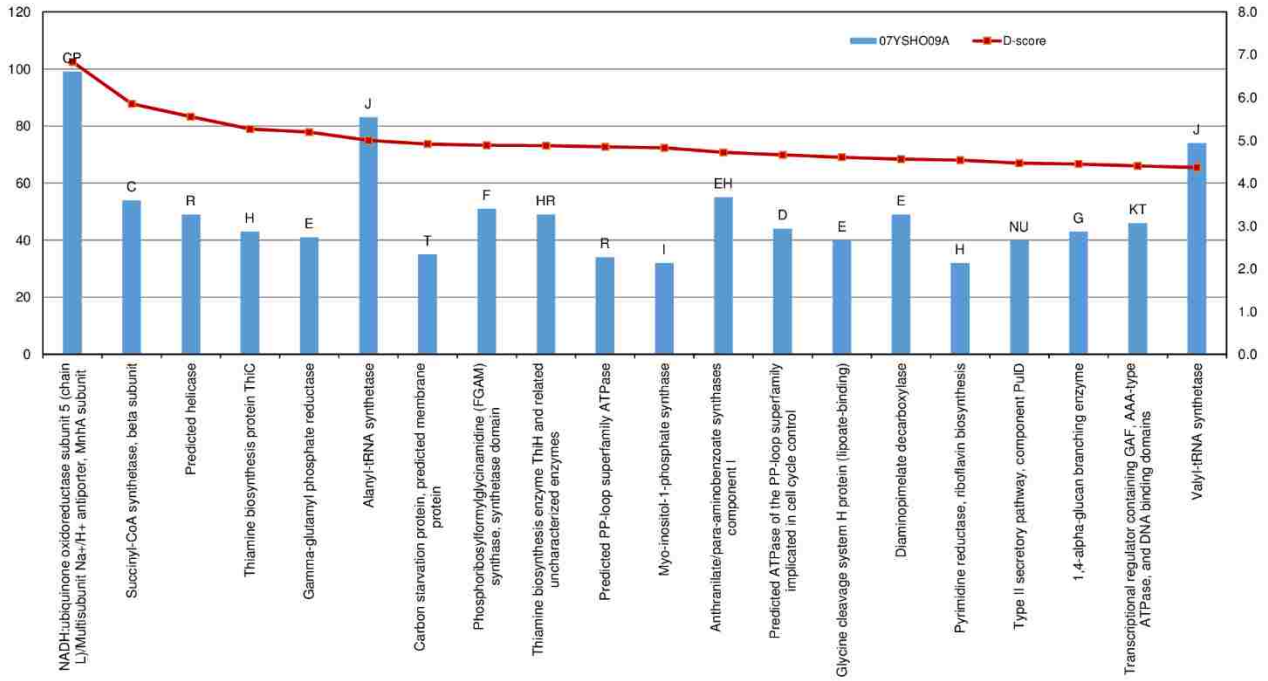


**Table S1.** General features of 20 selected YNP metagenomes for functional comparison. The table adapted from Inskeep et al. 2013.

| Site | Site abbreviation | T (ºC) | pH | DS (µM) | $S^0$ | Geobiological ecosystem type | IMG submission ID |
|---|---|---|---|---|---|---|---|
| White Creek | WC_6 | 48-50 | 8.2 | <1 | N | Phototrophic mats | 341 |
| Chocolate Pots | CP_7 | 52 | 6.2 | <1 | N | Phototrophic mats | 396 |
| Mushroom Spring | MS_15 | 60 | 8.2 | <1 | N | Phototrophic mats | 485 |
| Fairy Geyser | FG_16 | 36-38 | 9.1 | <1 | N | Phototrophic mats | 891 |
| Bath Lake Vista | BLVA_5 | 56-57 | 6.2 | 117 | Y | Phototrophic mats | 340 |
| Bath Lake Vista | BLVA_20 | 54-56 | 6.2 | 120 | Y | Phototrophic mats | 605 |
| Obsidian Pool Prime | OPP_17 | 56 | 5.7 | <1 | N | Bacterial OP Divisions | 572 |
| Dragon Spring | DS_9 | 70-72 | 3.1 | 80 | Y | Aquificales-rich 'filamentous-streamer' communities | 394 |
| 100 Spring Plain | OSP_14 | 72-74 | 3.5 | 10 | N | Aquificales-rich 'filamentous-streamer' communities | 344 |
| Mammoth Hot Spring | MHS_10 | 70-72 | 6.5 | 70 | Y | Aquificales-rich 'filamentous-streamer' communities | 888 |
| Calcite Springs | CS_12 | 76 | 7.8 | 105 | Y | Aquificales-rich 'filamentous-streamer' communities | 395 |
| Octopus Spring | OS_11 | 82 | 7.9 | <1 | N | Aquificales-rich 'filamentous-streamer' communities | 397 |
| Bechler Spring | BCH_13 | 82 | 7.8 | <1 | N | Aquificales-rich 'filamentous-streamer' communities | 343 |
| 100 Spring Plain | OSP_8 | 72 | 3.4 | <1 | N | Crenarchaeota,Mixed, Novel Archaea | 342 |
| Crater Hills | CH_1 | 76 | 2.6 | 2 | Y | Archaeal-dominated sediments | 392 |
| Nymph Lake | NL_2 | 88 | ~4 | 3 | Y | Archaeal-dominated sediments | 484 |
| Monarch Geyser | MG_3 | 78-80 | 4 | 25 | Y | Archaeal-dominated sediments | 875 |
| Cistern Spring | CIS_19 | 78-80 | 4.4 | 20 | Y | Archaeal-dominated sediments | 483 |
| Joseph's Coat Spring | JCHS_4 | 80 | 6.1 | 20 | Y | Archaeal-dominated sediments | 378 |
| Washburn Spring | WS_18 | 76 | 6.4 | 160 | Y | Archaeal-dominated sediments | 571 |

**Table S2.** Domain distribution on 07YSHO09A based on M5NR database.

| Domain | 07YSHO09A % |
|---|---|
| Archaea | 3.52 |
| Bacteria | 87.71 |
| Eukaryota | 0.01 |
| Viruses | 0.05 |
| Unassigned | 8.71 |

**Table S3.** Overrepresented COG functions in the metagenome 07YSHO09A compared to other YNP metagenomes from archaeal communities (OSP_8, CH_1, WS_18, CIS_19, NL_2, JCHS_4 and MG_3). Grey shading indicates significant D-scores.

| | | | Gene counts | D-scores[a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| COG | COG category | COG descirpiton | 07YSHO09A | Mean | OSP_8 | CH_1 | WS_18 | CIS_19 | NL_2 | JCHS_4 | MG_3 |
| COG1009 | CP | NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter, MnhA subunit | 99 | 5.7 | 7.12 | 4.72 | 7.61 | 5.24 | 4.3 | 5.68 | 4.94 |
| COG4770 | I | Acetyl/propionyl-CoA carboxylase, alpha subunit | 63 | 4.5 | 5.27 | 2.86 | 4.67 | 5.15 | 3.4 | 5.54 | 4.7 |
| COG0612 | R | Predicted Zn-dependent peptidases | 58 | 4.0 | 4.41 | 4.37 | 4.06 | 4.46 | 3.51 | 3.89 | 3.62 |
| COG0019 | E | Diaminopimelate decarboxylase | 49 | 4.0 | 4.3 | 4.02 | 4.67 | 4.3 | 2.86 | 4.29 | 3.46 |
| COG0863 | L | DNA modification methylase | 64 | 3.9 | 4.17 | 2.68 | 5.09 | 4.46 | 2.64 | 5.59 | 2.89 |
| COG0147 | EH | Anthranilate/para-aminobenzoate synthases component I | 55 | 3.9 | 3.84 | 3.2 | 5.32 | 4.47 | 3.4 | 3.32 | 3.63 |
| COG0013 | J | Alanyl-tRNA synthetase | 83 | 3.8 | 5.09 | 3.53 | 2.94 | 4.25 | 1.99 | 4.86 | 3.99 |
| COG0525 | J | Valyl-tRNA synthetase | 74 | 3.5 | 4.39 | 3.8 | 2.8 | 3.11 | 3.02 | 4.2 | 3.17 |
| COG1060 | HR | Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes | 49 | 3.4 | 3.38 | 4.02 | 4.8 | 2.89 | 3.49 | 3.27 | 2.04 |
| COG0029 | H | Aspartate oxidase | 38 | 3.3 | 3.61 | 1.99 | 5.25 | 3.51 | 2.71 | 2.82 | 3.15 |
| COG0045 | C | Succinyl-CoA synthetase, beta subunit | 54 | 3.2 | 3.27 | 2.65 | 5.22 | 2.37 | 3.06 | 2.89 | 2.78 |
| COG1007 | C | NADH:ubiquinone oxidoreductase subunit 2 (chain N) | 46 | 3.1 | 3.87 | 2.74 | 5.13 | 2.65 | 1.22 | 3.25 | 3.04 |
| COG0621 | J | 2-methylthioadenine synthetase | 66 | 3.1 | 4 | 2.34 | 2.43 | 3.54 | 2.72 | 3.05 | 3.37 |
| COG0649 | C | NADH:ubiquinone oxidoreductase 49 kD subunit 7 | 44 | 3.0 | 3.71 | 2.64 | 4.22 | 3.1 | 2.64 | 2.49 | 2.25 |
| COG0472 | M | UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase | 44 | 2.8 | 2.78 | 2.64 | 2.59 | 3.1 | 2.32 | 3.09 | 3.11 |
| COG0436 | E | Aspartate/tyrosine/aromatic aminotransferase | 87 | 2.7 | 2.43 | 4.09 | 3.36 | 3.06 | 2.81 | 1.41 | 1.87 |
| COG1190 | J | Lysyl-tRNA synthetase (class II) | 43 | 2.7 | 2.69 | 2.58 | 3.98 | 2.41 | 1.65 | 2.61 | 3.04 |
| COG2217 | P | Cation transport ATPase | 49 | 2.7 | 2.06 | 2.9 | 2.87 | 2.89 | 1.96 | 3.07 | 2.83 |
| COG1331 | O | Highly conserved protein containing a thioredoxin domain | 40 | 2.6 | 2.05 | 0.03 | 6.56 | 2.57 | 1.48 | 3.66 | 2.15 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| COG0558 | I | Phosphatidylglycerophosphate synthase | 35 | 2.5 | 3.12 | 1.79 | 2.7 | 3.03 | 1.85 | 2.57 | 2.42 |
| COG1110 | L | Reverse gyrase | 88 | 2.0 | 2.61 | 1.15 | 3.23 | 3.81 | -2.04 | 3.25 | 2.06 |

[a]P-value cutoffs used to identify significant D-scores (False Discovery Rate of 0.05): OSP_8, ≤ 2.38e-02; CH_1, ≤ 1.69e-02; WS_18, ≤ 2.13e-02; CIS_19, ≤ 2.26e-02; NL_2, ≤ 9.45e-03; JCHS_4, ≤ 2.36e-02; MG_3, ≤ 2.09e-02.

**Table S4.** Overrepresented COG functions in the metagenome 07YSHO09A compared to other YNP metagenomes from phototrophic communities (MS_15, FG_16, OPP_17, BLVA_20, WC_6 and CP_7). Grey shading indicates significant D-scores.

| | | | Gene counts | D-scores[a] | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| COG | COG category | COG descirpiton | 07YSHO09A | Mean | MS_15 | FG_16 | OPP_17 | BLVA_20 | WC_6 | CP_7 |
| COG1009 | CP | NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter, MnhA subunit | 99 | 6.8 | 8.5 | 6.12 | 6.23 | 7.2 | 6.81 | 6.14 |
| COG0045 | C | Succinyl-CoA synthetase, beta subunit | 54 | 5.9 | 5.68 | 6.06 | 5.97 | 6.04 | 6.39 | 4.98 |
| COG4889 | R | Predicted helicase | 49 | 5.6 | 5.82 | 5.9 | 5.21 | 6.67 | 5.93 | 3.77 |
| COG0422 | H | Thiamine biosynthesis protein ThiC | 43 | 5.3 | 6.05 | 4.97 | 4.49 | 5.22 | 5.88 | 4.96 |
| COG0014 | E | Gamma-glutamyl phosphate reductase | 41 | 5.2 | 5.5 | 4.43 | 5.27 | 5.68 | 5.31 | 4.95 |
| COG0013 | J | Alanyl-tRNA synthetase | 83 | 5.0 | 5.14 | 5.18 | 4.46 | 4.6 | 5.83 | 4.79 |
| COG1966 | T | Carbon starvation protein, predicted membrane protein | 35 | 4.9 | 4.98 | 4.59 | 4.49 | 5 | 5.88 | 4.54 |
| COG0046 | F | Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain | 51 | 4.9 | 4.25 | 5.03 | 4.22 | 5.9 | 5.48 | 4.4 |
| COG1060 | HR | Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes | 49 | 4.9 | 3.91 | 6.41 | 4.2 | 5.25 | 4.13 | 5.34 |
| COG0603 | R | Predicted PP-loop superfamily ATPase | 34 | 4.8 | 5.16 | 5.67 | 3.97 | 5.67 | 4.17 | 4.43 |
| COG1260 | I | Myo-inositol-1-phosphate synthase | 32 | 4.8 | 4.71 | 4.81 | 4.18 | 5.02 | 5.8 | 4.43 |
| COG0147 | EH | Anthranilate/para-aminobenzoate synthases component I | 55 | 4.7 | 4.61 | 4.88 | 4.43 | 5.3 | 5 | 4.06 |
| COG0037 | D | Predicted ATPase of the PP-loop superfamily implicated in cell cycle control | 44 | 4.7 | 4.83 | 5.07 | 4.06 | 4 | 5.62 | 4.36 |
| COG0509 | E | Glycine cleavage system H protein (lipoate-binding) | 40 | 4.6 | 5.23 | 4.98 | 4.02 | 4.72 | 4.19 | 4.48 |
| COG0019 | E | Diaminopimelate decarboxylase | 49 | 4.6 | 5.21 | 4.98 | 3.44 | 5.7 | 3.35 | 4.68 |
| COG1985 | H | Pyrimidine reductase, riboflavin biosynthesis | 32 | 4.5 | 5.2 | 5.01 | 4.18 | 4.82 | 3.57 | 4.43 |
| COG1450 | NU | Type II secretory pathway, component PulD | 40 | 4.5 | 4.43 | 3.71 | 3 | 5.39 | 5.77 | 4.48 |
| COG0296 | G | 1,4-alpha-glucan branching enzyme | 43 | 4.4 | 4.95 | 3.76 | 4.67 | 4.6 | 3.91 | 4.78 |
| COG3604 | KT | Transcriptional regulator containing GAF, AAA-type ATPase, and DNA binding domains | 46 | 4.4 | 5.59 | 3.17 | 4.08 | 1.79 | 6.35 | 5.42 |
| COG0525 | J | Valyl-tRNA synthetase | 74 | 4.4 | 5.6 | 4.95 | 3.01 | 3.28 | 4.56 | 4.78 |
| COG0843 | C | Heme/copper-type cytochrome/quinol oxidases, subunit 1 | 44 | 4.4 | 5.08 | 2.71 | 4.4 | 5.33 | 3.73 | 4.87 |
| COG2805 | NU | Tfp pilus assembly protein, pilus retraction ATPase PilT | 49 | 4.3 | 4.11 | 5.58 | 3.29 | 4.55 | 4.4 | 3.91 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| COG0776 | L | Bacterial nucleoid DNA-binding protein | 35 | 4.3 | 5.97 | 5.16 | 4.28 | 3.06 | 3.77 | 3.58 |
| COG1007 | C | NADH:ubiquinone oxidoreductase subunit 2 (chain N) | 46 | 4.3 | 4.96 | 3.43 | 3.76 | 4.5 | 3.95 | 5.06 |
| COG4147 | R | Predicted symporter | 33 | 4.3 | 4.85 | 3.64 | 4.73 | 4.39 | 4.61 | 3.36 |
| COG0123 | BQ | Deacetylases, including yeast histone deacetylase and acetoin utilization protein | 46 | 4.3 | 3.64 | 3.95 | 4.42 | 4.93 | 4.38 | 4.23 |
| COG1206 | J | NAD(FAD)-utilizing enzyme possibly involved in translation | 28 | 4.2 | 4.09 | 3.15 | 3.97 | 4.94 | 4.64 | 4.68 |
| COG1452 | M | Organic solvent tolerance protein OstA | 33 | 4.2 | 4.7 | 4.73 | 4.07 | 4.95 | 3.52 | 3.36 |
| COG0841 | V | Cation/multidrug efflux pump | 97 | 4.2 | 5.8 | 3.87 | 2.01 | 3.44 | 5.77 | 4.38 |
| COG0620 | E | Methionine synthase II (cobalamin-independent) | 29 | 4.2 | 4.25 | 4.44 | 3.85 | 4.44 | 4.77 | 3.45 |
| COG0847 | L | DNA polymerase III, epsilon subunit and related 3'-5' exonucleases | 26 | 4.1 | 5.66 | 4.49 | 3.98 | 3.63 | 3.13 | 3.96 |
| COG4770 | I | Acetyl/propionyl-CoA carboxylase, alpha subunit | 63 | 4.1 | 4.74 | 2.46 | 3.86 | 4.4 | 4.84 | 4.5 |
| COG0034 | F | Glutamine phosphoribosylpyrophosphate amidotransferase | 40 | 4.1 | 4.31 | 4.16 | 4.02 | 4.11 | 4.35 | 3.78 |
| COG0095 | H | Lipoate-protein ligase A | 31 | 4.1 | 4.88 | 4.9 | 2.67 | 3.44 | 4.79 | 3.69 |
| COG0260 | E | Leucyl aminopeptidase | 32 | 4.1 | 5.55 | 3.35 | 3.96 | 4.82 | 3.07 | 3.61 |
| COG0467 | T | RecA-superfamily ATPases implicated in signal transduction | 29 | 4.0 | 3.64 | 4.23 | 3.63 | 3.68 | 4.77 | 4.31 |
| COG0649 | C | NADH:ubiquinone oxidoreductase 49 kD subunit 7 | 44 | 4.0 | 4.95 | 3.2 | 3.73 | 4.28 | 3.88 | 4.03 |
| COG0067 | E | Glutamate synthase domain 1 | 31 | 4.0 | 4.4 | 3.38 | 3.85 | 4.5 | 3.99 | 3.69 |
| COG0497 | L | ATPase involved in DNA repair | 46 | 4.0 | 4.61 | 4.97 | 3.3 | 3.96 | 2.66 | 4.23 |
| COG0445 | D | NAD/FAD-utilizing enzyme apparently involved in cell division | 45 | 4.0 | 3.82 | 4.86 | 3.35 | 3.98 | 3.57 | 4.13 |
| COG0751 | J | Glycyl-tRNA synthetase, beta subunit | 38 | 4.0 | 4.96 | 3.47 | 3.11 | 4.02 | 5.37 | 2.77 |
| COG0741 | M | Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains) | 40 | 3.9 | 4.95 | 3.42 | 3 | 3.01 | 5.03 | 4.13 |
| COG2905 | T | Predicted signal-transduction protein containing cAMP-binding and CBS domains | 38 | 3.9 | 3.44 | 4.25 | 4.58 | 3.87 | 4.12 | 3.08 |
| COG0498 | E | Threonine synthase | 49 | 3.9 | 2.78 | 3.88 | 2.6 | 3.91 | 5.29 | 4.84 |
| COG0317 | TK | Guanosine polyphosphate pyrophosphohydrolases/synthetases | 55 | 3.9 | 4.42 | 4.62 | 3.17 | 4.18 | 3.84 | 3.02 |
| COG0552 | U | Signal recognition particle GTPase | 34 | 3.9 | 3.59 | 4.11 | 3.2 | 4 | 3.99 | 4.23 |
| COG2084 | I | 3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases | 29 | 3.8 | 3.79 | 3.65 | 3.21 | 4.24 | 3.92 | 4.09 |
| COG1729 | S | Uncharacterized protein conserved in bacteria | 51 | 3.8 | 4.05 | 4.22 | 3.34 | 3.64 | 3.19 | 4.4 |
| COG4775 | M | Outer membrane protein/protective antigen OMA87 | 62 | 3.7 | 5 | 5.41 | 3.15 | 3.77 | 3.07 | 2.09 |
| COG0370 | P | Fe2+ transport system protein B | 44 | 3.7 | 2.7 | 4.16 | 3.57 | 3.48 | 3.88 | 4.7 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| COG2270 | R | Permeases of the major facilitator superfamily | 33 | 3.7 | 4.4 | 3.99 | 3.27 | 3.39 | 3.69 | 3.73 |
| COG0419 | L | ATPase involved in DNA repair | 39 | 3.7 | 3.69 | 4.69 | 4.1 | 4.61 | 3.03 | 2.29 |
| COG1353 | R | Predicted hydrolase of the HD superfamily (permuted catalytic motifs) | 38 | 3.7 | 3.11 | 4.41 | 2.95 | 3.43 | 5.37 | 3.08 |
| COG0558 | I | Phosphatidylglycerophosphate synthase | 35 | 3.7 | 3.74 | 4.23 | 3.68 | 3.49 | 4.11 | 3.07 |
| COG1190 | J | Lysyl-tRNA synthetase (class II) | 43 | 3.7 | 3.67 | 4.34 | 3.31 | 2.75 | 3.77 | 4.43 |
| COG0050 | J | GTPases - translation elongation factors | 43 | 3.7 | 3.89 | 4.05 | 3.96 | 2.75 | 4.06 | 3.31 |
| COG0675 | L | Transposase and inactivated derivatives | 75 | 3.7 | 3.63 | 7.83 | 4.21 | 5.93 | -2.56 | 2.9 |
| COG1692 | S | Uncharacterized protein conserved in bacteria | 21 | 3.6 | 4.18 | 2.89 | 3.11 | 3.5 | 5.1 | 3.06 |
| COG0075 | E | Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase | 34 | 3.6 | 3.47 | 3.13 | 3.02 | 3.52 | 4.54 | 4.03 |
| COG0825 | I | Acetyl-CoA carboxylase alpha subunit | 32 | 3.6 | 3.82 | 3.68 | 3.35 | 4.82 | 3.07 | 2.88 |
| COG0452 | H | Phosphopantothenoylcysteine synthetase/decarboxylase | 35 | 3.6 | 3.25 | 3.73 | 2.96 | 4.12 | 4.11 | 3.41 |
| COG0492 | O | Thioredoxin reductase | 52 | 3.5 | 4.07 | 3.81 | 3.16 | 3.41 | 3.78 | 2.98 |
| COG1509 | E | Lysine 2,3-aminomutase | 23 | 3.5 | 3.77 | 3.85 | 2.89 | 3.59 | 3.99 | 3.1 |
| COG2262 | R | GTPases | 29 | 3.5 | 3.09 | 3.84 | 3 | 4.24 | 3.73 | 3.25 |
| COG0165 | E | Argininosuccinate lyase | 33 | 3.5 | 2.84 | 2.85 | 4.28 | 3.87 | 4.42 | 2.83 |
| COG1008 | C | NADH:ubiquinone oxidoreductase subunit 4 (chain M) | 52 | 3.5 | 4.07 | 3.57 | 3.57 | 4.74 | 2.5 | 2.61 |
| COG0618 | R | Exopolyphosphatase-related proteins | 38 | 3.5 | 3.33 | 3.93 | 3.81 | 3.29 | 3.19 | 3.4 |
| COG2870 | M | ADP-heptose synthase, bifunctional sugar kinase/adenylyltransferase | 37 | 3.5 | 4.55 | 3.35 | 3.35 | 3.74 | 2.93 | 2.97 |
| COG0743 | I | 1-deoxy-D-xylulose 5-phosphate reductoisomerase | 31 | 3.5 | 3.67 | 4.48 | 2.32 | 3.95 | 4.37 | 2.09 |
| COG0029 | H | Aspartate oxidase | 38 | 3.5 | 3 | 3.47 | 3.46 | 3.16 | 3.8 | 3.91 |
| COG1199 | KL | Rad3-related DNA helicases | 33 | 3.5 | 3.97 | 2.41 | 2.55 | 3.87 | 4.05 | 3.92 |
| COG1331 | O | Highly conserved protein containing a thioredoxin domain | 40 | 3.4 | 2.96 | 3.85 | 3 | 3.4 | 4.35 | 2.99 |
| COG0495 | J | Leucyl-tRNA synthetase | 62 | 3.4 | 4.27 | 4.79 | 2.68 | 1.56 | 4.15 | 3.07 |
| COG0021 | G | Transketolase | 36 | 3.4 | 4.68 | 3.69 | 2.89 | 2.64 | 3.25 | 3.35 |
| COG0677 | M | UDP-N-acetyl-D-mannosaminuronate dehydrogenase | 37 | 3.4 | 2.86 | 3.65 | 2.05 | 4.05 | 3.84 | 3.99 |
| COG1793 | L | ATP-dependent DNA ligase | 25 | 3.4 | 2.99 | 3.29 | 2.71 | 2.58 | 4.26 | 4.6 |
| COG0436 | E | Aspartate/tyrosine/aromatic aminotransferase | 87 | 3.4 | 4.48 | 3.49 | 3.38 | 4.06 | 2.24 | 2.63 |
| COG0026 | F | Phosphoribosylaminoimidazole carboxylase (NCAIR synthetase) | 28 | 3.4 | 4.26 | 3.15 | 3.74 | 4.94 | 1.61 | 2.56 |
| COG0016 | J | Phenylalanyl-tRNA synthetase alpha subunit | 35 | 3.4 | 3.74 | 4.41 | 2.96 | 2.78 | 3.6 | 2.74 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| COG2265 | J | SAM-dependent methyltransferases related to tRNA (uracil-5-)-methyltransferase | 39 | 3.4 | 4.68 | 3.16 | 3.39 | 3.7 | 3.31 | 1.88 |
| COG0794 | M | Predicted sugar phosphate isomerase involved in capsule formation | 23 | 3.3 | 3.42 | 4.58 | 3.37 | 2.8 | 2.89 | 2.88 |
| COG1912 | S | Uncharacterized conserved protein | 24 | 3.3 | 2.26 | 3.99 | 3.25 | 3.34 | 3.04 | 3.96 |
| COG0482 | J | Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase, contains the PP-loop ATPase domain | 41 | 3.3 | 2.99 | 4.75 | 2.06 | 2.75 | 3.99 | 3.25 |
| COG1995 | H | Pyridoxal phosphate biosynthesis protein | 26 | 3.3 | 3.61 | 4.49 | 2.43 | 3.63 | 3.33 | 2.29 |
| COG0312 | R | Predicted Zn-dependent proteases and their inactivated homologs | 42 | 3.3 | 3.22 | 5.19 | 1.11 | 3.77 | 3.66 | 2.77 |
| COG0820 | R | Predicted Fe-S-cluster redox enzyme | 33 | 3.3 | 3.08 | 4.53 | 2.91 | 2.51 | 3.03 | 3.54 |
| COG0403 | E | Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain | 36 | 3.3 | 4.01 | 4.01 | 2.73 | 3.18 | 3.25 | 2.39 |
| COG0169 | E | Shikimate 5-dehydrogenase | 28 | 3.2 | 3.63 | 4.1 | 2.49 | 2.86 | 3.22 | 2.93 |
| COG0243 | C | Anaerobic dehydrogenases, typically selenocysteine-containing | 60 | 3.2 | 2.98 | 0.85 | 3.87 | 2.5 | 5.59 | 3.36 |
| COG1538 | MU | Outer membrane protein | 62 | 3.2 | 5.29 | 3.77 | -0.25 | 3.37 | 4.04 | 2.85 |
| COG2896 | H | Molybdenum cofactor biosynthesis enzyme | 25 | 3.2 | 3.29 | 3.49 | 2.92 | 3.11 | 3.39 | 2.73 |
| COG1640 | G | 4-alpha-glucanotransferase | 29 | 3.1 | 3.5 | 2.95 | 3.21 | 3.17 | 3 | 3.06 |
| COG0177 | L | Predicted EndoIII-related endonuclease | 28 | 3.1 | 3.07 | 3.52 | 3.29 | 3.19 | 3.41 | 2.38 |
| COG2067 | I | Long-chain fatty acid transport protein | 20 | 3.1 | 3.8 | 2.94 | 2.47 | 3.12 | 2.85 | 3.43 |
| COG0345 | E | Pyrroline-5-carboxylate reductase | 24 | 3.1 | 2.67 | 3.77 | 2.36 | 3.95 | 3.04 | 2.8 |
| COG0647 | G | Predicted sugar phosphatases of the HAD superfamily | 22 | 3.1 | 2.61 | 2.64 | 3.24 | 3.03 | 3.61 | 3.45 |
| COG0240 | C | Glycerol-3-phosphate dehydrogenase | 25 | 3.1 | 3.77 | 3.91 | 2.5 | 3.68 | 3.59 | 1.12 |
| COG1459 | NU | Type II secretory pathway, component PulF | 48 | 3.1 | 2.66 | 3.52 | 1.13 | 2.97 | 4.16 | 4.12 |
| COG2877 | M | 3-deoxy-D-manno-octulosonic acid (KDO) 8-phosphate synthase | 23 | 3.1 | 4.14 | 3.4 | 2.66 | 3.18 | 2.69 | 2.45 |
| COG1778 | R | Low specificity phosphatase (HAD superfamily) | 18 | 3.1 | 4.31 | 3.59 | 2.68 | 2.78 | 2.52 | 2.61 |
| COG0364 | G | Glucose-6-phosphate 1-dehydrogenase | 25 | 3.1 | 3.44 | 2.55 | 2.71 | 3.29 | 2.8 | 3.6 |
| COG0069 | E | Glutamate synthase domain 2 | 30 | 3.1 | 2.74 | 3.08 | 2.93 | 3.14 | 3.49 | 2.99 |
| COG0351 | H | Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase | 23 | 3.0 | 2.94 | 2.8 | 2.44 | 3.38 | 3.31 | 3.34 |
| COG0161 | H | Adenosylmethionine-8-amino-7-oxononanoate aminotransferase | 25 | 3.0 | 4.48 | 2.05 | 3.15 | 2.75 | 2.62 | 3.15 |
| COG2041 | R | Sulfite oxidase and related enzymes | 22 | 3.0 | 3.77 | 2.45 | 2.52 | 3.88 | 3.16 | 2.31 |
| COG0846 | K | NAD-dependent protein deacetylases, SIR2 family | 24 | 3.0 | 3.12 | 3.34 | 3.02 | 3.14 | 3.04 | 2.39 |
| COG0439 | I | Biotin carboxylase | 25 | 3.0 | 3.14 | 3.69 | 3.38 | 3.68 | 2.44 | 1.62 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| COG0074 | C | Succinyl-CoA synthetase, alpha subunit | 26 | 3.0 | 2.6 | 2.36 | 2.84 | 3.08 | 3.94 | 3.07 |
| COG0150 | F | Phosphoribosylaminoimidazole (AIR) synthetase | 28 | 3.0 | 2.55 | 3.71 | 3.97 | 3.02 | 2.06 | 2.56 |
| COG2890 | J | Methylase of polypeptide chain release factors | 33 | 3.0 | 3.2 | 3.31 | 2.22 | 3.71 | 3.03 | 2.34 |
| COG1055 | P | Na+/H+ antiporter NhaD and related arsenite permeases | 34 | 2.9 | 2.87 | 3.77 | 2.33 | 3.37 | 3 | 2.3 |
| COG0138 | F | AICAR transformylase/IMP cyclohydrolase PurH (only IMP cyclohydrolase domain in Aful) | 42 | 2.9 | 2.63 | 2.36 | 4.04 | 2.28 | 3.95 | 2.35 |
| COG0496 | R | Predicted acid phosphatase | 26 | 2.9 | 3.61 | 3.24 | 2.63 | 2.73 | 2.08 | 3.28 |
| COG0003 | P | Oxyanion-translocating ATPase | 33 | 2.9 | 3.84 | 2.7 | 3.86 | 2.65 | 2.88 | 1.59 |
| COG0104 | F | Adenylosuccinate synthase | 31 | 2.9 | 3.27 | 3.38 | 2.67 | 2.11 | 3.1 | 2.94 |
| COG0803 | P | ABC-type metal ion transport system, periplasmic component/surface adhesin | 26 | 2.9 | 3.16 | 2.69 | 3.05 | 3.26 | 2.41 | 2.87 |
| COG0061 | G | Predicted sugar kinase | 28 | 2.9 | 2.93 | 3.33 | 2.68 | 3.02 | 2.7 | 2.74 |
| COG0129 | EG | Dihydroxyacid dehydratase/phosphogluconate dehydratase | 37 | 2.9 | 3.19 | 2.12 | 1.47 | 3.31 | 4.17 | 3.13 |
| COG0644 | C | Dehydrogenases (flavoproteins) | 43 | 2.9 | 2.67 | 2.71 | 2.27 | 4.9 | 2.45 | 2.32 |
| COG2894 | D | Septum formation inhibitor-activating ATPase | 17 | 2.9 | 3.88 | 2.66 | 2.53 | 2.83 | 2.35 | 2.99 |
| COG0441 | J | Threonyl-tRNA synthetase | 42 | 2.9 | 2.82 | 2.98 | 2.3 | 2.99 | 2.58 | 3.51 |
| COG1048 | C | Aconitase A | 33 | 2.9 | 4.25 | 2.85 | 3.09 | 2.65 | 1.48 | 2.83 |
| COG0134 | E | Indole-3-glycerol phosphate synthase | 22 | 2.8 | 2.92 | 3.25 | 2.3 | 3.88 | 1.98 | 2.52 |
| COG0458 | EF | Carbamoylphosphate synthase large subunit (split gene in MJ) | 46 | 2.8 | 2.08 | 3.17 | 2.3 | 3.09 | 3.54 | 2.64 |
| COG0279 | G | Phosphoheptose isomerase | 19 | 2.8 | 4.03 | 2.15 | 2.57 | 2.53 | 2.91 | 2.52 |
| COG0263 | E | Glutamate 5-kinase | 22 | 2.8 | 2.61 | 2.45 | 2.75 | 3.03 | 2.74 | 2.97 |
| COG0352 | H | Thiamine monophosphate synthase | 22 | 2.8 | 2.04 | 2.84 | 2.75 | 3.66 | 2.74 | 2.52 |
| COG0323 | L | DNA mismatch repair enzyme (predicted ATPase) | 39 | 2.7 | 3.24 | 2.89 | 2.73 | 3.15 | 1.98 | 2.44 |
| COG0119 | E | Isopropylmalate/homocitrate/citramalate synthases | 52 | 2.7 | 2.63 | 3.69 | 3.72 | 2.55 | 1.47 | 2.25 |
| COG0182 | J | Predicted translation initiation factor 2B subunit, eIF-2B alpha/beta/delta family | 27 | 2.7 | 2.39 | 2.84 | 2.36 | 2.88 | 2.56 | 3.2 |
| COG0005 | F | Purine nucleoside phosphorylase | 30 | 2.7 | 2.16 | 3.25 | 2.02 | 2.98 | 2.97 | 2.81 |
| COG0693 | R | Putative intracellular protease/amidase | 20 | 2.6 | 2.39 | 2.52 | 2.97 | 2.5 | 2.63 | 2.68 |
| COG0340 | H | Biotin-(acetyl-CoA carboxylase) ligase | 21 | 2.6 | 3.06 | 2.68 | 2.61 | 2.48 | 2.19 | 2.59 |
| COG0343 | J | Queuine/archaeosine tRNA-ribosyltransferase | 34 | 2.5 | 2.43 | 2.98 | 2.84 | 1.51 | 3 | 2.46 |
| COG1253 | R | Hemolysins and related proteins containing CBS domains | 41 | 2.5 | 2.22 | 2.61 | 2.34 | 2.87 | 2.34 | 2.8 |
| COG0040 | E | ATP phosphoribosyltransferase | 20 | 2.5 | 2.55 | 2.52 | 2.71 | 2.31 | 2.22 | 2.68 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| COG0815 | M | Apolipoprotein N-acyltransferase | 23 | 2.5 | 1.96 | 2.42 | 2.66 | 2.99 | 2.69 | 2.25 |
| COG0315 | H | Molybdenum cofactor biosynthesis enzyme | 17 | 2.5 | 2.63 | 2.9 | 1.31 | 3.07 | 2.35 | 2.45 |
| COG1053 | C | Succinate dehydrogenase/fumarate reductase, flavoprotein subunit | 33 | 2.3 | 2.5 | 2 | 2.55 | 2.25 | 2.29 | 2.5 |
| COG0209 | F | Ribonucleotide reductase, alpha subunit | 42 | 2.3 | 2.35 | 3.24 | 2.74 | 0.46 | 2.58 | 2.49 |

[a]P-value cutoffs used to identify significant D-scores (False Discovery Rate of 0.05): MS_15, ≤ 1.68e-02; FG_16, ≤ 1.69e-02; OPP_17, ≤ 1.18e-02; BLVA_20, ≤ 1.42e-02; WC_6, ≤ 1.63e-02; CP_7, ≤ 1.22e-02.

**Table S5.** Sequences related to specific functions and taxa within carbon fixation pathways in prokaryotes (rTCA) cycle using KEGG pathways.

| Number of reads assigned | EC number | Orthology KEGG | Enzyme name | Gene | Metagenome best hit organism | Class | E-value |
|---|---|---|---|---|---|---|---|
| 1045 | 2.7.9.2 | K01007 | pyruvate, water dikinase | *pps* | *Hydrogenobacter thermophilus TK-6* | Aquificae | 4e-86 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 2e-88 |
| 40 | 1.4.1.1 | K00259 | alanine dehydrogenase | *ald* | *Dictyoglomus turgidum DSM 6724* | Dictyoglomia | 6e-48 |
| | | | | | *Dictyoglomus thermophilum H-6-12* | Dictyoglomia | 2e-52 |
| | | | | | *Thermosediminibacter oceani DSM 16646* | Clostridia | 5e-25 |
| | | | | | *Solitalea canadensis DSM 3403* | Sphingobacteriia | 5e-23 |
| | | | | | *Haliangium ochraceum strain DSM 14365* | Deltaproteobacteria | 2e-8 |
| | | | | | *Nocardiopsis dassonvillei subsp. dassonvillei DSM 43111* | Actinobacteria | 5e-19 |
| | | | | | *Belliella baltica DSM 15883* | Cytophagia | 5e-62 |
| | | | | | *Marinilabilia sp. AK2* | Bacteroidia | 2e-29 |
| | | | | | *Clostridium bartlettii DSM 16795* | Clostridia | 2e-10 |
| | | | | | *Marivirga tractuosa DSM 4126* | Bacteroidetes | 3e-54 |
| | | | | | *Chloroherpeton thalassium ATCC 35110* | Chlorobi | 3e-26 |
| | | | | | *Idiomarina baltica OS145* | Gammaproteobacteria | 2e-8 |
| | | | | | *Selenomonas noxia ATCC 43541* | Negativicutes | 6e-37 |
| | | | | | *Alistipes shahii WAL 8301* | Bacteroidia | 6e-51 |
| | | | | | *Planctomyces brasiliensis DSM 5305* | Planctomycetia | 4e-10 |
| | | | | | *Parabacteroides goldsteinii CL02T12C30* | Bacteroidia | 3e-15 |
| 2421 | 1.2.7.1 | K00169 | pyruvate ferredoxin oxidoreductase, alpha subunit | *por*A | *Thermocrinis albus DSM 14484* | Aquificae | 5e-68 |
| | | | | | *Hydrogenobacter thermophilus TK-6* | Aquificae | 3e-82 |
| | | K00170 | pyruvate ferredoxin oxidoreductase, beta subunit | *por*B | *Thermocrinis albus DSM 14484* | Aquificae | 5e-96 |
| | | | | | *Hydrogenobacter thermophilus TK-6* | Aquificae | 1e-96 |
| | | K00171 | pyruvate ferredoxin oxidoreductase, delta subunit | *por*D | *Hydrogenobacter thermophilus TK-6* | Aquificae | 4e-37 |
| | | | | | *Thermodesulfobacterium* | Thermodesulfobacteria | 6e-22 |
| | | K00172 | pyruvate ferredoxin oxidoreductase, gamma subunit | *por*G | *Thermocrinis albus DSM 14484* | Aquificae | 2e-92 |
| | | | | | *Hydrogenobacter thermophilus TK-6* | Aquificae | 1e-54 |
| | | | | | *Sulfurihydrogenibium yellowstonense* | Aquificae | 7e-18 |
| 1895 | 6.2.1.1 | K01895 | acetyl-CoA synthetase | *ACSS* | *Thermodesulfatator indicus DSM 15286* | Thermodesulfobacteria | 3e-11 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 5e-99 |
| | | | | | *Thauera sp. MZ1T* | Betaproteobacteria | 1e-28 |
| | | | | | *Thermofilum pendens Hrk 5* | Thermoprotei | 3e-34 |
| | | | | | *Hydrogenivirga sp. 128-5-R1-1* | Aquificae | 1e-14 |
| | | | | | *Symbiobacterium thermophilum IAM 14863* | Clostridia | 7e-32 |
| | | | | | *Caldilinea aerophila DSM 14535* | Caldilineae | 8e-95 |
| | | | | | *Aquifex aeolicus VF5* | Aquificae | 7e-39 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 681 | 4.2.1.3 | K01681 | aconitate hydratase | *ACO* | *Thermodesulfobacterium* | Thermodesulfobacteria | 5e-24 |
| | | | | | *Hydrogenobaculum sp. HO* | Aquificae | 1e-16 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 8e-68 |
| | | | | | *Aquifex aeolicus VF5* | Aquificae | 2e-74 |
| | | | | | *Hydrogenobacter thermophilus TK-6* | Aquificae | 2e-43 |
| 481 | 1.1.1.42 | K00031 | isocitrate dehydrogenase | *IDH1, IDH2, icd* | *Hydrogenobacter thermophilus TK-6* | Aquificae | 4e-67 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 9e-67 |
| 138 | 1.2.7.3 | K00174 | 2-oxoglutarate ferredoxin oxidoreductase subunit alpha | *kor*A | *Macrococcus caseolyticus JCSC5402* | Bacilli | 3e-8 |
| | | | | | *Thermosphaera aggregans DSM 11486* | Thermoprotei | 3e-52 |
| | | | | | *Chloroflexus aggregans DSM 9485* | Chloroflexia | 8e-17 |
| | | | | | *Brevibacillus laterosporus* | Bacilli | 2e-9 |
| | | | | | *Aeropyrum pernix K1* | Thermoprotei | 2e-56 |
| | | | | | *Pyrococcus furiosus DSM 3638* | Thermococci | 4e-12 |
| | | | | | *Acidilobus saccharovorans 345-15* | Thermoprotei | 4e-22 |
| | | | | | *Thermodesulfobacterium* | Thermodesulfobacteria | 3e-23 |
| | | | | | *Mahella australiensis 50-1 BON* | Clostridia | 4e-9 |
| | | | | | *Bacillus cellulosilyticus DSM 2522* | Bacilli | 1e-19 |
| | | | | | *Gemella haemolysans* | Bacilli | 6e-10 |
| | | | | | *Bacillus bataviensis* | Bacilli | 3e-38 |
| | | | | | *Brevibacillus borstelensis* | Bacilli | 5e-21 |
| | | K00175 | 2-oxoglutarate ferredoxin oxidoreductase subunit beta | *kor*B | *Thermosphaera aggregans DSM 11486* | Thermoprotei | 6e-53 |
| | | | | | *Thermaerobacter subterraneus* | Clostridia | 6e-12 |
| | | | | | *Methanothermobacter marburgensis str. Marburg* | Methanobacteria | 2e-11 |
| | | | | | *Caldisphaera lagunensis DSM 15908* | Thermoprotei | 2e-20 |
| | | | | | *Clostridium ultunense* | Clostridia | 8e-36 |
| | | | | | *Thermoproteus uzoniensis 768-20* | Thermoprotei | 3e-22 |
| | | K00177 | 2-oxoglutarate ferredoxin oxidoreductase subunit gamma | *kor*C | *Thermodesulfobacterium* | Thermodesulfobacteria | 2e-21 |
| 1649 | 6.2.1.5 | K01902 | succinyl-CoA synthetase alpha subunit | *suc*D | *Thermocrinis albus DSM 14484* | Aquificae | 1e-81 |
| | | | | | *Aquifex aeolicus VF5* | Aquificae | 1e-72 |
| | | K01903 | succinyl-CoA synthetase beta subunit | *suc*C | *Hydrogenobacter thermophilus TK-6* | Aquificae | 1e-73 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 3e-78 |
| 1289 | 1.3.99.1 | K00239 | succinate dehydrogenase flavoprotein subunit | *sdh*A | *Hydrogenobacter thermophilus TK-6* | Aquificae | 7e-66 |
| | | | | | *Aeropyrum pernix K1* | Thermoprotei | 7e-69 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 1e-65 |
| | | K00240 | succinate dehydrogenase iron-sulfur subunit | *sdh*B | *Thermocrinis albus DSM 14484* | Aquificae | 6e-42 |
| | | | | | *Hydrogenobacter thermophilus TK-6* | Aquificae | 2e-51 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | K00241 | succinate dehydrogenase cytochrome b556 subunit | *sdh*C | *Aeropyrum pernix K1* | Thermoprotei | 2e-33 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 2e-8 |
| | | K00245 | fumarate reductase iron-sulfur subunit | *frd*B | *Aquifex aeolicus VF5* | Aquificae | 3e-46 |
| | | | | | *Hydrogenivirga sp. 128-5-R1-1* | Aquificae | 2e-16 |
| | | | | | *Sulfurihydrogenibium azorense Az-Fu1* | Aquificae | 2e-7 |
| | | | | | *Hydrogenobacter thermophilus TK-6* | Aquificae | 7e-32 |
| | | | | | *Carboxydothermus hydrogenoformans Z-2901* | Clostridia | 2e-38 |
| 527 | 4.2.1.2 | K01677 | fumarate hydratase subunit alpha | *fum*A | *Thermocrinis albus DSM 14484* | Aquificae | 5e-78 |
| | | | | | *Hydrogenobacter thermophilus TK-6* | Aquificae | 5e-46 |
| | | K01678 | fumarate hydratase subunit beta | *fum*B | *Hydrogenivirga sp. 128-5-R1-1* | Aquificae | 5e-43 |
| | | | | | *Aquifex aeolicus VF5* | Aquificae | 6e-54 |
| | | | | | *Thermodesulfobacterium* | Thermodesulfobacteria | 3e-10 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 1e-18 |
| | | K01679 | fumarate hydratase, class II | *fum*C | *Aeropyrum pernix K1* | Thermoprotei | 1e-49 |
| | | | | | *Thermus thermophilus JL-18* | Deinococci | 3e-14 |
| | | | | | *Deferribacter desulfuricans SSM1* | Deferribacteres | 2e-31 |
| | | | | | *Thermaerobacter marianensis DSM 12885* | Clostridia | 2e-14 |
| | | | | | *Flexistipes sinusarabici DSM 4947* | Deferribacteres | 3e-43 |
| | | | | | *Calditerrivibrio nitroreducens DSM 19672* | Deferribacteres | 1e-23 |
| | | | | | *Rubrobacter xylanophilus DSM 9941* | Actinobacteria | 2e-37 |
| | | | | | *Bdellovibrio bacteriovorus str. Tiberius* | Deltaproteobacteria | 4e-30 |
| | | | | | *Spirochaeta smaragdinae DSM 11293* | Spirochaetia | 6e-26 |
| 358 | 1.1.1.37 | K00024 | malate dehydrogenase | *mdh* | *Hydrogenobacter thermophilus TK-6* | Aquificae | 7e-73 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 2e-79 |
| | | | | | *Acidobacterium capsulatum ATCC 51196* | Acidobacteriia | 4e-16 |
| | | | | | *Azospirillum brasilense Sp245* | Alphaproteobacteria | 3e-20 |
| | | | | | *Desulfurobacterium thermolithotrophum DSM 11699* | Aquificae | 6e-38 |
| | | | | | *Caldisphaera lagunensis DSM 15908* | Thermoprotei | 2e-54 |
| | | | | | *Desulfurivibrio alkaliphilus AHT2* | Deltaproteobacteria | 3e-9 |
| | | | | | *Thermodesulfobacterium* | Thermodesulfobacteria | 3e-29 |
| | | | | | *Gloeocapsa sp. PCC 7428* | Chroococcales | 8e-15 |
| 13 | 4.1.1.31 | K01595 | phosphoenolpyruvate carboxylase | *ppc* | *Pyrobaculum aerophilum str. IM2* | Thermoprotei | 8e-59 |
| | | | | | *Pyrobaculum islandicum DSM 4184* | Thermoprotei | 5e-53 |
| | | | | | *Pyrobaculum sp. 1860* | Thermoprotei | 5e-65 |

**Table S6.** Sequences related to specific functions and taxa within nitrogen cycle using KEGG pathways.

| Number of reads assigned | EC number | Orthology KEGG | Enzyme name | Gene | Metagenome best hit organism | Class | E-value |
|---|---|---|---|---|---|---|---|
| 7 | 1.7.99.4 | K00372 | assimilatory nitrate reductase catalytic subunit | *nas*A | *Sulfurihydrogenibium yellowstonense SS-5* | Aquificae | 2e-73 |
| | | K00374 | nitrate reductase gamma subunit | *nar*I | *Pyrobaculum aerophilum str. IM2* | Thermoprotei | 2e-58 |
| 37 | 1.7.99.7 | K04561 | nitric oxide reductase subunit B | *nor*B | *Pyrobaculum aerophilum str. IM2* | Thermoprotei | 8e-66 |
| | | | | | *Pyrobaculum sp. 1860* | Thermoprotei | 1e-32 |
| | | K02305 | nitric oxide reductase subunit C | *nor*C | *Hydrogenobacter thermophilus TK-6* | Aquificae | 3e-74 |
| | | K04748 | nitric oxide reductase NorQ protein | *nor*Q | *Thermocrinis albus DSM 14484* | Aquificae | 5e-86 |
| 12 | 1.7.7.1 | K00366 | ferredoxin-nitrite reductase | *nir*A | *Hydrogenobacter thermophilus TK-6* | Aquificae | 7e-34 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 1e-21 |
| | 1.7.7.4 | K00362 | nitrite reductase large subunit | *nir*B | *Thermocrinis albus DSM 14484* | Aquificae | 6e-84 |
| 483 | 1.13.12.16 | K00459 | nitronate monooxygenase | *ncd2* | *Thermocrinis albus DSM 14484* | Aquificae | 1e-91 |
| | | | | | *Hydrogenobacter thermophilus TK-6* | Aquificae | 1e-57 |

**Table S7.** Sequences related to specific functions and taxa within sulfur cycle using KEGG pathways.

| Number of reads assigned | EC number | Orthology KEGG | Enzyme name | Gene | Metagenome best hit organism | Class | E-value |
|---|---|---|---|---|---|---|---|
| 224 | 2.5.1.47 | K10150 | cysteine synthase | *cys*O | *Thermosphaera aggregans DSM 11486* | Thermoprotei | 3e-60 |
| | | | | | *Aeropyrum pernix K1* | Thermoprotei | 6e-28 |
| | | | | | *Pyrobaculum aerophilum str. IM2* | Thermoprotei | 6e-68 |
| | | | | | *Pyrobaculum arsenaticum DSM 13514* | Thermoprotei | 2e-73 |
| | | K01738 | cysteine synthase A | *cys*K | *Chlorobium limicola DSM 245* | Chlorobia | 2e-21 |
| | | | | | *Pyrobaculum arsenaticum DSM 13514* | Thermoprotei | 6e-39 |
| | | | | | *Pyrobaculum islandicum DSM 4184* | Thermoprotei | 2e-31 |
| | | | | | *Dehalococcoides ethenogenes 195* | Dehalococcoidia | 2e-73 |
| | | K12339 | cysteine synthase B | *cys*M | *Hydrogenobacter thermophilus TK-6* | Aquificae | 4e-78 |
| | | | | | *Thermocrinis albus DSM 14484* | Aquificae | 2e-54 |
| | | | | | Hydrogenivirga sp. 128-5-R1-1 | Aquificae | 6e-9 |
| | | | | | Aquifex aeolicus VF5 | Aquificae | 2e-68 |
| | | | | | Hydrogenobaculum sp. Y04AAS1 | Aquificae | 2e-6 |
| 6 | 2.5.1.65 | K10150 | cysteine synthase | *cys*O | *Thermosphaera aggregans DSM 11486* | Thermoprotei | 3e-60 |
| | | | | | *Aeropyrum pernix K1* | Thermoprotei | 6e-28 |
| | | | | | *Pyrobaculum aerophilum str. IM2* | Thermoprotei | 6e-68 |
| | | | | | *Pyrobaculum arsenaticum DSM 13514* | Thermoprotei | 2e-73 |
| 11 | 2.7.7.4 | K00958 | sulfate adenylyltransferase | *sat* | *Pyrobaculum arsenaticum DSM 13514* | Thermoprotei | 1e-88 |
| | | | | | *Thermodesulfatator indicus DSM 15286* | Thermodesulfobacteria | 3e-64 |
| | | | | | *Pyrobaculum oguniense TE7* | Thermoprotei | 7e-90 |
| | | | | | *Pyrobaculum aerophilum str. IM2* | Thermoprotei | 1e-36 |
| 22 | 1.8.99.2 | K00394 | adenylylsulfate reductase, subunit A | aprA | *Pyrobaculum sp. 1860* | Thermoprotei | 2e-98 |
| | | | | | *Pyrobaculum arsenaticum DSM 13514* | Thermoprotei | 1e-58 |
| | | | | | *Thermodesulfobacterium geofontis OPF15* | Thermodesulfobacteria | 2e-85 |
| | | K00395 | adenylylsulfate reductase, subunit B | aprB | *Thermodesulfobacterium geofontis OPF15* | Thermodesulfobacteria | 9e-71 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | *Pyrobaculum calidifontis JCM 11548* | Thermoprotei | 1e-31 |
| 8 | 1.8.4.8 | K00390 | phosphoadenosine phosphosulfate reductase | *cys*H | *Pyrobaculum islandicum DSM 4184* | Thermoprotei | 5e-55 |
| | | | | | *Pyrobaculum oguniense TE7* | Thermoprotei | 4e-86 |
| | | | | | *Thermosphaera aggregans DSM* | Thermoprotei | 3e-95 |
| | | | | | *Pyrobaculum arsenaticum DSM 13514* | Thermoprotei | 7e-73 |

**Figure S1.** Geographic maps showing sampling locations.

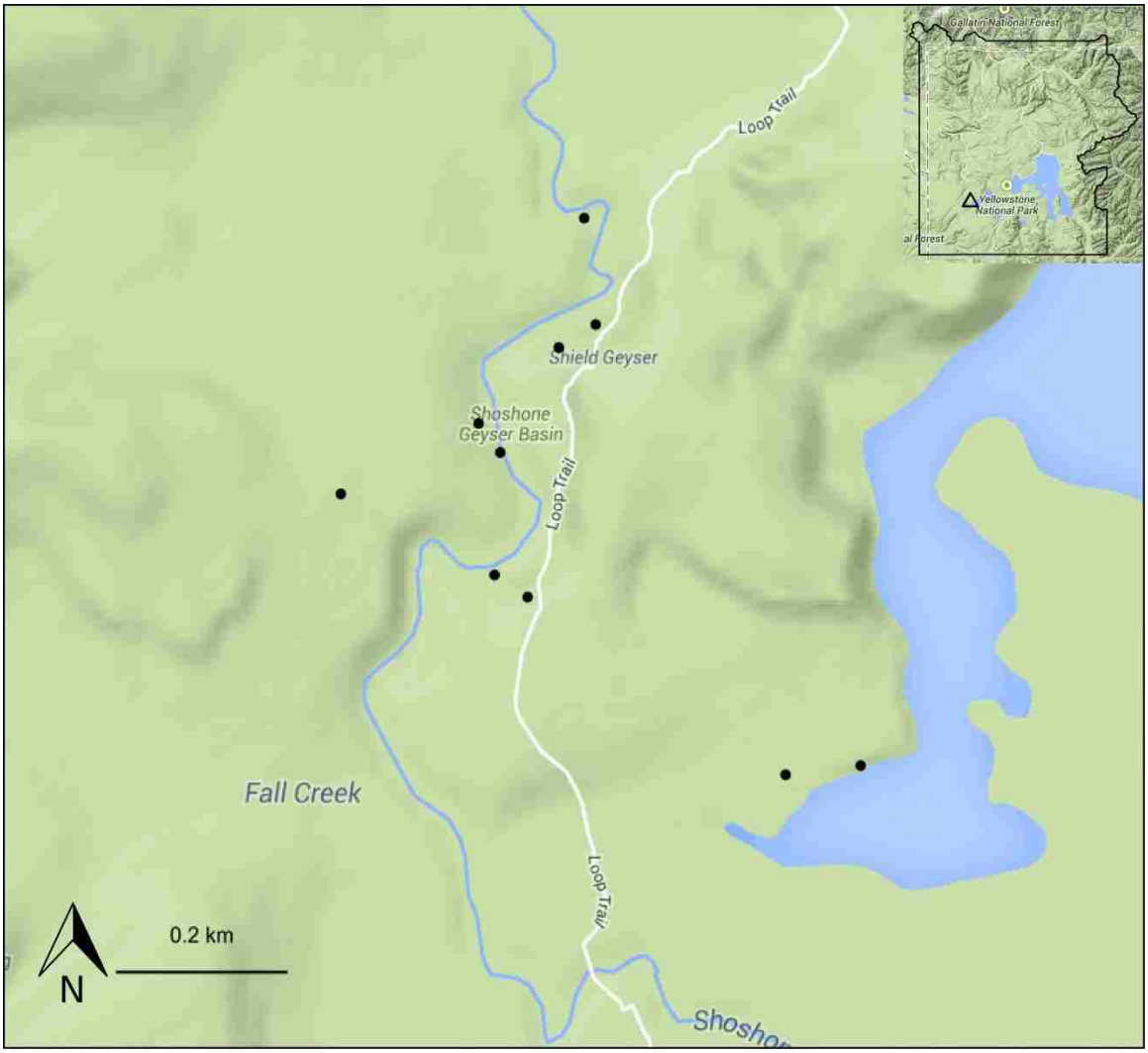


**Figure S2.** Taxonomic classification and comparison of 16S rRNA gene and metagenomic sequences.

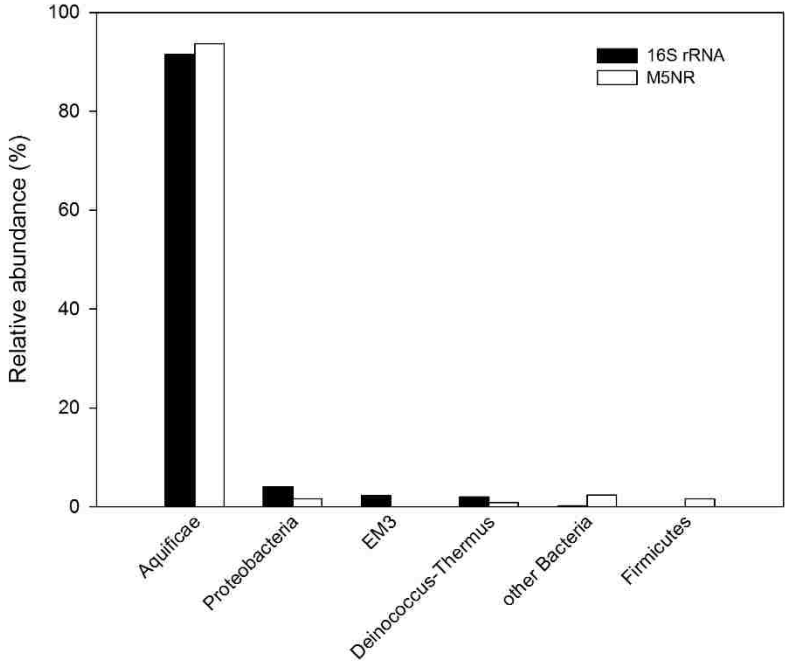


**Figure S3.** Taxonomic assignment of metagenomic reads from the site 07YSHO09A related to energy metabolism (KEGG identifiers). For Bacteria, only phyla with relative abundance of more than 1% are shown. Phyla with relative abundance of less than 1% are shown as other Bacteria.

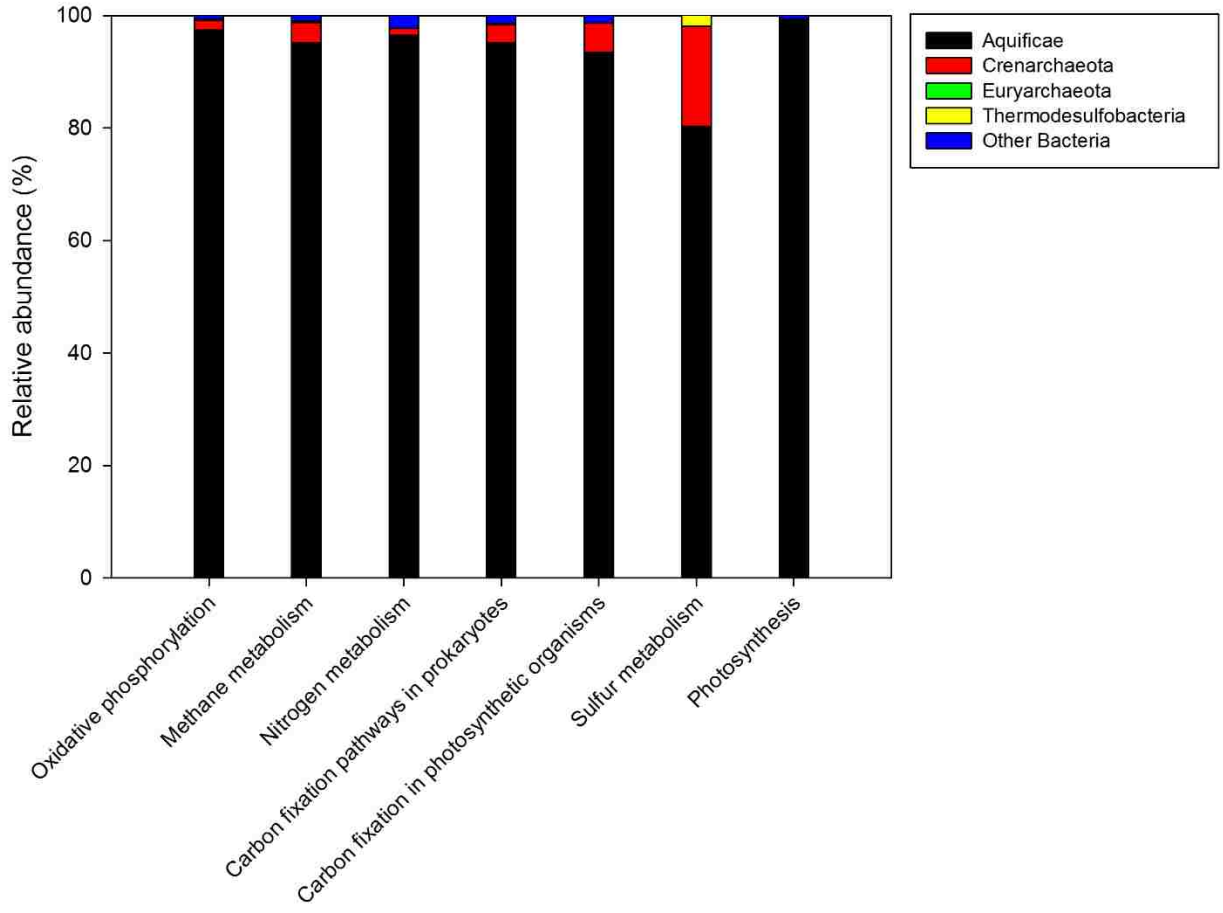


**Figure S4.** Partial carbon fixation pathways in prokaryotes (rTCA) identified by KEGG affiliation of the sequences from the site 07YSHO09A. Boxes indicate the KEGG identifiers and numbers in gray circles indicate the number of sequences assigned to the KEGG function.

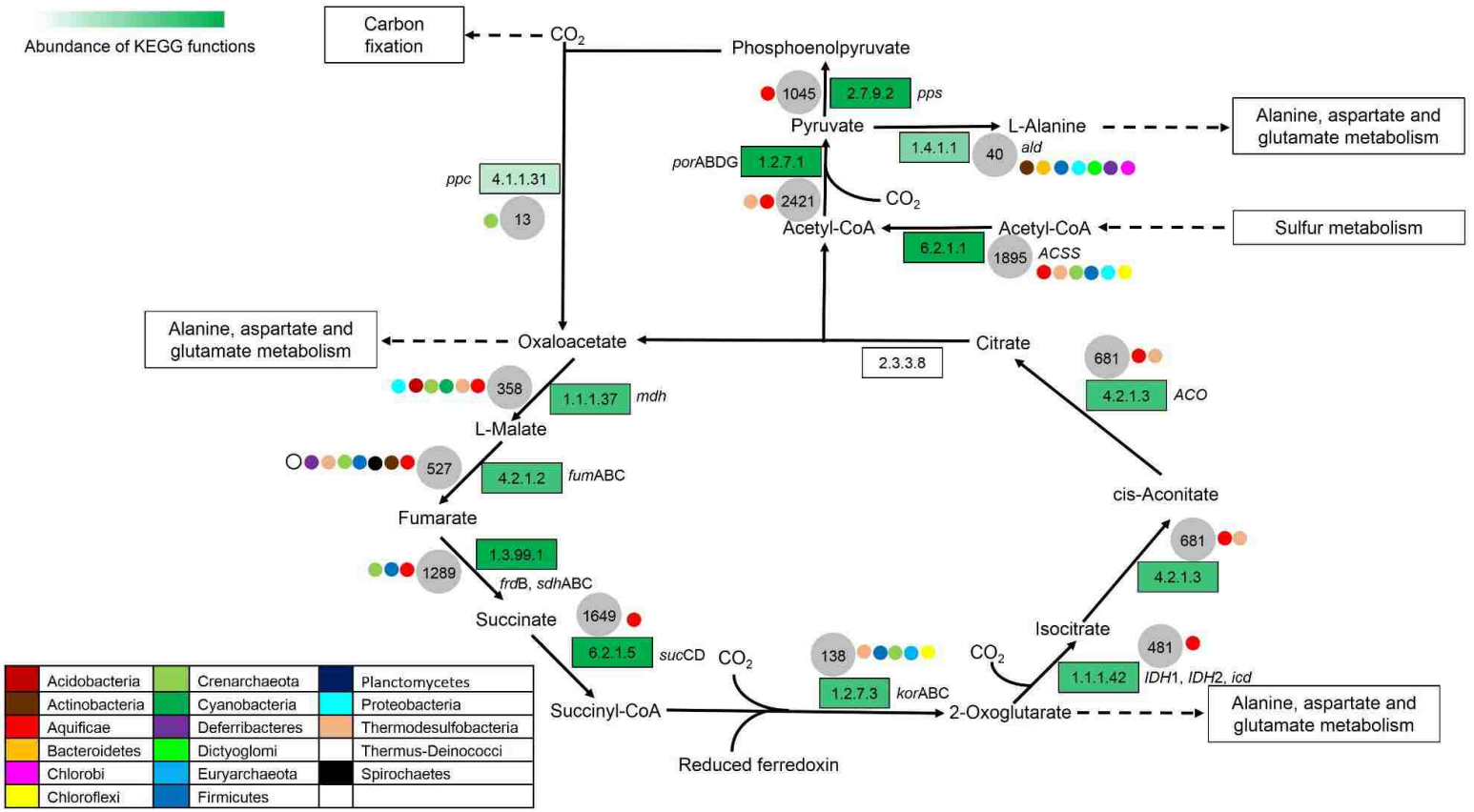


**Figure S5.** Partial nitrogen pathways identified by KEGG affiliation of the sequences from the site 07YSHO09A. Boxes indicate the KEGG identifiers and numbers in gray circles indicate the number of sequences assigned to the KEGG function.

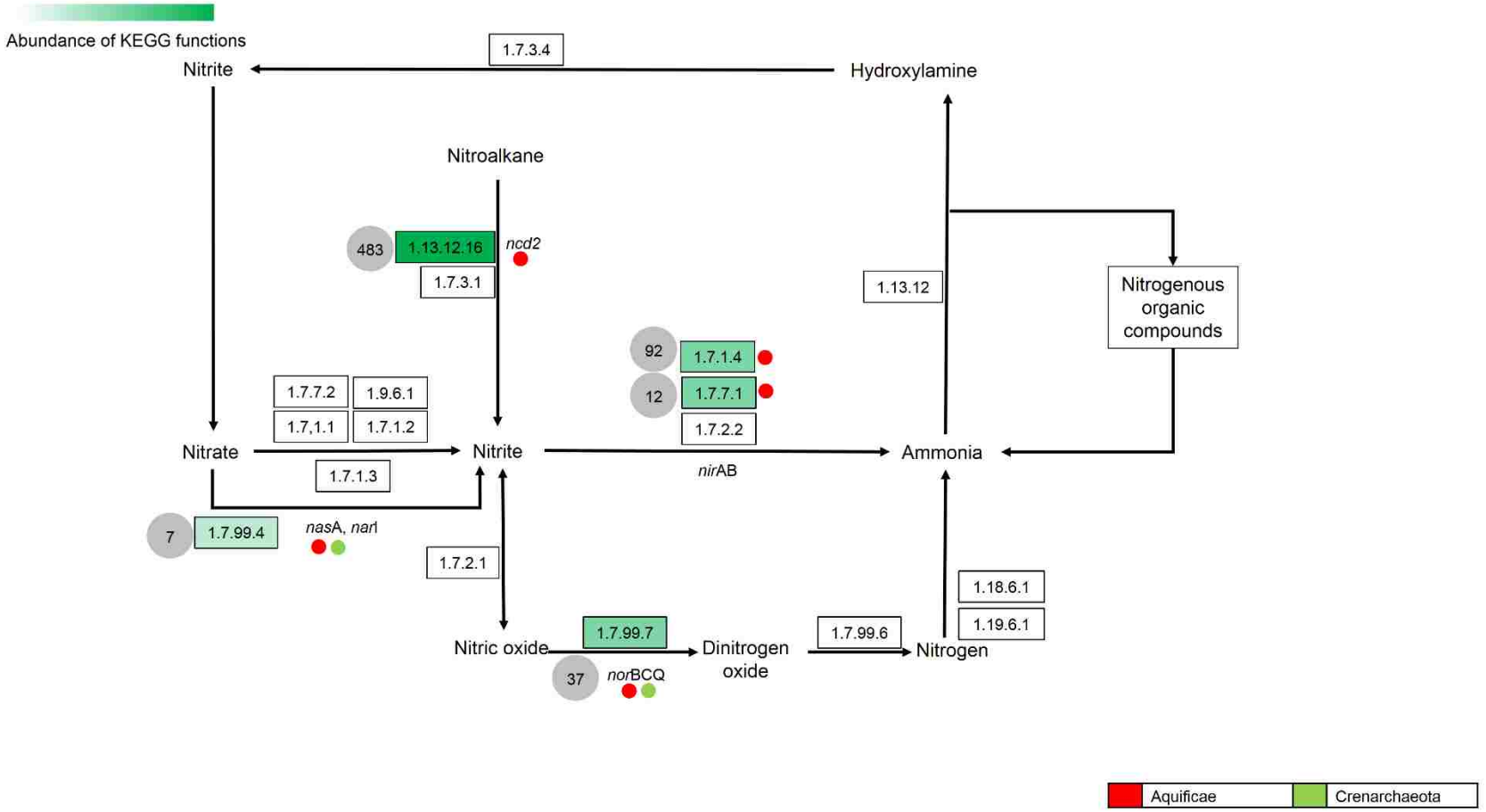
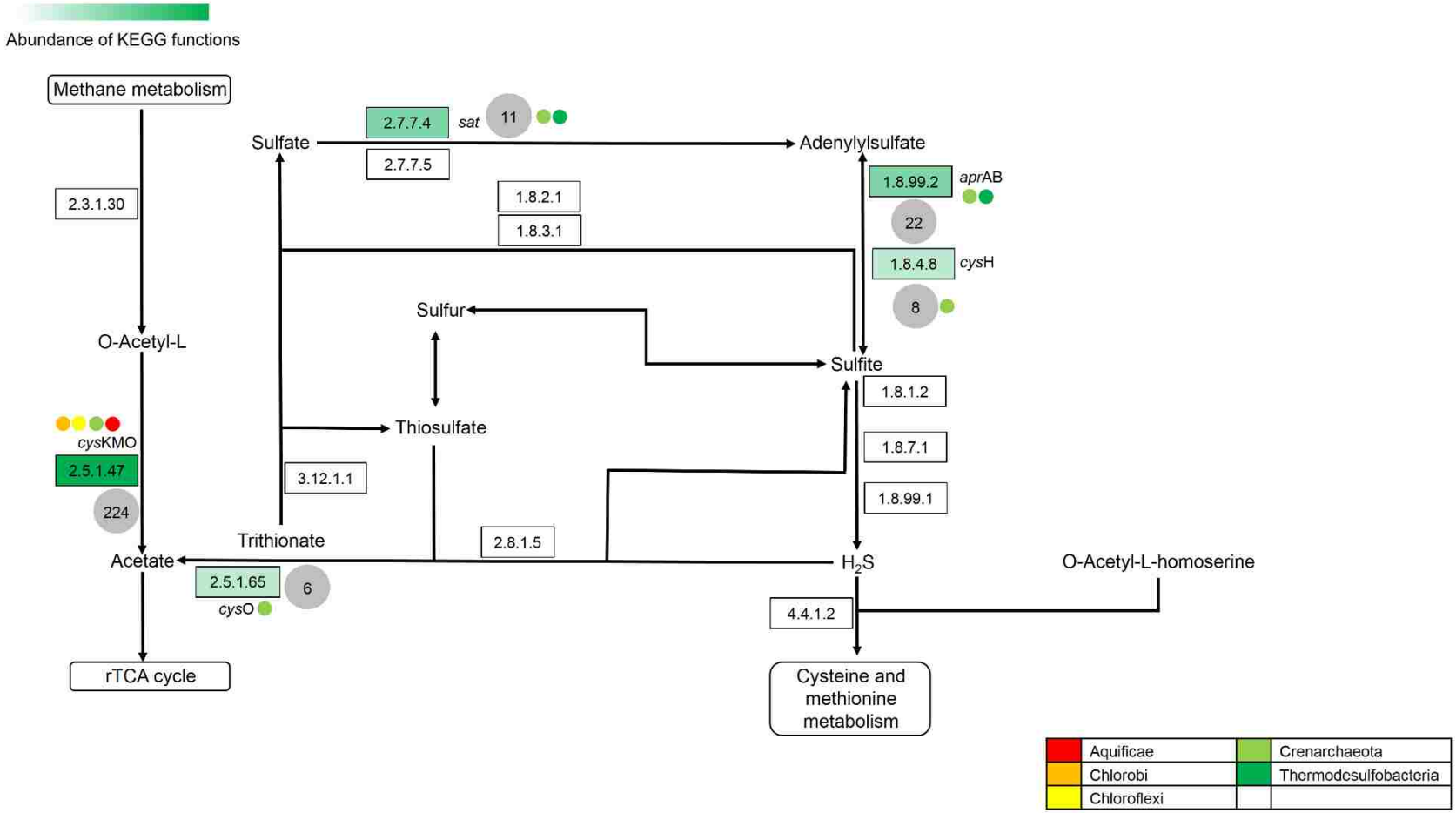


**Figure S6.** Partial sulfur pathways identified by KEGG affiliation of the sequences from the site 07YSHO09A. Boxes indicate the KEGG identifiers and numbers in gray circles indicate the number of sequences assigned to the KEGG function.



# Chapter 5

## Summary

This research is an important complementary YNP survey on the effect of different environmental variables on microbial taxonomic and functional diversity using recent high-throughput sequencing technologies. This work expands our current understanding of microbial ecology in different geothermal communities of YNP. We assessed the patterns of microbial composition and functions in two thermal spring types that were previously unexplored from a geobiological viewpoint: the pH 4 and low temperature site, and the high temperature and pH site.

Our results reported here suggest that the succession of predominant bacterial phyla is systematically correlated with temperature in circumneutral thermal springs. At high temperatures (81-87 ℃), *Aquificae* and EM3 predominated, whereas within the temperature range of photosynthesis upper limit (73-75 ℃) microbial communities were dominated by *Deinococcus-Thermus* and *Armatimonadetes*. Intermediate temperatures (63-68 ℃) and low temperatures (40-57 ℃) both favored *Cyanobacteria* and *Chloroflexi*, but bacterial communities within the low temperature range were more diverse. Furthermore, Chao1 richness and phylogenetic diversity in phototrophic communities were significantly higher than those in non-phototrophic communities were.

The wide variety of abiotic factors are known to contribute to microbial diversity in YNP thermal springs. By examining the metagenomic reads retrieved from the unexplored pH 4 and low temperature site, we outlined the first scenario regarding the microbial composition and functions in the pH 4 and low temperature YNP thermal spring. We found that the ecosystem type of this spring was similar to phototrophic communities; a high proportion of Chloroflexi, Bacteroidetes, Proteobacteria and Firmicutes were observed. This community was enriched in the COG functions related to energy production and conversion, transcription and carbohydrate transport, possibly to result from high microbial dynamics. The metabolic pathway analysis indicated the presence of assimilatory and dissimilatory nitrate reduction. Genes involved in the sulfur metabolism were mostly related to the reduction of sulfate to adenylylsulfate, sulfite and $H_2S$.

Finally, we investigated microbial diversity and community composition in 10 filamentous springs with similar physiochemical properties in the Shoshone area. Although all these samples were collected from springs that were geochemically very similar, we found that the relatively small variation of environmental variables such as temperature, pH and conductivity among the springs could shape microbial composition. Metagenomic analysis of the Aquificaceae dominant site indicated that functions related to aerobic respiration and amino acid synthesis were enriched. The results discussed here suggest that the rTCA cycle plays a dominant role in *in situ* energy metabolism.

The culture-independent research of microbial ecology in terrestrial thermal springs has been largely improved and benefited from the revolution of molecular techniques during the last decade. In future, metagenomics and metatranscriptomics

sequencing needs to be conducted using other high-throughput sequencing technologies such as PacBio and Illumina to confirm our results generated from 454 pyrosequencing. Furthermore, the microbial activities and interactions are yet to be further investigated using stable isotope analysis, metaproteomics and field experiments.