



GENETICS, GENOMICS AND –OMICS OF THERMOPHILES, 2nd Edition

EDITED BY: Kian Mau Goh, Kok-Gan Chan, Rajesh Kumar Sani,
Edgardo Rubén Donati and Anna-Louise Reysenbach
PUBLISHED IN: Frontiers in Microbiology



frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-904-9

DOI 10.3389/978-2-88945-904-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

GENETICS, GENOMICS AND –OMICS OF THERMOPHILES, 2nd Edition

Topic Editors:

Kian Mau Goh, Universiti Teknologi Malaysia, Malaysia

Kok-Gan Chan, University of Malaya, Malaysia

Rajesh Kumar Sani, South Dakota School of Mines and Technology, USA

Edgardo Rubén Donati, Universidad Nacional de La Plata, Argentina

Anna-Louise Reysenbach, Portland State University, USA

Publisher's note: In this 2nd edition, the following article has been updated: Irla M, Heggset TM, Nærdal I, Paul L, Haugen T, Le SB, Brautaset T and Wendisch VF (2016) Genome-Based Genetic Tool Development for *Bacillus methanolicus*: Theta- and Rolling Circle-Replicating Plasmids for Inducible Gene Expression and Application to Methanol-Based Cadaverine Production. *Front. Microbiol.* 7:1481.

doi: 10.3389/fmicb.2016.01481

Citation: Goh, K. M., Chan, K.-G., Sani, R. K., Donati, E. R., Reysenbach, A.-L., eds. (2019). *Genetics, Genomics and –Omics of Thermophiles*, 2nd Edition. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-904-9

Table of Contents

05 Editorial: Genetics, Genomics and –Omics of Thermophiles

Kian Mau Goh, Kok-Gan Chan, Rajesh Kumar Sani, Edgardo Rubén Donati and Anna-Louise Reysenbach

CHAPTER 1

METAGENOME OVERVIEW AND THERMOZYME APPLICATIONS

07 Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozyymes

María-Eugenia DeCastro, Esther Rodríguez-Belmonte and María-Isabel González-Siso

28 EstDZ3: A New Esterolytic Enzyme Exhibiting Remarkable Thermostability

Dimitra Zarafeta, Zalan Szabo, Danai Moschidi, Hien Phan, Evangelia D. Chrysina, Xu Peng, Colin J. Ingham, Fragiskos N. Kolis and Georgios Skretas

CHAPTER 2

MICROBIAL DIVERSITY AND METAGENOMICS

42 The Dark Side of the Mushroom Spring Microbial Mat: Life in the Shadow of Chlorophototrophs. I. Microbial Diversity Based on 16S rRNA Gene Amplicons and Metagenomic Sequencing

Vera Thiel, Jason M. Wood, Millie T. Olsen, Marcus Tank, Christian G. Klatt, David M. Ward and Donald A. Bryant

67 Metagenomic Analysis of Hot Springs in Central India Reveals Hydrocarbon Degrading Thermophiles and Pathways Essential for Survival in Extreme Environments

Rituja Saxena, Darshan B. Dhakan, Parul Mittal, Prashant Waiker, Anirban Chowdhury, Arundhuti Ghatak and Vineet K. Sharma

CHAPTER 3

THERMOPHILES GENOME

84 Aerobic Lineage of the Oxidative Stress Response Protein Rubrerythrin Emerged in an Ancient Microaerobic, (Hyper)Thermophilic Environment

Juan P. Cardenas, Raquel Quatrini and David S. Holmes

94 Gene Turnover Contributes to the Evolutionary Adaptation of Acidithiobacillus caldus: Insights From Comparative Genomics

Xian Zhang, Xueduan Liu, Qiang He, Weiling Dong, Xiaoxia Zhang, Fenliang Fan, Deliang Peng, Wenkun Huang and Huaqun Yin

107 Genome Analysis of a New Rhodothermaceae Strain Isolated From a Hot Spring

Kian Mau Goh, Kok-Gan Chan, Soon Wee Lim, Kok Jun Liew, Chia Sing Chan, Mohd Shahir Shamsir, Robson Ee and Tan-Guan-Sheng Adrian

- 117** *Genome Analysis of Thermosulfurimonas dismutans, the First Thermophilic Sulfur-Disproportionating Bacterium of the Phylum Thermodesulfobacteria*
Andrey V. Mardanov, Alexey V. Beletsky, Vitaly V. Kadnikov, Alexander I. Slobodkin and Nikolai V. Ravin
- 125** *Genome Sequencing of Sulfolobus sp. A20 From Costa Rica and Comparative Analyses of the Putative Pathways of Carbon, Nitrogen, and Sulfur Metabolism in Various Sulfolobus Strains*
Xin Dai, Haina Wang, Zhenfeng Zhang, Kuan Li, Xiaoling Zhang, Marielos Mora-López, Chengying Jiang, Chang Liu, Li Wang, Yaxin Zhu, Walter Hernández-Ascencio, Zhiyang Dong and Li Huang
- 138** *Genome-Based Genetic Tool Development for Bacillus methanolicus: Theta- and Rolling Circle-Replicating Plasmids for Inducible Gene Expression and Application to Methanol-Based Cadaverine Production*
Marta Irla, Tonje M. B. Heggeset, Ingemar Nærødal, Lidia Paul, Tone Haugen, Simone B. Le, Trygve Brautaset and Volker F. Wendisch
- 151** *The Complete Genome Sequence of Hyperthermophile Dictyoglomus turgidum DSM 6724™ Reveals a Specialized Carbohydrate Fermentor*
Phillip J. Brumm, Krishne Gowda, Frank T. Robb and David A. Mead



Editorial: Genetics, Genomics and –Omics of Thermophiles

Kian Mau Goh^{1*}, Kok-Gan Chan², Rajesh Kumar Sani³, Edgardo Rubén Donati⁴ and Anna-Louise Reysenbach⁵

¹ Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia, ² Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia,

³ Department of Chemical and Biological Engineering, South Dakota School of Mines and Technology, Rapid City, SD, USA,

⁴ CINDEFI (CCT, La Plata-CONICET, UNLP), Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina,

⁵ Department of Biology, Portland State University, Portland, OR, USA

Keywords: comparative genomics, extremophile, hot spring, hyperthermophile, thermozyme, metagenome

Editorial on the Research Topic

Genetics, Genomics and –Omics of Thermophiles

OPEN ACCESS

Edited by:

Jesse G. Dillon,
California State University,
Long Beach, USA

Reviewed by:

Jesse G. Dillon,
California State University,
Long Beach, USA
Matthew Schrenk,
Michigan State University, USA

*Correspondence:

Kian Mau Goh
gohkianmau@utm.my

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 09 February 2017

Accepted: 17 March 2017

Published: 03 April 2017

Citation:

Goh KM, Chan K-G, Sani RK,
Donati ER and Reysenbach A-L
(2017) Editorial: Genetics, Genomics
and –Omics of Thermophiles.
Front. Microbiol. 8:560.
doi: 10.3389/fmicb.2017.00560

Thermophilic Archaea and Bacteria occupy heated environments. Advancement of next-generation sequencing (NGS), single-cell analyses, and combinations of –omics and microscopic technologies have resulted in the discovery of new thermophiles. This e-book consists of a review, and 10 original articles authored by 94 authors. The main aim of this Research Topic of Frontiers in Microbiology was to provide a platform for researchers to describe recent findings on the ecology of thermophiles using NGS, functional genomics, comparative genomics, gene evolution, and extremozyme discovery.

The review by DeCastro et al. discussed the approaches currently available in assessing the taxonomy and functional metagenomics of thermophiles in high temperature environments. The review also provides limitations or challenges for each approach in the discovery of novel thermozymes that include lipolytic enzymes, glycosidases, proteases, and oxidoreductases.

Nearly 50 years ago, Thomas Brock was among the earliest researchers who elucidated the existence of living organisms in hot springs in Yellowstone National Park, YNP (Brock, 1967). In this e-book, Thiel et al. revisited Mushroom Spring (60°C) and examined the microbial diversity in the orange-colored undermat using NGS shotgun sequencing and 16S rRNA amplicon analyses. The phylum *Chloroflexi* dominated 49% of total OTUs, followed by *Thermotogae*, *Armatimonadetes* (previously known as candidate division OP10), *Aquificae*, *Cyanobacteria*, *Atribacteria* (candidate phylum OP-9/JS1), *Nitrospirae* and others. Thiel et al. showed that the dominant taxon, *Roseoflexus*, had high microdiversity of the 16S rRNA gene sequences which most likely represent different ecotypes with specific ecological adaptations. In a separate article, Saxena et al. performed shotgun metagenomic and 16S rRNA amplicons sequencing from samples collected from three Indian hot springs (43.5–98°C, pH 7.5–7.8). The alpha- and beta-diversity of thermophiles in seven distinct sites were compared and the authors concluded that the temperatures significantly affected the microbial community structure. These sites were dominated by phyla *Proteobacteria*, *Thermi*, *Chloroflexi*, *Bacteroidetes*, *Firmicutes*, and *Thermotogae*. Data from shotgun metagenome sequencing were used to assess hydrocarbon degradation pathways in the Anthoni hot spring. One of the interesting insights from Saxena et al. is that all enzymes involved in a particular hydrocarbon degradation pathway were not found in a single microbial species; therefore, the degradation could only be completed by consortium members of the microbial community.

Genomes of several hyper- and thermophilic bacteria were sequenced and reported in this e-book. *Dictyoglomus turgidum* has an optimum growth temperature (OGT) of 72°C. Brumm et al. analyzed genome content of *D. turgidum* from multiple aspects including metabolic pathways, polysaccharides degradation and transport, energy generation, DNA repair and recombination, and stress responses. *D. turgidum* has an abundance of glycosyl hydrolases, 16 of which were examined for their activities. *D. turgidum* can utilize most plant-based polysaccharides, except crystalline cellulose. Zarafeta et al. isolated a *Dictyoglomus* sp. Ch5.6.S from an *in situ* enrichment culture containing xanthan gum established in a Yunnan hot spring. A new hyperthermostable esterolytic enzyme (EstDZ3) was identified from the genome sequence. The EstDZ3 is likely a carboxylesterase as it reacts best on fatty acid esters with short to medium chain lengths. The enzyme exhibited a half-life of more than 24 h when incubated at 80°C. Clearly, both articles suggested *Dictyoglomus* is an interesting genus with biotechnological potential.

Three articles in this Research Topic provide novel insights into sulfur-metabolizing prokaryotes. Zhang et al. studied the evolution of six *Acidithiobacillus caldus* strains using comparative genomic approaches. The authors identified many mobile genetic elements and showed that gene gains and losses all drive the genomic diversification in this species. Dai et al. compared the genome of *Sulfolobus* sp. A20 with the *Sulfolobus solfataricus*, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, and *Sulfolobus tokodaii*, and identified 1,801 core genes sequences. Genes in central carbon metabolism and ammonium assimilation are highly conserved in *Sulfolobus*. The genes for sulfur oxygenase/reductase and inorganic nitrogen utilization which are less conserved probably due to the presence or remnants of insertion sequence elements. The anaerobic *Thermosulfurimonas dismutans* S95^T was isolated from a deep-sea hydrothermal vent by Mardanov et al. and they showed that *T. dismutans* had the sulfur-disproportionating capability even without the need of direct contact of the cells to solid elemental sulfur. It is therefore likely that the soluble glutathione persulfide is the actual substrate entering the disproportionation pathway. Mardanov et al. proposed a model of sulfur metabolism and related pathways in *T. dismutans*.

Goh et al. reported the isolation and genome description of a new bacterium strain RA (OGT: 50–60°C) in the *Rhodothermaceae*. The strain RA is likely to be a new genus due to its low 16S rRNA and housekeeping genes (e.g., *recA*, *rpoD*, and *gyrB*) similarity to other genera in *Rhodothermaceae*. Goh et al. also compared the genome of this bacterium with *Rhodothermus marinus* DSM 4252^T and *Salinibacter ruber* DSM^T and showed that it has putative genes for adaptation to osmotic stress and survival in a high sulfidic hot springs. Using phylogenetic analyses and sequence similarity networks, Cardenas et al. reported the phylogenomic analysis of 2,631 representative rubrerythrins, a group of proteins involved in oxidative stress defense. The authors proposed that “aerobic-type” rubrerythrins underwent a separate adaptation process than that of the “cyanobacterial group.” Lastly, Irla et al. compared four different plasmids that were able to replicate in *Bacillus methanolicus*. They reported the effects of copy number, expression levels and stability of these plasmids in *B. methanolicus*. The article provided new tools for genetic engineering of *B. methanolicus*. We hope that this e-book can stimulate the research community to integrate –omics and bioinformatics tools in understanding the biology of heated environments and thermophiles.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

KG is supported by the UTM GUP grants (14H67 and 15H50). KC gratefully acknowledges the financial support provided by University of Malaya—Ministry of Higher Education High Impact Research Grant (UM.C/625/1/HIR/MOHE/CHAN/01 Grant No. A-000001-50001 and UM.C/625/1/HIR/MOHE/CHAN/14/1 Grant No. H-50001-A000027). RS gratefully acknowledges the financial support provided by National Aeronautics and Space Administration (Grant # NNX16AQ98A). ED is thankful to grant PICT 2013-0630. This was supported by grants to AR (NASA grant #NNX16AJ66G and NSF DEB 1134877).

REFERENCES

Brock, T. D. (1967). Life at high temperatures. *Science* 158, 1012–1019.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Goh, Chan, Sani, Donati and Reysenbach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozyms

María-Eugenia DeCastro, Esther Rodríguez-Belmonte and María-Isabel González-Siso *

Grupo EXPRELA, Centro de Investigaciones Científicas Avanzadas (CICA), Departamento de Biología Celular e Molecular, Facultad de Ciencias, Universidade da Coruña, A Coruña, Spain

OPEN ACCESS

Edited by:

Kian Mau Goh,
Universiti Teknologi Malaysia, Malaysia

Reviewed by:

Alexander V. Lebedinsky,
Winogradsky Institute of Microbiology,
Russia

Jeremy Dodsworth,
California State University, USA

Rup Lal,
University of Delhi, India

*Correspondence:

María-Isabel González-Siso
migs@udc.es

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 28 July 2016

Accepted: 12 September 2016

Published: 27 September 2016

Citation:

DeCastro M-E, Rodríguez-Belmonte E
and González-Siso M-I (2016)
Metagenomics of Thermophiles with a
Focus on Discovery of Novel
Thermozyms.
Front. Microbiol. 7:1521.
doi: 10.3389/fmicb.2016.01521

Microbial populations living in environments with temperatures above 50°C (thermophiles) have been widely studied, increasing our knowledge in the composition and function of these ecological communities. Since these populations express a broad number of heat-resistant enzymes (thermozymes), they also represent an important source for novel biocatalysts that can be potentially used in industrial processes. The integrated study of the whole-community DNA from an environment, known as metagenomics, coupled with the development of next generation sequencing (NGS) technologies, has allowed the generation of large amounts of data from thermophiles. In this review, we summarize the main approaches commonly utilized for assessing the taxonomic and functional diversity of thermophiles through metagenomics, including several bioinformatics tools and some metagenome-derived methods to isolate their thermozyms.

Keywords: metagenomics, thermophiles, thermozyms, bioinformatics, NGS

INTRODUCTION

Thermophiles (growing optimally at 50°C or higher), extreme thermophiles (65–79°C) and hyperthermophiles (above 80°C), categories defined per Wagner and Wiegel (2008), are naturally found in various geothermally heated regions of Earth such as hot springs and deep-sea hydrothermal vents. They can also be present in decaying organic matter like compost and in some man-made environments. Besides the high temperatures, many of these environments are characterized by extreme pH or anoxia. The adaptation to these harsh habitats explains the high genomic and metabolic flexibility of microbial communities in these ecosystems (Badhai et al., 2015) and makes thermophiles and their thermostable proteins very suitable for some industrial and biotechnological applications. Therefore, screening for novel biocatalysts from extremophiles has become a very important field. In the last few years, novel thermostable polymerases (Moser et al., 2012; Schoenfeld et al., 2013), beta-galactosidases (Wang et al., 2014), esterases (Fuciños et al., 2014), and xylanases (Shi et al., 2013), among others, have been described and characterized, opening a new horizon in biotechnology.

Apart from the bioprospecting purposes, the analysis of these high-temperature ecosystems and their inhabitants can improve our understanding of microbial diversity from an ecological point of view and increase our knowledge of heat-tolerance adaptation (Lewin et al., 2013). Additionally, the study of thermophiles provides a better comprehension about the origin and evolution of earliest life, as they are considered to be phenotypically most similar to microorganisms present on the primitive Earth (Farmer, 1998; Stetter, 2006). In addition to the bacterial and archaeal communities,

there is an increasing interest in the study of the viral populations living in high-temperature ecosystems, as viruses are reported to be the main predators of prokaryotes in such environments (Breitbart et al., 2004), participating in the biogeochemical cycles and being important exchangers of genetic information (Rohwer et al., 2009).

The first studies of these extremophiles required their cultivation and isolation (Morrison and Tanner, 1922; Brock and Freeze, 1969; Fiala and Stetter, 1986; Prokofeva et al., 2005; De la Torre et al., 2008). Although these techniques have been improved (Tsudome et al., 2009; Pham and Kim, 2012), the growth of thermophiles under laboratory conditions is still a limitation for the insights into the microbial diversity. The evolution of high-throughput DNA sequencing has enabled the development and improvement of metagenomics: the genomic analysis of a population of microorganisms (Handelsman, 2004). Different high-temperature ecosystems like hot springs (Schoenfeld et al., 2008; Gupta et al., 2012; Ghelani et al., 2015; López-López et al., 2015b; Sangwan et al., 2015), deserts (Neveu et al., 2011; Fancello et al., 2012; Adriaenssens et al., 2015), compost (Martins et al., 2013; Verma et al., 2013), hydrocarbon reservoirs (de Vasconcellos et al., 2010; Kotlar et al., 2011), hydrothermal vents (Anderson et al., 2011, 2014), or a biogas plant (Ilmberger et al., 2012) have been analyzed using this metagenomic approach. These whole community DNA based studies were initially focused to answering the question “who are there” and now have shifted to finding out “what are they doing,” allowing us the access to the natural microbial communities and their metabolic potential (Kumar et al., 2015).

DIVERSITY ANALYSIS OF THERMOPHILES

Targeted Metagenomics

The universality of the 16S rRNA genes makes them an ideal target for phylogenetic analysis and taxonomic classification (Olsen et al., 1986). Schmidt et al. (1991) were the pioneers in performing a community characterization based on metagenome amplified 16S rRNA genes. Since then, the diversity of other natural microbial communities started to be studied using this approach. Jim's Black Pool hot spring, in Yellowstone National Park (YNP), is reported to be the first metagenome-derived analysis of a high-temperature environment based on 16S rRNA gene profiling (Barns et al., 1994).

Initially, these studies required the amplification of the 16S rRNA genes followed by either denaturing gradient gel electrophoresis (DGGE, Muyzer et al., 1993) and sequencing or by cloning of the amplicons. In this case, the libraries obtained were screened using direct Sanger sequencing or restriction fragment length polymorphism (RFLP) analysis (Liu et al., 1997; Baker et al., 2001), to select and sequence those clones with unique patterns (Figure 1). As an example, the effect of pH, temperature, and sulfide in the hyperthermophilic microbial communities living in hot springs of northern Thailand was determined with the amplification of complete 16S rRNA genes followed by DGGE separation and sequencing (Purcell et al., 2007). In a different study, RFLP analysis and sequencing of clones with unique RFLP patterns was used to reveal the presence

of abundant novel *Bacteria* and *Archaea* sequences in a 16S rRNA gene clone library prepared from the 55°C water and sediments of Boiling Spring Lake in California, USA (Wilson et al., 2008).

With the development of next generation sequencing (NGS) technologies, more samples can be analyzed at lower sequencing cost and time, improving the production of 16S rRNA gene-based biodiversity studies. Additionally, the use of NGS allows to recover more information about the taxonomy of the sample, as reflected by Song et al. (2013), who obtained greater detail in the community structures from 16 Yunnan and Tibetan hot springs with high throughput 454-pyrosequencing than previous studies using conventional clone library and DGGE (Song et al., 2010). These analyses often rely on a partial sequence of 16S rRNA genes, as the read length of most NGS platforms is relatively short. For this purpose, primers designed for amplification of variable regions of 16S rRNA, like the V4–V8 (Hedlund et al., 2013; Huang et al., 2013), or the V3–V4 (Chan et al., 2015) are used. In the last few years, a high amount of extreme temperature environments have been analyzed with this procedure, especially hot springs, some of which are summarized in Table 1. Thanks to this strategy, a large number of 16S rRNA sequences have been produced and deposited in public databases like the Ribosomal Database Project (RDP, Cole et al., 2014) or the SILVA database (Quast et al., 2013).

Even when the process of generating and sequencing the libraries is relatively fast, this PCR-based approach is biased due to limitations of primers, PCR artifacts like chimeras (Ashelford et al., 2005) and inhibitors that could be present in the sample hindering the amplification (Urbietta et al., 2015). Although there are some previous studies focused on primer design to acquire a high coverage rate (Wang and Qian, 2009), difficulties of the primers in recognizing all the 16S rRNA sequences have been described (Cai et al., 2013), leading to the unequal amplification of species 16S rRNA genes. Furthermore, analysis of 16S rRNA sequences can result in misidentification of the taxonomy, as closely related species may harbor nearly identical 16S rRNA genes. In addition, an overestimation of the community diversity could occur since sporadic cases of distant horizontal transfer of the 16S rRNA gene have been inferred from comparisons of these genes within and between individual genomes (Yap et al., 1999; Acinas et al., 2004).

The most used taxonomically informative genomic marker in targeted metagenomics is 16S rRNA, but there are other signature sequences that have been used to study the diversity of thermophiles such as internal transcribed spacer regions (ITS, Ferris et al., 2003) or 18S rRNA genes (Wilson et al., 2008), as well as different protein-coding genes such as *aoxB* gene fragment, which encodes the catalytic subunit of As(III) oxidase, employed by Sharma et al. (2015) in combination to 16S rRNA to assess the microbial diversity of the Soldhar hot spring in India.

Apart from the above mentioned amplicon-targeting strategy, in some studies a sequence capture technique coupled with NGS is driven to enrich the targeted sequences present in the metagenome. Captured metagenomics involves custom-designed hybridization-based oligonucleotide probes that hybridize with the metagenomic libraries followed by the sequencing of the probe-bound DNA fragments. Denonfoux et al. (2013) firstly

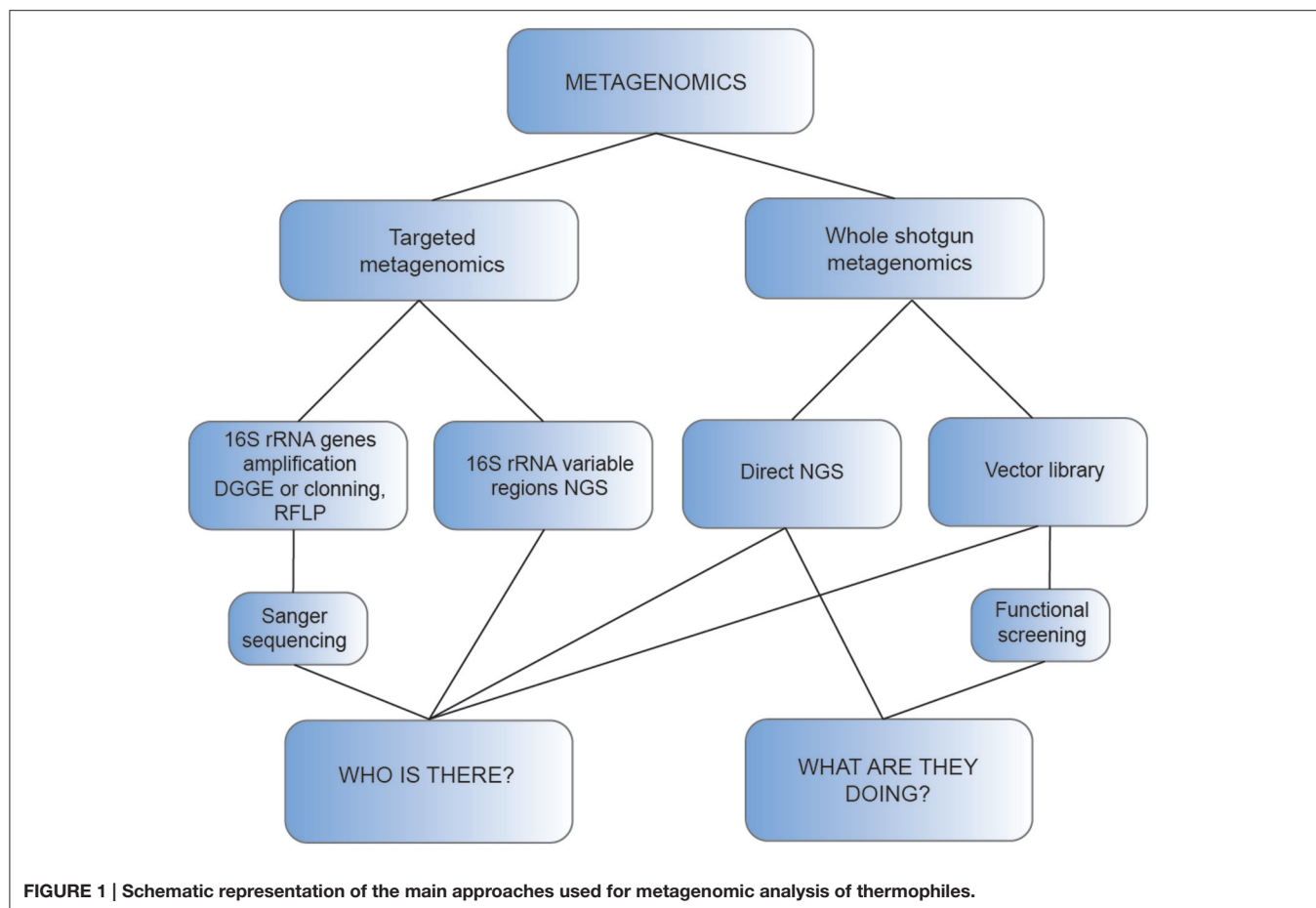


TABLE 1 | Examples of hot springs studied using the amplification of the variable regions of 16S rRNA.

| Hot spring | Type of sample | pH | Temperature (°C) | Sequencer | Region amplified | References |
|--|-----------------------------------|-----------|------------------|------------------|------------------|-------------------------|
| Siloam, Limpopo, South Africa | Water | 9.5 | 63 | Roche 454 GS FLX | V4–V7 | Tekere et al., 2011 |
| Lake Bogoria, Kenya | Water, sediment and microbial mat | 8.9–9.5 | 40–80 | Roche 454 GS FLX | V3–V4 | Dadheech et al., 2013 |
| Arzakan and Jermuk, Armenia | Water and sediment | 7.20–7.50 | 40–53 | Roche 454 GS FLX | V4–V8 | Hedlund et al., 2013 |
| Bacon Manito Geothermal Field, Philippines | Sediment | 3.72–6.58 | 60–92 | Roche 454 GS FLX | V4–V8 | Huang et al., 2013 |
| Furnas Valley, São Miguel, Azores | Water, sediment and microbial mat | 2.5–8 | 51–92 | Roche 454 GS FLX | V2–V3 | Sahm et al., 2013 |
| Yunnan province and Tibet, China | Sediment | 3.2–8.6 | 47–96 | Roche 454 GS FLX | V4 | Song et al., 2013 |
| Zavarzin, Uzon Caldera, Kamchatka, Russia | Microbial mat | 6.6 | 56–58 | Roche 454 GS FLX | V3 | Rozanov et al., 2014 |
| Sungai Klah, Malaysia | Water and sediment | 8 | 75–85 | Illumina MiSeq | V3–V4 | Chan et al., 2015 |
| Jakrem, Meghalaya, India | Microbial mat | – | – | Illumina | V3 | Panda et al., 2015 |
| Odisha, Deulajhari, India | Sediment | 7.14–7.83 | 43–55 | Illumina GAII | V3–V4 | Singh and Subudhi, 2016 |

used this procedure to explore the methanogen diversity in Lake Pavin (Frech Massif Central), showing that this GC-independent procedure is less biased and can detect broader diversity than traditional amplicon sequencing. The same approach has been used to enhance the capture of functional genes coding for

carbohydrate-active enzymes and proteases in agricultural soils (Manoharan et al., 2015), and could also be an interesting tool to study thermophilic populations.

Another method for targeted metagenomics enrichment is stable isotope probing (SIP) in which the environmental

microorganisms are grown in the presence of substrates labeled with isotopes. As a consequence of metabolic activity, the isotope (usually ^{13}C or ^{15}N) is incorporated into the nucleic acids of the microbes metabolizing the substrate, increasing the density of DNA or RNA that can be after separated from unlabelled ones (Coyotzi et al., 2016). The high-density community DNA is then used as template to amplify by PCR the 16S rRNA sequences (Brady et al., 2015) and/or some functional genes involved in the selected metabolic pathway, thus allowing the study of the microorganisms that are actively participating in the processes of interest. Gerbl et al. (2014) used this technique to assess the microbial populations implicated in the carbon cycle in the Franz Josef Quelle radioactive thermal spring (Austrian Central Alps).

Although the strategies of targeted metagenomics can be used to infer the taxonomic diversity of the community (16S rRNA gene profiling) or particular aspects of its functional diversity, a broader view of functional diversity, i.e., a more exhaustive answer to the question “what are they doing,” is provided by shotgun metagenomics (Figure 1).

Shotgun Metagenomics

Random sequencing of metagenomic DNA using high-throughput sequencing technology is becoming increasingly common. In this approach, DNA is extracted from the whole community and subsequently sheared into small fragments that are independently sequenced. At present, this is considered the most accurate method for assessing the structure of an environmental microbial community, since it does not comprise any selection and reduces technical biases, especially the ones introduced by amplification of the 16S rRNA gene (Lewin et al., 2013). Shah et al. (2011) compared bacterial communities analyzed with both 16S rRNA and whole shotgun metagenomics, revealing that the taxonomy derived from these two different approaches cannot be directly compared. This study also proposed that low abundance species are best identified through 16S rRNA gene sequencing. Therefore, some high-temperature studies use, in parallel, both techniques to assess the taxonomic composition of the microbial community (Dadheech et al., 2013; Klatt et al., 2013; Chan et al., 2015).

The biodiversity of several hot environments such as oil reservoirs (Kotlar et al., 2011), compost (Martins et al., 2013), or hot springs (Zamora et al., 2015; Mehetre et al., 2016), was studied using shotgun metagenomics sequencing. Some of them are summarized in Table 2.

Development of NGS has greatly enhanced this approach. The most widely used platforms for this kind of analysis in high temperature environments are Illumina and Roche 454 (Table 2). Illumina currently offers the highest throughput per run and the lowest cost per-base (Liu et al., 2012), generating read lengths up to 300 bp. On the other hand, Roche 454 gives longer reads (1 kb maximum), which are easier to map to a reference genome; however it is more expensive and has lower throughput (van Dijk et al., 2014). Even though they have substantial differences (Kumar et al., 2015), some studies have demonstrated that the information recovered from both sequencing platforms is comparable when analyzing the biodiversity of the same sample (Luo et al., 2012).

The main limitations of shotgun metagenome sequencing include its relatively expensive setup cost and the requirement of very high computing power for data storage, retrieval, and analysis. Another important drawback of this approach is that high quality whole community DNA is needed, which makes the extraction a critical step in the process of generating metagenomic data. Therefore, some studies have focused on the improvement of metagenomic DNA extraction from thermal environments (Mitchell and Takacs-Vesbach, 2008; Li et al., 2013a; Gupta et al., 2016). Nowadays the NGS platforms allow sequencing with low inputs of DNA, nevertheless in some cases it is necessary to amplify the metagenomic DNA to obtain enough quantity for preparing the sequencing libraries. As an example, Nakai et al. (2011) used multiple displacement amplification with Phi29 to sequence the metagenome of the hydrothermal fluid of the Mariana Trough, an active back-arc basin in the western Pacific Ocean. This amplification step is frequently required to generate viral metagenomic libraries, introducing a subsequent bias (Kim and Bae, 2011), as the extraction of enough high quality viral nucleic acids is a difficult process that usually relies on virus concentration methods.

To assess the taxonomic diversity with the short metagenomic reads obtained after sequencing, there are several non-exclusive approximations that can be done: analyzing taxonomically informative marker genes, grouping sequences into defined taxonomic groups (binning) or/and assembling sequences into definite genomes (Sharpton, 2014).

As mentioned before, the most frequently used taxonomically informative marker genes are rRNA genes or protein-coding genes that tend to be single copy and common to microbial genomes. In this approach, those reads that are homologs to the marker gene are identified in the sequences of the metagenome and annotated using sequence or phylogenetic similarity to the marker gene database sequences. Bioinformatics applications for this purpose include MetaPhyler (Liu et al., 2010), EMIRGE (Miller et al., 2011), and AMPHORA (Wu and Scott, 2012). Gladden et al. (2011) used EMIRGE to reconstruct near full-length small subunit (SSU) rRNA genes from metagenomic Illumina sequences to determine the taxonomy of compost-derived microbial consortia adapted to switchgrass at 60°C, finding a low-diversity community with predominance of *Rhodothermus marinus* and *Thermus thermophilus*. In another study, Klatt et al. (2011) used AMPHORA to identify the phylogenetic and functional marker genes in the assemblies of several hot springs cyanobacterial metagenomes from YNP. These studies allowed the discovery of novel chlorophototrophic bacteria belonging to uncharacterized lineages within the order Chlorobiales and within the Kingdom Chloroflexi. In a similar approach, Lin et al. (2015) and Colman et al. (2016) used a 16S rRNA gene-based diversity method blasting the metagenomic reads against the SILVA reference database to characterize bacterial populations in Shi-Huang-Ping acidic hot spring (Taiwan) and in two thermal springs in YNP, respectively.

Taxonomic binning is defined as the process of grouping reads or contigs and assigning them to operational taxonomic units, depending on information such as sequence similarity, sequence composition or read coverage (Dröge and McHardy,

TABLE 2 | Examples of high temperature environments studied with shotgun metagenomics.

| Environment | Location | Type of sample/s | pH | Temperature (°C) | Sequencer | Total reads | Size (Mbp) | References |
|----------------------------|---------------------------------|-----------------------------------|----------|------------------|-----------------|---------------|------------|---------------------------------|
| Hot spring | Yellowstone National Park, USA | Microbial mat | – | 60–65 | Sanger | 161,976 | 167 | Klatt et al., 2011 |
| | Yellowstone National Park, USA | Microbial mat | 6.2–9.1 | 40–60 | Sanger | – | 320.6 | Klatt et al., 2013 |
| | Yellowstone National Park, USA | Microbial mat | 3.5 | 60–78 | Roche 454 | – | – | Kozubal et al., 2013 |
| | Yellowstone National Park, USA | Microbial mat | 2.5–7.8 | 65–80 | Sanger | 75,000 | 60 | Inskeep et al., 2010 |
| | Yellowstone National Park, USA | Microbial mat and sediment | 2.5–6.4 | 70–85 | Sanger | – | 250 | Inskeep et al., 2013 |
| | Yellowstone National Park, USA | Water | 1.8 | 79 | Roche 454 | 1,604,079 | – | Menzel et al., 2015 |
| | Yellowstone National Park, USA | Water | 3.5–4.0 | 92 | Roche 454 | 420,726 | – | Menzel et al., 2015 |
| | Yellowstone National Park, USA | Microbial mat | 7.9 | 80–82 | Sanger | – | 1.29 | Colman et al., 2016 |
| | Los Azufres, Mexico | Sediment | 3.6 | 75 | Illumina Gallx | 6,000,792 | 216 | Servín-Garcidueñas et al., 2013 |
| | Lake Bogoria, Kenya | Water, sediment and microbial mat | 8.9–9.5 | 40–80 | Roche 454 | 24,567 | 12.7 | Dadheech et al., 2013 |
| | São Miguel, Azores | Water, sediment and microbial mat | 2.5–8 | 51–92 | Roche 454 | – | – | Sahm et al., 2013 |
| | Champagne pool, New Zealand | Water and sediment | 5.5–6.9 | 45–75 | Illumina MiSeq | 4,623,251 | – | Hug et al., 2014 |
| | Long Valley Caldera, California | Microbial mat | – | 50–80 | Illumina MiSeq | – | – | Stamps et al., 2014 |
| | Odisha, India | Water and sediment | 7.2–7.4 | 40–58 | Roche 454 GS | – | 71.26 | Badhai et al., 2015 |
| | Sungai Klah, Malaysia | Water and sediment | 8.00 | 75–85 | Illumina HiSeq | 5,527,175,000 | – | Chan et al., 2015 |
| | Shi-Huang-Ping, Taiwan | Water | 2.5 | 69 | Illumina HiSeq | 557,415,266 | – | Lin et al., 2015 |
| | Lobios, Ourense, Spain | Water | 8.2 | 76 | Illumina HiSeq | 11,982,436 | – | López-López et al., 2015b |
| | Tuwa, India | Water | 8.2–9 | 54–65 | Ion Torrent PGM | 541,379 | 98.7 | Mangrola et al., 2015a |
| | Lasundra, India | Water | 6.0 | 42–52 | Ion Torrent PGM | 606,867 | 98.6 | Mangrola et al., 2015b |
| | Eryuan, China | Sediment | 7.0 | 65 | Illumina HiSeq | 10,360,000 | – | Menzel et al., 2015 |
| | Uzon Caldera, Russia | Water and sediment | 5.8–6.0 | 61–64 | Roche 454 | 660,054 | – | Menzel et al., 2015 |
| | Pozzuoli, Italy | Water and sediment | 3.0 | 76 | Illumina HiSeq | 10,060,000 | – | Menzel et al., 2015 |
| | Pisciarelly, Italy | Sediment | 5.5 | 86 | Roche 454 | 876,681 | – | Menzel et al., 2015 |
| | Grensdalur, Iceland | Water and sediment | 5.0 | 85–90 | Illumina HiSeq | 10,330,000 | – | Menzel et al., 2015 |
| | Krisuvik, Iceland | Water and sediment | 3.5–4.0 | 90 | Illumina HiSeq | 10,050,000 | – | Menzel et al., 2015 |
| | Parvati River, India | Water and microbial mat | 7.1–7.4 | 93–52 | Illumina Gallx | 78,891,278 | – | Sangwan et al., 2015 |
| | Unkeshwar, India | Water | 7.3 | 50–60 | Illumina HiSeq | 848,096 | 212.87 | Mehetre et al., 2016 |
| Deep sea hydrothermal vent | Mid-Atlantic Ridge | Microbial mat | 9.0–11.0 | 90 | Illumina HiSeq | 46,361 | 35 | Brazelton and Baross, 2009 |
| | Juan de Fuca Ridge | Sulfide chimney | 2.0–3.0 | 316 | Roche 454 | 308,034 | 71 | Xie et al., 2011 |

(Continued)

TABLE 2 | Continued

| Environment | Location | Type of sample/s | pH | Temperature (°C) | Sequencer | Total reads | Size (Mbp) | References |
|----------------|------------------------------------|----------------------|-----------|------------------|----------------------|-------------|------------|-------------------------|
| | Guaymas Baisin, Gulf of California | Black-smoker chimney | 6.0 | 190 | Roche 454 | 512,83 | 196.38 | He et al., 2013 |
| | Juan de Fuca Ridge | Vent fluid | – | 125 | Roche 454 | 808,051 | – | Anderson et al., 2014 |
| Biogas reactor | Germany | Microbial mat | 6.64–8.11 | 55 | Roche 454 | 303,493 | 120.50 | Rademacher et al., 2012 |
| | Linköping, Sweden | Digester material | 7.5–8.1 | 50–55 | Roche 454 | 250,478 | – | Sundberg et al., 2013 |
| | Romania | Sludge | 6.94–7.62 | 55 | Ion Torrent PGM | 300,000 | – | Pap et al., 2015 |
| Oil reservoir | Norwegian sea | Oil, water and gas | – | 85 | Roche 454 | 702,607 | 345 | Kotlar et al., 2011 |
| | Gippsland Basin, Australia | Formation water | 7.2 | 120 | Roche 454 | – | – | Li et al., 2013a |
| Compost | São Paulo Zoo Park, Brazil | Compost material | 7 | 65.8–67.2 | Roche 454 | 3,167,044 | 836 | Martins et al., 2013 |
| Gold mine | South Africa | Fracture water | 9.3 | 60 | Sanger and Roche 454 | 500,008 | – | Chivian et al., 2008 |

In those studies comprising several samples, the total reads and size reflected is just the one of the sample with higher values.

2012). Metagenomic sequences can be binned based on their sequence similarity to a database of taxonomically annotated sequences using tools like MEGAN (Huson et al., 2011) or MGRAST, a public resource for the automatic phylogenetic and functional analysis of metagenomes (Meyer et al., 2008). MEGAN bases its taxonomic classification on the NCBI taxonomy using BLAST. With this tool, Klatt et al. (2013) assessed the community structure of six phototrophic microbial mat communities in YNP and Badhai et al. (2015) revealed the dominance of Bacteria over Archaea in four geothermal springs in Odisha, India. Taxonomic binning can be done with assembled or unassembled reads, although assessing taxonomic abundance with assembled data can lead to a miscalculation of the abundance of some taxa, as contigs are treated as a single sequence in most downstream analysis, hindering the analytical tools to accurately quantify the abundance of the taxon (Sharpton, 2014).

Assembly is described as the process of merging individual metagenomic reads into longer pieces of contiguous sequences (contigs) based on overlapping sequences and paired read information (Dröge and McHardy, 2012). Bioinformatic implements like MetaVelvet (Namiki et al., 2012) or IDBA-UD (Peng et al., 2012) have been used in the assembly of whole shotgun metagenome reads to study the taxonomical composition of different high-temperature environments. For example, MetaVelvet was applied in the study of eight globally distributed hot springs by Menzel et al. (2015) and IDBA-UD in the analysis of the community composition of Sungai Klah hot spring in Malaysia (Chan et al., 2015). This step can simplify bioinformatic analysis, but it may also produce chimeras, therefore researchers often bin reads and assemble each bin independently to

decrease the probability of generating chimeras (Sharpton, 2014).

In recent studies, the integration of assembly and taxonomic binning by sequence composition allowed the reconstruction of several partial genomes from high-temperature environments such as the genome of a novel archaeal Rudivirus obtained from a Mexican hot spring, (Servín-Garcidueñas et al., 2013) or the draft genome sequence of *Thermoanaerobacter* sp. strain A7A, reconstructed from the metagenome of a 102°C hydrocarbon reservoir in the Bass Strait, Australia (Li et al., 2013b). Using a similar approach, Sangwan et al. (2015), reconstructed the genome of the bacterial predator *Bdellovibrio* ArHS, with the metagenomic assembly of the microbial mats of an arsenic rich hot spring in the Parvati river valley (Manikaran, India). Also, Sharma et al. (2016) combining genomic and metagenomic data, used two *Cellulosimicrobium cellulans* genomes derived from metagenomics, to study the evolution of pathogenicity across the species of *C. cellulans*.

FUNCTIONAL ANALYSIS OF THERMOPHILES

Sequence-Based Function Prediction

The metagenomic reads obtained from shotgun sequencing of an environmental DNA can be annotated with functions to determine the functional diversity of the microbial community. This usually comprises two steps: identifying metagenomic reads that contain protein coding sequences (gene prediction), and comparing the coding sequences to a database of genes, proteins, protein families, or metabolic pathways (gene annotation) (Sharpton, 2014). Some frequently used databases for functional

annotation are the SEED annotation system (Overbeek et al., 2014), the KEGG orthology (KO) database (Kanehisa et al., 2016) or the Pfam database, based on hidden Markov models (HMM) to classify in accordance with the protein domains (Finn et al., 2015). There are several robust web resources that can be easily used to perform gene prediction, database search, family classification, and annotation, including MG-RAST (Meyer et al., 2008), IMG/M (Markowitz et al., 2014), or SUPER-FOCUS (Silva et al., 2015). Considerable functional profiles of thermophilic populations have been based on these tools such as the study of the microbiota of Tuwa hot spring in India (Mangrola et al., 2015a) in which the functional annotation was performed using the MG-RAST pipeline. In this study, a high number of annotated features were classified as unknown function, suggesting the potential source of novel microbial species and their products. Similar results were found in the metagenome of Unkeshwar, another hot spring in India, where pathway annotation was done using KEGG (Mehetre et al., 2016). For each contig sequence, the assignment of KO numbers obtained from known reference hits was done, revealing up to 20% unclassified sequences. These results reflect a promising world of undiscovered proteins that could be explored to find new catalysts for biotechnological applications.

In this approach, it is important to consider that, despite the information given by functional annotation of the metagenomic sequences; the presence of a gene on a metagenome does not mean that it is expressed. Therefore, functional metagenomics, metatranscriptomics, and metaproteomics assays are necessary to assess the real community functional activity. To increase the probability of finding active functional genes involved in a substrate uptake and transformation, some studies use a substrate-induced enrichment of the community before the metagenomic DNA extraction. After, these genes can be detected either by sequence (Graham et al., 2011; Wang et al., 2016) or by functional metagenomics (Chow et al., 2012). Using this procedure, Graham et al. (2011) found and characterized an hyperthermophilic cellulase in an archaeal community, obtained by growth at 90°C of the sediment of a geothermal source enriched with crystalline cellulose.

Another important limitation of shotgun metagenomics is that the databases may be subjected to phylogenetic biases, as some communities are more accurately or more exhaustively annotated than others (Chistoserdova, 2010).

Functional Metagenomics

Function-based metagenomics relies on the construction of metagenomic libraries by cloning environmental DNA into expression vectors and propagating them in the appropriate hosts, followed by activity-based screening. After an active clone is identified, the sequence of the clone is determined, the gene of interest is amplified and cloned with the subsequent expression and characterization of the product to explore its biotechnological potential (Figure 2). This technique has the advantage of not requiring the cultivation of the native microorganisms or previous sequence information of known

genes, thus representing a valuable approach for mining enzymes with new features.

The use of functional metagenomics allows the discovery of novel enzymes whose functions would not be predicted based on DNA sequence. This approach complements sequence-based metagenomics as the information from function-based analyses can be used to annotate genomes and metagenomes derived exclusively from sequence-based analyses (Lam et al., 2015). Therefore, several investigations in thermal environments combine sequencing methods (taxonomical and functional characterization) with functional screening of clones (Chen et al., 2007; Wemheuer et al., 2013; Leis et al., 2015; López-López et al., 2015b).

Depending on the size of the insert, functional metagenomics can be explored using fosmids (35–45 kb insert), BACs (~200 kb insert), cosmids (30–42 kb insert), or plasmids (<10 kb insert). Bigger inserts are more likely to contain complete genes and operons, allowing the expression of more enzymes. A great number of high temperature functional metagenomics studies use the commercial vector pCC1FOS (Table 3), which allows inserts up to 40 kb, and it is available in a toolkit to simplify the library construction. More information about this vector is compiled in Lam et al. (2015) review.

There are several technically challenging steps in library construction. Mainly, the high quality and length of the metagenomic DNA required for proceeding to the ligation into the vector and the need of obtaining a high proportion of clones in order to cover all the variability of the microbial community. This limitation is particularly important in soil studies, where it has been reported that contaminants like humic acids are present in metagenomic DNA extracts, interfering with the subsequent enzymatic reactions. Therefore, the widely extended method of soil DNA extraction established by Zhou et al. (1996), is usually accomplished with further purification of the sample that can lead to a loss of DNA yield. Some studies show that the humic acids can be easily removed by gel electrophoresis of the metagenomic DNA followed by gel extraction, as humic acids migrate faster than the large metagenomic DNA (Kwon et al., 2010). This simple approach was used to construct a Turpan Basin soil metagenomic library for a functional screening of thermostable beta-galactosidases (Wang et al., 2014). Alternatively, to avoid contaminating the circulating buffer, electrophoresis can be paused after humic acids have formed a front, excising the part of the gel containing the humic acids, and replacing it with fresh gel (Cheng et al., 2014).

Another important drawback that compromises the functional metagenomics approach is the selection of the expression host. Although the commonly used *E. coli* strains have relaxed requirements for promoter recognition and translation initiation, some genes from environmental samples may not be efficiently expressed due to differences in codon usage, transcription and/or translation initiation signals, protein-folding elements, post-translational modifications, or toxicity of the active enzyme (Uchiyama and Miyazaki, 2009). This problem could be even worse when the proteins expressed need special conditions to be active, such as high temperatures, considering that mesophiles, like *E. coli*, do not

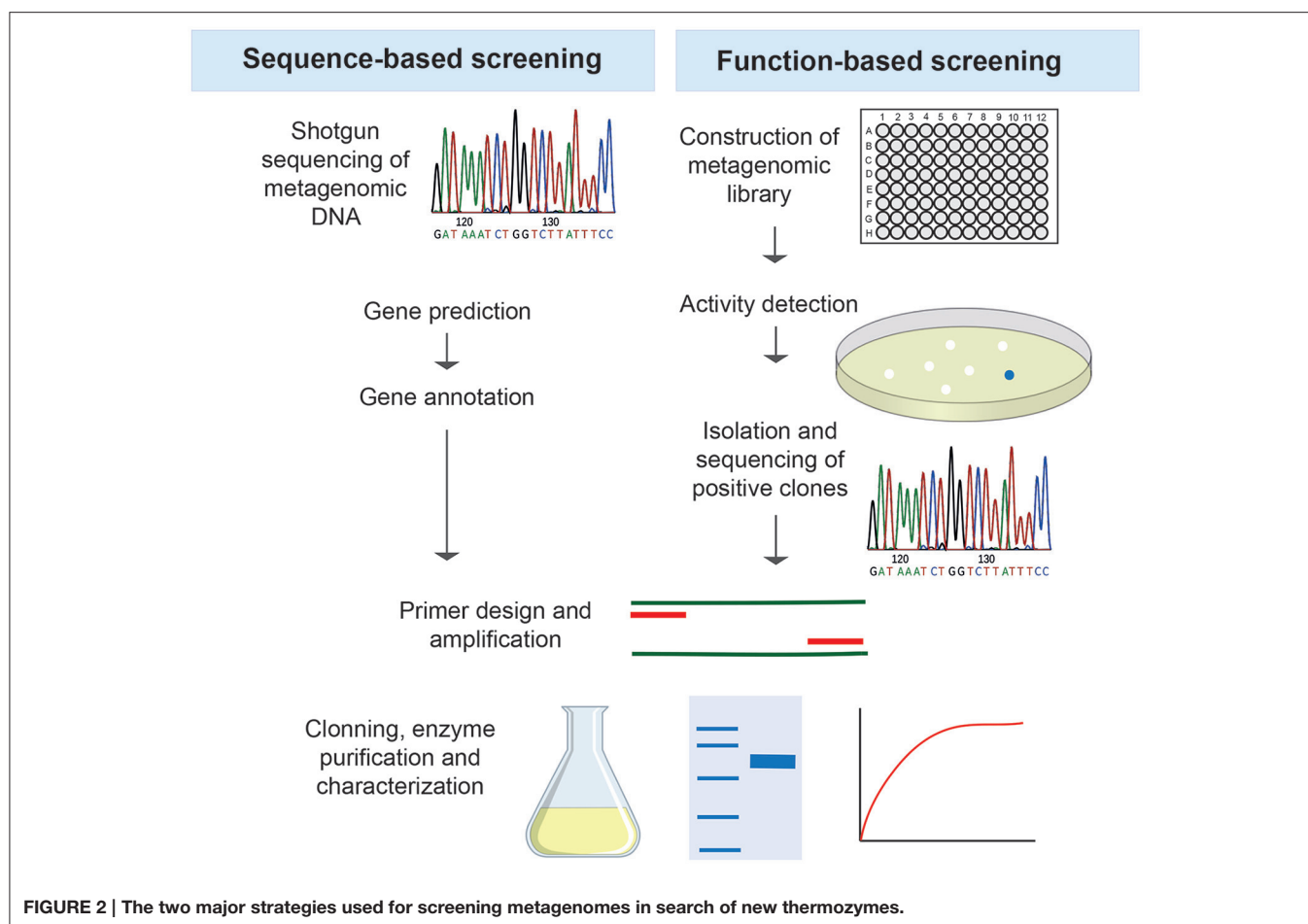


FIGURE 2 | The two major strategies used for screening metagenomes in search of new thermostases.

survive at these high temperature conditions. Accordingly, an alternative expression host may be required to overcome the heterologous expression of some genes derived from hot environments and thus, identify a broader range of enzymes. The thermophilic bacterium *T. thermophilus* has been proposed as a good candidate for function-based detection of thermostases. In a recent functional screening to detect esterases, Leis et al. (2015) constructed two large insert fosmid metagenomic libraries of compost and hot spring water using pCT3FK, a pCC1FOS derived *T. thermophilus*/*E. coli* shuttle fosmid (Angelov et al., 2009), in *T. thermophilus* and compared them to the same libraries expressed in *E. coli*. Only two esterases were found at 60°C in the libraries generated in *E. coli* while 5 different esterases were discovered in the same libraries expressed in *T. thermophilus*. Therefore, this could be a suitable system to improve the detection of metagenome-derived thermostases. The main restriction of this approach is that pCT3FK integrates into *T. thermophilus* chromosomal DNA. In fact, the genomes of the positive clones isolated by Leis et al. (2015) were completely sequenced before proceeding with the PCR amplification and cloning of the candidate genes, with the consequent cost of time and money. Other versatile broad-host-range cosmids that have been used in a soil study (pJC8 and pJC24) allow the phenotypic screening of the library in bacteria such as *Bacillus* and in the

yeast *Saccharomyces* (Cheng et al., 2014). The selection of the appropriate substrate for the functional screening is also a crucial step in this approach, as the substrate may cause biases in the selection of the activities of interest. Recent studies suggest that the initial selection of active clones with general substrates should be followed by a more specific one to improve the effectiveness of the detection (Ferrer et al., 2016). Other biases and limitations of functional metagenomics and strategies for its improvement have been previously reviewed by Ferrer et al. (2005) and Ekkers et al. (2012).

Some hot environments where function-based screening of microbial communities have been done include hot springs (López-López et al., 2015b), deserts (Neveu et al., 2011), petroleum reservoirs (de Vasconcellos et al., 2010), or human-made environments like a biogas plant (Ilmberger et al., 2012), demonstrating the potential of functional metagenomics as a very important source of new thermostases.

METAGENOME-DERIVED THERMOSTASES

Many industrial processes require elevated temperatures to take place. Thus, microorganisms surviving at temperatures above 55°C represent an important source of biotechnological richness for high temperature bioprocesses by producing a large variety

TABLE 3 | Examples of thermozymes obtained by functional metagenomics.

| Family | Activity | Metagenomic DNA source | Vector | Host | Substrate | Total No. clones | Positive clones | References |
|-------------------|------------------------|---|---------------------|-----------------------------|---|------------------|-----------------|---------------------------|
| Lypolitic enzymes | Lipase | Biomass sequencing fed-batch reactor | pAFS2 | <i>E. coli</i> LE392MP | Tributyrin | 10,000 | 10 | Meilleur et al., 2009 |
| | Lipase | Oil field soil | pZErO-2 | <i>E. coli</i> TOP10 | Tributyrin, rhodamine B | 83,000 | 1 | Fan et al., 2011 |
| | Lipase | Botanical Garden soil, Hamburg, Germany | pSuperCos | <i>E. coli</i> EPI100 | Tributyrin | 6500 | 2 | Chow et al., 2012 |
| | Lipase | Industrial wastewater treatment plant | pBeloBAC11 | <i>E. coli</i> DH10B | Tributyrin, rhodamine B | 40,000 | 12 | Sharma et al., 2012 |
| | Lipase | Activated sludge | pUC18 | <i>E. coli</i> DH5 α | Tricaprylin | 130,000 | 1 | Roh and Schmid, 2013 |
| | Lipase/Esterase | Hot spring, Kamchatka Peninsula | pCR-XL-TOPO | <i>E. coli</i> DH5 α | Tributyrin | – | 3 | Wemheuer et al., 2013 |
| | Lipase/Esterase | Oil polluted mud flat | pUC19 | <i>E. coli</i> DH5 α | Glyceryl trioctanoate | 3000 | 2 | Kim et al., 2015 |
| | Esterase | Poly(DL-lactic acid) disks in compost | pUC18 | <i>E. coli</i> DH10B | PLA solution | 40,000 | 7 | Mayumi et al., 2008 |
| | Esterase | Compost | pCC1FOS | <i>E. coli</i> EPI300T1R | Tributyrin, gum arabic | 13,000 | 10 | Kang et al., 2011 |
| | Esterase | Turpan Basin soil, China | pUC118 | <i>E. coli</i> TOP10 | Caprylate | 21,000 | 3 | Fan et al., 2012a |
| | Esterase | Turpan Basin soil, China | pUC118 | <i>E. coli</i> DH5 α | Caprylate | 26,000 | 1 | Fan et al., 2012b |
| | Esterase | Activated sludge | pBluescript SK+ | <i>E. coli</i> DH5 α | Tributyrin | 40,000 | 1 | Shao et al., 2013 |
| | Esterase | Turpan Basin soil, China | pUC118 | <i>E. coli</i> DH5 α | Tributyrin | 200,000 | 19 | Wang et al., 2013 |
| | Esterase | Deep-sea hydrothermal field, East Pacific | pUC18 | <i>E. coli</i> XL1-blue | Tributyrin | – | – | Zhu et al., 2013 |
| | Esterase | Guayas Basin smoker chimmeney | pCC2FOS | <i>E. coli</i> EPI300T1R | Tributyrin | 18,000 | 7 | Fu et al., 2015 |
| | Esterase | Oil polluted mud flat | pUC19 | <i>E. coli</i> DH5 α | Tributyrin | 3000 | 7 | Kim et al., 2015 |
| | Esterase | Hot spring, Furnas, Azores | pCT3FK | <i>T. thermophilus</i> BL03 | Tributyrin | 7968 | 6 | Leis et al., 2015 |
| | Esterase | Hot spring water Lobios, Galicia, Spain | pCC1FOS | <i>E. coli</i> EPI300T1R | Tributyrin | 11,600 | 6 | López-López et al., 2015b |
| Glycosidase | Cellulase | Compost | pCC2FOS | <i>E. coli</i> EPI300T1R | CMC Trypan blue | 12,380 | 2 | Kwon et al., 2010 |
| | Cellulase | Biogas plant and elephant feces | pCC1FOS | <i>E. coli</i> EPI300T1R | CMC Congo red | 29,000 | 14 | Ilmberger et al., 2012 |
| | Cellulase | Activated sludge | pUC18 | <i>E. coli</i> DH5 α | CMC | 32,000 | 4 | Sharma et al., 2012 |
| | β -glucosidase | Hot spring water, mud and sediment | pCR-XL-TOPO | <i>E. coli</i> TOP10 | Esculin hydrate, ferricammonium citrate | – | 1 | Schröder et al., 2014 |
| | β -glucosidase | Thermophilic methanogenic digester | pCC1FOS | <i>E. coli</i> EPI300T1R | Esculin hydrate, ferricammonium citrate | – | – | Wang et al., 2015 |
| | β -glucosidase | Termite hindgut | pUC118 | <i>E. coli</i> DH5 α | Esculin hydrate, ferricammonium citrate | 800,000 | 13 | Gao et al., 2016 |
| | β -galactosidase | Hot spring, Hymalaya, India | pSMART LCamp | <i>E. coli</i> DH10B | X-gal | 10,000 | 1 | Gupta et al., 2012 |
| | β -galactosidase | Turpan Basin soil, China | pUC19 Δ lacZ | <i>E. coli</i> DH5 α | X-gal | 8000 | 1 | Zhang et al., 2013 |

(Continued)

TABLE 3 | Continued

| Family | Activity | Metagenomic DNA source | Vector | Host | Substrate | Total No. clones | Positive clones | References |
|----------------|------------------------|------------------------------------|---------------------|-----------------------------|--|------------------|-----------------|------------------------------|
| | β -galactosidase | Turpan Basin soil, China | pUC19 Δ lacZ | <i>E. coli</i> DH5 α | X-gal | 700,000 | 1 | Wang et al., 2014 |
| | Xylanase | Compost | pCC2FOS | <i>E. coli</i> EPI300T1R | Xylan | 12,380 | 5 | Kwon et al., 2010 |
| | Endoxylanase | Compost-soil | p18 GFP | <i>E. coli</i> DH10B | RBB-xylan | 180,000 | 1 | Verma et al., 2013 |
| | β -xylosidase | Thermophilic methanogenic digester | pCC1FOS | <i>E. coli</i> EPI300T1R | Esculin hydrate, ferric ammonium citrate | – | – | Wang et al., 2015 |
| | Various | Raw and torrefied wheat straw | pCC1FOS | <i>E. coli</i> EPI300T1R | X-fuc, X-gal, X-Xyl, X-Man, X-cel, X-glu | 44,000 | 71 | Maruthamuthu et al., 2016 |
| Amylase | α -amylase | Western Ghats soil | pCC1FOS | <i>E. coli</i> EPI300T1R | Potassium iodide solution | 76,000 | 1 | Vidya et al., 2011 |
| | α -amylase | Biogas reactor | pBK-CMV | <i>E. coli</i> XL0LR | AZCL-amylose | 2000 | 1 | Jabbour et al., 2013 |
| Phosphatase | Phytase | Soil | pCC1FOS | <i>E. coli</i> EPI300T1R | Phytate | 14,000 | 28 | Tan et al., 2014 |
| Protease | Serine protease | Sand Gobi desert | pCC1FOS | <i>E. coli</i> EPI300T1R | Skim milk | 17,000 | 16 | Neveu et al., 2011 |
| | Serine protease | Sand Death Valley desert | pBSKII+ | <i>E. coli</i> DH10B | Skim milk | 30,000 | 1 | Neveu et al., 2011 |
| | Serine protease | Soil | pHT01 | <i>E. coli</i> DH10B | AZCL-casein | – | 1 | Biver et al., 2013 |
| | Patatin-like protein | Hot spring, Kamchatka | pCR-XL-TOPO | <i>E. coli</i> DH5 α | Skim milk | – | 1 | Wemheuer et al., 2013 |
| | Serine protease | Hot spring, Chumathang Ladakh | pUC18 | <i>E. coli</i> DH5 α | Casein | 9000 | 1 | Singh et al., 2015 |
| Oxidoreductase | Extradiol dioxygenase | Activated sludge | pCC1FOS | <i>E. coli</i> EPI300T1R | Catechol | 96,000 | 91 | Suenaga et al., 2007 |
| | | Oil reservoir | pCC2FOS | <i>E. coli</i> EPI300T1R | Hexadecane | 5000 | 72 | de Vasconcellos et al., 2010 |
| | Dioxygenase | Activated sludge | pBel0BAC11 | <i>E. coli</i> DH10B | Phenol, catechol | 40,000 | 4 | Sharma et al., 2012 |

of biocatalysts. Biotechnological processes carried out at high temperatures provide numerous benefits such as higher solubility of reagents, and reduced risk of microbial contamination (Mirete et al., 2016). From an industrial point of view, thermozymes possess certain advantages over their mesophilic counterparts as they are active and efficient under high temperatures, extreme pH values, high substrate concentrations, and high pressure (Sarmiento et al., 2015). Some of them are also highly resistant to denaturing agents and organic solvents (Fan et al., 2011; Roh and Schmid, 2013). In addition, thermozymes are easier to separate from heat-labile proteins during purification steps as reported by Pessela et al. (2004). As a result, high temperature-active enzymes can be potentially used in diverse industrial and biotechnological applications including food, paper and textile processing, chemical synthesis and the production of pharmaceuticals.

Some thermostable enzymes are still recovered by isolation from thermophilic microorganisms (Shi et al., 2013; Fuciños et al., 2014; Sen et al., 2016), however metagenomics has opened a new important field in the discovery of novel biocatalysts and has been revealed as a promising mining strategy of resources for the biotechnological and pharmaceutical industry. There are two different ways of screening a metagenome in search of thermozymes: a sequence-based approach and a function-based approach (Figure 2).

Sequence-based screening methods rely on the prior knowledge of conserved sequences of domains/proteins/families of interest. It involves primer designing followed by amplification and cloning of the metagenomic genes. The main drawback of this approach is its failure to detect fundamentally different novel genes, as it cannot discover non-homologous enzymes. Some potential biocatalysts have been isolated mining metagenomic

TABLE 4 | Examples of thermozymes isolated by sequence-based screening of metagenomes.

| Enzyme | Source of metagenomic DNA | References |
|--|--|------------------------|
| Cellulase | Long-term dry thermophilic methanogenic digester | Wang et al., 2015 |
| Xylanase | Cow dung compost | Sun et al., 2015 |
| Xylanase | Long-term dry thermophilic methanogenic digester | Wang et al., 2015 |
| Endoxylanase | Compost | Dougherty et al., 2012 |
| β -galactosidase | Hot spring water, Yongtai | Liu et al., 2015 |
| β -xylosidase/ α -arabinofuranosidase | Compost | Dougherty et al., 2012 |
| β -xylosidase | Long-term dry thermophilic methanogenic digester | Wang et al., 2015 |
| Pectinase | Hot spring soil, Manikaran, India | Singh et al., 2012 |
| α -fucosidase | Compost | Dougherty et al., 2012 |
| Phytase (phosphatase) | Insect-cultivated fungus gardens | Tan et al., 2016 |
| Limonene 1,2-epoxide hydrolase | Hot springs water | Ferrandi et al., 2015 |
| Nitrilase | Atlantis II Deep Brine Pool, Red Sea | Sonbol et al., 2016 |
| α -amylase | Hydrothermal vent, Juan de la Fuca | Wang et al., 2011 |
| Dehydrochlorinase | Soil contaminated with HCH-isomers | Macwan et al., 2011 |
| Polymerase | Hot spring 3173 Pol | Moser et al., 2012 |

sequences in prospecting for genes coding thermozymes (Table 4). Namely, a gene encoding a thermostable pectinase was isolated from a soil metagenome sample collected from hot springs of Manikaran (India), using a PCR-based cloning strategy with primers designed based on known sequences of pectinase genes from other species (Singh et al., 2012). The recombinant protein is proposed to be of great use in industrial processes due to its activity over a broad pH range. Thanks to this search based on sequence homology to related gene families, 22 putative ORFs (open reading frames) were identified from a switchgrass-adapted compost community finding a bi-functional β -xylosidase/ α -arabinofuranosidase that maintained ~75% of its activity after 16 h at 60°C (Dougherty et al., 2012). The same sequence-based approach was used by Ferrandi et al. (2015) who discovered, cloned and characterized two novel limonene-1,2-epoxide hydrolases (LEHs) with an *in-silico* screening of the LEHs sequences in the assembled contigs from hot spring metagenomes.

The function-based metagenomic screening is the most important way to discover novel thermozymes as it doesn't rely on the sequence. The main advantage of directly screening for enzymatic activities from metagenome libraries is that it gives access to previously unknown genes and their encoded enzymes. Thus, some completely new thermozymes that couldn't be found by sequence screening have been discovered, like the unusual glycosyltransferase-like enzyme with β -galactosidase activity recovered by Wang et al. (2013) from a Turpan Basin soil metagenomic library. Function-based metagenomic screening has allowed the discovery of a wide range of thermozymes (Table 3). In this review, we focus on the recovery of the functional-derived thermostable metagenomic enzymes that are mostly used in biocatalysis and industrial sectors, such as lipolytic enzymes, glycosidases, proteases, and oxidoreductases (Böhnke and Perner, 2015).

Lipolytic Enzymes

Lipolytic enzymes, comprising esterases (EC 3.1.1.1) and lipases (EC 3.1.1.3), are extensively distributed in microorganisms, plants, and animals. They catalyze the hydrolysis, synthesis, or transesterification of ester bonds. At present, these enzymes represent about 20% of commercialized enzymes for industrial use (López-López et al., 2015a), as they have great potential in several industrial processes such as production of biodegradable polymers, detergents, food flavoring, oil biodegradation, or waste treatment, among others (Anobom et al., 2014). Therefore, a considerable number of functional metagenomics studies are focused on mining thermal environments in search for these enzymes (Table 3).

Lipases are generally defined as carboxylesterases hydrolyzing water-insoluble (acyl chain length >10) triglycerides, with trioleoylglycerol as the standard substrate. In contrast, esterases catalyze the hydrolysis of short-chain esters (acyl chain length <10) with tributylglycerols (tributyrin) as the standard substrate, although lipases are also capable of hydrolyzing esterase substrates (Rhee et al., 2005). At least 200 different substrates have been successfully applied in assays for functional selection of esterase/lipase biocatalysts in metagenomic clone libraries (Ferrer et al., 2016), including the widely used tributyrin (Rhee et al., 2005; Meilleur et al., 2009; López-López et al., 2015b), and p-nitrophenyl (NP) acetate (Wang et al., 2013). Meilleur et al. (2009) isolated a new alkali-thermostable lipase with an optimal activity at 60°C and pH 10.5 by functional screening of a metagenomic cosmid library from the biomass produced in a gelatin enriched fed-batch reactor. Another gene coding for a thermostable esterase was detected by functional screening of fosmid environmental DNA libraries constructed with metagenomes from thermal environmental samples of Indonesia (Rhee et al., 2005). The recombinant esterase was active from 30 up to 95°C with an optimal pH of approximately 6.0. Mayumi et al. (2008), generated a metagenomic library with the community DNA extracted from biodegradable polyester

poly(lactic acid) (PLA) disks buried in compost and found a PLA depolymerase that had an esterase domain. Purified enzyme showed the highest activity at 70°C and degraded not only PLA, but also various aliphatic polyesters, tributyrin, and p-NP esters. As mentioned before, those enzymes able of retaining activity even in the presence of organic solvents are considered very interesting for industrial applications. A new thermophilic organic solvent-tolerant and halotolerant esterase with an optimum pH and temperature of 7.0 and 50°C, respectively, was found in the functional screening of a soil metagenomic library with 48,000 clones (Wang et al., 2013).

Apart from these above cited sources, metagenomic esterases, and lipases have been isolated by functional screening of other hot environments like deep-sea hydrothermal vents (Zhu et al., 2013) and hot springs (López-López et al., 2015b) as shown in **Table 3**. A more extensive review of metagenome derived extremophilic lipolytic enzymes can be found in López-López et al. (2014).

Glycosidases

The enzymes that hydrolyze glycosidic bonds between two or more sugars or a sugar and a nonsugar moiety within carbohydrates or oligosaccharides are known as glycosyl hydrolases (GHs) or glycosidases (Sathya and Khan, 2014). There are 115 GH families, collected in the Carbohydrate Active enZYme database (CAZy; <http://www.cazy.org>) (Lombard et al., 2014), including a broad number of enzymes like cellulases, β -galactosidases, amylases, and pectinases.

Cellulases encompass a group of complex enzymes conformed by endo- β -1,4 glucanases, cellobiohydrolases, celloextrinases, and β -glucosidases. These enzymes work together to degrade cellulose into simple sugars and their thermostable representatives could be used in biofuel production from lignocellulosic biomass (Bhalla et al., 2013). Several substrates can be employed in plate-based screens for the functional detection of clones harboring cellulase activity, such as carboxymethyl-cellulose in combination with trypan blue, Gram's iodine, or Congo Red. Meddeb-Mouelhi et al. (2014), found that Gram's iodine may lead to the identification of false positives, making Congo Red a more suitable dye for this approach. Using Congo Red dye as a colorimetric substrate, Ilmberger et al. (2012) obtained two fosmid clones derived from a carboxymethyl-cellulose (CMC)-enriched library from a biogas plant. These two fosmids were designated as pFosCelA2 and pFosCelA3, encoding two thermostable cellulases with significant activities in the presence of 30% (v/v) ionic liquids (ILs). This is an interesting property for the cellulose degradation, as cellulose could increase its solubility in the ILs.

From the group of cellulases, β -glucosidases have attracted considerable attention in recent years due to their important roles in various biotechnological processes such as hydrolysis of isoflavone glucosides or the production of fuel ethanol from agricultural residues (Singhania et al., 2013). Other uses of β -glucosidases include the cleavage of phenolic and phytoestrogen glucosides from fruits and vegetables for medical applications or to enhance the quality of beverages. An archaeal β -glucosidase (Bgl1) showing activity toward cellobiose,

cellotriose, and lactose was isolated from a metagenome from a hydrothermal spring in the island of Sjo Miguel (Azores, Portugal) (Schröder et al., 2014).

β -Galactosidases (EC 3.2.1.23), which hydrolyze lactose to glucose and galactose, have two main applications in the food industry: the production of low-lactose milk and dairy products for lactose intolerant people and the generation of galactooligosaccharides from lactose by the transgalactosylation reaction. These enzymes can be also used in the revalorisation of cheese whey (Becerra et al., 2015), a by-product of the dairy industry with a high organic load that can be considered a pollutant.

The most widely used substrate for the β -galactosidase screening, 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal), is the substrate providing, in some cases, the lowest number of positive hits in relation to the total number of clones screened (Ferrer et al., 2016). Usually, the positive clones capable of hydrolyzing the X-gal are further tested against ortho-NP- β -galactoside (ONPG) and lactose (Wierzbicka-Woś et al., 2013). Mayor drawbacks for the use of β -galactosidase in industrial processes is the inhibition by the reaction products, leading to a decrease in the reaction rates or even to stop the enzymatic reaction completely. The thermostable β -galactosidase (Gal308) discovered by Zhang et al. (2013) exhibited high tolerance to galactose and glucose with the highest activity at 78°C, an optimum pH of 6.8 and high enzyme activity with lactose as substrate. The authors suggest that these properties would make it a good candidate for the production of low-lactose milk and dairy products. Another novel and thermostable alkalophilic β -D-galactosidase with an optimum temperature at 65°C and with high transglycosylation activity was identified through functional screening of a metagenomic library from a hot spring in northern Himalayan region of India (Gupta et al., 2012).

Xylans, made of β -1,4 linked xylopyranoses as a linear backbone with branches, constitute the second most significant group of polysaccharides in plant cell walls and are degraded by xylanases (Sathya and Khan, 2014). These hemicellulolytic enzymes are mostly used as biobleaching agents in the paper and pulp industry. The discovery of thermostable and alkali-stable xylanases has become an important goal in this field since this process requires high temperatures and alkali media, but this is not the only application of thermostable xylanases (Kumar et al., 2016). Functional metagenomics of hot environments represents an interesting source of xylanases. As an example, a novel alkali-stable and thermostable GH-11 endoxylanase encoding gene (*Mxyl*), was isolated by functional screening of a compost-soil metagenome (Verma et al., 2013). The thermostability of this enzyme was subsequently engineered by directed site mutagenesis (Verma and Satyanarayana, 2013).

Amylases are known as enzymes that catalyze the hydrolysis of starch into sugars (Sundarram et al., 2014). A novel and thermostable amylase with the highest activity at 90°C was retrieved from a black smoker chimney by combining fosmid library construction with pyrosequencing (Wang et al., 2011). Another α -amylase was isolated in the functional screening of a metagenomic library of Western Ghats soil constructed in pCC1FOS. This amylase retained 30% activity after incubation

for 60 min at 80°C and had an optimal pH of 5.0 and could be potentially used in some industrial processes like liquefaction and saccharification of starch in food industry, or formulation of enzymatic detergents and removing starch from textiles (Vidya et al., 2011).

Proteases

Proteases are protein-hydrolyzing enzymes classified into acidic, neutral, or alkaline groups, based on their optimum pH. They can also be classified into aspartic, cysteine, glutamic, metallo, serine, and threonine protease types based on the amino acids present in their active sites (Singh et al., 2015). These enzymes are widely used in various industries such as detergent, food, and leather (Haddar et al., 2009; George et al., 2014). A thermotolerant, alkali-stable and oxidation resistant protease (CHpro1) was found by functional screening of a metagenomic library constructed from sediments of hot springs in Chumathang area of Ladakh, India (Singh et al., 2015). This enzyme, that showed optimum activity at pH 11 and stability in high alkaline range, could be especially interesting for the detergent industry, as the pH of laundry detergents is generally in the range of 9.0–12.0. This property, in addition to the resistance in the presence of detergent compounds, like oxidizing agents, and the possibility of working at high wash temperatures (optimum activity at 80°C), makes it a very suitable detergent protease.

Oxidoreductases

These enzymes catalyze oxidation-reduction reactions, in which hydrogen or oxygen atoms or electrons are transferred between molecules and are important biocatalysts for several industrial processes. From a pharmaceutical point of view, oxidoreductases can act like quorum-quenching enzymes, degrading signal molecules to block quorum-sensing-dependent infection, as reported by Bijtenhoorn et al. (2011), who found a soil-derived dehydrogenase/reductase implicated in the decreasing of *Pseudomonas aeruginosa* biofilm formation and virulence of *Caenorhabditis elegans*. These enzymes can also be used in food industry as they catalyze oxidation-reduction reactions that can play an important role in taste, flavor and nutritional value of aliments such as virgin olive oil (Peres et al., 2015). Another relevant application of oxidoreductases is their role in decomposing specific recalcitrant contaminants by precipitation or by transforming them to other products, leading to a better final treatment of the waste. Some oxidoreductases that can be used for this purpose include peroxidases, polyphenol oxidases, and estradiol dioxygenases (EDOs) (Durán and Esposito, 2000). Suenaga et al. (2007) constructed a metagenomic library from activated sludge used to treat coke plant wastewater containing various organic pollutants like phenol, mono- and polycyclic nitrogen-containing aromatics or aromatic hydrocarbons, among others. The library was screened for EDOs, using catechol as a substrate, yielding 91 EDO-positive clones, 38 of them were sequenced in order to conduct similarity searches using BLASTX. A polyphenol oxidase enzyme, with alkaline laccase activity and highly soluble expression, showing

the optimum activity of 55°C, was isolated from a functional screening of DNA from mangrove soil (Ye et al., 2010).

COMPARATIVE METAGENOMICS

The increasing number of metagenomes from high-temperature environments sequenced and the possibility of generating more sequences with a lower cost of time and money has enabled the comparison of metagenomic sequences between and within environments, opening a new field in metagenomics. Comparative metagenomics can enlighten how the microbial community taxa or the metabolic potential vary between sampling locations or time points, as well as explain the influence of several factors, such as high temperatures, in the taxonomical and functional composition of an ecosystem. Comparison of metagenomic data recovered from different high temperature habitats indicates that these communities are different with respect to species abundance and microbial composition. However, some groups of species are more commonly represented, for example, bacterial taxa such as *Thermotoga*, *Deinococcus-Thermus*, and *Proteobacteria*, as well as Archaea, like *Methanococcus*, *Thermoprotei*, and *Thermococcus* (Lewin et al., 2013). The comparison between metagenomes derived from six distantly located hot springs of varying temperature and pH revealed a wide distribution of four archaeal viral families, *Ampullaviridae*, *Bicaudaviridae*, *Lipothrixviridae*, and *Rudiviridae* (Gudbergsdóttir et al., 2016). Even though the important role of viruses in high temperature ecosystems has been demonstrated, the comparative studies are limited since the diversity of thermophilic viruses in many hot environments remains unknown, as revealed by Adriaenssens et al. (2015) in the Namib desert hypoliths metagenome, where the majority of the viral sequence reads were classified as unknown.

Comparative metagenomics can also increase our insight into the adaptation of microorganisms to high temperature environments. Xie et al. (2011) compared the sequences obtained from a fosmid metagenomic library of a black smoker chimney 4143-1 in the Mothra hydrothermal vent field at the Juan de Fuca Ridge with metagenomes of different environments, including a biofilm of a carbonate chimney from the Lost City hydrothermal vent field (90°C, pH 9–11 fluids). This study revealed that the deep-sea vent chimneys are highly enriched in genes for mismatch repair and homologous recombination, and exhibited a high proportion of transposases. These enzymes, which are critical in horizontal gene transfer, were also abundantly found when comparing the metagenomic data obtained from three different deep-sea hydrothermal vent chimneys (He et al., 2013). This fact supports the previous hypothesis that horizontal gene transfer may be common in the deep-sea vent chimney biosphere and could be an important source of phenotypic diversity (Brazelton and Baross, 2009).

Other comparative studies show that, apart from temperature, pH is also an important factor in the composition of microbial communities. A comparison of the biodiversity and community composition in eight geographically remote hot springs (temperature range between 61 and 92°C and pH

between 1.8 and 7) showed a decrease in biodiversity with increasing temperature and decreasing pH (Menzel et al., 2015). The loss of biodiversity in hot environments with low pH was also observed by Song et al. (2013), showing a more diverse bacterial population in non-acidic hot springs than in acidic hot springs from the Yunan Province (China).

IMG/M (Markowitz et al., 2014) and MG-RAST (Meyer et al., 2008) are two frequently used metagenomics pipelines to easily perform comparative analysis of microbial communities, and can be explored to find sequences of different high temperature environments, since a considerable number of metagenomes are deposited in their databases. MG-RAST has about 240 thousand data sets containing over 800 billion sequences and more than 36 thousand public metagenomes, including 225 metagenomes (0.61%) from different thermophilic biomes with temperatures ranging from 52 to 122°C obtained by whole shotgun and/or amplicon sequencing (data publicly available at MG-RAST server on August 2016).

Usually, these studies require statistical tools to explore multivariate data, like principal component analysis (PCA), in order to compare and contrast metagenomes from different environments. PCA is one of the most widely used statistical analyses for genomic data as it is a simple and robust data reduction technique that can be applied to large data sets. A more exhaustive description of some of the statistical analysis that can be used to compare metagenomes can be found in the study by Dinsdale et al. (2013). In this study, the metabolic functions of 212 metagenomes, including six different hot springs, were compared between and within environments using different statistical methods. Several tools like STAMP (statistical analysis of metagenomics profiles, Parks et al., 2014) and PRIMER-E can be used for this purpose, allowing the statistical analysis of multivariate data.

FUTURE PERSPECTIVES

High-Throughput Screening Methods

Although the new ultra-fast sequencing technologies quickly generate a remarkable number of target gene candidates, functional assays are still needed to confirm them. These assays for protein function represent one of the most reliable and invaluable tools for mining target genes. Thus, developing of high-throughput screening (HTS) methods and improved chromogenic substrates for the detection of thermozymes (Kračun et al., 2015) is a priority for reducing the time invested in primary screening. HTS techniques increase the success of function-based metagenomic screens since they compensate for the often low hit rates in such screens (Ekkers et al., 2012). Apart from conventional high throughput screens, which use microtiter plate wells to store a large number of clones (Ko et al., 2013), microarray-based technologies coupled with microfluidic devices, cell compartmentalization, flow cytometry, and cell sorting are arising as promising new technologies for this purpose (Najah et al., 2014; Meier et al., 2015; Vidal-Melgosa et al., 2015). Microfluidic technologies are of undeniable interest when it comes to reaching screening rates of a million clones per day (Ufarté et al., 2015). This screening method

generally uses fluorogenic substrates (Najah et al., 2013) and it is based on the encapsulation of single clones of the metagenomic library in droplets, followed by the substrate induced gene-expression screening and the fluorescence-activated cell sorting to isolate plasmidic clones containing the genes of interest (Colin et al., 2015; Hosokawa et al., 2015). The main advantages of this ultra-fast screening method are the small volume required (usually picoliters to femtoliters) and the capability of detecting intracellular, extracellular, and membrane proteins. This approach could be used for the screening of thermozymes, as droplets can be incubated at high temperature before proceeding to the screening and fluorescence sorting. In this regard, there is an ongoing FP7 Marie Curie Action named HOTDROPS that involves four companies and four academic partners (including the authors' group) aimed to develop a microfluidics-based ultra-high-throughput platform for the selection of thermozymes from metagenomics and directed evolution libraries.

Advanced Sequencing

Until recently, most of the sequences collected in reference databases were related to humans and their pathogens. Currently, the advances in sequencing technologies have enabled the generation of considerable amounts of longer reads in less time. This fact, in addition to the lower per base cost and the development of metagenomics, has produced a relevant increase in the number of genomes sequenced and annotated deposited in databases like GenBank, thus covering a high range of microorganisms from a wide variety of habitats, including high temperature environments. Therefore, the bias in the databases toward microorganisms with clinical or pathogenic interest is decreasing, allowing a better analysis of the populations with metagenomics. Furthermore, metagenomics is becoming a tool in reach of many laboratories with the recent release of new cheaper and smaller devices such as the Oxford Nanopore MinION, a USB flash drive-size sequencer that measures deviations in electrical current as a single DNA strand passes through a protein nanopore (Bayley, 2015). However, this technology presents high error rates compared to the others (Goodwin et al., 2015) and still has to be improved.

Altogether, these breakthrough developments make metagenomics a more affordable and robust tool to explore the taxonomy and the functional diversity of microbial communities. Nevertheless, the complexity of microbial species, together with the limitations of the technology to cover fully whole genome sequences, still pose a great challenge for metagenome research. NGS technologies have limitations and remain at least an order of magnitude more expensive than other conventional microbiological assays, thus samples often must be individually barcoded and pooled into single runs to decrease costs. All these deficiencies will probably disappear as technologies continue developing like they did in the last years, from the end of the human genome sequencing project in 2003 (Collins et al., 2003), up to now.

Advances in Bioinformatics

Due to the massive amount of metagenome data generated in the last 10 years, infrastructural developments associated

with managing and serving sequence data are needed. Additionally, the fast growth in the size of data complicates its storage, organization, and distribution. As the volume of metagenomics data keeps growing, new assemblers have been developed, namely MEGAHIT that can assemble large and complex metagenomics data in a time and cost-efficient way, especially on a single-node server (Li et al., 2015).

New bioinformatic pipelines designed to support researchers involved in functional and taxonomic studies of environmental microbial communities have been released like BioMaS (Fosso et al., 2015), DUDes (Piro et al., 2016), or MOCAT2 (Kultima et al., 2016), among others.

Since there is an increasing number of complex communities sequenced, improved statistical methodology is needed, especially to enhance comparative studies where a large number of covariates (e.g., environmental

or host physiological parameters) are collected for each sample.

AUTHOR CONTRIBUTIONS

MD did all the data gathering and write-up. ER and MG supervised and reviewed the manuscript, providing comments and guidance during the manuscript development.

FUNDING

Funding both from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 324439, and from the Xunta de Galicia (Consolidación D.O.G. 10-10-2012, Contract Number: 2012/118) co-financed by FEDER. The work of MD was supported by a FPU fellowship (Ministerio de Educación Cultura y Deporte) FPU12/05050.

REFERENCES

- Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., and Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J. Bacteriol.* 186, 2629–2635. doi: 10.1128/JB.186.9.2629-2635.2004
- Adriaenssens, E. M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M., et al. (2015). Metagenomic analysis of the viral community in namib desert hypoliths. *Environ. Microbiol.* 17, 480–495. doi: 10.1111/1462-2920.12528
- Anderson, R. E., Brazelton, W. J., Baross, J. A., Altschul, S., Gish, W., Miller, W., et al. (2011). Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* 77, 120–133. doi: 10.1111/j.1574-6941.2011.01090.x
- Anderson, R. E., Sogin, M. L., Baross, J. A., Anderson, R., Beltrán, M., Hallam, S., et al. (2014). Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics. *PLoS ONE* 9:e109696. doi: 10.1371/journal.pone.0109696
- Angelov, A., Mientus, M., Liebl, S., and Liebl, W. (2009). A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Syst. Appl. Microbiol.* 32, 177–185. doi: 10.1016/j.syapm.2008.01.003
- Anobom, C. D., Pinheiro, A. S., De-Andrade, R. A., Aguiéiras, E. C. G., Andrade, G. C., Moura, M. V., et al. (2014). From structure to catalysis: recent developments in the biotechnological applications of lipases. *Biomed Res. Int.* 2014:684506. doi: 10.1155/2014/684506
- Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., and Weightman, A. J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl. Environ. Microbiol.* 71, 7724–7736. doi: 10.1128/AEM.71.12.7724-7736.2005
- Badhai, J., Ghosh, T. S., and Das, S. K. (2015). Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha, India. *Front. Microbiol.* 6:1166. doi: 10.3389/fmicb.2015.01166
- Baker, G., Gaffar, S., Cowan, D., and Suharto, R. (2001). Bacterial community analysis of Indonesian hot springs. *FEMS Microbiol. Lett.* 200, 103–109. doi: 10.1016/S0378-1097(01)00207-5
- 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. *Proc. Natl. Acad. Sci. U.S.A.* 91, 1609–1613. doi: 10.1073/pnas.91.5.1609
- Bayley, H. (2015). Nanopore sequencing: from imagination to reality. *Clin. Chem.* 61, 25–31. doi: 10.1373/clinchem.2014.223016
- Becerra, M., Cerdán, M. E., and González-Siso, M. I. (2015). Biobutanol from cheese whey. *Microb. Cell Fact.* 14, 27. doi: 10.1186/s12934-015-0200-1
- Bhalla, A., Bansal, N., Kumar, S., Bischoff, K. M., and Sani, R. K. (2013). Bioresource technology improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresour. Technol.* 128, 751–759. doi: 10.1016/j.biortech.2012.10.145
- Bijtenhoorn, P., Mayerhofer, H., Müller-Dieckmann, J., Utpatel, C., Schipper, C., Hornung, C., et al. (2011). A novel metagenomic short-chain dehydrogenase/reductase attenuates *Pseudomonas aeruginosa* biofilm formation and virulence on *Caenorhabditis elegans*. *PLoS ONE* 6:e26278. doi: 10.1371/journal.pone.0026278
- Biver, S., Portetelle, D., and Vandenbol, M. (2013). Characterization of a new oxidant-stable serine protease isolated by functional metagenomics. *Springerplus* 2:410. doi: 10.1186/2193-1801-2-410
- Böhnke, S., and Perner, M. (2015). A function-based screen for seeking RubisCO active clones from metagenomes: novel enzymes influencing RubisCO activity. *ISME J.* 9, 735–745. doi: 10.1038/ismej.2014.163
- Brady, A. L., Sharp, C. E., Grasby, S. E., and Dunfield, P. F. (2015). Anaerobic carboxydrotrophic bacteria in geothermal springs identified using stable isotope probing. *Front. Microbiol.* 6:897. doi: 10.3389/fmicb.2015.00897
- Brazelton, W. J., and Baross, J. A. (2009). Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J.* 3, 1420–1424. doi: 10.1038/ismej.2009.79
- Breitbart, M., Wegley, L., Leeds, S., Schoenfeld, T., and Rohwer, F. (2004). Phage community dynamics in hot springs. *Appl. Environ. Microbiol.* 70, 1633–1640. doi: 10.1128/AEM.70.3.1633-1640.2004
- Brock, T. D., and Freeze, H. (1969). *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J. Bacteriol.* 98, 289–297.
- Cai, L., Ye, L., Tong, A. H. Y., Lok, S., Zhang, T., Handelsman, J., et al. (2013). Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS ONE* 8:e53649. doi: 10.1371/journal.pone.0053649
- Chan, C. S., Chan, K.-G., Tay, Y.-L., Chua, Y.-H., and Goh, K. M. (2015). Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front. Microbiol.* 6:177. doi: 10.3389/fmicb.2015.00177
- Chen, Z. W., Liu, Y. Y., Wu, J. F., She, Q., Jiang, C. Y., and Liu, S. J. (2007). Novel bacterial sulfur oxygenase reductases from bioreactors treating gold-bearing concentrates. *Appl. Microbiol. Biotechnol.* 74, 688–698. doi: 10.1007/s00253-006-0691-0
- Cheng, J., Pinnell, L., Engel, K., Neufeld, J. D., and Charles, T. C. (2014). Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. *J. Microbiol. Methods* 99, 27–34. doi: 10.1016/j.mimet.2014.01.015
- Chistoserdova, L. (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol. Lett.* 32, 1351–1359. doi: 10.1007/s10529-010-0306-9

- Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., et al. (2008). Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* 322, 275–278. doi: 10.1126/science.1155495
- Chow, J., Kovacic, F., Dall'Antonia, Y., Krauss, U., Fersini, F., Schmeisser, C., et al. (2012). The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. *PLoS ONE* 7:e47665. doi: 10.1371/journal.pone.0047665
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, 1–10. doi: 10.1093/nar/gkt1244
- Colin, P.-Y., Kintsas, B., Gielen, F., Miton, C. M., Fischer, G., Mohamed, M. F., et al. (2015). Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun.* 6:10008. doi: 10.1038/ncomms10008
- Collins, F. S., Morgan, M., Patrinos, A., Watson, J. D., Olson, M. V., Collins, F. S., et al. (2003). The human genome project: lessons from large-scale biology. *Science* 300, 286–290. doi: 10.1126/science.1084564
- Colman, D. R., Jay, Z. J., Inskeep, W. P., Jennings, R. deM., Maas, K. R., Rusch, D. B., et al. (2016). Novel, deep-branching heterotrophic bacterial populations recovered from thermal spring metagenomes. *Front. Microbiol.* 7:304. doi: 10.3389/fmicb.2016.00304
- Coyotzi, S., Pratscher, J., Murrell, J. C., and Neufeld, J. D. (2016). Targeted metagenomics of active microbial populations with stable-isotope probing. *Curr. Opin. Biotechnol.* 41, 1–8. doi: 10.1016/j.copbio.2016.02.017
- Dadheech, P. K., Glöckner, G., Casper, P., Kotut, K., Mazzoni, C. J., Mbedi, S., et al. (2013). Cyanobacterial diversity in the hot spring, pelagic and benthic habitats of a tropical soda lake. *FEMS Microbiol. Ecol.* 85, 389–401. doi: 10.1111/1574-6941.12128
- De la Torre, J. R., Walker, C. B., Ingalls, A. E., Könneke, M., and Stahl, D. A. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ. Microbiol.* 10, 810–818. doi: 10.1111/j.1462-2920.2007.01506.x
- Denonfoux, J., Parisot, N., Dugat-Bony, E., Biderre-Petit, C., Boucher, D., Morgavi, D. P., et al. (2013). Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Res.* 20, 185–196. doi: 10.1093/dnares/dst001
- Dinsdale, E. A., Edwards, R. A., Bailey, B. A., Tuba, I., Akhter, S., McNair, K., et al. (2013). Multivariate analysis of functional metagenomes. *Front. Genet.* 4:41. doi: 10.3389/fgene.2013.00041
- Dougherty, M. J., D'haeseleer, P., Hazen, T. C., Simmons, B. A., Adams, P. D., and Hadi, M. Z. (2012). Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. *BMC Biotechnol.* 12:38. doi: 10.1186/1472-6750-12-38
- Dröge, J., and McHardy, A. C. (2012). Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Brief. Bioinform.* 13, 646–655. doi: 10.1093/bib/bbs031
- Durán, N., and Esposito, E. (2000). Potential applications of oxidative enzymes and phenoloxidase-like compounds in wastewater and soil treatment: a review. *Appl. Catal. B Environ.* 28, 83–99. doi: 10.1016/S0926-3373(00)00168-5
- Ekkers, D. M., Cretioiu, M. S., Kielak, A. M., and Van Elsas, J. D. (2012). The great screen anomaly—a new frontier in product discovery through functional metagenomics. *Appl. Microbiol. Biotechnol.* 93, 1005–1020. doi: 10.1007/s00253-011-3804-3
- Fan, X., Liu, X., Huang, R., and Liu, Y. (2012a). Identification and characterization of a novel thermostable pyrethroid-hydrolyzing enzyme isolated through metagenomic approach. *Microb. Cell Fact.* 11:33. doi: 10.1186/1475-2859-11-33
- Fan, X., Liu, X., and Liu, Y. (2012b). The cloning and characterization of one novel metagenome-derived thermostable esterase acting on N-acylhomoserine lactones. *J. Mol. Catal. B Enzym.* 83, 29–37. doi: 10.1016/j.molcatb.2012.07.006
- Fan, X., Liu, X., Wang, K., Wang, S., Huang, R., and Liu, Y. (2011). Highly soluble expression and molecular characterization of an organic solvent-stable and thermotolerant lipase originating from the metagenome. *J. Mol. Catal. B Enzym.* 72, 319–325. doi: 10.1016/j.molcatb.2011.07.009
- Fancello, L., Trape, S., Robert, C., Boyer, M., Popgeorgiev, N., Raoult, D., et al. (2012). Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J.* 7, 359–369. doi: 10.1038/ismej.2012.101
- Farmer, J. (1998). Thermophiles, early biosphere evolution, and the origin of life on Earth: implications for the exobiological exploration of Mars. *J. Geophys. Res.* 103, 457–461. doi: 10.1029/98JE01542
- Ferrandi, E. E., Sayer, C., Isupov, M. N., Annovazzi, C., Marchesi, C., Iacobone, G., et al. (2015). Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. *FEBS J.* 282, 2879–2894. doi: 10.1111/febs.13328
- Ferrer, M., Martínez-Abarca, F., and Golyshin, P. N. (2005). Mining genomes and “metagenomes” for novel catalysts. *Curr. Opin. Biotechnol.* 16, 588–593. doi: 10.1016/j.copbio.2005.09.001
- Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol.* 9, 22–34. doi: 10.1111/1751-7915.12309
- Ferris, M. J., Kühl, M., Wieland, A., and Ward, D. M. (2003). Cyanobacterial ecotypes in different optical microenvironments of a 68°C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. *Appl. Environ. Microbiol.* 69, 2893–2898. doi: 10.1128/AEM.69.5.2893-2898.2003
- Fiala, G., and Stetter, K. O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch. Microbiol.* 145, 56–61. doi: 10.1007/BF00413027
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Fosso, B., Santamaria, M., Marzano, M., Alonso-Alemán, D., Valiente, G., Donvito, G., et al. (2015). BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS. *BMC Bioinformatics* 16:203. doi: 10.1186/s12859-015-0595-z
- Fu, L., He, Y., Xu, F., Ma, Q., Wang, F., and Xu, J. (2015). Characterization of a novel thermostable patatin-like protein from a Guaymas basin metagenomic library. *Extremophiles* 19, 829–840. doi: 10.1007/s00792-015-0758-x
- Fuciños, P., Atanes, E., López-López, O., Solaroli, M., Cerdán, M. E., González-Siso, M. I., et al. (2014). Cloning, expression, purification and characterization of an oligomeric His-tagged thermophilic esterase from *Thermus thermophilus* HB27. *Process Biochem.* 49, 927–935. doi: 10.1016/j.procbio.2014.03.006
- Gao, G., Wang, A., Gong, B., Li, Q., Liu, Y., He, Z., et al. (2016). A novel metagenome-derived gene cluster from termite hindgut: encoding phosphotransferase system components and high glucose tolerant glucosidase. *Enzyme Microb. Technol.* 84, 24–31. doi: 10.1016/j.enzmictec.2015.12.005
- George, N., Singh, P., Kumar, V., Puri, N., and Gupta, N. (2014). Approach to ecofriendly leather: characterization and application of an alkaline protease for chemical free dehairing of skins and hides at pilot scale. *J. Clean. Prod.* 79, 249–257. doi: 10.1016/j.jclepro.2014.05.046
- Gerbl, F. W., Weidler, G. W., Wanek, W., Erhardt, A., and Stan-Lotter, H. (2014). Thaumarchaeal ammonium oxidation and evidence for a nitrogen cycle in a subsurface radioactive thermal spring in the Austrian Central Alps. *Front. Microbiol.* 5:225. doi: 10.3389/fmicb.2014.00225
- Ghelani, A., Patel, R., Mangrola, A., and Dudhagara, P. (2015). Cultivation-independent comprehensive survey of bacterial diversity in Tulsi Shyam Hot Springs, India. *Genomics Data* 4, 54–56. doi: 10.1016/j.gdata.2015.03.003
- Gladden, J. M., Allgaier, M., Miller, C. S., Hazen, T. C., VanderGheynst, J. S., Hugenholtz, P., et al. (2011). Glycoside hydrolase activities of thermophilic bacterial consortia adapted to switchgrass. *Appl. Environ. Microbiol.* 77, 5804–5812. doi: 10.1128/AEM.00032-11
- Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M., and McCombie, W. R. (2015). Oxford Nanopore sequencing and *de novo* assembly of a eukaryotic genome. *Genome Res.* 25, 1750–1756. doi: 10.1101/013490
- Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., et al. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat. Commun.* 2:375. doi: 10.1038/ncomms1373
- Gudbergssdóttir, S. R., Menzel, P., Krogh, A., Young, M., and Peng, X. (2016). Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. *Environ. Microbiol.* 18, 863–874. doi: 10.1111/1462-2920.13079

- Gupta, P., Manjula, A., Rajendhran, J., Gunasekaran, P., and Vakhlu, J. (2016). Comparison of metagenomic DNA extraction methods for soil sediments of high elevation Puga hot spring in Ladakh, India to explore bacterial diversity. *Geomicrobiol. J.* doi: 10.1080/01490451.2015.1128995. [Epub ahead of print].
- Gupta, R., Govil, T., Capalash, N., and Sharma, P. (2012). Characterization of a glycoside hydrolase family 1 β -galactosidase from hot spring metagenome with transglycosylation activity. *Appl. Biochem. Biotechnol.* 168, 1681–1693. doi: 10.1007/s12010-012-9889-z
- Haddar, A., Agrebi, R., Bougatef, A., Hmidet, N., Sellami-kamoun, A., and Nasri, M. (2009). Bioresource technology two detergent stable alkaline serine-proteases from *Bacillus mojavensis* A21: purification, characterization and potential application as a laundry detergent additive. *Bioresour. Technol.* 100, 3366–3373. doi: 10.1016/j.biortech.2009.01.061
- Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–685. doi: 10.1128/MBR.68.4.669
- He, Y., Xiao, X., and Wang, F. (2013). Metagenome reveals potential microbial degradation of hydrocarbon coupled with sulfate reduction in an oil-immersed chimney from Guaymas Basin. *Front. Microbiol.* 4:148. doi: 10.3389/fmicb.2013.00148
- Hedlund, B. P., Dodsworth, J. A., Cole, J. K., and Panosyan, H. H. (2013). An integrated study reveals diverse methanogens, Thaumarchaeota, and yet uncultivated archaeal lineages in Armenian hot springs. *Antonie van Leeuwenhoek* 104, 71–82. doi: 10.1007/s10482-013-9927-z
- Hosokawa, M., Hoshino, Y., Nishikawa, Y., Hirose, T., Yoon, D. H., Mori, T., et al. (2015). Droplet-based microfluidics for high-throughput screening of a metagenomic library for isolation of microbial enzymes. *Biosens. Bioelectron.* 67, 379–385. doi: 10.1016/j.bios.2014.08.059
- 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 Microbiol. Ecol.* 85, 452–464. doi: 10.1111/1574-6941.12134
- Hug, K., Maher, W. A., Stott, M. B., Krikowa, F., Foster, S., and Moreau, J. W. (2014). Microbial contributions to coupled arsenic and sulfur cycling in the acid-sulfide hot spring Champagne Pool, New Zealand. *Front. Microbiol.* 5:569. doi: 10.3389/fmicb.2014.00569
- Huson, D., Mitra, S., and Ruscheweyh, H. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552–1560. doi: 10.1101/gr.120618.111
- Ilmberger, N., Meske, D., Juergensen, J., Schulte, M., Barthen, P., Rabausch, U., et al. (2012). Metagenomic cellulases highly tolerant towards the presence of ionic liquids - Linking thermostability and halotolerance. *Appl. Microbiol. Biotechnol.* 95, 135–146. doi: 10.1007/s00253-011-3732-2
- Inskeep, W. P., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Rusch, D. B., Tringe, S. G., 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. doi: 10.3389/fmicb.2013.00095
- 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. doi: 10.1371/journal.pone.0009773
- Jabbour, D., Sorger, A., and Sahm, K. (2013). A highly thermoactive and salt-tolerant α -amylase isolated from a pilot-plant biogas reactor. *Appl. Microbiol. Biotechnol.* 97, 2971–2978. doi: 10.1007/s00253-012-4194-x
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Kang, C.-H., Oh, K.-H., Lee, M.-H., Oh, T.-K., Kim, B. H., and Yoon, J.-H. (2011). A novel family VII esterase with industrial potential from compost metagenomic library. *Microb. Cell Fact.* 10:41. doi: 10.1186/1475-2859-10-41
- Kim, H. J., Jeong, Y. S., Jung, W. K., Kim, S. K., Lee, H. W., Kahng, H. Y., et al. (2015). Characterization of novel family IV esterase and family I.3 lipase from an oil-polluted mud flat metagenome. *Mol. Biotechnol.* 57, 781–792. doi: 10.1007/s12033-015-9871-4
- Kim, K. H., and Bae, J. W. (2011). Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* 77, 7663–7668. doi: 10.1128/AEM.00289-11
- 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. doi: 10.3389/fmicb.2013.00106
- Klatt, C. G., Wood, J. M., Rusch, D. B., Bateson, M. M., Hamamura, N., Heidelberg, J. F., et al. (2011). Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J.* 5, 1262–1278. doi: 10.1038/ismej.2011.73
- Ko, K. C., Han, Y., Cheong, D. E., Choi, J. H., and Song, J. J. (2013). Strategy for screening metagenomic resources for exocellulase activity using a robotic, high-throughput screening system. *J. Microbiol. Methods* 94, 311–316. doi: 10.1016/j.mimet.2013.07.010
- Kotlar, H. K., Lewin, A., Johansen, J., Throne-Holst, M., Haverkamp, T., Markussen, S., et al. (2011). High coverage sequencing of DNA from microorganisms living in an oil reservoir 2.5 kilometres subsurface. *Environ. Microbiol. Rep.* 3, 674–681. doi: 10.1111/j.1758-2229.2011.00279.x
- Kozubal, M. A., Romine, M., Jennings, R. deM., 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. doi: 10.1038/ismej.2012.132
- Kračun, S. K., Schückel, J., Westereng, B., Thygesen, L. G., Monrad, R. N., Eijssink, V. G. H., et al. (2015). A new generation of versatile chromogenic substrates for high-throughput analysis of biomass-degrading enzymes. *Biotechnol. Biofuels* 8:70. doi: 10.1186/s13068-015-0250-y
- Kultima, J. R., Coelho, L. P., Forslund, K., Huerta-Cepas, J., Li, S. S., Driessen, M., et al. (2016). MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* 32, 2520–2523. doi: 10.1093/bioinformatics/btw183
- Kumar, S., Krishnani, K. K., Bhushan, B., and Brahmane, M. P. (2015). Metagenomics: retrospect and prospects in high throughput age. *Biotechnol. Res. Int.* 2015:121735. doi: 10.1155/2015/121735
- Kumar, V., Marin-Navarro, J., and Shukla, P. (2016). Thermostable microbial xylanases for pulp and paper industries: trends, applications and further perspectives. *World J. Microbiol. Biotechnol.* 32, 34. doi: 10.1007/s11274-015-2005-0
- Kwon, E. J., Jeong, Y. S., Kim, Y. H., Kim, S. K., Na, H. B., Kim, J., et al. (2010). Construction of a metagenomic library from compost and screening of cellulase- and xylanase-positive clones. *J. Appl. Biol. Chem.* 53, 702–708. doi: 10.3839/jksabc.2010.106
- Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D., and Charles, T. C. (2015). Current and future resources for functional metagenomics. *Front. Microbiol.* 6:1196. doi: 10.3389/fmicb.2015.01196
- Leis, B., Angelov, A., Mientus, M., Li, H., Pham, V. T. T., Lauinger, B., et al. (2015). Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Front. Microbiol.* 6:275. doi: 10.3389/fmicb.2015.00275
- Lewin, A., Wentzel, A., and Valla, S. (2013). Metagenomics of microbial life in extreme temperature environments. *Curr. Opin. Biotechnol.* 24, 516–525. doi: 10.1016/j.copbio.2012.10.012
- Li, A., Chu, Y., Wang, X., Ren, L., Yu, J., Liu, X., et al. (2013a). A pyrosequencing-based metagenomic study of methane-producing microbial community in solid-state biogas reactor. *Biotechnol. Biofuels* 6:3. doi: 10.1186/1754-6834-6-3
- Li, D., Greenfield, P., Rosewarne, C. P., and Midgley, J. (2013b). Draft genome sequence of *Thermoanaerobacter* sp. Strain A7A, reconstructed from a metagenome obtained from a high-temperature hydrocarbon reservoir in the Bass Strait, Australia. *Genome Announc.* 1, e00701-13. doi: 10.1128/genomeA.00701-13
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Lin, K.-H., Liao, B.-Y., Chang, H.-W., Huang, S.-W., Chang, T.-Y., Yang, C.-Y., et al. (2015). Metabolic characteristics of dominant microbes and key rare

- species from an acidic hot spring in Taiwan revealed by metagenomics. *BMC Genomics* 16:1029. doi: 10.1186/s12864-015-2230-9
- Liu, B., Gibbons, T., Ghodsi, M., and Pop, M. (2010). "MetaPhyler: taxonomic profiling for metagenomic sequences," *2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Hong Kong), 95–100.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012:251364. doi: 10.1155/2012/251364
- Liu, W. T., Marsh, T. L., Cheng, H., and Forney, L. J. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* 63, 4516–4522.
- Liu, Z., Zhao, C., Deng, Y., Huang, Y., and Liu, B. (2015). Characterization of a thermostable recombinant β -galactosidase from a thermophilic anaerobic bacterial consortium YTY-70. *Biotechnol. Biotechnol. Equip.* 29, 547–554. doi: 10.1080/13102818.2015.1015244
- Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. doi: 10.1093/nar/gkt1178
- López-López, O., Cerdán, M. E., and González Siso, M. I. (2014). New extremophilic lipases and esterases from metagenomics. *Curr. Protein Pept. Sci.* 15, 445–455. doi: 10.2174/1389203715666140228153801
- López-López, O., Cerdán, M.-E., and González-Siso, M.-I. (2015a). *Thermus thermophilus* as a source of thermostable lipolytic enzymes. *Microorganisms* 3, 792–808. doi: 10.3390/microorganisms3040792
- López-López, O., Knapik, K., Cerdán, M. E., and González-Siso, M. I. (2015b). Metagenomics of an alkaline hot spring in Galicia (Spain): microbial diversity analysis and screening for novel lipolytic enzymes. *Front. Microbiol.* 6:1291. doi: 10.3389/fmicb.2015.01291
- Luo, C., Tsementzi, D., Kyripides, N., Read, T., Konstantinidis, K. T., Nelson, K., et al. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE* 7:e30087. doi: 10.1371/journal.pone.0030087
- Macwan, A. S., Javed, S., and Kumar, A. (2011). Isolation of a novel thermostable dehydrochlorinase (LinA) from a soil metagenome. *3 Biotech* 1, 193–198. doi: 10.1007/s13205-011-0012-x
- Mangrola, A., Dudhagara, P., Koringa, P., Joshi, C. G., Parmar, M., and Patel, R. (2015a). Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. *Genomics Data* 4, 153–155. doi: 10.1016/j.gdata.2015.04.014
- Mangrola, A. V., Dudhagara, P., Koringa, P., Joshi, C. G., and Patel, R. K. (2015b). Shotgun metagenomic sequencing based microbial diversity assessment of Lasundra hot spring, India. *Genomics Data* 4, 73–75. doi: 10.1016/j.gdata.2015.03.005
- Manoharan, L., Kushwaha, S. K., Hedlund, K., and Åhrén, D. (2015). Captured metagenomics: large-scale targeting of genes based on "sequence capture" reveals functional diversity in soils. *DNA Res.* 22, 451–460. doi: 10.1093/dnares/dsv026
- Markowitz, V. M., Chen, I. M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., et al. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 42, 568–573. doi: 10.1093/nar/gkt919
- Martins, L. F., Antunes, L. P., Pascon, R. C., de Oliveira, J. C. F., Digiampietri, L. A., Barbosa, D., et al. (2013). Metagenomic analysis of a tropical composting operation at the São Paulo zoo park reveals diversity of biomass degradation functions and organisms. *PLoS ONE* 8:e61928. doi: 10.1371/journal.pone.0061928
- Maruthamuthu, M., Jiménez, D. J., Stevens, P., and Van Elsland, J. D. (2016). A multi-substrate approach for functional metagenomics-based screening for (hemi) cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes. *BMC Genomics* 17:86. doi: 10.1186/s12864-016-2404-0
- Mayumi, D., Akutsu-Shigeno, Y., Uchiyama, H., Nomura, N., and Nakajima-Kambe, T. (2008). Identification and characterization of novel poly(DL-lactic acid) depolymerases from metagenome. *Appl. Microbiol. Biotechnol.* 79, 743–750. doi: 10.1007/s00253-008-1477-3
- Meddeb-Mouelhi, F., Kelly, J., and Beauregard, M. (2014). Enzyme and Microbial Technology A comparison of plate assay methods for detecting extracellular cellulase and xylanase activity. *Enzyme Microb. Technol.* 66, 16–19. doi: 10.1016/j.enzymtec.2014.07.004
- Mehetre, G. T., Paranjpe, A. S., Dastager, S. G., and Dharne, M. S. (2016). Complete metagenome sequencing based bacterial diversity and functional insights from basaltic hot spring of Unkeshwar, Maharashtra, India. *Genomics Data* 7, 140–143. doi: 10.1016/j.gdata.2015.12.031
- Meier, M. J., Paterson, E. S., and Lambert, I. B. (2015). Use of substrate-induced gene expression in metagenomic analysis of an aromatic hydrocarbon-contaminated soil. *Appl. Environ. Microbiol.* 82, 897–909. doi: 10.1128/AEM.03306-15
- Meilleur, C., Hupé, J. F., Juteau, P., and Shareck, F. (2009). Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. *J. Ind. Microbiol. Biotechnol.* 36, 853–861. doi: 10.1007/s10295-009-0562-7
- Menzel, P., Gudbergssdóttir, S. R., Rike, A. G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* 70, 411–424. doi: 10.1007/s00248-015-0576-9
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., et al. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386
- Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W., Banfield, J. F., Pace, N., et al. (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* 12:R44. doi: 10.1186/gb-2011-12-5-r44
- Mirete, S., Morgante, V., and González-Pastor, J. E. (2016). Functional metagenomics of extreme environments. *Curr. Opin. Biotechnol.* 38, 143–149. doi: 10.1016/j.copbio.2016.01.017
- Mitchell, K. R., and Takacs-Vesbach, C. D. (2008). A comparison of methods for total community DNA preservation and extraction from various thermal environments. *J. Ind. Microbiol. Biotechnol.* 35, 1139–1147. doi: 10.1007/s10295-008-0393-y
- Morrison, L. E., and Tanner, F. W. (1922). Studies on Thermophilic Bacteria: I. Aerobic Thermophilic Bacteria from Water. *J. Bacteriol.* 7, 343–366.
- Moser, M. J., DiFrancesco, R. A., Gowda, K., Klingele, A. J., Sugar, D. R., Stocki, S., et al. (2012). Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS ONE* 7:e38371. doi: 10.1371/journal.pone.0038371
- Muyzer, G., De Waal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Env. Microbiol.* 59, 695–700.
- Najah, M., Calbrix, R., Mahendra-Wijaya, I. P., Beneyton, T., Griffiths, A. D., and Drevelle, A. (2014). Droplet-based microfluidics platform for ultra-high-throughput bioprospecting of cellulolytic microorganisms. *Chem. Biol.* 21, 1722–1732. doi: 10.1016/j.chembiol.2014.10.020
- Najah, M., Mayot, E., Mahendra-Wijaya, I. P., Griffiths, A. D., Ladame, S., and Drevelle, A. (2013). New glycosidase substrates for droplet-based microfluidic screening. *Anal. Chem.* 85, 9807–9814. doi: 10.1021/ac4022709
- Nakai, R., Abe, T., Takeyama, H., and Naganuma, T. (2011). Metagenomic analysis of 0.2- μ m-passable microorganisms in deep-sea hydrothermal fluid. *Mar. Biotechnol.* 13, 900–908. doi: 10.1007/s10126-010-9351-6
- Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40:e155. doi: 10.1093/nar/gks678
- Neveu, J., Regard, C., and Dubow, M. S. (2011). Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. *Appl. Microbiol. Biotechnol.* 91, 635–644. doi: 10.1007/s00253-011-3256-9
- Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., and Stahl, D. A. (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* 40, 337–365. doi: 10.1146/annurev.mi.40.100186.002005
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, 206–214. doi: 10.1093/nar/gkt1226
- Panda, A. K., Bisht, S. S., Kumar, N. S., and De Mandal, S. (2015). Investigations on microbial diversity of Jakrem hot spring, Meghalaya, India

- using cultivation-independent approach. *Genomics Data* 4, 156–157. doi: 10.1016/j.gdata.2015.04.016
- Pap, B., Györkei, Á., Boboescu, I. Z., Nagy, I. K., Bíró, T., Kondorosi, É., et al. (2015). Temperature-dependent transformation of biogas-producing microbial communities points to the increased importance of hydrogenotrophic methanogenesis under thermophilic operation. *Bioresour. Technol.* 177, 375–380. doi: 10.1016/j.biortech.2014.11.021
- Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494
- Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174
- Peres, F., Martins, L. L., and Ferreira-Dias, S. (2015). Influence of enzymes and technology on virgin olive oil composition. *Crit. Rev. Food Sci. Nutr.* doi: 10.1080/10408398.2015.1092107. [Epub ahead of print].
- Pessela, B. C. C., Torres, R., Fuentes, M., Mateo, C., Filho, M., Carrascosa, A. V., et al. (2004). A simple strategy for the purification of large thermophilic proteins overexpressed in mesophilic microorganisms: application to multimeric enzymes from *Thermus* sp. strain T2 expressed in *Escherichia coli*. *Biotechnol. Prog.* 20, 1507–1511. doi: 10.1021/bp049785t
- Pham, V. H. T., and Kim, J. (2012). Cultivation of unculturable soil bacteria. *Trends Biotechnol.* 30, 475–484. doi: 10.1016/j.tibtech.2012.05.007
- Piro, V. C., Lindner, M. S., and Renard, B. Y. (2016). DUDS: a top-down taxonomic profiler for metagenomics. *Bioinformatics* 32, 2272–2280. doi: 10.1093/bioinformatics/btw150
- Prokofeva, M. I., Kublanov, I. V., Nercessian, O., Tourova, T. P., Kolganova, T. V., Lebedinsky, A. V., et al. (2005). Cultivated anaerobic acidophilic/acidotolerant thermophiles from terrestrial and deep-sea hydrothermal habitats. *Extremophiles* 9, 437–448. doi: 10.1007/s00792-005-0461-4
- Purcell, D., Sompong, U., Yim, L. C., Barraclough, T. G., Peerapornpisal, Y., and Pointing, S. B. (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. doi: 10.1111/j.1574-6941.2007.00302.x
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, 590–596. doi: 10.1093/nar/gks1219
- Rademacher, A., Zakrzewski, M., Schlüter, A., Schönborg, M., Szczepanowski, R., Goesmann, A., et al. (2012). Characterization of microbial biofilms in a thermophilic biogas system by high-throughput metagenome sequencing. *FEMS Microbiol. Ecol.* 79, 785–799. doi: 10.1111/j.1574-6941.2011.01265.x
- Rhee, J.-K., Ahn, D.-G., Kim, Y.-G., and Oh, J.-W. (2005). New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl. Environ. Microbiol.* 71, 817–825. doi: 10.1128/AEM.71.2.817-825.2005
- Roh, C., and Schmid, R. D. (2013). Isolation of an organic solvent-tolerant lipolytic enzyme from uncultivated microorganism. *Appl. Biochem. Biotechnol.* 171, 1750–1758. doi: 10.1007/s12010-013-0464-z
- Rohwer, F., Prangishvili, D., and Lindell, D. (2009). Roles of viruses in the environment. *Environ. Microbiol.* 11, 2771–2774. doi: 10.1111/j.1462-2920.2009.02101.x
- Rozanov, A. S., Bryanskaya, A. V., Malup, T. K., Meshcheryakova, I. A., Lazareva, E. V., Taran, O. P., et al. (2014). Molecular analysis of the benthos microbial community in Zavarzin thermal spring (Uzon Caldera, Kamchatka, Russia). *BMC Genomics* 15:S12. doi: 10.1186/1471-2164-15-S12-S12
- Sahm, K., John, P., Nacke, H., Wemheuer, B., Grote, R., Daniel, R., et al. (2013). High abundance of heterotrophic prokaryotes in hydrothermal springs of the Azores as revealed by a network of 16S rRNA gene-based methods. *Extremophiles* 17, 649–662. doi: 10.1007/s00792-013-0548-2
- Sangwan, N., Lambert, C., Sharma, A., Gupta, V., Khurana, P., Khurana, J. P., et al. (2015). Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator-prey genotypes. *Environ. Microbiol. Rep.* 7, 812–823. doi: 10.1111/1758-2229.12297
- Sarmiento, F., Peralta, R., and Blamey, J. M. (2015). Cold and hot extremozymes: industrial relevance and current trends. *Front. Bioeng. Biotechnol.* 3:148. doi: 10.3389/fbioe.2015.00148
- Sathya, T. A., and Khan, M. (2014). Diversity of glycosyl hydrolase enzymes from metagenome and their application in food industry. *J. Food Sci.* 79, R2149–R2156. doi: 10.1111/1750-3841.12677
- Schmidt, T. M., Delong, E. F., and Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* 173, 4371–4378.
- Schoenfeld, T., Patterson, M., Richardson, P. M., Wommack, K. E., Young, M., and Mead, D. (2008). Assembly of viral metagenomes from yellowstone hot springs. *Appl. Environ. Microbiol.* 74, 4164–4174. doi: 10.1128/AEM.02598-07
- Schoenfeld, T. W., Murugapiran, S. K., Dodsworth, J. A., Floyd, S., Lodes, M., Mead, D. A., et al. (2013). Lateral gene transfer of family A DNA polymerases between thermophilic viruses, aquificae, and apicomplexa. *Mol. Biol. Evol.* 30, 1653–1664. doi: 10.1093/molbev/mst078
- Schröder, C., Elleuche, S., Blank, S., and Antranikian, G. (2014). Characterization of a heat-active archaeal β -glucosidase from a hydrothermal spring metagenome. *Enzyme Microb. Technol.* 57, 48–54. doi: 10.1016/j.enzmictec.2014.01.010
- Sen, S. K., Jana, A., Bandyopadhyay, P., Das Mohapatra, P. K., and Raut, S. (2016). Thermostable amylase production from hot spring isolate *Exiguobacterium* sp: a promising agent for natural detergents. *Sustain. Chem. Pharm.* 3, 59–68. doi: 10.1016/j.scp.2016.04.002
- Servín-Garcidueñas, L. E., Peng, X., Garrett, R. A., and Martínez-Romero, E. (2013). Genome sequence of a novel archaeal rudivirus recovered from a mexican hot spring. *Genome Announc.* 1, e00040-12. doi: 10.1128/genomeA.00040-12
- Shah, N., Tang, H., Doak, T. G., and Ye, Y. (2011). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac. Symp. Biocomput.* 17, 165–176. doi: 10.1142/9789814335058
- Shao, H., Xu, L., and Yan, Y. (2013). Isolation and characterization of a thermostable esterase from a metagenomic library. *J. Ind. Microbiol. Biotechnol.* 40, 1211–1222. doi: 10.1007/s10295-013-1317-z
- Sharma, A., Gilbert, J. A., and Lal, R. (2016). (Meta)genomic insights into the pathogenome of *Cellulosimicrobium cellulans*. *Sci. Rep.* 6:25527. doi: 10.1038/srep25527
- Sharma, A., Jani, K., Shouche, Y. S., and Pandey, A. (2015). Microbial diversity of the Solihar hot spring, India, assessed by analyzing 16S rRNA and protein-coding genes. *Ann. Microbiol.* 65, 1323–1332. doi: 10.1007/s13213-014-0970-4
- Sharma, N., Tanksale, H., Kapley, A., and Purohit, H. J. (2012). Mining the metagenome of activated biomass of an industrial wastewater treatment plant by a novel method. *Indian J. Microbiol.* 52, 538–543. doi: 10.1007/s12088-012-0263-1
- Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* 5:209. doi: 10.3389/fpls.2014.00209
- Shi, H., Zhang, Y., Li, X., Huang, Y., Wang, L., Wang, Y., et al. (2013). A novel highly thermostable xylanase stimulated by Ca^{2+} from *Thermotoga thermarum*: cloning, expression and characterization. *Biotechnol. Biofuels* 6:26. doi: 10.1186/1754-6834-6-26
- Silva, G. G. Z., Green, K. T., Dutilh, B. E., and Edwards, R. A. (2015). SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* 32, 354–361. doi: 10.1093/bioinformatics/btv584
- Singh, A., and Subudhi, E. (2016). Profiling of microbial community of Odisha hot spring based on metagenomic sequencing. *Genomics Data* 7, 187–188. doi: 10.1016/j.gdata.2016.01.004
- Singh, R., Chopra, C., and Kumar, V. (2015). Purification and characterization of CHpro1, a thermotolerant, alkali-stable and oxidation-resisting protease of Chumathang hot spring. *Sci. Bull.* 60, 1252–1260. doi: 10.1007/s11434-015-0834-8
- Singh, R., Dhawan, S., Singh, K., and Kaur, J. (2012). Cloning, expression and characterization of a metagenome derived thermoactive/thermostable pectinase. *Mol. Biol. Rep.* 39, 8353–8361. doi: 10.1007/s11033-012-1685-x
- Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C., and Pandey, A. (2013). Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour. Technol.* 127, 500–507. doi: 10.1016/j.biortech.2012.09.012

- Sonbol, S. A., Ferreira, A. J. S., and Siam, R. (2016). Red Sea Atlantis II brine pool nitrilase with unique thermostability profile and heavy metal tolerance. *BMC Biotechnol.* 16:14. doi: 10.1186/s12896-016-0244-2
- 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. doi: 10.1007/s00792-010-0307-6
- 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. *Environ. Microbiol.* 15, 1160–1175. doi: 10.1111/1462-2920.12025
- Stamps, B. W., Corsetti, F. A., Spear, J. R., and Stevenson, B. S. (2014). Draft genome of a novel Chlorobi member assembled by tetranucleotide binning of a hot spring metagenome. *Genome Announc.* 2, e00897–e00814. doi: 10.1128/genomeA.00897-14.Copyright
- Stetter, K. O. (2006). Hyperthermophiles in the history of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1837–1842. doi: 10.1098/rstb.2006.1907
- Suenaga, H., Ohnuki, T., and Miyazaki, K. (2007). Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ. Microbiol.* 9, 2289–2297. doi: 10.1111/j.1462-2920.2007.01342.x
- Sun, M. Z., Zheng, H. C., Meng, L. C., Sun, J. S., Song, H., Bao, Y. J., et al. (2015). Direct cloning, expression of a thermostable xylanase gene from the metagenomic DNA of cow dung compost and enzymatic production of xylooligosaccharides from corncob. *Biotechnol. Lett.* 37, 1877–1886. doi: 10.1007/s10529-015-1857-6
- Sundarram, A., Pandurangappa, T., and Murthy, K. (2014). α -amylase production and applications: a review. *J. Appl. Environ. Microbiol.* 2, 166–175. doi: 10.12691/jaem-2-4-10
- Sundberg, C., Al-Soud, W. A., Larsson, M., Alm, E., Yekta, S. S., Svensson, B. H., et al. (2013). 454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters. *FEMS Microbiol. Ecol.* 85, 612–626. doi: 10.1111/1574-6941.12148
- Tan, H., Mooij, M. J., Barret, M., Hegarty, P. M., Harington, C., Dobson, A. D. W., et al. (2014). Identification of novel phytase genes from an agricultural soil-derived metagenome. *J. Microbiol. Biotechnol.* 24, 113–118. doi: 10.4014/jmb.1307.07007
- Tan, H., Wu, X., Xie, L., and Huang, Z. (2016). Identification and characterization of a mesophilic phytase highly resilient to high-temperatures from a fungus-garden associated metagenome. *Appl. Microbiol. Biotechnol.* 100, 2225–2241. doi: 10.1007/s00253-015-7097-9
- Tekere, M., Lötter, A., Olivier, J., Jonker, N., and Venter, S. (2011). Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa. *African J. Biotechnol.* 10, 18005–18012. doi: 10.5897/AJB11.899
- Tsudome, M., Deguchi, S., Tsujii, K., Ito, S., and Horikoshi, K. (2009). Versatile solidified nanofibrous cellulose-containing media for growth of extremophiles. *Appl. Environ. Microbiol.* 75, 4616–4619. doi: 10.1128/AEM.00519-09
- Uchiyama, T., and Miyazaki, K. (2009). Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr. Opin. Biotechnol.* 20, 616–622. doi: 10.1016/j.copbio.2009.09.010
- Ufarté, L., Potocki-Veronese, G., and Laville, É. (2015). Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology. *Front. Microbiol.* 6:563. doi: 10.3389/fmicb.2015.00563
- Urbietta, M. S., Donati, E. R., Chan, K. G., Shahar, S., Sin, L. L., and Goh, K. M. (2015). Thermophiles in the genomic era: biodiversity, science, and applications. *Biotechnol. Adv.* 33, 633–647. doi: 10.1016/j.biotechadv.2015.04.007
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi: 10.1016/j.tig.2014.07.001
- de Vasconcellos, S. P., Angolini, C. F. F., García, I. N. S., Martins Dellagnezze, B., da Silva, C. C., Marsaioli, A. J., et al. (2010). Screening for hydrocarbon biodegraders in a metagenomic clone library derived from Brazilian petroleum reservoirs. *Org. Geochem.* 41, 1067–1073. doi: 10.1016/j.orggeochem.2010.08.003
- Verma, D., Kawarabayasi, Y., Miyazaki, K., and Satyanarayana, T. (2013). Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (Mxyl) retrieved from compost-soil metagenome. *PLoS ONE* 8:e52459. doi: 10.1371/journal.pone.0052459
- Verma, D., and Satyanarayana, T. (2013). Improvement in thermostability of metagenomic GH11 endoxylanase (Mxyl) by site-directed mutagenesis and its applicability in paper pulp bleaching process. *J. Ind. Microbiol. Biotechnol.* 40, 1373–1381. doi: 10.1007/s10295-013-1347-6
- Vidal-Melgosa, S., Pedersen, H. L., Schückel, J., Arnal, G., Dumon, C., Amby, D. B., et al. (2015). A new versatile microarray-based method for high throughput screening of carbohydrate-active enzymes. *J. Biol. Chem.* 290, 9020–9036. doi: 10.1074/jbc.M114.630673
- Vidya, J., Swaroop, S., Singh, S., Alex, D., Sukumaran, R., and Pandey, A. (2011). Isolation and characterization of a novel α -amylase from a metagenomic library of Western Ghats of Kerala, India. *Biologia* 66, 939–944. doi: 10.2478/s11756-011-0126-y
- Wagner, I. D., and Wiegand, J. (2008). Diversity of thermophilic anaerobes. *Ann. N.Y. Acad. Sci.* 1125, 1–43. doi: 10.1196/annals.1419.029
- Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., et al. (2016). Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* 9:22. doi: 10.1186/s13068-016-0440-2
- Wang, H., Gong, Y., Xie, W., Xiao, W., Wang, J., Zheng, Y., et al. (2011). Identification and characterization of a novel thermostable gh-57 gene from metagenomic fosmid library of the Juan de Fuca ridge hydrothermal vent. *Appl. Biochem. Biotechnol.* 164, 1323–1338. doi: 10.1007/s12010-011-9215-1
- Wang, M., Lai, G. L., Nie, Y., Geng, S., Liu, L., Zhu, B., et al. (2015). Synergistic function of four novel thermostable glycoside hydrolases from a long-term enriched thermophilic methanogenic digester. *Front. Microbiol.* 6:509. doi: 10.3389/fmicb.2015.00509
- Wang, S. D., Guo, G. S., Li, L., Cao, L. C., Tong, L., Ren, G. H., et al. (2014). Identification and characterization of an unusual glycosyltransferase-like enzyme with β -galactosidase activity from a soil metagenomic library. *Enzyme Microb. Technol.* 57, 26–35. doi: 10.1016/j.enzmictec.2014.01.007
- Wang, S., Wang, K., Li, L., and Liu, Y. (2013). Isolation and characterization of a novel organic solvent-tolerant and halotolerant esterase from a soil metagenomic library. *J. Mol. Catal. B Enzym.* 95, 1–8. doi: 10.1016/j.molcatb.2013.05.015
- Wang, Y., and Qian, P. Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS ONE* 4:e7401. doi: 10.1371/journal.pone.0007401
- Wemheuer, B., Taube, R., Akyol, P., Wemheuer, F., and Daniel, R. (2013). Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka peninsula. *Archaea* 2013:136714. doi: 10.1155/2013/136714
- Wierzbicka-Woś, A., Bartasun, P., Cieślinski, H., and Kur, J. (2013). Cloning and characterization of a novel cold-active glycoside hydrolase family 1 enzyme with β -glucosidase, β -fucosidase and β -galactosidase activities. *BMC Biotechnol.* 13:22. doi: 10.1186/1472-6750-13-22
- Wilson, M. S., Siering, P. L., White, C. L., Hauser, M. E., and Bartles, A. N. (2008). Novel archaea and bacteria dominate stable microbial communities in North America's largest hot spring. *Microb. Ecol.* 56, 292–305. doi: 10.1007/s00248-007-9347-6
- Wu, M., and Scott, A. J. (2012). Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28, 1033–1034. doi: 10.1093/bioinformatics/bts079
- Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S. M., Meng, J., et al. (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J.* 5, 414–426. doi: 10.1038/ismej.2010.144
- Yap, W. H., Zhang, Z., and Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* 181, 5201–5209.
- Ye, M., Li, G., Liang, W. Q., and Liu, Y. H. (2010). Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. *Appl. Microbiol. Biotechnol.* 87, 1023–1031. doi: 10.1007/s00253-010-2507-5
- Zamora, M. A., Pinzón, A., Zambrano, M. M., Restrepo, S., Broadbelt, L. J., Moura, M., et al. (2015). A comparison between functional frequency and metabolic flows framed by biogeochemical cycles in metagenomes: the case

- of “El Coquito” hot spring located at Colombia’s national Nevados park. *Ecol. Modell.* 313, 259–265. doi: 10.1016/j.ecolmodel.2015.06.041
- Zhang, X., Li, H., Li, C.-J., Ma, T., Li, G., and Liu, Y.-H. (2013). Metagenomic approach for the isolation of a thermostable β -galactosidase with high tolerance of galactose and glucose from soil samples of Turpan Basin. *BMC Microbiol.* 13:237. doi: 10.1186/1471-2180-13-237
- Zhou, J., Bruns, M. A., and Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62, 316–322.
- Zhu, Y., Li, J., Cai, H., Ni, H., Xiao, A., and Hou, L. (2013). Characterization of a new and thermostable esterase from a metagenomic library. *Microbiol. Res.* 168, 589–597. doi: 10.1016/j.micres.2013.04.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 DeCastro, Rodríguez-Belmonte and González-Siso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



EstDZ3: A New Esterolytic Enzyme Exhibiting Remarkable Thermostability

Dimitra Zarafeta^{1,2†}, Zalan Szabo^{3†}, Danai Moschidi², Hien Phan⁴, Evangelia D. Chrysina¹, Xu Peng⁴, Colin J. Ingham³, Fragiskos N. Kolisis^{2*} and Georgios Skretas^{1*}

¹ Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation, Athens, Greece,

² Laboratory of Biotechnology, School of Chemical Engineering, National Technical University of Athens, Athens, Greece,

³ MicroDish B.V., Utrecht, Netherlands, ⁴ Danish Archaea Centre, Department of Biology, Copenhagen University, Copenhagen, Denmark

OPEN ACCESS

Edited by:

Kok Gan Chan,
University of Malaya, Malaysia

Reviewed by:

Hugh Morgan,
University of Waikato, New Zealand
Filip Meersman,
University College London, UK

*Correspondence:

Fragiskos N. Kolisis
kolisis@chemeng.ntua.gr
Georgios Skretas
gskretas@eie.gr

[†] These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 02 August 2016

Accepted: 24 October 2016

Published: 16 November 2016

Citation:

Zarafeta D, Szabo Z, Moschidi D,
Phan H, Chrysina ED, Peng X,
Ingham CJ, Kolisis FN and Skretas G
(2016) EstDZ3: A New Esterolytic
Enzyme Exhibiting Remarkable
Thermostability.
Front. Microbiol. 7:1779.
doi: 10.3389/fmicb.2016.01779

Lipolytic enzymes that retain high levels of catalytic activity when exposed to a variety of denaturing conditions are of high importance for a number of biotechnological applications. In this study, we aimed to identify new lipolytic enzymes, which are highly resistant to prolonged exposure to elevated temperatures. To achieve this, we searched for genes encoding for such proteins in the genomes of a microbial consortium residing in a hot spring located in China. After performing functional genomic screening on a bacterium of the genus *Dictyoglomus*, which was isolated from this hot spring following *in situ* enrichment, we identified a new esterolytic enzyme, termed EstDZ3. Detailed biochemical characterization of the recombinant enzyme, revealed that it constitutes a slightly alkalophilic and highly active esterase against esters of fatty acids with short to medium chain lengths. Importantly, EstDZ3 exhibits remarkable thermostability, as it retains high levels of catalytic activity after exposure to temperatures as high as 95°C for several hours. Furthermore, it exhibits very good stability against exposure to high concentrations of a variety of organic solvents. Interestingly, EstDZ3 was found to have very little similarity to previously characterized esterolytic enzymes. Computational modeling of the three-dimensional structure of this new enzyme predicted that it exhibits a typical α/β hydrolase fold that seems to include a “subdomain insertion”, which is similar to the one present in its closest homolog of known function and structure, the cinnamoyl esterase Lj0536 from *Lactobacillus johnsonii*. As it was found in the case of Lj0536, this structural feature is expected to be an important determinant of the catalytic properties of EstDZ3. The high levels of esterolytic activity of EstDZ3, combined with its remarkable thermostability and good stability against a range of organic solvents and other denaturing agents, render this new enzyme a candidate biocatalyst for high-temperature biotechnological applications.

Keywords: hyperthermostability, esterase, *Dictyoglomus*, functional genomics, biocatalysis, biotechnology

INTRODUCTION

Lipolytic enzymes (EC 3.1.1.x) catalyze the hydrolysis of ester bonds in lipids, and depending on their substrate preference, they are divided in two main classes, carboxylesterases (EC 3.1.1.1) and lipases (EC 3.1.1.3) (Brockerhoff, 2012). Carboxylesterases show specificity toward short to medium fatty acid chain lengths and water-soluble substrates, whereas lipases toward long-chained and water-insoluble ones (Bornscheuer, 2002; Brockerhoff, 2012). In non-aqueous media, many of these enzymes are capable of performing the inverse reaction and catalyze the synthesis of ester bonds (Bornscheuer and Kazlauskas, 2006). These characteristics, complemented by their ability to modify a very broad range of substrates with high chemo-, regio-, and enantio-selectivity, render lipolytic enzymes a very attractive class of catalysts for conducting biotransformations (Bornscheuer, 2002). Industrial applications of esterases and lipases are diverse and include the preparation of chiral compounds, the de-inking of paper pulps, the degradation of plastics, the synthesis of fine chemicals and flavoring agents, etc. (Zamost et al., 1991; Vieille and Zeikus, 2001; Kirk et al., 2002). Probably the most characteristic example of an industrially relevant esterolytic enzyme is that of the naproxen carboxylesterase from *Bacillus subtilis*, which is utilized for the biocatalytic synthesis of the non-steroidal drug naproxen (Quax and Broekhuizen, 1994).

In industrial settings, esterases and lipases are often required to perform well under harsh conditions. These include high temperatures, significant concentrations of organic solvents, metal ions, surfactants, and other agents known to cause protein denaturation and enzyme inactivation (Hough and Danson, 1999). Consequently, stability against elevated temperatures and tolerance to protein-destabilizing conditions in general, is a crucial prerequisite before the broad industrial use of this type of enzymes can be realized. During the last two decades, a growing number of thermostable enzymes that catalyze ester bond hydrolysis at elevated temperatures have been reported, mainly due to the employment of metagenomic analyses. However, hyperthermostable enzymes, i.e., enzymes that exhibit high levels of catalytic activity at temperatures above 80°C, are rarer and not many examples of such biocatalysts have been discovered and characterized.

In order to obtain new hyperthermostable enzymes, there are two main strategies, which are typically employed. The first one is protein engineering, either through rational design or directed evolution (Bornscheuer and Pohl, 2001; Dalby, 2011; Bornscheuer et al., 2012). In this approach, a mesophilic protein is optimized for stability against thermal denaturation through protein modeling-guided amino acid substitutions (rational design) or via random mutagenesis and genetic screening (directed evolution) (Matthews et al., 1987; Bornscheuer and Pohl, 2001; Lutz, 2010). The second strategy relies on the identification of naturally occurring enzymes, which have evolved to withstand high-temperature exposure by retaining their proper folding and catalytic activity under such conditions (Lorenz et al., 2002). In nature, hyperthermostable enzymes are often encoded in the genomes of hyperthermophilic

microorganisms that grow optimally at temperatures above 80°C. Such organisms are encountered in all types of terrestrial and marine hot environments, and are represented only by bacterial and archaeal species. Hyperthermostable enzymes encoded in such genomes can be identified by screening genomic or metagenomic material, either bioinformatically (Levisson et al., 2007; Graham et al., 2011; Zarafeta et al., 2016) or through functional genomic screening (Ewis et al., 2004; Tirawongsaroj et al., 2008; Peng et al., 2011).

In this study, we have identified a new esterolytic enzyme, termed EstDZ3. Detailed biochemical characterization of the recombinant protein revealed that it constitutes a highly active esterase with preference toward short to medium acyl chain length substrates. The most outstanding feature of EstDZ3 is its remarkable thermostability, as it was found to retain high levels of catalytic activity even after exposure to near boiling temperatures for several hours. Importantly, this enzyme is also highly stable when exposed to high concentrations of organic solvents for extended periods of time. EstDZ3 originates from a bacterium of the genus *Dictyoglomus* and its amino acid sequence exhibits very low homology to functionally characterized proteins. Structural modeling of the new enzyme predicted that it exhibits a typical α/β hydrolase fold, which seems to include a “subdomain insertion” similar to the one present in its closest homolog of known structure, the cinnamoyl esterase Lj0536 from *Lactobacillus johnsonii*. As it was found in the case of Lj0536, this “subdomain insertion” is expected to be an important determinant of the catalytic properties of this new enzyme. The high levels of esterolytic activity of EstDZ3, combined with its remarkable thermostability and good stability against a range of organic solvents and other denaturing agents, render this new enzyme a candidate biocatalyst for high-temperature biotechnological applications.

Environmental Sampling, Clone Isolation, and Expression Library Construction

In a previous attempt to isolate biomass-degrading thermophilic organisms, an *in situ* enrichment culture containing xanthan gum was established in a hot spring located at the Eryuan region of Yunnan, China (Menzel et al., 2015). The temperature of the sampling site when the sample was collected was 83°C and the pH about 7. After 10 days of incubation in the hot spring, a sample was collected and sealed immediately. This sample was then diluted and cultivated anaerobically at 78 and 83°C in the laboratory, as described in the “Materials and Methods” section. After three sequential passages in the same medium, the culture appeared homogeneous in morphology, with only rod-shaped cells of similar dimension visible under the microscope. Finally, the culture was diluted serially until single colonies were obtained from anaerobic Gelrite™ bottles.

A single colony, termed Ch5.6.S, was subsequently isolated and cultivated under anaerobic conditions in glucose-containing medium to avoid interference of xanthan gum with DNA extraction. Sequencing of the gene encoding for the 16S rRNA revealed a 98% nucleotide identity with that of *Dictyoglomus thermophilum*, thus indicating that the isolated clone belongs

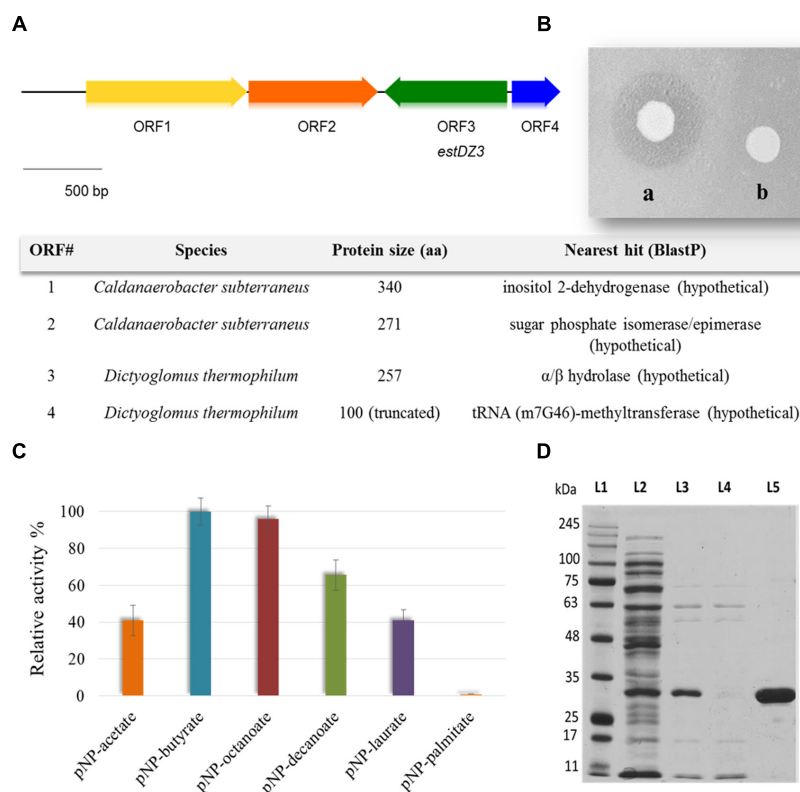


FIGURE 1 | Discovery, identification and purification of EstDZ3. (A) Map of the 3.3 kb *ch2* insert. ORFs indicated by arrows were annotated using BlastP and the closest homolog of each ORF is listed in the corresponding table (bottom). (B) *E. coli* BL21(DE3) cells carrying plasmid pLATE52-EstDZ3 (a) or empty vector (b), and grown on tributyrin-enriched agar containing 0.1 mM IPTG, the inducer of *estDZ3* overexpression. (C) Initial substrate profiling of the esterolytic activity of EstDZ3 using clarified lysates of BL21(DE3) cells overexpressing *estDZ3*. Cell lysates were assayed against pNP esters of fatty acids with acyl chain lengths varying from C2 to C16. The relative enzymic activity was measured spectrophotometrically at 410 nm. The reported values correspond to the mean value from three independent experiments performed in triplicate and the error bars to one standard deviation from the mean value. (D) SDS-PAGE analysis of EstDZ3, after heat-treatment and IMAC purification. Protein bands were visualized by Coomassie staining. L1: molecular weight marker; L2: soluble lysate of *E. coli* cells producing EstDZ3; L3: soluble lysate of the same cells after heat treatment; L4: flow-through after loading of the IMAC column with the lysate corresponding to L3; L5: eluted EstDZ3 after imidazole addition. The predicted molecular weight of the recombinant enzyme is 31.6 kDa.

to the *Dictyoglomus* genus. Then, genomic DNA derived from Ch5.6.S was isolated, partially digested, and fragments with sizes larger than 2 kb were cloned into the vector pUC18 to form a genomic library. The diversity of the generated library was ~300,000 independent clones as estimated by the number of colonies that appeared after plating serial dilutions of the transformed *Escherichia coli* cells.

Library Screening and Discovery of EstDZ3

The generated Ch5.6.S genomic library was transformed into electro-competent *E. coli* cells and was screened for sequences exhibiting lipolytic activity by plating onto LB agar medium containing 0.1% tributyrin (Lawrence et al., 1967). After 3 days of incubation at 37°C, a zone of clearance was observed around two colonies, indicating tributyrin hydrolysis. The positive clones were re-streaked on fresh LB-tributyrin agar plates and lipolytic activity was confirmed for one of them, termed Ch2.1. The plasmid isolated from Ch2.1 was purified and the

contained insert, termed *ch2*, was sequenced and found to correspond to a 3.3-kb DNA fragment, comprising four open reading frames (ORFs) that coded for the following putative proteins: (i) a hypothetical inositol 2-dehydrogenase from *Caldanaerobacter subterraneus* (ORF1), (ii) a hypothetical sugar phosphate isomerase/epimerase from *C. subterraneus* (ORF2), (iii) a predicted α/β hydrolase from *D. thermophilum* (ORF3), and (iv) a partial predicted tRNA (m7G46)-methyltransferase from *D. thermophilum* (ORF4) (Figure 1A).

Since the predicted α/β hydrolase was present in the selected clone as a full-length ORF and was also likely to confer the observed lipolytic activity, the corresponding gene, termed *estDZ3*, was cloned into the expression vector pLATE52 to form plasmid pLATE52-EstDZ3, which was used for heterologous expression of *estDZ3* in *E. coli*. A zone of clearance was observed around bacterial cells carrying pLATE52-EstDZ3 when grown onto tributyrin-enriched agar, in contrast to the same cells carrying an empty vector (Figure 1B), thus demonstrating that *estDZ3* is the gene responsible for the phenotype observed in the initial screen and suggests that *estDZ3* encodes for a protein

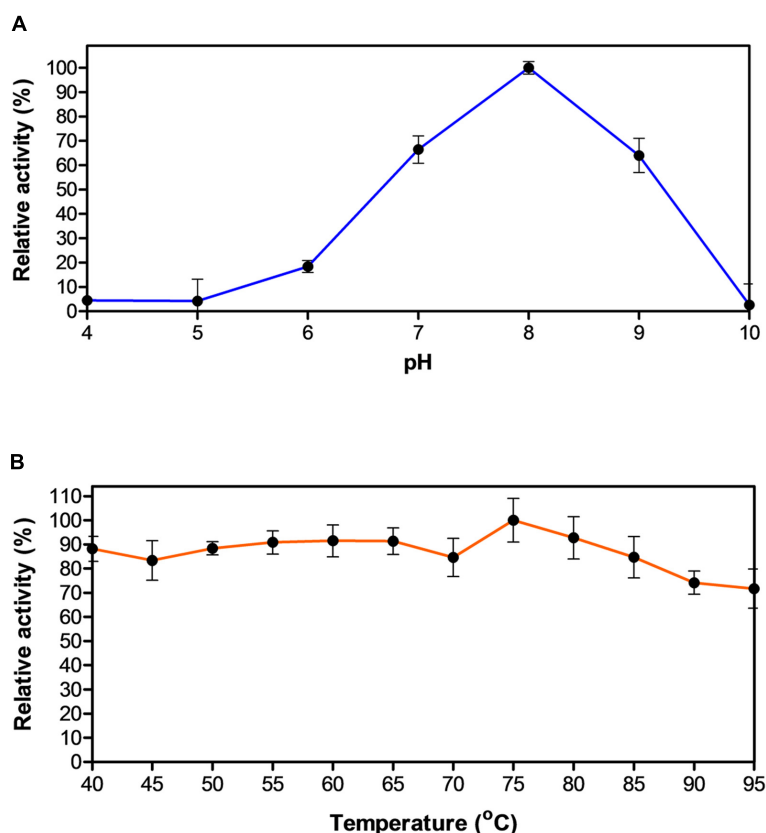


FIGURE 2 | Effect of pH and temperature on the activity of EstDZ3. (A) Effect of pH on the activity of EstDZ3. Enzymic activity was measured in the standard reaction at 40°C for 5 min at pH values ranging from 4 to 10. **(B)** Effect of temperature on the activity of EstDZ3. Enzymic activity was measured at temperatures ranging from 40 to 95°C and pH 8. The reported values correspond to the mean value from three independent experiments performed in triplicate and the error bars to one standard deviation from the mean value.

with hydrolytic activity against tributyrin. Furthermore, when the same cell lysates were assayed for their ability to hydrolyze *p*-nitrophenyl butyrate colorimetrically, the characteristic yellow color of *p*-nitrophenol (pNP), which is indicative of ester bond cleavage, was observed only when *estDZ3* was expressed (Figure 1C), thus confirming that EstDZ3 is an esterolytic enzyme.

An initial substrate preference test, using soluble lysates from *estDZ3*-expressing cells and pNP esters derived from fatty acids with a range of carbon chain lengths, demonstrated that EstDZ3 has a preference for short to medium size aliphatic chains (C2–C12), while its activity is barely detectable for C16 (Figure 1C). This suggests that EstDZ3 acts as a carboxylesterase rather than a lipase.

Biochemical Characterization of EstDZ3

In order to study the biochemical properties of EstDZ3, the enzyme was produced heterologously in *E. coli* and purified in soluble form. *E. coli* BL21(DE3) cells transformed with pLATE52-EstDZ3 were grown in liquid LB cultures and the production of EstDZ3 was induced by the addition of isopropyl- β -D-thiogalactoside (IPTG) as described in the “Materials and Methods” section. The recombinant protein

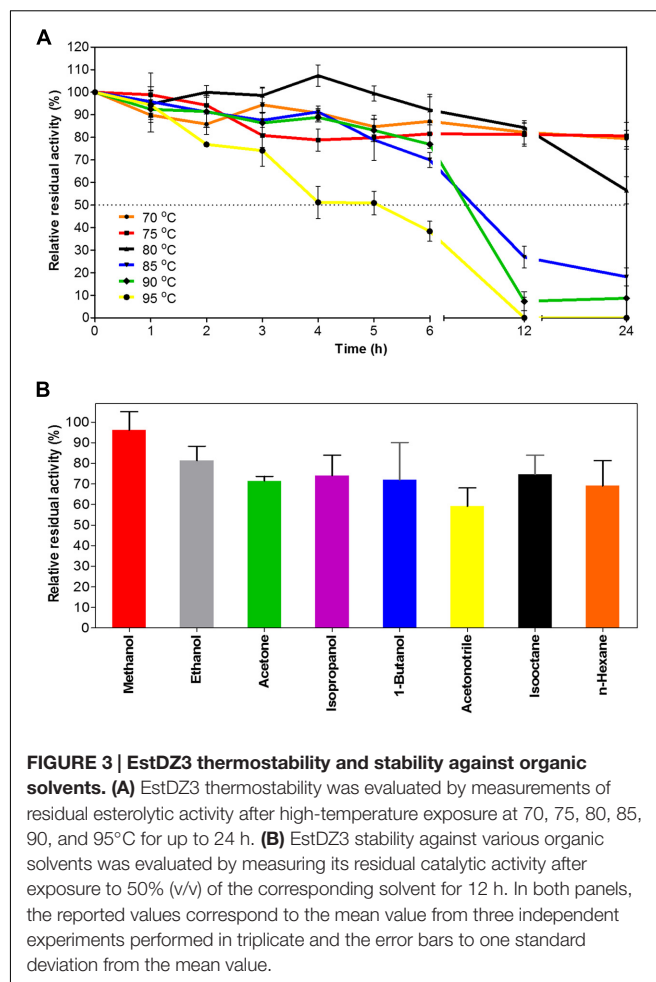
accumulated primarily in the soluble fraction of the bacterial lysate and was purified by an initial heat-treatment step, followed by immobilized metal affinity chromatography (IMAC) to near homogeneity as evaluated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 1D).

Biochemical characterization of EstDZ3 was carried out as described in the “Materials and Methods” section using pNP-butyrates as a substrate. First, we determined the optimal pH for EstDZ3 esterolytic activity, which was assayed within the value range of 4–10 at 40°C. Significant levels of catalytic activity were recorded at pH values 7–9, with an optimum at pH 8 (Figure 2A). Below pH 7 and above pH 9, the esterolytic activity of EstDZ3 was rapidly diminished. Measurements of its relative catalytic activity at different temperatures, on the other hand, revealed that EstDZ3 has a very broad temperature range of action, as its esterolytic activity remained practically unchanged at temperatures between 40 and 95°C (Figure 2B). This type of “flat” temperature profile is quite rare but has been observed previously for esterolytic and other hydrolytic enzymes as well (Aygan et al., 2008; Novototskaya-Vlasova et al., 2012). Thus, EstDZ3 is a slightly alkalophilic and highly thermotolerant esterase.

To study the substrate specificity of EstDZ3 in more detail, we determined the catalytic parameters of EstDZ3 using a range of esters of fatty acids with carbon chain lengths, varying from C2 to C12, with pNP. EstDZ3-mediated hydrolysis of these substrates followed Michaelis–Menten kinetics and revealed that the new enzyme shows preference toward short to medium chain-length substrates (C2 and C4) (Table 1). The highest catalytic efficiency was detected for pNP-butyrate (C4) with a k_{cat}/K_m value of $12,464 \text{ s}^{-1}\cdot\text{mM}^{-1}$ and decreased with increasing substrate chain length for C8 and C10, while the enzyme was found to be inactive against substrates with chains longer than C12 (Table 1; Figure 1C). EstDZ3 was also found capable of hydrolyzing efficiently the longer-chain substrate pNP-laurate (C12) (Table 1). However, we believe that this is probably an artifact of the presence of a poly-histidine tag in the recombinant enzyme, which may be causing the specificity of the enzyme to shift toward more hydrophobic substrates as observed in a number of previous studies (Lee et al., 1999; Peng et al., 2011). Collectively, these results demonstrate that EstDZ3 acts as an esterase rather than a lipase.

Performance of EstDZ3 When Exposed to High Temperatures, High Concentrations of Organic Solvents and Other Denaturing Agents

When exposed to high temperatures for prolonged periods of time, EstDZ3 retained very high stability, as determined by measurements of residual levels of its catalytic activity. At 70 and 75°C, EstDZ3 esterolytic activity was practically unchanged even after 24 h of incubation, while when incubated at 80°C, the enzyme exhibited a half-life of more than 24 h (Figure 3A). Importantly, EstDZ3 exhibited significant levels of esterolytic activity for several hours even after incubation at temperatures as high as 95°C (Figure 3A). Furthermore, EstDZ3 exhibited exquisite stability against high concentrations of a variety of organic solvents. More specifically, EstDZ3 activity was found to be practically unaffected after the enzyme had been exposed to 50% (v/v) methanol for 12 h (Figure 3B). Similarly, when this enzyme was exposed to the same concentration of ethanol, acetone, 1-butanol, isooctane, isopropanol and n-hexane for the same period of time, its residual activity was decreased by less than 30%. Finally, after exposure to 50% acetonitrile, EstDZ3 was found capable of retaining about 60% of its maximal activity (Figure 3B). These results demonstrate that EstDZ3 is an esterolytic enzyme with remarkable kinetic thermostability



and very good stability against prolonged exposure to high concentrations of organic solvents.

Subsequently, we studied the effects of a range of metal ions, reducing agents and detergents on the catalytic efficiency of EstDZ3. The esterolytic activity of the enzyme was practically unaffected by the addition of a variety of mono- and divalent metals such as Na^+ , K^+ , Li^{2+} , Mn^{+2} , and Mg^{+2} at 1 mM concentration (Table 2). The addition of 1 mM Ca^{2+} and Fe^{2+} resulted in a minor decrease in its catalytic activity by about 20%, while the presence of Cu^{2+} and Zn^{+2} at the same concentration resulted in significant EstDZ3 inactivation by about 60 and 50%, respectively. When the chelating agent ethylenediaminetetraacetic acid (EDTA) was added to the

TABLE 1 | Kinetic parameters of EstDZ3-mediated hydrolysis against various pNP-esters.

| Substrate (pNP ester) | K_m (mM) | V_{max} ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$) | k_{cat} (s^{-1}) | k_{cat}/K_m ($\text{s}^{-1}\cdot\text{mM}^{-1}$) |
|-----------------------|-----------------|--|-------------------------------|--|
| Acetate (C2) | 0.30 ± 0.07 | 355.3 ± 32.9 | 740 | 2,428 |
| Butyrate (C4) | 0.15 ± 0.02 | 906.4 ± 46.0 | 1,888 | 12,464 |
| Caprylate (C8) | 0.19 ± 0.01 | 500.6 ± 11.6 | 104 | 557 |
| Caprate (C10) | 0.17 ± 0.01 | 386.7 ± 8.2 | 80 | 471 |
| Laurate (C12) | 0.61 ± 0.14 | 357.1 ± 45.1 | 743 | 1,268 |

TABLE 2 | Effect of metal ions, surfactants, and other chemicals on the esterolytic activity of EstDZ3.

| Metal ion or chemical agent | Concentration | Relative activity (%) |
|-----------------------------|---------------|-----------------------|
| None | – | 100.0 ± 4.1 |
| K ⁺ | 1 mM | 93.1 ± 6.7 |
| Mn ²⁺ | 1 mM | 101.4 ± 6.2 |
| Ca ²⁺ | 1 mM | 78.4 ± 3.0 |
| Zn ²⁺ | 1 mM | 45.7 ± 5.7 |
| Li ²⁺ | 1 mM | 90.8 ± 3.7 |
| Mg ²⁺ | 1 mM | 90.1 ± 4.2 |
| Na ⁺ | 1 mM | 90.4 ± 5.5 |
| Fe ³⁺ | 1 mM | 78.3 ± 3.5 |
| Cu ²⁺ | 1 mM | 40.6 ± 3.4 |
| EDTA | 1 mM | 96.7 ± 5.1 |
| PMSF | 1 mM | 33.8 ± 5.8 |
| Triton X-100 | 1% (v/v) | 67.4 ± 1.5 |
| Tween 20 | 1% (v/v) | 70.5 ± 5.4 |
| Tween 80 | 1% (v/v) | 73.2 ± 4.7 |
| SDS | 1% (w/v) | 21.4 ± 2.5 |

reaction mixture at 1 mM, no significant change in the enzyme's activity was observed, thus indicating that the EstDZ3 fold and/or catalytic activity does not depend on a metal co-factor, as observed, for example, in certain cases of esterolytic enzymes that resemble metallo- β -lactamases (Hermoso et al., 2005; Lagartera et al., 2005). The addition of surfactants, such as Triton X-100, Tween 20 and Tween 80 at 1% caused a reduction in the enzyme's activity by approximately 30%, while the addition of SDS at the same concentration caused almost complete inactivation. Finally, addition of the serine hydrolase inhibitor phenylmethylsulfonyl fluoride (PMSF) resulted in a dramatic decrease in EstDZ3 activity, thus indicating that a serine residue is involved in the catalytic mechanism of this new enzyme (Table 2). These levels of tolerance against the presence of metals and detergents are typical for thermostable enzymes (Peng et al., 2011; López et al., 2014).

Finally, EstDZ3 was found to have good tolerance against a variety of organic solvents. In the presence of 10% ethanol, acetone, and acetonitrile, the activity of EstDZ3 was slightly stimulated, whereas methanol addition at 10% had a minor inhibitory effect (Table 3). When either butanol, hexane or isooctane were added at 10%, EstDZ3 retained about half of its maximal catalytic activity, while isopropanol addition at the same concentration caused complete inactivation. When the concentration of methanol, ethanol, acetone, acetonitrile, isooctane, and hexane was raised to 30%, the enzyme exhibited low, but detectable levels of activity, whereas the addition of 1-butanol at the same concentration resulted in almost complete inactivation of the enzyme's esterolytic activity (Table 3).

Homology Analysis and Structural Modeling of EstDZ3

First, the amino acid sequence of EstDZ3 was analyzed with SignalP (Petersen et al., 2011) to detect the possible presence of protein export-signaling sequences. No such sequences were detected, thus indicating that EstDZ3 is not an exported/secreted

TABLE 3 | Effect of organic solvents on the esterolytic activity of EstDZ3.

| Organic solvent | Concentration (v/v) | Relative activity (%) |
|------------------|---------------------|-----------------------|
| Methanol | 10% | 93.0 ± 6.7 |
| | 30% | 26.5 ± 8.3 |
| Ethanol | 10% | 114.1 ± 8.7 |
| | 30% | 23.8 ± 1.5 |
| Acetone | 10% | 109.6 ± 3.3 |
| | 30% | 28.1 ± 7.3 |
| Isopropanol | 10% | 2.8 ± 9.1 |
| | 30% | 50.4 ± 7.9 |
| 1-Butanol | 10% | 8.3 ± 4.6 |
| | 30% | 109.6 ± 9.2 |
| Acetonitrile | 10% | 29.5 ± 5.4 |
| | 30% | 53.6 ± 2.9 |
| Isooctane | 10% | 33.6 ± 7.1 |
| | 30% | 51.2 ± 2.1 |
| <i>n</i> -Hexane | 10% | 25.8 ± 7.1 |
| | 30% | |

enzyme. Then, its sequence was analyzed with BlastP against the Non-Redundant (NR) protein sequences database, the UniProtKB/SwissProt database and the Protein Data Bank (PDB). The BlastP-embedded NCBI conserved protein domain search predicted that EstDZ3 belongs to the α/β hydrolase family 5, while NR analysis revealed that EstDZ3 is identical to a putative *Dictyoglomus thermophilum* α/β hydrolase (Accession no. WP_012548346). Analysis against Uniprot/SwissProt indicated that the closest sequence homolog of EstDZ3, which has been characterized functionally, is an arylesterase from *Pseudomonas fluorescens* (sequence identity 23%, coverage 78%, Accession no. P22862.4) (Choi et al., 1990). The structure of this arylesterase has also been determined via X-ray crystallography (PDB code: 1VA4) (Cheeseman et al., 2004). The other hits from Uniprot/SwissProt included the putative peptidase YtmA from *Bacillus subtilis* subsp. *subtilis* str. 168 (Lapidus et al., 1997) (27% identity, 88% coverage), a dihydropseudooxynicotine hydrolase from *Paenarthrobacter nicotinovorans* (Baitsch et al., 2001) (23% identity, 47% coverage) and other proteins, which were either uncharacterized or with very low query coverage (<17%). On the other hand, a BlastP search against PDB yielded that the closest sequence homolog of EstDZ3, which has been characterized both biochemically and structurally is the cinnamoyl esterase Lj0536 originating from *Lactobacillus johnsonii* (identity 29%, coverage 96%, PDB code: 3PF8) (Lai et al., 2011). The rest of the PDB hits included esterases and peptidases of bacterial and archaeal origin, such as the aforementioned *Pseudomonas fluorescens* aryl esterase and a *Pyrococcus horikoshii* acylaminoacyl peptidase (PDB code: 4HXE) (Menyhárd et al., 2013).

Multiple alignment of the amino acid sequence of EstDZ3 with the top seven sequences of natural proteins of known 3D structure obtained from the PDB BlastP search, indicated that EstDZ3 contains a catalytic triad comprising the residues Ser114, Asp202, and His233 (numbering for EstDZ3), which is absolutely conserved in the sequences of its homologs (Figure 4). Furthermore, the sequence of EstDZ3 also contains the GXSGXG catalytic motif, which is very characteristic for esterolytic

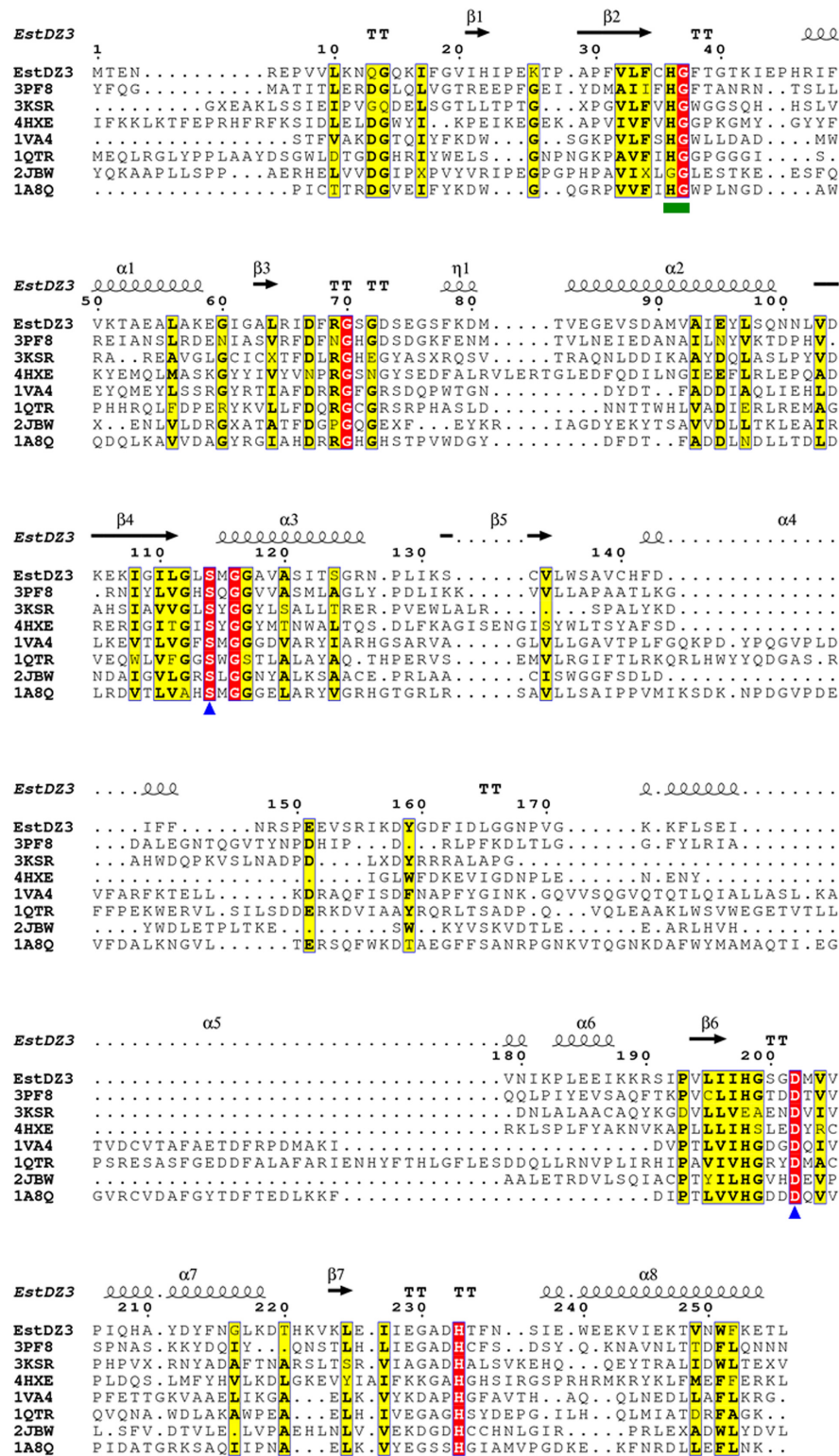


FIGURE 4 | Multiple sequence alignment of EstDZ3 and homologs with known three-dimensional (3D) structure. The absolutely conserved amino acids are highlighted in red and similar ones in yellow. The catalytic residues, Ser114, Asp202, and His233 are indicated by blue triangles. The conserved His36-Gly37 dipeptide, which participates in the formation of the oxyanion hole during ester hydrolysis, is indicated by a green square. Elements of the predicted EstDZ3 secondary structure are denoted as α (α helix), β (β sheet), η (random coil), and T (β turn). Sequence alignment was performed using Clustal Omega (Sievers et al., 2011) and illustrated by ESPript (Robert and Gouet, 2014).

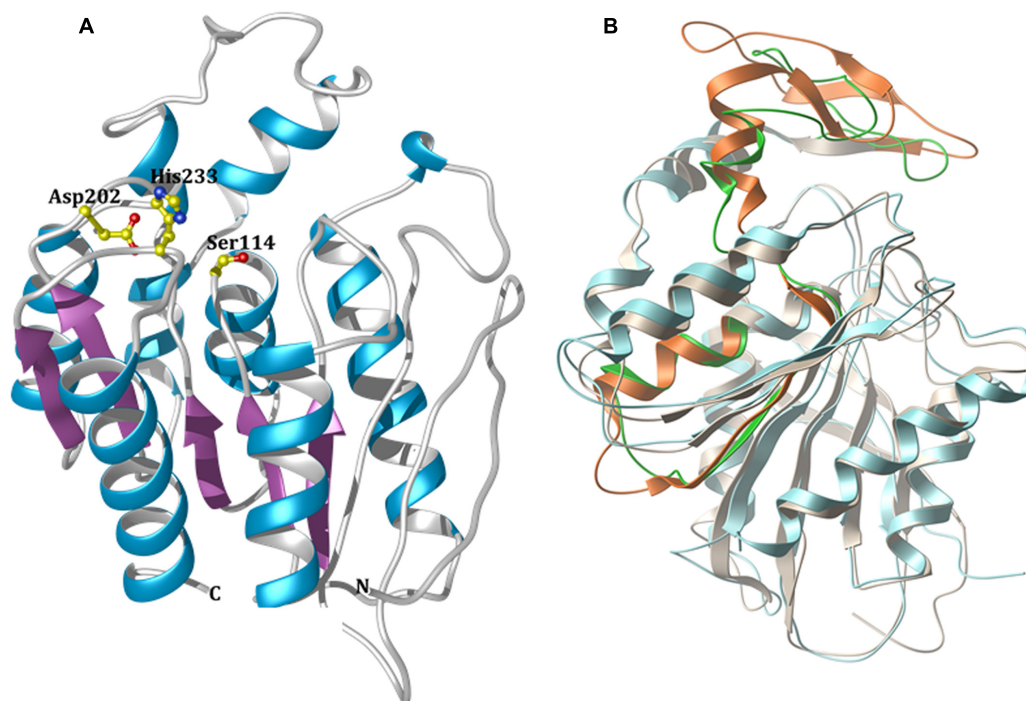


FIGURE 5 | Modeled 3D structure of EstDZ3. (A) The modeled 3D structure of EstDZ3. Residues Ser114, Asp202, and His233, which are predicted to form the catalytic triad, are indicated in ball-and-stick representation. The figure was prepared using *Chimera* (Pettersen et al., 2004). **(B)** Superposition of the predicted EstDZ3 structure (shown in cyan) with that of its closest structural homolog, the cinnamoyl esterase Lj0536 (PDB code: 3PF8, shown in gray). The “insertion subdomain” of Lj0536 is depicted in orange, while the corresponding region in EstDZ3 is indicated in green. The figure was prepared using *MolSoft* (MolSoft LLC, 2000).

enzymes (Bornscheuer, 2002). Finally, the dipeptide His-Gly, which is known to contribute to the formation of the oxyanion hole during ester hydrolysis (Wei et al., 1999; Kim et al., 2013), is also present in the sequence of EstDZ3 (His36-Gly37, EstDZ3 numbering) and conserved within all of the aligned sequences (Figure 4).

Modeling studies to predict the three-dimensional (3D) structure of EstDZ3 were performed using the I-TASSER suite (Yang et al., 2015). I-TASSER applies iterative threading assembly simulations, coupled with secondary structure enhanced Profile-Profile threading alignment and *ab initio* Monte Carlo simulations for unaligned regions. The top-ten threading templates selected by I-TASSER included esterases and peptidases, such as the *P. horikoshii* acylaminoacyl peptidase mentioned above (PDB code: 4HXE) (Menyhárd et al., 2013), the Est1E feruloyl esterase from *Butyrivibrio proteoclasticus* (PDB code: 2WTM) (Goldstone et al., 2010) and an acylaminoacyl peptidase from *Aeropyrum pernix* (PDB code: 2HU8) (Kiss et al., 2007), with sequence identities ranging from 16 to 24% and alignment coverage ranging from 84 to 96%. The presence of acylaminoacyl peptidases among the resulting threading templates is not surprising, since this type of enzymes share common sequence, structural, and functional characteristics with esterolytic enzymes. More specifically, acylaminoacyl peptidases resemble lipolytic enzymes more than classical serine proteases in terms of sequence and structure, as they also carry

the GX SXG motif and adopt an α/β hydrolase fold that includes the catalytic triad Ser-Asp-His in the same sequential order that is encountered in lipases (Polgár, 1992). Furthermore, acylaminoacyl peptidases have been reported in some cases to exhibit esterolytic activity, which may surpass their peptidolytic efficiency (Polgár, 1992; Wang et al., 2006). This catalytic promiscuity has been attributed to the fact that acylaminoacyl peptidases are evolutionarily related to microbial esterases and/or lipases (Polgár, 1992). The modeled 3D structure of EstDZ3 is presented in Figure 5A.

The predicted EstDZ3 structure exhibits a typical α/β hydrolase fold (Figure 5A), which is characteristic for the vast majority of esterolytic enzymes (Bornscheuer, 2002; Brockerhoff, 2012). This provides support for the initial prediction from the NCBI conserved protein domain search that EstDZ3 belongs to the α/β hydrolase family 5. The residues Ser114, Asp202 and His233 are predicted to be located at the catalytic site, with Ser114 at the core of the highly conserved GX SXG catalytic motif (Wei et al., 1999). This is in agreement with the sequence alignment of EstDZ3 and its homologs with known 3D structure (Figure 4). Participation of a serine residue in the catalytic mechanism is additionally supported by the fact that the presence of the serine hydrolase-specific inhibitor PMSF (Smith et al., 1999) resulted in a dramatic reduction of the EstDZ3 esterolytic activity (Table 2).

Superposition of the predicted model structure of EstDZ3 with its closest sequence homolog of known structure and

function, the cinnamoyl esterase Lj0536 (PDB code: 3PF8), using their corresponding secondary structure elements, demonstrated that they share a common architecture (**Figure 5B**). Compared to enzymes of the same family, the structure of Lj0536 is characterized by the “insertion” of an α/β “subdomain,” which has been found to be important for the catalytic profile of this esterolytic enzyme (Lai et al., 2011). This “subdomain insertion” appears to be present also in the predicted structure of EstDZ3. The overall conformation of this “subdomain” appears to be highly similar in the two related enzymes, differing only in the β -sheet region of the α/β “subdomain” of Lj0536 (**Figure 5B**). In the case of Lj0536, it has been found that the conformation of the “subdomain insertion” is an important determinant of the substrate specificity of this enzyme and of its close structural homologs. More specifically, when compared to its closest structural homologs in terms of substrate specificity, Lj0536 resembled mostly the aforementioned esterase Est1E from *B. proteoclasticus* (PDB code: 2WTM), which also contained a mixed α/β “subdomain” with very similar conformation (Lai et al., 2011). On the other hand, the rest of the close structural homologs of Lj0536, which contained all- α -helical “subdomains” with conformations that deviated significantly from that of the corresponding region in Lj0536, exhibited also divergent substrate specificities (Lai et al., 2011). These results suggest strongly that the presence of the “insertion subdomain” in EstDZ3 and the conformation adopted by this region of the protein is expected to be an important determinant of the catalytic properties of this new enzyme. Our computational prediction of the conformation of the “inserted subdomain” in this particular region in the EstDZ3 structure is not of sufficient accuracy – primarily due to the low sequence homology of EstDZ3 with previously studied proteins – to allow the characterization of this domain also as a mixed α/β one as in Lj0536 and Est1E, or not. The experimental determination of the 3D structure of EstDZ3 is expected to provide definitive answers to these questions.

Rationalizing the remarkable thermostability of EstDZ3 is difficult at this point. There are a number of sequence characteristics, which have been found to contribute to increased enzyme resistance against heat-induced destabilization. These include the presence of Tyr and Arg residues at higher frequencies and the presence of Ser and Cys residues at lower ones in thermophilic enzymes compared to their mesophilic counterparts (Kumar et al., 2000; Vieille and Zeikus, 2001). The sequence of EstDZ3, however, is comprised of only 1.4% Tyr and 3.2% Arg, frequencies which are lower than the average number for these amino acids in mesophilic proteins (Kumar et al., 2000). From a structural point of view, the presence of hydrogen bonds and salt bridges in surface-exposed residues and the formation of disulfides are additional factors, which have been shown to be very important contributors to enhanced thermostability in a number of cases (Kumar et al., 2000; Trivedi et al., 2006). Again, the determination of the 3D structure of EstDZ3 via X-ray crystallography, which is currently underway in our laboratories, is expected to provide explanations about the molecular determinants of the remarkable thermostability of EstDZ3.

DISCUSSION

The first hyperthermostable carboxylesterase was isolated from the thermoacidophilic archaeon *Sulfolobus acidocaldarius* and characterized biochemically back in 1988 (Sobek and Görisch, 1988). Since then, more hyperthermostable lipolytic enzymes have been isolated from a small number of hyperthermophiles. Quite surprisingly, very few of these are used nowadays for industrial biotransformations. Most esterases used in the industry are mesophilic, presumably due to the fact that this type of enzymes were the first to be identified and studied more extensively (Levisson et al., 2009). New enzymes with improved properties, which are continuously being discovered via metagenomic and functional genomic approaches, are being introduced into industrial biocatalytic processes only with very low frequencies. In a recent comprehensive review, Ferrer et al. (2016) have attempted to provide explanations about this apparent paradox, and have suggested three main causes: (i) the optimization phase for industrial biocatalysts is both time-consuming and expensive; (ii) the industrial criteria for the selection of appropriate biocatalysts are very strict; and (iii) patent violation restrictions are often encountered. However, as the list of new thermostable/hyperthermostable and overall tolerant lipolytic enzymes is growing, novel biocatalysts that meet the criteria for industrial use are expected to make their way into biotechnological applications. Furthermore, the discovery and characterization of a large number of enzymes with the ability to fold and retain high levels of catalytic activity under extreme conditions will broaden our understanding of their evolutionary occurrence and stabilization mechanisms and will guide future protein engineering efforts.

In this study, we have identified a new hyperthermostable esterolytic enzyme, termed EstDZ3. EstDZ3 originates from a bacterium that belongs to the *Dictyoglomus* genus and exhibits low homology to known proteins, as its closest related enzyme, which has been functionally and structurally characterized, is the cinnamoyl esterase Lj0536 from *L. johnsonii* (identity 29%, coverage 96%, PDB code: 3PF8) (Lai et al., 2011). Biochemical characterization revealed that EstDZ3 exhibits a preference toward esters of fatty acids with short to medium chain lengths, such as pNP-butyrate, indicating that it acts as a carboxylesterase rather than a lipase. Similarly to the vast majority of thermophilic esterases, EstDZ3 functions optimally at a basic pH. At its optimal conditions for ester bond hydrolysis and against its preferred model substrates, EstDZ3 presented high levels of catalytic efficiency ($k_{\text{cat}}/K_m = 12,464 \text{ s}^{-1} \cdot \text{mM}^{-1}$ for pNP-butyrate). Compared to the 20 esterases that have been assayed against pNP-butyrate and deposited in the BRENDA database (Schomburg et al., 2004), the catalytic efficiency of EstDZ3 is among the highest ones. More specifically, 16 out of those 20 esterases exhibited catalytic efficiencies that were one or two orders of magnitude lower than that of EstDZ3. On the contrary, comparison with the rest of the four more active esterases, EstDZ3 was found to exhibit a k_{cat}/K_m value that is only twofold to threefold lower. EstDZ3 preference for short and medium acyl chain length substrates, such as butyric acid-based esters, complemented by its high catalytic

efficiency and excellent thermostability, could be of great value for the dairy product and flavor industries (Saerens et al., 2008).

Many esterolytic enzymes lose their ability to efficiently hydrolyze esters in the presence of organic solvents, a phenomenon occurring primarily due to solvent-induced enzyme denaturation (Klibanov, 2001). On the other hand, thermostable enzymes are often capable of retaining their inherent rigidity not only when exposed to high temperatures but also against other denaturing agents, such as organic solvents (Sayer et al., 2016). EstDZ3 was found to be very stable against exposure to organic solvents, as it was capable of retaining more than 60% of its maximal activity after being exposed to high concentrations of methanol, ethanol, acetone, isopropanol, 1-butanol, acetonitrile, isooctane, and *n*-hexane for 12 h. Compared to the esterase Pf_Est from *Pyrococcus furiosus*, one of the most recently discovered hyperthermostable esterolytic enzymes (Mandelli et al., 2016), EstDZ3 exhibited higher stability when being exposed to methanol, ethanol, and isopropanol (relative residual activity 96, 81, and 74% after 12 h for EstDZ3 versus 39, 51, and 52% after 30 min for Pf_Est, respectively) (Mandelli et al., 2016), while the solvent stability of EstDZ3 resembled more those of the recently discovered organic solvent-tolerant lipase LipXO (Mo et al., 2016).

Importantly, EstDZ3 was found to exhibit remarkable thermostability, as it retained high levels of catalytic activity after exposure to temperatures as high as 95°C for several hours. Comparison with other esterases listed in a previous extensive review of enzymes derived from hyperthermophilic organisms, indicated that EstDZ3 is among the 10 most thermostable ones (Levisson et al., 2009). Only esterases of archaeal origin, such as an esterase/acylpeptide hydrolase from *Aeropyrum pernix* (Gao et al., 2003), esterases EstA and EstB from *Picrophilus torridus* (Hess et al., 2008), and four esterases from the archaeal genera *Pyrococcus* (Cornec et al., 1998; Ikeda and Clark, 1998) and *Sulfolobus* (Huddleston et al., 1995; Park et al., 2008) were reported to exhibit higher thermostability than that of EstDZ3. Among the listed bacterial esterases, EstDZ3 appears to possess the highest catalytic efficiency.

During the recent years, additional hyperthermostable esterolytic enzymes have been discovered. Some characteristic examples are a hyperthermostable lipase from *Bacillus sonorensis* 4R, which exhibits a half-life of about 2 h at 90°C (Bhosale et al., 2016), the xylan-esterase AxeA from *Thermotoga maritima* with a half-life of about 13 h at 98°C (Drzewiecki et al., 2010), and the esterase EstW from the soil bacterium *Streptomyces lividans* TK64 with a half-life of 12 h at 95°C (Wang et al., 2015). The latter esterase, for which kinetic parameters have been determined, exhibits 50-fold lower catalytic efficiency compared to EstDZ3 against pNP-butyrate, and 10-fold lower against pNP-acetate, which is the preferred substrate for EstW (Wang et al., 2015).

The predicted model structure of EstDZ3 has provided preliminary insights on structural features that may be

important for its function. Ongoing structural studies of the new enzyme will shed light into its physiological function and elucidate its role as a potential biocatalyst for industrial biotransformations that require high operation temperatures.

MATERIALS AND METHODS

Reagents and Chemicals

All chemical reagents used in this study were purchased from Sigma-Aldrich. All molecular biology related products (restriction enzymes, protein markers, etc.) were from New England Biolabs unless stated otherwise.

Environmental Sampling and Colony Isolation

A 10-ml glass bottle containing 0.1 g of xanthan gum was filled with water from a hot spring located at the Eryuan region of Yunnan, China (83°C, pH 7), sealed with an anaerobic cap carrying two needles for circulation and immersed into the hot spring. After 10 days, the bottle was collected, sealed immediately and transported to the laboratory for further cultivation. The anaerobically prepared medium contained 1 g xanthan gum (KELTROL® T, food grade, Lot#2F5898K, CP Kelco), 0.13 g (NH₄)₂SO₄, 0.28 g KH₂PO₄ and 0.25 g MgSO₄ as well as trace elements (Na₂MoO₄·2H₂O, 0.025 mg; CaCl₂·2H₂O, 0.01 mg; FeCl₃, 0.28 mg; CuSO₄, 0.016 mg; MnSO₄·H₂O, 2.2 mg; H₃BO₃, 0.5 mg; ZnSO₄·7H₂O, 0.5 mg; CoCl₂·6H₂O, 0.05 mg) and vitamins (biotin 2 mg, folic acid 2 mg, pyridoxine hydrochloride 10 mg, riboflavin 5 mg, thiamine 5 mg, nicotinic acid 5 mg, cobalamin 0.1 mg, *p*-aminobenzoic acid 5 mg, lipoic acid 5 mg) per L of aqueous solution. The *in situ* enrichment was diluted 100-fold in the medium and cultivated anaerobically at 78 and 83°C. Single colonies were obtained by mixing equal volumes of serially diluted cultures and pre-warmed 1% Phytigel (Sigma-Aldrich, cat # P8170), solidification at room temperature, and incubation at 78°C. Visible single colonies were extracted from the solid medium and transferred to liquid cultures.

Expression Library Construction

A single Ch5.6.S clone was incubated in the same medium as mentioned above, except that xanthan gum was replaced by glucose (2 g/L). Cells were harvested, DNA was extracted, and about 20 µg of genomic DNA were digested in a 400-µl reaction containing Bsp143I and Hin1III (0.02 unit/µl each) at 37°C for 30 min. The enzymes were inactivated at 70°C for 10 min and the digested DNA was precipitated and resuspended in TE buffer before gel extraction. Fragments of >2 kb were selected and mixed with pUC18 vector previously digested BamHI and SphI in a 20-µl ligation reaction (250–400 ng genomic DNA fragments, 50 ng vector, 5 units T4 DNA ligase). After overnight ligation at 14°C, 1 µl of the ligation mixture was electroporated into *E. coli* 10G SUPREME cells (Lucigen) according to the manufacturer's

instructions. After 1 h incubation at 37°C, 10 μ l of the cells were plated onto a LB agar plate containing 100 μ g/ml ampicillin to estimate the size of the constructed library, and 1 ml cells were transferred to a 50 ml LB liquid culture containing ampicillin for overnight shaking at 37°C. The cells from the resulting culture were stored in 10% glycerol at –80°C.

Screening of the Expression Library

Samples of the Ch5.6.S expression library transformed into *E. coli* strain NEB10-beta were stored in LB medium containing 20% glycerol at –80°C at a cell density of approximately 2×10^8 cells/ml. For screening, this stock was diluted to 3×10^4 cells/ml in LB medium and plated onto 145 mm round Petri dishes containing screening medium (LB agar containing 100 μ g/ml ampicillin and 0.1% tributyrin) at a density of 10,000 colonies/plate. The plates were incubated at 37°C and the formation of zones of clearance around the colonies was monitored. Colonies that produced clear halos were purified by re-streaking on fresh screening medium. One positive clone was obtained, the corresponding plasmid was isolated, and the insert was sequenced by primer walking using the vector-specific primer M13-RP (5'-CAGGAAACAGCTATGAC-3') and, subsequently, the insert specific primer O-034 (5'-CCGAAGAAGTGTCGAGAA-3').

Plasmid Construction

The coding sequence of EstDZ3 was amplified by polymerase chain reaction using the forward primer Ch2.1_f_52 (5'-GG TTGGGAATTGCAATGACTGAAAATAGAGAACCG-3') and the reverse primer Ch2.1_r_11 (5'-GGAGATGGGAAGTCA TTATAATGTTTCTTTAAACCAATTTACAG-3'). The PCR product was cloned into the pLATE52 vector using the aLICator ligation-independent cloning kit (Thermo Scientific) according to the manufacturer's protocol to form plasmid pLATE52–EstDZ3.

Protein Expression and Purification

BL21(DE3) cells carrying the pLATE52–EstDZ3 plasmid were grown in LB broth containing 100 μ g/ml ampicillin at 37°C under constant shaking until the culture reached an optical density at 600 nm (OD600) of about 0.5. At that point, the expression of *estDZ3* was induced by the addition of 0.2 mM isopropyl- β -D-thiogalactoside (IPTG) followed by overnight incubation at 25°C with shaking. For EstDZ3 purification, the cells from a 500-mL culture grown in a 2-L shake flask were harvested, washed, re-suspended in 10 mL equilibration buffer NPI20 supplemented with 1% Triton X-100, and lysed by brief sonication steps on ice. The cell extract was clarified by centrifugation at $10,000 \times g$ for 15 min at 4°C and the supernatant was incubated for 30 min at 80°C in order to denature other soluble proteins. After this heat-treatment step, the precipitated material was removed by centrifugation at $10,000 \times g$ for 15 min at 4°C. The supernatant was collected and combined with 0.5 mL Ni-NTA agarose beads (Qiagen) and shaken mildly for 1 h at 4°C. The mixture was then loaded onto a 5 mL polypropylene column (Thermo Scientific),

the flow-through was discarded, and the column was washed with 10 mL of NPI20 wash buffer containing 1% Triton X-100. Next, Triton X-100 was washed away by passing 10 mL of standard NPI20 wash buffer. EstDZ3 was eluted using NPI200 elution buffer. All buffers used for purification were prepared according to the manufacturer's protocol. Imidazole was subsequently removed from this protein preparation using a Sephadex G-25 M PD10 column (GE Healthcare). Protein concentration was estimated according to the assay described by Bradford (Bradford, 1976) using bovine serum albumin as a standard. The purified protein was visualized by SDS-PAGE analysis.

Enzyme Activity Assays

For the biochemical characterization of EstDZ3, the catalytic activity of the enzyme was determined by quantification of the amount of pNP released from pNP-ester substrates by photometric measurement at 410 nm. The standard reaction mixture consisted of 25 mM Tris-HCl pH 8 buffer with 0.05% Triton X-100, 2 mM pNP-butyrate and 2 μ g/mL enzyme and was carried out for 5 min at 75°C on a MJ Research thermal cycler, with a pre-incubation setting of the buffer to the target temperature before the enzyme was added. The reactions were terminated by placement on ice and absorbance was measured immediately using a Safire II-Basic plate reader (Tecan). Enzymic activity was recorded by measuring the absorbance of released pNP at 410 nm. All measurements were corrected for non-enzymic hydrolysis of the substrate using control reactions, where no enzyme was added and pNP standard curves were used for the calculation of the enzyme's activity. For the temperature tolerance assay, the buffer was pre-heated and adjusted to pH 8 for each temperature tested. For the substrate specificity experiments, a range of different pNP-fatty acyl esters, such as acetate (C2), butyrate (C4), octanoate (C8), decanoate (C10) and laurate (C12) were used in concentrations ranging from 0.1 to 1 mM. For the initial substrate specificity experiments, clarified lysates of cells producing EstDZ3 were used, while blank reactions were conducted using lysate of cells carrying an empty vector. Data analysis and curve fitting to the Michaelis-Menten equation was performed using the Graphpad Prism 5 software. For the determination of the enzyme's optimal pH, reactions were carried out at 40°C in 25 mM acetate, PIPES, Tris-HCl and glycine buffers for pH values 4–6, 7, 8–9, and 10, respectively. The extinction coefficient of pNP was determined under each reaction condition prior to the measurement. Temperature profiling of EstDZ3 was performed by incubating the standard reaction at temperatures ranging from 40 to 95°C, after the buffer was heated and titrated to the correct pH. Residual activity assays were performed by incubating the enzyme at high temperatures or 50% solvent concentration and subsequently measuring its activity into the standard reaction. Maximal (100%) enzyme activity corresponds to the activity of an enzyme sample that was not exposed to any of the tested denaturing conditions. In the case of solvent stability experiment, the incubation medium was vigorously agitated during the 12 h incubation time, and subsequently

it was diluted to remove the solvent before assaying the enzyme. The assays for the determination of EstDZ3 tolerance in the presence of metal ions, detergents and organic solvents were also executed in the standard reaction with the only difference being the addition of the agents at the specified concentrations. Blanks for this experiment consisted of the same reaction mix, including the tested agent, but without the addition of enzyme. All measurements were obtained from at least three independent experiments carried out in triplicates.

Homology Analysis and Structural Modeling Studies of EstDZ3

The EstDZ3 sequence was submitted to a similarity search analysis using BLASTp (Altschul et al., 1990) against the NR, Uniprot/SwissProt and PDB databases, and the embedded NCBI's conserved domain search (Marchler-Bauer et al., 2014). The results obtained from the PDB search (including natural enzymes and excluding engineered ones) were aligned using Clustal Omega (Sievers et al., 2011) and illustrated with ESPrpt (Robert and Gouet, 2014). Modeling of the 3D structure of EstDZ3 (residues 1 to 256) was performed with I-TASSER (Yang et al., 2015). Out of the top-five predictive models prepared, the first one with C-score 0.3 and TM-score 0.75 ± 0.10 was selected. Superposition of the modeled structure with the closest structural homolog was performed by molecular graphics software COOT (Emsley et al., 2015) using the secondary structure elements. Molecular visualization of the modeled structure was performed with Chimera (Pettersen et al., 2004).

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Aygan, A., Arikan, B., Korkmaz, H., Dinçer, S., and Çolak, Ö. (2008). Highly thermostable and alkaline α -amylase from a halotolerant-alkaliphilic *Bacillus* sp. AB68. *Braz. J. Microbiol.* 39, 547–553.
- Baitsch, D., Sandu, C., Brandsch, R., and Igloi, G. L. (2001). Gene cluster on pAO1 of *Arthrobacter nicotinovorans* involved in degradation of the plant alkaloid nicotine: cloning, purification, and characterization of 2,6-dihydroxypyridine 3-hydroxylase. *J. Bacteriol.* 183, 5262–5267. doi: 10.1128/JB.183.18.5262-5267.2001
- Bhosale, H., Shaheen, U., and Kadam, T. (2016). Characterization of a hyperthermostable alkaline lipase from *Bacillus sonorensis* 4R. *Enzyme Res.* 2016:4170684. doi: 10.1155/2016/4170684
- Bornscheuer, U., Huisman, G., Kazlauskas, R., Lutz, S., Moore, J., and Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature* 485, 185–194. doi: 10.1038/nature11117
- Bornscheuer, U. T. (2002). Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol. Rev.* 26, 73–81. doi: 10.1111/j.1574-6976.2002.tb00599.x
- Bornscheuer, U. T., and Kazlauskas, R. J. (2006). *Hydrolases in Organic Synthesis: Regio- and Stereoselective Biotransformations*. Hoboken, NJ: John Wiley & Sons.
- Bornscheuer, U. T., and Pohl, M. (2001). Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* 5, 137–143. doi: 10.1016/S1367-5931(00)00182-4

ACCESSION NUMBERS

The *estDZ3* nucleotide sequence and the *ch2* insert sequence have been deposited in GenBank under accession codes KX557297 and KX557298, respectively.

AUTHOR CONTRIBUTIONS

DZ, ZS, FK, and GS designed the project; DZ, ZS, HP, XP, FK, and GS designed the research; DZ, ZS, DM, HP, and EC performed the research; DZ, ZS, DM, HP, EC, XP, CI, FK, and GS analyzed the data; XP, CI, FK, and GS supervised the research; DZ, ZS, and GS wrote the paper with contributions from XP and EC. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

This work was carried out in the framework of the HotZyme Project (<http://hotzyme.com>, grant agreement no. 265933) financed by the European Union Seventh Framework Programme FP7/2007-2013, a collaborative programme whose aim was the use of genomic and metagenomic approaches to identify new thermostable hydrolases from diverse hot environments with improved performances and/or novel functionalities for industrial biotransformations. DZ was also supported by a Ph.D. fellowship from the Greek State Scholarships Foundation (Idryma Kratikon Ypotofion-IKY) in the framework of the Excellence IKY-Siemens Program, which is co-financed by the European Social Fund and the Greek Government.

- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3
- Brocknerhoff, H. (2012). *Lipolytic Enzymes*. Amsterdam: Elsevier.
- Cheeseman, J. D., Tocilj, A., Park, S., Schrag, J. D., and Kazlauskas, R. J. (2004). Structure of an aryl esterase from *Pseudomonas fluorescens*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 1237–1243. doi: 10.1107/S0907444904010522
- Choi, K. D., Jeohn, G. H., Rhee, J. S., and Yoo, O. J. (1990). Cloning and nucleotide sequence of an esterase gene from *Pseudomonas fluorescens* and expression of the gene in *Escherichia coli*. *Agric. Biol. Chem.* 54, 2039–2045. doi: 10.1080/00021369.1990.10870247
- Cornec, L., Robineau, J., Rolland, J. L., Dietrich, J., and Barbier, G. (1998). Thermostable esterases screened on hyperthermophilic archaeal and bacterial strains isolated from deep-sea hydrothermal vents: characterization of esterase activity of a hyperthermophilic archaeum, *Pyrococcus abyssi*. *J. Mar. Biotechnol.* 6, 104–110.
- Dalby, P. A. (2011). Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* 21, 473–480. doi: 10.1016/j.sbi.2011.05.003
- Drzewiecki, K., Angelov, A., Ballschmiter, M., Tiefenbach, K. J., Sterner, R., and Liebl, W. (2010). Hyperthermostable acetyl xylan esterase. *Microb. Biotechnol.* 3, 84–92. doi: 10.1111/j.1751-7915.2009.00150.x
- Emsley, P., Brunger, A. T., and Lütteke, T. (2015). Tools to assist determination and validation of carbohydrate 3D structure data. *Glycoinformatics* 1273, 229–240.
- Ewis, H. E., Abdelal, A. T., and Lu, C.-D. (2004). Molecular cloning and characterization of two thermostable carboxyl esterases from *Geobacillus stearothermophilus*. *Gene* 329, 187–195. doi: 10.1016/j.gene.2003.12.029

- Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol.* 9, 22–34. doi: 10.1111/1751-7915.12309
- Gao, R., Feng, Y., Ishikawa, K., Ishida, H., Ando, S., Kosugi, Y., et al. (2003). Cloning, purification and properties of a hyperthermophilic esterase from archaeon *Aeropyrum pernix* K1. *J. Mol. Catal. B: Enzymatic* 24, 1–8. doi: 10.1016/S1381-1177(03)00064-X
- Goldstone, D. C., Villas-Bôas, S. G., Till, M., Kelly, W. J., Attwood, G. T., and Arcus, V. L. (2010). Structural and functional characterization of a promiscuous feruloyl esterase (Est1E) from the rumen bacterium *Butyrivibrio proteoclasticus*. *Proteins* 78, 1457–1469.
- Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., et al. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat. Commun.* 2, 375. doi: 10.1038/ncomms1373
- Hermoso, J. A., Lagartera, L., González, A., Stelter, M., García, P., Martínez-Ripoll, M., et al. (2005). Insights into pneumococcal pathogenesis from the crystal structure of the modular teichoic acid phosphorylcholine esterase Pce. *Nat. Struct. Mol. Biol.* 12, 533–538. doi: 10.1038/nsmb940
- Hess, M., Katzer, M., and Antranikian, G. (2008). Extremely thermostable esterases from the thermoacidophilic euryarchaeon *Picrophilus torridus*. *Extremophiles* 12, 351–364. doi: 10.1007/s00792-008-0139-9
- Hough, D. W., and Danson, M. J. (1999). Extremozymes. *Curr. Opin. Chem. Biol.* 3, 39–46. doi: 10.1016/S1367-5931(99)80008-8
- Huddleston, S., Yallop, C., and Charalambous, B. (1995). The identification and partial characterization of a novel inducible extracellular thermostable esterase from the archaeon *Sulfolobus shibatae*. *Biochem. Biophys. Res. Commun.* 216, 495–500. doi: 10.1006/bbrc.1995.2650
- Ikeda, M., and Clark, D. S. (1998). Molecular cloning of extremely thermostable esterase gene from hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli*. *Biotechnol. Bioeng.* 57, 624–629. doi: 10.1002/(SICI)1097-0290(19980305)57:5<624::AID-BIT15>3.0.CO;2-B
- Kim, Y.-O., Park, I.-S., Kim, H.-K., Nam, B.-H., Kong, H. J., Kim, W.-J., et al. (2013). Shewanella sp. Ke75 esterase with specificity for p-nitrophenyl butyrate: gene cloning and characterization. *J. Korean Soc. Appl. Biol. Chem.* 56, 55–62.
- Kirk, O., Borchert, T. V., and Fuglsang, C. C. (2002). Industrial enzyme applications. *Curr. Opin. Biotechnol.* 13, 345–351. doi: 10.1016/S0958-1669(02)00328-2
- Kiss, A. L., Hornung, B., Rádi, K., Gengeliczi, Z., Sztáray, B., Juhász, T., et al. (2007). The acylaminoacyl peptidase from *Aeropyrum pernix* K1 thought to be an exopeptidase displays endopeptidase activity. *J. Mol. Biol.* 368, 509–520. doi: 10.1016/j.jmb.2007.02.025
- Klibanov, A. M. (2001). Improving enzymes by using them in organic solvents. *Nature* 409, 241–246. doi: 10.1038/35051719
- Kumar, S., Tsai, C.-J., and Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Eng.* 13, 179–191. doi: 10.1093/protein/13.3.179
- Lagartera, L., González, A., Hermoso, J. A., Saiz, J. L., García, P., García, J. L., et al. (2005). Pneumococcal phosphorylcholine esterase, Pce, contains a metal binuclear center that is essential for substrate binding and catalysis. *Protein Sci.* 14, 3013–3024. doi: 10.1110/ps.051575005
- Lai, K.-K., Stogios, P. J., Vu, C., Xu, X., Cui, H., Molloy, S., et al. (2011). An inserted α/β subdomain shapes the catalytic pocket of *Lactobacillus johnsonii* cinnamoyl esterase. *PLoS ONE* 6:e23269. doi: 10.1371/journal.pone.0023269
- Lapidus, A., Galleron, N., Sorokin, A., and Ehrlich, S. D. (1997). Sequencing and functional annotation of the *Bacillus subtilis* genes in the 200 kb *rrnB-dnaB* region. *Microbiology* 143, 3431–3441. doi: 10.1099/00221287-143-11-3431
- Lawrence, R. C., Fryer, T. F., and Reiter, B. (1967). Rapid method for the quantitative estimation of microbial lipases. *Nature* 213, 1264–1265. doi: 10.1038/2131264a0
- Lee, Y.-L., Su, M.-S., Huang, T.-H., and Shaw, J.-F. (1999). C-terminal His-tagging results in substrate specificity changes of the thioesterase I from *Escherichia coli*. *J. Am. Oil Chem. Soc.* 76, 1113–1118. doi: 10.1007/s11746-999-0082-7
- Levisson, M., van der Oost, J., and Kengen, S. W. (2007). Characterization and structural modeling of a new type of thermostable esterase from *Thermotoga maritima*. *FEBS J.* 274, 2832–2842. doi: 10.1111/j.1742-4658.2007.05817.x
- Levisson, M., van der Oost, J., and Kengen, S. W. (2009). Carboxylic ester hydrolases from hyperthermophiles. *Extremophiles* 13, 567–581. doi: 10.1007/s00792-009-0260-4
- López, G., Chow, J., Bongen, P., Pietruszka, J., Streit, W. R., et al. (2014). A novel thermoalkalostable esterase from *Acidicaldus* sp. strain USBA-GBX-499 with enantioselectivity isolated from an acidic hot springs of Colombian Andes. *App. Microbiol. Biotechnol.* 98, 8603–8616. doi: 10.1007/s00253-014-5775-7
- Lorenz, P., Liebeton, K., Niehaus, F., and Eck, J. (2002). Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr. Opin. Biotechnol.* 13, 572–577. doi: 10.1016/S0958-1669(02)00345-2
- Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* 21, 734–743. doi: 10.1016/j.copbio.2010.08.011
- Mandelli, F., Gonçalves, T., Gandin, C., Oliveira, A., Neto, M. O., and Squina, F. (2016). Characterization and Low-resolution structure of an extremely thermostable esterase of potential biotechnological interest from *Pyrococcus furiosus*. *Mol. Biotechnol.* 1273, 229–240.
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2014). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 1. doi: 10.1093/nar/gku1221
- Matthews, B., Nicholson, H., and Becktel, W. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Nat. Acad. Sci. U.S.A.* 84, 6663–6667. doi: 10.1073/pnas.84.19.6663
- Menyhárd, D. K., Kiss-Szemán, A., Tichy-Rács, É., Hornung, B., Rádi, K., Szeltnér, Z., et al. (2013). A Self-compartmentalizing Hexamer Serine Protease from *Pyrococcus horikoshii* substrate selection achieved through multimerization. *J. Biol. Chem.* 288, 17884–17894. doi: 10.1074/jbc.M113.451534
- Menzel, P., Gudbergdóttir, S. R., Rike, A. G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* 70, 411–424. doi: 10.1007/s00248-015-0576-9
- Mo, Q., Liu, A., Guo, H., Zhang, Y., and Li, M. (2016). A novel thermostable and organic solvent-tolerant lipase from *Xanthomonas oryzae* pv. *oryzae* YB103: screening, purification and characterization. *Extremophiles* 20, 157–165. doi: 10.1007/s00792-016-0809-y
- MolSoft LLC (2000). *2.8 Program Manual*. San Diego, CA: MolSoft LLC.
- Novototskaya-Vlasova, K., Petrovskaya, L., Yakimov, S., and Gilichinsky, D. (2012). Cloning, purification, and characterization of a cold-adapted esterase produced by *Psychrobacter cryohalolentis* K5T from Siberian cryopeg. *FEMS Microbiol. Ecol.* 82, 367–375. doi: 10.1111/j.1574-6941.2012.01385.x
- Park, Y.-J., Yoon, S.-J., and Lee, H.-B. (2008). A novel thermostable arylesterase from the archaeon *Sulfolobus solfataricus* P1: purification, characterization, and expression. *J. Bacteriol.* 190, 8086–8095. doi: 10.1128/JB.00803-08
- Peng, Q., Zhang, X., Shang, M., Wang, X., Wang, G., Li, B., et al. (2011). A novel esterase gene cloned from a metagenomic library from neritic sediments of the South China Sea. *Microb. Cell Fact.* 10:95. doi: 10.1186/1475-2859-10-95
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comp. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Polgár, L. (1992). Structural relationship between lipases and peptidases of the prolyl oligopeptidase family. *FEBS Lett.* 311, 281–284. doi: 10.1016/0014-5793(92)81120-B
- Quax, W., and Broekhuizen, C. (1994). Development of a new *Bacillus* carboxyl esterase for use in the resolution of chiral drugs. *Appl. Microbiol. Biotechnol.* 41, 425–431. doi: 10.1007/BF00939031
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new END script server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316
- Saerens, K., Descamps, D., and Dewettinck, K. (2008). Release of short chain fatty acids from cream lipids by commercial lipases and esterases. *Biotechnol. Lett.* 30, 311–315. doi: 10.1007/s10529-007-9541-0

- Sayer, C., Finnigan, W., Isupov, M. N., Levisson, M., Kengen, S. W., van der Oost, J., et al. (2016). Structural and biochemical characterisation of *Archaeoglobus fulgidus* esterase reveals a bound CoA molecule in the vicinity of the active site. *Sci. Rep.* 6:25542. doi: 10.1038/srep25542
- Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., et al. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* 32(Suppl. 1), D431–D433. doi: 10.1093/nar/gkh081
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi: 10.1038/msb.2011.75
- Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M., and Baker, E. N. (1999). Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of Bacillus Ak. 1 protease at 1.8 Å resolution. *J. Mol. Biol.* 294, 1027–1040. doi: 10.1006/jmbi.1999.3291
- Sobek, H., and Görisch, H. (1988). Purification and characterization of a heat-stable esterase from the thermoacidophilic archaeobacterium *Sulfolobus acidocaldarius*. *Biochem. J.* 250, 453–458. doi: 10.1042/bj2500453
- Tirawongsaroj, P., Sriprang, R., Harnpicharnchai, P., Thongaram, T., Champreda, V., Tanapongpipat, S., et al. (2008). Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *J. Biotechnol.* 133, 42–49. doi: 10.1016/j.jbiotec.2007.08.046
- Trivedi, S., Gehlot, H., and Rao, S. (2006). Protein thermostability in Archaea and Eubacteria. *Genet. Mol. Res.* 5, 816–827.
- Vieille, C., and Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microb. Mol. Biol. Rev.* 65, 1–43. doi: 10.1128/MMBR.65.1.1-43.2001
- Wang, B., Wang, A., Cao, Z., and Zhu, G. (2015). Characterization of a novel highly thermostable esterase from the gram-positive soil bacterium *Streptomyces lividans* TK64. *Biotechnol. Appl. Biochem.* 63, 305–343.
- Wang, Q., Yang, G., Liu, Y., and Feng, Y. (2006). Discrimination of esterase and peptidase activities of acylaminoacyl peptidase from hyperthermophilic *Aeropyrum pernix* K1 by a single mutation. *J. Biol. Chem.* 281, 18618–18625. doi: 10.1074/jbc.M601015200
- Wei, Y., Contreras, J. A., Sheffield, P., Osterlund, T., Derewenda, U., Kneusel, R., et al. (1999). Crystal structure of brefeldin A esterase, a bacterial homolog of the mammalian hormone-sensitive lipase. *Nat. Struct. Mol. Biol.* 6, 340–345. doi: 10.1038/7576
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* 12, 7–8. doi: 10.1038/nmeth.3213
- Zamost, B. L., Nielsen, H. K., and Starnes, R. L. (1991). Thermostable enzymes for industrial applications. *J. Ind. Microbiol.* 8, 71–81. doi: 10.1007/BF01578757
- Zarafeta, D., Kissas, D., Sayer, C., Gudbergsdottir, S. R., Ladoukakis, E., Isupov, M. N., et al. (2016). Discovery and characterization of a thermostable and highly halotolerant GH5 cellulase from an icelandic hot spring isolate. *PLoS ONE* 11:e0146454. doi: 10.1371/journal.pone.0146454

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zarafeta, Szabo, Moschidi, Phan, Chrysina, Peng, Ingham, Kolis and Skretas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Dark Side of the Mushroom Spring Microbial Mat: Life in the Shadow of Chlorophototrophs. I. Microbial Diversity Based on 16S rRNA Gene Amplicons and Metagenomic Sequencing

Vera Thiel^{1*†}, Jason M. Wood², Millie T. Olsen², Marcus Tank^{1†}, Christian G. Klatt^{2,3}, David M. Ward² and Donald A. Bryant^{1,4*}

OPEN ACCESS

Edited by:

Anna-Louise Reysenbach,
Portland State University, USA

Reviewed by:

Charles K. Lee,
University of Waikato, New Zealand
Anirban Chakraborty,
University of Calgary, Canada
Wesley Douglas Swingley,
Northern Illinois University, USA

*Correspondence:

Vera Thiel
vthiel@tmu.ac.jp;
Donald A. Bryant
dab14@psu.edu

†Present Address:

Vera Thiel and Marcus Tank,
Department of Biological Sciences,
Tokyo Metropolitan University,
Hachioji, Japan

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 07 March 2016

Accepted: 27 May 2016

Published: 17 June 2016

Citation:

Thiel V, Wood JM, Olsen MT, Tank M,
Klatt CG, Ward DM and Bryant DA
(2016) The Dark Side of the
Mushroom Spring Microbial Mat: Life
in the Shadow of Chlorophototrophs.
I. Microbial Diversity Based on 16S
rRNA Gene Amplicons and
Metagenomic Sequencing.
Front. Microbiol. 7:919.
doi: 10.3389/fmicb.2016.00919

¹ Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA,

² Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT, USA, ³ Agricultural Research Service, United States Department of Agriculture, University of Minnesota, Saint Paul, MN, USA, ⁴ Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT, USA

Microbial-mat communities in the effluent channels of Octopus and Mushroom Springs within the Lower Geyser Basin at Yellowstone National Park have been studied for nearly 50 years. The emphasis has mostly focused on the chlorophototrophic bacterial organisms of the phyla *Cyanobacteria* and *Chloroflexi*. In contrast, the diversity and metabolic functions of the heterotrophic community in the microoxic/anoxic region of the mat are not well understood. In this study we analyzed the orange-colored undermat of the microbial community of Mushroom Spring using metagenomic and rRNA-amplicon (iTag) analyses. Our analyses disclosed a highly diverse community exhibiting a high degree of unevenness, strongly dominated by a single taxon, the filamentous anoxygenic phototroph, *Roseiflexus* spp. The second most abundant organisms belonged to the *Thermotogae*, which have been hypothesized to be a major source of H₂ from fermentation that could enable photomixotrophic metabolism by *Chloroflexus* and *Roseiflexus* spp. Other abundant organisms include two members of the *Armatimonadetes* (OP10); *Thermocrinis* sp.; and phototrophic and heterotrophic members of the *Chloroflexi*. Further, an *Atribacteria* (OP9/JS1) member; a sulfate-reducing *Thermodesulfovibrio* sp.; a *Planctomycetes* member; a member of the EM3 group tentatively affiliated with the *Thermotogae*, as well as a putative member of the *Arminicenantetes* (OP8) represented $\geq 1\%$ of the reads. Archaea were not abundant in the iTag analysis, and no metagenomic bin representing an archaeon was identified. A high microdiversity of 16S rRNA gene sequences was identified for the dominant taxon, *Roseiflexus* spp. Previous studies demonstrated that highly similar *Synechococcus* variants in the upper layer of the mats represent ecological species populations with specific ecological adaptations. This study suggests that similar putative ecotypes specifically adapted to different niches occur within the undermat community, particularly for *Roseiflexus* spp.

Keywords: hot spring, microbial community, microbial diversity, extreme environments, phototrophic bacteria

INTRODUCTION

Microbial mat communities inhabiting the effluent channels of Octopus and Mushroom Springs within the Lower Geyser Basin at Yellowstone National Park (YNP) have been studied for nearly 50 years (Brock, 1967; Ward et al., 2012). In these studies, the chlorophototrophic bacterial populations, i.e., chlorophyll-based phototrophs including members of the *Cyanobacteria*, *Chloroflexi* and the newly discovered *Chloracidobacterium* (*Cab.*) *thermophilum* and “*Candidatus Thermochlorobacter* (Tcb.) *aerophilum*,” have generally been the main focus (Bauld and Brock, 1973; Nold and Ward, 1996; Bryant et al., 2007; van der Meer et al., 2007; Steunou et al., 2008; Becraft et al., 2011; Klatt et al., 2011, 2013b; Liu et al., 2011, 2012; Tank and Bryant, 2015a,b). In contrast, the diversity and metabolic functions of the heterotrophic community in the microoxic/anoxic region of the mat are not well understood.

Using cultivation-based methods, early studies focused on the dominant *Cyanobacteria* and phototrophic *Chloroflexi* (Bauld and Brock, 1973; Bateson and Ward, 1988). Over time, these studies were extended by a variety of molecular methods with increasing molecular resolution. A pioneering molecular study targeting 16S rRNA gene sequences directly indicated a greater diversity of uncultivated bacteria in the mat than previously realized (Ward et al., 1990). However, only recently have metagenomic (Klatt et al., 2011), metatranscriptomic (Liu et al., 2011, 2012; Klatt et al., 2013b) and metatranscriptomic (Kim et al., 2015) analyses led to a holistic overview, in terms of the organisms present and their functional potentials, of the major taxa inhabiting the upper 2 mm of the 60–65°C regions of the Mushroom Spring microbial mats (**Figure 1**). The microbial community of the upper green mat layer contains eight dominant bacterial populations, of which six are chlorophototrophs (Klatt et al., 2011). Oxygenic cyanobacteria from the genus *Synechococcus* have been shown to be the predominant primary producers in these communities by *in situ* studies of bicarbonate fixation and nitrogen fixation (Steunou et al., 2008) using stable and radioactive isotopes (Bateson and Ward, 1988; Nübel et al., 2002; van der Meer et al., 2007). In addition, anoxygenic photoheterotrophic members of the *Roseiflexus* spp. have been suggested to perform inorganic carbon fixation (van der Meer et al., 2003, 2005, 2007, 2010; Klatt et al., 2007, 2013b). *Synechococcus* spp. fix CO₂ and synthesize and excrete metabolites that are then consumed by (photo)heterotrophic members of the community, including several members of the *Chloroflexi*, and presumably *Roseiflexus* spp. (Anderson et al., 1987; Bateson and Ward, 1988; Kim et al., 2015). Collectively, cyanobacteria and *Roseiflexus* spp. account for the majority of the biomass of the upper 0–2 mm portion of the mat community. Two additional members of the phylum *Chloroflexi*, *Chloroflexus* sp. and an apparently phototrophic, “*Anaerolineae*-like” organism (“*Ca. Roseilinea gracile*”; Tank et al., in press), as well as two recently discovered aerobic/microaerophilic, anoxygenic photoheterotrophs, *Cab. thermophilum* (Bryant et al., 2007; Garcia Costas et al., 2012a,b; Tank and Bryant, 2015a,b) and “*Ca. Tcb. aerophilum*” (Liu et al., 2012), also occur in the upper photic layer of the mat.



FIGURE 1 | Sampling site at Mushroom Spring, Yellowstone National Park, and microbial mat core (adapted from Kim et al., 2015).

Early studies on the processes and organisms involved in aerobic and anaerobic decomposition of the mat have been discussed in a review by Ward et al. (1992; and earlier papers cited therein). Since the discovery of the aerobic heterotroph *Thermus aquaticus* (Brock and Freeze, 1969) many aerobic (e.g., *Thermomicrobium roseum*; Jackson et al., 1973) and anaerobic fermentative and sulfate-reducing bacteria were cultivated from these mats (e.g., *Bacillus stearothermophilis*, *Thermoanaerobium brockii*, *Thermoanaerobacter ethanolicus*, *Thermodesulfotobacterium commune*; see Ward et al., 1992 for primary references). Many of the latter were sought with the hope that thermophiles would be useful for biofuel production. However, critical review indicated that most of these isolates had not been cultivated from highly diluted mat samples, and thus their importance to the community remained unknown (see Ward et al., 1998). Indeed, with one exception, *Thermomicrobium roseum* (Wu et al., 2009), the genomes of these organisms did not recruit reads with high identity values from metagenomic analysis of the upper mat layer (Klatt et al., 2011). Only two low-abundance, unidentified heterotrophic bacteria lacking the genes needed to synthesize chlorophyll (Chl) were detected in the upper mat community represented by metagenomic bins (Klatt et al., 2011). Nevertheless, heterotrophs, together with the photoheterotrophic and photomixotrophic community members, can be considered potential consumers of metabolites produced by cyanobacteria and possibly other mat inhabitants. In more recent years, the activity and diversity of sulfate-reducing bacteria of the microbial mats have been more intensively studied. Dillon et al. (2007) showed that an active sulfur cycle occurs in the mat community despite very low sulfate concentrations. The highest rates of sulfate respiration were reportedly associated with *Thermodesulfovibrio*-like organisms and were measured close to the surface of the mat late in the day when photosynthetic oxygen production had ceased. Additionally, methane production has been detected in numerous alkaline siliceous hot spring microbial mats in YNP

(Ward, 1978; Sandbeck and Ward, 1981, 1982). Methanogenic archaea ($\sim 10^7$ to 10^8 ml $^{-1}$) have been enumerated in small cores of Octopus Spring mats, which in combination with the detection of low levels of archaeal lipids, suggests that methanogenesis occurred *in situ* in those mats (Ward, 1978; Sandbeck and Ward, 1981; Ward et al., 1985). The relative rarity of these organisms compared to *Synechococcus* (on the order of 1% or less) suggests that these terminal anaerobes receive little of the energy recycled during decomposition of the mat (Ward et al., 1989).

The first revolution of molecular microbial ecology enabled the study of uncultured bacterial diversity through amplification, sequencing and phylogenetic analysis of ribosomal RNA genes (Olsen et al., 1986; Ward et al., 1990; Amann et al., 1995; Hugenholtz and Pace, 1996; Hugenholtz et al., 1998a; Pace, 2009). Through such studies, our perspective on microbial diversity has increased enormously over the past three decades, and the impact of culture-independent studies on the emerging view of bacterial diversity cannot be overstated (Hugenholtz et al., 1998a). Ward and coworkers reported the presence of a number of uncultured bacterial lineages in their first molecular microbial diversity study of the mat community of Octopus Spring (Ward et al., 1990). Over the course of the past 25 years, several of those initially unidentified ribosomal RNA sequences have been associated with chlorophototrophic mat members (OS-A and B with *Synechococcus* spp., OS-C with *Roseiflexus* sp., OS-D with *Cab. thermophilum*, and OS-E with “*Ca. Tcb. aerophilum*”), whereas many others (OS-F, OS-G, OS-H, OS-K, OS-L, OS-M, OS-N, OS-R) still have not been identified and were not detected in the metagenome of the upper green layer (Klatt et al., 2011).

“Red-layer” communities, which may often be “orange” in color as is the case for the mats of Mushroom Spring, have been shown to contain novel chlorophototrophs (Boomer et al., 2000, 2002), whose pigments exhibit unusual *in vivo* absorption spectra (Boomer et al., 2000), but these communities have not yet been studied in detail. As part of a comparative study of YNP hot spring microbial mat communities, a 45-Mbp metagenome based on Sanger sequencing revealed some initial insights into the composition of the undermat microbial community of Mushroom Spring (Klatt et al., 2013a). Compared to the upper green layer, fewer *Synechococcus* spp., a greater number of *Roseiflexus* spp., and several presumed anaerobic or fermentative organisms within the *Bacteroidetes* and *Thermodesulfobacteria* were identified. The undermat community contained a *Thermotoga*-like population as well as several low G+C organisms that could not be characterized (Klatt et al., 2013a). Low coverage and a small number of long scaffolds above the threshold used in most clustering analyses (>10 kb) limited the application of metagenomic binning approaches (Klatt et al., 2013a) and indicated that additional studies with much deeper sequencing would be needed to define the undermat community.

The overall goal of this research is to investigate the complete microbial mat community at Mushroom Spring and to develop a comprehensive understanding of the microbial ecology of the microbial mats of this hot spring. The specific objectives of this study were to analyze the orange-colored undermat community, to identify those organisms that are present, and

to facilitate an active integration of these mostly heterotrophic members into models of the mat community. This paper describes the composition and diversity of the Mushroom Spring undermat community based on rRNA-amplicon (iTag) and deep metagenomic sequencing analyses, with an initial focus on the identity and taxonomic diversity of the community members. A description of the metabolic potential and putative interactions, including a metabolic description of the entire microbial mat community, will be published separately.

MATERIALS AND METHODS

The samples were collected on August 10th, 2011 from a chlorophototrophic microbial mat in an effluent channel of the siliceous and slightly alkaline Mushroom Spring in YNP, WY (USA). The samples were collected using a #4 cork borer at a site where the water above the mat was 60°C (Figure 1). The microbial mat is made up of an upper green layer (1–2 mm thick), which mainly consists of different chlorophototrophic bacteria, and an orange-colored undermat layer (Figure 1). Genomic DNA was extracted from the orange-colored undermat layer (~ 3 –5 mm depth; DNA from below this level was too degraded to analyze). The metagenome as well as 16S rRNA gene PCR amplicons were sequenced at the DOE Joint Genome Institute (JGI) using HiSeq and MiSeq Illumina technologies. The iTag sequences were analyzed at two different identity levels. All reads were clustered into operational taxonomic units (OTUs) with 97% sequence identity cutoff by using USEARCH, but they were also analyzed after dereplication (i.e., clustered by 100% nt identity, see Supplementary Materials). RDP Classifier (Wang et al., 2007; Cole et al., 2009), BLAST searches (Altschul et al., 1990) and phylogenetic analyses (Ludwig et al., 2004) were used to identify sequences. Microdiversity was assessed using the number of highly abundant dereplicated sequences, and the “oligotyping pipeline” (<http://merenlab.org/projects/oligotyping/>). HiSeq metagenomic reads were assembled and then clustered into bins by oligonucleotide frequency pattern analyses using ESOM (Dick et al., 2009). Metagenomic bins were treated as partial genomes of single taxa and were taxonomically affiliated using Amphoranet (<http://pitgroup.org/amphoranet/>; Kerepesi et al., 2014) to assess the phylogenetic marker genes present in each bin. Detailed descriptions of the methods for DNA extraction, library construction, sequencing, and data analyses are found in the Supplementary Materials.

RESULTS

We used deep sequencing of rRNA gene amplicons (iTags) and total environmental DNA to study the subsurface community of the chlorophototrophic microbial mat at Mushroom Spring. We describe the diversity and community composition on both levels, based on “OTUs” (Figures 2A, 3A, Table 1 and Table S1) and based on “dereplicated iTag” sequences (Figures 2B, 3B, Table 2) in Section “16S rRNA Gene Amplicons (iTags),” as well as on metagenomic bins obtained based on

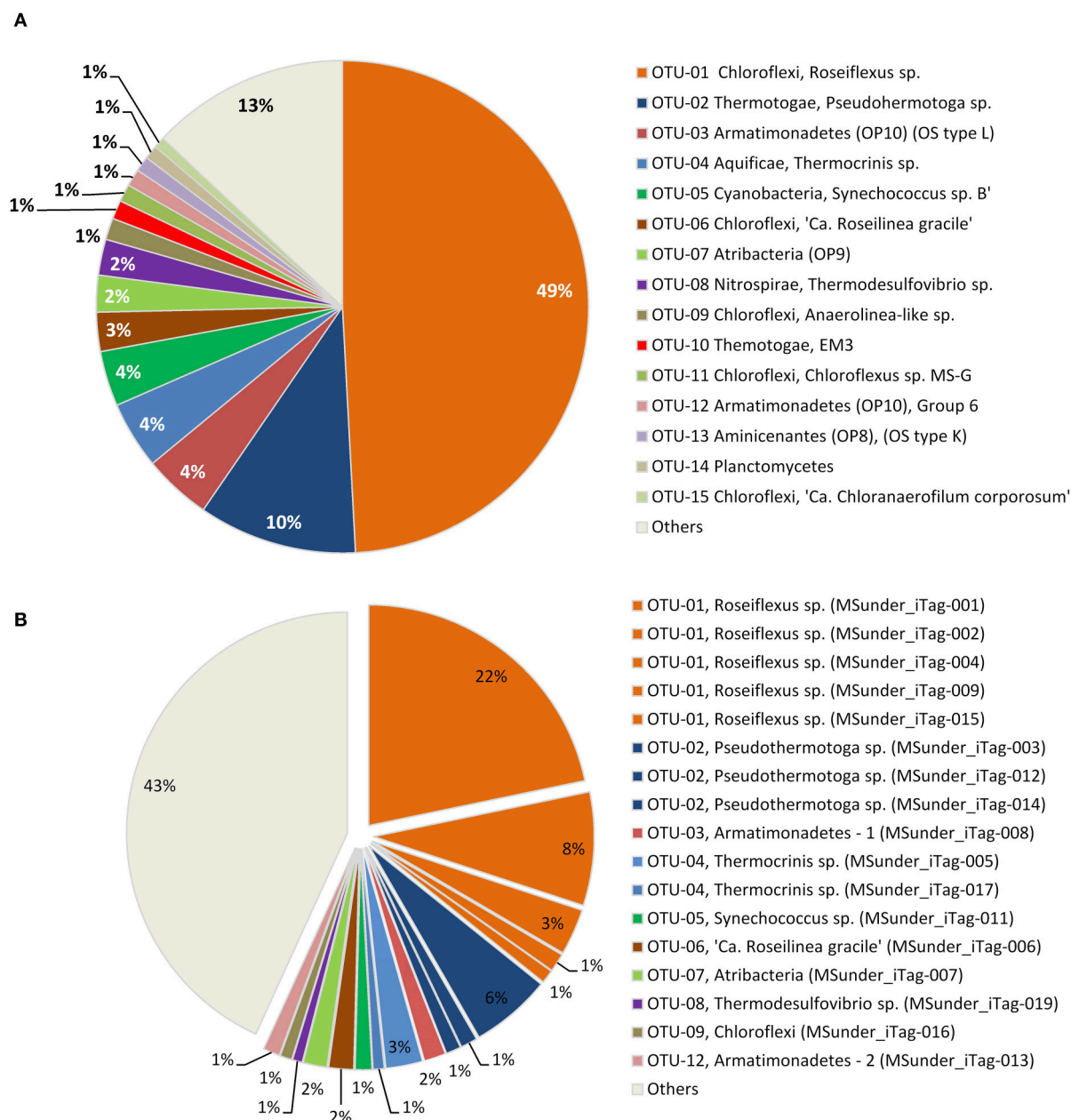
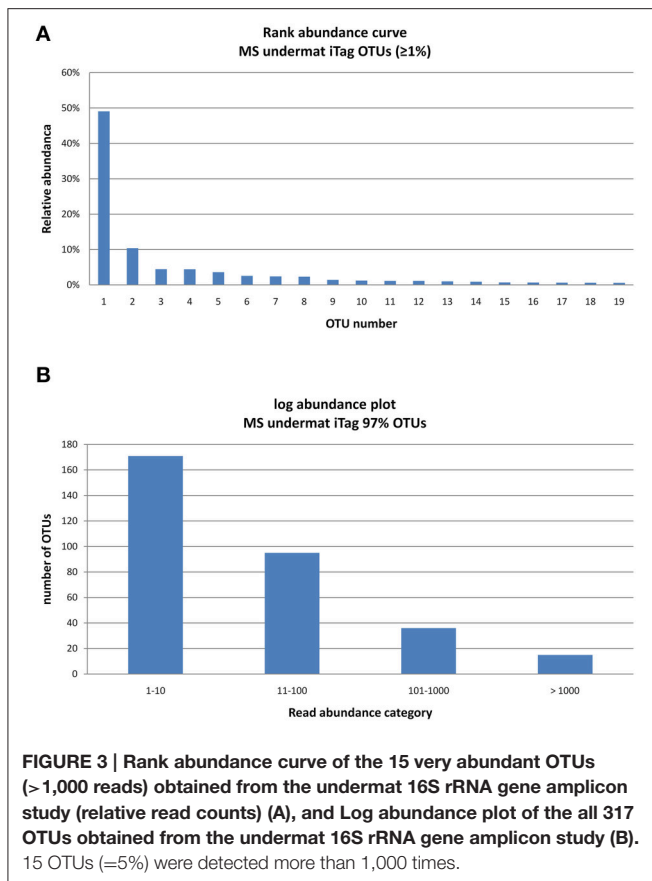


FIGURE 2 | Relative abundance of (A) the 15 most abundant 97% OTUs, and (B) the 17 most abundant dereplicated iTag sequences in the Mushroom Spring undermat 16S rRNA gene amplicon (iTag) analysis. All less abundant OTUs (<1,000 reads each) are shown combined as "Others."

oligonucleotide frequency patterns in Section "Metagenome Sequencing" (Figure 4, Table 3). An overview of the most important taxa detected in each phylum will be presented in Section "Overview of Phyla and Taxa Detected in the Mushroom Spring Undermat." Each iTag OTU was found to represent a variable number of dereplicated iTag sequences, which is interpreted as representing different degrees of microdiversity within a taxon (Figure 2B, Table 1). Members of 20 different phyla were identified (Figure 5 and Figure S1, Table S1). Organisms of the phylum *Chloroflexi* dominated the microbial

undermat community in both read abundance and diversity (Tables 1, 2, and Table S1, Figures 2A,B). Thirteen out of seventeen members of the microbial mat detected in previous 16S rRNA gene sequence cloning and DGGE studies (OS types, Table 4; Ward et al., 1990, 1992; Weller et al., 1992; Ferris et al., 1996b, 1997; Ferris and Ward, 1997), as well as relatives of ribosomal sequence types derived from a previous undermat study (Klatt et al., 2013a, Figure 5 and Figure S1) were detected in this study and thus confirmed as members of a compositionally and temporally stable microbial community.



16S rRNA Gene Amplicons (iTags)

Sequencing of partial 16S rRNA genes resulted in 139,326 total and 30,861 dereplicated (i.e., unique) reads after quality control. Abundance values of dereplicated reads varied between 1 and 30,285, with an average of 5.4 reads per sequence.

Diversity Based on OTUs

The 16S rRNA gene amplicon reads clustered into 317 OTUs of $\geq 97\%$ nt identity, with abundances between 1 and 68,369 reads per OTU (Table S1). The community was characterized by a low degree of evenness (Figure 3A). The majority of the OTUs were present in low abundance; only 15 OTUs (5% of the taxa) were represented by 1,000 or more reads (Figure 3B). Due to the high number of singleton sequences, the estimated richness based on Chao1 ($S_{\text{Chao1}} = S_{\text{obs}} + (\text{no. of singletons}^2) / (2 * \text{no. of doubletons})$) (Chao, 1984) was rather high, Chao1 = 369.74; a lower value of Chao1 = 220.9 was obtained in a previous study (Klatt et al., 2013a). In contrast, the Simpson's Reciprocal Index ($D = \frac{\sum n(n-1)}{N(N-1)}$) obtained in this study is considerably lower than in previous studies (3.85 in this study vs. 37.5; Klatt et al., 2013a), reflecting the low evenness and strong dominance of only a few OTUs in the amplicon study. While an identity cut-off of 97% for rRNA gene sequences is often used to demarcate species (Stackebrandt and Goebel, 1994; Schloss and Handelsman, 2005; Koepfel and Wu, 2013), this is an arbitrary value that does

not necessarily correlate with any species definition. Here, we refer to OTUs as “taxa,” use the term “populations” mainly for dereplicated iTag sequences, and discuss our understanding of the bacterial species concept in Section “Discussion.”

Most Abundant Taxa Based on OTUs

When considering OTU sequences based on 97% nt sequence identity, 15 OTUs were identified with $>1,000$ reads each, varying in abundance between 1,008 and 68,369 reads (Table 1). These are considered to represent highly abundant taxa and thus are likely to represent key members of the Mushroom Spring undermat community. However, the threshold of 1,000 reads was arbitrarily chosen and does not necessarily correlate with activity or ecological importance. We will focus the discussion on the “very abundant” taxa listed in Table 1, but will also include selected “abundant” and “less abundant” OTUs with read abundances of ≥ 100 and less, respectively (Table S1).

The 16S rRNA gene amplicons of the microbial undermat community were dominated by sequences derived from *Roseiflexus* spp. (Figure 2A, OTU-1, 49%) with the second most abundant sequences belonging to a *Pseudothermotoga* sp. (OTU-2, 10%). An unidentified *Armatimonadetes* (formerly known as OP10) bacterium (OTU-3), a member of the *Aquificae* (OTU-4), as well as the sequences derived from member of the *Cyanobacteria* each represented $\sim 4\%$ of the sequences (Table 1). On the basis of *psaA* sequences the cyanobacterial sequences can be classified as belonging to ecotype populations of *Synechococcus* detected in the upper green layer of the mat and are considered likely to arise from buried surface populations that are not expected to represent metabolically active constituents of the undermat community. The sixth most abundant OTU was identified as a phototrophic member of the phylum *Chloroflexi*, which had previously been detected in the upper green layer using metagenome analysis and identified as the first phototrophic “*Anaerolineae*-like” *Chloroflexi*; it has provisionally been named “*Ca. Roseilinea gracile*” (Klatt et al., 2011, 2013b; Tank et al., in press). Additional abundant OTUs were affiliated with the *Atribacteria* (OP9) *Nitrospirae*, *Planctomycetes* and several phototrophic and non-phototrophic members of the phylum *Chloroflexi* (Table 1). Three of the fifteen most abundant OTU sequences from the undermat amplicon study represented sequences obtained from the mats of Octopus Spring in previous 16S rRNA gene surveys (OS-B: *Synechococcus* sp. Type B; OS-C: *Roseiflexus* sp. RS-1; and OS-L: *Armatimonadetes* member OTU-3) (Table 4, Ward et al., 1990, 1992; Ferris et al., 1996a; van der Meer et al., 2010).

Most Abundant Populations Based on Dereplicated iTag Sequences

Seventeen dereplicated iTag sequences, representing members of the nine most abundant OTUs, were each detected more than 1,000 times, and in total represent more than half of all iTag reads recovered in this study (Table 2, Figure 2B). These sequences probably correspond to the most abundant “populations” (in contrast to “taxa” for OTUs). Five of these very abundant dereplicated iTag sequences belong to a single OTU representing the most abundant taxon, *Roseiflexus* spp. (Figure 2B, Table 2

TABLE 1 | Most abundant OTUs (97% nt identity), number of reads and relative abundance, microdiversity in terms of represented dereplicated iTag sequences, corresponding metagenome sequences and next relatives determined by BLAST search.

| OTU-name | Reads | Rel. abundance (%) | No. of derep. iTags | | Metagenome 16S rRNA | | | Relevant Blast hits | Acc. No. | Ident (%) |
|----------|--------|--------------------|---------------------|------|---------------------|---------------|-------------|---|-----------|-----------|
| | | | Total | >100 | Scaffold ID | Avg. coverage | Length (bp) | | | |
| OTU-01 | 68,369 | 49 | 6,193 | 24 | 101,6681 | 2,842 | 1,357 | <i>Roseiflexus</i> sp. RS-1 | NR_074197 | 100 |
| OTU-02 | 14,480 | 10 | 1,664 | 3 | 1,000,336 | 458 | 1,426 | <i>Thermotoga hypogaea</i> NBRC 106472 (T) | AP014508 | 98 |
| OTU-03 | 6,203 | 4 | 2,082 | 9 | 1,062,246 | 951 | 710 | Unc. bacterium clone SM2D03 | AF445720 | 100 |
| OTU-04 | 6,160 | 4 | 899 | 3 | 1,002,657 | 262 | 1,434 | <i>Eubacterium</i> sp. (OS type L) | L04707 | 98 |
| | | | | | | | | clone YNP_SBC_BP2A_B2 | HM448202 | 99 |
| | | | | | | | | <i>Thermocrinis ruber</i> strain DSM 23557(T) | CP007028 | 97 |
| OTU-05 | 5,046 | 4 | 1,289 | 7 | 1,008,881 | 105 | 1,002 | <i>Synechococcus</i> sp. JA-2-3B'a(2-13) | CP000240 | 100 |
| OTU-06 | 3,580 | 3 | 675 | 1 | 1,001,696 | 423 | 1,364 | clone YNP_SBC_BP4_B2 | HM448255 | 98 |
| | | | | | | | | <i>Thermanaerothermox daxisensis</i> GNS-1 (T) | HM596746 | 87 |
| OTU-07 | 3,350 | 2 | 564 | 1 | 1,014,288 | 189 | 1,444 | OP9 bacterium clone TP29 | EF205555 | 98 |
| | | | | | | | | <i>Ammonifex thiophilus</i> strain SR(T) | EF554597 | 83 |
| OTU-08 | 3,283 | 2 | 705 | 2 | 1,000,748 | 135 | 1,413 | clone SMD-B01 | AB477993 | 99 |
| | | | | | | | | <i>Thermodesulfobrio yellowstonii</i> DSM11347(T) | NR_074345 | 96 |
| OTU-09 | 1,981 | 1 | 449 | 2 | 1,000,273 | 129 | 1,389 | clone NY-30 | KC290430 | 94 |
| | | | | | | | | <i>Thermanaerothermox daxisensis</i> GNS-1 (T) | HM596746 | 91 |
| OTU-10 | 1,715 | 1 | 510 | 6 | 1,001,962 | 67 | 1,406 | EM3 clone OPB88 | AF027006 | 99 |
| | | | | | | | | <i>Rhodothermus marinus</i> SG0.5JP17-172 (T) | CP003029 | 82 |
| OTU-11 | 1,594 | 1 | 521 | 3 | 1,030,146 | 83 | 1,309 | <i>Chloroflexus</i> sp. MS-G | KR230107 | 99 |
| | | | | | | | | <i>Chloroflexus aurantiacus</i> J-10-fl (T) | CP000909 | 95 |
| OTU-12 | 1,569 | 1 | 477 | 3 | 1,003,586 | 802 | 1,415 | OP10 clone OPB80 | AF027089 | 94 |
| | | | | | | | | <i>Fimbrimonas ginsengisoli</i> Gsoil 348 (T) | CP007139 | 81 |
| OTU-13 | 1,392 | 1 | 306 | 2 | 1,015,572 | 36 | 1,497 | clone bac67 | HM184963 | 95 |
| | | | | | | | | <i>Thermoanaerobaculum aquaticum</i> MP-01 (T) | NR_109681 | 88 |
| OTU-14 | 1,260 | 1 | 322 | 3 | 1,003,293 | 60 | 1,420 | clone TP5 | EF205581 | 99 |
| | | | | | | | | <i>Thermogutta terrifontis</i> R1 (T) | KC867694 | 90 |
| OTU-15 | 1,008 | 1 | 391 | 3 | 1,021,867 | 132 | 1,299 | clone OB17 | EF429491 | 98 |
| | | | | | | | | <i>Chloroflexus aurantiacus</i> J-10-fl(T) | CP000909 | 90 |

OTU-1, MSunder_iTags-1, 2, 4, 9, and 15). The third most abundant sequence (MSunder_iTag-3), as well as two additional abundant, dereplicated sequences (MSunder_iTags-12 and 14), were representatives of *Pseudothermotoga* spp. (OTU-2), the second most abundant taxon. The *Armatimonadetes* (OTU-3) and a member of the phylum *Aquificae* (OTU-4) contained two slightly different, highly abundant dereplicated iTag sequences each, whereas the other OTUs (OTUs 5–9) had only one very abundant dereplicated iTag sequence. With regard to the single dereplicated iTag sequences, cyanobacteria derived from the green upper layer of the mat community are represented

by the eleventh most abundant iTag sequence, and thus the ten most abundant dereplicated iTag sequences (representing eight OTUs) are considered to represent the most abundant populations in the undermat community (MSunder_iTag-1 through MSunder_iTag-10; **Table 2**).

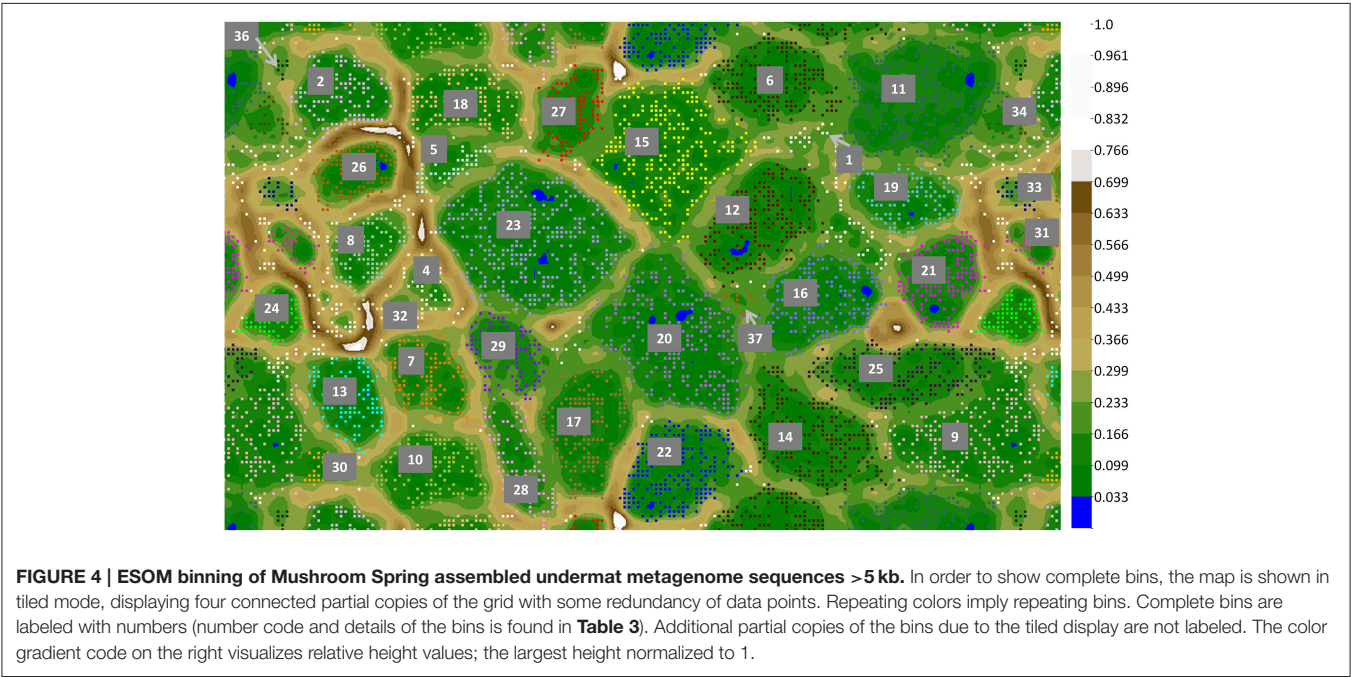
Microdiversity

We used different methods to assess the degree of sequence heterogeneity and microdiversity within the microbial undermat community. Based on the number of different dereplicated iTag sequences within one 97% OTU, a high degree of diversity was

TABLE 2 | Most abundant dereplicated iTag sequences (100% nucleotide identity) detected in the Mushroom Spring undermat.

| MSunder_iTag (dereplicated iTag) | Reads | Relative abundance (%) | Phylum | Genus | OTU |
|----------------------------------|--------|------------------------|-----------------|---------------------|-----|
| MSunder_iTag-1 | 30,285 | 21.70 | Chloroflexi | Roseiflexus | 1 |
| MSunder_iTag-2 | 11,586 | 8.30 | Chloroflexi | Roseiflexus | 1 |
| MSunder_iTag-3 | 8,257 | 5.90 | Thermotogae | Pseudothermotoga | 2 |
| MSunder_iTag-4 | 4,712 | 3.40 | Chloroflexi | Roseiflexus | 1 |
| MSunder_iTag-5 | 3,760 | 2.70 | Aquificae | Thermocrinis | 4 |
| MSunder_iTag-6 | 2,551 | 1.80 | Chloroflexi | Ca. Roseilinea | 6 |
| MSunder_iTag-7 | 2,436 | 1.70 | Atribacteria | | 7 |
| MSunder_iTag-8 | 2,229 | 1.60 | Armatimonadetes | | 3 |
| MSunder_iTag-9 | 1,881 | 1.40 | Chloroflexi | Roseiflexus | 1 |
| MSunder_iTag-10 | 1,721 | 1.20 | Nitrospirae | Thermodesulfobivrio | 8 |
| MSunder_iTag-11 | 1,716 | 1.20 | Cyanobacteria | Synechococcus | 5 |
| MSunder_iTag-12 | 1,695 | 1.20 | Thermotogae | Pseudothermotoga | 2 |
| MSunder_iTag-13 | 1,542 | 1.10 | Armatimonadetes | | 3 |
| MSunder_iTag-14 | 1,370 | 1.00 | Thermotogae | Pseudothermotoga | 2 |
| MSunder_iTag-15 | 1,220 | 0.90 | Chloroflexi | Roseiflexus | 1 |
| MSunder_iTag-16 | 1,161 | 0.80 | Chloroflexi | | 9 |
| MSunder_iTag-17 | 1,027 | 0.70 | Aquificae | Thermocrinis | 4 |

Read numbers, relative abundance, taxonomic affiliation and OTU affiliation are provided.



indicated, especially for the most abundant OTU, *Roseiflexus* spp. We detected 6,193 total dereplicated iTag sequences, 24 of which had >100 reads (Table 1). A similar microdiversity was identified by the oligotyping approach, and was also suggested by a high number of very similar but non-identical clone sequences obtained in a previous study (Klatt et al., 2013a; Figure 5A, and Figure S2, Table S2). Based on ten distinct nucleotide positions, 246 different oligotypes were identified, of which 55 were represented by >10 reads, 23 by >100 reads and nine by > 1,000 reads in the combined dataset (which consisted

of ~39,000 upper green layer reads and 75,000 undermat reads). The total “purity scores” of 0.95 and 0.86 for >100 and >10 reads, respectively, indicates a good separation for the highly abundant oligotypes, but also implies further low abundance oligotypes in the samples. Differences in diversity and abundance of oligotypes between the upper green layer and the undermat were detected, e.g., for the most abundant *Roseiflexus*-oligotypes (Table S2, Figure S2). In general the undermat is more diverse. The upper green layer for example contains a lower number of highly abundant oligotypes (six oligotypes >1% of

TABLE 3 | Metagenome bins recovered based on tetranucleotide frequencies.

| Bin* | OTU | Identity | 16S rRNA | Contigs | Size [Mb] | Marker genes |
|----------------|--------|---|--------------------|---------|-----------|--------------|
| 1 | OTU-01 | <i>Roseiflexus</i> sp. | No | 18 | 0.20 | None |
| 2 | OTU-02 | <i>Thermotoga</i> sp. | Yes, scaffold00336 | 142 | 1.85 | 30 |
| 3 ^S | OTU-03 | <i>Armatimonadetes</i> (OP10) | Yes | 81 | 2.9 | 32 |
| 4 | OTU-04 | <i>Aquificae</i> / <i>Thermocrinis</i> sp. | Yes, scaffold02657 | 44 | 0.22 | 5 |
| 5 | OTU-05 | <i>Synechococcus</i> sp. B' | No | 68 | 0.43 | 13 |
| 6 | OTU-06 | " <i>Ca. Roseilinea gracile</i> " | No | 188 | 1.71 | 25 |
| 7 | OTU-07 | <i>Atribacteria</i> (OP9) | No | 136 | 1.45 | 26 |
| 8 | OTU-08 | <i>Thermodesulfobivrio</i> sp. | Yes, scaffold00748 | 119 | 1.59 | 17 |
| 9 | OTU-09 | <i>Chloroflexi</i> , <i>Anerolineae</i> | Yes, scaffold00273 | 220 | 2.62 | 26 |
| 10 | OTU-10 | <i>Thermotogae</i> -EM3/OPS-2 | Yes, scaffold01962 | 157 | 1.38 | 18 |
| 11 | OTU-11 | <i>Chloroflexus</i> sp. MS-G | No | 336 | 3.1 | 21 |
| 12 | OTU-12 | <i>Armatimonadetes</i> _Gp6 (OP10) | Yes, scaffold03586 | 200 | 1.72 | 24 |
| 13 | OTU-13 | <i>Aminicenantes</i> (OP8) | No | 94 | 2.54 | 30 |
| 14 | OTU-14 | <i>Planctomycetes</i> | Yes, scaffold03293 | 258 | 1.90 | 19 |
| 15 | OTU-15 | " <i>Ca. Chloranaerofilum corporosum</i> " | No | 299 | 2.26 | 19 |
| 16 | OTU-17 | <i>Chloracidobacterium thermophilum</i> | No | 220 | 2.07 | 21 |
| 17 | OTU-18 | <i>Armatimonadetes</i> _Gp 2 | Yes, scaffold00584 | 207 | 2.64 | 31 |
| 18 | OTU-21 | <i>Meiothermus</i> sp. | No | 155 | 1.31 | 8 |
| 19 | OTU-24 | <i>Chlorobi</i> -lineage 5: OPB56 | Yes, scaffold02638 | 137 | 1.17 | 18 |
| 20 | OTU-36 | <i>Acidobacterium</i> , OPB3 | Yes, scaffold01343 | 253 | 2.60 | 6 |
| 21 | OTU-38 | " <i>Ca. Thermochlorobacter aerophilum</i> " | No | 198 | 2.09 | 27 |
| 22 | OTU-46 | <i>Elioraea</i> sp. | No | 229 | 2.02 | 26 |
| 23 | | <i>Planctomycetes</i> | No | 385 | 3.8 | 29 |
| 24 | | <i>Ignavibacteriaceae</i> | No | 75 | 2.50 | 31 |
| 25 | OTU-31 | <i>Chloroflexi</i> , <i>Bellilinea</i> sp. | No | 237 | 2.46 | 24 |
| 26 | | Unidentified | No | 130 | 1.78 | 28 |
| 27 | | Unidentified | No | 96 | 0.85 | 2 |
| 28 | | Unidentified | No | 110 | 0.84 | 3 |
| 29 | | Unidentified | No | 98 | 0.66 | 5 |
| 30 | | Unidentified | No | 22 | 0.27 | None |
| 31 | | Unidentified | No | 38 | 0.26 | 2 |
| 32 | OTU-26 | <i>Thermodesulfobacteria</i> | No | 27 | 0.18 | 12 |
| 33 | | Unidentified | No | 23 | 0.17 | none |
| 34 | | Unidentified | No | 23 | 0.17 | none |
| 35 | | Unidentified | No | 17 | 0.11 | 8 |
| 36 | | Unidentified | No | 12 | 0.09 | none |
| 37 | | <i>Chloroflexi</i> , <i>Dehalococcoides</i> -like | No | 8 | 0.06 | none |

*Numbers refer to ESOM bins shown in **Figure 4**. ^S Bin 3 was obtained from an enrichment culture, not from the undermat metagenome.

all *Roseiflexus* sequences), whereas the undermat is more diverse with nine oligotypes >1% (**Table S2**, **Figure S2**). Notably, the most abundant oligotypes are present in both samples in similar abundances. One oligotype dominates both datasets (48% in the upper layer vs. 54% in the undermat). The second most abundant oligotype "CTCTACGGGC" is more abundant in the upper layer (32 vs. 20% of the reads), whereas the third is more abundant in the undermat (9 vs. 6%, **Table S2**). In general the undermat is more diverse and some oligotypes show distinct differences. For example, the difference in the entropy figures from upper green layer and undermat after separate analyses (considerably lower entropy at pos. 104 and 109 in upper layer; **Figures S2B,C**) are indicative of a lower abundance of two oligotypes in the

upper layer, namely "CCCCGCGTGC" (2.13% in undermat, 0.19% in upper layer) and "CCCCGCGGGC" (1.02 vs. 0.21%) (**Table S2**).

A high degree of microdiversity was also indicated for other OTUs obtained in this study, e.g., OTU-3 (*Armatimonadetes* member, OS type L) and OTU-5 (*Synechococcus* spp.) Overall, the twelve most abundant OTUs also exhibited the highest number of unique amplicon sequences, indicating a correlation between microdiversity and sequencing depth (**Table 1**). However, the number of abundant dereplicated sequences, i.e., putative ecotypes did not show the self-correlation with sequencing depth, but correlated with the metagenome assembly quality; a high microdiversity was suggested to be interfering with the sequence

TABLE 4 | OS-type sequences from previous studies (Ward et al., 1990, 1992; Weller et al., 1992; Ferris et al., 1996b, 1997; Ferris and Ward, 1997) and the corresponding sequences obtained in this study.

| OS type | Acc. Nos. | Phylum | Identity | OTU | Metagenome sequences [JGI24185J35167_] |
|---------|----------------------|----------------------------|--|---------|---|
| A | X52544 | Cyanobacteria | <i>Synechococcus</i> sp. A | OTU-22 | 10825551 (280 bp) 10154254 (272 bp) 10858251 (546 bp) |
| B | M62776 X52545 | Cyanobacteria | <i>Synechococcus</i> sp. B' | OTU-5 | 12370211 (272 bp) |
| C | X52546 U42421 U90433 | <i>Chloroflexi</i> | <i>Roseiflexus</i> sp. RS-1 | OTU-1 | 10166812 (1357 bp) |
| D | X52547 | <i>Acidobacteria</i> | <i>Chloracidobacterium thermophilum</i> | OTU-17 | 10032584 (1383 bp) |
| E | X52548 U42419 U42420 | <i>Chlorobi</i> | <i>Ca. Thermochlorobacter aerophilum</i> | OTU-38 | 10007615 (1379 bp) |
| F | X52549 | <i>Chlorobi</i> | <i>OPB56-like Chlorobi</i> | OTU-262 | No scaffold |
| G | X52550 | <i>Betaproteobacteria</i> | <i>Hydrogenophilus</i> sp. | OTU-101 | No scaffold |
| H | X52551 | <i>Spirichaeetes</i> | <i>Spirochaeta</i> sp. | OTU-41 | No scaffold |
| I | L04709 | Cyanobacteria | <i>Leptolyngbya</i> sp. | no OTU | No scaffold |
| J | L04710 | Cyanobacteria | <i>Synechococcus</i> sp. | no OTU | 1045912 (920 bp) |
| K | L04711 | <i>Acidobacteria</i> | <i>Thermoanaerobaculum</i> sp. | OTU-61 | 10096124 (954 bp) 10037885 (587 bp) |
| L | L04707 | <i>Armatimonadetes</i> | Uncultured <i>Armatimonadetes</i> (OP10) | OTU-3 | 11621781 (295 bp) 10622461 (710 bp) 12054061 (298 bp) |
| M | L04708 | <i>Chlorobi</i> | Unidentified <i>Ignavibacteriaceae</i> | OTU-34 | 1068906 (652 bp) 1010292 (1038 bp) |
| N | L05931 L05930 | <i>Betaproteobacteria</i> | <i>Tepidimonas</i> sp. | No OTU | No scaffold |
| O | L04706 | <i>Alphaproteobacteria</i> | <i>Elioraea</i> sp. | OTU-46 | 10309321 (1157 bp) 10459122 (309 bp) |
| Q | U42422 | <i>Chloroflexi</i> | <i>Candidatus Chlorothrix halophila</i> | No OTU | No scaffold |
| R | U46750 | <i>Betaproteobacteria</i> | Uncultured <i>Beta-Proteobacterium</i> | OTU-172 | No scaffold |

assembly. Very few contigs with >5 kb length were assembled for the OTUs with the highest microdiversity (OTU-1 and OTU-3).

Metagenome Sequencing

One full lane of Illumina HiSeq sequencing led to 176,741,874 quality-passed reads. 169,595,919 (96%) of these reads were assembled into a 232-Mb metagenome comprising 315,154 total contigs with a maximum scaffold length of 158 kb and a N/L50 value of 32,529/1.24 kb, which defines the number of fragments at or above the Length50 cutoff. There were 13,766 contigs >2.5 kb, 5,362 contigs >5 kb, and 1,665 >10 kb. Contigs >50 kb ($n = 38$) accounted for 1.14% of all assembled sequence data.

Metagenome Bins

Binning of the metagenome contigs based on tetranucleotide frequency patterns resulted in 36 clusters (Table 3, Figure 4). An additional bin, representing OTU-3 from the iTag study of the undermat, was obtained from an cyanobacterial enrichment culture metagenome (Olsen et al., 2015). Thus, 37 partial genomes, 26 of which contained ≥ 1 Mb of sequence information, were found by this method (Table 3). Twenty-six of the bins were identified taxonomically, and 22 could be affiliated with abundant OTUs. A specific cut-off with regard to taxonomic levels or sequence threshold cannot be given for the represented populations. However, previous studies, as well as joint binning of the sequences from the presented study with reference genomes, suggest that genomes derived from bacterial populations with 16S rRNA gene sequences identities of $\geq 96\%$ do

not separate into distinct bins (data not shown; Klatt et al., 2011). In this study, the cyanobacterial genomes of *Synechococcus* Types A and B' (97% 16S rRNA nt identity), and within the *Chloroflexi*, *Roseiflexus castenholzii* and *Roseiflexus* sp. RS-1 (95.6% 16S rRNA nt identity) as well as *Chloroflexus aurantiacus* J-10-fl and *Chloroflexus* sp. MS-G (95.7% 16S rRNA nt identity) genomes clustered in single bins containing sequences of both genomes, respectively. All other included *Chloroflexi* reference genomes (<94% 16S rRNA nt identity) clustered in separate but sometimes adjacent bins. The occurrence of several metagenomic bins affiliated with the *Chloroflexi* as well as the separate clustering of the included *Chloroflexi* reference genomes, provides an estimate of the ability of this approach to discriminate and resolve among different members of the same phylum. Based on these observations, as well as 16S rRNA OTU similarities found in this study displaying values of either <95% or >96.8% nt identity, we expect genomes of populations sharing <95% 16S rRNA sequence identity to be represented by distinct metagenomic bins, whereas OTUs of >96.8% similarity would probably be represented by a single partial genome (i.e., metagenomic bin).

Overview of phyla and Taxa Detected in the Mushroom Spring Undermat

In the following paragraphs we will describe selected taxa from each phylum detected in the undermat community based on combined information of iTag and metagenomic sequence data. The phyla and members thereof are presented in the order of abundance, starting with the most abundant phylum and the

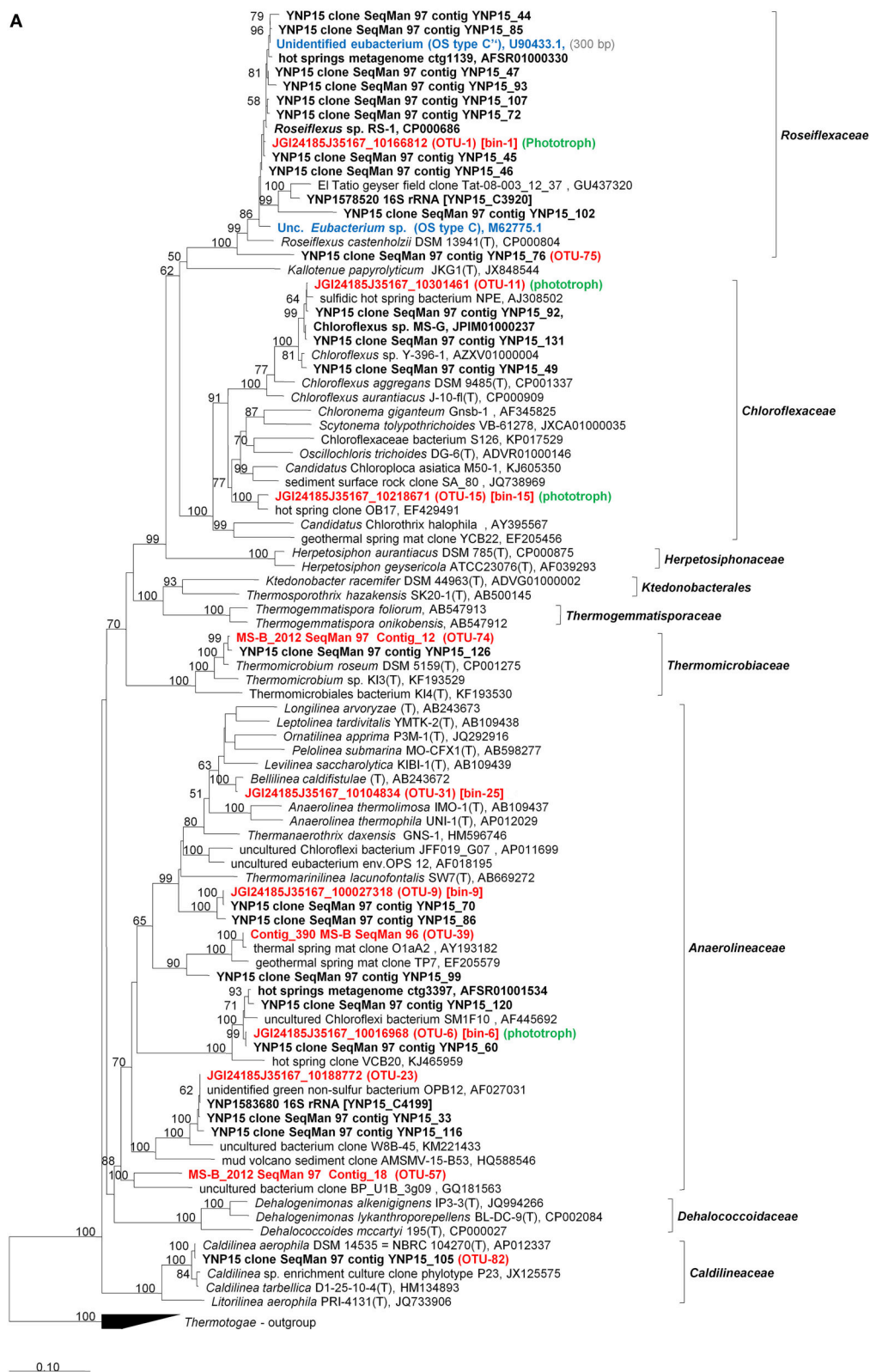




FIGURE 5 | Phylogenetic tree based on 16S rRNA gene sequences showing the phylogenetic relationship between members of the phylum Chloroflexi (A) and Bacteroidetes-Chlorobi (B) phyla and sequences obtained from the Mushroom Spring microbial undermat community. The tree was generated based on the Maximum Likelihood method using the phyML software included in the ARB package. Percentage numbers on nodes refer to 100 bootstrap pseudoreplicates conducted. Only values >50% are shown. Bold sequences were obtained from Mushroom or Octopus Spring in this or previous studies. Red bold labels indicate sequences obtained in this study. Blue bold labels indicate "OS type" sequences from previous studies. OTU numbers shown refer to the most abundant OTU represented by the sequence. Only sequences with length >1,000 bp were used for phylogenetic calculations. Sequence length <1,000 bp are given in (gray) in the labels and corresponding sequences were added using the Parsimony method without changing tree topology.

most abundant member, respectively. Taxonomic identification was always based on the longest 16S rRNA sequence available, in conjunction with phylogenetic marker genes. Information on additional taxa and phyla can be found in the phylogenetic trees and the Supplemental Materials (**Figure 5**, and **Figure S1**, **Table S1**). Phylogenetic analyses based on 16S rRNA sequences extracted from metagenomic data identified >50 members of 20 different phyla (**Figure 5** and **Figure S1**), most of which could also be affiliated with iTag sequences obtained in the amplicon study.

Chloroflexi

Members of the phylum *Chloroflexi* were the most diverse group of organisms present in the microbial undermat community. Overall, 41 OTUs were affiliated with the phylum *Chloroflexi* (**Table S1**), and twelve *Chloroflexi* sequences were identified phylogenetically (**Figure 5A**). Five of the fifteen most abundant OTUs (>1,000 reads), as well as four abundant OTUs with ≥ 100 reads, were identified as members of the *Chloroflexi* (**Table S1**, **Figure 5A**). Based on the metagenomic information for these taxa, four out of five very abundant *Chloroflexi* are chlorophototrophic members of this phylum (OTUs-1, 6, 11, and 15; see **Figure 5A**), while one is a putative chemoheterotroph (OTU-9). Three additional abundant OTUs also are associated with putatively chemoheterotrophic members of this phylum (OTUs 23, 31, and 39). Thirty-two less abundant OTUs were also affiliated with the phylum *Chloroflexi* (**Table S1**, **Figure 5A**).

Binning of the assembled metagenomic data yielded only a very small partial genome for *Roseiflexus* spp., the most abundant and most diverse OTU in the undermat (Bin-1; **Figures 2, 3, 5**, **Tables 1–3** and **Table S1**). Bin-1 did not contain any phylogenetic marker genes but was identified by high nucleotide sequence identities ($92 \pm 5\%$; range 79–100%) to the *Roseiflexus* sp. RS-1 genome (CP000686, 5.8 Mb, van der Meer et al., 2010). The *Roseiflexus* sp. RS-1 genome recruited 23,534 contigs from the metagenome ($\geq 85\%$ nt identity and ≥ 75 coverage), of which 13,329 contigs showed sequence identity of $\geq 95\%$. Only 12 of those contigs were >5 kb in length, sharing a minimum of 94.52% nt identity with the *Roseiflexus* sp. RS-1 genome sequence. *Roseiflexus* sp. RS-1 is a filamentous anoxygenic phototroph that synthesizes bacteriochlorophyll (BChl) *a* but not BChl *c*. It was previously isolated from Mushroom Spring and was affiliated with OS Type C sequences obtained in early molecular studies (Ward et al., 1990; Ferris et al., 1996b, 1997; Ferris and Ward, 1997). In addition to BChl *a*-containing photosynthetic reaction centers, the genome of this organism encodes xanthorhodopsin, which was also detected in the undermat metagenome (RoseRS_2966, GenBank Acc. no. ABQ91330.1; JGI24185J3567_10248071), and indicates a possible additional use of light energy (Choi et al., 2014). The small number of long contigs affiliated with this OTU, in combination with the broad coverage range from $31\times$ to $1,557\times$, reflects a high microdiversity as well as the high abundance of the core genome sequences.

A 1,364-bp partial 16S rRNA sequence identified OTU-6 as a member of the *Chloroflexi*, which is most closely related to uncultured members in streamer biofilm-producing communities in YNP hot springs (**Table 3**; Meyer-Dombard et al., 2011). It represents an uncultured chlorophototrophic *Anaerolineae*-like organism, which was also identified in the upper green layer of the Mushroom Spring microbial mat in a previous metagenomic analysis (Klatt et al., 2011). Despite the absence of a 16S rRNA gene, Bin-6 was identified to represent OTU-6 based on $93 \pm 5.6\%$ average nt identity to Cluster 6 from the upper layer metagenome (Klatt et al., 2011), which did contain a ribosomal RNA sequence with 98% identity to OTU-6, as well as 99% sequence identity to a 16S rRNA sequence detected in the metagenome of this study. When first reported by Klatt et al. (2011), this uncultured organism was identified as “*Anaerolineae*-like,” with *Anaerolinea thermophila* strain UNI-1 being its closest cultivated and described relative (85% nt identity, Sekiguchi et al., 2003). At the time of this writing [February 2016], a BLAST search identified *Thermanaerothermox daxensis* strain GNS-1^T (Grégoire et al., 2011) and *Thermomarinilinea lacunofontalis* strain SW7 (Nunoura et al., 2013) as the closest isolated relatives with a 16S rRNA sequence identity value of 87% (**Table 1**). Phylogenetic analysis based on the full-length 16S rRNA sequences supports a phylogenetic affiliation to the *Anaerolineales* as well as a more distant relationship to known chlorophototrophic *Chloroflexi* (**Figure 5A**). Genes annotated within this metagenomic bin suggest that, like *Roseiflexus* spp., this anoxygenic chlorophototroph has the potential to produce BChl *a* but probably doesn’t contain BChl *c* or chlorosomes, although it does possess a putative xanthorhodopsin-like gene (Klatt et al., 2011). Thin short filaments possibly representing this *Anaerolineae*-like phototrophic *Chloroflexi*, tentatively named “*Ca. Roseilinea gracile*” (Tank et al., in press), have been observed in fresh mat samples and enrichment cultures. They exhibit BChl *a* but not BChl *c* autofluorescence.

OTU-09 is represented by Bin-9 and was also identified as being derived from a member of a cluster of uncultured *Chloroflexi* within the *Anaerolineae* (**Figure 5A**). However, based on the absence of photosynthesis-related genes in the corresponding metagenomic bin and the absence of unassigned photosynthesis-related genes in the remaining unbinned contigs, the organisms corresponding to OTU-09 are not predicted to be chlorophototrophs.

A close relative of *Chloroflexus* sp. strain MS-G, a chlorophototrophic member of the *Chloroflexi* that was previously isolated from this mat (Thiel et al., 2014b), is represented by OTU-11 and Bin-11 in this study. Like strain MS-G, OTU-11 is predicted to be an anoxygenic phototroph containing type-2 (quinone-type) photosynthetic reaction centers, light-harvesting complex 1 and chlorosomes based on a metagenomic bin of 3.1 Mb, with an average read coverage of $30\times$ (Bin-11, **Table 3**). The bin contained 21 phylogenetic marker genes, all of which share amino acid sequence identity values of 98.7 to 100% with sequences from *Chloroflexus* sp. MS-G (**Table 3**). The organism representing OTU-11/Bin-11 and strain MS-G share 98.3% 16S rRNA and $94 \pm 6\%$ overall genomic nucleotide sequence identity, respectively.

A third anoxygenic phototrophic *Chloroflexi* is represented by OTU-15 and Bin-15. Phylogenetic analysis and BLAST search results indicate this organism to be only distantly related to other chlorophototrophic *Chloroflexus* spp., displaying 90–91% 16S rRNA sequence identity to *Oscillochloris trichoides*, *Chloroflexus aurantiacus* J-10-fl and “*Candidatus Chloroploca asiatica*.” The organism associated with these sequences presumably represents a novel genus of chlorophototrophic *Chloroflexi* within the family *Chloroflexaceae* (Figure 5A). Based on the conserved signature indels that are specific for different groups within the *Chloroflexi* as described by Gupta et al. (2013), this filamentous anoxygenic phototroph is affiliated with the proposed order of “green nonsulfur bacteria,” *Chloroflexales*, suborder *Chloroflexineae*, but is distinct from all known members of the genera *Chloroflexus* and *Oscillochloris*. The functional gene content of the associated metagenome bin (Bin-15) indicates that this organism has the capacity to synthesize BChl *a* and *c*. A filamentous BChl *a*- and BChl *c*-producing isolate similar to *Oscillochloris* sp. has been obtained in enrichment cultures, and tentatively named “*Candidatus Chloranaerofilum corporosum*” (Tank et al., in press).

Thermomicrobium roseum, phylum *Chloroflexi*, which had previously been isolated from the mats (Jackson et al., 1973), was detected in the metagenome in this study and a previous 16S rRNA cloning study (Klatt et al., 2013a), but *T. roseum* was only present in low numbers based on the analysis of iTAG amplicons (OTU-74, 44 reads, Table S1, Figure 5A).

Thermotogae

Only two OTUs, OTU-2, and OTU-107, were identified as members of the phylum *Thermotogae* by the RDP classifier (Table S1). OTU-2 represents the second most abundant species-level iTAG sequence and the corresponding metagenomic 16S rRNA sequence is 99% identical to that of *Pseudothermotoga hypogea*, formerly known as *Thermotoga hypogea* (Fardeau et al., 1997; Bhandari and Gupta, 2014). Bin-2 sequences, which represent this *Pseudothermotoga* sp. OTU-2 mat member (Table 3), show high similarities (98–100% aa sequence identities) to sequences obtained from a previous metagenomic study by Klatt et al. (2013a; IMG/M OID 2015219002), and form a single cluster with the genome sequence of *Pseudothermotoga hypogea* DSM 11164 in the metagenome binning analysis, which indicates the high similarity of these two genomes. OTU-107 shares 99% nt sequence identity to *Fervidobacterium pennivorans* strain DSM 9078 as well as to *Fervidobacterium* sp. isolated from YNP (Sullivan et al., unpublished, AY151268) but is represented by only 20 reads (Table S1, Figure S1). In addition, several sequences were affiliated with group EM3, which has tentatively been placed in the *Thermotogae* (Reysenbach et al., 2000) (Table S1, Figure S1). OTU-10 was misidentified as a member of the *Chlorobi* by RDP classifier, but actually represents the most abundant EM3 population and shares highest similarities with hot spring clones OPB88 (AF027006, Hugenholtz et al., 1998b) and OPS2 (AF018187, Graber et al., unpublished) from YNP with 99 and 98% 16S rRNA nt identity, respectively. Bin-10 representing this OTU was identified based on the presence of a matching 16S rRNA gene (Table 3). Phylogenetic affiliations

of the phylogenetic marker genes were uncertain with most of the sequences only being assigned to the kingdom (“bacteria”) and phylum level (“*Bacteroidetes*,” “*Chlorobi*,” “*Deinococcus-Thermus*,” “*Chloroflexi*,” or “*Thermotogae*,” respectively), which indicates a high degree of novelty for this uncultured organism. Sequences similar to the ones in this metagenomic bin have previously been detected in the oxic upper green layer of the mat community (Klatt et al., 2011). The sequences formed unidentified Cluster 8 in the previous study, which were associated with an uncultivated, putatively heterotrophic bacterium. Bin-10 and Cluster 8 sequences formed a single bin when included in the analysis. A BLASTn comparison revealed an average nucleotide identity of $97 \pm 3\%$ between sequences of the previous cluster and the sequences in the bin from this study.

Armatimonadetes (OP10)

Uncultivated members of the Candidate phylum OP10, now named *Armatimonadetes* (Tamaki et al., 2011; Lee et al., 2013), were first detected in Obsidian Pool in YNP (Hugenholtz et al., 1998b). The undermat community at Mushroom Spring also contains a considerable diversity of members of this phylum. Two of the most highly abundant OTUs, OTUs 3, and 12, were identified as members of the *Armatimonadetes*. In addition, two abundant (OTUs 18 and 33) and nine less abundant iTAG OTUs were identified as members of this phylum (Table S1). Partial genomes were identified for OTUs-3, 12, and 18 (Table 3, Figure S1).

Despite the high abundance of *Armatimonadetes* member OTU-3 sequences in the amplicon study and the presence of a partial 16S rRNA sequence with high coverage (951×; JGI24185J35167_1062246), no corresponding bin was obtained in the undermat metagenome. Serendipitously, a highly similar organism (99% 16S rRNA sequence identity) was identified as a chemoheterotrophic contaminant in a cyanobacterial enrichment culture obtained from these mats in the Ward laboratory at Montana State University (unpublished data). A partial genome of this enrichment contaminant was obtained by binning the assembled contigs of the corresponding enrichment culture metagenome (Bin-3, Table 3). This enrichment partial genome recruited 17,252 sequences (a total of 11 Mb of sequence data) from the undermat metagenome displaying $90.5 \pm 7.5\%$ nt id (covering min. 80% of the metagenome scaffold). OTU-3 amplicon sequences were also detected in the upper green layer in lower numbers (4.5 vs. 0.8% relative abundance; Table S1) and a partial genome of this organism was also detected as an unidentified heterotroph Cluster 7 in the upper layer metagenome (Klatt et al., 2011). The partial genome of the upper layer displayed similar identity values of $90.3 \pm 7.5\%$ to the enrichment culture metagenome bin and $94.6 \pm 5.3\%$ to sequences in the undermat metagenome, and formed a single ESOM bin with the partial genome obtained from the enrichment culture (data not shown). OTU-3 was phylogenetically identified as belonging to the “OS-L clade” within the uncharacterized group 7 of the phylum *Armatimonadetes* (Lee et al., 2013) (Figure S1). Clade OS-L is named after the first sequence of this clade, OS Type L, obtained from a DGGE study of enrichment cultures from microbial mats in Octopus Spring (Ward et al.,

1992), with which the 16S rRNA genes in both Bin-3 from the enrichment culture and the undermat metagenome share 98% nt identity (L04707). So far, no isolated representative has been reported for this phylogenetic group. The presence of all 31 bacterial phylogenetic marker genes in the bin suggests that it contains a nearly complete genome (Table 3). Genes encoded in the partial genome, in combination with its occurrence in an enrichment with oxygenic cyanobacteria, indicates that this organism probably exhibits an aerobic or microaerobic lifestyle, similar to the other isolated members of the *Armatimonadetes* (Lee et al., 2011; Tamaki et al., 2011; Im et al., 2012). A considerable microdiversity was suggested by the presence of nine abundant iTag sequences (Table 1) as well as the diversity of partial, flagellum-associated genes affiliated with this organism, which were present on short contigs in the metagenome. Additionally, thirteen closely related 16S rRNA sequences were derived from a previous undermat 16S rRNA cloning study (Klatt et al., 2013a). These sequences show high identity values (>97%) to the OTU-3 sequence as well as to each other (assembly based on 97% nt sequence identity, Figure S1A) and also reflect a high microdiversity of these organisms. Similar to the situation found for *Roseiflexus* spp. (see above), the high microdiversity suggested for this taxon probably caused assembly difficulties, which may explain why no metagenomic bin was recovered directly from the undermat metagenome.

Aquificae

Of four OTUs identified as belonging to members of the *Aquificae* (Table S1), only OTU-4 was detected in significant numbers (Table S1). The corresponding 1,434-bp rRNA metagenomic sequence is 99% nt identical to clone sequences previously obtained from YNP hot spring habitats (*Thermocrinis* sp. clone YNP_SBC_BP2A_B2, HM448202, Meyer-Dombard et al., 2011), as well as to the YNP isolate *Thermocrinis* sp. P2L2B (AJ320219, Eder and Huber, 2002). The closest described relative is *Thermocrinis ruber* DSM 23557, which was isolated from Octopus Spring and which has a 16S rRNA sequence that shares 97% nt identity to the one found in this study (Huber et al., 1998) (Figure S1). Correlating to the high microdiversity detected for this OTU (Table 1), only a small partial genome was identified in the binning analysis of the metagenome (Bin-4, Table 3). The presence of at least two closely related populations in the undermat community is indicated by two highly similar (96% amino acid identity), *Thermocrinis*-like *soxB* genes; these genes are located on three individual scaffolds in the metagenome, each [gene-1, ~270× coverage: JGI24185J35167_10446912, JGI24185J35167_10438521, JGI24185J35167_10819822; gene-2, ~70× coverage: JGI24185J35167_10446972, JGI24185J35167_10438611, JGI24185J35167_10820392], which also suggests problems with sequence assembly that could be related to microdiversity.

Cyanobacteria

The two major photoautotrophic primary producers of the upper green layer, *Synechococcus* spp. Type A and Type B', were also abundant members of the undermat by iTag analysis (OTUs 5 and 22, Table 1, Table S1). Seventeen additional but less abundant iTag OTUs (each ≤25 reads, representing <0.05%

of the total iTag sequences) were assigned to cyanobacteria (Table S1). At the temperature sampled in this study (60°C), members of *Synechococcus* sp. Type B' (OS Type B', Table 4) are the predominant organisms (Klatt et al., 2011; Liu et al., 2011) and were also detected in this study (OTU-5, Bin-5, Table 3). *Synechococcus* sp. A (OS Type A, Table 4) sequences were detected in lower abundance (OTU-22, Table S1). The small size of Bin-5 (Table 3) reflects a low number of long and well-assembled contigs (68 contigs, 5,005–12,792 bp; 18× to 96× coverage) in comparison to a total of 3,353 contigs identified as having their origins in members of the *Cyanobacteria* in the metagenome (440 to 12,792 bp). Local BLASTn analysis and reference guided assembly using the genome sequence of *Synechococcus* sp. Type B' as query (applying a 95% nt identity threshold) identified 4,898 contigs as belonging to these organisms. The low assembly quality is indicative of high microdiversity as indicated by the presence of seven abundant iTag sequences (Table 1). Recent studies have found that a high number of ecotype populations occur within this cyanobacterial population, displaying variations in gene content and sequence as well as differences in gene arrangement (Becraft et al., 2011; Olsen et al., 2015). Genome sequences of several ecotypes isolated from the dominant cyanobacteria from Mushroom Spring are now available, and these provide comprehensive insights into the physiological and metabolic capacities of the oxygenic chlorophototrophs in the mat (Bhaya et al., 2007; Nowack et al., 2015; Olsen et al., 2015).

Atribacteria (OP-9/JS1)

The phylum *Atribacteria*, formerly known as Candidate phylum OP-9/JS1, exhibited low diversity. Of two OTUs identified as belonging to members of this phylum, only OTU-7 was detected in significant numbers in the iTag analysis (Table S1, Figure S1). OTU-7 represented 2.4% of all iTag reads and was represented by only a single abundant dereplicated iTag sequence (Table 1). Bin-7 contained a partial genome of this uncultured bacterium, as identified by the full-length 16S rRNA sequence which shared 99% and 98% sequence identity to *Atribacteria* clones OPB72 and TP29 obtained from hot springs in YNP and Tibet, respectively (Hugenholtz et al., 1998b; Lau et al., 2009). The affiliated metagenomic bin indicates an anaerobic, fermentative lifestyle for this member of the *Atribacteria* (data not shown), which is similar to properties deduced from single-cell genome sequences previously obtained from members of the *Atribacteria* (Dodsworth et al., 2013; Nobu et al., 2016).

Nitrospirae

iTag analysis identified seven *Nitrospirae* OTUs in the undermat community, of which only one, OTU-8, was abundant (Table S1). Bin-8 was assigned to this *Thermodesulfovibrio* sp.-like mat member based on presence of the corresponding 16S rRNA sequence (Figure 4, Table 3). OTU-8 represented ~2.0% (3,283 reads) of all iTag sequences (Table 1), and the full 16S rRNA sequence was most closely related to a clone sequence obtained from geothermal groundwater (99%, clone: SMD-B01, NCBI acc. no. AB477993, Kimura et al., 2010) and to *Thermodesulfovibrio yellowstonii* strain DSM 11347, as the closest isolated relative (96%, NCBI acc. no. CP001147, Henry et al., 1994; Bhatnagar

et al., 2015). Bin-8 contained scaffolds with coverage values ranging from 29 to 135, which possibly reflects two different populations with different abundances. This was also suggested by the different read numbers of two abundant, dereplicated iTag sequences (OTU-8, iTag-10, 1,721 reads; and iTag-28, 602 reads; **Table S1**). The partial genome suggests sulfate-reducing metabolism for this organism, similar to *T. yellowstoneii*, which was isolated from thermal vent water in Yellowstone Lake, Wyoming, USA (Henry et al., 1994; Bhatnagar et al., 2015). The *dsrAB* gene sequences associated with dissimilatory sulfate-reduction of this uncultured organism have previously been detected in the Mushroom Spring microbial mat, and the corresponding *Thermodesulfobivrio*-like organism was associated with the sulfate reduction activity measured in the mat (Dillon et al., 2007). OTU-8 has been detected in both the upper and lower parts of the mat (**Table 5**, **Table S1**), possibly indicating that these organisms are not restricted to the undermat; this is further supported by the finding of *Thermodesulfobivrio*-like sequences also in the green upper layer metagenome in a previous study (Klatt et al., 2011).

Aminicenantes (OP8)

The *Aminicenantes* (Candidate phylum OP8) was represented by only a single taxon, OTU-13, and its corresponding metagenomic Bin-13, which contains a 1,497-bp 16S rRNA gene sequence (**Table 1**, **Table S1**). Notably, OTU-13 amplicon sequences were found exclusively in the undermat community (**Table S1**). Although the iTag sequence shared 99% nt identity to the uncultured *Aminicenantes* bacterium clone OPB95 obtained from a Yellowstone hot spring (AF027060, Hugenholtz et al., 1998b), the full-length sequence showed only 95% nt identity to that sequence. No isolated bacterium shares more than 88% nt identity with this uncultured organism. 16S rRNA gene sequence surveys indicated that members of the *Aminicenantes* are ubiquitously present in many different habitats and across many environmental parameters (temperature, salinity, and oxygen tension) (Farag et al., 2014). They usually represent only a small fraction (<1%) of microbial communities, but have been found to be more abundant in anoxic environments (Farag et al., 2014).

Planctomycetes

Five abundant iTag OTUs were identified as belonging to members of the phylum *Planctomycetes* (**Table S1**), the very abundant OTU-14 (1,260 reads), as well as four less abundant OTUs (OTUs-16, 19, 49, and 51, **Table S1**). Twelve additional *Planctomycetes* sequences were found in very low abundance (**Table S1**).

Bin-14 contained a partial genome for *Planctomycetes* member OTU-14 and was identified based on the corresponding full-length 16S rRNA sequence as well as nineteen phylogenetic marker genes (**Table 3**, **Figure S1**). An uncultured hot spring-associated bacterium from a neutral 61°C geothermal hot-spring mat in Tibet, clone TP5, was identified as closest relative (EF205581, 99%, Lau et al., 2009). The microaerophilic, facultatively anaerobic, thermophilic *Planctomycetes* strain, *Thermogutta terrifontis* strain R1^T (KC867694, Slobodkina et al., 2014), with 90% sequence identity, is the most closely related isolated relative (**Table 1**). Based on the number of phylogenetic

marker genes present in the metagenome bin, and because of the large sizes of available *Planctomycetes* genomes (3.8–9.7 Mb for those in JGI/IMG as of December 2015), we expect the 1.87-Mb bin to represent no more than 60% of the genome. The presence of the iTag sequences for this OTU almost exclusively in the undermat sample (a single read was found in iTag analysis of upper green layer; **Table 5**, **Table S1**) suggests that this organism lives exclusively in the orange-colored undermat and possibly in its deeper regions below 3 mm, where mainly anoxic conditions occur and persist (Nübel et al., 2002; Jensen et al., 2011).

Bin-23 was also identified as derived from a member of the *Planctomycetes*, but could not be directly affiliated with any iTag sequence(s) due to absence of an rRNA sequence in the bin (**Table 3**).

Acidobacteria

Thirteen OTUs representing four different members of the *Acidobacteria* were identified in the Mushroom Spring undermat community, and two of them were abundant with >100 reads (**Table S1**). OTU-17 was a member of group 4 of the *Acidobacteria* and was identified as *Cab. aerophilum* (Tank and Bryant, 2015a,b). Bin-16 (**Table 3**) contained a partial genome for this unique microaerophilic, chlorophototrophic member of the phylum *Acidobacteria*, which was first identified in the phototrophic mats of Mushroom and Octopus Spring and corresponds to the OS Type D sequences from earlier studies (Ward et al., 1990, 1992; Bryant et al., 2007; Tank and Bryant, 2015a,b).

OTU-36, as well as the less abundant OTU-72, were members of *Acidobacteria* group 3 and were identified as *Solibacter*-like organisms. Bin-20 was associated with OTU-36 by the presence of a 16S rRNA-containing scaffold as well as by the presence of six phylogenetic marker genes (**Table 3**). All six phylogenetic marker genes indicated an affiliation with the *Acidobacteria* and four of them specifically with the candidate species, “*Ca. Solibacter usitatus*” (Challacombe et al., 2011). Phylogenetic analysis supported the affiliation and placed the sequence in subgroup 3 of the *Acidobacteria*, closely related to Yellowstone clone OPB3 (98%, AF027004, Hugenholtz et al., 1998b) and “*Ca. Solibacter usitatus*” Ellin6076 as the closest named relative (**Table 1**, **Figure S1**). The low number of phylogenetic marker genes indicates that this member of the *Acidobacteria* has a large genome, only a part of which is included in the metagenomic bin. This correlates well with the fact that “*Ca. Solibacter usitatus*” Ellin6076 has an exceptionally large, 9.97-Mb genome (Challacombe et al., 2011).

The fourth member of the phylum *Acidobacteria* corresponded to a less abundant OTU (OTU-61, 70 reads = 0.1%) and was represented by two partial 16S rRNA sequences in the metagenome. These sequences and the represented uncultured organisms were affiliated with OS Type K sequences from previous studies (**Table 4**, Ward et al., 1992; Weller et al., 1992).

Proteobacteria

Four abundant OTUs were affiliated with the phylum *Proteobacteria* by the RDP classifier, one of which was misidentified as *Proteobacteria* and rather represents a

Brevinema-like member of the *Spriochoeta* (OTU-35), two of which were *Deltaproteobacteria* (OTUs-40 and 44), and one of which was an *Alphaproteobacterium* (OTU-46). Twenty-nine additional, low-abundance OTUs were affiliated with *Proteobacteria* by RDP classifier (Table S1). Sequences for 16S rRNAs of two *Alpha*-, two *Beta*- and one *Delta*-*Proteobacteria* were found in the metagenome (Figure S1C). The abundant *deltaproteobacterial* sequence (OTU-44) was closely affiliated to a sequence obtained in a previous metagenome study (Figure S1C, Klatt et al., 2013a). Although the *Deltaproteobacteria* are commonly known to include members with sulfate-reducing metabolism, and sulfate-reduction has been shown in the microbial mat at Mushroom Spring (Dillon et al., 2007), *deltaproteobacterial dsrAB* genes were not identified in this nor any previous study. No metagenomic bin was affiliated with a *Deltaproteobacterium*.

The abundant *Alphaproteobacterium* (OTU-46) was identified as an *Elioraea* sp. within the *Rhodospirilliales*, which corresponds to OS Type O obtained in previous studies (Figure S1C, Table 4, Ward et al., 1992). The corresponding partial genome (Bin-22, Figure 4, Table 3) as well as the genome for the closest relative, *Elioraea tepidiphila* DSM 17972 (NCBI acc. no. NZ_KB899965.1), contain genes for anoxygenic photosynthesis. Although chlorophototrophy has not been described for *Elioraea tepidiphila* (Albuquerque et al., 2008), the ability to synthesize BChl *a* is predicted for the OTU-46 population in the undermat community. A BChl *a* containing strain, “*Candidatus* *Elioraea thermophilum*,” was isolated from the mat, which shares 99.8% and 99.2% sequence identity with the 16S rRNA sequences from the metagenome and amplicon study, respectively (Figure S1C, Tank et al., in press). A low abundance *Alphaproteobacterium* sequence (OTU-121, 16 reads) was identified as belonging to a *Roseomonas/Rhodovarius*-like organism, for which an isolate has been obtained from Mushroom Spring and which has tentatively been named “*Candidatus* *Roseovibrio tepidum*” (Figure S1C, Tank et al., in press). The isolate exhibits BChl *a* autofluorescence suggesting a phototrophic lifestyle, which is further strengthened by the presence of low coverage, unidentified *alphaproteobacterial pufLM* sequences in the metagenome (scaffold JGI24185J35167_1024732, genes 2 and 3, 20× coverage). Only a single described *Roseomonas* sp., *R. aestuarii*, has been reported to produce BChl *a*, but no *pufLM* sequences are available for that isolate (Venkata Ramana et al., 2010). Furthermore, two low-abundance OTUs (OTUs-101 and 154) showed the same phylogenetic affiliation (*Hydrogenophilus* sp., *Betaproteobacteria*) as OS type G from previous studies (Ward et al., 1990, 1992). The OS Type R sequence (NCBI acc. no. U46750, unpublished) represented an unidentified *Betaproteobacterium* and a similar, low-abundance *iTag* sequence (OTU-172) was detected in this study (Table 4, Figure S1C).

Bacteroidetes-Chlorobi

The RDP classifier identified twenty and eight different OTU sequences belonging to members of the phyla *Chlorobi* and *Bacteroidetes*, respectively. Seven OTUs affiliated with the

Chlorobi were abundant with read numbers >100, and one was very abundant with >1,000 reads (Table S1). However, the most abundant “*Chlorobi*” sequence (OTU-10) was mis-classified and represents an *Thermotogae/EM3* group member (see above, Table 1, Figure 5B). The other abundant *Chlorobi* sequences were affiliated with the proposed family *Thermochlorobacteriaceae* (OTU-38) (Liu et al., 2012), “*Chlorobi* lineage 5” = “OPB56 group” (OTUs 24, 27, and 29) (Iino et al., 2010; Hiras et al., 2015) and “*Chlorobi* lineage 2” = “SM1H02 group” (OTUs 34 and 45) (Iino et al., 2010; <http://www.arb-silva.de/browser/ssu-121/AY555793>, named after clone SM1H02, Genbank acc. no. AF445702). Bin-19 (Table 3) was identified as a partial genome representing OTU-24, a representative of OPB56, a subgroup of the *Chlorobi* with predicted chemoheterotrophic lifestyle that was first detected in YNP (Hugenholtz et al., 1998b; Hiras et al., 2015, Table 3). A low abundance OTU in the OPB56, OTU-262, was identified as a probable representative of the OS Type F sequences from previous studies (Table 4, Ward et al., 1990, 1992). The first aerobic, phototrophic member of the *Chlorobi*, “*Ca. Tcb. aerophilum*,” which belongs to the proposed family *Thermochlorobacteriaceae* and was identified in the upper green layer of the microbial mat by previous metagenomic analyses (Liu et al., 2012), is represented by OTU-38 (Table 1), and was identified as OS Type E in previous studies (Ward et al., 1990, 1992; Ferris et al., 1996b). Bin-21 is derived from this novel phototroph (Table 3) and supports its characterization as a chlorophototroph that synthesizes type-1 reaction centers and chlorosomes, similar to cultivated relatives among the green sulfur bacteria, but which is otherwise very different physiologically. “*Ca. Tcb. aerophilum*” is proposed to be an aerobic photoheterotroph that cannot oxidize sulfur compounds, cannot fix N₂, and does not fix CO₂ (Liu et al., 2012).

Bin-24 (Table 3) does not contain a 16S rRNA sequence, but was affiliated with a putative member of the *Bacteroidetes-Chlorobi* group based on phylogenetic marker genes. It is most closely related to heterotrophic members of the *Chlorobi*, in the family *Ignavibacteriaceae* (Liu et al., 2012; Kadnikov et al., 2013) and is presumably affiliated with OTUs-34 or 45 in the *Chlorobi* Lineage 5/group SM1H02 (Figure 5B). All genes needed for dissimilatory sulfate reduction are present in the partial genome and indicate that this organism is putatively the first sulfate-reducing member of the *Bacteroidetes-Chlorobi* group. These results will be described in detail elsewhere (Thiel et al., in preparation). The OS Type M sequences obtained in previous studies (Ward et al., 1992) are affiliated with OTU-34 as well as with two partial 16S rRNA sequences from the metagenome (Table 4) within the SM1H02 (*Chlorobi* Lineage 2) group.

Only low abundance OTUs were affiliated with the *Bacteroidetes* (Table S1). Many of them were closely related to clone sequences obtained in a previous undermat study, and some also represented partial 16S rRNA sequences from the metagenome (Figure 5B, Klatt et al., 2013a). *Schleiferia thermophila*, a strain of which has been isolated from Octopus Spring microbial mats (Thiel et al., 2014a), was not detected in this study.

Deinococcus-Thermus/Thermi

Of two different members of the phylum *Thermi* identified in this study, only *Meiothermus* sp. was abundant in the undermat community (OTU-21, 656 reads), whereas sequences of *Thermus* spp. were only present in low numbers in the iTag study (Table S1, Figure S1C). Members of both genera have been isolated from these mat communities (Brock and Freeze, 1969; Ward et al., 1997; Thiel et al., 2015). OTU-21 was identified as a relative of *Meiothermus ruber*, a member of which, strain A, has previously been isolated from an enrichment culture originally obtained from the microbial mats at Octopus Spring and whose genome has been sequenced (Thiel et al., 2015). Tetranucleotide frequency-based binning of contigs > 10 kb led to a 1.3-Mb partial genome (Bin-18, Table 3) for this moderately thermophilic, aerobic, and heterotrophic bacterium. The *Meiothermus* sp. 16S rRNA sequences obtained from the metagenome share 96.7% nt sequence identity with *M. ruber* strains A and DSM1279^T. Sequences of Bin-18 shared 84.5% ($\pm 4.5\%$) with the *M. ruber* strain A genome and 84.2% ($\pm 4.5\%$) with *M. ruber* DSM1279^T. Although the (partial) genome sequences of the isolate and the metagenome bin clusters overlap, some separation was visible when the sequences of both organisms were included in the binning analyses (data not shown).

Archaea

Although methanogenesis has been demonstrated in several mats of alkaline siliceous hot springs, including Mushroom Spring (Ward, 1978; Sandbeck and Ward, 1982), and methane has been shown to accumulate in the water above the Mushroom Spring mat in darkness (Kim et al., 2015), iTag sequencing only identified a few partial 16S rRNA sequences as potentially derived from methanogenic *Archaea* (OTUs-143, 151, 162, 192, and 244; ≤ 11 reads = $\leq 0.01\%$, Table S1). Phylogenetic analysis confirmed affiliation to the *Euryarchaeota* for four of them (OTUs-143, 151, 162, and 192, Figure S1A) and three of the sequences were detected in a cloning experiment from a previous study (Klatt et al., 2013a); thus, methanogenic archaea seem to be present in the mat over time, although in very low abundance. One OTU, OTU-151 with 10 reads but no representative sequence in the metagenome, shows high similarity (99% nt id) with the 16S rRNA sequence of the methanogenic archaeon *Methanothermobacter thermoautotrophicus*, strains of which have been isolated from these mats previously (former *Methanobacter thermoautotrophicum*; Sandbeck and Ward, 1982). Further, a single, low coverage *mcrA* gene encoding a methyl-coenzyme M reductase alpha subunit was present in the metagenome (JGI24185J35167_11200021, 7 \times coverage) possibly indicating methanogenic metabolism in at least one of the archaeal mat members. Two slightly more abundant 16S rRNA sequences affiliated with ammonia-oxidizing *Archaea* were detected (Table S1). One (OTU-60, 72 reads) was related to “*Candidatus Nitrosocaldus yellowstonii*,” which was also identified in an enrichment culture from Octopus Spring mat in previous studies (De La Torre et al., 2008). The other, OTU-67 represents a member of a putatively novel archaeal phylum/division, related to “*Candidatus Caldiarchaeum subterraneum*” (Nunoura et al., 2011). Another less abundant

iTag sequence, similar to that of an archaeal 16S rRNA sequence recovered from the undermat metagenome previously (Klatt et al., 2013a), was also detected in the iTag analysis (OTU-125, 15 reads), but not in the metagenome of this study (Figure S1A). None of the metagenomic bins could be identified as belonging to *Archaea*, and only a few contigs with low coverage values, showed high identities to known archaeal sequences. Thus, our metagenomic and 16S rRNA gene amplicon studies indicate a very low abundance of *Archaea*, of which sequences related to ammonia-oxidizing *Archaea* seem to be more abundant than possible methanogenic *Archaea*. The low abundance of archaeal sequences is consistent with the low relative abundance of archaeal lipids in previous studies, which had been discussed to be related to the energy flows through the trophic structure of the community (Ward et al., 1989).

Firmicutes

Although *Anoxybacillus* spp. are common members of cyanobacterial enrichment cultures from these environments (e.g., Nowack, 2014; Olsen et al., 2015; Tank and Bryant, 2015b), no evidence for this organism was found in the metagenome nor the iTag analysis. Twenty-four OTUs were classified as belonging to members of the *Firmicutes*, of which two (OTUs-251 and 255) were predicted to be *Bacillus* sp.; however, they shared highest sequence similarity to the type strains of *Syntrophothermus lipocalidus* and *Acetomicrobium faecale* (both clostridia). None of the 16S rRNA genes retrieved from the metagenome could be affiliated with the *Firmicutes*. In addition, none of the metagenomic scaffolds were affiliated with *Anoxybacillus* spp. No sequence from an *Anoxybacillus* sp. was identified by BLASTn analysis of the metagenome using the partial genome sequence obtained from the *Anoxybacillus* sp. MT isolated from an enrichment culture from Octopus Spring (Thiel et al. in prep), nor the “phylogenetic distribution of genes by BLAST percent identities” tool implemented in the JGI/IMG website.

DISCUSSION

In this study we analyzed the orange undermat of the microbial mat community at 60°C in Mushroom Spring YNP by 16S rRNA gene amplicon and metagenomic sequencing. Only eight major organismal populations were identified in the upper green layer by genomic, metagenomic and metatranscriptomic analysis (Klatt et al., 2011; Liu et al., 2011). A higher diversity had been speculated to occur in the undermat community (Klatt et al., 2013a). In this study the undermat was found to be a highly diverse but uneven bacterial community, which could be related to the trophic structure associated with mat-decomposing organisms, as hypothesized to explain the variable abundances of lipid biomarkers (Ward et al., 1989) and 16S rRNA sequences (Ward et al., 1998). Out of 317 OTUs, the 15 most abundant ones represent 87% of all iTag sequences, and the single most abundant OTU comprises nearly half of all iTag reads. More than 44 abundant taxa, as defined by read numbers of >100 in the iTag analysis, were detected in the orange-colored undermat at Mushroom Spring. The phylum *Chloroflexi* displayed the highest diversity with nine abundant and 41 total taxon-specific 16S

rRNA sequences (OTUs) found. All of the taxa found in the upper mat by Klatt et al. (2011) were also identified in the undermat.

In this study we analyzed the composition and diversity of the microbial community based on 16S rRNA gene sequences, which cannot easily be translated into species populations. However, relatively high 16S rRNA sequence diversity was found in this study, not only on the OTU level but particularly within the dereplicated iTags, which suggests that this microbial mat community is not simple. Previous observations that closely related cyanobacterial 16S rRNA sequences were differently distributed along environmental gradients (Ferris and Ward, 1997; Ramsing et al., 2000) prompted consideration of the Stable Ecotype Model of species and speciation (Cohan and Perry, 2007), which postulates that some microorganisms exist as ecological species occupying distinct niches (Ward, 1998; Ward and Cohan, 2005). Studies with more rapidly evolving protein-encoding loci led to the prediction of numerous ecotypes with identical or nearly identical 16S rRNA sequences (Ferris et al., 2003; Becraft et al., 2011, 2015; Melendrez et al., 2011). The existence of temperature- and light-adapted *Synechococcus* ecotypes has been demonstrated by obtaining representative strains and studying their temperature and light preferences as well as their genomes, (Allewalt et al., 2006; Nowack et al., 2015; Olsen et al., 2015). A similar microdiversity and existence of putative ecotypes is suggested by this study for members of the undermat community, and in particular for *Roseiflexus* spp., the most dominant member in the undermat. The presence of unique 16S rRNA genotypes in the undermat (this study) and at different temperatures (Ferris and Ward, 1997), supports this inference. In addition to the high diversity of OTUs within the phylum *Chloroflexi*, a high microdiversity was found for *Roseiflexus* spp. by the presence of 24 abundant and a total of 6,193 dereplicated *Roseiflexus* sp. iTag sequences, which is further supported by a preliminary analysis of *pufLM* amplicon sequence data (J. Wood and D. Ward, unpublished data).

The microbial mat as a living and active biological system has been shown to be constantly growing (Doemel and Brock, 1977). In this study we observed phototrophic taxa known from the upper layer in the undermat. Analyses of *psaA* sequences sampled in this metagenomic study suggest that the *Synechococcus* populations observed match species found in the upper mat and thus likely occur in the undermat as a consequence of burial. In contrast, similar analyses of *pufLM* sequences as well as oligotyping suggest that *Roseiflexus* populations in the undermat are a mixture of those found in the upper green mat layers and those uniquely found in the undermat (Table 5, Figure S2, Wood et al., unpublished). The detection of identical dereplicated iTag and oligotype sequences in both layers might indicate burial. However, the detection of oligotypes and dereplicated iTag sequences with higher relative abundance in the undermat strongly suggests the existence of putative ecotypes specifically adapted to niches in the undermat. Further it is important to note that specifically adapted ecotypes can be so closely related that they have the identical 16S rRNA gene sequence, and can only be detected using more rapidly evolving genes (Becraft et al., 2011, 2015). For other organisms, a greater relative abundance, or exclusive presence in the lower part of the mat, is indicated

by the relative number of 16S rRNA gene amplicon reads between the upper layer and undermat samples. For example, *Pseudothermotoga* spp. OTU-2, *Armatimonadetes* member OTU-3, *Thermocrinis* spp. OTU-4, *Chloroflexi* members OTU-6, 9, and 15, as well as the *Atribacteria* member OTU-7, the *Aminicenantes* member OTU-13, and *Planctomycetes* member OTU-14, are found in much higher relative abundance in the undermat (Table 5, Table S1). Future transcriptomic studies will assess which of the detected populations correspond to the highest transcriptional activities based on gene expression. The presence of aerobic, microaerobic and anaerobic organisms detected in this study indicate a possible layered distribution along the steep and fluctuating oxygen gradient and shows that some oxygen is available during the day below a depth of 2 mm in the microbial mat, as previously suggested by microelectrode measurements (Revsbech and Ward, 1984; Nübel et al., 2002; Jensen et al., 2011). Whereas aerobic bacteria and facultative anaerobes are expected to live in the transition zone adjacent to the upper green layer, abundant anaerobic members of the undermat community, e.g., *Pseudothermotoga* sp. OTU-2 and *Atribacteria* member OTU-7 can be expected to be active members mainly in the community below a depth of 3 mm, where anoxic conditions are expected to persist throughout the day (Nübel et al., 2002; Becraft et al., 2011; Jensen et al., 2011). Despite the anaerobic lifestyle of sulfate reduction, *Thermodesulfovibrio* sp. OTU-8 was detected in higher abundance in the upper layer, which might indicate some degree of oxygen tolerance and diel activity patterns, i.e., primary sulfate-reducing activity under anoxic conditions in the afternoon or at night as measured by Dillon et al. (2007). An *Aminicenantes* (OP8) member (OTU-13), a *Planctomycetes* member (OTU-14) and an *Oscillochloris*-like chlorophototrophic member of the *Chloroflexi*, “*Ca. Chloranaerofilum corporosum*” (OTU-15) (Tank et al., in press) were exclusively detected in the undermat by iTag analysis, which suggests that they have an anaerobic lifestyle in the deeper layers of the undermat. However, “*Ca. Chloranaerofilum corporosum*” is expected to be a phototroph, and only a limited amount of light reaches deep into the undermat. Thus, a layered structure of the microbial community, as has been demonstrated in the upper green layer (Ramsing et al., 2000; Becraft et al., 2011), can only be hypothesized for the undermat at this time. Further studies are needed to determine the distribution of the members of the undermat community.

All seven chlorophototrophs identified in previous genomic and metagenomic studies of the upper green layer were also present in the undermat metagenome (Table 5; Klatt et al., 2011; Liu et al., 2011). *Roseiflexus* spp. and “*Candidatus Roseilinea gracile*” showed higher relative abundance in the undermat, whereas the other phototrophs are present in lower relative abundance in comparison to the upper green layer of the mat (Table 5, Table S1). Three additional phototrophic bacteria were detected in the microbial mat for the first time in this study (“*Candidatus Chloranaerofilum corporosum*” OTU-15, as well as two phototrophic *Alphaproteobacteria*, “*Candidatus Elioraea thermophila*” OTU-46, and “*Candidatus Roseovibrio tepidum*” OTU-121; Tank et al., in press). A total of sixteen phototrophic bacterial taxa representing six different phyla have now been

TABLE 5 | Overview of community composition detected by the different methods used in this study (iTag, metagenome 16S rRNA, metagenome binning) and relative abundances in undermat and upper layer iTag sequencing study.

| iTag OTU | Identity | Undermat metagenome 16S rRNA | Undermat metagenomic bin | iTag MS undermat (%) | iTag MS upper green layer (%) | Upper green layer metagenomic bin ^a |
|----------|---|--------------------------------|--------------------------|----------------------|-------------------------------|--|
| OTU-01 | <i>Chloroflexi</i> , <i>Roseiflexus</i> sp., OS Type C | Yes, scaffold 166812 | Yes, bin-1 [very small] | 49.1 | 33.9 | Yes |
| OTU-02 | <i>Thermotogae</i> , <i>Pseudothermotoga</i> sp. | Yes, scaffold 336 | Yes, bin-2 | 10.4 | 1.5 | No |
| OTU-03 | OP10/ <i>Armatimonadetes</i> _Gp7, OS Type L | Yes, scaffold 622461 | No [enrichment bin-3] | 4.5 | 0.8 | Yes |
| OTU-04 | <i>Aquificae</i> , <i>Thermocrinis</i> sp. | Yes, scaffold 2657 | Yes, bin-4 | 4.4 | 0.3 | No |
| OTU-05 | <i>Cyanobacteria</i> , <i>Synechococcus</i> sp. B', OS Type B | Yes, scaffold 8881 | Yes, bin-5 | 3.6 | 37.4 | Yes |
| OTU-06 | <i>Chloroflexi</i> , <i>Anaerolineae</i> -like' phototroph | Yes, scaffold 1696 | Yes, bin-6 | 2.6 | 1.2 | Yes |
| OTU-07 | OP9/ <i>Atribacteria</i> | Yes, scaffold 14288 | Yes, bin-7 | 2.4 | 0.7 | No |
| OTU-08 | <i>Nitrospirae</i> , <i>Thermodesulfobivrio</i> sp. | Yes, scaffold 00748 | Yes, bin-8 | 2.4 | 2.7 | (Yes ^b) |
| OTU-09 | <i>Chloroflexi</i> , <i>Anerolineae</i> | Yes, scaffold 273 | Yes, bin-9 | 1.4 | 0.7 | No |
| OTU-10 | <i>Thermotogae</i> , EM3/OPS-2 | Yes, scaffold 1962 | Yes, bin-10 | 1.2 | 3.1 | Yes |
| OTU-11 | <i>Chloroflexus</i> sp. MS-G | Yes, scaffold 301461 | Yes, bin-11 | 1.1 | 1.2 | Yes |
| OTU-12 | OP10/ <i>Armatimonadetes</i> _Gp6 | Yes, scaffold 3586 | Yes, bin-12 | 1.1 | 0.5 | No |
| OTU-13 | OP-8/ <i>Aminicenantes</i> | Yes, scaffold 32931 | Yes, bin-13 | 1.0 | 0.0 | No |
| OTU-14 | <i>Planctomycetes</i> | Yes, scaffold 3293 | Yes, bin-14 | 0.9 | 0.0 | No |
| OTU-15 | <i>Chloroflexi</i> | Yes, scaffold 218671 | Yes, bin-15 | 0.7 | 0.0 | No |
| OTU-16 | <i>Planctomycetes</i> , <i>Gemmata</i> | Yes, scaffold 261011 | Maybe, bin-23 | 0.7 | 0.2 | No |
| OTU-17 | <i>Chloracidobacterium thermophilum</i> | Yes, scaffold 32584 | Yes, bin-16 | 0.7 | 5.2 | No |
| OTU-18 | <i>Armatimonadetes</i> _Gp 2 | Yes, scaffold 584 | Yes, bin-17 | 0.6 | 0.0 | No |
| OTU-19 | <i>Planctomycetes</i> , uncultured | Yes, scaffold 11289 | No | 0.6 | 0.0 | No |
| OTU-20 | <i>Elusimicrobia</i> , uncultured | Yes, scaffold 20130 | No | 0.5 | 0.1 | Maybe |
| OTU-21 | <i>Deinococcus-Thermus</i> , <i>Meiothermus</i> sp. | Yes, scaffolds 8341 and 25957 | Yes, bin-18 | 0.5 | 0.4 | No |
| OTU-22 | <i>Cyanobacteria</i> , <i>Synechococcus</i> sp. A, OS Type A | Yes, scaffold 85825 | No | 0.5 | 0.5 | No |
| OTU-23 | <i>Chloroflexi</i> , <i>Anaerolineae</i> , OPB12 | Yes, scaffold 18877 | No | 0.5 | 0.0 | No |
| OTU-24 | <i>Chlorobi</i> , Lineage 5: OPB56 | Yes, scaffold 2638 | Yes, bin-19 | 0.5 | 2.0 | No |
| OTU-25 | <i>Spirochaetae</i> , <i>Leptospiraceae</i> | Yes, scaffold 4665 | No | 0.4 | 0.1 | No |
| OTU-26 | <i>Thermodesulfobacteria</i> , <i>Caldimicrobium</i> | Yes, scaffold 193547 | Yes, bin-32 | 0.4 | 0.4 | No |
| OTU-27 | <i>Chlorobi</i> , Lineage 5: OPB56 | Yes, scaffold 104947 | No | 0.4 | 1.0 | No |
| OTU-28 | <i>Thermotogae</i> , EM3 | Yes, scaffold 49539 | No | 0.4 | 0.0 | No |
| OTU-29 | <i>Chlorobi</i> , Lineage 5: OPB56 | Yes, scaffold 6663 | No | 0.4 | 0.3 | No |
| OTU-30 | <i>Thermotogae</i> , EM3 | Yes, scaffold 26263 | No | 0.4 | 0.0 | No |
| OTU-31 | <i>Chloroflexi</i> , <i>Bellilinea</i> | Yes, scaffold 10483 | Yes, bin-25 | 0.3 | 0.1 | No |
| OTU-32 | <i>Spirochaetae</i> , <i>Exilispira</i> | Yes, scaffolds 5322 and 20220 | No | 0.2 | 1.1 | No |
| OTU-33 | OP10/ <i>Armatimonadetes</i> _Gp7, OS Type L | Yes, scaffold 62246 | No | 0.2 | 0.0 | No |
| OTU-34 | <i>Chlorobi</i> , <i>Ignavibacteriaceae</i> | Yes, scaffolds 10292 and 68906 | Maybe, bin-24 | 0.2 | 0.2 | No |
| OTU-35 | <i>Spirochaetae</i> , <i>Brevinema</i> | No | No | 0.2 | 0.3 | No |
| OTU-36 | <i>Acidobacterium</i> , OPB3 | Yes, scaffolds 1343 and 2292 | Yes, bin-20 | 0.2 | 0.0 | No |

(Continued)

TABLE 5 | Continued

| iTag OTU | Identity | Undermat metagenome 16S rRNA | Undermat metagenomic bin | iTag MS undermat (%) | iTag MS upper green layer (%) | Upper green layer metagenomic bin ^a |
|----------|--|----------------------------------|--------------------------|----------------------|-------------------------------|--|
| OTU-37 | <i>Verrucomicrobia</i> , uncultured | Yes, scaffold 25955 | No | 0.2 | 0.0 | No |
| OTU-38 | <i>Chlorobi</i> , “Ca. Thermochlorobacter aerophilum” | Yes, scaffold 761 | Yes, bin-21 | 0.2 | 2.2 | No |
| OTU-39 | <i>Chloroflexi</i> , <i>Anaerolineae</i> | Yes, scaffold 145123 | No | 0.2 | 0.1 | No |
| OTU-40 | <i>Deltaproteobacteria</i> , <i>Desulfarculales</i> , uncultured | Yes, scaffold 112567 | No | 0.2 | 0.0 | No |
| OTU-41 | <i>Spirochaetae</i> , <i>Spirochaeta</i> | Yes, scaffold 19076 | No | 0.2 | 0.1 | No |
| OTU-42 | <i>Acetothermia</i> , uncultured | (Yes, scaffold 169700)* | No | 0.1 | 0.0 | No |
| OTU-43 | <i>Spirochaetae</i> , <i>Leptospiraceae</i> , uncultured | (Yes, scaffold 205406)* | No | 0.1 | 0.0 | No |
| OTU-44 | <i>Deltaproteobacteria</i> , <i>Synthrophorhabdus</i> -like | Yes, scaffolds 233149 and 241554 | No | 0.1 | 0.0 | No |
| OTU-45 | <i>Chlorobi</i> , <i>Ignavibacteriaceae</i> | Yes, scaffolds 104127 and 141772 | Maybe, bin-24 | 0.1 | 0.0 | No |
| OTU-46 | <i>Elioraea</i> sp. | Yes, scaffold 23894 | Yes, bin-22 | 0.1 | 0.0 | No |

*based on phylogenetic analysis, no overlap of sequences.

^aKlatt et al. (2011).

^bno metagenomic bin, but related sequences recruited by reference genomes.

detected in the Mushroom Spring microbial mat (Tank et al., in press). Additionally, the discovery of multiple organisms with genes encoding xanthorhodopsin raises new questions about the role of retinal-based phototrophy (retinalphototrophy; Bryant and Frigaard, 2006) or signaling in the undermat. This will be addressed in more detail elsewhere (Thiel et al., in preparation). The unidentified Cluster 8 previously detected in the upper layer metagenome was identified again here as OTU-10, an organism affiliated with the group EM3, which has tentatively been placed in the phylum *Thermotogae* (Reysenbach et al., 1994; Klatt et al., 2013a). The second unidentified heterotroph previously detected in the upper layer metagenome, Cluster 7 (Klatt et al., 2011), was identified as an *Armatimonadetes* member OTU-3. Due to a high microdiversity of this organism in the microbial mat sample, identification was only possible by a serendipitous finding of a closely related organism in an enrichment culture.

CONCLUSIONS

In this study we analyzed the community composition and diversity of the orange-colored undermat of Mushroom Spring, an alkaline hot spring in YNP (WY, USA) by 16S rRNA gene amplicon and metagenomic analyses. Despite a long history of research on the microbial mats at Mushroom and Octopus Springs (Brock, 1967; Ward et al., 1998, 2012; Kim et al., 2015), these mats still harbor the potential for many novel discoveries. Members of the genus *Roseiflexus* dominated a fairly diverse but uneven microbial community, and metagenomic analysis identified several novel organisms with unusual traits. Many unidentified 16S rRNA sequences recovered from these environments in previous studies were detected and phylogenetically identified. Other organisms, which have been cultured from either Mushroom or Octopus Spring,

were not detected, once again illustrating the inherent bias of untargeted cultivation experiments. A more detailed analysis of the metagenome, focusing on the metabolic potential of the mat members and their putative interactions, will be published elsewhere (Thiel et al., in preparation). Studies of microbial ecology, diversity, species evolution and interspecies interactions are still subjects of ongoing research with many open questions to be addressed. Comparisons of species in both upper and lower mat and a diel-transcriptomic analysis that will hopefully reveal gene expression activity within the undermat community that will allow us to distinguish between active and inactive members of the community defined in this study, and should provide information on the temporal pattern of gene expression in the undermat. Depth-dependent distributions of OTU populations that may represent putative ecotypes will also be addressed in future studies.

ACCESSION NUMBERS

16S rRNA gene sequences of iTag OTUs as well as assembled clone sequences have been deposited in GenBank (Acc. nos. KU860141–KU860455 [iTag OTUs]; KX213895–KX214032 [clone OTUs]). Complete metagenome data are available in the Integrated Microbial Genomes with Microbiome Samples (IMG/M, <https://img.jgi.doe.gov/>) database, taxon object IDs 3300002493, 3300005452 and 2015219002.

AUTHOR CONTRIBUTIONS

VT conducted sequence analysis after assembly for both amplicon and metagenome sequences, including phylogenetic analysis and phylogenetic marker genes analysis of metagenome bins. JW conducted initial tetranucleotide binning analyses,

reference targeted mapping studies and contributed to discussion and manuscript. Sampling and DNA extraction from the hot-spring microbial mat and enrichment cultures was conducted by MO, who also wrote corresponding sections in the manuscript and contributed to the discussion of results. MT isolated and identified all cultures mentioned in the manuscript, contributed to writing the manuscript and discussing the results. CK conducted 16S rRNA cloning and sequencing from undermat samples from a previous time point, analyzed those sequences and contributed to manuscript and discussion. Sequencing, quality check, assembly and dereplication of amplicon and the metagenome was conducted by JGI staff. DW and DB planned the experiments, acquired funding, organized and led field excursions and provided scientific infrastructure. VT, DW, and DB wrote the manuscript.

FUNDING

This study was partly funded by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the Department of Energy through Grant DE-FG02-94ER20137. DB and DW additionally acknowledge support from the NASA Exobiology program (NX09AM87G). This work was also partly supported by the U. S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program 395 (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL) under a subcontract to DB. The nucleotide sequencing was performed as part of a Community Sequencing Program (Project CSP-411) and was performed by the U.S. Department of Energy Joint Genome Institute, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

ACKNOWLEDGMENTS

The authors would like to thank all of the JGI staff members who contributed to obtaining the sequence data. The materials

used in this study were collected under permit #YELL-SCI-0129 held by DW and administered under the authority of Yellowstone National Park. The authors especially thank Christie Hendrix and Stacey Gunther for their advice and assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00919>

Figure S1 | Phylogenetic tree based on 16S rRNA gene sequences showing the phylogenetic relationship between sequences obtained from the Mushroom Spring microbial undermat community and cultivated and uncultivated relatives in the phyla Armatimonadetes OP10, Aminicenantes OP8, Nitrospirae, Thermodesulfobacteria, Microgenomates OP11, Aquificae, SM2F11 and Archaea (A), Planctomycetes, Verrucomicrobia, Spirochaeta, Acidobacteria, Atribacteria OP9/JS1, Elusimicrobia and Cyanobacteria (B), and Proteobacteria, Thermotogae/EM3 and Thermi (C).

The trees were generated based on the Maximum Likelihood method using the phyML software included in the ARB package. Percentage numbers on nodes refer to 100 bootstrap pseudoreplicates conducted. Only values >50% are shown. Bold sequences were obtained from Mushroom or Octopus Spring in this or previous studies. Red bold labels indicate sequences obtained in this study. Blue bold labels indicate "OS type" sequences from previous studies. OTU numbers shown refer to the most abundant OTU represented by the sequence. Only sequences with length >1,000 bp were used for phylogenetic calculations. Sequence length <1,000 bp are given in (gray) in the labels and corresponding sequences were added using the Parsimony method without changing tree topology.

Figure S2 | Entropy distribution in *Roseiflexus*-like 16S rRNA gene amplicon sequences in the combined dataset (A), and the undermat (B), and upper layer (C). The difference in the entropy figures from upper layer (C) and undermat (B) analyzed separately, specifically the considerably lower entropy at pos. 104 and 109 in upper layer, are indicative of a lower abundance of two oligotypes in the upper layer, namely "CCCCGCGTGC" (2.13% in undermat, 0.19% in upper layer) and "CCCCGCGGGC" (1.02 vs. 0.21%) (Table S2).

Table S1 | OTUs obtained from 16S rRNA V4 iTag sequencing. Read numbers, relative abundance and number of total and abundant dereplicated iTag sequences are stated. Classification are based on RDP classifier.

Table S2 | Read counts and relative abundance of the 23 most abundant *Roseiflexus*-like oligotypes in undermat and upper layer as determined in the combined amplicon dataset (>100 reads total).

REFERENCES

- Albuquerque, L., Rainey, F. A., Nobre, M. F., and da Costa, M. S. (2008). *Elioraea tepidiphila* gen. nov., sp. nov., a slightly thermophilic member of the Alphaproteobacteria. *Int. J. Syst. Evol. Microbiol.* 58, 773–778. doi: 10.1099/ijs.0.65294-0
- Allewalt, J. P., Bateson, M. M., Revsbech, N. P., Slack, K., and Ward, D. M. (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. doi: 10.1128/AEM.72.1.544
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Amann, R. L., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169.
- Anderson, K. L., Tayne, T. A., and Ward, D. M. (1987). Formation and fate of fermentation products in hot spring cyanobacterial mats. *Appl. Environ. Microbiol.* 53, 2343–2352.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Bateson, M. M., and Ward, D. M. (1988). Photoexcretion and fate of glycolate in a hot spring cyanobacterial mat. *Appl. Environ. Microbiol.* 54, 1738–1743.
- Bauld, J., and Brock, T. D. (1973). Ecological studies of *Chloroflexus*, a gliding photosynthetic bacterium. *Arch. Mikrobiol.* 92, 267–284. doi: 10.1007/BF00409281
- Becraft, E. D., Cohan, F. M., Kuehl, M., Jensen, S. I., and Ward, D. M. (2011). Fine-scale distribution patterns of *Synechococcus* ecological diversity in the microbial mat of Mushroom Spring, Yellowstone National Park. *Appl. Env. Microbiol.* 77, 7689–7697. doi: 10.1128/AEM.05927-11
- Becraft, E. D., Wood, J. M., Rusch, D. B., Kühl, M., Jensen, S. I., Bryant, D. A., et al. (2015). The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative

- ecotypes of *Synechococcus* inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. *Front. Microbiol.* 6:590. doi: 10.3389/fmicb.2015.00590
- Bhandari, V., and Gupta, R. S. (2014). Molecular signatures for the phylum (class) *Thermotogae* and a proposal for its division into three orders (*Thermotogales*, *Kosmotogales* ord. nov. and *Petrotogales* ord. nov.) containing four families (*Thermotogaceae*, *Fervidobacteriaceae* fam. nov., *Kosmotogaceae* fam. nov. and *Petrotogaceae* fam. nov.) and a new genus *Pseudothermotoga* gen. nov. with five new combinations. *Antonie van Leeuwenhoek* 105, 143–168. doi: 10.1007/s10482-013-0062-7
- Bhatnagar, S., Badger, J. H., Madupu, R., Khouri, H. M., Connor, E. M. O., Robb, F. T., et al. (2015). Genome sequence of the sulfate-reducing thermophilic bacterium *Thermodesulfovibrio yellowstonii* strain DSM 11347^T (Phylum *Nitrospirae*). *Genome Announc.* 3, e01489–e01414. doi: 10.1128/genomeA.01489-14
- Bhaya, D., Grossman, A. R., Steunou, A.-S., Khuri, N., Cohan, F. M., Hamamura, N., et al. (2007). Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J.* 1, 703–713. doi: 10.1038/ismej.2007.46
- Boomer, S. M., Lodge, D. P., Dutton, B. E., and 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. doi: 10.1128/AEM.68.1.346-355.2002
- Boomer, S. M., Pierson, B. K., Austinhirst, R., and Castenholz, R. W. (2000). Characterization of novel bacteriochlorophyll a-containing red filaments from alkaline hot springs in Yellowstone National Park. *Arch. Microbiol.* 174, 152–161. doi: 10.1007/s002030000189
- Brock, T. D. (1967). Micro-organisms adapted to high temperatures. *Nature* 214, 882–885.
- Brock, T. D., and Freeze, H. (1969). *Thermus aquaticus* gen. nov. and sp. nov., a non-sporulating extreme thermophile. *J. Bacteriol.* 98, 289–297.
- 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. doi: 10.1126/science.1143236
- Bryant, D. A., and Frigaard, N.-U. (2006). Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol.* 14, 488–496. doi: 10.1016/j.tim.2006.09.001
- Caporaso, J. G., Bittinger, K., and Bushman, F. D., et al. (2010b). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26, 266–267. doi: 10.1093/bioinformatics/btp636
- Caporaso, J. G., Kuczynski, J., and Stombaugh, J., et al. (2010a). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7, 335–336. doi: 10.1038/nmeth.f.303.QIIME
- Castenholz, R. W. (1969). Thermophilic blue-green algae and the thermal environment. *Bacteriol. Rev.* 33, 476–504.
- Challacombe, J. F., Eichorst, S. A., Hauser, L., Land, M., Xie, G., and Kuske, C. R. (2011). Biological consequences of ancient gene acquisition and duplication in the large genome of “*Candidatus* Solibacter usitatus” Ellin6076. *PLoS ONE* 6:e24882. doi: 10.1371/journal.pone.0024882
- Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* 11, 265–270.
- Choi, A. R., Shi, L., Brown, L. S., and Jung, K.-H. (2014). Cyanobacterial light-driven proton pump, *Gloeobacter* rhodopsin: complementarity between rhodopsin-based energy production and photosynthesis. *PLoS ONE* 9:e110643. doi: 10.1371/journal.pone.0110643
- Cohan, F. M., and Perry, E. B. (2007). A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* 17, 373–386. doi: 10.1016/j.cub.2007.03.032
- Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucl. Acids Res.* 37, D141–D145. doi: 10.1093/nar/gkn879
- De La Torre, J. R., Walker, C. B., Ingalls, A. E., Könneke, M., and Stahl, D. A. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ. Microbiol.* 10, 810–818. doi: 10.1111/j.1462-2920.2007.01506.x
- Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A. P., et al. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* 10, R85. doi: 10.1186/gb-2009-10-8-r85
- Dillon, J. G., Fishbain, S., Miller, S. R., Bebout, B. M., Habicht, K. S., Webb, S. M., et al. (2007). High rates of sulfate reduction in a low-sulfate hot spring microbial mat are driven by a low level of diversity of sulfate-respiring microorganisms. *Appl. Environ. Microbiol.* 73, 5218–5226. doi: 10.1128/AEM.00357-07
- Dodsworth, J. A., Blainey, P. C., Murugapiran, S. K., Wesley, D., Ross, C. A., Tringe, S. G., et al. (2013). Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for member of the OP9 lineage. *Nat Commun.* 4, 1854. doi: 10.1038/ncomms2884
- Doemel, W. N., and Brock, T. D. (1977). Structure, growth, and decomposition of laminated algal-bacterial mats in alkaline hot springs. *Appl. Environ. Microbiol.* 34, 433–452.
- Eder, W., and Huber, R. (2002). New isolates and physiological properties of the *Aquificales* and description of *Thermocrinis albus* sp. nov. *Extremophiles* 6, 309–318. doi: 10.1007/s00792-001-0259-y
- Farag, I. F., Davis, J. P., Youssef, N. H., and Elshahed, M. S. (2014). Global patterns of abundance, diversity and community structure of the *Aminicenantes* (Candidate phylum OP8). *PLoS ONE* 9:e92139. doi: 10.1371/journal.pone.0092139
- Fardeau, M.-L., Ollivier, B., Patel, B. K. C., Magot, M., Thomas, P., and Rimbault, A. (1997). *Thermotoga hypogea* sp. nov., a xylanolytic, thermophilic bacterium from an oil-producing well. *Int. J. Syst. Bacteriol.* 47, 1013–1019.
- Ferris, M. J., Kocczynski, E. D., Bateson, M. M., and Ward, D. M. (1996a). Enrichment culture and microscopy conceal diverse thermophilic *Synechococcus* populations in a single hot spring mat habitat. *Microbiology* 62, 1045–1050.
- Ferris, M. J., Kuehl, M., Wieland, A., and Ward, D. M. (2003). Cyanobacterial ecotypes in different optical microenvironments of a 68°C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. *Appl. Environ. Microbiol.* 69, 2893–2898. doi: 10.1128/AEM.69.5.2893
- Ferris, M. J., Muyzer, G., and Ward, D. M. (1996b). Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Appl. Environ. Microbiol.* 62, 340–346.
- Ferris, M. J., Nold, S. C., Revsbech, N. P., and Ward, D. M. (1997). Population structure and physiological changes within a hot spring microbial mat community following disturbance. *Appl. Environ. Microbiol.* 63, 1367–1374.
- Ferris, M. J., and Ward, D. M. (1997). Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 63, 1375–1381.
- Garcia Costas, A. M., Liu, Z., Tomsho, L. P., Schuster, S. C., Ward, D. M., and Bryant, D. A. (2012a). Complete genome of “*Candidatus* Chloracidobacterium thermophilum”, a chlorophyll-based photoheterotroph belonging to the phylum *Acidobacteria*. *Environ. Microbiol.* 14, 177–190. doi: 10.1111/j.1462-2920.2011.02592.x
- Garcia Costas, A. M., Tsukatani, Y., Rijpsstra, W. I. C., Schouten, S., Welander, P. V., Summons, R. E., et al. (2012b). Identification of the bacteriochlorophylls, carotenoids, quinones, lipids, and hopanoids of “*Candidatus* Chloracidobacterium thermophilum.” *J. Bacteriol.* 194, 1158–1168. doi: 10.1128/JB.06421-11
- Grégoire, P., Fardeau, M.-L., Joseph, M., Guasco, S., Hamaide, F., Biasutti, S., et al. (2011). Isolation and characterization of *Thermanaerothermophilus daxensis* gen. nov., sp. nov., a thermophilic anaerobic bacterium pertaining to the phylum Chloroflexi, isolated from a deep hot aquifer in the Aquitaine Basin. *Syst. Appl. Microbiol.* 34, 494–497. doi: 10.1016/j.syapm.2011.02.004
- Gupta, R., Chander, P., and George, S. (2013). Phylogenetic framework and molecular signatures for the class Chloroflexi and its different clades: proposal for division of the class *Chloroflexi* class. nov. into the suborder Chloroflexineae subord. nov., consisting of the emended family Oscillochloridaceae and the family Chloroflexaceae fam. nov., and the suborder Roseiflexinae subord. nov., containing the family Roseiflexaceae fam. nov. *Antonie van Leeuwenhoek* 103, 99–119. doi: 10.1007/s10482-012-9790-3
- Henry, E. A., Devereux, R., Maki, J. S., Gilmour, C. C., Woese, C. R., Mandelco, L., et al. (1994). Characterization of a new thermophilic sulfate-reducing bacterium *Thermodesulfovibrio yellowstoneii*, gen. nov. and sp. nov.: its

- phylogenetic relationship to *Thermodesulfobacterium commune* and their origins deep within the bacterial domain. *Arch. Microbiol.* 161, 62–69.
- Hiras, J., Wu, Y.-W., Eichorst, S. A., Simmons, B. A., and Singer, S. W. (2015). Refining the phylum *Chlorobi* by resolving the phylogeny and metabolic potential of the representative of a deeply branching, uncultivated lineage. *ISME J.* 9, 1751–17370. doi: 10.1038/ismej.2015.158
- Huber, R., Eder, W., Heldwein, S., Huber, H., Rachel, R., Karl, O., et al. (1998). *Thermocrinis ruber* gen. nov., sp. nov., a pink-filament-forming hyperthermophilic bacterium isolated from Yellowstone National Park. *Appl. Environ. Microbiol.* 64, 3576–3583.
- Hugenholtz, P., Goebel, B. M., and Pace, N. R. (1998a). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180, 4765–4774.
- Hugenholtz, P., and Pace, N. R. (1996). Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends Biotechnol.* 14, 190–197. doi: 10.1016/0167-7799(96)10025-1
- Hugenholtz, P., Pitulle, C., Hershberger, K. L., and Pace, N. R. (1998b). Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* 180, 366–376.
- Iino, T., Mori, K., Uchino, Y., Nakagawa, T., Harayama, S., and Suzuki, K.-I. (2010). *Ignavibacterium album* gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from microbial mats at a terrestrial hot spring and proposal of *Ignavibacteria classis* nov., for a novel lineage at the periphery of green sulfur bacteria. *Int. J. Syst. Evol. Microbiol.* 69, 1376–1382. doi: 10.1099/ijs.0.012484-0
- Im, W.-T., Hu, Z.-Y., Kim, K.-H., Rhee, S.-K., Meng, H., Lee, S.-T., et al. (2012). Description of *Fimbrimonas ginsengisoli* gen. nov., sp. nov. within the *Fimbrimonadidae* class nov., of the phylum *Armatimonadetes*. *Antonie Van Leeuwenhoek* 102, 307–317. doi: 10.1007/s10482-012-9739-6
- Jackson, T. J., Ramaley, R. F., and Meinschein, W. G. (1973). *Thermomicrobium*, a new genus of extremely thermophilic bacteria. *Int. J. Syst. Bacteriol.* 23, 28–36. doi: 10.1099/00207713-23-1-28
- Jensen, S. I., Steunou, A.-S., Bhaya, D., Kühl, M., and Grossman, A. R. (2011). *In situ* dynamics of O₂, pH and cyanobacterial transcripts associated with CCM, photosynthesis and detoxification of ROS. *ISME J.* 5, 317–328. doi: 10.1038/ismej.2010.131
- Kadnikov, V. V., Mardanov, A. V., Podosokorskaya, O. A., Gavrillov, S. N., Kublanov, I. V., Beletsky, A. V., et al. (2013). Genomic analysis of *Melioribacter roseus*, facultatively anaerobic organotrophic bacterium representing a novel deep lineage within *Bacteroidetes/Chlorobi* group. *PLoS ONE* 8:e53047. doi: 10.1371/journal.pone.0053047
- Kerepesi, C., Bánky, D., and Grolmusz, V. (2014). AmphoraNet: the webserver implementation of the AMPHORA2 metagenomic workflow suite. *Gene* 533, 538–540. doi: 10.1016/j.gene.2013.10.015
- Kim, Y.-M., Nowack, S., Olsen, M. T., Becraft, E. D., Wood, J. M., Thiel, V., et al. (2015). Diel metabolomics analysis of a hot spring chlorophototrophic microbial mat leads to new hypotheses of community member metabolisms. *Front. Microbiol.* 6:209. doi: 10.3389/fmicb.2015.00209
- Kimura, H., Nashimoto, H., Shimizu, M., Hattori, S., Yamada, K., Koba, K., et al. (2010). Microbial methane production in deep aquifer associated with the accretionary prism in Japan. *ISME J.* 4, 531–541. doi: 10.1038/ismej.2009.132
- Klatt, C. G., Bryant, D. A., and Ward, D. M. (2007). Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats. *Environ. Microbiol.* 9, 2067–2078. doi: 10.1111/j.1462-2920.2007.01323.x
- Klatt, C. G., Inskeep, W. P., Herrgard, M. J., Jay, Z. J., Rusch, D. B., Tringe, S. G., et al. (2013a). Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front. Microbiol.* 4:106. doi: 10.3389/fmicb.2013.00106
- Klatt, C. G., Liu, Z., Ludwig, M., Kühl, M., Jensen, S. I., Bryant, D. A., et al. (2013b). Temporal metatranscriptomic patterning in phototrophic *Chloroflexi* inhabiting a microbial mat in a geothermal spring. *ISME J.* 7, 1775–1789. doi: 10.1038/ismej.2013.52
- Klatt, C. G., Wood, J. M., Rusch, D. B., Bateson, M. M., Hamamura, N., Heidelberg, J. F., et al. (2011). Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J.* 5, 1262–1278. doi: 10.1038/ismej.2011.73
- Koeppel, A. F., and Wu, M. (2013). Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucl. Acids Res.* 41, 5175–5188. doi: 10.1093/nar/gkt241
- Lau, M. C. Y., Aitchison, J. C., and Pointing, S. B. (2009). Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles* 13, 139–149. doi: 10.1007/s00792-008-0205-3
- Lee, K. C.-Y., Dunfield, P. F., Morgan, X. C., Crowe, M. A., Houghton, K. M., Vyssotski, M., 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 OP1. *Int. J. Syst. Evol. Microbiol.* 61, 2482–2490. doi: 10.1099/ijs.0.027235-0
- Lee, K. C.-Y., Herbold, W., Dunfield, P. F., Morgan, X. C., McDonald, I. R., and Stott, M. B. (2013). Phylogenetic delineation of the novel phylum *Armatimonadetes* (former candidate division OP10) and definition of two novel candidate divisions. *Appl. Environ. Microbiol.* 79, 2484–2487. doi: 10.1128/AEM.03333-12
- Liu, Z., Klatt, C. G., Ludwig, M., Rusch, D. B., Jensen, S. I., Kühl, M., et al. (2012). “*Candidatus Thermochlorobacter aerophilum*,” an aerobic chlorophototrophic member of the phylum *Chlorobi* defined by metagenomics and metatranscriptomics. *ISME J.* 6, 1869–1882. doi: 10.1038/ismej.2012.24
- Liu, Z., Klatt, C. G., Wood, J. M., Rusch, D. B., Ludwig, M., Wittekindt, N., et al. (2011). Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. *ISME J.* 5, 1279–1290. doi: 10.1038/ismej.2011.37
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, et al. (2004). ARB: a software environment for sequence data. *Nucl. Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293
- Markowitz, V. M., Chen, I. M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., et al. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 42, 568–573. doi: 10.1093/nar/gkt919
- Melendrez, M. C., Lange, R. K., Cohan, F. M., and Ward, D. M. (2011). Influence of molecular resolution on sequence-based discovery of ecological diversity among *Synechococcus* populations in an alkaline siliceous hot spring microbial mat. *Appl. Environ. Microbiol.* 77, 1359–1367. doi: 10.1128/AEM.02032-10
- Meyer-Dombard, D. R., Swingle, W., Raymond, J., Havig, J., Shock, E. L., and Summons, R. E. (2011). Hydrothermal ecotones and streamer biofilm communities in the Lower Geyser Basin, Yellowstone National Park. *Environ. Microbiol.* 13, 2216–2231. doi: 10.1111/j.1462-2920.2011.02476.x
- Nobu, M. K., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., Gies, E. A., Webster, G., et al. (2016). Phylogeny and physiology of candidate phylum “*Atribacteria*” (OP9/JS1) inferred from cultivation-independent genomics. *ISME J.* 10, 273–286. doi: 10.1038/ismej.2015.97
- Nold, S. C., and Ward, D. M. (1996). Photosynthate partitioning and fermentation in hot spring microbial mat communities. *Appl. Environ. Microbiol.* 62, 4598–4607.
- Nowack, S., Olsen, M. T., Schaible, G. A., Becraft, E. D., Shen, G., Klapper, I., et al. (2015). The molecular dimension of microbial species: 2. *Synechococcus* strains representative of putative ecotypes inhabiting different depths in the Mushroom Spring microbial mat exhibit different adaptive and acclimative responses to light. *Front. Microbiol.* 6:626. doi: 10.3389/fmicb.2015.00626
- Nowack, S. P. (2014). *Niche Character in a Temporally Varying Environment*. Ph. D. thesis, Montana State University. doi: 10.1017/CBO9781107415324.004
- Nübel, U., Bateson, M. M., Vandieken, V., Kühl, M., and Ward, D. M. (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. doi: 10.1128/AEM.68.9.4593
- Nunoura, T., Hirai, M., Miyazaki, M., Kazama, H., Makita, H., Hirayama, H., et al. (2013). Isolation and characterization of a thermophilic, obligately anaerobic and heterotrophic marine *Chloroflexi* bacterium from a *Chloroflexi*-dominated microbial community associated with a Japanese shallow hydrothermal system, and proposal for *Thermomarinilinea lacunofontalis* gen. nov., sp. nov. *Microbes Environ.* 28, 228–235. doi: 10.1264/jsme2.ME12193
- Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., et al. (2011). Insights into the evolution of *Archaea* and eukaryotic protein modifier

- systems revealed by the genome of a novel archaeal group. *Nucl. Acids Res.* 39, 3204–3223. doi: 10.1093/nar/gkq1228
- Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., and Stahl, D. A. (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* 40, 337–365.
- Olsen, M. T., Nowack, S., Wood, J. M., Becraft, E. D., LaButti, K., Lipzen, A., et al. (2015). The molecular dimension of microbial species: 3. Comparative genomics of *Synechococcus* strains with different light responses and in situ diel transcription patterns of associated putative ecotypes in the Mushroom Spring microbial mat. *Front. Microbiol.* 6:604. doi: 10.3389/fmicb.2015.00604
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, 1–9. doi: 10.1093/nar/gkt1226
- Pace, N. R. (2009). Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* 73, 565–576. doi: 10.1128/MMBR.00033-09
- 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. *Appl. Environ. Microbiol.* 66, 1038–1049. doi: 10.1128/AEM.66.3.1038-1049.2000
- Revsbech, N. P., and Ward, D. M. (1984). Microelectrode studies of interstitial water chemistry and photosynthetic activity in a hot spring microbial mat. *Appl. Environ. Microbiol.* 48, 270–275.
- Reysenbach, A., Longnecker, K., and Kirshtein, J. (2000). Novel bacterial and archaeal lineages from an *in situ* growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Appl. Environ. Microbiol.* 66, 3798–3806. doi: 10.1128/AEM.66.9.3798-3806.2000
- Reysenbach, A., Wickham, G. S., and Pace, N. R. (1994). Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl. Environ. Microbiol.* 60, 2113–2119.
- Sandbeck, K. A., and Ward, D. M. (1981). Fate of immediate methane precursors in low-sulfate, hot-spring algal-bacterial mats. *Appl. Environ. Microbiol.* 41, 775–782.
- Sandbeck, K. A., and Ward, D. M. (1982). Temperature adaptations in the terminal processes of anaerobic decomposition of yellowstone national park and icelandic hot spring microbial mats. *Appl. Environ. Microbiol.* 44, 844–851.
- Schloss, P. D., and Handelsman, J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* 71, 1501–1506. doi: 10.1128/AEM.71.3.1501
- Sekiguchi, Y., Yamada, T., Hanada, S., Ohashi, A., Harada, H., and Kamagata, Y. (2003). *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. *Int. J. Syst. Evol. Microbiol.* 53, 1843–1851. doi: 10.1099/ijso.0.02699-0
- Slobodkina, G. B., Kovaleva, O. L., Miroshnichenko, M. L., Slobodkin, A. I., Kolganova, T. V., Novikov, A. A., et al. (2014). *Thermogutta terrifontis* gen. nov., sp. nov. and *Thermogutta hypogea* sp. nov., thermophilic anaerobic representatives of the phylum Planctomycetes. *Int. J. Syst. Evol. Microbiol.* 65, 760–765. doi: 10.1099/ijso.0.000009
- Stackebrandt, E., and Goebel, B. M. (1994). Taxonomic Note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44, 846–849. doi: 10.1099/00207713-44-4-846
- Steunou, A.-S., Jensen, S. I., Brecht, E., Becraft, E. D., Bateson, M. M., Kilian, O., et al. (2008). Regulation of *nif* gene expression and the energetics of N₂ fixation over the diel cycle in a hot spring microbial mat. *ISME J.* 2, 364–378. doi: 10.1038/ismej.2007.117
- Tamaki, H., Tanaka, Y., Matsuzawa, H., Muramatsu, M., Meng, X.-Y., Hanada, S., et al. (2011). *Armatimonas rosea* gen. nov., sp. nov., of a novel bacterial phylum, *Armatimonadetes* phyl. nov., formally called the candidate phylum OP10. *Int. J. Syst. Evol. Microbiol.* 61, 1442–1447. doi: 10.1099/ijso.0.025643-0
- Tank, M., and Bryant, D. A. (2015a). *Chloracidobacterium thermophilum* gen. nov., sp. nov.: an anoxygenic microaerophilic chlorophototrophic acidobacterium. *Int. J. Syst. Evol. Microbiol.* 65, 1426–1430. doi: 10.1099/ijso.0.000113
- Tank, M., and Bryant, D. A. (2015b). Nutrient requirements and growth physiology of the phototrophic acidobacterium, *Chloracidobacterium thermophilum*. *Front. Microbiol.* 6:226. doi: 10.3389/fmicb.2015.00226
- Tank, M., Thiel, V., Ward, D. M., and Bryant, D. A. (in press). “A panoply of phototrophs: a photomicrographic overview of chlorophototrophs found in the microbial mats of alkaline siliceous hot springs in Yellowstone National Park, WY, USA,” in *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects*, ed P. C. Hallenbeck (Berlin: Springer).
- Thiel, V., Hamilton, T. L., Tomsho, L. P., Burhans, R., Gay, S. E., Ramaley, R. F., et al. (2014a). Draft genome sequence of the moderately thermophilic bacterium *Schleiferia thermophila* strain Yellowstone (*Bacteroidetes*). *Genome Announc.* 2, e00860–e00814. doi: 10.1128/genomeA.00860-14
- Thiel, V., Hamilton, T. L., Tomsho, L. P., Burhans, R., Gay, S. E., Schuster, S. C., et al. (2014b). Draft genome sequence of a sulfide-oxidizing, autotrophic filamentous anoxygenic phototrophic bacterium, *Chloroflexus* sp. strain MS-G (*Chloroflexi*). *Genome Announc.* 2, e00872–e00814. doi: 10.1128/genomeA.00872-14
- Thiel, V., Tomsho, L. P., Burhans, R., Gay, S. E., Schuster, S. C., Ward, D. M., et al. (2015). Draft genome sequence of the *Deinococcus-Thermus* bacterium *Meiothermus ruber* strain A. *Genome Announc.* 3, e00202–e00215. doi: 10.1128/genomeA.00202-15
- van der Meer, M. T. J., Klatt, C. G., Wood, J., Bryant, D. A., Bateson, M. M., Lammerts, L., et al. (2010). Cultivation and genomic, nutritional, and lipid biomarker characterization of *Roseiflexus* strains closely related to predominant in situ populations inhabiting Yellowstone hot spring microbial mats. *J. Bacteriol.* 192, 3033–3042. doi: 10.1128/JB.01610-09
- van der Meer, M. T. J., Schouten, S., Damsté, J. S. S., and Ward, D. M. (2007). Impact of carbon metabolism on 13C signatures of cyanobacteria and green non-sulfur-like bacteria inhabiting a microbial mat from an alkaline siliceous hot spring in Yellowstone National Park (USA). *Environ. Microbiol.* 9, 482–491. doi: 10.1111/j.1462-2920.2006.01165.x
- van der Meer, M. T. J., Schouten, S., Jaap, S., Damsté, S., Leeuw, J. W., De Ward, D. M., et al. (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. doi: 10.1128/AEM.69.10.6000-6006.2003
- van der Meer, M. T. J., Schouten, S., Mary, 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. *Appl. Environ. Microbiol.* 71, 3978–3986. doi: 10.1128/AEM.71.7.3978-3986.2005
- Venkata Ramana, V., Sasikala, C., Takaichi, S., and Ramana, C. V. (2010). *Roseomonas aestuarii* sp. nov., a bacteriochlorophyll-a containing alphaproteobacterium isolated from an estuarine habitat of India. *Syst. Appl. Microbiol.* 33, 198–203. doi: 10.1016/j.syapm.2009.09.004
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07
- Ward, D. M. (1978). Thermophilic methanogenesis in a hot-spring algal-bacterial mat (71 to 30 degrees C). *Appl. Environ. Microbiol.* 35, 1019–1026.
- Ward, D. M. (1998). A natural species concept for prokaryotes. *Curr. Opin. Microbiol.* 1, 271–277.
- Ward, D. M., Bateson, M. M., Weller, R., and Ruff-Roberts, A. L. (1992). “Ribosomal RNA analysis of microorganisms as they occur in nature,” in *Advances in Microbial Ecology*, ed K. C. Marshall (New York, NY: Springer US), 219–286.
- Ward, D. M., Brassell, S. C., and Eglinton, G. (1985). Archaeobacterial lipids in hot-spring microbial mats. *Nature* 318, 162–163. doi: 10.1038/314731a0
- Ward, D. M., Castenholz, R. W., and Miller, S. R. (2012). “Ecology of cyanobacteria II: cyanobacteria in geothermal habitats,” in *Ecology of Cyanobacteria II*, ed B. A. Whitton (Dordrecht: Springer), 39–63.
- Ward, D. M., and Cohan, F. M. (2005). “Microbial diversity in hot spring cyanobacterial mats: pattern and prediction,” in *Geothermal Biology and Geochemistry in Yellowstone National Park*, eds W. P. Inskeep and T. McDermott (Bozeman: Thermal Biology Institute), 185–202.

- 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. *Microbiol. Mol. Biol. Rev.* 62, 1353–1370.
- Ward, D. M., Santegoeds, C. M., Nold, S. C., Ramsing, N. B., Ferris, M. J., and Bateson, M. M. (1997). Biodiversity within hot spring microbial mat communities: molecular monitoring of enrichment cultures. *Antonie van Leeuwenhoek* 71, 143–150. doi: 10.1023/A:1000131426164
- Ward, D. M. J., Shiea, J., Zeng, Y. B., Dobson, G., Brassell, S., and Eglinton, G. (1989). "Lipid biochemical markers and the composition of microbial mats," in *Microbial Mats: Physiological Ecology of Benthic Microbial Communities*, eds Y. Cohen and E. Rosenberg (Washington, DC: American Society For Microbiology), 439–454.
- Ward, D. M., Weller, R., and Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 346, 183–187. doi: 10.1038/346183a0
- Weller, R., Bateson, M. M., Heimbuch, B. K., Kopczynski, E. D., and Ward, D. M. (1992). Uncultivated cyanobacteria, *Chloroflexus*-like inhabitants, and spirochete-like inhabitants of a hot spring microbial mat. *Appl. Environ. Microbiol.* 58, 3964–3969.
- Wu, D., Raymond, J., Wu, M., Chatterji, S., Ren, Q., Graham, J. E., et al. (2009). Complete genome sequence of the aerobic CO-oxidizing thermophile *Thermomicrobium roseum*. *PLoS ONE* 4:e4207. doi: 10.1371/journal.pone.0004207
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Thiel, Wood, Olsen, Tank, Klatt, Ward and Bryant. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Metagenomic Analysis of Hot Springs in Central India Reveals Hydrocarbon Degrading Thermophiles and Pathways Essential for Survival in Extreme Environments

Rituja Saxena^{1†}, Darshan B. Dhakan^{1†}, Parul Mittal^{1†}, Prashant Waiker¹, Anirban Chowdhury², Arundhuti Ghatak² and Vineet K. Sharma^{1*}

¹ Metagenomics and Systems Biology Laboratory, Department of Biological Sciences, Indian Institute of Science Education and Research, Bhopal, India, ² Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research, Bhopal, India

OPEN ACCESS

Edited by:

Kian Mau Goh,
Universiti Teknologi Malaysia, Malaysia

Reviewed by:

Alexandre Soares Rosado,
Federal University of Rio de Janeiro,
Brazil

Mariana Lozada,
CESIMAR (CENPAT-CONICET),
Argentina

*Correspondence:

Vineet K. Sharma
vineetks@iiserb.ac.in

[†] These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 05 August 2016

Accepted: 15 December 2016

Published: 05 January 2017

Citation:

Saxena R, Dhakan DB, Mittal P,
Waiker P, Chowdhury A, Ghatak A and
Sharma VK (2017) Metagenomic
Analysis of Hot Springs in Central
India Reveals Hydrocarbon Degrading
Thermophiles and Pathways Essential
for Survival in Extreme Environments.
Front. Microbiol. 7:2123.
doi: 10.3389/fmicb.2016.02123

Extreme ecosystems such as hot springs are of great interest as a source of novel extremophilic species, enzymes, metabolic functions for survival and biotechnological products. India harbors hundreds of hot springs, the majority of which are not yet explored and require comprehensive studies to unravel their unknown and untapped phylogenetic and functional diversity. The aim of this study was to perform a large-scale metagenomic analysis of three major hot springs located in central India namely, Badi Anthoni, Chhoti Anthoni, and Tattapani at two geographically distinct regions (Anthoni and Tattapani), to uncover the resident microbial community and their metabolic traits. Samples were collected from seven distinct sites of the three hot spring locations with temperature ranging from 43.5 to 98°C. The 16S rRNA gene amplicon sequencing of V3 hypervariable region and shotgun metagenome sequencing uncovered a unique taxonomic and metabolic diversity of the resident thermophilic microbial community in these hot springs. Genes associated with hydrocarbon degradation pathways, such as benzoate, xylene, toluene, and benzene were observed to be abundant in the Anthoni hot springs (43.5–55°C), dominated by *Pseudomonas stutzeri* and *Acidovorax* sp., suggesting the presence of chemoorganotrophic thermophilic community with the ability to utilize complex hydrocarbons as a source of energy. A high abundance of genes belonging to methane metabolism pathway was observed at Chhoti Anthoni hot spring, where methane is reported to constitute >80% of all the emitted gases, which was marked by the high abundance of *Methylococcus capsulatus*. The Tattapani hot spring, with a high-temperature range (61.5–98°C), displayed a lower microbial diversity and was primarily dominated by a nitrate-reducing archaeal species *Pyrobaculum aerophilum*. A higher abundance of cell metabolism pathways essential for the microbial survival in extreme conditions was observed at Tattapani. Taken together, the results of this study reveal a novel consortium of microbes, genes, and pathways associated with the hot spring environment.

Keywords: metagenomics, extremophiles, thermophiles, Indian hot springs, hydrocarbon degradation, Anthoni, Tattapani

INTRODUCTION

Extremophilic microorganisms are known to survive in diverse extreme conditions, such as high or low temperatures, high salinity, acidic and alkaline pH-values, and high radiation (Mirete et al., 2016). Among these extremophilic microbes, thermophiles, and hyper-thermophiles have the ability to survive in environments with very high temperature such as in hot springs, with the help of enzymes that remain catalytically active under such conditions. Detailed genomic analysis of many novel thermophiles isolated from hot springs has revealed their potential applications in industrial and biotechnological processes (Deckert et al., 1998; Sharma et al., 2014; Poli et al., 2015; Saxena et al., 2015; Dhakan et al., 2016). The remarkable genomic versatility and complexity of such largely unculturable extremophilic communities can be accessed using metagenomics and next-generation sequencing technologies (Lopez-Lopez et al., 2013), and has led to the discovery of many novel thermophilic bacterial and viral high-quality genomes with several prospective applications (Colman et al., 2016; Eloe-Fadros et al., 2016; Gudbergdottir et al., 2016).

A time point metagenomic study carried out for 3 years in three alkaline hot springs of Yellowstone National Park (44–75°C) showed variations in the resident bacterial population at the three sites which had different temperatures. However, no significant changes were observed in microbial diversity in the same samples collected at different time points (Bowen De Leon et al., 2013). The elemental analysis of these hot springs revealed elevated levels of sodium, chloride and fluoride, and absence of iron, cobalt, silver, and other heavy metals. The metagenomic analysis revealed the presence of many photosynthetic bacteria (*Cyanobacteria* and *Chloroflexi*) and methanogenic archaea (*Methanomassiliicoccus* and *Methanocella*) in the hot spring samples. In another study based on 16S rRNA amplicon sequencing of metagenomic samples from Sungai Klah (SK) hot spring in Malaysia (50–110°C, pH 7–9), Firmicutes and Proteobacteria were found as the most abundant phyla (Chan et al., 2015). Elements like aluminum, arsenic, boron, chloride, fluoride, iron, and magnesium were found at higher levels in the hot spring sample, whereas other heavy metals such as lead, mercury, chromium, copper, etc., were below the limits of quantification. The presence of genes for sulfur, carbon, and nitrogen metabolism suggested metabolic and functional diversity among the microbiome species. The enrichment of carbon metabolism pathway in the SK hot spring was attributed to the high total organic content due to plant litter observed at this site. The major taxa found to be dominant in other alkaline hot springs globally were *Fischerella*, *Leptolyngbya*, *Geitlerinema* (Coman et al., 2013), *Stenotrophomonas*, and *Aquaspirillum* (Tekere et al., 2011). However, no significant correlation has been observed between the microbial diversity and geochemical characteristics of the hot springs in the above-mentioned studies.

Metagenomic analysis of an acidic hot spring, *El coquito* in the Colombian Andes, reported the presence of transposase-like sequences involved in horizontal gene transfer and genes for DNA repair system (Jimenez et al., 2012). The hot springs showed a higher proportion of Gammaproteobacteria and Alphaproteobacteria. In other reports focusing on acidic

hot springs, the major taxa found to be dominant were *Verrucomicrobia* and *Acidithiobacillus* spp. (Mardanov et al., 2011; Menzel et al., 2015). Many studies have also revealed the presence of genes encoding for enzymes of biotechnological interest, such as hydrolases, xylanases, proteases, galactosidases, and lipases (Ferrandi et al., 2015; Littlechild, 2015). Some studies focusing on the metagenomic analysis of hot springs located in India have reported *Bacillus licheniformis* (Mangrola et al., 2015), *Bacillus megaterium*, *Bacillus sporothermodurans*, *Hydrogenobacter* sp., *Thermus thermophilus*, *Thermus Brockianus* (Bhatia et al., 2015), *Clostridium bifermentans*, *Clostridium lituseburensense* (Ghelani et al., 2015), *Opitutus terrae*, *Rhodococcus erythropolis*, and *Cellovibrio mixtus* (Mehetre et al., 2016) as major bacterial genera. Genes for stress responses and metabolism of aromatic and other organic compounds have been identified by preliminary functional analysis of these sites (Mangrola et al., 2015; Mehetre et al., 2016).

The Geological Survey of India has identified about 340 hot springs located in different parts of India, which are characterized by their orogenic activities (Chandrasekharam, 2005; Bisht et al., 2011). All these hot springs have been classified and grouped into six geothermal provinces on the basis of their geo-tectonic setup. Geothermal resources along Son-Narmada lineament viz. Anthoni-Samoni and Tattapani form the most promising resource base in central India (Shanker, 1986). This is one of the most important lineaments/rifted structure of the sub-continent. It runs across the country in an almost East-West direction and has a long history of tectonic reactivation (Pal and Bhimasankaram, 1976). It contains several known thermal spring areas, the most interesting one being those situated at Tattapani and Anthoni (Bisht et al., 2011). Given the large size and geographical diversity of India, the metagenomic studies from Indian hot springs are still in the infancy stage and more detailed and comprehensive studies are essential to unravel the unknown and untapped microbial and functional diversity of this region. The aim of the study was to perform a comprehensive metagenomic analysis of samples collected from seven different sites at three hot spring locations (Badi Anthoni, Chhoti Anthoni, and Tattapani) situated in two distinct geographical regions, Anthoni and Tattapani, in central India. The hot springs in this study have a temperature range from 43.5 to 98°C and neutral to slightly alkaline pH-values (7.0–7.8). Although, hot springs with similar temperature and pH-values have been studied globally for their phylogenetic and functional characteristics, the geographical location of the currently studied hot springs makes them unique in their geochemical setup. 16S rRNA V3 hypervariable region amplicon sequencing and shotgun metagenome sequencing of all the samples was carried out using Illumina NextSeq 500 for the exploration of microbial communities in the sample and gaining new insights into genes, enzymes, and metabolic pathways contributing to their survival in the thermophilic environment.

MATERIALS AND METHODS

Site Description and Sample Collection

The hot springs namely Badi Anthoni and Chhoti Anthoni are located (22.65° N latitude and 78.36° E longitude) ~2 km apart from each other, near the hill station of Pachmarhi in the state of

Madhya Pradesh, India (**Supplementary Figure 1**). Anthoni hot springs are aligned along a prominent geological fracture zone running through that region. A surface hot spring site is located in the Chhoti Anthoni region. In addition, a few boreholes to a depth of ~635 m have been drilled here as a part of petroleum oil exploration and to study the temperature regime and rock segments by the Geological Survey of India, in which presence of inflammable gases (~80% methane) has also been observed (Pandey and Negi, 1995; Sarolkar, 2010; Vaidya et al., 2015). The presence of interlayer basic silts and volcanic tuffs underlain by basic intrusive rock are also reported under these boreholes (Pandey and Negi, 1995). Water samples were collected from two sites at Chhoti Anthoni region, (i) sample from a depth of ~1 m from the surface of the hot spring (referred to as “CAN”) which had a temperature of 43.5°C, and (ii) sample from the free flowing water of the outlet of a borehole (temperature 52.1°C) which was among the several boreholes drilled (reported to be up to 635 m in depth) at Chhoti Anthoni site and was labeled as CAP (Chhoti Anthoni Petroleum). The third sample was collected at a depth of ~1 m from the surface of Badi Anthoni (BAN) sites.

The Tattapani hot spring field is ~700 km away from the Anthoni hot springs and is located in the Sarguja district of Chhattisgarh state, India (23.41°N latitude and 83.39°E longitude). The temperature of different reservoirs at this site has been reported to be as high as $230 \pm 40^\circ\text{C}$ at a depth of 2 km and $112 \pm 30^\circ\text{C}$ at a depth of 1 km (Vaidya et al., 2015). Among the 26 boreholes that had been drilled here by the Geological survey of India, the ones with distinctively high temperatures (61.5–98°C, up to 325 m deep) were chosen for sample collection (Sarolkar and Das, 2006). Multiple water samples were collected from different physical locations (TAT-1, TAT-2, TAT-3, and TAT-4) and depths, and were pooled to make four distinct samples from each site (~5 L each).

Thus, including three samples from Anthoni and four samples from Tattapani, a total of seven samples were collected. All the water samples were collected in separate sterile plastic carboys (2 L volume) which were rinsed with 0.05% bleach solution for disinfection. A total of 5 L water sample was collected from each site and brought to the laboratory within 12–18 h of collection at 4°C. The samples were stored at -20°C and processed for the extraction of metagenomic DNA within a week. The sample description and physicochemical properties recorded on-site are summarized in **Table 1**.

The seven samples considered in this study were divided into two groups, “Anthoni” and “Tattapani,” based on their geographical location for analysis. Hence, the samples collected from Chhoti Anthoni (CAP and CAN) and Badi Anthoni (BAN) sites are referred to as “Anthoni” and the samples collected from the Tattapani hot spring location (TAT-1, TAT-2, TAT-3, and TAT-4) are collectively referred to as “Tattapani” in this study.

Elemental Analysis of the Sampling Sites

Approximately 250 ml of each water sample was preserved on-site by mixing with 1:100 v/v 5% HNO₃ for elemental analysis. Dissolved major and trace elements were analyzed using an Inductively Coupled Plasma Mass Spectrometer (ICPMS, iCAPQ—Thermo Fisher Scientific, USA) in accordance with the

TABLE 1 | Physico-chemical properties of the hot spring samples recorded on-site.

| | Sample name | Temperature (°C) | pH | Total dissolved solids (ppm) |
|---------------------|-----------------------|-------------------|------------------|------------------------------|
| Anthoni | BAN | 55 | 7.8 | 690 |
| | CAN | 43.5 | 7.5 | 590 |
| | CAP | 52.1 | 7.8 | 620 |
| | Mean | 50.2 ± 5.98 | 7.475 ± 0.34 | 720 ± 116.62 |
| Tattapani | TAT-1 | 98 | 7.5 | 880 |
| | TAT-2 | 61.5 | 7.6 | 600 |
| | TAT-3 | 69 | 7 | 700 |
| | TAT-4 | 67 | 7.8 | 700 |
| | Mean | 73.88 ± 16.39 | 7.7 ± 0.17 | 633.3 ± 51.31 |
| Mann-Whitney U-test | Tattapani vs. Anthoni | 0.057 | 0.571 | 0.171 |

Six samples (BAN, CAN, TAT-1, TAT-2, TAT-3, and TAT-4) were collected from a depth of about one meter from the surface and one sample (CAP) was collected from the outlet of a ~635 m deep borehole.

United States Geological Survey protocol (Hannigan and Basu, 1998; Hannigan et al., 2001) at the Indian Institute of Science Education and Research (IISER) Bhopal, India in a class 10,000 clean lab with class 1000 clear zones. The following elements were analyzed in each sample—Li, Be, B, Mg, Al, Si, K, Ca, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Se, Sr, Mo, Cd, Cs, Ba, La, Ce, Pb, Hg, and S. All samples were spiked with an internal standard of 10 ppb In, Re, and Bi for internal calibration and the final solution was an undiluted solution in 2% supra pure grade HNO₃. Ten, hundred, and thousand ppb solutions of the sample elements were prepared and standardized against high-grade multi-elemental standards. These solutions were then used as standards for measurements of the seven water samples of this study. Helium was used as a collision gas to reduce interference of argon oxide ions. Suprapure HCl (5%) was used as a backwash during analyses of Hg and S, while 5% supra pure HNO₃ was used as a backwash for all other elements. Both Hg and S were measured as separate individual experiments and blanks were measured in between each sample for these two elements to ensure zero memory from previous samples. Analytical uncertainties are <5% for all elements analyzed for this study.

Metagenomic DNA Extraction

In the lab, each sample was filtered through a 1.2 µm pore size membrane filter to remove any traces of debris and coarse particles from the collected water. The filtrate was then passed through a 0.2 µm pore size membrane filter to entrap the prokaryotic cells on the filter. The filter membrane was cut into pieces and subjected to metagenomic DNA extraction using Metagenomic DNA isolation kit for water (Epicentre, Wisconsin, USA) according to the manufacturer’s protocol with some modifications to increase the yield and purity of the extracted DNA sample. The modifications included the addition of 100 µl 5M NaCl to 700 µl of isopropanol for efficient precipitation

of DNA. The DNA pellet was resuspended in 50 μ l of 10 mM Tris (pH 8.5) and evaluated on Genova nanodrop microspectrophotometer (Jenway, Bibby Scientific Limited, UK) for purity and Qubit 2.0 fluorometer using Qubit HS dsDNA kit (Life technologies, USA) for quantification.

16S rRNA and Shotgun Metagenome Sequencing

The purified DNA samples were used as a template for generating the 16S rDNA V3 amplicon library. The primers used for amplification were 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGAGGCAGCAG3' (341F-ADA) and 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTACCGCGGCTGCTGGC3' (534R-ADA). The underlined regions in the above sequences are the Illumina Nextera XT adapter overhangs, whereas the non-underlined regions are the primer sequences known to target eubacterial 16S rRNA V3 region (Wang and Qian, 2009; Soergel et al., 2012). The optimized PCR conditions were: initial denaturation at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 69°C for 30 s, extension at 72°C for 30 s, and a final extension cycle at 72°C for 5 min. Recombinant *Taq* DNA polymerase (Life technologies, USA) was used and 5% DMSO was added to the master mix to enhance the concentration of amplified product from the GC-rich metagenomic template. The amplified products were evaluated on 2% w/v agarose gel, purified using Agencourt Ampure XP kit (Beckman Coulter, USA) and amplicon libraries were prepared by following the Illumina 16S metagenomic library preparation guide. The shotgun metagenomic libraries were prepared using Illumina Nextera XT sample preparation kit (Illumina Inc., USA) using the manufacturer's protocol. Both the libraries were evaluated on 2100 Bioanalyzer using Bioanalyzer DNA 1000 kit for amplicon and High Sensitivity DNA kit for metagenome (Agilent, USA) to estimate the library size. Libraries were quantified using Qubit dsDNA HS kit on a Qubit 2.0 fluorometer (Life technologies, USA) and KAPA SYBR FAST qPCR Master mix and Illumina standards and primer premix (KAPA Biosystems, USA) as per the Illumina suggested protocol. Equal concentration of libraries was loaded on Illumina NextSeq 500 platform using NextSeq 500/550 v2 sequencing reagent kit (Illumina Inc., USA) and 150 bp paired-end sequencing of both types of libraries was performed at the Next-Generation Sequencing (NGS) Facility, IISER Bhopal, India.

Analysis of Amplicon Reads

The amplicon reads were trimmed from the ends to remove ambiguous bases using NGSQC tool kit (Patel and Jain, 2012) and the reads with ≥ 3 ambiguous bases were removed. The paired-end reads were assembled together into single reads using FLASH (Magoc and Salzberg, 2011). The reads having $\geq 80\%$ bases above Q30 were quality filtered and primer sequences were removed using cutadapt (Martin, 2011). The high-quality reads were then clustered and assigned by closed-reference OTU (Operational Taxonomic Unit) picking protocol (pick_closed_reference_otus.py) of QIIME v1.9 (Caporaso et al., 2010) using Greengenes database v13_5 (DeSantis et al., 2006) as a reference at $\geq 97\%$ identity. The *de novo* OTU picking

protocol (pick_otus.py) was adopted for the sequences that failed closed-reference OTU picking and were aligned against the Greengenes database using BLAT (Kent, 2002). The assignment of representative sequences from these OTUs was carried out by an in-house Perl script using Lowest Common Ancestor (LCA) approach (Chaudhary et al., 2015). The OTUs with low abundance (≤ 100 sequences) were filtered out to remove noise.

Analysis of Metagenomic Reads

The whole genome shotgun reads were filtered using NGSQC tool kit by removing the reads with ambiguous bases and further selecting the reads having $\geq 80\%$ bases above Q30. The high-quality metagenomic paired-end reads obtained from each sample were then assembled using MetaVelvet at a k-mer size of 77 bp (Namiki et al., 2012). The contigs obtained after *de novo* assembly were used for the prediction of open reading frames (ORFs) using MetaGeneMark (Zhu et al., 2010). The predicted genes obtained from all the samples were combined and clustered using cd-hit at 95% identity and 90% coverage length. To prepare a comprehensive dataset of metagenomic genes, the non-redundant gene repertoire generated from hot spring samples was combined with a non-redundant gene set from genes obtained from 2785 reference genomes from NCBI and genes from reference genomes in JGI genome portal (Grigoriev et al., 2012). The reads from each sample were mapped to this combined gene pool using Bowtie 2 (Langmead and Salzberg, 2012) for quantification of genes. The paired-end reads which mapped to the same gene, and cases where only one read from the paired-end read mapped on a gene and the other read remained unmapped, were considered and quantified. The genes having a total count of < 10 were removed to avoid ambiguous genes in the analysis. The mapping of reads resulted in a total of 438,157 genes (count ≥ 10) including all samples, of which 355,675 genes were previously identified in assembled contigs from hot spring dataset, whereas 82,483 genes could not be predicted in assembled contigs and were identified through mapping of reads directly to the gene repertoire prepared using predicted genes from hot springs data, NCBI, and JGI genes. The genes were further annotated by alignment using BLAST 2.2.6 (Altschul et al., 1990) against eggNOG 4.0 (Powell et al., 2014), KEGG Database version 2011 (Kanehisa and Goto, 2000), and also by KEGG Automated Annotation Server (KAAS; Moriya et al., 2007). The information on pathway and KO was updated by retrieving the data from KEGG web-server in September 2015. The genes with best hits (bit-score ≥ 60 and $e \leq 10^{-6}$) were assigned with KO and eggNOG annotation, respectively, and were used for further analysis. The abundance of genes assigned to same KO IDs was summed and the relative abundance of each KO was calculated for each sample. Similarly, the abundance of genes assigned to same eggNOG IDs was summed and the relative abundance of each eggNOG was calculated for each sample.

Quantification of genomes was performed by mapping the reads to reference genomes using Bowtie 2 with default parameters. The reads which mapped to the reference genomes with $\geq 90\%$ identity and with both the pairs mapping concordantly were considered as a hit. The abundance of

identified genomes was further normalized by reference genome size and total number of reads in each sample (Gupta et al., 2016). To determine the taxonomic origin of metagenomic ORFs, the predicted ORFs were aligned using BLASTN against 2785 NCBI reference genomes. The hits were filtered using an *e*-value cut-off of $\leq 10^{-5}$ and alignment coverage $\geq 80\%$ of the reference gene length. In the case of multiple best hits with equal identity and scores, LCA method was used for assignment. An identity threshold of 95% was used for assignment up to species level, 85% for genus level assignment and 65% for phylum level. A total of 1,53,098 genes (35%) could be assigned with taxonomy at least up to phylum level based on LCA algorithm.

Statistical Analysis

For amplicon reads, alpha diversity metrics including observed species and Shannon index were calculated at each rarefaction depth beginning from 100 sequences per sample to 2.2 million sequences per sample ($n = 10$ times) at a step size of 0.1 million using QIIME v1.9. Pielou's evenness was calculated using Vegan package in R to identify the distribution of species with respect to their proportion at each site (Oksanen et al., 2013). The abundance of various phyla and genera in samples was calculated and those which showed $\geq 1\%$ abundance were plotted. Furthermore, the genera which had $\geq 1\%$ abundance and showed significant (Wilcoxon Rank Sum test, $p \leq 0.05$) difference in their abundance in two locations (Anhoni and Tattapani) were plotted. The taxonomic tree was constructed for genera having $\geq 1\%$ abundance using GraPhlAn (Asnicar et al., 2015) to understand the taxonomic differences in different samples along with their abundances. Unweighted UniFrac distances were calculated on rarefied OTU table using QIIME v1.9 and were plotted on PCA (Principal Component Analysis) plot. To identify the discriminatory taxa based on the two locations, Random Forest analysis was carried out using randomForest package in R (Liaw and Wiener, 2002). In order to identify changes in diversity due to differences in temperature, multiple comparisons using Tukey's test were performed for number of observed species across the three groups of samples, i.e., moderately high temperature (40–55°C), high temperature (55–75°C) and extremely high temperature ($\geq 75^\circ\text{C}$).

The relative abundance of KOs and eggNOGs calculated in each sample from individual metagenomic reads was used for further statistical analysis. Hellinger distances were calculated to estimate the beta diversity between samples. A total of 4749 KOs (with ≥ 10 counts) were used for further analysis and the odds ratio was calculated for a comparison between the two sites. Pathway analysis was performed using Reporter Features algorithm (Oliveira et al., 2008) to identify the pathways significantly enriched at both the locations due to temperature and other factors. The species identified from metagenomes were clustered using hierarchical clustering algorithm based on their proportions in each sample. The clustering pattern and species abundance (normalized) of each sample were plotted as a heatmap. The proportions of species were also correlated with temperature and those with significant Spearman's Rank correlation coefficient (FDR or False Discovery Rate adjusted $p \leq 0.05$) and higher correlations ($\rho \geq 0.7$ or ≤ -0.7) were plotted.

Moreover, to find associations of species with temperature groups (moderately high, high, and extremely high temperature) on ordinations, PCA analysis was performed using Biplots with species plotted as vectors in ordinations.

Nucleotide Sequence Data Deposition

The nucleotide paired-end sequences have been deposited in NCBI SRA database with accession ids: SRR3961733, SRR3961734, SRR3961739, SRR3961740, SRR3961741, SRR3961742, and SRR3961743 for Whole Genome Sequencing (WGS) reads, and SRR3961735, SRR3961736, SRR3961737, SRR3961738, SRR3961744, SRR3961745, and SRR3961746 for 16S rRNA (V3 region) amplicon reads under the BioProject PRJNA335670.

RESULTS

Physico-Chemical Analysis of the Sampling Sites

The temperature at the Anhoni hot springs ranged between 43.5 and 55°C, with pH-values between 7.5 and 7.8 at the three sites (Table 1). Tattapani hot springs had a much higher temperature than Anhoni and ranged between 61.5 and 98°C, with pH-values between 7 and 7.8 at the four different hot spring sites. The total dissolved solids (measured in ppm) varied between 590 to 690 at Anhoni and 600 to 880 at Tattapani. All these measurements were carried out on-site. The Anhoni hot spring samples showed high concentrations of Co, La, Fe, Hg, and Si (Supplementary Table 1 and Supplementary Figure 2). However, Pb, Zn, Ni, and B were observed to be high in Tattapani samples, indicating high levels of heavy metals in that site. Hg and S were observed in high concentration in the Anhoni samples as compared to Tattapani samples.

Amplicon and Metagenomic Analysis

A total of 21,881,886 high quality reads from 16S rRNA V3 region obtained from seven samples, each consisting of 2.3–4 million reads, were analyzed (Supplementary Table 2). A total of 2196 OTUs were obtained on closed-reference OTU picking. Since 5,242,917 reads (24% of sequences) could not be clustered and assigned using closed-reference OTU picking method, *de novo* OTU picking was performed for the remaining reads followed by taxonomic assignment. A total of 4018 OTUs (clustered at $\geq 97\%$ identity) were obtained using both closed reference and *de novo* OTU picking methods, of which 90 OTUs (0.3% of total sequences) could not be assigned at any taxonomic level. A total of 99,156,908 high-quality metagenomic reads [$14,165,272 \pm 3,481,846$ (mean \pm SD)] were obtained from the seven samples. The assembly of reads resulted into 373,793 contigs (mean = 53,399; Supplementary Table 3) and a total of 438,157 genes (count ≥ 10) were identified in all samples. The gene repertoire from all hot spring samples was classified into KEGG Orthologous groups and eggNOG Orthologous groups. A total of 308,855 (70.49%) genes were annotated with 4749 KO IDs, and 313,798 (71.6%) genes were annotated with 31,794 eggNOG IDs.

Hot Springs at Higher Temperatures Displayed Lower Microbial Diversity and Gene Content

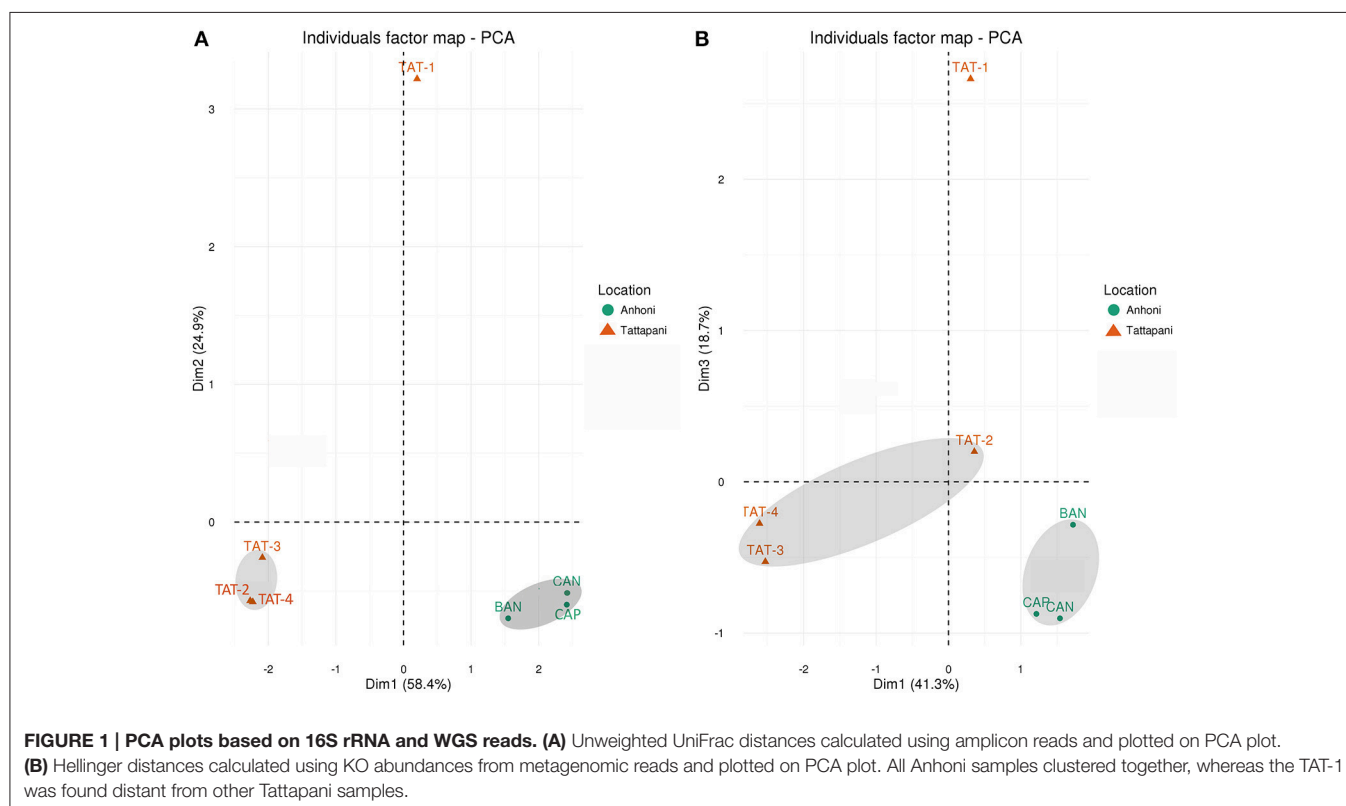
Alpha diversity analysis using Observed Species, Shannon Diversity Index, and Pielou's evenness was carried out to determine the species richness and evenness in the samples. TAT-1 sample having the highest temperature among all samples showed the lowest species richness and diversity as compared to all other sites at each rarefying depth (**Supplementary Figure 3**). Beta diversity analysis was carried out using Unweighted UniFrac distances. Lower UniFrac distances and close clustering was observed among the Anthoni samples, and also among three of the four Tattapani samples on the PCA plot (**Figure 1A**). TAT-1 sample from Tattapani showed higher UniFrac distance from all other sites. The region-specific clustering of Anthoni and Tattapani samples on the PCA indicates phylogenetic similarity in samples obtained from a similar region. It also highlights the variation in microbial community of the two locations due to the differences in geographical regions and observed temperature.

Shannon diversity index was used to estimate the gene diversity within the samples and showed a significantly lower ($p \leq 0.05$) gene diversity in Tattapani as compared to Anthoni (**Supplementary Table 4**). The Tattapani samples showed separate clustering from Anthoni samples using Hellinger Distances based on KO proportions and eggNOG proportions (**Figure 1B** and **Supplementary Figure 4**). The Tattapani samples showed higher Hellinger distances, which could be due to the wide temperature range. TAT-1, which has the

highest temperature showed the largest distance from other samples collected from the same geographical location. This suggests that temperature might be playing a significant role in determining the metabolic potential of the microbial community in this region. Tukey's test showed a significant reduction ($p \leq 0.001$) in the number of observed species at an extremely high temperature ($\geq 75^\circ\text{C}$) compared to other two groups of moderately high ($40\text{--}55^\circ\text{C}$) and high temperature ($55\text{--}75^\circ\text{C}$; **Supplementary Figure 5**).

Microbial Community Structure

Based on the amplicon analysis, Proteobacteria was found to be the most abundant (52–99%) phyla in all samples, except TAT-3 (19.5%) which had the highest abundance (59.97%) of phylum Thermi (**Figure 2A**). Thermotogae was also among the abundant phyla in TAT-2 (14.2%), TAT-3 (9.2%), and TAT-4 (26.8%). Firmicutes were found abundant in all samples ($\sim 4\text{--}10\%$), except CAP and TAT-1. At the genus level, *Tepidimonas* was the most abundant genus in BAN, CAN, and TAT-2 (56–67%) and was also abundant in TAT-4 (15.4%; **Figure 2B**). However, *Flavobacterium* was the most abundant in CAP (27.1%), *Thermus* in TAT-3 (60%) and *Acinetobacter* in TAT-1 (92.1%). TAT-4 also showed a higher abundance of *Fervidobacterium* (26.8%) and *Paracoccus* (24.4%). *Pseudomonas* was abundant in Anthoni samples and was noticeably high in CAP (8.5%). Species identification was carried out using WGS reads, and the species with $\geq 5\%$ abundance were plotted in bar plots (**Figure 2C**). A higher abundance of *Acidovorax* sp. was observed in BAN (54.6%) and CAN



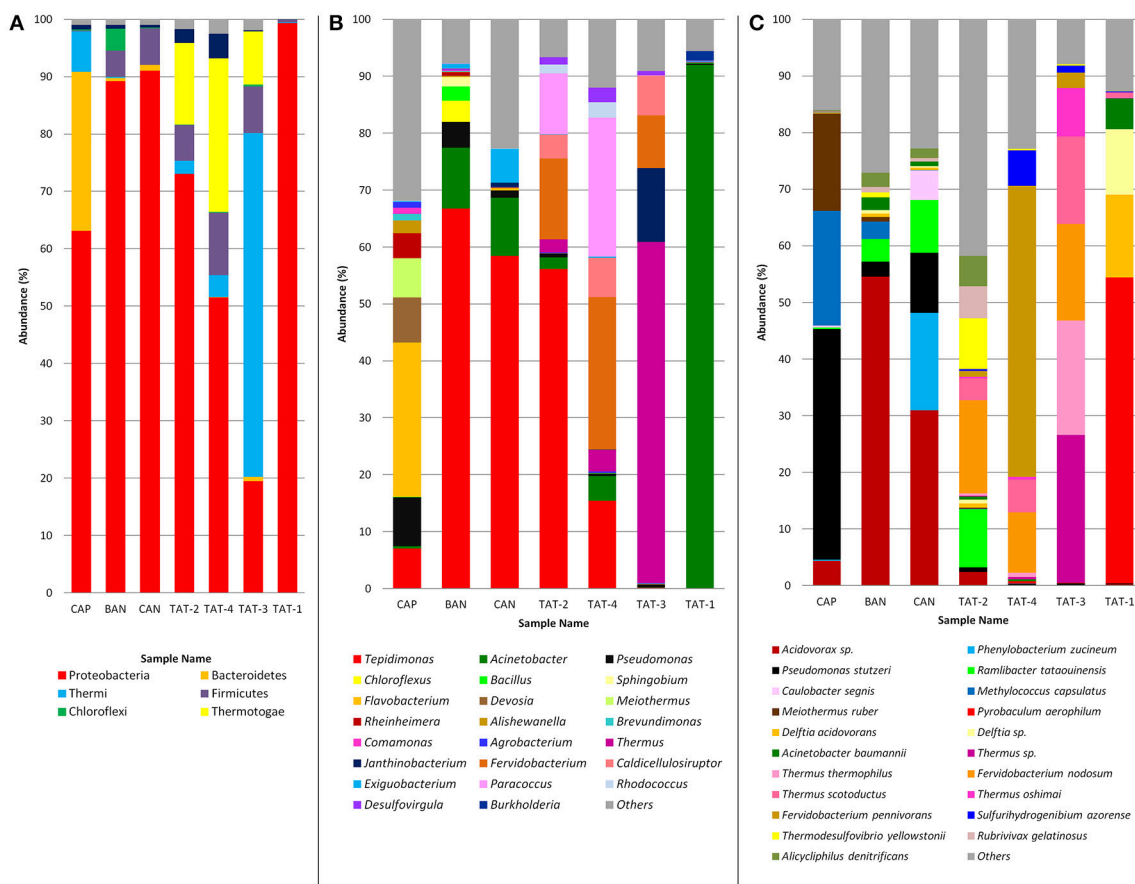


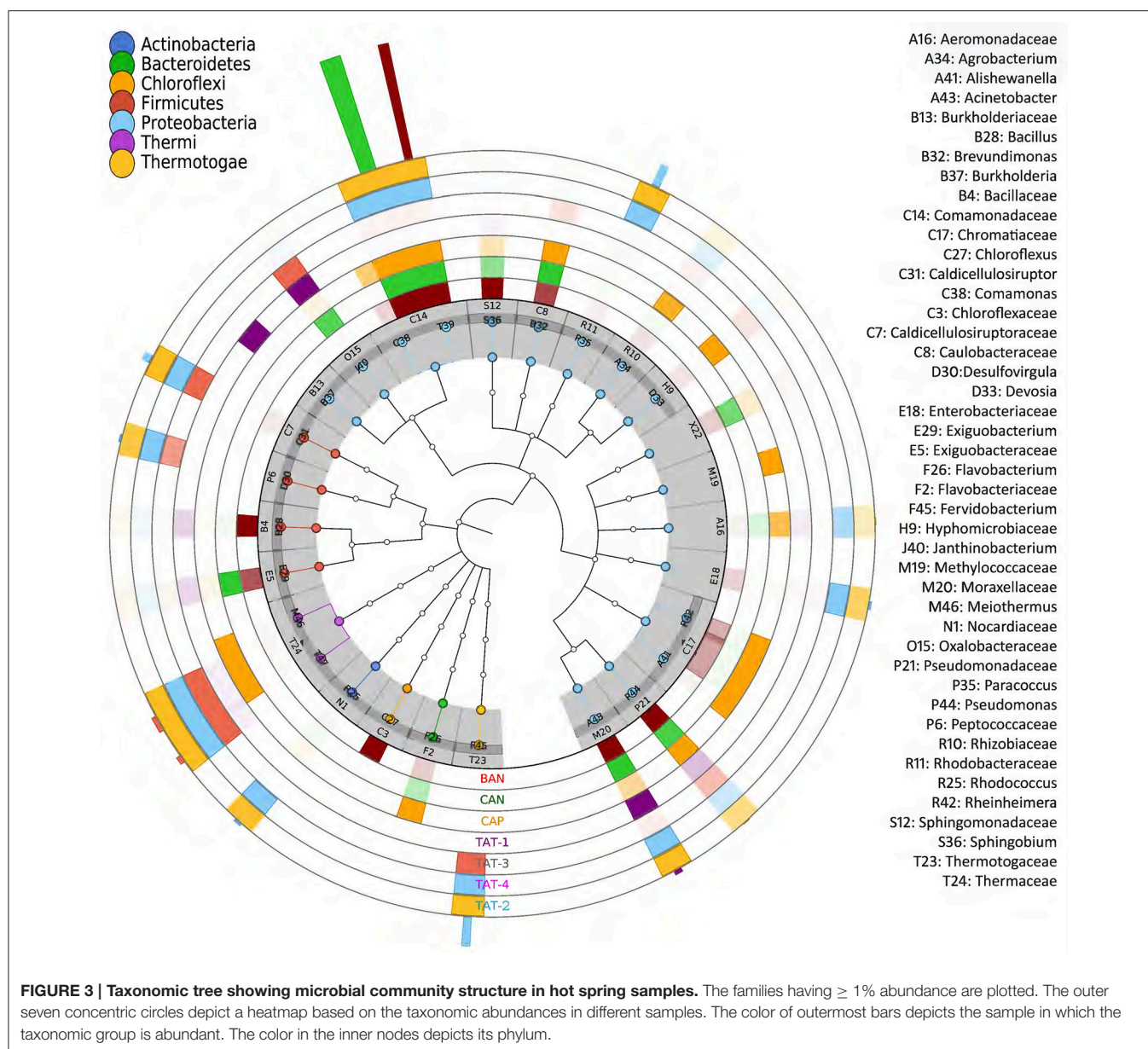
FIGURE 2 | Relative abundance of microbial community at different taxonomic levels. (A) Phylum abundance in hot springs samples. Phylum abundance $\geq 1\%$ are plotted in the bar plot. **(B)** Genus abundance in hot springs samples. Genus abundance $\geq 1\%$ are plotted in the bar plot. **(C)** Species abundance in hot springs samples. Species abundance $\geq 5\%$ are plotted in the bar plot. **(A,B)** are prepared from the results from 16S rRNA gene (V3 amplicon) sequencing and C is prepared from the taxonomic analysis of metagenomic reads.

(30.9%) samples, and *Pseudomonas stutzeri* was abundant in CAP (40.8%). An archaeal species, *Pyrobaculum aerophilum* was abundant (54.0%) in the TAT-1 sample. Two *Fervidobacterium* species, *Fervidobacterium pennivorans* (51.4% in TAT-4) and *Fervidobacterium nodosum* (16.5% in TAT-2; 17.1% in TAT-3 and 10.9% in TAT-4), were observed as abundant in other Tattapani samples. Three known *Thermus* species, *Thermus thermophilus* (20.2%), *Thermus scotoductus* (15.5%), and *Thermus oshimai* (8.6%) and an unknown *Thermus* species (26.1%) were found abundant in TAT-3. *Ramlibacter tataouinensis* was also among the abundant species in TAT-2, CAN, and BAN. This species is a cyst-producing aerobic chemoautotroph and is commonly found to be associated with dry environments (Heulin et al., 2003). The presence of this soil-dwelling bacterium in a hot spring water sample is intriguing.

Among the Anthoni samples, BAN and CAN showed similar taxonomic profiles, whereas CAP displayed a different taxonomic composition. Similarly, TAT-2 and TAT-4 from Tattapani displayed similar taxonomic profile, whereas TAT-1 and TAT-3 had a strikingly different taxonomic structure (Figure 3). *Desulfovibrula*, *Fervidobacterium*, and *Thermus* were

found to be significantly associated ($p \leq 0.05$) with Tattapani, whereas *Flavobacterium*, *Pseudomonas*, and *Rheinheimera* were found to be significantly associated ($p \leq 0.05$) with Anthoni (Supplementary Figure 6).

The heatmap depicting the hierarchical clustering of species and samples using species proportions at each site (Figure 4) indicated the close clustering of three samples from Tattapani (TAT-2, TAT-3, and TAT-4), and similarly, the three samples from Anthoni (BAN, CAN, and CAP) were found to cluster separately. Interestingly, TAT-1 was observed to be the farthest from all samples in the dendrogram suggesting the existence of a unique microbial population enriched in archaeal species in this region. In order to find associations of species with temperature, the Spearman's rank correlation coefficient was calculated for each species with temperature and those with significant correlation coefficients (FDR adjusted $p \leq 0.05$; $\rho \geq 0.7$ or ≤ -0.7) were plotted (Figure 5). *Pyrobaculum aerophilum* and *Pyrobaculum arseniticum* correlated positively with temperature, whereas *P. stutzeri*, *Methylococcus capsulatus*, and *Caulobacter* spp. correlated negatively showing their association with lower temperature sites. The association of *P. aerophilum* and *Delftia*



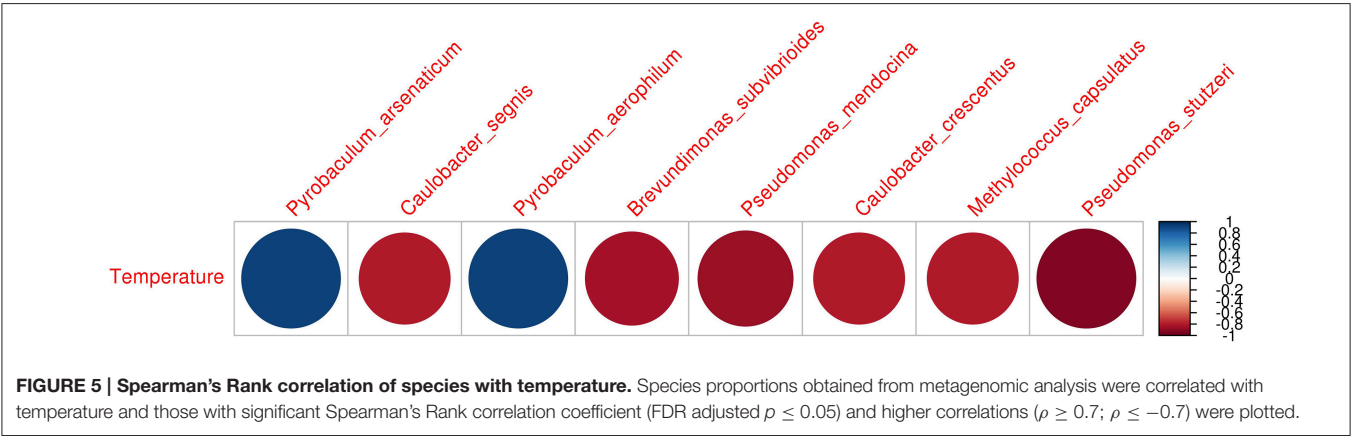
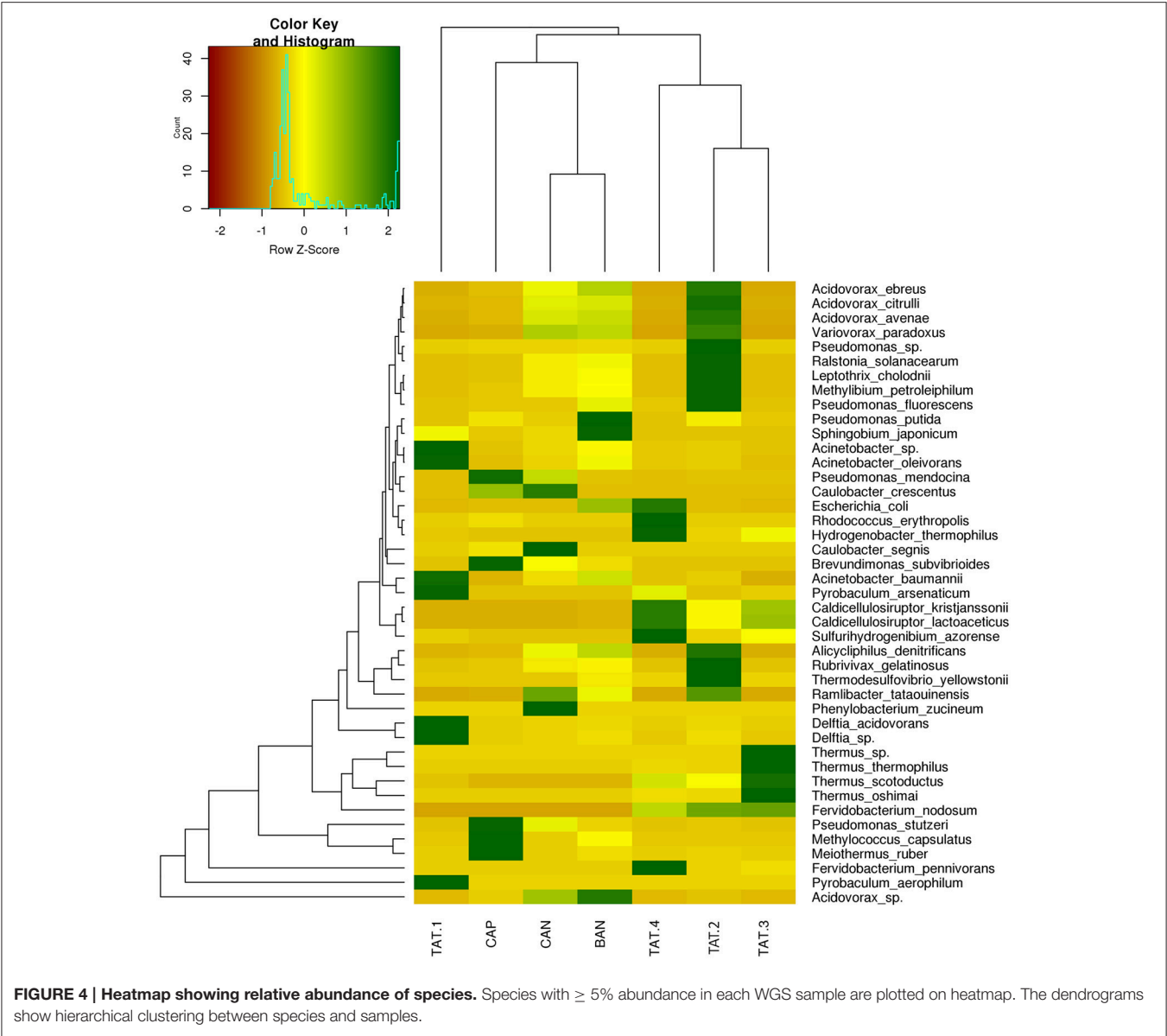
species was observed with extremely high temperature, and *Thermus thermophilus* and other thermophilic species were found to be associated with high temperature when plotted in ordinations using Biplot (Figure 6). A clear distinction was also observed with *M. capsulatus* and *Meiothermus* showing association only with CAP, making this site different from other hot springs. Species association also varied between TAT-2, TAT-3, and TAT-4 suggesting unique species diversity at each site.

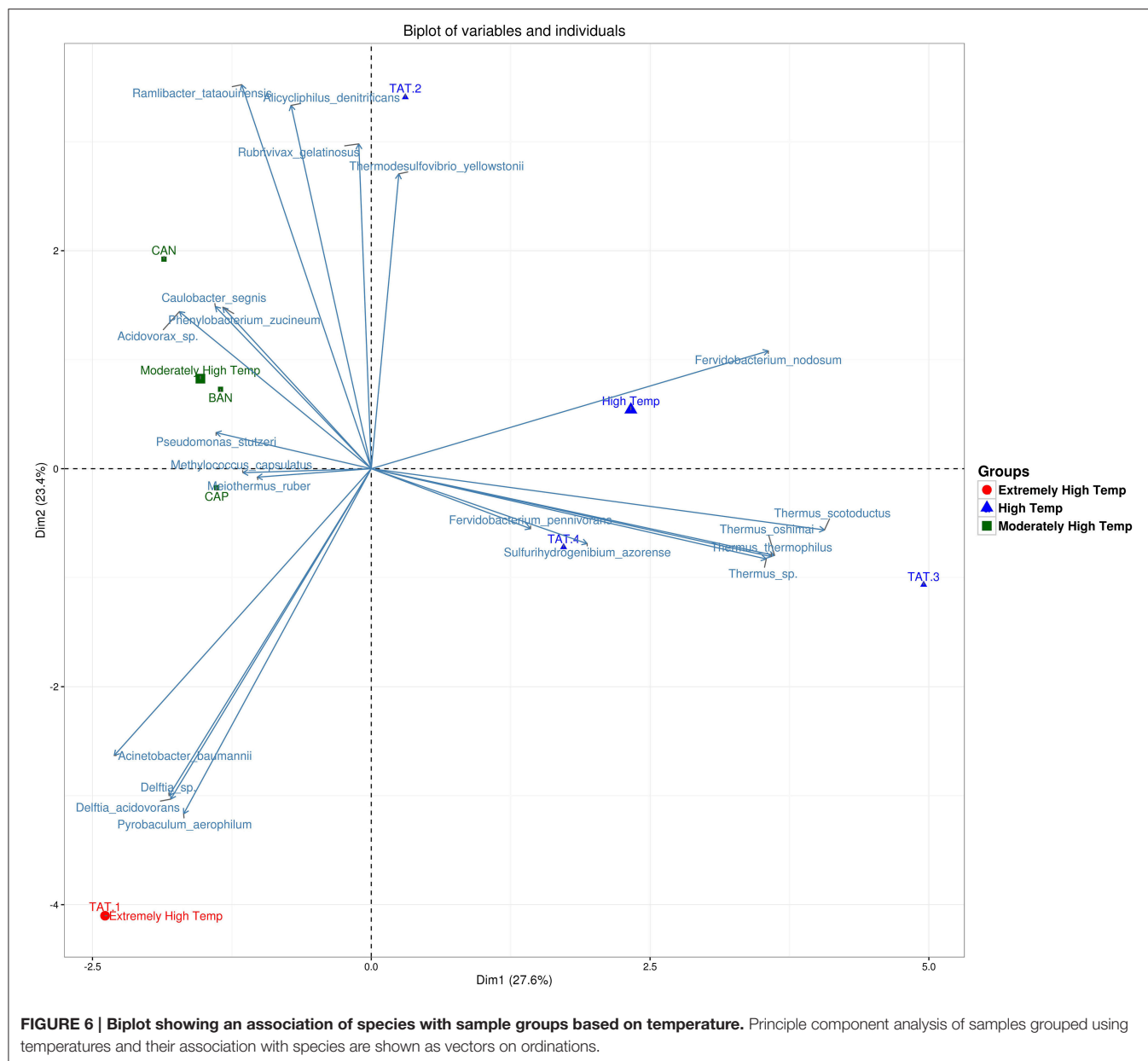
A total of 24 OTUs were observed as discriminatory among Anthoni and Tattapani samples using Random Forest analysis (Supplementary Table 5), out of which nine OTUs belonged to *Pseudomonas* (*P. stutzeri*, *P. alcaligenes*, and unknown species of *Pseudomonas*). OTUs belonging to genus *Pseudomonas*, *Acinetobacter*, *Exiguobacterium*, and *Vogesella*

were discriminatory for Anthoni, whereas *Paenibacillus*, *Planomicrobium*, and *Devosia* were found as discriminatory genus for Tattapani.

Identification of Functions in Anthoni and Tattapani

The analysis of eggNOG classes revealed interesting trends with respect to the enrichment of functional categories (Supplementary Figure 7). The functions for inorganic ion transport, lipid metabolism, and secondary metabolites were enriched in Anthoni, whereas energy production, nucleotide metabolism, cell cycle control, replication, and post-translational modifications were enriched in Tattapani. It is apparent that the functions and pathways related to nucleotide metabolism, DNA





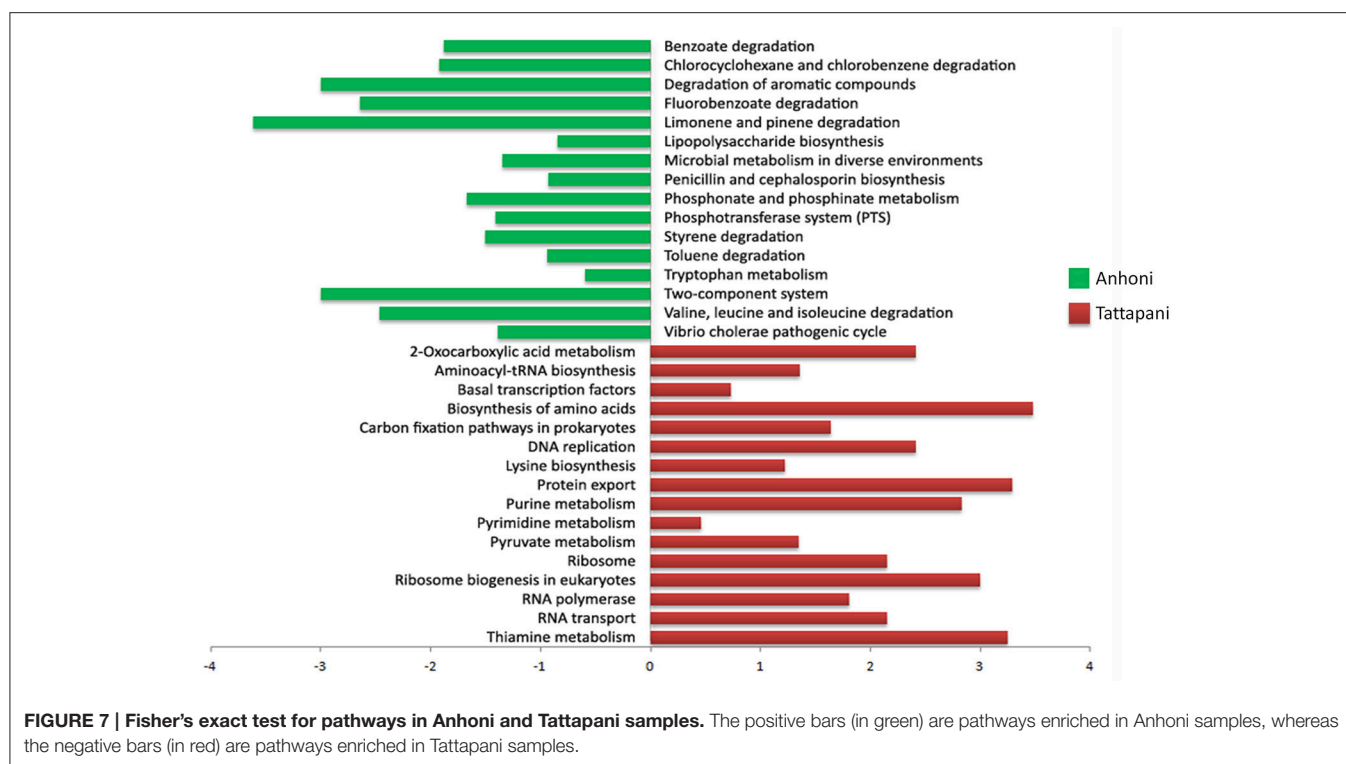
replication and energy production were enriched (FDR adjusted $p \leq 0.05$) in samples obtained from higher temperature hot springs (Tattapani).

A comparison of KEGG pathways using the Odds ratio for enrichment, and further by Fisher's exact test on the proportion of enriched KOs belonging to a pathway in the two locations revealed several major pathways significantly associated with Anthoni and Tattapani (Figure 7). Log odds ratios calculated for significantly discriminatory pathways through Fisher's exact test revealed pathways for degradation of hydrocarbons such as benzoate, toluene, xylene, fluorobenzoate, chlorocyclohexane, chlorobenzene, etc., as significantly ($p \leq 0.05$) associated with Anthoni, whereas DNA replication, purine, and pyrimidine metabolism were significantly associated with Tattapani, which

corroborates with the results of functional analysis using eggNOG classes for Tattapani.

Enriched KEGG Metabolic Pathways in Tattapani

A detailed analysis of KEGG pathways in Tattapani showed that pathways for biosynthesis of secondary metabolites, amino acid biosynthesis, nitrogen metabolism, and other cellular functions were abundant in this location (Figure 8). The higher abundance of TCA cycle and oxidative phosphorylation pathways revealed by KEGG pathway analysis and abundance of energy production pathway revealed by eggNOG analysis shows that aerobic respiration is the major mechanism for energy generation by the microbial community at this site. The high abundance of



P. aerophilum in TAT-1 (Figure 2C) and its earlier reported nitrate reducing properties (Afshar et al., 2001) corroborates with the observed nitrogen metabolism pathway at this site.

Degradation of Hydrocarbons in Anthoni

One of the interesting findings of this study is the presence of hydrocarbon degradation pathways in Anthoni Samples. We performed a comprehensive analysis of these pathways to identify the various intermediates and end products, microbes harboring these pathways and their relative abundance in the CAP, BAN, and CAN samples which are summarized in Figure 9. Methane metabolism was found to be highly enriched in CAP which is reported to have a high proportion (>80%) of methane in previous studies (Sarolkar, 2015). The high (20.27%) abundance of *M. capsulatus* which has the ability to utilize methane as a source of energy corroborates with the enrichment of methane metabolism pathway at this site (Ward et al., 2004). The genes involved in methane metabolism were present in the metagenomic dataset and could be annotated with this genome by taxonomic assignment (Supplementary Figure 8). Oxidation of methane to methanol and subsequently to formaldehyde is performed by this microbe which is further utilized in other downstream pathways including conversion to Acetyl-CoA for energy production.

Complete metabolic pathways for the degradation of toluene, xylene, benzoate and benzene were found in various microbes identified in Anthoni samples (Supplementary Figures 9–11). Enzymes for toluene degradation pathway were found as most abundant in BAN and were also present in CAP and CAN samples, contributed mainly by *Acidovorax* sp. and *Alicyclophilus denitrificans* (Figure 9). The enzymes for benzoate

and xylene degradation, which are downstream pathways of toluene degradation, were higher in CAP as compared to BAN and CAN samples and were predominantly contributed by *P. stutzeri* and *Acinetobacter baumannii*. Benzene on conversion to phenol and subsequently to catechol can be degraded via benzoate degradation pathway and the enzymes for this pathway were mainly contributed by *Acidovorax* sp. and *Alicyclophilus denitrificans*. Xylene degradation pathway was also found to be abundant in Anthoni samples and was contributed by microbial community mainly comprising of *P. stutzeri*, *Acidovorax* sp., and *Rhodococcus erythropolis*. It is interesting to note that most of the above microbes involved in the metabolism of complex hydrocarbons do not possess all enzymes required in the respective pathway. However, together as a community, they could complete the set of enzymes of a pathway, and thus, achieve the ability to carry out complex metabolism. For example, in the case of benzoate degradation pathway in CAP, 3-oxoadipate CoA-transferase enzyme was not present in *P. stutzeri*, whereas all other seven enzymes (out of the total eight enzymes) involved in the pathway were present. Another microorganism, *Acidovorax* sp., which was abundant in the same microbial community is found to harbor this enzyme and perhaps contribute toward the completion of the pathway (Supplementary Figure 12). Similarly, in xylene degradation pathway, individual microbes including unknown microbial species contribute different steps in xylene degradation and complete the pathway as a community. Taken together, it is apparent that Anthoni samples possess a unique microbial community with the ability to utilize various hydrocarbons as a source of energy via aerobic respiration. By mapping these enzymes to their respective pathways, it was observed that benzoate and xylene are utilized and converted to

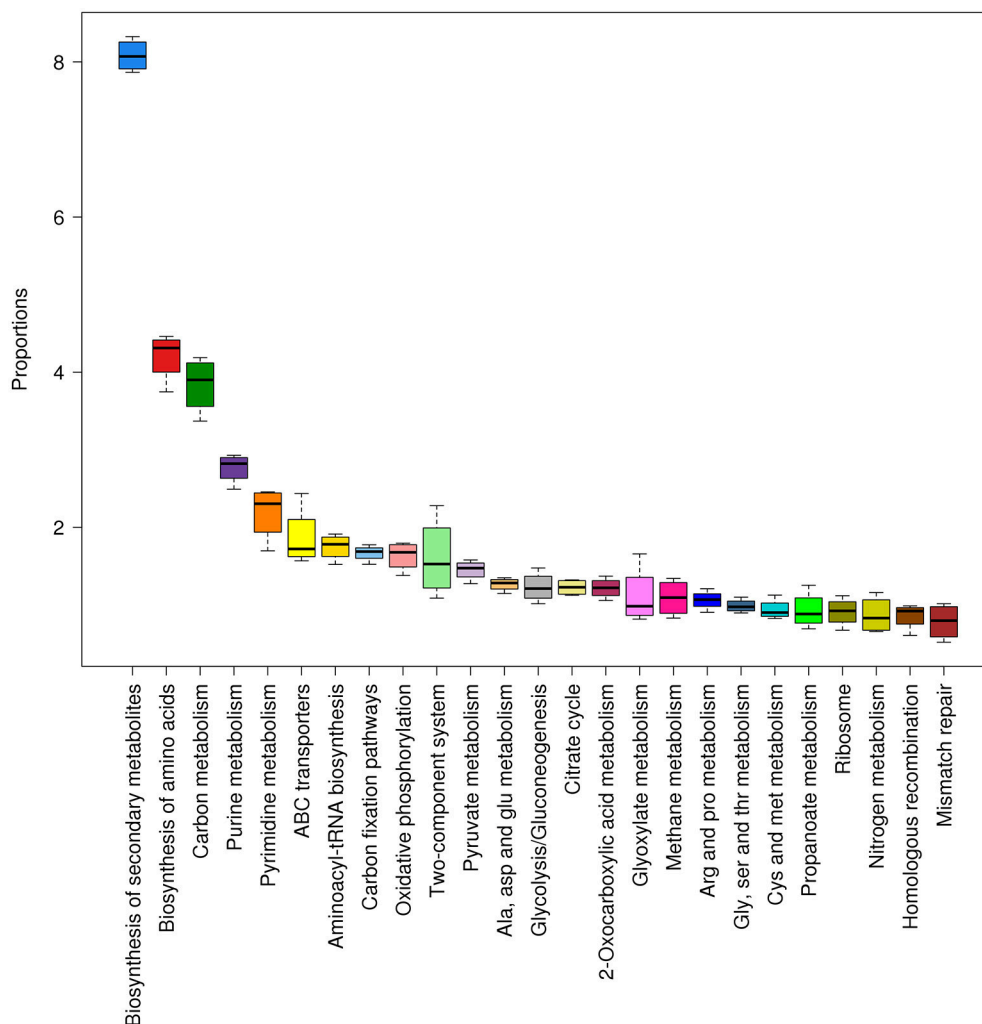


FIGURE 8 | Abundant KEGG pathways found in Tattapani samples. The pathways and their proportions in Tattapani samples are shown using box plots ($\geq 1\%$ proportion). The samples showed a higher abundance of pathways for general cell functions and certain specific functions such as nitrogen metabolism and methane metabolism.

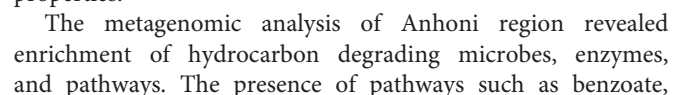
Succinyl-CoA and Acetyl-CoA, respectively, which then enters the TCA cycle to generate energy (**Supplementary Figure 13**).

DISCUSSION

The hot springs of Anthoni and Tattapani are located in the margins of Gondwana coalfields of India and have been reported as important geothermal resources (Pandey and Negi, 1995). The presence of volcanic tuffs and interlayer basic silts, which are flat intrusions of igneous rocks formed between the pre-existing layers of rocks, has also been reported under these regions. The volcanic sediments are known to constitute hydrocarbons and other organic compounds and also release organic gases (Farooqui et al., 2009). The long-term seepage of hydrocarbons, either as macroseepage or as microseepage can bring about a diverse array of mineralogical and chemical changes that are favored by the development of near-surface oxidation or reduction zones (Khan and Jacobson, 2008). Surface

geochemical methods are based on the premise that hydrocarbon gas components migrate from sub-surface petroleum reserves through faults and fractures and leave their signatures in near surface soils (Price, 1986; Tedesco, 2012). Petroleum microseepage in the soil causes several chemical reactions modifying the oxidation-reduction potential of the soil which plays an important role in the mobility of elements. It can be inferred that in Anthoni, a lot of igneous bodies cut the sedimentary bodies and water present deep along these joints gets heated from the mantle thus, finding space to come to the surface. As they travel up and cool they start to deposit the elements with lower solubility i.e., the hydrocarbons, progressively. The same water continues to cycle in this process and brings elements associated with hot springs from the mantle to the crust, making it enriched in hydrocarbons and other inflammable compounds.

Anomalous amounts of vanadium, chromium, nickel, cobalt, manganese, mercury, copper, molybdenum, uranium, zinc, lead,



toluene, and xylene degradation at Anthoni, specifically in CAP sample, underscores the potential of the inherent microbial community to metabolize hydrocarbons as a source of energy. *P. stutzeri* and *Acidovorax* sp. which are known to harbor the enzymes for hydrocarbon degradation were found abundant at Anthoni. *P. stutzeri* is among the important alkane-degrading microorganisms reported for the bioremediation of crude oil, oil derivatives and aliphatic hydrocarbons (Lalucat et al., 2006). Potential degradation of toxic hydrocarbons such as, benzene, xylene and benzoate by *P. stutzeri* are well-reported in the literature and is also observed in the results of this study. *Acidovorax* sp. which was among the abundant species in CAP is also reported for its hydrocarbon degrading properties at high temperatures (Singleton et al., 2009). Microbes belonging to phylum Proteobacteria found abundant at this site are known to be highly abundant in petroleum-contaminated terrestrial and aquatic environments (Kimes et al., 2013; Bao et al., 2016). A 16S rRNA-based study from a hydrothermal vent (270–325°C) with similar geochemistry at the Gulf of California, where the organic matter in the sediments are converted to aliphatic and aromatic hydrocarbons under high temperature, suggested a resident sulfate-reducing bacterial population, such as *Desulfobacter*, *Desulforhabdus*, *Thermodesulforhabdus*, etc. (Dhillon et al., 2003).

The known high percentage of methane gases at Anthoni corroborates well with the presence of methanotrophic species such as *M. capsulatus* (Sarolkar, 2015). The higher abundance of this pathway, particularly in CAP, shows the importance of this site for the isolation of novel methanotrophic species. *M. capsulatus* is known to convert formaldehyde through Ribulose-P pathway to form Acetyl-CoA for subsequent energy generation pathways which were evident in our data (Supplementary Figure 8; Ward et al., 2004). Our data also suggests the conversion of formaldehyde to formate which then enters the Acetyl-CoA pathway. However, this pathway is not reported to be carried out by *M. capsulatus* (Ward et al., 2004), indicating the presence of other methanotrophs in CAP. *Meiothermus ruber*, another abundant species in CAP is a thermophile observed to be associated with moderately high temperature (Tindall et al., 2010). Methane-producing and sulfur-utilizing thermophilic bacteria belonging to the genus, such as *Thermococcus*, *Acinetobacter*, *Pseudomonas*, and *Methylobacterium* have also been reported to be associated with high-temperature petroleum reservoirs (50–120°C) in California (Orphan et al., 2000).

One of the interesting insights from this study is that all enzymes involved in a particular hydrocarbon degradation pathway were not found in a single microbial species; however, as a microbial community, the entire set of genes for the degradation of that hydrocarbon could be completed. Furthermore, the results also hint toward the existence of novel species in these hot springs which might harbor the complete degradation pathways and can be confirmed by their isolation and sequencing, therefore, providing leads for further studies. The species and enzymes identified in this study and future studies from this site could be used as promising bioremediation agents in oil spills and other hydrocarbon-contaminated regions (Kimes et al., 2013; Gaur et al., 2014).

Tattapani hot spring having an extremely high temperature (61.5–98°C) displayed a diverse thermophilic population and their pathway analysis suggested the functional adaptations of these thermophiles to survive at high temperatures. Extreme environments such as hot springs with very high temperature are commonly found to be inhabited by chemolithoautotrophic thermophilic species majorly from archaeobacteria, and from Aquifex and Thermotoga phyla in eubacteria (Stetter, 1996, 1999). The high abundance of *P. aerophilum* in this environment is an interesting finding as it is a unique hyperthermophilic archaeal species reported extensively for its nitrate reducing properties with the help of a molybdenum-associated nitrate reductase which is distinct from other mesophilic nitrate reductases (Afshar et al., 2001). *P. aerophilum* also found to be abundant in Manikaran hot spring in India (96°C), is also known to grow anaerobically by dissimilatory nitrate reduction, which is an exclusive catabolic feature of this species (Afshar et al., 2001; Bhatia et al., 2015). *Fervidobacterium thermophilum* and *Fervidobacterium pennivorans* are among the other thermophilic eubacteria observed to be abundant in Tattapani, and are reported to produce thermostable cellulases and keratinases (fervidolysin) respectively, which have potential application in the conversion of biomass to biofuels at high temperature and other biotechnological processes (Kim et al., 2004; Wang et al., 2010). Overall, a high species diversity was observed at the sites with moderately high temperature (40–55°C) i.e., from the Anthoni site, compared to the site at extremely high temperature i.e., TAT-1 (98°C; **Supplementary Figure 5**). The presence of hydrocarbons at Anthoni site could be a contributing factor for the increase in species diversity at this site, as it provides a rich carbon source to the microorganisms thus, helping them to survive despite high temperature (Head et al., 2006). Large community diversity is generally observed in petroleum or hydrocarbon-contaminated environments and is also observed in the Anthoni samples (Kimes et al., 2013; Bao et al., 2016).

The presence of genes for replication and DNA repair pathways appear to be important in Tattapani microbial community to survive in adverse physical conditions such as, high temperature, which cause considerable damage to DNA. The enrichment of genes associated with DNA repair systems and homologous recombination have been well-reported in extreme environments (Xie et al., 2011; Jimenez et al., 2012). Some recent studies have shown that the evolutionary rate of microbial communities is governed by the environmental conditions (Gupta and Sharma, 2015), and microbes in extreme habitats evolve faster with extensive DNA repair system and high mutation rates to cope with the deleterious effects of environment on their genomes as compared to those in stable environments (Li et al., 2014). Despite having a higher abundance of pathways such as replication and nucleotide metabolism, lower species richness was observed in Tattapani which indicates that very few microbial species are able to adapt and survive in extreme conditions.

The present metagenomic exploration of the hot springs located in central India is perhaps the largest comprehensive metagenomic study of any geographical location in India including hot springs, carried out using both 16S rRNA amplicon and shotgun sequencing. It has provided novel insights into the taxonomic, functional, and metabolic diversity of the

unique microbial community surviving in these hot springs. The presence of chemoorganotrophic thermophiles degrading complex hydrocarbons at the Anthoni hot springs and extensive survival mechanisms at the Tattapani hot springs are novel revelations of the study. The present analysis is however limited by the availability of known species of microbes in reference datasets and opens up new opportunities to decipher the genomes of yet unknown microorganisms in these hot springs.

AUTHOR CONTRIBUTIONS

VS and RS conceived the idea. RS and PW designed and performed the experiments and sequencing. AC and AG carried out the elemental analysis and interpreted its results. DD and PM performed the computational analysis. RS, DD, PM, AG, and VS analyzed the data and interpreted the results. RS, DD, and PM prepared all the figures and tables. RS, DD, PM, AG, and VS wrote and reviewed the manuscript.

ACKNOWLEDGMENTS

We thank MHRD, Govt of India, funded Centre for Research on Environment and Sustainable Technologies (CREST) at IISER Bhopal for providing financial support. However, the views expressed in this manuscript are that of the authors alone and no approval of the same, explicit or implicit, by MHRD should be assumed. We thank NGS and HPC Facility at IISER Bhopal for carrying out the sequencing runs and computational analysis, respectively. RS and DD acknowledge Department of Science and Technology-INSPIRE and University Grants Commission (UGC), Government of India, respectively, for providing research fellowships. We thank Ankita Roy and Vibhuti Shastri for reading the manuscript and providing valuable suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.02123/full#supplementary-material>

Supplementary Figure 1 | Geographical location and photographs of sampling sites of Anthoni and Tattapani hot springs. The given map was retrieved from Google Maps (2016).

Supplementary Figure 2 | Graphical representation of elemental composition at the seven sample sites.

Supplementary Figure 3 | Alpha diversity in the Hot spring samples. (A) Shannon index and (B) Observed Species were calculated by rarefying from 100 to 2.2 million sequences at a step size of 0.1 million. This analysis was carried out using amplicon reads.

REFERENCES

Afshar, S., Johnson, E., de Vries, S., and Schroder, I. (2001). Properties of a thermostable nitrate reductase from the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *J. Bacteriol.* 183, 5491–5495. doi: 10.1128/JB.183.19.5491-5495.2001

Supplementary Figure 4 | PCA plot prepared using Hellinger distances based on eggNOG proportions.

Supplementary Figure 5 | Multiple comparisons of number of observed species with moderately high, high and extremely high-temperature locations. Multiple comparisons using Tukey's test were performed between samples grouped on the basis of temperature (moderately high, high, and extremely high). The mean \pm SD are plotted with significant variations shown as *** $p \leq 0.001$.

Supplementary Figure 6 | Significantly discriminatory genera ($p \leq 0.05$) in Anthoni and Tattapani sites.

Supplementary Figure 7 | eggNOG functional classes and their distributions in the two sites. The box plot shows the abundance of eggNOG functional categories in Anthoni and Tattapani.

Supplementary Figure 8 | A pathway map of methane metabolism showing the observed KOs (highlighted in red) in this dataset. Methane is oxidized to methanol via methane monooxygenase and converted to formaldehyde with the help of methanol dehydrogenase. The KOs which are commonly present in all three sites (CAP, CAN, and BAN) of the Anthoni region are shown.

Supplementary Figure 9 | A pathway map of toluene degradation showing the observed KOs (highlighted in red) in this dataset. Toluene is degraded to Benzoate and 3-methylcatechol which enters the benzoate degradation and xylene degradation pathways respectively for further downstream processes. The KOs which are commonly present in all three sites (CAP, CAN, and BAN) of the Anthoni region are shown.

Supplementary Figure 10 | A pathway map of xylene degradation showing the KOs (highlighted in red) observed in our dataset. Xylene is degraded and converted to Acetyl-CoA by the community microbes and enters the TCA cycle for energy generation. The KOs which are commonly present in all three sites (CAP, CAN, and BAN) of the Anthoni region are shown.

Supplementary Figure 11 | A pathway map of benzoate degradation showing the observed KOs (highlighted in red) in this dataset. Benzoate is degraded and converted to intermediates such as Pyruvate and Succinyl-CoA, finally entering into TCA cycle for energy generation. The KOs which are commonly present in all three sites (CAP, CAN, and BAN) of the Anthoni region are shown.

Supplementary Figure 12 | Bar charts showing the proportion of key enzymes of benzoate pathway observed in this dataset along with their taxonomic annotations.

Supplementary Figure 13 | Utilization of hydrocarbons by the microbial community for generation of energy. The hydrocarbons are utilized as a substrate by the chemoorganotrophic thermophiles to produce energy. The degradation pathways which were observed to be completely present in Anthoni samples are shown with some of the intermediates and their end-products.

Supplementary Table 1 | Elemental analysis of the hot spring samples.

Supplementary Table 2 | Number of 16S rRNA (V3 hypervariable region) amplicon reads obtained per sample.

Supplementary Table 3 | Sequencing and assembly statistics of the hot spring samples obtained after *de novo* assembly. The size of contigs ranged from 1817 to 189,789 bp indicating a taxonomic diversity across different samples.

Supplementary Table 4 | The alpha diversity metrics calculated for 16SrRNA and metagenomics reads are shown for both sites (mean \pm SD).

Supplementary Table 5 | Discriminatory taxa of Anthoni and Tattapani identified using Random Forest Analysis.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C., and Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* 3:e1029. doi: 10.7717/peerj.1029

- Bao, Y.-J., Xu, Z., Li, Y., Yao, Z., Sun, J., and Song, H. (2016). High-throughput metagenomic analysis of petroleum-contaminated soil microbiome reveals the versatility in xenobiotic aromatics metabolism. *J. Environ. Sci.* doi: 10.1016/j.jes.2016.08.022. Available online at: <http://www.sciencedirect.com/science/article/pii/S1001074216306155>
- Bhatia, S., Batra, N., Pathak, A., Green, S. J., Joshi, A., and Chauhan, A. (2015). Metagenomic evaluation of bacterial and archaeal diversity in the geothermal hot springs of manikaran, India. *Genome Announc.* 3:e01544-14. doi: 10.1128/genomeA.01544-14
- Bisht, S. S., Das, N. N., and Tripathy, N. K. (2011). Indian hot-water springs: a bird's eye view. *J. Ener. Environ. Carbon Credits* 1, 1–15.
- Bowen, H. J. M. (1979). *Environmental Chemistry of the Elements*. London; New York, NY: Academic Press.
- Bowen De Leon, K., 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. *Front. Microbiol.* 4:330. doi: 10.3389/fmicb.2013.00330
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Chan, C. S., Chan, K. G., Tay, Y. L., Chua, Y. H., and Goh, K. M. (2015). Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front. Microbiol.* 6:177. doi: 10.3389/fmicb.2015.00177
- Chandrasekharam, D. (2005). "Geothermal energy resources of India: past and the present," in *Proceedings World Geothermal Congress* (Antalya).
- Chaudhary, N., Sharma, A. K., Agarwal, P., Gupta, A., and Sharma, V. K. (2015). 16S classifier: a tool for fast and accurate taxonomic classification of 16S rRNA hypervariable regions in metagenomic datasets. *PLoS ONE* 10:e0116106. doi: 10.1371/journal.pone.0116106
- Colman, D. R., Jay, Z. J., Inskeep, W. P., Jennings, R., Maas, K. R., Rusch, D. B., et al. (2016). Novel, deep-branching heterotrophic bacterial populations recovered from thermal spring metagenomes. *Front. Microbiol.* 7:304. doi: 10.3389/fmicb.2016.00304
- Coman, C., Druga, B., Hegedus, A., Sicora, C., and Dragos, N. (2013). Archaeal and bacterial diversity in two hot spring microbial mats from a geothermal region in Romania. *Extremophiles* 17, 523–534. doi: 10.1007/s00792-013-0537-5
- Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., et al. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392, 353–358. doi: 10.1038/32831
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05
- Dhakan, D. B., Saxena, R., Chaudhary, N., and Sharma, V. K. (2016). Draft genome sequence of *Tepidimonas taiwanensis* strain MB2, a chemolithotrophic thermophile isolated from a hot spring in Central India. *Genome Announc.* 4:e01723-15. doi: 10.1128/genomeA.01723-15
- Dhillon, A., Teske, A., Dillon, J., Stahl, D. A., and Sogin, M. L. (2003). Molecular characterization of sulfate-reducing bacteria in the Guaymas Basin. *Appl. Environ. Microbiol.* 69, 2765–2772. doi: 10.1128/AEM.69.5.2765-2772.2003
- Duchscherer, W. (1983). Geochemical hydrocarbon exploration—a new/old exploration tool. *J. Geochem. Explor.* 19, 335–336. doi: 10.1016/0375-6742(83)90025-0
- Eloe-Fadrosh, E. A., Paez-Espino, D., Jarett, J., Dunfield, P. F., Hedlund, B. P., Dekas, A. E., et al. (2016). Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* 7:10476. doi: 10.1038/ncomms10476
- Farooqui, M., Hou, H., Li, G., Machin, N., Neville, T., Pal, A., et al. (2009). Evaluating volcanic reservoirs. *Oilfield Rev.* 21, 36–47.
- Ferrandi, E. E., Sayer, C., Isupov, M. N., Annovazzi, C., Marchesi, C., Iacobone, G., et al. (2015). Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. *FEBS J.* 282, 2879–2894. doi: 10.1111/febs.13328
- Gaur, N., Flora, G., Yadav, M., and Tiwari, A. (2014). A review with recent advancements on bioremediation-based abolition of heavy metals. *Environ. Sci.* 16, 180–193. doi: 10.1039/c3em00491k
- Ghelani, A., Patel, R., Mangrola, A., and Dudhagara, P. (2015). Cultivation-independent comprehensive survey of bacterial diversity in Tulsi Shyam Hot Springs, India. *Genom Data* 4, 54–56. doi: 10.1016/j.gdata.2015.03.003
- Google Maps (2016). *Anhoni, Madhya Pradesh to Tattapani*, Chhattisgarh. Available online at: <https://www.google.co.in/maps/dir/Tattapani,+Chhattisgarh/Anhoni,+Madhya+Pradesh+480559/@23.4008419,79.7856623,6z/data=!4m13!4m12!1m5!1m1!1s0x398be11dce5307e3:0x583e563bac1615ab!2m2!1d83.6840385!2d23.6985978!1m5!1m1!1s0x397e35860faf2e99:0x8f442263b36683f8!2m2!1d78.6058242!2d22.5879396>
- Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., et al. (2012). The genome portal of the department of energy joint genome institute. *Nucleic Acids Res.* 40, D26–D32. doi: 10.1093/nar/gkr947
- Gudbergdottir, S. R., Menzel, P., Krogh, A., Young, M., and Peng, X. (2016). Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. *Environ. Microbiol.* 18, 863–874. doi: 10.1111/1462-2920.13079
- Gupta, A., Kumar, S., Prasoodanan, V. P., Harish, K., Sharma, A. K., and Sharma, V. K. (2016). Reconstruction of bacterial and viral genomes from multiple metagenomes. *Front. Microbiol.* 7:469. doi: 10.3389/fmicb.2016.00469
- Gupta, A., and Sharma, V. K. (2015). Using the taxon-specific genes for the taxonomic classification of bacterial genomes. *BMC Genomics* 16:1. doi: 10.1186/s12864-015-1542-0
- Hannigan, R. E., and Basu, A. R. (1998). "Late diagenetic trace element remobilization in organic-rich black shales of the Taconic foreland basin of Quebec, Ontario and New York," in *Shales and Mudstones II*, eds J. Schieber, W. Zimmerle, and P. S. Sethi (Zurich: Schweizerbart'sche), 209–234.
- Hannigan, R. E., Basu, A. R., and Teichmann, F. (2001). Mantle reservoir geochemistry from statistical analysis of ICP-MS trace element data of equatorial mid-Atlantic MORB glasses. *Chem. Geol.* 175, 397–428. doi: 10.1016/S0009-2541(00)00335-1
- Head, I. M., Jones, D. M., and Røling, W. F. (2006). Marine microorganisms make a meal of oil. *Nat. Rev. Microbiol.* 4, 173–182. doi: 10.1038/nrmicro1348
- Heulin, T., Barakat, M., Christen, R., Lesourd, M., Sutra, L., De Luca, G., et al. (2003). *Ramlibacter tataouinensis* gen. nov., sp. nov., and *Ramlibacter henchirensis* sp. nov., cyst-producing bacteria isolated from subsurface soil in Tunisia. *Int. J. Syst. Evol. Microbiol.* 53(Pt 2), 589–594. doi: 10.1099/ijs.0.02482-0
- Jimenez, D. J., Andreote, F. D., Chaves, D., Montana, J. S., Osorio-Forero, C., Junca, H., et al. (2012). Structural and functional insights from the metagenome of an acidic hot spring microbial planktonic community in the Colombian Andes. *PLoS ONE* 7:e52069. doi: 10.1371/journal.pone.0052069
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kanoksilapatham, W., Keawram, P., Gonzalez, J. M., and Robb, F. T. (2015). Isolation, characterization, and survival strategies of *Thermotoga* sp. strain PD524, a hyperthermophile from a hot spring in Northern Thailand. *Extremophiles* 19, 853–861. doi: 10.1007/s00792-015-0761-2
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202
- Khan, S. D., and Jacobson, S. (2008). Remote sensing and geochemistry for detecting hydrocarbon microseepage. *Geol. Soc. Am. Bull.* 120, 96–105. doi: 10.1130/0016-7606(2008)120[96:RSAGFD]2.0.CO;2
- Kim, J.-S., Kluskens, L. D., de Vos, W. M., Huber, R., and van der Oost, J. (2004). Crystal structure of fervidolysin from *Fervidobacterium pennivorans*, a keratinolytic enzyme related to subtilisin. *J. Mol. Biol.* 335, 787–797. doi: 10.1016/j.jmb.2003.11.006
- Kimes, N. E., Callaghan, A. V., Aktas, D. F., Smith, W. L., Sunner, J., Golding, B. T., et al. (2013). Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front. Microbiol.* 4:50. doi: 10.3389/fmicb.2013.00050
- Lalucat, J., Bennasar, A., Bosch, R., Garcia-Valdes, E., and Palleroni, N. J. (2006). Biology of *Pseudomonas stutzeri*. *Microbiol. Mol. Biol. Rev.* 70, 510–547. doi: 10.1128/MMBR.00047-05
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, S. J., Hua, Z. S., Huang, L. N., Li, J., Shi, S. H., Chen, L. X., et al. (2014). Microbial communities evolve faster in extreme environments. *Sci. Rep.* 4:6205. doi: 10.1038/srep06205

- Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.
- Littlechild, J. A. (2015). Enzymes from extreme environments and their industrial applications. *Front. Bioeng. Biotechnol.* 3:161. doi: 10.3389/fbioe.2015.00161
- Lopez-Lopez, O., Cerdan, M. E., and Gonzalez-Siso, M. I. (2013). Hot spring metagenomics. *Life* 3, 308–320. doi: 10.3390/life3020308
- Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507
- Mangrola, A., Dudhagara, P., Koringa, P., Joshi, C. G., Parmar, M., and Patel, R. (2015). Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. *Genom Data* 4, 153–155. doi: 10.1016/j.gdata.2015.04.014
- Mardanov, A. V., Gumerov, V. M., Beletsky, A. V., Perevalova, A. A., Karpov, G. A., Bonch-Osmolovskaya, E. A., et al. (2011). Un archaea dominate in the thermal groundwater of Uzon Caldera, Kamchatka. *Extremophiles* 15, 365–372. doi: 10.1007/s00792-011-0368-1
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Mehetre, G. T., Paranjpe, A. S., Dastager, S. G., and Dharne, M. S. (2016). Complete metagenome sequencing based bacterial diversity and functional insights from basaltic hot spring of Unkeshwar, Maharashtra, India. *Genom Data* 7, 140–143. doi: 10.1016/j.gdata.2015.12.031
- Menzel, P., Gudbergdottir, S. R., Rike, A. G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* 70, 411–424. doi: 10.1007/s00248-015-0576-9
- Mirete, S., Morgante, V., and Gonzalez-Pastor, J. E. (2016). Functional metagenomics of extreme environments. *Curr. Opin. Biotechnol.* 38, 143–149. doi: 10.1016/j.copbio.2016.01.017
- Mongenot, T., Tribouillard, N.-P., Desprairies, A., Lallier-Vergès, E., and Laggoun-Defarge, F. (1996). Trace elements as palaeoenvironmental markers in strongly mature hydrocarbon source rocks: the Cretaceous La Luna Formation of Venezuela. *Sediment. Geol.* 103, 23–37. doi: 10.1016/0037-0738(95)00078-X
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40, e155. doi: 10.1093/nar/gks678
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., et al. (2013). *Package 'Vegan'*. Community Ecology Package, Version 2.
- Oliveira, A. P., Patil, K. R., and Nielsen, J. (2008). Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Syst. Biol.* 2:17. doi: 10.1186/1752-0509-2-17
- Orphan, V. J., Taylor, L. T., Hafenbradl, D., and Delong, E. F. (2000). Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Appl. Environ. Microbiol.* 66, 700–711. doi: 10.1128/AEM.66.2.700-711.2000
- Pal, P., and Bhimasankaram, V. (1976). Tectonics of the Narmada-Son-Brahmaputra lineament. *Misc. Publ. Geol. Surv. Ind* 34, 133–140.
- Pandey, O., and Negi, J. (1995). "Geothermal fields of India: a latest update," in *Proceedings World Geothermal Congress* (Florence), 163–171.
- Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619. doi: 10.1371/journal.pone.0030619
- Poli, A., Nicolaus, B., Chan, K. G., Kahar, U. M., Chan, C. S., and Goh, K. M. (2015). Genome Sequence of *Anoxybacillus thermarum* AF/04^T, isolated from the euganean hot springs in Abano Terme, Italy. *Genome Announc.* 3:e00490-15. doi: 10.1128/genomeA.00490-15
- Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., et al. (2014). eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* 42, D231–D239. doi: 10.1093/nar/gkt1253
- Price, L. C. (1986). "A critical overview and proposed working model of surface geochemical exploration," in *Unconventional Methods in Exploration for Petroleum and Natural Gas, Symposium IV*. Dallas, TX: Southern Methodist University Press.
- Sarolkar, P. (2010). Exploration strategy for hot springs associated with gondwana coalfields in India. *Proc. World Geother. Congr.* 74, 68.
- Sarolkar, P. (2015). "Strategy for development of geothermal resources in India," in *Proceedings World Geothermal Congress* (Melbourne, VIC).
- Sarolkar, P., and Das, A. (2006). "Reservoir studies at Tatapani geothermal field, Surguja district, India," in *Proceedings of 31st Workshop on Geothermal Reservoir Engineering* (Stanford, CA, Stanford University, SGP-TR-179).
- Saxena, R., Chaudhary, N., Dhakan, D. B., and Sharma, V. K. (2015). Draft genome sequence of *Gulbenkiania mobilis* strain MB1, a sulfur-metabolizing thermophile isolated from a hot spring in Central India. *Genome Announc.* 3:e01295-15. doi: 10.1128/genomeA.01295-15
- Shanker, R. (1986). Scope of utilisation of geothermal energy for area development in backward, Hilly and Tribal regions of India. *Indian Miner.* 40, 49–61.
- Sharma, A., Hira, P., Shakarad, M., and Lal, R. (2014). Draft genome sequence of *Cellulosimicrobium* sp. Strain MM, isolated from arsenic-rich microbial mats of a himalayan hot spring. *Genome Announc.* 2:e01020-14. doi: 10.1128/genomeA.01020-14
- Singleton, D. R., Ramirez, L. G., and Aitken, M. D. (2009). Characterization of a polycyclic aromatic hydrocarbon degradation gene cluster in a phenanthrene-degrading *Acidovorax* strain. *Appl. Environ. Microbiol.* 75, 2613–2620. doi: 10.1128/AEM.01955-08
- Soergel, D. A., Dey, N., Knight, R., and Brenner, S. E. (2012). Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J.* 6, 1440–1444. doi: 10.1038/ismej.2011.208
- Stetter, K. O. (1996). Hyperthermophilic procaryotes. *FEMS Microbiol. Rev.* 18, 149–158. doi: 10.1111/j.1574-6976.1996.tb00233.x
- Stetter, K. O. (1999). Extremophiles and their adaptation to hot environments. *FEBS Lett.* 452, 22–25. doi: 10.1016/S0014-5793(99)00663-8
- Tedesco, S. A. (2012). *Surface Geochemistry in Petroleum Exploration*. New York, NY: Springer Science & Business Media.
- Tekere, M., Lötter, A., Olivier, J., Jonker, N., and Venter, S. (2011). Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa. *Afr. J. Biotechnol.* 10, 18005–18012. doi: 10.5897/AJB11.899
- Tindall, B. J., Sikorski, J., Lucas, S., Goltsman, E., Copeland, A., Del Rio, T. G., et al. (2010). Complete genome sequence of *Meiothermus ruber* type strain (21 T). *Stand. Genomic Sci.* 3, 26. doi: 10.4056/signs.1032748
- Vaidya, D., Shah, M., Sircar, A., Sahajpal, S., and Dhale, S. (2015). Geothermal energy: exploration efforts in India. *Int. J. Latest Res. Sci. Technol.* 4, 61–69.
- Wang, Y., and Qian, P. Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS ONE* 4:e7401. doi: 10.1371/journal.pone.007401
- Wang, Y., Wang, X., Tang, R., Yu, S., Zheng, B., and Feng, Y. (2010). A novel thermostable cellulase from *Fervidobacterium nodosum*. *J. Mol. Catal. B Enzym.* 66, 294–301. doi: 10.1016/j.molcatb.2010.06.006
- Ward, N., Larsen, O., Sakwa, J., Bruseth, L., Khouri, H., Durkin, A. S., et al. (2004). Genomic insights into methanotrophy: the complete genome sequence of *Methylococcus capsulatus* (Bath). *PLoS Biol.* 2:e303. doi: 10.1371/journal.pbio.0020303
- Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S. M., Meng, J., et al. (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J.* 5, 414–426. doi: 10.1038/ismej.2010.144
- Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38:e132. doi: 10.1093/nar/gkq275

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Saxena, Dhakan, Mittal, Waiker, Chowdhury, Ghatak and Sharma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Aerobic Lineage of the Oxidative Stress Response Protein Rubrerythrin Emerged in an Ancient Microaerobic, (Hyper)Thermophilic Environment

Juan P. Cardenas^{1,2†}, Raquel Quatrini³ and David S. Holmes^{1,2*}

¹ Center for Bioinformatics and Genome Biology, Fundacion Ciencia & Vida, Santiago, Chile, ² Facultad de Ciencias Biologicas, Universidad Andres Bello, Santiago, Chile, ³ Laboratory of Microbial Ecophysiology, Fundación Ciencia & Vida, Santiago, Chile

OPEN ACCESS

Edited by:

Kian Mau Goh,
Universiti Teknologi Malaysia,
Malaysia

Reviewed by:

William P. Inskeep,
Montana State University, USA
Amy Michele Grunden,
North Carolina State University, USA
Xi-Ying Zhang,
Shandong University, China

*Correspondence:

David S. Holmes
dsholmes2000@yahoo.com

† Present address:

Juan P. Cardenas,
uBiome, Inc., San Francisco, CA,
USA

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 03 August 2016

Accepted: 31 October 2016

Published: 18 November 2016

Citation:

Cardenas JP, Quatrini R and
Holmes DS (2016) Aerobic Lineage
of the Oxidative Stress Response
Protein Rubrerythrin Emerged in an
Ancient Microaerobic,
(Hyper)Thermophilic Environment.
Front. Microbiol. 7:1822.
doi: 10.3389/fmicb.2016.01822

Rubrerythrins (RBRs) are non-heme di-iron proteins belonging to the ferritin-like superfamily. They are involved in oxidative stress defense as peroxide scavengers in a wide range of organisms. The vast majority of RBRs, including classical forms of this protein, contain a C-terminal rubredoxin-like domain involved in electron transport that is used during catalysis in anaerobic conditions. Rubredoxin is an ancient and large protein family of short length (<100 residues) that contains a Fe-S center involved in electron transfer. However, functional forms of the enzyme lacking the rubredoxin-like domain have been reported (e.g., sulerythrin and ferriperoxin). In this study, phylogenomic evidence is presented that suggests that a complete lineage of rubrerythrins, lacking the rubredoxin-like domain, arose in an ancient microaerobic and (hyper)thermophilic environments in the ancestors of the Archaea Thermoproteales and Sulfolobales. This lineage (termed the “aerobic-type” lineage) subsequently evolved to become adapted to environments with progressively lower temperatures and higher oxygen concentrations via the acquisition of two co-localized genes, termed DUF3501 and RFO, encoding a conserved protein of unknown function and a predicted Fe-S oxidoreductase, respectively. Proposed Horizontal Gene Transfer events from these archaeal ancestors to Bacteria expanded the opportunities for further evolution of this RBR including adaption to lower temperatures. The second lineage (termed the cyanobacterial lineage) is proposed to have evolved in cyanobacterial ancestors, maybe in direct response to the production of oxygen via oxygenic photosynthesis during the Great Oxygen Event (GOE). It is hypothesized that both lineages of RBR emerged in a largely anaerobic world with “whiffs” of oxygen and that their subsequent independent evolutionary trajectories allowed microorganisms to transition from this anaerobic world to an aerobic one.

Keywords: rubrerythrin, evolution, phylogeny, comparative genomics, microaerophilic, hyperthermophiles, GOE, cyanobacteria

INTRODUCTION

The ability to combat oxidative stress is a widespread feature found in most organisms including many obligate anaerobic and microaerophilic organisms. The Great Oxygenation Event (GOE), that has been hypothesized to occur approximately 2.3 billion years ago (Lyons et al., 2014), would likely have initiated an adaptation process to attenuate the threatening exposure to increased levels of reactive oxygen species (ROS). The development of efficient ROS scavenging mechanisms would have facilitated the co-evolution of redox proteins that could take advantage of the energetically advantageous use of oxygen as a terminal electron acceptor.

Reactive oxygen species are partially reduced oxygen compounds that are produced as byproducts of oxygen reduction and, in many conditions, their exacerbated production can lead to severe stress and even cellular death (Imlay, 2003). Several mechanisms have evolved to mitigate the oxidative stress caused by ROS including direct mechanisms and indirect scavenging of ROS (Farr and Kogoma, 1991; Maaty et al., 2009; Zuber, 2009). The most stable ROS is hydrogen peroxide (H_2O_2) and a variety of enzymes have evolved to remove it, such as catalases, peroxiredoxins and other peroxidases (Mishra and Imlay, 2012). Among the H_2O_2 -scavenging enzymes, one variant with an important role in stress survival in several microorganisms, is rubrerythrin.

Rubrerythrin (RBR) is a member of the Ferritin-like superfamily (FLSF), consisting of proteins with a variety of different functions such as iron storage and/or iron detoxification (ferritins, Dps proteins, bacterioferritins), ubiquinone biosynthesis (COQ7 proteins), and radical scavenging (Mn-catalases, RBRs; Andrews, 2010). These proteins possess a four-helical bundle, configured in two anti-parallel helix pairs, forming a di-iron center, mediated by the coordination of (at least) six highly conserved residues among the four-helix structure (Andrews, 2010). In RBR, the di-iron center is responsible for the reduction of H_2O_2 and organic hydroperoxide (Dillard et al., 2011). The role of RBR in the reduction of H_2O_2 has been experimentally verified in several different organisms including aerobes (Sato et al., 2012), cyanobacteria (Zhao et al., 2007), and also obligate anaerobes (Weinberg et al., 2004). The classic RBR, found in the obligate anaerobe *Desulfovibrio vulgaris* contains, in addition to the FLSF domain, a C-terminal domain related to the rubredoxin family, putatively involved in electron transfer during catalysis (Andrews, 2010). This feature is conserved in most representatives of this family (Iyer et al., 2005), including a version of RBR with its rubredoxin domain in the opposite orientation with respect to the classic version called “reverse RBR” (Andrews, 2010). This suggests that the short rubredoxin-like domain has an important role in RBR function.

Despite the strong conservation of the rubredoxin-like domain in RBR proteins, there are also functional versions of this enzyme without the rubredoxin-like domain such as sulerythrin and ferriperoxin from *Sulfolobus tokodaii* (Wakagi, 2003) and *Hydrogenobacter thermophilus* (Sato et al., 2012). This one-domain RBR, composed only of the FLSF domain, was found to be widely distributed in microbial organisms, raising questions

about how it evolved and whether the lack of the C-terminal rubredoxin-family domain had functional implications for the catalysis of H_2O_2 . However, in obligate anaerobes, a functional and physical association between RBR and other proteins was demonstrated, suggesting that some other proteins might supply the missing catalytic function to the one-domain RBR (Weinberg et al., 2004; Le Fourn et al., 2011).

In this report, we have undertaken a phylogenomic analysis of RBRs using a number of techniques including conventional phylogenetic approaches, sequence similarity networks and comparisons of genomic neighborhoods. The objective was to derive a plausible trajectory of the evolution of RBRs and, if possible, to link this trajectory with postulated changes of temperature and atmospheric oxygen concentrations during the early stages of the earth over 3 billion years ago.

MATERIALS AND METHODS

Compilation of RBR Sequences

Rubrerythrin sequences were obtained from NCBI non-redundant (NR) database using a two-step filter: first, the NR database was analyzed using *HMMsearch* (*HMMer* version 3.0) against PF02915 (PFAM domain for Rubrerythrin, E -value $< 10^{-6}$) followed by a RPS-BLAST search against COG, recovering all proteins with significant similarity to COG1592 (E -value $< 10^{-10}$); all sequences with lengths less than 100 residues and/or with COG coverage values less than 70% were discarded. **Supplementary Table S1** lists the sequences found using this strategy.

Sequence Similarity Network

The sequence similarity network elaboration was as described by Atkinson et al. (2009). A total of 4527 RBR sequences found by the aforementioned method were clustered using CD-HIT software (Fu et al., 2012), resulting in 2631 representatives comprising groups defined by 90% identity. These filtered sequences were analyzed in a BLASTp-all-versus-all round (no-filter, default parameters) with a threshold E -value of 10^{-35} . The pairwise bit scores were used as measure of distance for the network, visualized in Cytoscape 3.0.2 using *Organic layout*.

Phylogenetic Analyses

Protein sequences were aligned using MAFFT (Katoh and Standley, 2014) with the L-INS strategy. The phylogenetic analyses were performed using either maximum likelihood (ML) or Bayesian inferences (BI). For ML, the trees were elaborated in PhyML, version 3.1 (Guindon et al., 2009) with different sets of parameters for each case. For BI, the trees were elaborated in MrBayes version 3.2.6 (Huelsenbeck and Ronquist, 2001), with different parameters for each case (see “Results”). The selected substitution models for both ML and BI analyses were selected using ModelGenerator (Keane et al., 2006).

For the “aerobic-type” RBR tree, the alignment containing 127 ungapped and unambiguous sites for 335 taxa was analyzed by MrBayes with five substitution rate categories and gamma-distributed rate variation, using LG as the prior model. Bayesian

analysis was run for five million generations (in two independent runs, using four chains and a heating parameter of 0.1); trees were saved every 200 generations and posterior probabilities calculated after discarding the first 33% of trees.

For the DUF3501 and the rubrerythrin-associated Fe-S oxidoreductase (RFO) protein families (see below), the trees were constructed using PhyML. For DUF3501, the alignment containing 118 ungapped and unambiguous sites for 258 taxa was used; the ML-based tree was performed using the LG model with four substitution rate categories and gamma-distributed rate variation, with a proportion of invariable sites. The tree topology search was performed by the combination of NNIs and SPRs strategies and the approximate likelihood ratio was computed by the SH-like branch support test (Anisimova et al., 2011). The RFO tree was computed using similar parameters, from an alignment containing 289 ungapped and unambiguous sites for 206 taxa. Additionally, a tree of the “cyanobacterial group” RBR orthologs was computed from an alignment containing 289 ungapped and unambiguous sites for 206 taxa, using the same configuration used for the two latter phylogenetic analyses.

Other Analyses

Genomic contexts for the genes encoding rubrerythrin were obtained from IMG-JGI (Markowitz et al., 2012), using a set of 4461 complete prokaryotic genomes. Protein domain analyses were carried out using InterProScan (Quevillon et al., 2005). Co-localized genes with RBR were searched in the complete prokaryotic genome collection using BLAST (Altschul et al., 1997).

RESULTS AND DISCUSSION

Identification of Different Rubrerythrin Groups

In order to study the evolution of RBR forms, the first step was the preparation of a trustworthy set of these proteins. The search of putative RBR homologs was made using a dual criterion selection that combines the sensibility of Hidden Markov Model (HMM) based search (PF02915) with a confirmative application of RPS-BLAST against the COG1592 family (corresponding to the RBR family). This double filter allowed a discrimination to be made between RBRs and other members of the FLSF, including members of the Mg²⁺-protoporphyrin IX monomethyl ester cyclase subfamily, which were also included in the HMM profile for PF02915.

A sequence similarity network approach was implemented in order to identify subfamilies inside the RBR family. Similarity networks are acceptable approximations for the study of huge protein families, although they are not a replacement for phylogenetic studies (Atkinson et al., 2009). The use of a given *E*-value (or score) threshold in an “all-vs-all” BLAST permits the recovery of clusters that can be visualized in a network fashion, where each node is an homolog or family member, and each edge is the measure of the pairwise score obtained by BLAST. The distances between different node groups are inversely proportional to the value of their pairwise scores: higher

BLAST pairwise scores between multiple sequences promote their clustering, and vice versa.

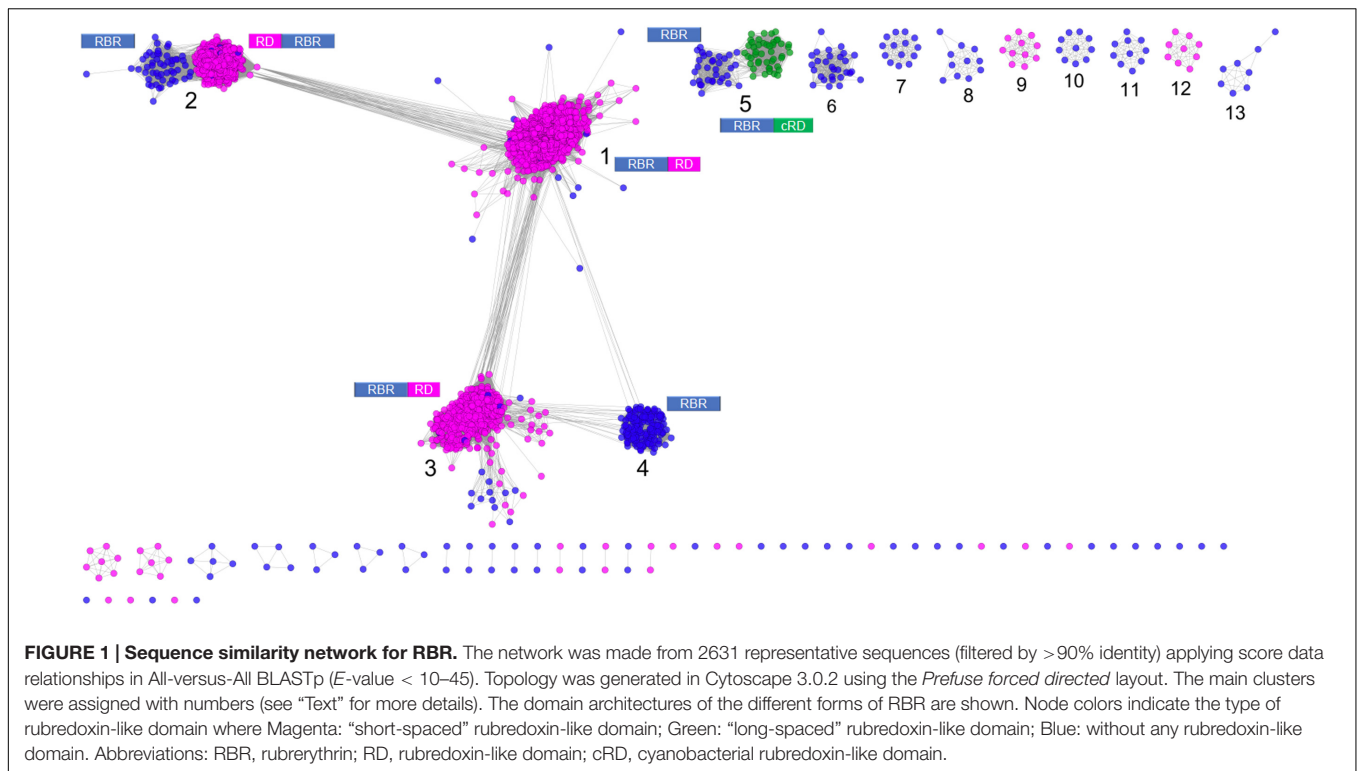
Figure 1 shows a sequence similarity network for the RBR orthologs retrieved from the NR database following the strategy outlined above. From this network (comprising 2631 representative genes), a set of node clusters or groups can be inferred. Four large groups (1–4) connected in a major network were recovered using an *E*-value threshold ($E < 10^{-45}$). In addition, there are nine minor isolated groups (5–13), and several smaller groups with disconnected nodes, indicating that, at the BLAST *E*-value used, those sequences have no similarity with other members of the RBR protein set. Group 5 is included with groups 1–4 in subsequent analyses. Groups 1–5 comprise ~93% of the sequences.

Group 1 (**Figure 1**) corresponds to the principal group of canonical rubrerythrins, and is comprised almost completely (98.7 % of 1336 sequences) by members with the rubredoxin-like domain with a relatively short spacing between the two C-x-x-C motifs (10–14 residues, we term “short-spaced”). Taxonomic information (**Supplementary Figure S1**) indicates that this group is dominated by members of *Clostridia* (44.5%) and δ/ϵ -*Proteobacteria* classes (11.0%) and the *Bacteroidetes* phylum (10.2%). Comparison with PDB database showed that the classical RBR protein from *D. vulgaris* (Swissprot accession P24931, PDB entry 1B71) is a member of this group. Therefore, we propose that group 1 is termed the “Classical RBR” group. The group is populated predominately by obligate anaerobic microorganisms.

Group 2 (**Figure 1**) consists of two sub-clusters: the first (85.3% of the total sequences) exhibits the rubredoxin domain as do group 1 rubredoxins, but unlike group 1 this rubredoxin domain is located in the N terminal of the protein as previously observed in *Clostridium acetobutylicum* (Riebe et al., 2009). The other sub-cluster (14.7% of the total sequences) does not have a rubredoxin domain. Group 2 was dominated obligate anaerobes such as members of the *Clostridia* class (64.5 %) and *Bacteroidetes* phylum (10.5%; see **Supplementary Figure S1**). We propose that group 2 be termed the “Reverse-type” RBR group.

Group 3 (**Figure 1**) is mainly composed of RBR with the C-terminal rubredoxin-like domain and, according to taxonomic information, is dominated by members of δ/ϵ -*Proteobacteria* (25.97%) and *Clostridia* (20.7%) classes and the *Euryarchaeota* phylum (17.31%). Comparison with PDB showed that a previously crystallized RBR from *Pyrococcus furiosus* belongs to this group (Uniprot accession Q9UWP7, PDB entry 1NNQ); this RBR contains canonical hydrogen peroxide reductase activity (Dillard et al., 2011). We propose that this group is termed the “Classical group B.” Like group 1, both the “reverse” (group 2) and the “classical-B” types of RBR (group 3) are enriched in taxonomic groups dominated by obligate anaerobes.

Group 4 (**Figure 1**) is composed exclusively of genes lacking the rubredoxin-like domain. This group is dominated by members of $\alpha/\beta/\gamma$ -*Proteobacteria* (61.1%), *Actinobacteria* (9.0%), and *Crenarchaeota* (including members of *Sulfolobales* and *Thermoproteales* orders, 8.3%), as well as bacterial members from *Aquificae* (including the ferriperoxin from *H. thermophilus*), *Nitrospirae* and some few members of



Firmicutes, Deltaproteobacteria, and archaeal members of Thermoplasmatales (from the Euryarchaeota phylum). Unlike the three aforementioned groups, members of group 4 are dispersed in taxonomic groups dominated by (facultative) aerobic organisms. Sulrerythrin from *S. tokodaii* (Uniprot accession F9VPE5; PDB entry 1J30) and a rubrerythrin from *Burkholderia pseudomallei* (Uniprot accession Q3JK2; PDB entry 4DI0) have three dimensional structures in the PDB. The presence of RBRs without the rubredoxin-like domain is distributed in different sequence clusters of this network, suggesting that these proteins arose independently more than once in evolution. The detection of a well-defined group composed only of members without the aforementioned domain, and additionally, that those homologs are mostly associated with aerobic organisms, strongly suggest that this group is a complete and distinctive evolutionary lineage with common properties and we term it the “aerobic-type” group of RBR.

Group 5 (Figure 1) contains two well-compartmentalized sub-clusters; one of them (comprising 41.8% of the sequences) lacks the rubredoxin-like domain. Members of the other sub-cluster (comprising the 58.2% of the sequences) is distinguished by having a C-terminal rubredoxin-like domain, with a longer spacing between the cysteines motifs (27–32 residues) compared with the C-terminal domain of the classical RBR (10–14 residues). Taxonomically, the group 5 is dominated by members of Cyanobacteria (59.5%) and $\alpha/\beta/\gamma$ -Proteobacteria (36.7%). The “long-spaced” rubredoxin-like domain is present in almost all cyanobacterial RBRs and has a different evolutionary origin from the other RBRs. Due to the prevalence of cyanobacterial RBR genes in group 5, we term this the “Cyanobacterial-type” RBR.

Given their evolutionary properties, the phylogenetic analysis of this group, and the origin of the “long” rubredoxin domain will be covered below.

Phylogenetic and Genomic Analyses of the “Aerobic-Type” RBR

In order to analyze the evolutionary history of this proposed lineage, a phylogenetic tree was constructed using BI (Figure 2A in summary and Supplementary Figure S2 in detail). The “aerobic-type” RBR phylogenetic tree (using an appropriate outgroup) showed that the closest clade to the root of the group comprises a set of orthologs from archaeal members (Figure 2B in summary and Supplementary Figure S2 in detail) from the *Vulcanisaeta*, *Caldivirga*, and *Thermoproteus* genera of the Thermoproteales order (Itoh et al., 1999; Mavromatis et al., 2010; Gumerov et al., 2011; Mardanov et al., 2011). All of the extant members of this group are hyperthermophiles and are slightly acidophilic (pH 3.0–6.5). They grow only in anaerobic or microaerobic conditions. For example, a well-supported clade contains orthologs from members of Actinobacteria (species from *Acidithrix*, *Ferrimicrobium*, and other related groups), Nitrospirae (species from *Leptospirillum* and *Nitrospira*), and Clostridia (*Sulfobacillus*), associated with extremely acidic (pH < 3) environments (Cardenas et al., 2016). Another well-supported clade contains only orthologs from members of the Sulfolobales order (from *Sulfolobus*, *Metallosphaera*, and *Acidianus* genera). It is interesting to note that this clade is paraphyletic with respect to other clades containing other orthologs from Archaea (for example, the clades containing orthologs from Thermoproteales or Thermoplasmatales). This

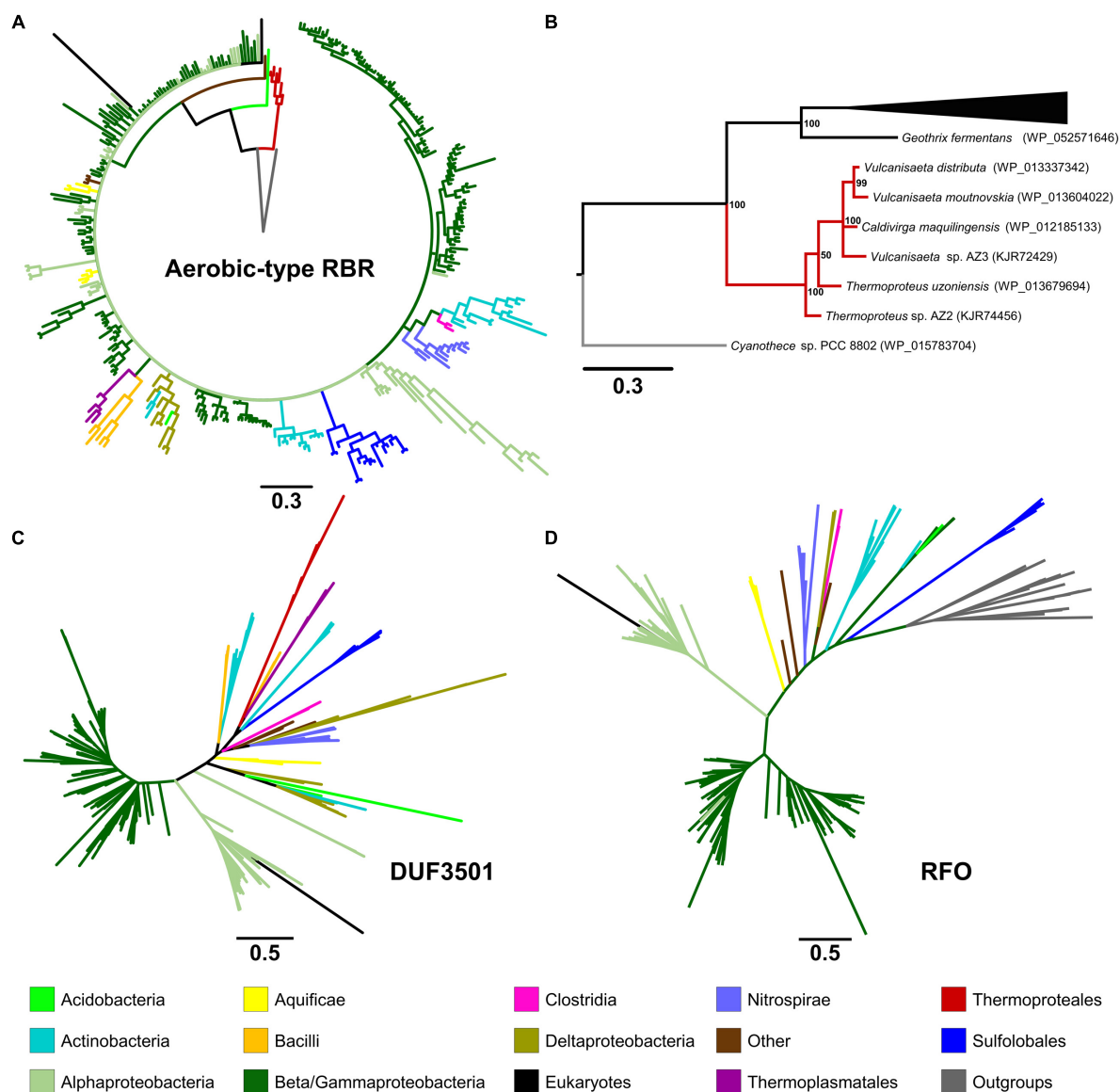


FIGURE 2 | Phylogenetic trees constructed in this study. The circular tree for representative members of the “aerobic-type” RBR (A) and (B) with a phylogram representation of the closest clade to the root of the tree (B). The figure also includes unrooted versions of the trees made from the protein sequences of DUF3501 (C) and RFO (D). For the tree of the “aerobic-type” RBR group, the sequence of a RBR from *Cyanothece* sp. PCC 8802 (a cyanobacterial RBR) was used as an outgroup taxon, whereas the tree for RFO used the sequence of GlpC from *Escherichia coli* as outgroup. For each case, the line length represents the respective value for phylogenetic distance. Organism names and more detailed trees are displayed in **Supplementary Figures S2–S5**.

suggests that the evolutionary history of this lineage of RBR involved multiple different horizontal transferences within archaeal organisms. Another noticeable observation inferable from this tree is the expansion of this RBR lineage inside in the Alpha-, Beta- and Gammaproteobacteria. Most importantly, the phylogenetic tree of RBR suggests that the most ancient group (the closest to the root) is composed of members from the Thermoproteales; this is consistent with the idea that this form of RBR evolved in a microaerobic and thermophilic environment.

Gene Context Analyses of Genes/Domains Predicted to Be Related to “Aerobic-Type” RBR Function

We have undertaken an analysis of the gene contexts of the “aerobic-type” of RBR in order to shed light on its evolution. A comparative visualization of RBR gene neighborhoods among complete genomes was carried out using the IMG-JGI database (Markowitz et al., 2012). The gene neighborhood comparison (Figure 3) showed that the majority of the RBR genes from

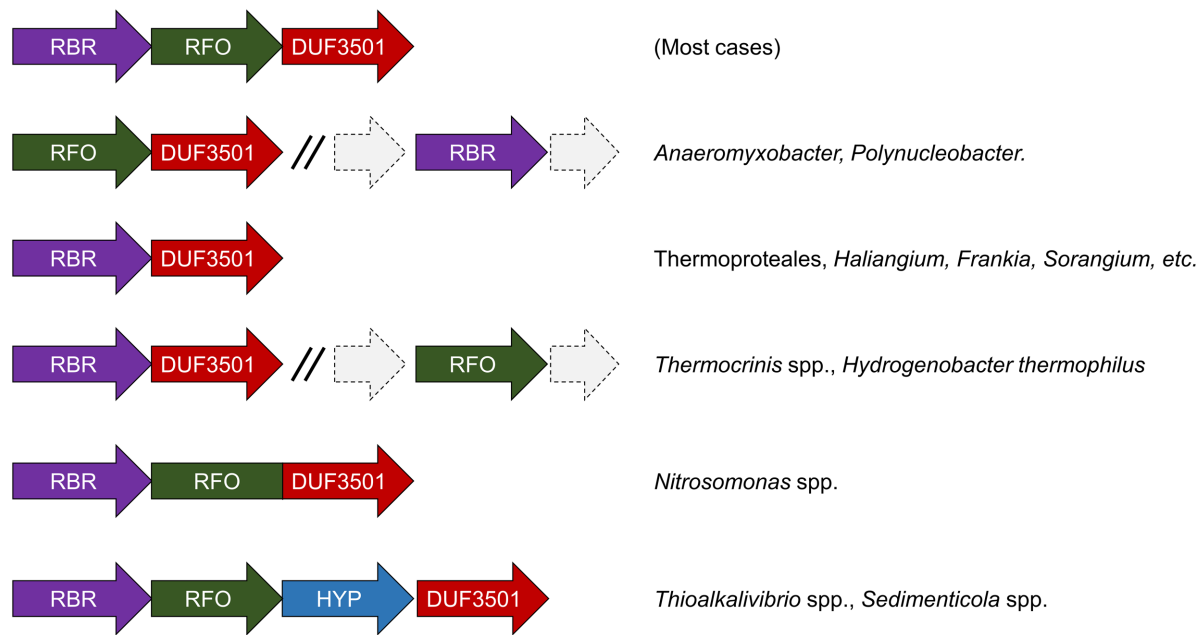


FIGURE 3 | Genomic neighborhoods of “aerobic” RBR in sequenced genomes. RBR, aerobic-type rubrerythrin; DUF3501, conserved protein of unknown function; RFO, Fe-S oxidoreductase; HYP, hypothetical protein.

the “aerobic-type” were co-localized in a conserved gene cluster with a gene encoding a protein of unknown function (DUF3501, pfam12007), and in several other cases, also co-localized with a gene predicted to encode a member of the COG0247 family (*Fe-S oxidoreductase*). The role of the DUF3501 protein is unknown. However, the function of the COG0247 family protein has been suggested to be related to electron transfer inside protein complexes, since it contains a conserved CCG domain (pfam02754) and a cysteine rich domain. The latter is present in several oxidoreductase complexes related to energy metabolism under anaerobic conditions such as subunit GlpC in the anaerobic glycerol-3-phosphate dehydrogenase (Cole et al., 1988) and subunits of the HdrD/E from the heterodisulfide reductase complex found in methanogens (Kunkel et al., 1997). We propose that this COG0247 family protein member associated with RBR be termed RFO (Rubrerythrin-associated Fe-S Oxidoreductase). In some instances, DUF3501- and RFO are fused into one gene (e.g., *Nitrosomonas* spp., **Figure 3**). In other organisms (*Thermocrinis* spp. and *H. thermophilus*), the RFO-encoding gene is separated from the other two genes. It is interesting to note that other members of the deeply rooted Aquificae phylum (such as *Hydrogenobaculum*) have the more common three-gene arrangement. This difference in gene cluster structure among members of the same phylum may be a reflection of the antiquity of the events that resulted in these different gene arrangements.

Additional comparisons suggest that the genes encoding DUF3501 and RFO are not only strongly co-localized with the “aerobic-type” RBR, but also are co-occurrent. An examination of a set of 4461 completed prokaryotic genomes from the IMG-JGI database showed that DUF3501 and RFO genes are not associated with RBR in only one organism (*Burkholderia cepacia* AMMD)

and only in one other genome (*Azoarcus* sp. CIB) a DUF3501-encoding gene was detected without the other two, RFO and RBR (**Supplementary Table S2**). The exclusive co-occurrence of RBR – DUF3501 was found in 25 organisms, and the triple co-occurrence of RBR, DUF3501, and RFO was detected in 185 genomes. Interestingly, the co-occurrence of RBR – DUF3501 was detected in the members of Thermoproteales, suggesting that this association could be as ancient as the last common ancestor of this group, supporting the contention that the origin of “aerobic-type” RBR lineage, occurred in a high-temperature and microaerobic environment.

Phylogenetic Analysis of DUF3501 and RFO Families

In order to investigate the possible co-evolution of the “aerobic-type” RBR with the DUF3501 and RFO genes, phylogenetic trees for DUF3501 (**Figure 2C** in summary and **Supplementary Figure S3** in detail) and RFO (**Figure 2D** in summary and **Supplementary Figure S4** in detail) were constructed. These trees are consistent with the late expansion of the respective protein families into the Alpha-, Beta- and Gammaproteobacteria, and with the observed expansion of the co-localizing “aerobic-type” RBR. The case of DUF3501 also showed the non-monophyletic association between the archaeal orthologs from the three different orders (Thermoproteales, Sulfolobales, and Thermoplasmatales), seen previously in the RBR tree. In the case of the RFO tree, BLAST searches of orthologs for this phylogenetic tree showed that the most closely related proteins are GlpCs that encode the subunit C of the anaerobic glycerol-3-phosphate dehydrogenase (involved in glycerol degradation under anaerobic conditions). The

application of GlpC as an outgroup showed that the most ancient group corresponds to Sulfolobales, an archaeal order composed of aerobic thermoacidophiles (Huber and Stetter, 2015), suggesting that both GlpC as well as the “aerobic-type” RBR, arose in high-temperature aerobic environments. Both phylogenetic trees for DUF3501 and RFO supporting the contention of the co-evolution of DUF3501 and RFO with the “aerobic-type” RBR.

The “Cyanobacterial Group”: A Parallel Adaptation of Rubrerythrin to an Aerobic Environment

Information obtained from phylogenomic analyses of the “aerobic-type” RBR and its co-occurrent genes suggests an evolutionary trajectory driven by the adaptation to aerobic environments, initiated from an (hyper)thermophilic ancestor. However, another group in the RBR sequence similarity network was also detected (the “cyanobacterial group,” group 5 from **Figure 1**) that was not connected to the other four groups and was composed mostly of members from Cyanobacteria (phototrophic oxygen-producers) and $\alpha/\beta/\gamma$ -Proteobacteria (facultative anaerobes), raising the question about how this protein group evolved. A phylogenetic tree for members of the “cyanobacterial RBR” group was made using ML (**Supplementary Figure S5**). This analysis showed two well-defined clades: a clade containing the cyanobacterial protein and a clade containing the proteobacterial orthologs. The members of the cyanobacterial clade contained a fusion of the FLSF of RBR with a “long-spaced” rubredoxin domain (data not shown). Based on the well-defined phylogenetic bifurcation and the coherent pattern of inheritance, the most parsimonious explanation is that the original RBR ancestor for this group was a one-domain RBR (as the “aerobic-type” RBR is), but in Cyanobacteria, the protein was fused with the “long-spaced” rubredoxin domain. We hypothesize that this adaptation had a relationship with the exposure to oxygen associated with the development of oxygenic photosynthesis in Cyanobacterial ancestor during GOE.

Evolution of the “Aerobic” and “Cyanobacterial” Rubrerythrins in the Early Earth: A Model

We propose a model (**Figure 4**) for the evolution of the “aerobic” and “cyanobacterial” rubrerythrins that is consistent with evidence provided by network information, phylogenetic trees and gene/domain contexts of the different forms of rubrerythrins. It is suggested that the “classical” form of rubrerythrin arose by gene fusion of a FLSF domain and a rubredoxin-like domain in a hot and anaerobic environment. This form is predicted to have existed in LUCA (Weiss et al., 2016). Extant obligate thermophilic and mesophilic anaerobes retain this form today [e.g., *Thermotoga maritima* (Lakhal et al., 2011) and *D. vulgaris* (Lumppio et al., 1997), respectively].

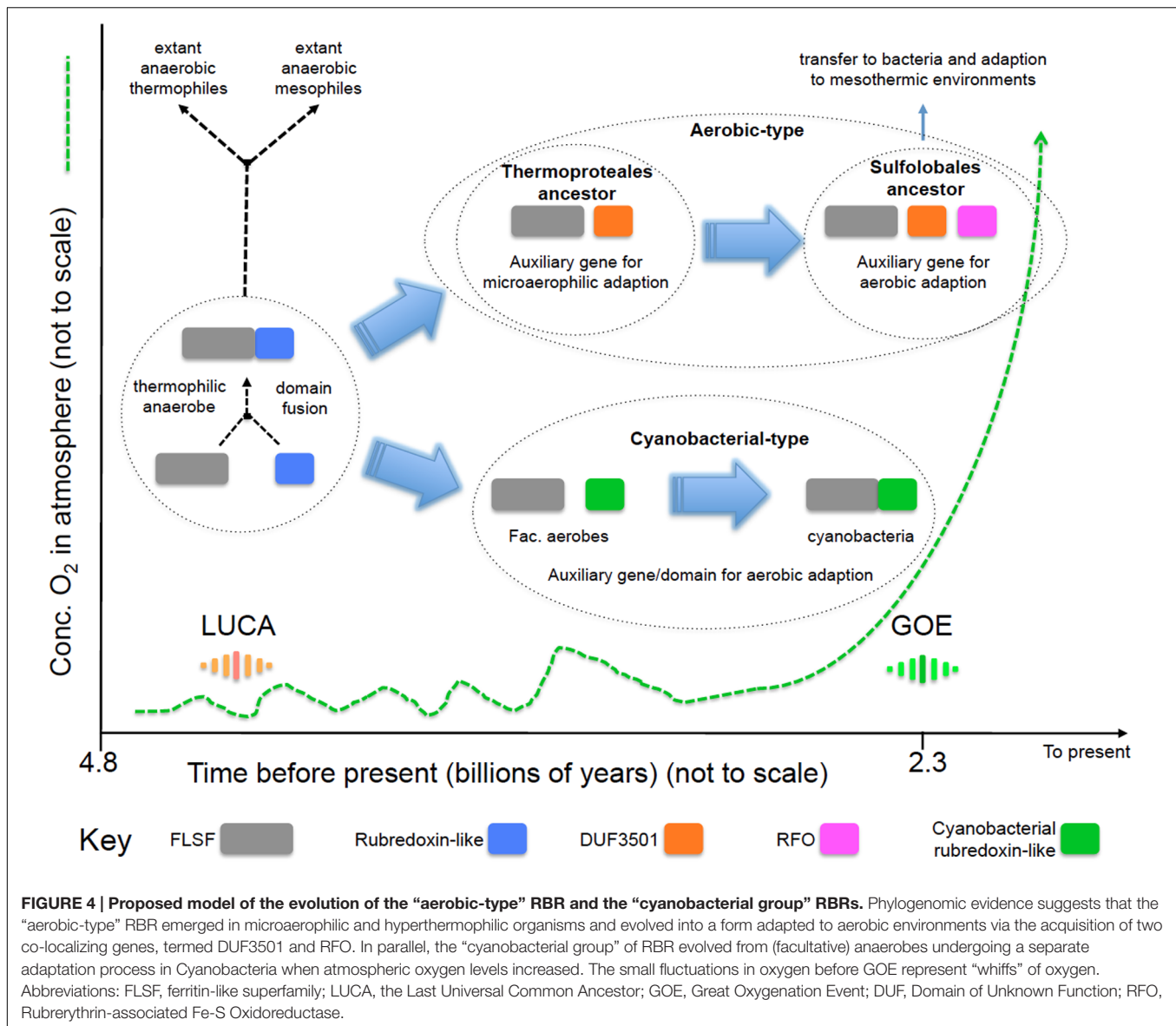
Two separate lines of evolution then occurred. In the first, that is postulated to have occurred in the ancestors of the thermophilic Archaeal group Thermoproteales, we hypothesize that the loss of the rubredoxin-like domain was compensated for by the

acquisition of DUF3501. This association is proposed to have provided protection in microaerobic environments that might have arisen before GOE due to exposure to low levels of oxygen. These low levels and perhaps transitory increments of oxygen could have originated from biotic or abiotic reactions, such as enzymatic ROS-detoxifying reactions or water UV photolysis (Martin and Sousa, 2016). It is probable that during those “oxygen whiffs,” many proteins would be the object of selection pressure to survive in oxygen that could force them to lose or gain new features such as protein domains (Nasir et al., 2014) or physiological activities. In addition, the ancient hot environment of the Archaean eon could have helped to accelerate the evolutionary rate of change of these proteins, as postulated to be occurring in current extreme environments (Li et al., 2014). Subsequently, an additional gene, RFO, was added that provided full function in aerobic environments perhaps resulting from GOE. This step is proposed to have occurred in the ancestors of the thermophilic Archaea Sulfolobales. This triple gene form was passed by HGT to the Bacteria and underwent further adaptation to work in mesothermic conditions (e.g., $<40^{\circ}\text{C}$). It is not known why the additional DUF3501 and RFO genes are not found fused as domains to the FLSF domain but rather they remain as conserved co-localizing but separate genes. Perhaps this architecture provides more opportunities for differential expression of the genes.

The second line, involved the replacement of the rubredoxin-like domain with a separately evolved cyanobacterial rubredoxin-like domain. It is hypothesized that this promoted the function of RBR in high levels of oxygen resulting from oxygenic photosynthesis that is thought to have evolved in the cyanobacterial lineage about the time of GOE.

Oxidative stress is an ancient and widespread phenomenon: several of the extant obligate anaerobes exhibit mechanisms to combat the presence of ROS (Jenney et al., 1999). One of those mechanisms is the use of RBR. This protein has been suggested to be an ancient protein (Andrews, 2010) predicted to be present in LUCA that was itself hypothesized to be an obligate anaerobe and thermophile (Martin and Sousa, 2016). The proposed antiquity of RBR makes it an interesting protein to be analyzed, establishing a case study of the effect of the arise of oxygen in the adaptation and origin of new protein (sub)families. It is interesting to note that the vast majority of the known RBRs have the “short-spaced” rubredoxin-like domain, even in a “reverse” form (**Figure 1**). This strongly suggest that, regardless that the very last common ancestor of all rubrerythrins lacked this rubredoxin domain, this fusion was implemented very early in evolution, and maybe even in different independent events (e.g., generating the “classic” and “reverse” RBR separately), and those successful, two-domain RBRs have been inherited by extant aerobic microorganisms.

There is phylogenomic evidence of ancient, massive gene birth events. Large-scale phylogenetic reconstruction of more than 3,000 gene families predicted a massive event of birth and loss of gene families occurring approximately $3.3\text{--}2.9 \times 10^9$ years ago (David and Alm, 2011), coinciding with or perhaps preceding, with the estimated time of GOE (Bekker, 2014). Another study (using a similar strategy) suggests that aerobic metabolism appeared about 2.9×10^9 years ago (Wang et al.,



2011). In that context, it is probable that this “aerobic-type” lineage of RBR, as well as their co-occurrent genes, could arise before GOE or just as GOE was starting. Regarding the case of the “cyanobacterial group” of RBR, the evidence provided by the sequence similarity network (group 5, **Figure 1**) suggests that this group of RBRs has a different evolutionary pathway from the “aerobic-form.”

It is interesting to note that the 3-genes form of RBR was most likely to have evolved in the ancestors of two archaeal orders: Thermoproteales and Sulfolobales. Both those two taxonomic orders belong to the Crenarchaeota phylum, and have a proposed ancient point of divergence (Gribaldo and Brochier-Armanet, 2006). Whereas DUF3501 may have arisen in temperatures over 85° C (temperature associated with the lifestyle of extant members of Thermoproteales), the RFO gene could have arisen in temperatures between 60–80°C (associated

with the lifestyle of extant Sulfolobales). We hypothesize that the acquisition of DUF3501 was part of an adaptation for function in combined microaerobic and hyperthermic conditions, whereas RFO evolving in direct response to increasing oxygen presence. One observation, consistent with this supposition, is that the RFO gene is not found in microaerobic organisms, such as members of Thermoproteales and some members of Deltaproteobacteria.

We suggest that, since RFO belongs to a group of proteins potentially involved in electron transfer, its linkage with the “aerobic-type” RBR can compensate for the loss of the rubredoxin-like domain, a component of the anaerobic RBRs that has a proposed role in electron transfer during catalysis (Kurtz, 2006). However, this also raises another question: how is the catalysis of electron transfer in the RBR form found in Thermoproteales, where DUF3501 but not RFO is present? This question remains to be experimentally addressed.

The proposed origin and evolution of the “aerobic-type” RBR is consistent with the hypothesis that early life evolved in a hot, anaerobic environment (Martin and Sousa, 2016). It also supports the idea that oxidative stress mechanisms could have been present before GOE. It also provides an exquisite opportunity to study the evolutionary trajectory of proteins as they adapt to increasing oxygen levels and to econiches with lower temperatures.

AUTHOR CONTRIBUTIONS

JC, RQ, and DH conceived the project. JC designed and carried out the experiments. All authors analyzed the data. JC wrote the first draft of the manuscript. All authors contributed to other drafts of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENT

This work was supported by Conicyt Basal CCTE PFB16 and Fondecyt 1130683 and 1140048.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01822/full#supplementary-material>

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Andrews, S. C. (2010). The Ferritin-like superfamily: evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim. Biophys. Acta* 1800, 691–705. doi: 10.1016/j.bbagen.2010.05.010
- Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., and Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* 60, 685–699. doi: 10.1093/sysbio/syr041
- Atkinson, H. J., Morris, J. H., Ferrin, T. E., and Babbitt, P. C. (2009). Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE* 4:e4345. doi: 10.1371/journal.pone.0004345
- Bekker, A. (2014). “Great oxygenation event,” in *Encyclopedia of Astrobiology*, eds R. Amils, M. Gargaud, J. Cernicharo Quintanilla, J. H. Cleaves, M. W. Irvine, D. Pinti, et al. (Berlin: Springer), 1–9.
- Cardenas, J. P., Quatrini, R., and Holmes, D. S. (2016). Genomic and metagenomic challenges and opportunities for bioleaching: a mini-review. *Res. Microbiol.* 167, 529–538. doi: 10.1016/j.resmic.2016.06.007
- Cole, S. T., Eiglmeier, K., Ahmed, S., Honore, N., Elmes, L., Anderson, W. F., et al. (1988). Nucleotide sequence and gene-polypeptide relationships of the glpABC operon encoding the anaerobic sn-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol.* 170, 2448–2456.
- David, L. A., and Alm, E. J. (2011). Rapid evolutionary innovation during an Archaeal genetic expansion. *Nature* 469, 93–96. doi: 10.1038/nature09649
- Dillard, B. D., Demick, J. M., Adams, M. W., and Lanzilotta, W. N. (2011). A cryo-crystallographic time course for peroxide reduction by rubrerythrin

FIGURE S1 | Phylogenetic distribution of rubrerythrin (RBR) types belonging to groups 1–5 (see “Text” of article). Sequences were retrieved from the network shown in **Figure 1** and were sorted by taxonomic origin (Phylum or Class) and counted.

FIGURE S2 | Phylogenetic distribution of the “aerobic” group of RBRs. The tree was elaborated as specified in “Material and Methods.” Each taxon is tagged with an NCBI Accession number shown in parentheses.

FIGURE S3 | Phylogenetic distribution of members of the DUF3501 protein family. The tree was elaborated as specified in “Material and Methods.” The branch support value is symbolized by the color of each node circle (color gradient legend shown). This tree is unrooted. Each taxon is tagged with an NCBI Accession number shown in parentheses.

FIGURE S4 | Phylogenetic distribution of members of the RFO protein family. The tree was elaborated as specified in “Material and Methods.” The branch support value is symbolized by the color of each node circle (color gradient legend shown). This tree was rooted using the sequence of GlpC from *Escherichia coli* as an outgroup. Each taxon is tagged with an NCBI Accession number shown in parentheses.

FIGURE S5 | Phylogenetic distribution of members of the “cyanobacterial group” of RBRs. The tree was elaborated as specified in “Material and Methods.” The branch support value is symbolized by the color of each node circle (color gradient legend shown). In the base of each great clade, the domain fusion/separation is specified. This tree was rooted using the sequence of an “aerobic-type” RBR derived from *Vulcanisaeta distributa* as an outgroup. Each taxon is tagged with an NCBI Accession number shown in parentheses. Abbreviation: RB, rubredoxin.

TABLE S1 | List of rubrerythrin (RBR) sequences found in NR database.

The table contains the NCBI accessions, protein length, and description for each sequence.

TABLE S2 | List of candidate aerobic rubrerythrins, DUF3501 and RFO in prokaryotic finished genomes from IMG-JGI database. This table includes a Venn diagram displaying the co-occurrence of these genes.

from *Pyrococcus furiosus*. *J. Biol. Inorg. Chem.* 16, 949–959. doi: 10.1007/s00775-011-0795-6

Farr, S. B., and Kogoma, T. (1991). Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev.* 55, 561–585.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gribaldo, S., and Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state of the art. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1007–1022. doi: 10.1098/rstb.2006.1841

Guindon, S., Delsuc, F., Dufayard, J. F., and Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* 537, 113–137. doi: 10.1007/978-1-59745-251-9_6

Gumerov, V. M., Mardanov, A. V., Beletsky, A. V., Prokofeva, M. I., Bonch-Osmolovskaya, E. A., Ravin, N. V., et al. (2011). Complete genome sequence of “*Vulcanisaeta moutnovskia*” strain 768-28, a novel member of the hyperthermophilic crenarchaeal genus *Vulcanisaeta*. *J. Bacteriol.* 193, 2355–2356. doi: 10.1128/JB.00237-11

Huber, H., and Stetter, K. O. (2015). “Sulfolobales,” in *Bergey’s Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (Chichester: John Wiley & Sons, Ltd).

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Imlay, J. A. (2003). Pathways of oxidative damage. *Annu. Rev. Microbiol.* 57, 395–418. doi: 10.1146/annurev.micro.57.030502.090938

Itoh, T., Suzuki, K., Sanchez, P. C., and Nakase, T. (1999). *Caldivirga maquilensis* gen. nov., sp. nov., a new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. *Int. J. Syst. Bacteriol.* 49(Pt 3), 1157–1163. doi: 10.1099/00207713-49-3-1157

- Iyer, R. B., Silaghi-Dumitrescu, R., Kurtz, D. M. Jr., and Lanzilotta, W. N. (2005). High-resolution crystal structures of *Desulfovibrio vulgaris* (Hildenborough) nigerythrin: facile, redox-dependent iron movement, domain interface variability, and peroxidase activity in the rubrerythrins. *J. Biol. Inorg. Chem.* 10, 407–416. doi: 10.1007/s00775-005-0650-8
- Jenney, F. E. Jr., Verhagen, M. F., Cui, X., and Adams, M. W. (1999). Anaerobic microbes: oxygen detoxification without superoxide dismutase. *Science* 286, 306–309. doi: 10.1126/science.286.5438.306
- Katoh, K., and Standley, D. M. (2014). MAFFT: iterative refinement and additional methods. *Methods Mol. Biol.* 1079, 131–146. doi: 10.1007/978-1-62703-646-7_8
- Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J., and McInerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* 6:1–17. doi: 10.1186/1471-2148-6-29
- Kunkel, A., Vaupel, M., Heim, S., Thauer, R. K., and Hedderich, R. (1997). Heterodisulfide reductase from methanol-grown cells of *Methanosarcina barkeri* is not a flavoenzyme. *Eur. J. Biochem.* 244, 226–234. doi: 10.1111/j.1432-1033.1997.00226.x
- Kurtz, D. M. Jr. (2006). Avoiding high-valent iron intermediates: superoxide reductase and rubrerythrin. *J. Inorg. Biochem.* 100, 679–693. doi: 10.1016/j.jinorgbio.2005.12.017
- Lakhal, R., Auria, R., Davidson, S., Ollivier, B., Durand, M. C., Dolla, A., et al. (2011). Oxygen uptake rates in the hyperthermophilic anaerobe *Thermotoga maritima* grown in a bioreactor under controlled oxygen exposure: clues to its defence strategy against oxidative stress. *Arch. Microbiol.* 193, 429–438. doi: 10.1007/s00203-011-0687-8
- Le Fourn, C., Brasseur, G., Brochier-Armanet, C., Pieulle, L., Brioukhanov, A., Ollivier, B., et al. (2011). An oxygen reduction chain in the hyperthermophilic anaerobe *Thermotoga maritima* highlights horizontal gene transfer between Thermococcales and Thermotogales. *Environ. Microbiol.* 13, 2132–2145. doi: 10.1111/j.1462-2920.2011.02439.x
- Li, S. J., Hua, Z. S., Huang, L. N., Li, J., Shi, S. H., Chen, L. X., et al. (2014). Microbial communities evolve faster in extreme environments. *Sci. Rep.* 4:6205. doi: 10.1038/srep06205
- Lumppio, H. L., Shenvi, N. V., Garg, R. P., Summers, A. O., and Kurtz, D. M. Jr. (1997). A rubrerythrin operon and nigerythrin gene in *Desulfovibrio vulgaris* (Hildenborough). *J. Bacteriol.* 179, 4607–4615.
- Lyons, T. W., Reinhard, C. T., and Planavsky, N. J. (2014). The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 506, 307–315. doi: 10.1038/nature13068
- Maaty, W. S., Wiedenheft, B., Tarlykov, P., Schaff, N., Heinemann, J., Robison-Cox, J., et al. (2009). Something old, something new, something borrowed; how the thermoacidophilic archaeon *Sulfolobus solfataricus* responds to oxidative stress. *PLoS ONE* 4:e6964. doi: 10.1371/journal.pone.0006964
- Mardanov, A. V., Gumerov, V. M., Beletsky, A. V., Prokofeva, M. I., Bonch-Osmolovskaya, E. A., Ravin, N. V., et al. (2011). Complete genome sequence of the thermoacidophilic crenarchaeon *Thermoproteus uzoniensis* 768-20. *J. Bacteriol.* 193, 3156–3157. doi: 10.1128/JB.00409-11
- Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40, D115–D122. doi: 10.1093/nar/ gkr1044
- Martin, W. F., and Sousa, F. L. (2016). Early microbial evolution: the age of anaerobes. *Cold Spring Harb. Perspect. Biol.* 8:a018127. doi: 10.1101/cshperspect.a018127
- Mavromatis, K., Sikorski, J., Pabst, E., Teshima, H., Lapidus, A., Lucas, S., et al. (2010). Complete genome sequence of *Vulcanisaeta distributa* type strain (IC-017). *Stand Genomic Sci* 3, 117–125. doi: 10.4056/signs.1113067
- Mishra, S., and Imlay, J. (2012). Why do bacteria use so many enzymes to scavenge hydrogen peroxide? *Arch. Biochem. Biophys.* 525, 145–160. doi: 10.1016/j.abb.2012.04.014
- Nasir, A., Kim, K. M., and Caetano-Anolles, G. (2014). Global patterns of protein domain gain and loss in superkingdoms. *PLoS Comput. Biol.* 10:e1003452. doi: 10.1371/journal.pcbi.1003452
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33, W116–W120. doi: 10.1093/nar/gki442
- Riebe, O., Fischer, R. J., Wampler, D. A., Kurtz, D. M. Jr., and Bahl, H. (2009). Pathway for H₂O₂ and O₂ detoxification in *Clostridium acetobutylicum*. *Microbiology* 155, 16–24. doi: 10.1099/mic.0.022756-0
- Sato, Y., Kameya, M., Fushinobu, S., Wakagi, T., Arai, H., Ishii, M., et al. (2012). A novel enzymatic system against oxidative stress in the thermophilic hydrogen-oxidizing bacterium *Hydrogenobacter thermophilus*. *PLoS ONE* 7:e34825. doi: 10.1371/journal.pone.0034825
- Wakagi, T. (2003). Sulerythrin, the smallest member of the rubrerythrin family, from a strictly aerobic and thermoacidophilic archaeon, *Sulfolobus tokodaii* strain 7. *FEMS Microbiol. Lett.* 222, 33–37. doi: 10.1016/S0378-1097(03)00233-7
- Wang, M., Jiang, Y.-Y., Kim, K. M., Qu, G., Ji, H.-F., Mittenthal, J. E., et al. (2011). A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. *Mol. Biol. Evol.* 28, 567–582. doi: 10.1093/molbev/msq232
- Weinberg, M. V., Jenney, F. E. Jr., Cui, X., and Adams, M. W. (2004). Rubrerythrin from the hyperthermophilic archaeon *Pyrococcus furiosus* is a rubredoxin-dependent, iron-containing peroxidase. *J. Bacteriol.* 186, 7888–7895. doi: 10.1128/JB.186.23.7888-7895.2004
- Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., et al. (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1:16116. doi: 10.1038/nmicrobiol.2016.116
- Zhao, W., Ye, Z., and Zhao, J. (2007). RbrA, a cyanobacterial rubrerythrin, functions as a FNR-dependent peroxidase in heterocysts in protection of nitrogenase from damage by hydrogen peroxide in *Anabaena* sp. *PCC 7120*. *Mol. Microbiol.* 66, 1219–1230.
- Zuber, P. (2009). Management of oxidative stress in *Bacillus*. *Annu. Rev. Microbiol.* 63, 575–597. doi: 10.1146/annurev.micro.091208.073241

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Cardenas, Quatrini and Holmes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Gene Turnover Contributes to the Evolutionary Adaptation of *Acidithiobacillus caldus*: Insights from Comparative Genomics

Xian Zhang^{1,2}, Xueduan Liu^{1,2}, Qiang He³, Weiling Dong¹, Xiaoxia Zhang⁴, Fenliang Fan⁵, Deliang Peng⁶, Wenkun Huang⁶ and Huaqun Yin^{1,2*}

¹ School of Minerals Processing and Bioengineering, Central South University, Changsha, China, ² Key Laboratory of Biometallurgy of Ministry of Education, Central South University, Changsha, China, ³ Department of Civil and Environmental Engineering, the University of Tennessee, Knoxville, TN, USA, ⁴ Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China, ⁵ Key Laboratory of Plant Nutrition and Fertilizer, Chinese Academy of Agricultural Sciences, Beijing, China, ⁶ State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:

Edgardo Donati,
National University of La Plata,
Argentina

Reviewed by:

Om V. Singh,
University of Pittsburgh, USA
Jorge Hernan Valdes,
Center for Genomics
and Bioinformatics/Universidad
Mayor, Chile

*Correspondence:

Huaqun Yin
yinhuaqun@gmail.com

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 27 June 2016

Accepted: 22 November 2016

Published: 06 December 2016

Citation:

Zhang X, Liu X, He Q, Dong W,
Zhang X, Fan F, Peng D, Huang W
and Yin H (2016) Gene Turnover
Contributes to the Evolutionary
Adaptation of *Acidithiobacillus caldus*:
Insights from Comparative Genomics.
Front. Microbiol. 7:1960.
doi: 10.3389/fmicb.2016.01960

Acidithiobacillus caldus is an extremely acidophilic sulfur-oxidizer with specialized characteristics, such as tolerance to low pH and heavy metal resistance. To gain novel insights into its genetic complexity, we chosen six *A. caldus* strains for comparative survey. All strains analyzed in this study differ in geographic origins as well as in ecological preferences. Based on phylogenomic analysis, we clustered the six *A. caldus* strains isolated from various ecological niches into two groups: group 1 strains with smaller genomes and group 2 strains with larger genomes. We found no obvious intraspecific divergence with respect to predicted genes that are related to central metabolism and stress management strategies between these two groups. Although numerous highly homogeneous genes were observed, high genetic diversity was also detected. Preliminary inspection provided a first glimpse of the potential correlation between intraspecific diversity at the genome level and environmental variation, especially geochemical conditions. Evolutionary genetic analyses further showed evidence that the difference in environmental conditions might be a crucial factor to drive the divergent evolution of *A. caldus* species. We identified a diverse pool of mobile genetic elements including insertion sequences and genomic islands, which suggests a high frequency of genetic exchange in these harsh habitats. Comprehensive analysis revealed that gene gains and losses were both dominant evolutionary forces that directed the genomic diversification of *A. caldus* species. For instance, horizontal gene transfer and gene duplication events in group 2 strains might contribute to an increase in microbial DNA content and novel functions. Moreover, genomes undergo extensive changes in group 1 strains such as removal of potential non-functional DNA, which results in the formation of compact and streamlined genomes. Taken together, the findings presented herein show highly frequent gene turnover of *A. caldus* species that inhabit extremely acidic environments, and shed new light on the contribution of gene turnover to the evolutionary adaptation of acidophiles.

Keywords: *Acidithiobacillus caldus*, comparative genomics, intraspecific diversity, gene turnover, evolutionary adaptation

INTRODUCTION

Acidithiobacillus caldus (formerly *Thiobacillus caldus*), a moderately thermophilic, obligately chemolithoautotrophic, and extremely acidophilic sulfur-oxidizing bacterium (Hallberg and Lindström, 1994, 1996), is of interest for its potential role in industrial bioleaching (Rawlings, 1998; Dopson and Lindström, 1999). *A. caldus* exploits elemental sulfur and a wide range of reduced inorganic sulfur compounds at moderately high temperatures to support autotrophic growth (Mangold et al., 2011; Chen et al., 2012). It is the primary member of a consortium of sulfur oxidizers in different toxic-laden acidic environments, which are termed “extreme environments,” including coal pile and spoil, gold-bearing reactor operation, as well as low-grade copper bioleaching heap (Valdes et al., 2009; You et al., 2011; Zhang et al., 2016c). Considering that *A. caldus* inhabits harsh environments for prolonged periods and accommodates both sudden stress changes and long-term stress conditions in various habitats, gene flow and genetic drift might frequently occur. As such, the flexible gene repertoire generated by gene exchange has imparted *A. caldus* with extensive genetic material for diversification of function and phenotype. Therefore, research focusing on the correlation between genomic changes and evolutionary adaptation is of great interest.

The accumulation of genomic changes underlying evolutionary adaptation has often been viewed as a complex process, and has been subject to many influences and complications (Barrick et al., 2009). As stated by Carretero-Paulet et al. (2015), homologous genes derived from newly formed subgenomes might undergo asymmetric fractionation via mutational events, which include nucleotide substitutions, gene gains and losses, and changes in genomic structure and organization (Librado et al., 2014). In terms of neutral mutation theory, mutations underlying gene and genome evolution, though not necessarily beneficial, should accumulate at a constant rate by drift (Kimura, 1984). Another view is that the substitution rates for beneficial and deleterious mutations depend on environmental selection, as well as population size and structure (Gillespie, 1991; Ohta, 1992). For many years, the crucial role of gene and genome duplications (namely, neofunctionalization and subfunctionalization) in governing organismal evolution has been acknowledged (Ohno, 1970; Force et al., 1999; Innan and Kondrashov, 2010; Kulmuni et al., 2013). Only in recent decades has great attention been paid to the molecular mechanisms of gene loss (deletion or pseudogenization) as a pervasive source of genetic change, which is believed to be another key evolutionary event that causes adaptive phenotypic diversity (Albalat and Cañestro, 2016). In recent years, a number of analytical methods for population genomics and molecular evolution have provided substantial evidence to determine the relative contribution of diverse evolutionary forces, which shape genome organization, architecture, and diversity in response to environmental perturbations (Librado et al., 2014). In eukaryotes, gene family evolution has often been modeled after a phylogenetic birth-and-death (BD) process (Nei and Rooney, 2005). This BD model, though suitable to account for single-gene duplications, might

not be appropriate for calculating gene turnover rates given that horizontal gene transfer (HGT) events occur in certain organisms (Librado et al., 2014). However, an alternative gain-and-death (GD) stochastic model in a maximum-likelihood statistical framework was applied to circumvent this limitation (Librado et al., 2012). Unlike the birth process, gains in the developed GD model can accommodate all kinds of gene acquisitions, irrespective of their original source, even including HGT (Librado et al., 2014). In this study, we are interested in whether the aforementioned theoretical and analytical approaches can be applied to explain the relationship between genetic change and adaptive evolution of *A. caldus* inhabiting extraordinarily extreme environments.

Members of *A. caldus* species are ubiquitous throughout many sulfur-rich acidic environments worldwide (Table 1), indicating their adaptation to various niches with high concentrations of toxic substrates, such as coal spoil, gold-bearing bioleaching reactor, and copper mine tailing. In recent years, revolutionary technologies and tools have allowed for the rapid characterization of microbial genome sequences (MacLean et al., 2009; Metzker, 2010). Accurate analyses of gene family evolution have been made possible owing to the increasing availability of closely related genomes (Hahn et al., 2007; Sánchez-Gracia et al., 2009; Vieira and Rozas, 2011). Furthermore, acquisition of numerous additional genomes has fuelled a new field termed comparative genomics (Jacobsen et al., 2011), which is useful for investigating microbial genome evolution and even mechanisms for speciation (González et al., 2014; Justice et al., 2014; Ullrich et al., 2016). Comparative surveys based on the available genomes of the two *A. caldus* strains ATCC 51756 and SM-1 have revealed that both strains harbor a relatively high proportion of unique gene complements (Acuña et al., 2013). These gene complements represent a diverse pool of mobile genetic elements, including insertion sequences (ISs), genomic islands (GIs), and integrative conjugative and mobilizable elements. Yet, limited information is available on the contribution of diverse evolutionary forces to the genomic diversification of *A. caldus*. Given this knowledge gap, we have isolated and sequenced four new *A. caldus* strains from different geographic origins (Table 1).

In this study, we estimated the phylogenetic relationships of *A. caldus* strains based on their genomic sequences (four newly sequenced genomes and two existing genomes from a public database), and performed an exhaustive study of the GD dynamics, with special focus on genetic exchange underlying evolutionary adaptation. These findings, to some extent, highlight the role of gene turnover in the evolutionary diversification of *A. caldus* and adaptation to specific lifestyles and environmental niches.

MATERIALS AND METHODS

DNA Sequencing and Bioinformatics Analysis

Genome sequences for six strains were retrieved in this study, including *A. caldus* ATCC 51756, SM-1, DX, S1, ZBY, and ZJ.

TABLE 1 | General features of sequenced chromosomes in *A. caldus* strains.

| Organism | <i>A. caldus</i> SM-1 | <i>A. caldus</i> ATCC 51756 | <i>A. caldus</i> S1 | <i>A. caldus</i> DX | <i>A. caldus</i> ZBY | <i>A. caldus</i> ZJ |
|------------------------------------|---|--------------------------------------|------------------------------------|--------------------------------------|---|-------------------------------------|
| Geographic origin | Gold-bearing bioleaching reactor, China | Coal spoil at the Kingsbury Mine, UK | Coal heap drainage, Jiangxi, China | Copper mine tailings, Jiangxi, China | Copper mine tailings, Chambishi, Zambia | Copper mine tailings, Fujian, China |
| Status | Complete | Complete | Draft | Draft | Draft | Draft |
| Accession number | NC_015850 | NZ_CP005986 | LZYH000000000 | LZYE000000000 | LZYF000000000 | LZYG000000000 |
| Total bases (bp) | 2,932,225 | 2,777,717 | 2,792,792 | 3,122,206 | 3,160,074 | 3,143,077 |
| Completeness* | | | 89.75 | 98.76 | 98.76 | 98.14 |
| Coverage | 38× | 120× | 92× | 95× | 89× | 76× |
| GC content (%) | 61.32 | 61.72 | 60.90 | 61.01 | 60.98 | 61.00 |
| Number of contigs | 1 | 1 | 1,208 | 390 | 414 | 386 |
| Maximum sequence length | 2,932,225 | 2,777,717 | 26,396 | 102,019 | 77,380 | 74,790 |
| Minimum sequence length | 2,932,225 | 2,777,717 | 200 | 207 | 201 | 206 |
| N50 (bp) | 2,932,225 | 2,777,717 | 4,735 | 22,157 | 18,983 | 18,308 |
| N90 (bp) | 2,932,225 | 2,777,717 | 617 | 4,321 | 4,123 | 2,291 |
| Number of rRNA operon (5s-16s-23s) | 2 | 2 | 1 | 1 | 1 | 1 |
| Number of tRNA | 47 | 49 | 32 | 46 | 47 | 46 |
| Number of coding sequences | 2,833 | 2,699 | 2,874 | 2,942 | 3,017 | 2,984 |
| Proteins with predicted function | 2,042 | 2,008 | 1,860 | 2,109 | 2,161 | 2,144 |
| Reference | You et al. (2011) | Valdes et al. (2009) | This study | This study | This study | This study |

*Genome completeness was estimated using the CheckM. Strains SM-1 and ATCC 51756 with complete genome were excluded.

Of these bacteria, the type strain ATCC 51756 was isolated from a coal spoil in Kingsbury, UK (Marsh and Norris, 1983), strain SM-1 was from an industrial reactor used in bioleaching operation (Liu et al., 2007), and the other strains (DX, S1, ZBY, and ZJ) were obtained from the China Center for Type Culture Collection. More details for geographic origins of these four new strains were shown in **Table 1**. Genome sequences of strains ATCC 51756 and SM-1, including chromosomal and plasmid sequences, were downloaded from the GenBank database. For strains DX, S1, ZBY, and ZJ, chromosomal DNA was sequenced by an Illumina MiSeq sequencer (Illumina, Inc., USA), using the paired-end sequencing approach with an average DNA insert size of 300 bp and typical read-length of 150 bp. Subsequently, bioinformatics analysis of raw sequences was performed as described previously (Yin et al., 2014), primarily including quality control, genome assembly, computational prediction of coding sequences (CDS) and other genome features such as rRNA and tRNA, as well as functional assignments against public databases (NCBI-nr and COG). Genome completeness of each strain was also estimated using the program CheckM (Parks et al., 2015). Additionally, circular maps showing chromosome architecture were drawn using the Circos software (Krzywinski et al., 2009).

Intergenomic distance scores were calculated using the web service Genome-to-Genome Distance Calculator (GGDC) 2.1 (Meier-Kolthoff et al., 2013). The distance $d(X, Y)$

between genome X and Y was calculated according to the formula:

$$d(X, Y) = 1 - \frac{2 \cdot I_{XY}}{H_{XY} + H_{YX}}$$

(1)

in which, I_{XY} denotes the sum of identical base pairs over all high-scoring segment pairs (HSPs, which are intergenomic matches), while H_{XY} and/or H_{YX} denote the total length of all HSPs. Heatmap was shown using the software HemI (Deng et al., 2014).

16S Ribosomal RNA (rRNA) Gene-Based and Whole Genome-Based Phylogenetic Tree

Phylogenetic relationship based on 16S rRNA sequences of *Acidithiobacillus* strains was analyzed using MEGA v5.05 with neighbor-joining method. The robustness of clustering was evaluated by 1,000 bootstrap replicates. Additionally, the phylogenetic relationships between complete and draft genomes from *A. caldus* strains were estimated. We employed an online platform CVTree3 (Zuo and Hao, 2015) to construct the whole-genome based phylogenetic tree using a composition vector approach. This whole-genome-based and alignment-free prokaryotic phylogeny was validated by directly comparing our result with the taxonomy of these strains, as opposed to performing statistical resampling tests such as bootstrap or

jackknife. The genome sequence of *Acidithiobacillus ferrooxidans* ATCC 23270 was chosen as an outgroup. Subsequently, visualization of phylogenetic tree was executed using the MEGA v5.05 (Tamura et al., 2011).

Pan-Genome Analysis

Species diversity could be identified by analyzing gene repertoire across all strains of a species, i.e., the pan-genome (Tettelin et al., 2008). PanOCT v3.18 (Fouts et al., 2012) with a BLASTP all-against-all comparison of entire proteins (E -value $\leq 1e^{-5}$; sequence identity $\geq 50\%$) was used to identify shared and unique gene content. Subsequently, annotation of core genome and strain-specific genes was implemented using BLAST against the extended COG database (Franceschini et al., 2013).

Gene Family Evolution

Groups of orthologous sequences (orthogroups, herein referred to as gene families) in all six *A. caldus* strains were classified by clustering with OrthoFinder v0.4 (Emms and Kelly, 2015), using a Markov cluster algorithm. Transposable elements were excluded, given that these gene sequences might interfere with our analyses owing to lineage-specific expansions (Carretero-Paulet et al., 2015).

To analyze the evolutionary rates of gene families, we applied the developed computational program BadiRate v1.35 using a GD stochastic model (Librado et al., 2012). The gain (γ) and death (δ) rates of gene families were estimated using a branch-specific rates (GD-BR-ML) model assuming that each phylogenetic branch had its own specific turnover rate.

Mobile Gene Elements, Insertion Sequence Elements, Transposable Elements, and Genomic Islands

IS family annotation and transposase inspection was done by BLAST comparison (E -value $\leq 1e^{-5}$) against the ISFinder database with manual detection of the surrounding significant search hits (Siguier et al., 2006). The program SeqWord Genomic Island Sniffer (Bezuidt et al., 2009) was implemented to identify the putative horizontally transferred elements distributed in the chromosome of *A. caldus* ATCC 51756. Then, the prediction of genes in the putative horizontally transferred elements was performed using the MetaGeneAnnotator (Noguchi et al., 2008). For the other chromosomes, the computational tool IslandViewer 3 (Dhillon et al., 2015), which integrates three different prediction methods including IslandPick (Langille et al., 2008), IslandPath-DIMOB (Hsiao et al., 2003), and SIGI-HMM (Waack et al., 2006), was used to predict GIs. The GC content of GI sequences was calculated using the NGS QC Toolkit (Patel and Jain, 2012). Due to the high number of contigs, *A. caldus* S1 was excluded from the GI prediction.

Availability of Supporting Data

The data sets supporting our results in this study are available in the GenBank repository. These Whole Genome Shotgun projects of four newly sequenced *A. caldus* strains have been

deposited at the DDBJ/ENA/GenBank under the accession numbers LZYE000000000 (DX), LZYP000000000 (ZBY), LZYH000000000 (S1), and LZYG000000000 (ZJ). Additionally, the versions described in this paper are version LZYE01000000, LZYP01000000, LZYH01000000, and LZYG01000000, respectively.

RESULTS AND DISCUSSION

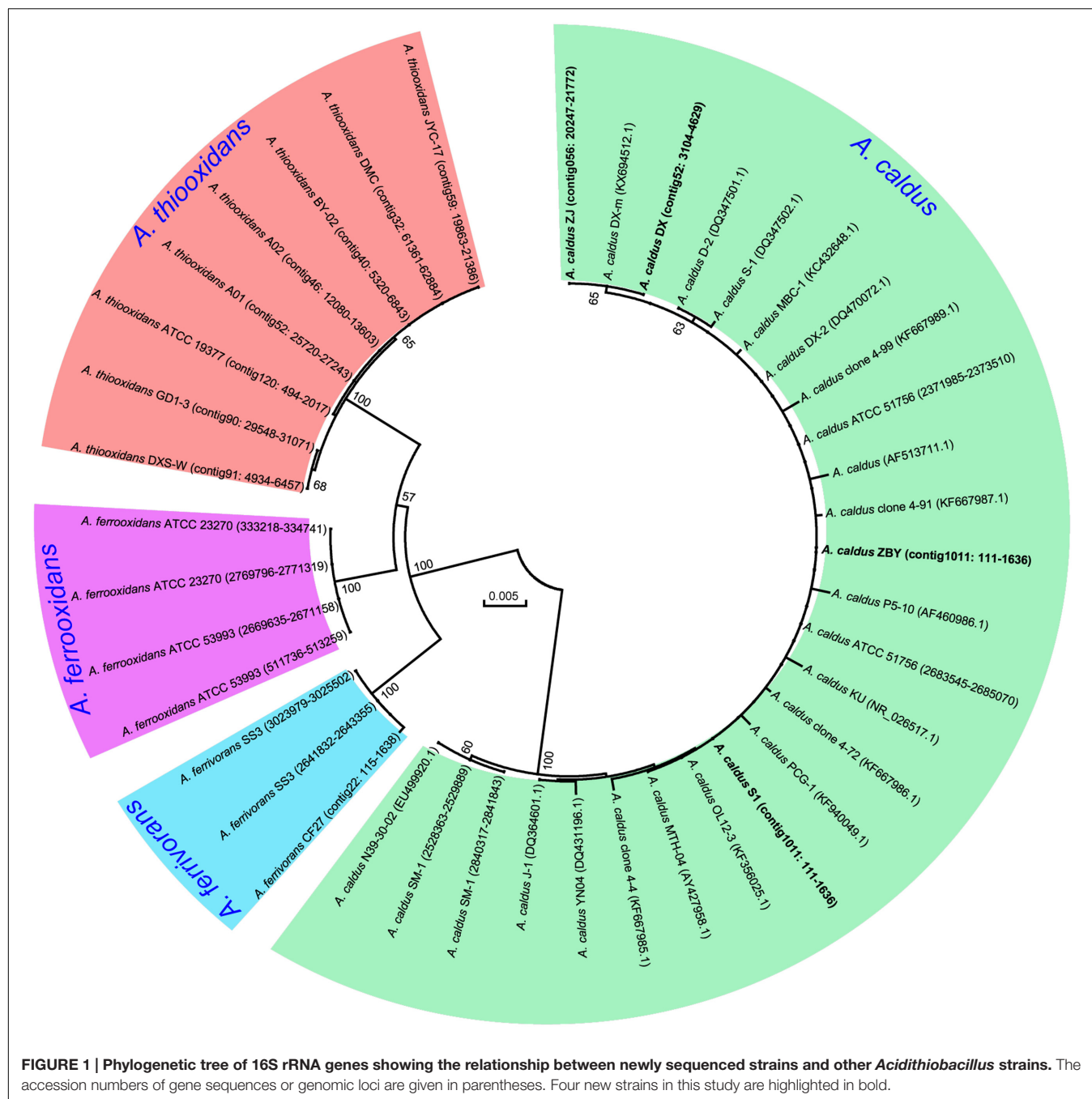
Overview of the *A. caldus* Chromosomes

The circular chromosomes of *A. caldus* strains varied from 2.78 to 3.16 Mb (Table 1). *A. caldus* strains DX, ZBY, and ZJ, which were isolated from a copper mine, possess larger chromosomes than the other strains inhabiting the divergent habitats. Genome-size variations in bacteria correspond to variations in gene number as bacterial genomes are tightly packed, and most sequences are functional protein-coding regions (Mira et al., 2001). Accordingly, strains with larger genome were predicted to harbor more CDSs compared to other strains in this study. Additionally, the evaluation of quality and completeness of genome assemblies supported the reliability of pan-genome analysis, although strain S1 had relatively low genome completeness in comparison with its closely related counterparts (Table 1).

In all *A. caldus* strains, the mean percentage GC content of these chromosomal DNAs (60.90–61.72% for all six strains) was much higher than that observed for other recognized *Acidithiobacillus* spp., e.g., *A. ferrooxidans*, *A. thiooxidans*, and *A. ferrivorans*. It might be reasonable considering that *A. caldus* species was known as the only known mesothermophile within the Acidithiobacillales (Acuña et al., 2013), and GC content of prokaryotic genomes was positively correlated with optimal growth temperature (Musto et al., 2004, 2006).

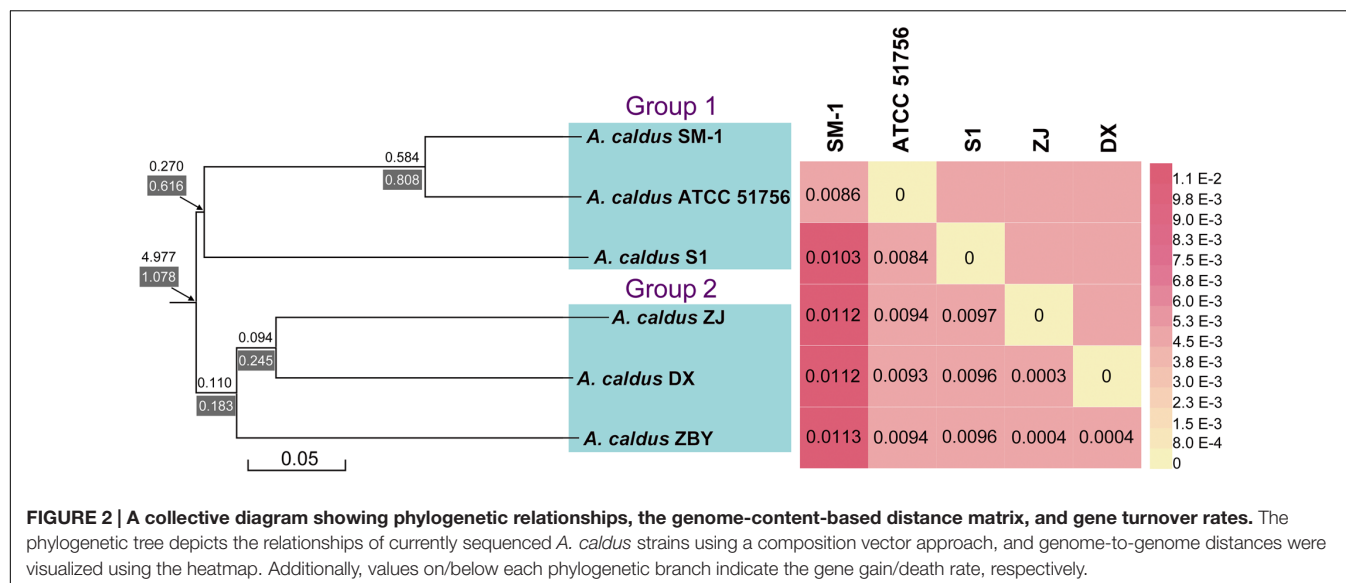
Evolutionary Relationship of *A. caldus* Strains

A phylogenetic tree based on 16S rRNA genes of *Acidithiobacillus* strains preliminarily demonstrated that these four newly sequenced strains in this study were taxonomically affiliated with *A. caldus* (Figure 1). To further identify the evolutionary relationships of *A. caldus* strains, a whole-genome-based and alignment-free phylogenetic tree was constructed (Figure 2). Additionally, GGDC analyses were employed to support the phylogenetic relationship. This phylogenomic tree showed that three strains isolated from the copper mine (namely, ZJ, DX, and ZBY) were clustered together (group 2 in Figure 2). Similarly, an earlier study reported that taxonomic clustering of six strains belonging to the genus *Novosphingobium* was generally influenced by their respective source of isolation (Gan et al., 2013). Further inspection revealed that the geographic distribution of strain ZBY was distinctively differed from those of the other two strains (ZJ and DX), and the genome-content-based distance matrix implied a slight evolutionary divergence (Figure 2). The correlation between intraspecific divergence and geographic distribution was also observed within the closely related *A. thiooxidans* species by comparative genomic



analysis (Zhang et al., 2016a). In the group 1 (**Figure 2**), interestingly, *A. caldus* SM-1 was obtained from a bioleaching reactor used for low grade gold-bearing minerals (Acuña et al., 2013), and the strain ATCC 51756 was isolated from a coal spoil; moreover, phylogenetic analysis revealed that these two strains were more closely related to each other than to the other four strains examined in this study. We therefore suspect that strain SM-1 might originally be isolated from an acidic setting similar to the habitat for ATCC 51756.

Ji et al. (2014) showed that the differences in adaptive evolution were attributable to different ecoiniche by genetically analyzing the marine and freshwater magnetospirilla. Accordingly, we propose that environmental variation, particularly geochemical conditions, might be a determinant of genomic diversity of *A. caldus* strains. From an alternative perspective, it appears that geographic distribution has less of an influence on hereditary variation in comparison with ecoiniche difference. The findings were consistent with an earlier study showing that environmental heterogeneity has relatively more



influence on microbial biogeography compared to geographic distance (Lin et al., 2013).

Gene Contents in *A. caldus* Strains

Gene prediction showed that the chromosomes of *A. caldus* strains contained 2,699 (ATCC 51756), 2,833 (SM-1), 2,874 (S1), 2,942 (DX), 3,017 (ZBY), and 2,984 (ZJ) predicted CDS. Functional analysis based on COG categories (Supplementary Table S1) revealed that the four most abundant functional categories within all *A. caldus* strains were “function unknown [S],” “replication, recombination, and repair [L],” “cell wall/membrane/envelope biogenesis [M],” and “energy production and conversion [C].” As reported by Silver and Phung (1996), high concentrations of toxic substrates such as heavy metals might cause a high rate of DNA damage. Thus, it was expected that CDS involved in COG category [L] would be abundant in *A. caldus* strains. Additionally, these data can also explain why this finding was distinct from previous studies analyzing the COG classification of other organisms such as marine magnetospirillum *Magnetospira* sp. QH-2 (Ji et al., 2014), given that the concentrations of potential toxic substrates in the extreme environment were much higher than those in the marine environment.

A previous study based on four genomes of “*Ferroplasma*” strains highlighted the most distinct differences in interspecific metabolisms (Ullrich et al., 2016). However, in our study the assignment of CDS to the COG classification revealed that no significant differences in the number of assigned CDS were observed between the six genomes (Supplementary Table S1), probably suggesting few group-specific metabolic traits.

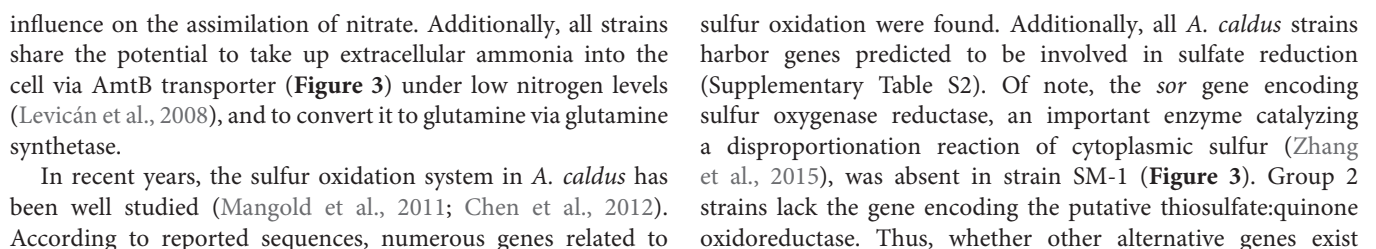
Comparison of Inferred Metabolic Traits and Niche Adaptation

Comparison of the Central Metabolism

In light of COG assignment aforementioned, we further observed CDS related to the predicted metabolic profiles. Compared with

other metabolic models reported in the literature, including carbon metabolism (You et al., 2011; Zhang et al., 2016b), nitrogen uptake (Levicán et al., 2008; Justice et al., 2014), and sulfur oxidation (Mangold et al., 2011; Chen et al., 2012; Yin et al., 2014), all strains in our study were predicted to contain numerous genes involved in central metabolism (Supplementary Table S2). The metabolic potentials of all strains were reconstructed and compared to each other for the identification of shared metabolic features as well as group- or strain-specific traits (Figure 3). Comprehensive analysis of these metabolism-related genes focuses on the main differences between the six *A. caldus* strains. As depicted in Figure 3, however, the evidence showed low intraspecific genetic diversity in the predicted metabolic profiles between *A. caldus* strains. A suite of genes involved in carbon assimilation were found in all strains. *A. caldus* fixes carbon dioxide via the classical Calvin–Benson–Bassham (CBB) cycle, and harbors a gene cluster predicted to encode carbon dioxide-concentrating protein (CcmK) with various copies, carboxysome shell protein (CsoS), carboxysomal shell carbonic anhydrase (CsoSCA), and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO; Supplementary Table S2). Moreover, *A. caldus* operates a complete Embden–Meyerhof pathway (EMP) or glycolysis, pentose phosphate pathway (PPP), and incomplete tricarboxylic acid (TCA) cycle, which lacks the 2-oxoglutarate dehydrogenase complex (Valdés et al., 2008b).

With respect to nitrogen uptake, although *A. caldus* lacks nitrogenase directing the fixation of molecular nitrogen (Valdés et al., 2008a), assimilation of nitrate, nitrite, and ammonia plays a critical role in meeting nitrogen requirements. *A. caldus* utilizes nitrate or nitrite via nitrate transporter (NRT) and nitrate/nitrite transporter (Nrt). However, NRT was not present in strain ATCC 51756. Genes associated with dissimilatory nitrate reduction were identified, while a gene involved in assimilatory nitrate reduction (*nasA*) was absent in strain ATCC 51756. Though absent, it appears that the non-existence of those genes had little



in these strains needs to be studied further. Similar to the well-studied model for electron transfer of *A. ferrooxidans* (Valdés et al., 2008a), *A. caldus* potentially employs the electron transfer pathway from sulfur oxidation to (1) various types of terminal oxidases to generate a proton gradient or (2) to NADH complex to produce reducing power (Chen et al., 2012).

To some extent, investigation of genes involved in central metabolism supported the results of COG assignment that there were no obvious intraspecific differences. In other words, comparison of intraspecific genomes showed that only slight differences were observed in metabolic profiles, at least in central metabolism.

Response to Environmental Stress

Microbial response to environmental stresses is always a critical issue in ecological fields (Yin et al., 2015). A long-term experiment with *Escherichia coli* revealed complex coupling between organismal adaptation and genome evolution, which occurred even in a constant environment (Barrick et al., 2009). In the context of the six *A. caldus* strains, bacterial adhesion, motility, heavy metal resistance, and organic solvent tolerance were taken into account (Figure 3). All strains share a core set of genes potentially related to environmental adaptation (Supplementary Table S2). The presence of genes encoding extracellular polymeric substances precursors and type IV pili in *A. caldus* suggests a cell adhesion on mineral surface. This trait provides a reaction space between cell and mineral surface, thereby increasing the dissolution of metal sulfides (Watling, 2006; González et al., 2013). Genes assigned to COG category [N] (cell mobility) and [T] (signal transduction) were also observed, but there were few differences between these two groups (Supplementary Table S1). A full suite of genes associated with flagellar assembly were found in all strains, suggesting that *A. caldus* strains had the capacity to swim across environmental gradients and to colonize new sites.

Extremely acidic environments, especially bioleaching systems, are regarded as having extremely high concentrations of soluble and potentially toxic substrates such as heavy metals, including arsenic, mercury, copper, and cadmium (Valdés et al., 2008a) and organic extractants, such as Lix984n (Zhou et al., 2012). A series of gene clusters potentially encoding functional enzymes were identified, suggesting that *A. caldus* has the ability to cope with high concentrations of heavy metal ions. As for organic solvent tolerance, a six-gene cluster, encoding ABC transporter ATP-binding protein, hypothetical protein, toluene tolerance protein, mce-related protein, toluene tolerance protein Ttg2B, and toluene ABC transporter ATP-binding protein, was found in all strains. Additionally, an *acrAB-tolC* operon potentially encoding AcrB (transporter AcrB/AcrD/AcrF family protein), AcrA (RND family efflux transporter MFP subunit), and TolC (outer membrane efflux protein) in each genome indicated that *A. caldus* can utilize the pumps associated with resistance-nodulation-cell division protein to transfer these organic substrates.

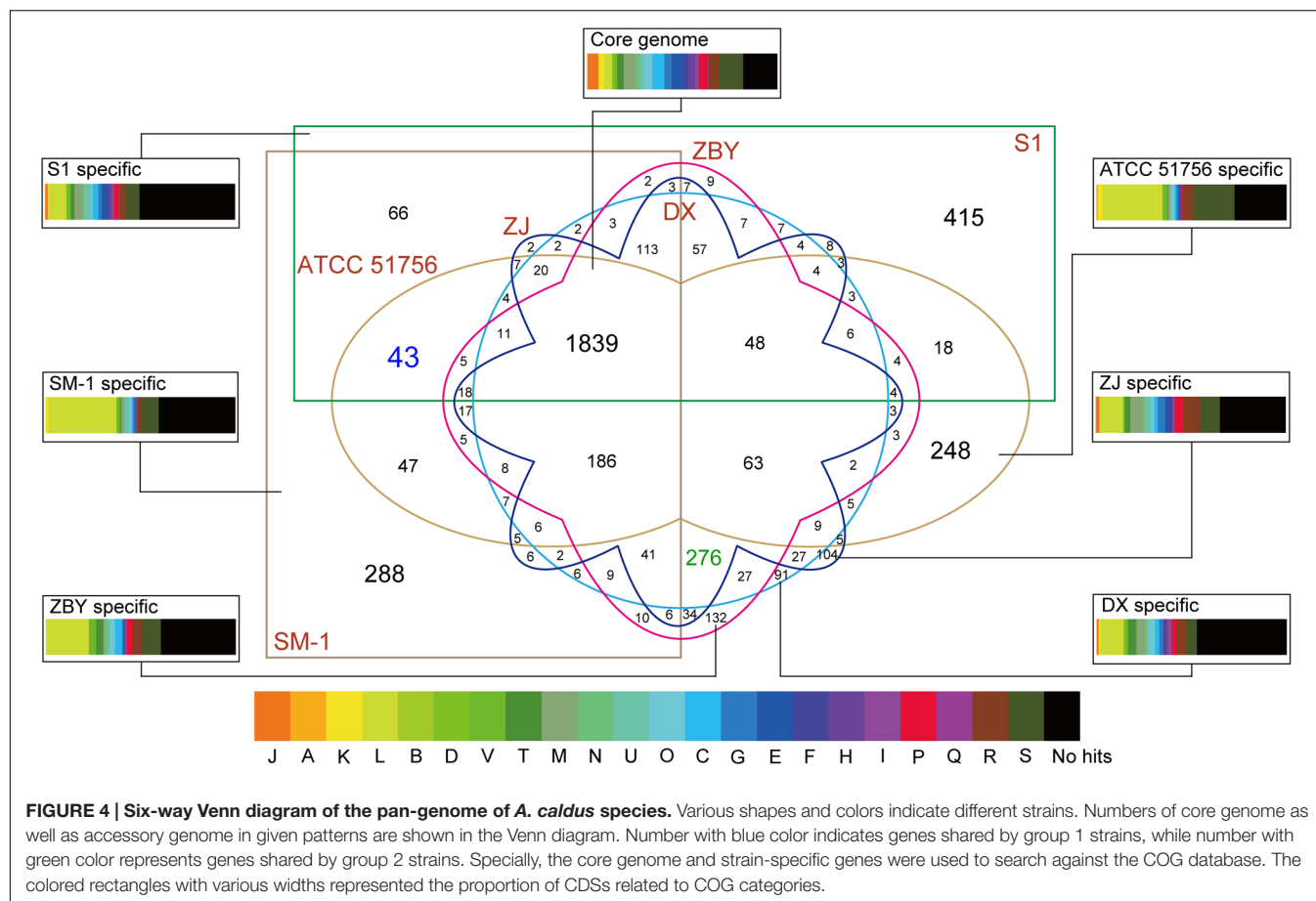
Pan-Genome Analysis

As shown above, numerous homologous genes associated with metabolic pathways as well as environmental adaptation were observed. To gain a deeper understanding of group- and strain-specific features, pan-genome analysis of *A. caldus* species was performed. A total of 4,424 CDS acquired from the four newly sequenced chromosomes plus two available chromosomes in the public database were clustered using the PanOCT. Pairwise BLAST comparisons indicated that 1,839 orthologs (41.57%) with a high percentage across all six strains were identified as the *A. caldus* core genome (Figure 4). The remaining variable 1,307 clusters were classified as the *A. caldus* accessory genome. Furthermore, strain-specific clusters were observed among the six *A. caldus* strains.

Functional assignment based on the core genome was employed to investigate the proportion of proteins in each COG category. As depicted in Figure 4, the core genome in *A. caldus* strains was commonly enriched in the COG category [M] (cell wall/membrane/envelope biogenesis; 6.36%). Additionally, our results showed that CDSs involving COG categories [C] (energy production and conversion; 6.30%) and [E] (amino acid transport and metabolism; 6.25%) were abundant. The large proportion of these genes indicated that energy utilization and uptake of nutrients in these strains might be more efficient to better adapt to the challenging environment. In other words, these findings were in line with an earlier report detailing that core genes provided functions that were essential to the basic lifestyle of the species (Medini et al., 2005).

Persistent genes encoding essential functions are stably maintained in genomes under constant selection (Nuñez et al., 2013), while dispensable or accessory genes are frequently gained or lost (Medini et al., 2005). Therefore, the accessory genome contributes to intraspecific diversity (Tettelin et al., 2008). Here, we identified many transposases by alignment of accessory genes against the NCBI-nr database (Supplementary Table S3), suggesting roles in shaping the evolution of protein families. Similarly, previously studies based on available genomes revealed that plentiful accessory genes were probably acquired by HGT (Tian et al., 2012; Sugawara et al., 2013). Additionally, it is particularly noteworthy that strain-specific genes were found to be enriched in the COG category [L] (replication, recombination and repair; Figure 4), thus supporting the view that the accessory genome confers selective advantages such as niche adaptation.

In particular, a total of 43 and 276 group-specific genes shared by group 1 and group 2 strains, respectively, were detected (Figure 4). Functional profiling based on COGs revealed that most of these predicted CDS were assigned to no COG category, probably indicating the existence of many group-specific CDS with unidentified function (Supplementary Table S4). Further inspection underscored that the abundant genes involved in certain COG categories, including [L] (replication, recombination, and repair), [M] (cell wall/membrane/envelope biogenesis), and [P] (inorganic ion transport and metabolism), might be necessary for the group 2 strains. A reasonable explanation is that copper bioleaching heap, the habitat for group 2 strains, has high concentrations of toxic metals (Zhang et al., 2016c). Microbes in such an extreme environment might



harbor potential strategies to cope with the chemical constraints of their natural functions. Additionally, COG categories [S] (general function prediction only) and [R] (function unknown) were relatively abundant in all groups, further highlighting the role of these unknown functional CDSs in genomic differentiation.

Mobile and Transposable Elements

Prediction and classification of transposable elements using ISFinder indicated that a large number of IS elements, which accounted for various proportions of the total CDS in each chromosome (ranging from 1.8 to 5.8%), were randomly distributed over the chromosomes of the *A. caldus* strains (Supplementary Table S5). Although the types of IS families were similar to each other, their distribution and relative abundance varied with each strain. Among them, some of these IS elements were identified to cluster in flexible chromosomal regions that did not satisfy the criteria of other putative mobile elements such as GIs; these findings were consistent with those from an earlier study (Acuña et al., 2013). As stated by Bentley and Parkhill (2004), the progressive loss of gene order in a prokaryotic genome might be attributed to several events including gene deletion, IS and repeat expansion, as well as recombination or rearrangement. Given this, *A. caldus* SM-1 as well as ATCC 51756 might have

higher genome plasticities compared with other closely related strains, mainly because of the acquisition of IS elements during evolution.

Aside from IS elements, the putative GI elements in all *A. caldus* strains were also identified. Results showed that several GIs ranging from 4 to 58 kb were widespread in the chromosomes of *A. caldus* strains (Supplementary Table S6). Additionally, most CDS in the GIs were annotated as hypothetical proteins. Further analyses showed the presence of integrases or mobile genetic elements such as transposase, thereby indicating that various putative GIs might be acquired via HGT. In light of the view that underscores the contribution of horizontal (lateral) gene transfer (HGT) in the expansion of gene repertoires of prokaryotes (Ochman et al., 2000; Gogarten et al., 2002; Treangen and Rocha, 2011), we inferred that the frequency of HGT was high in group 2 strains with larger genomes, conferring a predominant role in shaping their evolution and allowing the acquisition of novel adaptive functions. We emphasized the role of GIs in adaptation to specific lifestyles and environmental niches, considering that many GIs were highly relevant for niche-specific adaptation (Wu et al., 2011). Furthermore, numerous genes in *A. caldus* species might be obtained by genetic exchange as suggested by the presence of a large load of mobile genetic elements including IS elements, transposases, and GIs. Consequently,

TABLE 2 | Summary of genes assigned or unassigned to orthogroups in six *A. caldus* strains.

| Number | SM-1 | ATCC 51756 | S1 | ZJ | DX | ZBY |
|------------------------------------|-------|------------|-------|-------|-------|-------|
| Orthogroups | 2,514 | 2,422 | 2,395 | 2,818 | 2,800 | 2,828 |
| Genes in orthogroups | 2,709 | 2,538 | 2,477 | 2,918 | 2,886 | 2,942 |
| Genes unassigned to any orthogroup | 124 | 161 | 397 | 66 | 56 | 75 |

changes in genome structure and gene copy number might provide *A. caldus* strains with a survival advantage for rapid adaptation and survival in highly acidic and metal-laden environments.

Probabilistic Analysis of Gene Family Turnover

Gene families in *A. caldus* strains were classified as orthogroups using OrthoFinder (Table 2). Our classification identified up to 3,109 orthogroups (containing two or more genes in all selected strains), which included 16,470 sequences. There were fewer genes in group 1 strains with small genomes clustered into multigene orthogroups than in group 2 strains. However, these smaller genomes contained more unassigned genes than any orthogroups compared to the others. These results appear to be explained in part by intense fractionation pressure (Carretero-Paulet et al., 2015). In other words, multigene families in smaller genomes might be under continuous deletion pressure and, as a result, these genomes tend to be smaller in comparison with their counterparts in larger genomes.

BadiRate analysis, using a full likelihood method, was applied to examine the evolutionary dynamics of gene families across the *A. caldus* species, and to characterize the expansion and/or contraction of genomes. The statistical framework not only estimates GD rates in a decoupled manner, using two independent parameters (γ and δ), but also explicitly takes into account certain key features in prokaryotic evolution, such as HGT. A stochastic GD-BR-ML model statistically evaluating the turnover rates demonstrated that a large number of orthologous genes frequently undergoing high gain and/or death events have evolved from ancestral genes (Figure 2). Particularly, gene families in *A. caldus* species rapidly expand through gene gain (duplication) and slowly contract through gene death (deletion or pseudogenization), indicating that the extensive recruitment of genes involved in long-term evolution confers an ecological advantage for survival and proliferation under extremely acidic conditions. Moreover, the phylogenetic branch with the higher death rate ($\delta = 0.616$ and/or 0.808) indicated that group 1 strains with smaller genomes might be derived from free-living ancestors by the genome-reductive evolutionary process (Figure 2). Given that genome reduction coincided with the increase in frequency of mobile elements and repeated sequences (Moran, 2003), multiple IS elements identified in strain SM-1 and ATCC 51756 (Supplementary Table S5) might play a key role in mediating intrachromosomal recombination, thereby leading to rearrangements and gene loss. However, the dispensable genes in the above-mentioned microorganisms might suffer

extensive loss and non-functionalization. The compact genomes in the given organisms can perform essential functions for cellular survival and replication, as the loss of dispensable genes has little effect on bacterial fitness, at least under certain environmental conditions (Albalat and Cañestro, 2016). Despite their smaller or near-minimal size, all reduced genomes still retain the essential gene set, and are thereby able to support cellular life both in stable and changing circumstances (Moya et al., 2009). Therefore, small genomes in group 1 strains would be more tightly packed by selective reduction, and are thus more streamlined than their larger genome counterparts.

Gene turnover in group 2 strains was also estimated. As illustrated in Figure 2, phylogenetic branches showed a lower gene turnover rate in group 2 strains compared to that in group 1 strains. Additionally, we found that the rates of gene death were slightly higher than the gain turnover rates. There were two possible explanations for these results. The number of genes gained from HGT as well as gene duplication events might be significant enough to account for the increase of microbial DNA content and novel functions, and play a key role in evolution (Mira et al., 2001; Navarre et al., 2006). This hypothesis may also be supported by an earlier genetic study on the evolution of *Bacillus anthracis* virulence, which revealed that key genes that cause anthrax in this bacterium were identified as acquired by HGT (Zwick et al., 2012). However, a conceivable explanation underlying environment-dependent conditional dispensability indicates that genes in a given species would be dispensable if they were related to certain processes that were only required in a specific untested environments (Albalat and Cañestro, 2016). Of note, it is challenging to assess which genes are regarded as dispensable or essential components by coupling genotypes with phenotypes. In view of the complexity of environmental conditions in copper mines, low deletion pressures might provide microbes with a major fitness advantage for growth in adverse environments. Furthermore, large genomes in bacteria correspond to species that have the ability to tackle various environmental stimuli (Schneiker et al., 2007). Accordingly, large bacterial genomes might have an adaptive role in the evolution of group 2 strains.

CONCLUSION

Six chromosomes of the extreme acidophile *A. caldus* were valuable resources for the investigation of genetic diversity and evolutionary adaptation. A phylogenetic tree based on chromosomal sequences of *A. caldus* species showed a potential correlation between genomic diversity

and geochemical characteristics. Further analysis revealed that chemical constraint in respective natural habitat might be a determinant contributing to genetic diversification. Apparently, genetic analyses indicated that gene gain and loss were both dominant evolutionary forces in the adaptive evolution of *A. caldus* species. During adaptation to these adverse environmental conditions, GD rates varied in different settings, resulting in genomic differentiation and speciation. The compact and streamlined genomes might undergo selective deletion pressure, whereas large genomes had been extensively recruited by intraspecific or interspecific genetic exchange. These genome-guided findings in our study, to some extent, provide novel insights into the evolutionary adaptation of *A. caldus* species.

AUTHOR CONTRIBUTIONS

XnZ, XL, and QH conceived and designed the experiments. XnZ and WD performed the experiments. XnZ analyzed the data. XnZ wrote the paper. XoZ, FF, DP, WH, and HY revised the manuscript.

REFERENCES

- Acuña, L. G., Cárdenas, J. P., Covarrubias, P. C., Haristoy, J. J., Flores, R., Nuñez, H., et al. (2013). Architecture and gene repertoire of the flexible genome of the extreme acidophile *Acidithiobacillus caldus*. *PLoS ONE* 8:e78237. doi: 10.1371/journal.pone.0078237
- Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. *Nat. Rev. Genet.* 17, 379–391. doi: 10.1038/nrg.2016.39
- Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., et al. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461, 1243–1247. doi: 10.1038/nature08480
- Bentley, S. D., and Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* 38, 771–791. doi: 10.1146/annurev.genet.38.072902.094318
- Bezuidt, O., Lima-Mendez, G., and Reva, O. N. (2009). SeqWord gene island sniffer: a program to study the lateral genetic exchange among bacteria. *W. Acad. Sci. Eng. Techn.* 58, 410–415.
- Carretero-Paulet, L., Librado, P., Chang, T., Ibarra-Laclette, E., Herrera-Estrella, L., Rozas, J., et al. (2015). High gene family turnover rates and gene space adaptation in the compact genome of the carnivorous plant *Utricularia gibba*. *Mol. Biol. Evol.* 32, 1284–1295. doi: 10.1093/molbev/msv020
- Chen, L., Ren, Y., Lin, J., Liu, X., Pang, X., and Lin, J. (2012). *Acidithiobacillus caldus* sulfur oxidation model based on transcriptome analysis between the wild type and sulfur oxygenase reductase defective mutant. *PLoS ONE* 7:e39470. doi: 10.1371/journal.pone.0039470
- Deng, W., Wang, Y., Liu, Z., Cheng, H., and Xue, Y. (2014). HemI: a toolkit for illustrating heatmaps. *PLoS ONE* 9:e111988. doi: 10.1371/journal.pone.0111988
- Dhillon, B. K., Laird, M. R., Shay, J. A., Winsor, G. L., Lo, R., Nizam, F., et al. (2015). IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res.* 43, W104–W108. doi: 10.1093/nar/gkv401
- Dopson, M., and Lindström, E. B. (1999). Potential role of *Thiobacillus caldus* in arsenopyrite bioleaching. *Appl. Environ. Microbiol.* 65, 36–40.
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. doi: 10.1186/s13059-015-0721-2
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Fouts, D. E., Brinkac, L., Beck, E., Inman, J., and Sutton, G. (2012). PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Res.* 40, e172. doi: 10.1093/nar/gks757
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094
- Gan, H. M., Hudson, A. O., Rahman, A. Y. A., Chan, K. G., and Savka, M. A. (2013). Comparative genomic analysis of six bacteria belonging to the genus *Novosphingobium*: insights into marine adaptation, cell-cell signaling and bioremediation. *BMC Genomics* 14:431. doi: 10.1186/1471-2164-14-431
- Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. Oxford: Oxford University Press.
- Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238. doi: 10.1093/oxfordjournals.molbev.a004046
- González, A., Bellenberg, S., Mamani, S., Ruiz, L., Echeverría, A., Souler, L., et al. (2013). AHL signaling molecules with a large acyl chain enhance biofilm formation on sulfur and metal sulfides by the bioleaching bacterium *Acidithiobacillus ferrooxidans*. *Appl. Microbiol. Biotechnol.* 97, 3729–3737. doi: 10.1007/s00253-012-4229-3
- González, C., Yanquepe, M., Cardenas, J. P., Valdes, J., Quatrini, R., Holmes, D. S., et al. (2014). Genetic variability of psychrotolerant *Acidithiobacillus ferrivorans* revealed by (meta)genomic analysis. *Res. Microbiol.* 165, 726–734. doi: 10.1016/j.resmic.2014.08.005
- Hahn, M. W., Han, M. V., and Han, S. (2007). Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet.* 3:e197. doi: 10.1371/journal.pgen.0030197
- Hallberg, K. B., and Lindström, E. B. (1994). Characterization of *Thiobacillus caldus* sp. nov., a moderately thermophilic acidophile. *Microbiology* 140, 3451–3456. doi: 10.1099/13500872-140-12-3451
- Hallberg, K. B., and Lindström, E. B. (1996). Multiple serotypes of the moderate thermophile *Thiobacillus caldus*, a limitation of immunological assays for biomining microorganisms. *Appl. Environ. Microbiol.* 62, 4243–4246.
- Hsiao, W., Wan, I., Jones, S. J., and Brinkman, F. S. L. (2003). IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19, 418–420. doi: 10.1093/bioinformatics/btg004

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 31570113 and No. 41573072) and the Fundamental Research Funds for the Central Universities of Central South University (No. 2016zzts102).

ACKNOWLEDGMENTS

We thank Dr. Qichao Tu in Zhejiang University and Dr. Guanyun Wei in Nanjing Normal University for helpful discussion and suggestions. Also, we thank the National Center for Biotechnology Information (NCBI) for providing the genomic sequences of *A. caldus* strains ATCC 51756 and SM-1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01960/full#supplementary-material>

- Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689
- Jacobsen, A., Hendriksen, R. S., Aarestrup, F. M., Ussery, D. W., and Friis, C. (2011). The *Salmonella enterica* Pan-genome. *Microb. Ecol.* 62, 487–504. doi: 10.1007/s00248-011-9880-1
- Ji, B., Zhang, S., Arnoux, P., Rouy, Z., Alberto, F., Philippe, N., et al. (2014). Comparative genomic analysis provides insights into the evolution and niche adaptation of marine *Magnetospira* sp. QH-2 strain. *Environ. Microbiol.* 16, 525–544. doi: 10.1111/1462-2920.12180
- Justice, N. B., Norman, A., Brown, C. T., Singh, A., Thomas, B. C., and Banfield, J. F. (2014). Comparison of environmental and isolate *Sulfobacillus* genomes reveals diverse carbon, sulfur, nitrogen, and hydrogen metabolisms. *BMC Genomics* 15:1107. doi: 10.1186/1471-2164-15-1107
- Kimura, M. (1984). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kulmuni, J., Wurm, Y., and Pamilo, P. (2013). Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. *Heredity* 110, 538–547. doi: 10.1038/hdy.2012.122
- Langille, M. G., Hsiao, W. W., and Brinkman, F. S. (2008). Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics* 9:329. doi: 10.1186/1471-2105-9-329
- Levicán, G., Ugalde, J. A., Ehrenfeld, N., Maass, A., and Parada, P. (2008). Comparative genomic analysis of carbon and nitrogen assimilation mechanisms in three indigenous bioleaching bacteria: predictions and validations. *BMC Genomics* 9:581. doi: 10.1186/1471-2164-9-581
- Librado, P., Vieira, F. G., and Rozas, J. (2012). BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* 28, 279–281. doi: 10.1093/bioinformatics/btr623
- Librado, P., Vieira, F. G., Sánchez-Gracia, A., Kolokotronis, S., and Rozas, J. (2014). Mycobacterial phylogenomics: an enhanced method for gene turnover analysis reveals uneven levels of gene gain and loss among species and gene families. *Genome Biol. Evol.* 6, 1454–1465. doi: 10.1093/gbe/evu117
- Lin, W., Wang, Y., Gorby, Y., Neelson, K., and Pan, Y. (2013). Integrating niche-based process and spatial process in biogeography of magnetotactic bacteria. *Sci. Rep.* 3, 1643. doi: 10.1038/srep01643
- Liu, X., Lin, J., Zhang, Z., Bian, J., Zhao, Q., Liu, Y., et al. (2007). Construction of conjugative gene transfer system between *E. coli* and moderately thermophilic, extremely acidophilic *Acidithiobacillus caldus* MTH-04. *J. Microbiol. Biotechnol.* 17, 162–167.
- MacLean, D., Jones, J. D. G., and Studholme, D. J. (2009). Application of 'next-generation' sequencing technologies to microbial genetics. *Nat. Rev. Microbiol.* 7, 287–296. doi: 10.1038/nrmicro2122
- Mangold, S., Valdés, J., Holmes, D. S., and Dopson, M. (2011). Sulfur metabolism in the extreme acidophile *Acidithiobacillus caldus*. *Front. Microbiol.* 2:17. doi: 10.3389/fmicb.2011.00017
- Marsh, R. M., and Norris, P. R. (1983). The isolation of some thermophilic, autotrophic, iron-and sulfur-oxidizing bacteria. *FEMS Microbiol. Lett.* 17, 311–315. doi: 10.1111/j.1574-6968.1983.tb00426.x
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60
- Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626
- Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17, 589–596. doi: 10.1016/S0168-9525(01)02447-7
- Moran, N. A. (2003). Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr. Opin. Microbiol.* 6, 512–518. doi: 10.1016/j.mib.2003.08.001
- Moya, A., Gil, R., Latorre, A., Peretó, J., Garcillán-Barcia, M. P., and De La Cruz, F. (2009). Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol. Rev.* 33, 225–235. doi: 10.1111/j.1574-6976.2008.00151.x
- Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., and Bernardi, G. (2006). Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochem. Biophys. Res. Commun.* 347, 1–3. doi: 10.1016/j.bbrc.2006.06.054
- Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., and Bernardi, G. (2004). Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. *FEBS Lett.* 573, 73–77. doi: 10.1016/j.febslet.2004.07.056
- Navarre, W. W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S. J., et al. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* 313, 236–238. doi: 10.1126/science.1128794
- Nei, M., and Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39, 121–152. doi: 10.1146/annurev.genet.39.073003.112240
- Noguchi, H., Taniguchi, T., and Itoh, T. (2008). MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* 15, 387–396. doi: 10.1093/dnares/dsn027
- Nuñez, P. A., Romero, H., Farber, M. D., and Rocha, E. P. C. (2013). Natural selection for operons depends on genome size. *Genome Biol. Evol.* 5, 2242–2254. doi: 10.1093/gbe/evt174
- Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500
- Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer.
- Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23, 263–286. doi: 10.1146/annurev.es.23.110192.001403
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619. doi: 10.1371/journal.pone.0030619
- Rawlings, D. E. (1998). Industrial practice and the biology of leaching of metals from ores. *J. Ind. Microbiol. Biotechnol.* 20, 268–274. doi: 10.1038/sj.jim.2900522
- Sánchez-Gracia, A., Vieira, F. G., and Rozas, J. (2009). Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103, 208–216. doi: 10.1038/hdy.2009.55
- Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M. O., et al. (2007). Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* 25, 1281–1289. doi: 10.1038/nbt1354
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014
- Silver, S., and Phung, L. T. (1996). Bacterial heavy metal resistance: new surprises. *Annu. Rev. Microbiol.* 50, 753–789. doi: 10.1146/annurev.micro.50.1.753
- Sugawara, M., Epstein, B., Badgley, B. D., Unno, T., Xu, L., Reese, J., et al. (2013). Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biol.* 14, R17. doi: 10.1186/gb-2013-14-2-r17
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi: 10.1016/j.mib.2008.09.006
- Tian, C. F., Zhou, Y. J., Zhang, Y. M., Li, Q. Q., Zhang, Y. Z., Li, D. F., et al. (2012). Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8629–8634. doi: 10.1073/pnas.1120436109
- Treangen, T. J., and Rocha, E. P. C. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7:e1001284. doi: 10.1371/journal.pgen.1001284
- Ullrich, S. R., González, C., Poehlein, A., Tischler, J. S., Daniel, R., Schlömann, M., et al. (2016). Gene loss and horizontal gene transfer contributed to the genome

- evolution of the extreme acidophile “Ferrovum”. *Front. Microbiol.* 7:797. doi: 10.3389/fmicb.2016.00797
- Valdés, J., Pedroso, I., Quatrini, R., Dodson, R. J., Tettelin, H., Blake, R., et al. (2008a). *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Genomics* 9:597. doi: 10.1186/1471-2164-9-597
- Valdés, J., Pedroso, I., Quatrini, R., and Holmes, D. S. (2008b). Comparative genome analysis of *Acidithiobacillus ferrooxidans*, *A. thiooxidans* and *A. caldus*: Insights into their metabolism and ecophysiology. *Hydrometallurgy* 94, 180–184. doi: 10.1016/j.hydromet.2008.05.039
- Valdes, J., Quatrini, R., Hallberg, K., Dopson, M., Valenzuela, P. D. T., and Holmes, D. S. (2009). Draft genome sequence of the extremely acidophilic bacterium *Acidithiobacillus caldus* ATCC 51756 reveals metabolic versatility in the genus *Acidithiobacillus*. *J. Bacteriol.* 191, 5877–5878. doi: 10.1128/JB.00843-09
- Vieira, F. G., and Rozas, J. (2011). Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biol. Evol.* 3, 476–490. doi: 10.1093/gbe/evr033
- Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W. F., et al. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 7:142. doi: 10.1186/1471-2105-7-142
- Watling, H. R. (2006). The bioleaching of sulphide minerals with emphasis on copper sulphides—a review. *Hydrometallurgy* 84, 81–108. doi: 10.1016/j.hydromet.2006.05.001
- Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., and van der Lelie, D. (2011). Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol. Rev.* 35, 299–323. doi: 10.1111/j.1574-6976.2010.00249.x
- Yin, H., Niu, J., Ren, Y., Cong, J., Zhang, X., Fan, F., et al. (2015). An integrated insight into the response of sedimentary microbial communities to heavy metal contamination. *Sci. Rep.* 5, 14266. doi: 10.1038/srep14266
- Yin, H., Zhang, X., Li, X., He, Z., Liang, Y., Guo, X., et al. (2014). Whole-genome sequencing reveals novel insights into sulfur oxidation in the extremophile *Acidithiobacillus thiooxidans*. *BMC Microbiol.* 14:179. doi: 10.1186/1471-2180-14-179
- You, X. Y., Guo, X., Zheng, H. J., Zhang, M. J., Liu, L. J., Zhu, Y. Q., et al. (2011). Unraveling the *Acidithiobacillus caldus* complete genome and its central metabolisms for carbon assimilation. *J. Genet. Genomics* 38, 243–252. doi: 10.1016/j.jgg.2011.04.006
- Zhang, X., Feng, X., Tao, J., Ma, L., Xiao, Y., Liang, Y., et al. (2016a). Comparative genomics of the extreme acidophile *Acidithiobacillus thiooxidans* reveals intraspecific divergence and niche adaptation. *Int. J. Mol. Sci.* 17, 1355. doi: 10.3390/ijms17081355
- Zhang, X., Liu, X., Liang, Y., Fan, F., Zhang, X., and Yin, H. (2016b). Metabolic diversity and adaptive mechanisms of iron- and/or sulfur-oxidizing autotrophic acidophiles in extremely acidic environments. *Environ. Microbiol. Rep.* 8, 738–751. doi: 10.1111/1758-2229.12435
- Zhang, X., Niu, J., Liang, Y., Liu, X., and Yin, H. (2016c). Metagenome-scale analysis yields insights into the structure and function of microbial communities in a copper bioleaching heap. *BMC Genet.* 17:21. doi: 10.1186/s12863-016-0330-4
- Zhang, X., Huaqun, Y., Yili, L., Qiu, G., and Liu, X. (2015). Theoretical model of the structure and the reaction mechanisms of sulfur oxygenase reductase in *Acidithiobacillus thiooxidans*. *Adv. Mater. Res.* 1130, 67–70. doi: 10.4028/www.scientific.net/AMR.1130.67
- Zhou, Z., Fang, Y., Li, Q., Yin, H., Qin, W., Liang, Y., et al. (2012). Global transcriptional analysis of stress-response strategies in *Acidithiobacillus ferrooxidans* ATCC 23270 exposed to organic extractant-Lix984n. *World J. Microbiol. Biotechnol.* 28, 1045–1055. doi: 10.1007/s11274-011-0903-3
- Zuo, G., and Hao, B. (2015). CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. *Genomics Proteomics Bioinformatics* 13, 321–331. doi: 10.1016/j.gpb.2015.08.004
- Zwick, M. E., Joseph, S. J., Didelot, X., Chen, P. E., Bishop-Lilly, K. A., Stewart, A. C., et al. (2012). Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of *Bacillus anthracis*. *Genome Res.* 22, 1512–1524. doi: 10.1101/gr.134437.111

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhang, Liu, He, Dong, Zhang, Fan, Peng, Huang and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome Analysis of a New *Rhodothermaceae* Strain Isolated from a Hot Spring

Kian Mau Goh^{1*}, Kok-Gan Chan², Soon Wee Lim¹, Kok Jun Liew¹, Chia Sing Chan¹, Mohd Shahir Shamsir¹, Robson Ee² and Tan-Guan-Sheng Adrian²

¹ Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia, ² Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

OPEN ACCESS

Edited by:

Jesse G. Dillon,
California State University,
Long Beach, USA

Reviewed by:

Bradley Stevenson,
University of Oklahoma, USA
Xi-Ying Zhang,
Shandong University, China

*Correspondence:

Kian Mau Goh
gohkianmau@utm.my

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 13 April 2016

Accepted: 04 July 2016

Published: 14 July 2016

Citation:

Goh KM, Chan K-G, Lim SW, Liew KJ,
Chan CS, Shamsir MS, Ee R and
Adrian T-G-S (2016) Genome Analysis
of a New *Rhodothermaceae* Strain
Isolated from a Hot Spring.
Front. Microbiol. 7:1109.
doi: 10.3389/fmicb.2016.01109

A bacterial strain, designated RA, was isolated from water sample of a hot spring on Langkawi Island of Malaysia using marine agar. Strain RA is an aerophilic and thermophilic microorganism that grows optimally at 50–60°C and is capable of growing in marine broth containing 1–10% (w/v) NaCl. 16S rRNA gene sequence analysis demonstrated that this strain is most closely related (<90% sequence identity) to *Rhodothermaceae*, which currently comprises of six genera: *Rhodothermus* (two species), *Salinibacter* (three species), *Salisaeta* (one species), *Rubricoccus* (one species), *Rubrivirga* (one species), and *Longimonas* (one species). Notably, analysis of average nucleotide identity (ANI) values indicated that strain RA may represent the first member of a novel genus of *Rhodothermaceae*. The draft genome of strain RA is 4,616,094 bp with 3630 protein-coding gene sequences. Its GC content is 68.3%, which is higher than that of most other genomes of *Rhodothermaceae*. Strain RA has genes for sulfate permease and arylsulfatase to withstand the high sulfur and sulfate contents of the hot spring. Putative genes encoding proteins involved in adaptation to osmotic stress were identified which encode proteins namely Na⁺/H⁺ antiporters, a sodium/solute symporter, a sodium/glutamate symporter, trehalose synthase, malto-oligosyltrehalose synthase, choline-sulfatase, potassium uptake proteins (TrkA and TrkH), osmotically inducible protein C, and the K⁺ channel histidine kinase KdpD. Furthermore, genome description of strain RA and comparative genome studies in relation to other related genera provide an overview of the uniqueness of this bacterium.

Keywords: *Rhodothermaceae*, *Rhodothermus*, *Salinibacter*, *Salisaeta*, strain RA, *Rubricoccus*, *Rubrivirga*, *Longimonas*

INTRODUCTION

The family *Rhodothermaceae* is placed under the order *Cytophagales* of the class *Cytophagia* within the phylum *Bacteroidetes*. At the time of writing, *Rhodothermaceae* consisted of six genera: *Rhodothermus* (Alfredsson et al., 1988; Marteinsson et al., 2010), *Salinibacter* (Antón et al., 2002; Makhdoumi-Kakhki et al., 2012), *Salisaeta* (Vaisman and Oren, 2009), *Rubricoccus* (Park et al., 2011), *Rubrivirga* (Park et al., 2013), and the recently described genus *Longimonas* (Xia et al., 2015). *Rhodothermaceae* is a family of gram-negative, non-sporulating, chemoorganotrophic aerobes shaped as rods or cocci, and most strains are pigmented (Park et al., 2014).

Rhodothermus marinus was isolated from a submarine hot spring, at a depth of 2–3 m from the sea surface during low tide (Alfredsson et al., 1988), while *Rhodothermus profundus* was obtained from a deep-sea (2634 m depth) hydrothermal field (Marteinsson et al., 2010). *Rhodothermus obamensis* was initially described as a type strain of *Rhodothermus* (Sako et al., 1996) but was later regarded as a synonym of *R. marinus* owing to high 16S rRNA gene sequence similarity, high DNA–DNA reassociation values, and similar fatty acid profiles (Silva et al., 2000). For *R. marinus* DSM 4252^T (Nolan et al., 2009) complete genome is available in GenBank (CP001807). The genome of the sister strain *R. marinus* SG0.5JP17-172 (CP003029) is also available but has not been published. Draft genomes have been deposited for the other strains of *Rhodothermus*, which include *R. profundus* DSM 22212^T (BioProject ID: PRJNA303571), *R. marinus* JCM 9785^T (PRJDB841), and *R. marinus* SG0.5JP17-171 (PRJNA52953).

Members of the genus *Salinibacter* (*Salinibacter iranicus*, *Salinibacter luteus*, and *Salinibacter ruber*) were isolated from a hypersaline evaporating water body (Park et al., 2014). *S. ruber* requires a high salt concentration, exhibiting optimum growth in media with 15–25% (w/v) total salts (Antón et al., 2002). In addition, sodium chloride and magnesium chloride are vital for the growth of *S. ruber* (Antón et al., 2002). The complete genome of *S. ruber* DSM 13855^T (CP000159.1) has been assembled into one contig of 3,386,737 bp, while the genome of *S. ruber* M8 (PRJNA48827) is available in draft form. A species of another genus, the long, rod-shaped bacterium *Salisaeta longa* S4-4^T (= DSM 21114^T; ATTH00000000.1), was isolated from hypersaline water bodies formed from premixed Dead Sea and Red Sea water samples (Vaisman and Oren, 2009). Although few studies have been conducted on strain S4-4^T, its genome has been sequenced under the Community Science Program of the Joint Genome Institute (Project ID: 404303).

Rubricoccus marinus SG-29^T, a reddish, coccoid bacterium isolated from shallow water of the North Pacific Ocean, is the type strain of the type species of *Rubricoccus* (Park et al., 2011). The average GC content of strain SG-29^T (68.9 mol%, determined by HPLC) is the highest value known for *Rhodothermaceae*. Two years after the report of *Rubricoccus marinus*, the same group of researchers isolated another bacterium from deep seawater of the western North Pacific Ocean. They proposed a new genus, *Rubrivirga*, with *Rubrivirga marina* SAORIC-28^T as the sole species (Park et al., 2013). Recently, a new genus of *Rhodothermaceae*, *Longimonas*, was proposed (Xia et al., 2015). *Longimonas halophila* SYD6^T was obtained from a marine solar saltern on the coast of Weihai, China (Xia et al., 2015). The 16S rRNA gene sequence similarity between *L. halophila* SYD6^T and *S. longa* is less than 92% (Xia et al., 2015). Like most other strains of *Rhodothermaceae*, strain SYD6^T is red. At the time of writing, genome information for *Rubricoccus*, *Rubrivirga*, and *Longimonas* were not available.

Among the six aforementioned genera, *Rhodothermus* and *Salinibacter* are better studied (Mongodin et al., 2005; Peña et al., 2005, 2010; Rosselló-Mora et al., 2008; Oren, 2013; Oren et al., 2016). Because most *Rhodothermus* spp. can grow optimally at 65°C, numerous attempts to mine and examine thermostable proteins from this genus have been reported. Examples of these

enzymes include amylase and pullulanase (Gomes et al., 2003), α -L-arabinofuranosidase (Gomes et al., 2000), cellulase (Okano et al., 2014), cellobiose 2-epimerase (Ojima et al., 2011), chitinase (Hobel et al., 2005), and α -galactosidase (Blücher et al., 2000).

In this work, a bacterium designated as strain RA was isolated from a saline hot spring sample. Analyses of the genome, 16S rRNA gene, and housekeeping gene sequences suggested that the strain could represent a new genus of *Rhodothermaceae*. The purpose of this report is to describe the phenotypic characteristics of this strain and present its genome sequence.

MATERIALS AND METHODS

Water Sample Analyses and Bacterial Characterization

Three water samples were collected from a hot spring near the village of Ayer Hangat (6°25'22"N, 99°48'49"E) and analyzed by Allied Chemists Laboratory Sdn. Bhd. (Malaysia) within 3 days of sampling, in accordance with American Public Health Association and United States Environmental Protection Agency guidelines.

Marine agar (Conda, Torrejón de Ardoz Madrid, Spain) was adjusted to pH 7.6 with 3 M NaOH. A 100 μ L aliquot of a water sample was spread on the agar and incubated at 50°C. After 2 days of incubation, several colonies appeared on the agar. Three distinct colonies were repeatedly streaked on the same medium to obtain pure colonies as confirmed by a single morphology and size when examined directly using a DM300 light microscope (Leica Microsystems, Germany). 16S rRNA gene sequences of these isolates were amplified using the forward primer 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and the reverse primer 1525R (5'-AAGGAGGTGWTCCARCC-3') (Lane, 1991; Chai et al., 2012). Sequencing was performed at Malaysian First BASE service provider.

The isolated strains were Gram stained and then examined under a Leica DM300 light microscope. For investigation of endospore formation, endospore staining was conducted on bacterial colony obtained from a fresh culture plate as well as a week-old culture (Schaeffer and Fulton, 1933). The growth temperature range of strain RA was assessed by growing this bacterium in marine broth and were incubated at 4, 10, 37, 45, 50, 55, 60, 65, 70, and 75°C for up to 3 days. To determine the optimal pH which support the growth of this bacterium, strain RA was grown in non-buffered marine broths with various pH-values (pH 3.0–12 with interval pH-value of 0.5). The salt tolerance of strain RA was determined in marine broth [2% (w/v) NaCl] and half-strength marine broth [1% (w/v) NaCl], at increments of 1.0% (w/v) up to 20% w/v.

Carbohydrate utilization for strain RA was measured using API 50CHB/E test strips (bioMérieux, Marcy-l'Étoile, France). Catalase activity and oxidase activity were determined as described by Cowan and Steel (1965). Motility of strain RA was determined using semi-solid medium and bacterial cells were inoculated with a straight wire making a single stab down the center of the tube to about half the depth of the medium, followed by incubation at 50°C and were

examined for up to 3 days. Susceptibility to different antibiotics was determined by the disc diffusion method (Kirby–Bauer antibiotic testing) on marine agar at 50°C. Bacterial lawns were prepared by spreading the cells on plates with sterile cotton swabs dipped in colony suspensions adjusted to a 2.0 McFarland standard. The following antibiotics were tested: ampicillin, bacitracin, chloramphenicol, erythromycin, gentamicin, kanamycin, nalidixic acid, ciprofloxacin, colistin, penicillin, rifampicin, sulfamethoxazole, and vancomycin.

DNA Purification

Cells were scraped from marine agar plates and subjected to genomic DNA extraction using a Qiagen DNeasy Blood and Tissue Kit (Qiagen, Venlo, Netherlands), according to the manufacturer's instructions. A NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) was used to determine the purity (A_{260}/A_{280} ratio) and concentration of the DNA, which were 1.98 and 1400 ng μL^{-1} , respectively.

Genome Sequencing and Annotation

Whole-genome sequencing libraries were prepared using a Nextera DNA Sample Preparation Kit according to the manufacturer's guidelines (Illumina, Inc., San Diego, CA, USA). Paired-end (2×100 bp) sequencing was performed on the HiSeq 2500 platform using the HiSeq Rapid SBS Kit v2 (Illumina, Inc., San Diego, CA, USA). Adapter sequence removal, quality trimming, and de novo genome assembly were conducted using CLC Genomics Workbench version 7.5 (CLC Bio, Aarhus, Denmark). The resulting sequences were then annotated with NCBI Prokaryotic Genome Annotation Pipeline version 2.10, using the “best-placed reference protein set” method, and GeneMarkS+ version 3.1, which integrates information regarding protein alignments, frameshifted genes, non-coding RNA sequences, and DNA-specific statistical patterns typical of protein-coding and non-coding regions into gene predictions (Besemer and Borodovsky, 2005). Annotation and KEGG pathway prediction for strain RA were performed using the online service Pathosystems Resource Integration Center (PATRIC; Mao et al., 2015) as well as Integrated Microbial Genomes/Expert Review (IMG/ER; Markowitz et al., 2009). VirulenceFinder version 1.5 (threshold 85%, minimum length 60%; Joensen et al., 2014) and ResFinder version 2.1 (threshold 60%, minimum length 60%; Zankari et al., 2012) were used to predict the presence of putative virulence factors and antimicrobial resistance genes, respectively.

The following genomes were downloaded from GenBank: *Rhodothermus marinus* DSM 4252^T (CP001807.1), *S. ruber* DSM 13855^T (CP000159.1), and *S. longa* DSM 21114^T (ATTH00000000.1). Genome comparisons were performed using the average nucleotide identity (ANI) function of IMG/ER. The complete 16S rRNA gene sequence of strain RA was submitted to GenBank under accession number KU517707.

Phylogenetic Analysis

The 16S rRNA gene sequences phylogenetic tree was constructed using Neighbor-joining method with 1000 bootstrap replicates.

All the methods above were conducted using Molecular Evolutionary Genetics Analysis software (MEGA, Version 6.0; Tamura et al., 2013).

Data Access

Sequencing data for strain RA are available online as BioProject PRJNA308615, NCBI taxonomy ID 1779382, Patric genome id 1100069.3, and IMG/ER Taxon ID 2648501586 (GOLD ID Gs0118059). This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number LRDG00000000. The version described in this paper is version LRDG01000000. A Fasta file of the strain RA genome is available in the following figshare link (<https://figshare.com/s/60c1a70599d11684aaf7>). All Supplementary Materials are available in figshare.

RESULTS AND DISCUSSION

Site Description

The hot spring sampled in this study is located near Ayer Hangat (AH) on Langkawi Island, Malaysia. Owing to its relatively low temperature ($\sim 45^\circ\text{C}$), the hot spring has been used for balneotherapy for tourists and local residents. The water in this hot spring is trapped within a man-made pit and is therefore nearly stagnant and is slightly yellowish, and there is a thin biomat (<1 cm) on the sides of the pit that is light brown, green, and yellow. Water samples taken from this site contained high concentrations of Na^+ (7900 mg L^{-1}), Cl^- (13,800 mg L^{-1}), Mg^{2+} (390 mg L^{-1}), SO_4^{2-} (950 mg L^{-1}), S (480 mg L^{-1}), and CaCO_3 (5020 mg L^{-1}).

Cell Morphology and Biochemical Analyses

Three bacteria (designated as strains RA, RB, and RC) were isolated from a water sample taken at the AH hot spring. Based on 16S rRNA gene sequences, strains RB and RC had the closest match with *Bacillus* spp. and *Geobacillus* spp., respectively. Since strain RA had low sequence similarity to known *Rhodothermaceae* (see below), it was selected for subsequent analyses. The strain RA colonies appeared smooth, mucoid, and raised, with ~ 1 mm margins. Furthermore, while the majority of *Rhodothermaceae* strains form orange-reddish colonies (Park et al., 2014), the colonies of strain RA were cream colored (**Table 1**). Meanwhile, light microscopy analysis demonstrated that this isolate is a gram-negative, rod-shaped bacterium that can be arranged in pairs and chains or as lone cells. *Rhodothermus* spp. are known to occur singly, but not in chains or filaments (Park et al., 2014). Strain RA was motile and catalase positive and it grew well in marine broth and half-strength marine broth but failed to grow on Luria–Bertani, Reasoner's 2A, Mueller–Hinton, or nutrient agar, with or without supplementation with 2% (w/v) NaCl. While the microorganism grew at 37–60°C and pH 5.0–9.0, optimal growth occurred at 50–60°C and under circumneutral conditions. Strain RA was unable to grow at 65°C. Furthermore, strain RA exhibited growth in marine broth supplemented with 1–10% (w/v) NaCl. On marine agar, strain RA was susceptible to erythromycin (zone of inhibition of 25 mm around a 5 μg disc), rifampicin (25 mm

TABLE 1 | General features and information regarding the sequencing of the strain RA genome, according to the MGS (Minimum information about a genome sequence) recommendations.

| Items | Description |
|-----------------------------|--|
| GENERAL FEATURES | |
| Gram stain | Negative |
| Cell shape | Rod |
| Pigmentation | Cream |
| Motility | Motile |
| Temperature range; optimum | 37–60°C; 50–60°C |
| Salinity range; optimum | 1–10.0% NaCl (w/v); 1–5% NaCl (w/v) |
| pH range; optimum | 5.0–9.0; 7.6 |
| Carbon source | Varied |
| MIGS 4.0 DATA | |
| Submitted_to_insd | GenBank |
| investigation_type | bacteria_archaea |
| project_name | <i>Rhodothermaceae</i> bacterium RA genome sequencing |
| experimental_factor | NA |
| lat_lon | NA |
| Depth | NA |
| geo_loc_name | Langkawi, Malaysia |
| collection_date | 2014-10-10 |
| env_biome | Hot spring |
| env_feature | Water |
| env_material | Water with high NaCl content |
| env_package | Missing |
| num_replicons | NA |
| ref_biomaterial | NA |
| source_mat_id | NA |
| Pathogenicity | NA |
| biotic_relationship | Free-living |
| trophic_level | Chemoheterotroph |
| rel_to_oxygen | Aerobic |
| isol_growth_condt | Marine agar, pH 7.6, 50°C |
| GENOME ASSEMBLY DATA | |
| assembly_name | CLC bio CLC Genomics Workbench version 7.5 |
| finishing_strategy | <i>De novo</i> assembly to high quality draft; 121-fold coverage |
| sequencing Technology | HiSeq 2500 |

zone around a 5 µg disc), and penicillin (35 mm zone around a 10 U disc). Lastly, like *R. marinus* DSM 4252^T, strain RA is a non-spore-forming bacterium.

Phylogenetic Relationships

For phylogenetic analysis of strain RA, we sequenced the entire 16S rRNA coding region (1516 bp) of the microorganism and compared the resulting sequence against the GenBank and EzTaxon-e databases (Kim et al., 2012). These analyses indicated that the closest relative (99% identity, 98% coverage) of strain RA is the uncultured bacterial metagenome clone KSB113 (JX047086), which was identified in a marine hot spring at Kalianda Island, Indonesia (unpublished report). Moreover, the

genera closest to strain RA (86.5–89.3%) are those of the family *Rhodothermaceae*: *Rhodothermus* (Alfredsson et al., 1988; Sako et al., 1996; Marteinsson et al., 2010), *Salinibacter* (Antón et al., 2002; Makhdoumi-Kakhki et al., 2012), *Salisaeta* (Vaisman and Oren, 2009), *Rubricoccus* (Park et al., 2011), *Rubrivirga* (Park et al., 2013), and *Longimonas* (Xia et al., 2015; **Table 2, Figure 1**). The sequence identities of housekeeping genes of strain RA and species of the family *Rhodothermaceae* are also low, for instance (i) *recA*, 83% to *R. marinus* DSM 4252^T; (ii) *rpoD*, 85% to *R. marinus* DSM 4252^T and 82% to *S. ruber* DSM 13855^T; and (iii) *gyrB*, 79% to *R. marinus* DSM 4252^T and 79% to *S. ruber* DSM 13855^T.

We then compared the available genome sequences of several of these strains with the draft genome of strain RA (**Table 2**). In addition to exhibiting low 16S rRNA gene sequence identity (89.3%), strain RA and *R. marinus* DSM 4252^T had a low average ANI-value of 73.28 (**Table 2**). The ANI-value between strain RA and *R. marinus* SG0.5JP17-171 or *R. marinus* SG0.5JP17-172 was 73.23, while that between strain RA and *R. profundus* DSM 22212^T was 70.66. Meanwhile, the average ANI-values for strain RA when compared with *S. ruber* DSM 13855^T and *S. longa* DSM 21114^T were 71.85 and 71.47, respectively. Collectively, these data strongly suggest that strain RA is a novel strain of *Rhodothermaceae*.

Genomic Features of Strain RA

The draft genome of strain RA is 4,616,094 bp in length (**Table 3**), and the largest contig is 819,789 bp. The N75, N50, and N25-values are 89,274, 152,113, and 242,570 bp, respectively. The average coverage for the 91 contigs obtained was 120-fold. Notably, the genome of strain RA is larger than those of *Rhodothermus marinus* (3.3 Mbp), *S. ruber* (3.5–3.8 Mbp), and *S. longa* (3.1 Mbp; a draft genome with three contigs; **Table 2**). Furthermore, the genome is predicted to contain 3680 coding DNA sequences (CDS) and 50 rRNA genes, and it has an average GC content of 68.28%, which is markedly higher than that of *R. marinus* DSM 4252^T (64.3%; Nolan et al., 2009), *S. ruber* DSM 13855^T (66.2%; Mongodin et al., 2005), and *S. longa* DSM 21114^T (63.5%) but comparable to that of *R. marinus* KCTC 23197^T (68.9%; **Table 2**). Among the CDS, the functions of 2815 (76.49%) sequences could be predicted, while 879 (24.38%) of the total CDS are annotated as enzymes. Protein-coding genes related to clusters of orthologous groups (COGs) are shown in **Figure 2**.

Using the Phylogenetic Distribution of Genes function in IMG/ER, we determined the best hits for protein-coding genes in strain RA at a 60% identity cutoff. A total of 1408 genes (39% of the protein-coding genes) were found to be similar to genes in the phylum *Bacteroidetes*. The majority of these sequences are affiliated with class *Cytophagia* and unclassified classes, while only 52 genes are related to *Bacteroidia*, *Flavobacteriia*, and *Sphingobacteriia*. Table S1 shows that 1987 genes (54.7% of the protein-coding genes) were not assigned to any phylum, as they are specific to strain RA. When all protein-coding genes were examined at a 90% sequence identity threshold, 99.75% (3621) of the genes were classified as unassigned. This suggests that the gene sequences of strain RA are more diverse than those of its *Rhodothermaceae* counterparts (data not shown).

TABLE 2 | General data and comparison of strain RA and other Rhodothermaceae genera.

| Strain RA | <i>Rhodothermus marinus</i> DSM 4252 ^T | <i>Salinibacter ruber</i> DSM 13855 ^T | <i>Salisaeta longa</i> DSM 2114 ^T | <i>Rubricoccus marinus</i> KCTC 23197 ^T | <i>Rubrivirga marina</i> KCTC 23867 ^T | <i>Longimonas halophila</i> KCTC 42399 ^T |
|---|---|--|--|--|--|---|
| Taxonomy ID | 1,779,382 | 309,807 | 1,089,550 | 716,817 | 1,196,024 | n.a |
| Colony color | Cream | Red | Red | Red | Pale-red | Red |
| Cell shape | Rod | Rod | Rod | Cocci | Rod | Rod |
| NaCl range (%; w/v) | 1–10 | 11–30 | 5–20 | 1–5 | 1–5 | 4–25 |
| Temperature range | 37–60 | 37–47 | 37–46 | 5–37 | 10–37 | 20–50 |
| pH range | 5–9 | 6.5–8.0 | 6.5–8.5 | 5.0–9.0 | 6.0–9.0 | 6.5–8.5 |
| DNA GC% | 68.3 (genome) | 66.5 (HPLC); 66.2 (genome) | 62.9 (HPLC); 63.5 (genome) | 68.9 (HPLC) | 64.8–65.8 (HPLC) | 61.5 (HPLC) |
| 16S rRNA similarity to strain RA (%) | 100 | 86.50 | 88.83 | 88.6 | 88.78 | 88.14 |
| Genome size (bp) | 4,616,094 | 3,587,328 | 3,192,765 | – | – | – |
| Predicted coding DNA sequences | 3680 | 3086 | 2833 | – | – | – |
| Genome contigs no. | 91 | 1 | 3 | – | – | – |
| ANI against strain RA (%) | 100 | 71.85 | 71.47 | – | – | – |
| 16S rRNA accession number | KU517707 | NR_074895.1 | NR_044496.1 | NR_113053 | NR_109654 | KF955302.1 |
| Genome DDBJ/EMBL/GenBank accession number | LFDG00000000 | CP000159.1 | ATTH00000000.1 | – | – | – |
| GOLD ID | Gp0127843 | Gp0000409 | Gp0015133 | – | – | – |
| PATRIC ID | 1100069 | 309807 | 1089550 | – | – | – |
| KEGG complete genome T number | – | T00312 | – | – | – | – |
| References | This study | Alfredsson et al., 1988; Nolan et al., 2009 | Vaisman and Oren, 2009 | Park et al., 2011 | Park et al., 2013 | Xia et al., 2015 |

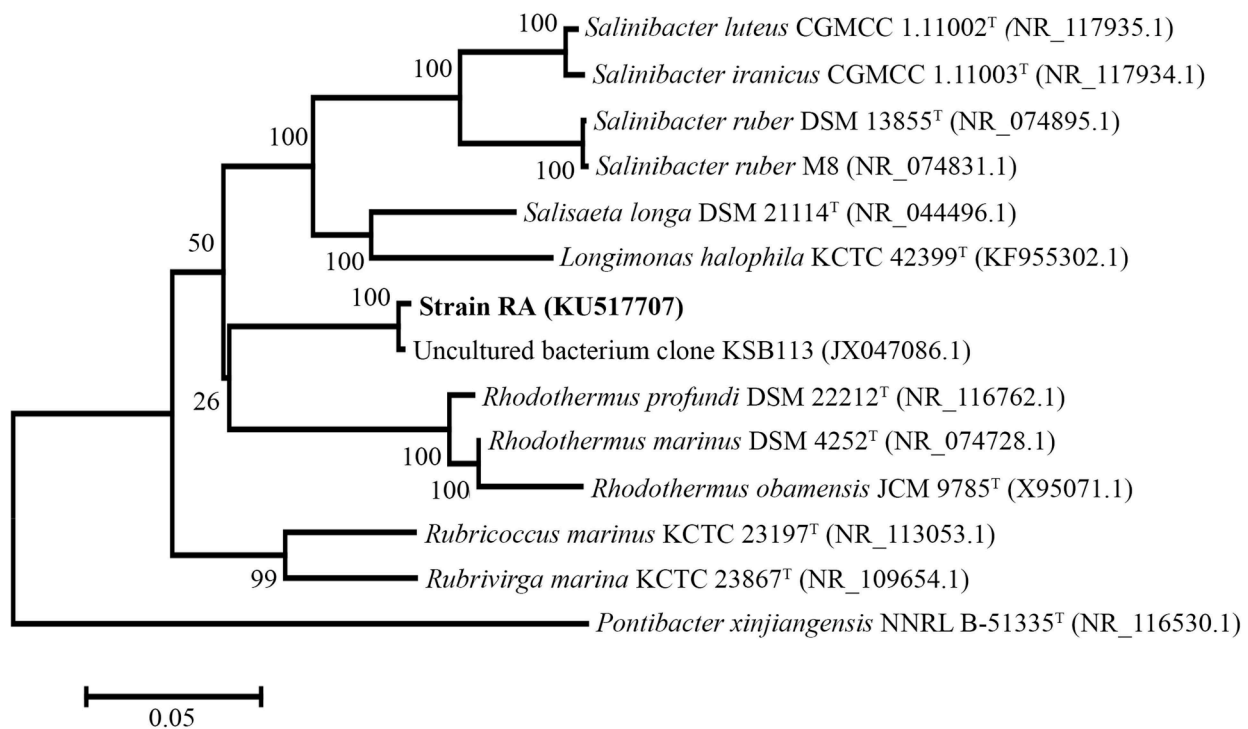


FIGURE 1 | Phylogenetic tree of strain RA and other *Rhodothermaceae* genera based on 16S rRNA sequencing. *Pontibacter xinjiangensis* NNRL B-51335^T was selected as the out group (<https://figshare.com/s/4f470ab9ce3727bf4392>).

Consistent with its observed motility, strain RA carries the genes necessary for complete flagellum assembly. Furthermore, genomic analysis using the antiSMASH predictor program (Blin et al., 2013) identified a gene cluster encoding a type III polyketide synthase. The key enzyme within this cluster has a structural domain related to chalcone synthase; thus, strain RA might use this enzyme to produce certain types of aromatic ketones, including chalcones, which possess important properties for industrial use (Singh et al., 2014).

Comparison of the Genome of Strain RA with those of Other *Rhodothermaceae* Strains

At the time of writing, the genomes of three *R. marinus* strains and two *S. ruber* strains, as well as the draft genome of *S. longa* DSM 21114^T, were publicly available. Unless otherwise specified, we chose to compare the genome of strain RA with the complete genome sequences of *R. marinus* DSM 4252^T and *S. ruber* DSM 13855^T.

RAST analysis identified 215 and 221 strain RA genes that are not carried by *R. marinus* DSM 4252^T and *S. ruber* DSM 13855^T, respectively (Table S2), including genes encoding DnaK-related proteins, cytochrome C oxidase subunit CcoN, superfamily II DNA/RNA helicases (SNF2 family), primosomal protein N' (replication factor Y) superfamily II helicase, β -ureidopropionase, fucose permease, gluconolactonase, α -L-fucosidase, malto-oligosyltrehalose

synthase, α -1,2-mannosidase, endo-1,4- β -xylanase, endoglucanase, choline-sulfatase, putrescine transport ATP-binding protein PotA, arylsulfatase, and sulfate permease. Of these, strain RA likely maintains the genes for sulfate permease and arylsulfatase to withstand the high sulfur and sulfate contents of the AH hot spring. Furthermore, the presence of these genes suggests that strain RA plays a role in the recycling of sulfur within the hot spring. Two cholesterol oxidase genes were also present in strain RA but not in closely related genera. Notably, cholesterol oxidases catalyze the conversion of cholesterol to cholest-4-en-3-one, which has been proposed as a potential drug candidate for treatment of amyotrophic lateral sclerosis (Bordet et al., 2007).

KEGG metabolic pathway analyses indicated that the starch and sucrose metabolism pathways of strain RA are similar to those of *Rhodothermus* and *Salinibacter* spp. According to these analyses, strain RA is capable of hydrolyzing starch to maltose, glucose, or trehalose and converting starch to glycogen. Meanwhile, all three genera encode the enzymes (EC 3.2.1.4 and 3.2.1.21) necessary for hydrolysis of cellulose to glucose (Figure S1).

Likewise, *R. marinus* DSM 4252^T, *S. ruber* DSM 13855^T, and strain RA each encode a type I system (TolC), a SecSRP pathway (SecD/F, SecY, SecE, YidC, FstY, Ffh), and a Tat pathway (TatC) for protein secretion. However, while strain RA also carries genes encoding type II (GspD, GspE, GspF) and type VI (VgrG, Hcp, IcmF, DotU, ClpV) secretion systems, virulence factor prediction analyses using the PATRIC database

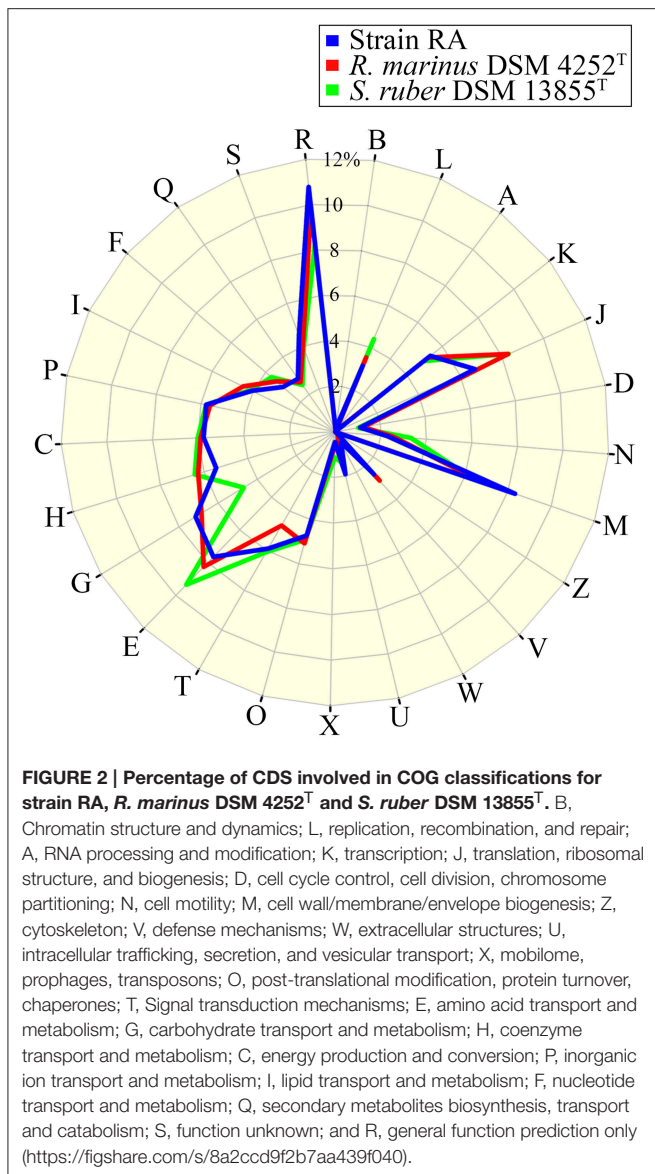
TABLE 3 | Genome statistics of strain RA.

| | Number | % of Total (%) |
|--|-----------|----------------|
| DNA, total number of bases | 4,616,094 | 100.00 |
| DNA coding number of bases | 4,160,472 | 90.13 |
| DNA G+C number of bases | 3,152,009 | 68.28 |
| DNA scaffolds | 61 | 100.00 |
| CRISPR count | 5 | |
| Genes total number | 3680 | 100.00 |
| Protein coding genes | 3630 | 98.64 |
| RNA genes | 50 | 1.36 |
| rRNA genes | 3 | 0.08 |
| 5S rRNA | 1 | 0.03 |
| 16S rRNA | 1 | 0.03 |
| 23S rRNA | 1 | 0.03 |
| tRNA genes | 44 | 1.20 |
| Other RNA genes | 3 | 0.08 |
| Protein coding genes with function prediction | 2815 | 76.49 |
| Without function prediction | 815 | 22.15 |
| Protein coding genes with enzymes | 897 | 24.38 |
| w/o enzymes but with candidate KO based enzymes | 50 | 1.36 |
| Protein coding genes connected to Transporter Classification | 365 | 9.92 |
| Protein coding genes connected to KEGG pathways3 | 912 | 24.78 |
| Not connected to KEGG pathways | 2718 | 73.86 |
| Protein coding genes connected to KEGG Orthology (KO) | 1587 | 43.12 |
| Not connected to KEGG Orthology (KO) | 2043 | 55.52 |
| Protein coding genes connected to MetaCyc pathways | 777 | 21.11 |
| Not connected to MetaCyc pathways | 2853 | 77.53 |
| Protein coding genes with COGs | 2222 | 60.38 |
| With KOGs | 713 | 19.38 |
| With Pfam | 2905 | 78.94 |
| With TIGRfam | 1144 | 31.09 |
| With InterPro | 1900 | 51.63 |
| With IMG Terms | 611 | 16.60 |
| With IMG Pathways | 231 | 6.28 |
| With IMG Parts List | 246 | 6.68 |
| In internal clusters | 734 | 19.95 |
| In chromosomal cassette | 3633 | 98.72 |
| Chromosomal cassettes | 309 | - |
| Biosynthetic clusters | 14 | - |
| Genes in biosynthetic clusters | 236 | 6.41 |
| Fused protein coding genes | 93 | 2.53 |
| Protein coding genes coding signal peptides | 527 | 14.32 |
| Protein coding genes coding transmembrane proteins | 881 | 23.94 |
| COG clusters | 1399 | 62.96 |
| KOG clusters | 485 | 21.83 |
| Pfam clusters | 1958 | 67.40 |
| TIGRfam clusters | 916 | 80.07 |

(Wattam et al., 2014), VirulenceFinder 1.5 (Joensen et al., 2014), and ResFinder 2.1 (Zankari et al., 2012) indicated that this strain is likely non-pathogenic.

Certain ABC transporters encoded by strain RA, including those specific for phosphate (PstS, PstA, PstB, PstC), lipoproteins (LolC, LolE, LolD), and iron-siderophores (FhuB, FhuC, FhuD),

are potentially produced by *R. marinus* DSM 4252^T and *S. ruber* DSM 13855^T. In addition, there were multiple transporters that were encoded in the genomes of strain RA and *R. marinus* DSM 4252^T but not in that of *S. ruber* DSM 13855^T, including transporters for molybdate (ModA, ModB), iron(III; AfuA, AfuB), zinc/manganese/iron (TroA,



TroB, TroC, TroD), raffinose/stachyose/melibiose (MsmK), sorbitol/mannitol (SmoK), cellobiose (MsiK), chitobiose (MsiK), arabinooligosaccharide (MsmX), and glucose/mannose (MalK; Table S2). Meanwhile, strain RA also encodes an RbsABC transporter for uptake of ribose/xylose. Lastly, strain RA contains genes encoding unique CusS, CusR, CusB, and CusA proteins that are predicted to play a role in copper and/or silver efflux, which are absent from all currently known *Rhodothermus* and *Salinibacter* spp.

Temperature Adaptation Genes of Strain RA

The average water temperature of the AH hot spring during sampling was 45°C; however, *in vitro* analysis demonstrated that strain RA is capable of thriving at temperatures up to 60°C. Meanwhile, analysis of the composition of the spring

water detected high levels of NaCl. We therefore endeavored to elucidate the mechanisms by which strain RA withstands thermal and osmotic stresses by characterizing the stress-related genes within the annotated genome.

The linear polyamines putrescine and spermidine have been associated with the thermophilicity of thermophiles (Takahashi and Kakehi, 2010; Goh et al., 2014). Notably, in addition to harboring the genes necessary to synthesize putrescine via *N*-carbamoylputrescine amidase, strain RA appears to be capable of taking up putrescine and spermidine from the environment using the PotA and PotC transporters. In contrast, the genes encoding PotA and PotC are not present in the genomes of *R. marinus* DSM 4252^T and *S. ruber* DSM 13855^T. A unique sequence in strain RA was annotated as a primosomal protein N' (replication factor Y) superfamily II helicase gene, and the corresponding protein sequence exhibited >58% sequence identity to PriA helicases deposited in GenBank. Several of these helicases have been proposed to play important roles in the thermostability of certain thermophiles (Goh et al., 2014). As mentioned earlier, the GC content of the strain RA genome is higher than that of most other genomes of *Rhodothermaceae*. In some cases, high genome and tRNA GC% levels are associated with increased optimal growth temperatures (Trivedi et al., 2005). Yet, the average tRNA GC content (61.4%) of strain RA is slightly lower than that of the rest of the genomes (68.3%). Additionally, the strain RA genome contains CDS that exhibit sequence similarity to genes encoding chaperones (DnaJ, DnaK, HtrA, metal chaperone) and the known heat shock proteins GroES and GroEL (HSP-60 family co-chaperones), and it contains three copies of the gene encoding the cold shock protein CspA. These CspA proteins exhibited the highest level of sequence identity to counterparts from *R. marinus* DSM 4252^T (67, 78, and 80%, respectively).

Osmotic Stress Adaptation Genes of Strain RA

We detected several genes encoding proteins involved in adaptation to osmotic stress within the strain RA genome, including Na⁺/H⁺ antiporters, a sodium/solute symporter, and a sodium/glutamate symporter (Table S3). Moreover, we detected three separate trehalose synthase genes (EC 5.4.99.16), which exhibited 64–65% identity to genes present in *S. longa* DSM 21114^T and *S. ruber* DSM 13855^T, while similar genes were absent from the *R. marinus* DSM 4252^T genome. Strain RA also harbors a gene predicted to encode malto-oligosyltrehalose trehalohydrolase (EC 5.3.2.141) that is 58% similar to a gene present in *S. ruber* DSM 13855^T. It also has a malto-oligosyltrehalose synthase (EC 5.4.99.15) gene that exhibits 58% sequence identity to a CDS present in *Trueperia radiovictrix* but has no homologs in the genomes of *S. longa* DSM 21114^T, *S. ruber* DSM 13855^T, and *R. marinus* DSM 4252^T. In addition, strain RA is predicted to be capable of biosynthesizing the osmoprotectant choline via a choline-sulfatase (EC 3.1.6.6). However, unlike *S. ruber* DSM 13855^T, both strain RA and *R. marinus* DSM 4252^T lack the genes encoding the osmoprotectant-associated transporters ProX, ProW, ProV, OpuBC, OpuBB, and OpuBA. Lastly, strain RA encodes the potassium uptake proteins TrkA

and TrkH, the osmotically inducible protein C, and the K⁺ channel histidine kinase KdpD but lacks genes encoding a glycine betaine transporter or proteins involved in ectoine transport/synthesis.

CONCLUSIONS

In this report, we have described the genome of strain RA, which was isolated from a saline hot spring in Malaysia. The draft genome of strain RA comprises 4,616,094 bp with a mean GC content of 68.3%. It contains 91 contigs with an N50 contig length of 152,113 bp. This genome is predicted to include 3630 protein-coding genes. At a 60% identity cutoff, a low percentage of these protein-coding genes are similar to genes in the phylum *Bacteroidetes*. The results of 16S rRNA gene sequencing, ANI-value and genome comparisons clearly indicate that this strain exhibits many differences from known genera of *Rhodothermaceae*. In the NCBI taxonomy database, the lineage for strain RA is classified as Bacteria; FCB group; Bacteroidetes/Chlorobi group; Bacteroidetes; Bacteroidetes Order II. Incertae sedis; *Rhodothermaceae*; unclassified *Rhodothermaceae*. In the near future, chemotaxonomic and phenotypic characterization of strain RA will be performed to further compare this strain with

other related type strains from the family *Rhodothermaceae*, and names for the genus and species will be proposed.

AUTHOR CONTRIBUTIONS

KG and KC designed the experiments and wrote manuscript; SL, KL, CC, RE, MS, and TA performed biochemical and sequencing experiments, and analyzed the genetic content of the bacterium.

ACKNOWLEDGMENTS

This work was supported by the University of Malaya via High Impact Research Grants [UM.C/625/1/HIR/MOHE/CHAN/01 (Grant No. A-000001-50001) and UM.C/625/1/HIR/MOHE/CHAN/14/1 (Grant No. H-50001-A000027)] awarded to KC. KG is grateful for funding received from Universiti Teknologi Malaysia GUP (Grant 09H98).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01109>

REFERENCES

- Alfredsson, G. A., Kristjansson, J. K., Hjörleifsdóttir, S., and Stetter, K. O. (1988). *Rhodothermus marinus*, gen. nov., sp. nov., a thermophilic, halophilic bacterium from submarine hot springs in Iceland. *Microbiology* 134, 299–306. doi: 10.1099/00221287-134-2-299
- Antón, J., Oren, A., Benlloch, S., Rodríguez-Valera, F., Amann, R., and Rosselló-Mora, R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the bacteria from saltern crystallizer ponds. *Int. J. Syst. Evol. Microbiol.* 52, 485–491. doi: 10.1099/00207713-52-2-485
- Besemer, J., and Borodovsky, M. (2005). GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* 33, W451–W454. doi: 10.1093/nar/gki487
- Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., et al. (2013). antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* 41, W204–W212. doi: 10.1093/nar/gkt449
- Blücher, A., Karlsson, E. N., and Holst, O. (2000). Substrate-dependent production and some properties of a thermostable, α -galactosidase from *Rhodothermus marinus*. *Biotechnol. Lett.* 22, 663–669. doi: 10.1023/A:1005627501609
- Bordet, T., Buisson, B., Michaud, M., Drouot, C., Galéa, P., Delaage, P., et al. (2007). Identification and characterization of cholest-4-en-3-one, oxime (TRO19622), a novel drug candidate for amyotrophic lateral sclerosis. *J. Pharmacol. Exp. Ther.* 322, 709–720. doi: 10.1124/jpet.107.123000
- Chai, Y. Y., Kahar, U. M., Salleh, M. M., Illias, R. M., and Goh, K. M. (2012). Isolation and characterization of pullulan-degrading *Anoxybacillus* species isolated from Malaysian hot springs. *Environ. Technol.* 33, 1231–1238. doi: 10.1080/09593330.2011.618935
- Cowan, S. T., and Steel, K. J. (1965). *Manual for the Identification of Medical Bacteria*. London: Cambridge University Press.
- Goh, K. M., Gan, H. M., Chan, K. G., Chan, G. F., Shahar, S., Chong, C. S., et al. (2014). Analysis of *Anoxybacillus* genomes from the aspects of lifestyle adaptations, prophage diversity, and carbohydrate metabolism. *PLoS ONE* 9:e90549. doi: 10.1371/journal.pone.0090549
- Gomes, I., Gomes, J., and Steiner, W. (2003). Highly thermostable amylase and pullulanase of the extreme thermophilic eubacterium *Rhodothermus marinus*: production and partial characterization. *Bioresour. Technol.* 90, 207–214. doi: 10.1016/S0960-8524(03)00110-X
- Gomes, J., Gomes, I., Terler, K., Gubala, N., Ditzelmüller, G., and Steiner, W. (2000). Optimisation of culture medium and conditions for α -L-arabinofuranosidase production by the extreme thermophilic eubacterium *Rhodothermus marinus*. *Enzyme Microb. Technol.* 27, 414–422. doi: 10.1016/S0141-0229(00)00229-5
- Hobel, C. F., Hreggvidsson, G. O., Marteinsson, V. T., Bahrani-Mougeot, F., Einarsson, J. M., and Kristjansson, J. K. (2005). Cloning, expression, and characterization of a highly thermostable family 18 chitinase from *Rhodothermus marinus*. *Extremophiles* 9, 53–64. doi: 10.1007/s00792-004-0422-3
- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., et al. (2012). Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* 62, 716–721. doi: 10.1099/ijls.0.038075-0
- Lane, D. J. (1991). “16S/23S rRNA sequencing,” in *Nucleic Acid Techniques in Bacterial Systematics*, eds E. Stackebrandt and M. Goodfellow (Chichester: Wiley), 115–117.
- Makhdomi-Kakhki, A., Amoozegar, M. A., and Ventosa, A. (2012). *Salinibacter iranicus* sp. nov. and *Salinibacter luteus* sp. nov., isolated from a salt lake, and emended descriptions of the genus *Salinibacter* and of *Salinibacter ruber*. *Int. J. Syst. Evol. Microbiol.* 62, 1521–1527. doi: 10.1099/ijls.0.031971-0
- Mao, C., Abraham, D., Wattam, A. R., Wilson, M. J., Shukla, M., Yoo, H. S., et al. (2015). Curation, integration and visualization of bacterial virulence factors in PATRIC. *Bioinformatics* 31, 252–258. doi: 10.1093/bioinformatics/btu631
- Markowitz, V. M., Mavromatis, K., Ivanova, N. N., Chen, I.-M. A., Chu, K., and Kyrpides, N. C. (2009). IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25, 2271–2278. doi: 10.1093/bioinformatics/btp393
- Marteinsson, V. T., Björnsdóttir, S. H., Bienvenu, N., Kristjansson, J. K., and Birrien, J. L. (2010). *Rhodothermus profundus* sp. nov., a thermophilic bacterium

- isolated from a deep-sea hydrothermal vent in the Pacific Ocean. *Int. J. Syst. Evol. Microbiol.* 60, 2729–2734. doi: 10.1099/ijs.0.012724-0
- Mongodin, E. F., Nelson, K., Daugherty, S., Deboy, R., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18147–18152. doi: 10.1073/pnas.0509073102
- Nolan, M., Tindall, B. J., Pomrenke, H., Lapidus, A., Copeland, A., Glavina Del Rio, T., et al. (2009). Complete genome sequence of *Rhodothermus marinus* type strain (R-10^T). *Stand. Genomic Sci.* 1, 283–291. doi: 10.4056/sigs.46736
- Ojima, T., Saburi, W., Sato, H., Yamamoto, T., Mori, H., and Matsui, H. (2011). Biochemical characterization of a thermophilic cellobiose 2-epimerase from a thermohalophilic bacterium, *Rhodothermus marinus* JCM9785. *Biosci. Biotechnol. Biochem.* 75, 2162–2168. doi: 10.1271/bbb.110456
- Okano, H., Ozaki, M., Kanaya, E., Kim, J.-J., Angkawidjaja, C., Koga, Y., et al. (2014). Structure and stability of metagenome-derived glycoside hydrolase family 12 cellulase (LC-CelA) a homolog of Cel12A from *Rhodothermus marinus*. *FEBS Open Bio* 4, 936–946. doi: 10.1016/j.fob.2014.10.013
- Oren, A. (2013). *Salinibacter*: an extremely halophilic bacterium with archaeal properties. *FEMS Microbiol. Lett.* 342, 1–9. doi: 10.1111/1574-6968.12094
- Oren, A., Abu-Ghosh, S., Argov, T., Kara-Ivanov, E., Shitrit, D., Volpert, A., et al. (2016). Expression and functioning of retinal-based proton pumps in a saltern crystallizer brine. *Extremophiles* 20, 69–77. doi: 10.1007/s00792-015-0798-2
- Park, S., Akira, Y., and Kogure, K. (2014). “The family Rhodothermaceae,” in *The Prokaryotes*, eds. E. Rosenberg, E. F. Delong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer-Verlag), 849–856.
- Park, S., Song, J., Yoshizawa, S., Choi, A., Cho, J. C., and Kogure, K. (2013). *Rubrivirga marina* gen. nov., sp. nov., a member of the family Rhodothermaceae isolated from deep seawater. *Int. J. Syst. Evol. Microbiol.* 63, 2229–2233. doi: 10.1099/ijs.0.046318-0
- Park, S., Yoshizawa, S., Kogure, K., and Yokota, A. (2011). *Rubricoccus marinus* gen. nov., sp. nov., of the family “Rhodothermaceae”, isolated from seawater. *Int. J. Syst. Evol. Microbiol.* 61, 2069–2072. doi: 10.1099/ijs.0.026294-0
- Peña, A., Teeling, H., Huerta-Cepas, J., Santos, F., Yáza, P., Brito-Echeverría, J., et al. (2010). Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting *Salinibacter ruber* strains. *ISME J.* 4, 882–895. doi: 10.1038/ismej.2010.6
- Peña, A., Valens, M., Santos, F., Buczolits, S., Antón, J., Kämpfer, P., et al. (2005). Intraspecific comparative analysis of the species *Salinibacter ruber*. *Extremophiles* 9, 151–161. doi: 10.1007/s00792-005-0430-y
- Roselló-Mora, R., Lucio, M., Peña, A., Brito-Echeverría, J., López-López, A., Valens-Vadell, M., et al. (2008). Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. *ISME J.* 2, 242–253. doi: 10.1038/ismej.2007.93
- Sako, Y., Takai, K., Ishida, Y., Uchida, A., and Katayama, Y. (1996). *Rhodothermus obamensis* sp. nov., a modern lineage of extremely thermophilic marine bacteria. *Int. J. Syst. Evol. Microbiol.* 46, 1099–1104. doi: 10.1099/00207713-46-4-1099
- Schaeffer, A. B., and Fulton, M. D. (1933). A simplified method of staining endospores. *Science* 77, 194–194. doi: 10.1126/science.77.1990.194
- Silva, Z., Horta, C., Da Costa, M. S., Chung, A. P., and Rainey, F. A. (2000). Polyphasic evidence for the reclassification of *Rhodothermus obamensis* Sako et al. 1996 as a member of the species *Rhodothermus marinus* Alfredsson et al. 1988. *Int. J. Syst. Evol. Microbiol.* 50, 1457–1461. doi: 10.1099/00207713-50-4-1457
- Singh, P., Anand, A., and Kumar, V. (2014). Recent developments in biological activities of chalcones: a mini review. *Eur. J. Med. Chem.* 85, 758–777. doi: 10.1016/j.ejmech.2014.08.033
- Takahashi, T., and Kakehi, J. I. (2010). Polyamines: ubiquitous polycations with unique roles in growth and stress responses. *Ann. Bot.* 105, 1–6. doi: 10.1093/aob/mcp259
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Trivedi, S., Rao, S. R., and Gehlot, H. S. (2005). Nucleic acid stability in thermophilic prokaryotes: a review. *J. Cell Mol. Biol.* 4, 61–69.
- Vaisman, N., and Oren, A. (2009). *Salisaeta longa* gen. nov., sp. nov., a red, halophilic member of the *Bacteroidetes*. *Int. J. Syst. Evol. Microbiol.* 59, 2571–2574. doi: 10.1099/ijs.0.010892-0
- Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 42, D581–D591. doi: 10.1093/nar/gkt1099
- Xia, J., Zhou, Y. X., Zhao, L. H., Chen, G. J., and Du, Z. J. (2015). *Longimonas halophila* gen. nov., sp. nov., isolated from a marine solar saltern. *Int. J. Syst. Evol. Microbiol.* 65, 2272–2276. doi: 10.1099/ijs.0.000247
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Goh, Chan, Lim, Liew, Chan, Shamsir, Ee and Adrian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome Analysis of *Thermosulfurimonas dismutans*, the First Thermophilic Sulfur-Disproportionating Bacterium of the Phylum *Thermodesulfobacteria*

Andrey V. Mardanov¹, Alexey V. Beletsky¹, Vitaly V. Kadnikov¹, Alexander I. Slobodkin² and Nikolai V. Ravin^{1*}

OPEN ACCESS

Edited by:

Kian Mau Goh,
Universiti Teknologi Malaysia,
Malaysia

Reviewed by:

Kai Waldemar Finster,
Aarhus University, Denmark
James F. Holden,
University of Massachusetts Amherst,
USA

*Correspondence:

Nikolai V. Ravin
nravin@biengi.ac.ru

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 05 May 2016

Accepted: 02 June 2016

Published: 17 June 2016

Citation:

Mardanov AV, Beletsky AV,
Kadnikov VV, Slobodkin AI and
Ravin NV (2016) Genome Analysis
of *Thermosulfurimonas dismutans*,
the First Thermophilic
Sulfur-Disproportionating Bacterium
of the Phylum *Thermodesulfobacteria*.
Front. Microbiol. 7:950.
doi: 10.3389/fmicb.2016.00950

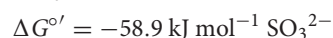
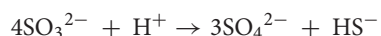
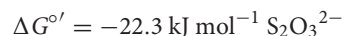
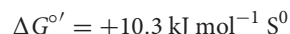
¹ Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia,
² Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia

Thermosulfurimonas dismutans S95^T, isolated from a deep-sea hydrothermal vent is the first bacterium of the phylum *Thermodesulfobacteria* reported to grow by the disproportionation of elemental sulfur, sulfite, or thiosulfate with carbon dioxide as the sole carbon source. In contrast to its phylogenetically close relatives, which are dissimilatory sulfate-reducers, *T. dismutans* is unable to grow by sulfate respiration. The features of this organism and its 2.1 Mb draft genome sequence are described in this report. Genome analysis revealed that the *T. dismutans* genome contains the set of genes for dissimilatory sulfate reduction including ATP sulfurylase, the AprA and B subunits of adenosine-5'-phosphosulfate reductase, and dissimilatory sulfite reductase. The oxidation of elemental sulfur to sulfite could be enabled by APS reductase-associated electron transfer complex QmoABC and heterodisulfide reductase. The genome also contains several membrane-linked molybdopterin oxidoreductases that are thought to be involved in sulfur metabolism as subunits of thiosulfate, polysulfide, or tetrathionate reductases. Nitrate could be used as an electron acceptor and reduced to ammonium, as indicated by the presence of periplasmic nitrate and nitrite reductases. Autotrophic carbon fixation is enabled by the Wood-Ljungdahl pathway, and the complete set of genes that is required for nitrogen fixation is also present in *T. dismutans*. Overall, our results provide genomic insights into energy and carbon metabolism of chemolithoautotrophic sulfur-disproportionating bacterium that could be important primary producer in microbial communities of deep-sea hydrothermal vents.

Keywords: sulfur disproportionation, thermophile, *Thermodesulfobacteria*, thiosulfate, genome sequence

INTRODUCTION

The biogeochemical sulfur cycle in the modern biosphere depends on activities of different anaerobic and aerobic microorganisms. One particular group of sulfur-metabolizing prokaryotes, i.e., the bacteria that disproportionate sulfur compounds, simultaneously perform the sulfur oxidation and reduction (Bak and Cypionka, 1987; Bak and Pfennig, 1987; Thamdrup et al., 1993). In this process, elemental sulfur, thiosulfate or sulfite each serves as both an electron donor and acceptor and became converted into sulfate and hydrogen sulfide:



The disproportionation of elemental sulfur is endergonic under standard conditions and proceeds only under low concentrations of hydrogen sulfide, which is achieved in the environment by the precipitation of sulfide with iron or by rapid oxidation. Disproportionation of inorganic sulfur compounds is of environmental significance in marine sediments (Jørgensen, 1990), and could be one of the earliest microbial processes dating back to 3.4 Ga (Philippot et al., 2007).

The process of disproportionation of inorganic sulfur compounds was described for about twenty species of the class *Deltaproteobacteria*, most of which are dissimilatory sulfate reducers (Finster, 2008). Among them, there are two thermophilic species, *Dissulfuribacter thermophilus* and *Dissulfurimicrobium hydrothermale* (Slobodkin et al., 2013, 2016). Outside *Deltaproteobacteria*, the ability to disproportionate sulfur compounds has been shown for three species of the phylum *Firmicutes* (genera *Desulfotomaculum* and *Dethiobacter*) and for the gamma-proteobacterium *Pantoea agglomerans* (Jackson and McInerney, 2000; Obraztsova et al., 2002; Nazina et al., 2005). Recently, the capacity for sulfur disproportionation has been reported for members of the phylum *Thermodesulfobacteria* – *Thermosulfurimonas dismutans* (Slobodkin et al., 2012) and *Caldimicrobium thiodismutans* (Kojima et al., 2016).

The metabolic pathways enabling disproportionation of thiosulfate and sulfite have been partly resolved in biochemical studies (Kramer and Cypionka, 1989; Frederiksen and Finster, 2003, 2004; Finster, 2008), but the enzymatic machinery of elemental sulfur disproportionation remains unclear. Complete genome sequences of several sulfur-disproportionating microorganisms are publically available; however, the analysis of genomic data in relation to mechanisms underlying the disproportionation of sulfur compounds has so far only been made for *Desulfocapsa sulfoexigens* (Finster et al., 2013).

Here, we present the results of sequencing and analysis of *Thermosulfurimonas dismutans* S95^T genome that provided

insights into the mechanisms of disproportionation of sulfur compounds. *T. dismutans*, isolated from deep-sea hydrothermal vent, is an anaerobic thermophilic bacterium which is able to grow chemolithoautotrophically by disproportionation of elemental sulfur, thiosulfate, and sulfite (Slobodkin et al., 2012). Unlike the majority of deltaproteobacterial sulfur disproportionators, *T. dismutans* is unable to respire sulfate. Elemental sulfur is abundant in some marine hydrothermal vents (Nakagawa et al., 2006), and its disproportionation by thermophilic bacteria could be an important process of primary production of organic matter in these ecosystems.

MATERIALS AND METHODS

Cultivation of *T. dismutans*

Thermosulfurimonas dismutans strain S95^T was isolated from a sample of the actively venting hydrothermal sulfidic chimney-like deposit located at the Mariner hydrothermal field (1910 m depth) on the Eastern Lau Spreading Center, Pacific Ocean and was maintained in the culture collection of the Laboratory of Hyperthermophilic Microbial Communities, Winogradsky Institute of Microbiology, Russian Academy of Sciences (Slobodkin et al., 2012). To obtain biomass for genome sequencing, the strain was grown in sealed bottles as previously described (Slobodkin et al., 2012) in anaerobic, bicarbonate-buffered marine liquid medium with 101 kPa of H₂:CO₂ (80%:20%) in the headspace and 10 mM Na₂S₂O₃ as the electron acceptor. The medium composition and preparation techniques were described previously (Slobodkin et al., 2012). The pH of the medium was 6.5–6.8 and the incubation temperature was 65°C. Cells were collected by centrifugation and then genomic DNA was isolated by SDS-CTAB method (Milligan, 1998). In order to determine if direct cell contact with sulfur is necessary for growth, *T. dismutans* was grown in the medium above except that thiosulfate was replaced with elemental sulfur entrapped into alginate beads. The technique for forming the sulfur alginate beads is as described previously (Gavrilov et al., 2012), except that a 1% (w/v) suspension of elemental sulfur (sublimed, Sigma) was used in place of ferrihydrite.

Genome Sequencing and Annotation

The *T. dismutans* S95^T genome was sequenced with a Roche Genome Sequencer (GS FLX), using the Titanium XL+ protocol for a shotgun genome library. The GS FLX run resulted in the generation of about 143 Mb of sequences with an average read length of 635 bp. The GS FLX reads were *de novo* assembled using Newbler Assembler version 2.9 (454 Life Sciences, Branford, CT, USA). The draft genome of *T. dismutans* S95^T consists of 61 contigs longer than 500 bp, with a total contig length of 2,119,932 bp.

Gene calling, annotation and analysis were performed for all contigs longer than 500 bp. Coding genes were annotated using the RAST server (Brettin et al., 2015). The annotation was manually corrected by searching the National Center for Biotechnology Information (NCBI) databases. The tRNAscan-SE tool (Lowe and Eddy, 1997) was used to find and annotate tRNA

genes, whereas ribosomal RNA genes were found by RNAMMER server (Lagesen et al., 2007). Signal peptides were predicted using Signal P v.4.1 for Gram-negative bacteria¹. The N-terminal twin-arginine translocation (Tat) signal peptides were predicted using PRED-TAT², the transmembrane helices – with TMHMM Server v. 2.0³.

For phylogenetic analysis of the catalytic A subunits of molybdopterine oxidoreductases the *T. dismutans* proteins TDIS_0362, TDIS_0614, TDIS_0652, TDIS_1010, and TDIS_1816 were used along with consensus sequences of the A subunits of tetrathionate reductases (Ttr), formate dehydrogenases (Fdh), thiosulfate or polysulfide reductases (Psr), and DMSO reductases (Dmsr), defined in Yanyushin et al. (2005). Amino acid sequences were aligned using MUSCLE (Edgar, 2004). Ambiguously aligned sites were removed using trimAl (Capella-Gutiérrez et al., 2009) before the phylogenetic reconstruction. The maximum likelihood phylogenetic tree was computed by PhyML 3.1 (Guindon et al., 2010), using the gamma model of rate heterogeneity (four discrete rate categories, an estimated alpha-parameter) and LG substitution matrix. The support values for the internal nodes were estimated by the approximate Bayesian method.

Nucleotide Sequence Accession Number

The annotated genome sequence of *T. dismutans* has been deposited in the GenBank database under accession no LWLG000000000.

RESULTS

General Features of the Genome

Sequencing and assembly of *T. dismutans* draft genome resulted in 61 contigs longer than 500 bp, with a total contig length of 2,119,932 bp. The G+C content of the genome is 50.1%. A single 16S-23S-5S rRNA operon and 48 tRNA genes coding for all of the 20 amino acids were identified.

Using a combination of coding potential prediction and similarity searches, 2159 protein-coding genes were predicted.

¹<http://www.cbs.dtu.dk/services/SignalP/>

²<http://www.compgen.org/tools/PRED-TAT/>

³<http://www.cbs.dtu.dk/services/TMHMM/>

TABLE 1 | General features of the genome.

| Attribute | Value |
|---------------------------------------|-----------|
| Size (bp) | 2,119,932 |
| G+C content (%) | 50.12 |
| Coding region (%) | 93.03 |
| Total genes | 2210 |
| rRNA genes | 3 |
| tRNA genes | 48 |
| Protein-coding genes, of them: | 2159 |
| Genes with predicted functions | 1458 |
| Genes assigned to COGs | 1443 |
| Genes with signal peptides | 86 |
| Genes specific to <i>T. dismutans</i> | 260 |

Of these, 1458 genes were functionally assigned with different degrees of generalization and confidence, while the function of the remaining 701 genes could not be predicted from the deduced amino acid sequences. The properties and the statistics of the genome are summarized in **Table 1**. Consistent with its affiliation to the phylum *Thermodesulfobacteria*, *T. dismutans* shares more than half of the proteome with that of its closest relative with sequenced genome, *Thermodesulfatator indicus* (1326 proteins).

Metabolism of Inorganic Sulfur Compounds

Although *T. dismutans* cannot grow by sulfate reduction, its genome contains the complete set of genes for dissimilatory sulfate reduction (Bradley et al., 2011; Pereira et al., 2011), including sulfate adenylyltransferase (TDIS_1516), manganese-dependent inorganic pyrophosphatase (TDIS_1154), APS reductase subunits AprA and AprB (TDIS_1513 and TDIS_1514), the subunits of dissimilatory sulfite reductase DsrABD (TDIS_1700, TDIS_1701, TDIS_1702), and distantly encoded DsrC (TDIS_1619). All of these predicted proteins lack signal peptides and transmembrane helices and were predicted to be located in the cytoplasm. The sulfate-reduction pathway could be linked to the membrane by sulfite reductase-associated electron transfer complex DsrMKJOP (TDIS_0546–TDIS_0542). Two of its subunits, DsrM and DsrP were predicted to contain transmembrane domains, while the iron-sulfur protein DsrO contains a Tat signal peptide enabling its translocation across the periplasmic membrane. A three-gene operon (TDIS_1512–TDIS_1510) encoding the subunits QmoA, QmoB, and QmoC of APS reductase-associated electron transfer complex QmoABC is located immediately downstream of the *aprBA* operon. The QmoA subunit contains a conserved FAD-binding site and the four cysteine cluster that binds an Fe–S center. QmoB contain FAD-binding site, 4Fe–4S double cluster binding domain and C-terminal domain similar to the delta subunit of methyl-viologen-reducing hydrogenase. The QmoC contains 4Fe–4S dicluster and transmembrane domain, thus linking the QmoABC complex to the cytoplasmic membrane. In sulfate reducing bacteria, QmoABC transfers electrons from the quinone pool to AprAB (Pires et al., 2003; Venceslau et al., 2010). Here, the electron transport could proceed in the opposite direction. This complex could also play a role in the oxidation of sulfur compounds to sulfite, as discussed below. The *T. dismutans* genome also encodes rodanese-like sulfurtransferase (TDIS_0247) that could participate in the thiosulfate and/or sulfur disproportionation, although the actual physiological role of this enzyme is unclear. Sulfur transport could be facilitated by sulfotransferase TDIS_0343, sulfur relay protein TusA (TDIS_0895) and integral membrane protein TDIS_0896.

Our additional physiological experiments revealed that *T. dismutans* is capable of sustained growth (at least four subsequent 5% v/v transfers) via disproportionation of elemental sulfur entrapped in alginate beads (a nominal molecular mass cutoff of 12 kDa). This finding indicates that the direct contact of the cells to solid elemental sulfur is not required for growth,

and the actual substrate for disproportionation is not the poorly soluble elemental sulfur, but most likely soluble polysulfides abiotically formed under these conditions.

Analysis of the *T. dismutans* genome revealed several membrane-linked oxidoreductases that could be involved in reduction of inorganic sulfur compounds. Among them there are four putative molybdopterin oxidoreductases of the Psr/Psh family. Such complexes typically consist of a molybdopterin-binding catalytic A subunit, an electron-transfer B subunit with an [Fe-S] cluster, and a membrane-anchor C subunit. Phylogenetic analysis of their catalytic A subunits (**Figure 1**) allowed to assign them the functions of tetrathionate reductase (TDIS_0362), formate dehydrogenase (TDIS_1010 and TDIS_1816), and thiosulfate or polysulfide reductases (TDIS_0614 and TDIS_0652). All of these catalytic subunits, except for formate dehydrogenase TDIS_1010, contain N-terminal twin-arginine translocation (Tat) signal peptides, indicating that these oxidoreductases operate on the periplasmic side of the membrane. The presence of hypothetical thiosulfate reductase, capable of producing sulfide and sulfite from thiosulfate, could explain the ability of *T. dismutans* to grow by disproportionation of thiosulfate.

The reduction of tetrathionate could be also enabled by the function of octaheme *c*-type cytochrome tetrathionate reductase (TDIS_1882). This is presumably a periplasmic protein linked to the cytoplasmic membrane by nearby encoded cytochrome *b* subunit (TDIS_1883) containing multiple transmembrane helices. *In vitro* studies of octaheme tetrathionate reductase from *Shewanella oneidensis* suggested a multifunctional role for these enzymes able to catalyze the reduction of tetrathionate, nitrite and hydroxylamine (Atkinson et al., 2007). The Sox system operating in aerobic sulfur oxidizing bacteria is missing in *T. dismutans*.

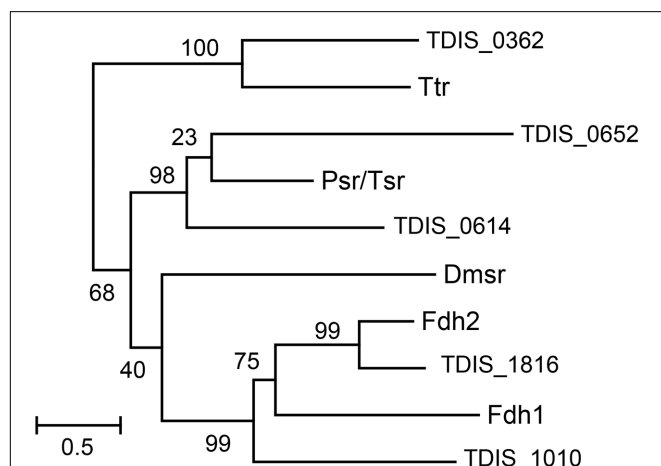


FIGURE 1 | Phylogenetic tree of the catalytic A subunits of molybdopterin oxidoreductases. Abbreviations: Ttr, tetrathionate reductases; Psr/Tsr, polysulfide and thiosulfate reductases; Dmsr, DMSO reductase; Fdh1 and Fdh2, formate dehydrogenases. Numbers at nodes represent the support values estimated by an approximate Bayesian method. The scale bar represents the average number of substitutions per site.

Alternative Electron Donors and Acceptors

Analysis of the *T. dismutans* genome revealed additional potential metabolic capabilities of this bacterium. The ability to use nitrate as an electron acceptor was suggested by the presence of an operon *napMADGH* (TDIS_0603- TDIS_0599) encoding periplasmic nitrate reductase. This complex includes the catalytic large subunit NapA, small tetraheme cytochrome *c* subunit NapM, electron transfer subunit NapG with 4Fe-4S double cluster binding domain, membrane anchor NapH and cytoplasmic chaperone NapD. The gene order is similar to that in the *napCMADGH* operon in sulfate and nitrate-reducing delta-proteobacterium *Desulfovibrio desulfuricans* (Marietou et al., 2005), although the small cytochrome NapC was not identified in *T. dismutans*. The nitrate reductase seems to localize in the periplasmic space, as indicated by the presence of N-terminal targeting sequences in NapM, NapA, and NapG subunits. Nitrite, produced from nitrate by this reductase could be further reduced to ammonium by dissimilatory nitrite reductase. This enzyme complex includes the multiheme membrane-bound cytochrome *c* subunit NrfH (TDIS_1141), the membrane subunit NrfD (TDIS_1142), the 4Fe-4S ferredoxin subunit NrfC (TDIS_1143), and the cytochrome *c* family protein with three heme motifs (TDIS_1144). The presence of N-terminal twin-arginine translocation (Tat) signal peptide in NrfC suggests the periplasmic location of nitrite reductase. We did not identify an apparent homolog of the catalytic NrfA subunit, but since the *nrf* operon is located at the end of a contig, this gene may be split and not found in the assembly. Alternatively, catalytic function could be performed by one of periplasmic multiheme cytochromes (e.g., TDIS_0311, TDIS_1112, and TDIS_1483) or the above-mentioned octaheme tetrathionate reductase.

The ability to reduce nitrate or nitrite was not reported in the original description of *T. dismutans* (Slobodkin et al., 2012), but the genome data prompted to reevaluate this trait. Indeed, our experiments have shown that *T. dismutans* is capable of growing with elemental sulfur as an electron donor and nitrate as an electron acceptor producing sulfate and ammonia (to be published elsewhere).

A four-gene cluster encodes two multiheme *c*-type cytochromes (TDIS_0609 and TDIS_0606), the iron-sulfur protein similar to B subunits of tetrathionate reductases (TDIS_0608) and the membrane anchor protein (TDIS_0607). The absence of N-terminal signal peptides in these proteins suggests that this oxidoreductase faces the cytoplasmic side of the inner membrane. The specificity of this complex could not be reliably predicted, but location of these genes close to the *nap* operon suggests that activity of this oxidoreductase could be coupled to nitrate reduction.

Two hydrogenases are encoded by the *T. dismutans* genome. An operon of genes TDIS_0913-TDIS_0915 encodes cytoplasmic methyl viologen-reducing hydrogenase MvhDGA. This enzyme, along with cytoplasmic CoB-CoM heterodisulfide reductase encoded by the nearby genes TDIS_0910-TDIS_0912 (*hdrCBA*) forms hydrogen:heterodisulfide oxidoreductase, which catalyzes the reduction of disulfide. In particular, the HdrB subunit

(TDIS_0911) contains two cysteine-rich domains with a 4Fe–4S cluster binding motif, involved in the reduction of disulfide bonds (Hamann et al., 2007).

The second hydrogenase, classified as group 1 NiFe enzyme (Vignais and Billoud, 2007), is a membrane-linked respiratory complex that could couple the oxidation of molecular hydrogen to the reduction of quinones and finally the terminal electron acceptor, probably, thiosulfate. The ability of *T. dismutans* to grow with molecular hydrogen as an electron donor and thiosulfate as an electron acceptor was reported in the original description (Slobodkin et al., 2012). The energy could be conserved in the form of a transmembrane proton gradient. This periplasmic complex is encoded close to the hydrogen:heterodisulfide oxidoreductase and includes the small subunit (TDIS_0916) carrying the Tat signal peptide, the large subunit (TDIS_0917) and membrane-bound cytochrome *b* subunit (TDIS_0918) transferring the electrons to the quinone pool.

The genome of *T. dismutans* suggests that formate could be used as an electron donor similar to hydrogen. Formate dehydrogenase of the molybdopterin oxidoreductase family is encoded by the genes TDIS_1816 (catalytic subunit A), TDIS_1818 (iron-sulfur subunit B) and TDIS_1819 (membrane anchor cytochrome *b* subunit C). The presence of an N-terminal Tat signal peptide in FdhA suggests that this formate dehydrogenase is oriented toward the periplasmic site of the internal membrane. However, in the series of additional experiments, we could not demonstrate the ability of *T. dismutans* to grow with formate as an electron donor and elemental sulfur, sulfate, thiosulfate, or nitrate as an electron acceptor.

Another component of the electron transfer chain is NADH-ubiquinone oxidoreductase consisting of the subunits NuoA, B, C, D, H, I, J, K, L, M, and N, encoded by the genes TDIS_1025–TDIS_1014. Genes encoding the subunits NuoEFG were not found in the genome indicating that NADH is likely not an electron donor for this complex. The presence of antiporter subunits suggests that the activity of this complex probably contributes to the generation of a transmembrane proton gradient that could be used by F₁F₀ ATP synthase for ATP production.

Central Metabolic Pathways

The *T. dismutans* genome encodes the complete Embden–Meyerhof pathway of glucose catabolism including glucokinase (TDIS_1571), glucose-6-phosphate isomerase (TDIS_1407), phosphofructokinase (TDIS_0184), fructose 1,6-bisphosphate aldolase (TDIS_0661), triosephosphate isomerase (TDIS_1029), glyceraldehyde-3 phosphate dehydrogenase (TDIS_2022, TDIS_2024, TDIS_2143), phosphoglycerate kinase (TDIS_2021), phosphoglycerate mutase (TDIS_0293), enolase (TDIS_1673) and pyruvate kinase (TDIS_1208). Taking into account that *T. dismutans* is unable to ferment sugars (Slobodkin et al., 2012), the glycolysis pathway probably operates in the reverse direction of gluconeogenesis. Consistently, the enzymes specifically catalyzing the reverse reactions are encoded: phosphoenolpyruvate synthase (TDIS_0764) and

fructose-1,6-bisphosphatase (TDIS_0690). The reversible conversion of pyruvate to acetyl-CoA could be performed by pyruvate:ferredoxin oxidoreductase encoded by the genes TDIS_0147–TDIS_0150.

Consistent with the inability of *T. dismutans* to use organic substrates as electron donors, its genomes do not encode the complete tricarboxylic acid cycle, as evidenced by the lack of genes for citrate synthase, succinyl CoA synthetase and succinate dehydrogenase.

Thermosulfurimonas dismutans is able to grow autotrophically without organic carbon sources (Slobodkin et al., 2012). Similarly to *D. sulfoexigens*, *T. dismutans* genome encodes a complete Wood–Ljungdahl (the acetyl-CoA reductive) pathway for the fixation of CO₂, including formate-tetrahydrofolate ligase (TDIS_0997), methylenetetrahydrofolate dehydrogenase/cyclohydrolase (TDIS_0998), methylenetetrahydrofolate reductase (TDIS_1006 and TDIS_0870), methyltetrahydrofolate:corrinoid iron–sulfur protein methyltransferase (TDIS_1009), and the CO dehydrogenase/acetyl CoA synthase complex (Ragsdale and Pierce, 2008). Formate dehydrogenase, the first enzyme of the methyl branch of this pathway is probably encoded by gene TDIS_1010. Contrary to the product of gene TDIS_1816, this formate dehydrogenase lack recognizable N-terminal targeting sequence and is probably located in the cytoplasm. The key enzymes of other known pathways of autotrophic carbon fixation, the reverse tricarboxylic acid cycle and the Calvin–Benson pathways were not identified.

Although the ability of *T. dismutans* to use N₂ gas as sole nitrogen source for growth was not analyzed at the original description (Slobodkin et al., 2012), its genome contains all genes necessary for nitrogen fixation, including the molybdenum–iron nitrogenase (genes TDIS_0750 and TDIS_0751 coding for subunits α and β , respectively), its regulatory and accessory proteins, all encoded in a single locus (genes TDIS_0746–TDIS_0754).

DISCUSSION

Thermosulfurimonas dismutans S95^T is the first known sulfur-disproportionating bacterium of the phylum *Thermodesulfobacteria*. Most representatives of this phylum are sulfate-reducing organisms (Zeikus et al., 1983; Moussard et al., 2004), with the exception of two species of the genus *Caldimicrobium* (Miroshnichenko et al., 2009; Kojima et al., 2016) and *Geothermobacterium ferrireducens* (Kashefi et al., 2002), which reduce thiosulfate and sulfur or Fe(III), respectively, and are incapable of dissimilatory sulfate reduction. *T. dismutans* also does not grow by sulfate respiration (Slobodkin et al., 2012).

Previous studies of enzymatic activities in deltaproteobacterium *Desulfocapsa sulfoexigens* indicate that sulfite is a key intermediate in the disproportionation of sulfur compounds (Frederiksen and Finster, 2003). The genome of *D. sulfoexigens* contains a full set of genes required for dissimilatory sulfate reduction and the reason why this bacterium does not respire sulfate remains unclear (Finster et al.,

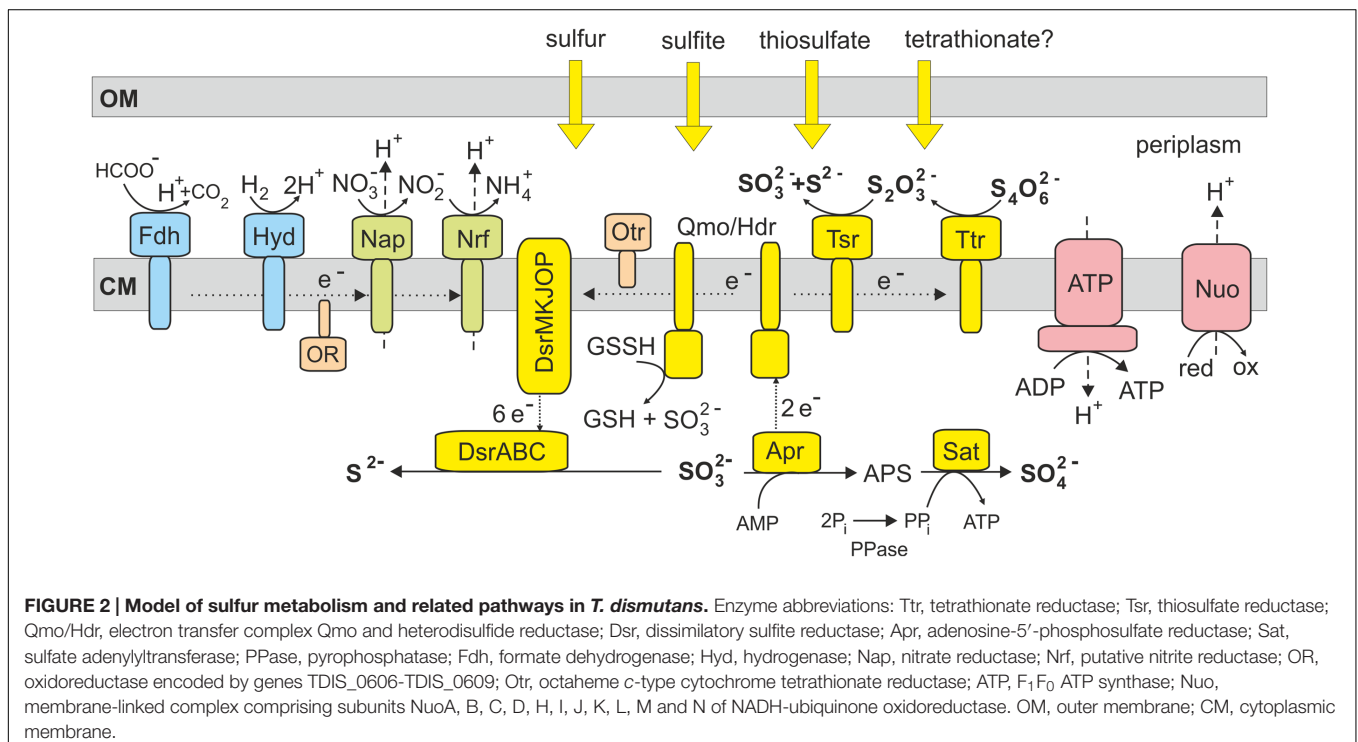
2013). Similar to *D. sulfoexigens*, the genome of *T. dismutans* also encodes the complete sulfate reduction pathway that may explain the ability to disproportionate sulfite. The oxidation of sulfite to sulfate could be enabled by reversal of the initial steps of the sulfate reduction pathway, performed by APS reductase and sulfate adenylyltransferase (Figure 2). These reverse reactions would result in ATP synthesis and the donation of electrons to the membrane quinone pool. An alternative hypothetical pathway of direct oxidation of sulfite to sulfate by sulfite oxidoreductase (Finster, 2008), found in some sulfur-oxidizing bacteria, seems to be absent in *T. dismutans* as well as in *D. sulfoexigens* (Finster et al., 2013). The reduction of sulfite to sulfide is likely enabled by the dissimilatory sulfite reductase and its accessory proteins, as in the typical sulfate reducers. Thus, *T. dismutans* makes ATP directly by substrate level phosphorylation and also with the aid of ATP synthetase consuming the proton-motive force generated by membrane-linked oxidoreductases.

Disproportionation of thiosulfate likely fits the same pathway with the addition of thiosulfate reductase that splits thiosulfate into sulfite and sulfide (Figure 2). The presence of tetrathionate reductase converting tetrathionate into thiosulfate suggests that *T. dismutans* could also grow by the disproportionation of tetrathionate, although this has not yet been studied.

To date, there is no conclusive information on the enzymatic pathways of elemental sulfur disproportionation. It is supposed that sulfite is an intermediate, although the corresponding enzymes(s) performing reactions with elemental sulfur itself were not identified (Frederiksen and Finster, 2003; Finster, 2008). The candidate genes were also not found in the *D. sulfoexigens* genome leading to the suggestion that the oxidation of elemental

sulfur to sulfite could depend on the activity of the adenylylsulfate reductase-associated electron transfer complex (Qmo) consisting of subunits A, B, and C, related to the subunits A and E of heterodisulfide reductase (Finster et al., 2013). In sulfate-reducing microorganisms, heterodisulfide reductase catalyzes the reversible reduction of disulfide bonds coupled to the generation of a proton motive force (Mander et al., 2004). Analysis of the *T. dismutans* genome revealed a similar *qmoABC* gene cluster (TDIS_1512- TDIS_1510). It was hypothesized that in the sulfur-oxidizing bacterium *Acidithiobacillus ferrooxidans* heterodisulfide reductase could oxidize disulfide intermediates to sulfite and donate electrons to the quinone pool (Quatrini et al., 2009). Taking into account that the heterodisulfide reductase catalytic site is actually located in HdrB, the involvement of cytoplasmic hydrogen:heterodisulfide oxidoreductase (*hdrCBA-mvhDGA*, genes TDIS_0910- TDIS_0915) in this reaction together with membrane-linked Qmo complex could be proposed. As in the case of *Acidithiobacillus* (Rohwerder and Sand, 2003), the actual substrate entering the disproportionation pathway in *T. dismutans* is probably not an elemental sulfur that is poorly soluble and cannot enter the cell, but soluble sulfane-sulfur compound glutathione persulfide (GSSH), which contains a disulfide bond that has been proposed to be cleaved by Qmo/Hdr to produce SO_3^{2-} and glutathione (GSH). Our observation that *T. dismutans* is able to grow via sulfur disproportionation without direct contact of the cells to solid elemental sulfur further supports this proposal.

Interestingly, *T. dismutans* and *D. sulfoexigens* have several common metabolic pathways besides those related to sulfur metabolism. Both bacteria can grow both autotrophically



and diazotrophically, which corresponds to the presence of a reverse acetyl-CoA pathway of CO₂ fixation and nitrogenase in their genomes. Both genomes suggest a potential for dissimilatory nitrate reduction coupled to elemental sulfur oxidation as an alternative or addition to sulfur-dependent metabolism thus linking sulfur and nitrogen cycles.

Overall, the genome sequence of *T. dismutans* provides new information about the metabolic pathways in this chemolithoautotrophic microorganism. *T. dismutans* was isolated from a chimney of a deep-sea hydrothermal vent where elemental sulfur is an abundant compound and thus bacterial sulfur disproportionation could represent an important process of primary production in such ecosystems. Genomic insights into energy and carbon metabolism of *T. dismutans* will stimulate and facilitate further biochemical and genetic studies required for the understanding of enzymatic pathways of microbial sulfur disproportionation.

REFERENCES

- Atkinson, S. J., Mowat, C. G., Reid, G. A., and Chapman, S. K. (2007). An octaheme c-type cytochrome from *Shewanella oneidensis* can reduce nitrite and hydroxylamine. *FEBS Lett.* 581, 3805–3808. doi: 10.1016/j.febslet.2007.07.005
- Bak, F., and Cypionka, H. (1987). A novel type of energy metabolism involving fermentation of inorganic sulphur compounds. *Nature* 326, 891–892. doi: 10.1038/326891a0
- Bak, F., and Pfennig, N. (1987). Chemolithotrophic growth of *Desulfovibrio sulfodismutans* sp. nov. by disproportionation of inorganic sulfur compounds. *Arch. Microbiol.* 147, 184–189. doi: 10.1007/BF00415282
- Bradley, A. S., Leavitt, W. D., and Johnston, D. T. (2011). Revisiting the dissimilatory sulfate reduction pathway. *Geobiology* 9, 446–457. doi: 10.1111/j.1472-4669.2011.00292.x
- Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5, 8365. doi: 10.1038/srep08365
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Finster, K. (2008). Microbiological disproportionation of inorganic sulfur compounds. *J. Sulfur Chem.* 29, 281–292. doi: 10.1080/17415990802105770
- Finster, K. W., Kjeldsen, K. U., Kube, M., Reinhardt, R., Musmann, M., Amann, R., et al. (2013). Complete genome sequence of *Desulfocapsa sulfexigens*, a marine deltaproteobacterium specialized in disproportionating inorganic sulfur compounds. *Stand. Genomic Sci.* 8, 58–68. doi: 10.4056/signs.777412
- Frederiksen, T. M., and Finster, K. (2003). Sulfite-oxido-reductase is involved in the oxidation of sulfite in *Desulfocapsa sulfexigens* during disproportionation of thiosulfate and elemental sulfur. *Biodegradation* 14, 189–198. doi: 10.1023/A:1024255830925
- Frederiksen, T. M., and Finster, K. (2004). The transformation of inorganic sulfur compounds and the assimilation of organic and inorganic carbon by the sulfur disproportionating bacterium *Desulfocapsa sulfexigens*. *Antonie Van Leeuwenhoek* 85, 141–149. doi: 10.1023/B:ANTO.0000020153.82679.f4
- Gavrilov, S. N., Lloyd, J. R., Kostrikin, N. A., and Slobodkin, A. I. (2012). Physiological mechanisms for dissimilatory reduction of poorly crystalline Fe(III) oxide by a thermophilic gram-positive bacterium *Carboxydotherrmus ferrireducens*. *Geomicrobiol. J.* 29, 804–819. doi: 10.1080/01490451.2011.635755
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

AUTHOR CONTRIBUTIONS

AM and NR designed the research project and wrote the paper; VK and AS performed the research; AB, AM, and NR analyzed the data.

ACKNOWLEDGMENTS

The work on the sequencing and analysis of *T. dismutans* genome was supported by the Russian Foundation for Basic Research (grant 13-04-40206 to AM) and by the program “Molecular and cellular biology” of the Russian Academy of Sciences. Microbiological studies of autotrophy and sulfur metabolism was supported by the Russian Science Foundation (grant 14-24-00165) and the Russian Foundation for Basic Research (grant 15-04-00405 to AS).

- Hamann, N., Mander, G. J., Shokes, J. E., Scott, R. A., Bennati, M., and Hedderich, R. (2007). A cysteine-rich CCG domain contains a novel [4Fe-4S] cluster binding motif as deduced from studies with subunit B of heterodisulfide reductase from *Methanothermobacter marburgensis*. *Biochemistry* 46, 12875–12885. doi: 10.1021/bi700679u
- Jackson, B. E., and McInerney, M. (2000). Thiosulfate disproportionation by *Desulfotomaculum thermoarcticum*. *Appl. Environ. Microbiol.* 66, 3650–3653. doi: 10.1128/AEM.66.8.3650-3653.2000
- Jørgensen, B. B. (1990). A thiosulfate shunt in the sulfur cycle of marine sediments. *Science* 249, 152–154. doi: 10.1126/science.249.4965.152
- Kashefi, K., Holmes, D. E., Reysenbach, A. L., and Lovley, D. R. (2002). Use of Fe(III) as an electron acceptor to recover previously uncultured hyperthermophiles: isolation and characterization of *Geothermobacterium ferrireducens* gen. nov., sp. nov. *Appl. Environ. Microbiol.* 68, 1735–1742. doi: 10.1128/AEM.68.4.1735-1742.2002
- Kojima, H., Umezawa, K., and Fukui, M. (2016). *Caldimicrobium thiodismutans* sp. nov., a sulfur-disproportionating bacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 66, 1828–1831. doi: 10.1099/ijsem.0.000947
- Kramer, M., and Cypionka, H. (1989). Sulfate formation via ATP sulfurylase in thiosulfate- and sulfite-disproportionating bacteria. *Arch. Microbiol.* 151, 232–237. doi: 10.1007/BF00413135
- Lagesen, K., Hallin, P. F., Rødland, E., Stærfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAMmer: consistent annotation of rRNA genes in genomic sequences. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955
- Mander, G. J., Pierik, A. J., Huber, H., and Hedderich, R. (2004). Two distinct heterodisulfide reductase-like enzymes in the sulfate-reducing archaeon *Archaeoglobus profundus*. *Eur. J. Biochem.* 271, 1106–1116.
- Marietou, A., Richardson, D., Cole, J., and Mohan, S. (2005). Nitrate reduction by *Desulfovibrio desulfuricans*: a periplasmic nitrate reductase system that lacks NapB, but includes a unique tetraheme c-type cytochrome, NapM. *FEMS Microbiol. Lett.* 248, 217–225. doi: 10.1016/j.femsle.2005.05.042
- Milligan, B. G. (1998). “Total DNA isolation,” in *Molecular Genetic Analysis of Population: A Practical Approach*, 2nd Edn, ed. A. R. Hoelzel (Oxford: Oxford University Press), 29–64.
- Miroshnichenko, M. L., Lebedinsky, A. V., Chernyh, N. A., Tourova, T. P., Kolganova, T. V., Spring, S., et al. (2009). *Caldimicrobium rimae* gen. nov., sp. nov., an extremely thermophilic, facultatively lithoautotrophic, anaerobic bacterium from the Uzon Caldera, Kamchatka. *Int. J. Syst. Evol. Microbiol.* 59, 1040–1044. doi: 10.1099/ijms.0.006072-0
- Moussard, H., L’Haridon, S., Tindall, B. J., Banta, A., Schumann, P., Stackebrandt, E., et al. (2004). *Thermodesulfatator indicus* gen. nov., sp. nov., a novel thermophilic chemolithoautotrophic sulfate-reducing bacterium

- isolated from the Central Indian Ridge. *Int. J. Syst. Evol. Microbiol.* 54, 227–233. doi: 10.1099/ijs.0.02669-0
- Nakagawa, T., Takai, K., Suzuki, Y., Hirayama, H., Konno, U., Tsunogai, U., et al. (2006). Geomicrobiological exploration and characterization of a novel deep-sea hydrothermal system at the TOTO caldera in the Mariana Volcanic Arc. *Environ. Microbiol.* 8, 37–49. doi: 10.1111/j.1462-2920.2005.00884.x
- Nazina, T. N., Rozanova, E. P., Belyakova, E. V., Lysenko, A. M., Poltarau, A. B., Tourova, T. P., et al. (2005). Description of “*Desulfotomaculum nigrificans* subsp. *salinus*” as a new species, *Desulfotomaculum salinum* sp. nov. *Microbiology* 74, 567–574.
- Obraztsova, A. Y., Francis, C. A., and Bradley, M. T. (2002). Sulfur disproportionation by the facultative anaerobe *Pantoea agglomerans* SP1 as a mechanism for chromium (VI) reduction. *Geomicrobiol. J.* 19, 121–132. doi: 10.1080/014904502317246219
- Pereira, I. A. C., Ramos, A. R., Grein, F., Marques, M. C., da Silva, S. M., and Venceslau, S. S. (2011). A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. *Front. Microbiol.* 2:69. doi: 10.3389/fmicb.2011.00069
- Philippot, P., Van Zuilen, M., Lepot, K., Thomazo, C., Farquhar, J., and Van Kranendonk, M. J. (2007). Early archaean microorganisms preferred elemental sulfur, not sulfate. *Science* 317, 1534–1537. doi: 10.1126/science.1145861
- Pires, R. H., Lourenco, A. I., Morais, F., Teixeira, M., Xavier, A. V., Saraiva, L. M., et al. (2003). A novel membrane-bound respiratory complex from *Desulfovibrio desulfuricans* ATCC 27774. *Biochim. Biophys. Acta* 1605, 67–82. doi: 10.1016/S0005-2728(03)00065-3
- Quatrini, R., Appia-Ayme, C., Denis, Y., Jedlicki, E., Holmes, D., and Bonnefoy, V. (2009). Extending the models for iron and sulfur oxidation in the extreme acidophile *Acidithiobacillus ferrooxidans*. *BMC Genomics* 10:394. doi: 10.1186/1471-2164-10-394
- Ragsdale, S. W., and Pierce, E. (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO₂ fixation. *Biochim. Biophys. Acta* 1784, 1873–1898. doi: 10.1016/j.bbapap.2008.08.012
- Rohwerder, T., and Sand, W. (2003). The sulfane sulfur of persulfides is the actual substrate of the sulfur-oxidizing enzymes from *Acidithiobacillus* and *Acidiphilium* spp. *Microbiology* 149, 1699–1710.
- Slobodkin, A. I., Reysenbach, A. L., Slobodkina, G. B., Baslerov, R. V., Kostrikina, N. A., Wagner, I. D., et al. (2012). *Thermosulfurimonas dismutans* gen. nov., sp. nov., an extremely thermophilic sulfur-disproportionating bacterium from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 62, 2565–2571. doi: 10.1099/ijs.0.034397-0
- Slobodkin, A. I., Reysenbach, A.-L., Slobodkina, G. B., Kolganova, T. V., Kostrikina, N. A., and Bonch-Osmolovskaya, E. A. (2013). *Dissulfuribacter thermophilus* gen. nov., sp. nov. a novel thermophilic autotrophic sulfur-disproportionating deeply-branching delta-proteobacterium from a deep-sea hydrothermal vent of the Eastern Lau Spreading Center. *Int. J. Syst. Evol. Microbiol.* 63, 1967–1971. doi: 10.1099/ijs.0.046938-0
- Slobodkin, A. I., Slobodkina, G. B., Panteleva, A. N., Chernyh, N. A., Novikov, A. A., and Bonch-Osmolovskaya, E. A. (2016). *Dissulfurimicrobium hydrothermale* gen. nov., sp. nov., a thermophilic, autotrophic, sulfur-disproportionating deltaproteobacterium isolated from a hydrothermal pond of Uzon Caldera, Kamchatka. *Int. J. Syst. Evol. Microbiol.* 66, 1022–1026. doi: 10.1099/ijsem.0.000828
- Thamdrup, B., Finster, K., Hansen, J. W., and Bak, F. (1993). Bacterial disproportionation of elemental sulfur coupled to chemical reduction of iron or manganese. *Appl. Environ. Microbiol.* 59, 101–108.
- Venceslau, S. S., Lino, R. R., and Pereira, I. A. (2010). The Qrc membrane complex, related to the alternative complex III, is a menaquinone reductase involved in sulfate respiration. *J. Biol. Chem.* 285, 22774–22783. doi: 10.1074/jbc.M110.124305
- Vignais, P. M., and Billoud, B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* 107, 4206–4272. doi: 10.1021/cr050196r
- Yanyushin, M. F., del Rosario, M. C., Brune, D. C., and Blankenship, R. E. (2005). New class of bacterial membrane oxidoreductases. *Biochemistry* 44, 10037–10045. doi: 10.1021/bi047267l
- Zeikus, J. G., Dawson, M. A., Thompson, T. E., Ingvorsent, K., and Hatchikian, E. C. (1983). Microbial ecology of volcanic sulphidogenesis: isolation and characterization of *Thermodesulfobacterium commune* gen. nov. and sp. nov. *J. Gen. Microbiol.* 129, 1159–1169.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Mardanov, Beletsky, Kadnikov, Slobodkin and Ravin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome Sequencing of *Sulfolobus* sp. A20 from Costa Rica and Comparative Analyses of the Putative Pathways of Carbon, Nitrogen, and Sulfur Metabolism in Various *Sulfolobus* Strains

Xin Dai^{1,2}, Haina Wang^{1,2}, Zhenfeng Zhang¹, Kuan Li³, Xiaoling Zhang³, Marielos Mora-López⁴, Chengying Jiang¹, Chang Liu^{1,2}, Li Wang¹, Yaxin Zhu¹, Walter Hernández-Ascencio⁴, Zhiyang Dong¹ and Li Huang^{1,2*}

OPEN ACCESS

Edited by:

Kian Mau Goh,
Universiti Teknologi Malaysia, Malaysia

Reviewed by:

Yutaka Kawarabayashi,
Kyushu University, Japan
Peter Redder,
Paul Sabatier University, France

*Correspondence:

Li Huang
huangl@sun.im.ac.cn

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 20 August 2016

Accepted: 14 November 2016

Published: 30 November 2016

Citation:

Dai X, Wang H, Zhang Z, Li K, Zhang X, Mora-López M, Jiang C, Liu C, Wang L, Zhu Y, Hernández-Ascencio W, Dong Z and Huang L (2016) Genome Sequencing of *Sulfolobus* sp. A20 from Costa Rica and Comparative Analyses of the Putative Pathways of Carbon, Nitrogen, and Sulfur Metabolism in Various *Sulfolobus* Strains. *Front. Microbiol.* 7:1902. doi: 10.3389/fmicb.2016.01902

¹ State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, ² College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, ³ State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, ⁴ Center for Research in Cell and Molecular Biology, Universidad de Costa Rica, San José, Costa Rica

The genome of *Sulfolobus* sp. A20 isolated from a hot spring in Costa Rica was sequenced. This circular genome of the strain is 2,688,317 bp in size and 34.8% in G+C content, and contains 2591 open reading frames (ORFs). Strain A20 shares ~95.6% identity at the 16S rRNA gene sequence level and <30% DNA-DNA hybridization (DDH) values with the most closely related known *Sulfolobus* species (i.e., *Sulfolobus islandicus* and *Sulfolobus solfataricus*), suggesting that it represents a novel *Sulfolobus* species. Comparison of the genome of strain A20 with those of the type strains of *S. solfataricus*, *Sulfolobus acidocaldarius*, *S. islandicus*, and *Sulfolobus tokodaii*, which were isolated from geographically separated areas, identified 1801 genes conserved among all *Sulfolobus* species analyzed (core genes). Comparative genome analyses show that central carbon metabolism in *Sulfolobus* is highly conserved, and enzymes involved in the Entner-Doudoroff pathway, the tricarboxylic acid cycle and the CO₂ fixation pathways are predominantly encoded by the core genes. All *Sulfolobus* species encode genes required for the conversion of ammonium into glutamate/glutamine. Some *Sulfolobus* strains have gained the ability to utilize additional nitrogen source such as nitrate (i.e., *S. islandicus* strain REY15A, LAL14/1, M14.25, and M16.27) or urea (i.e., *S. islandicus* HEV10/4, *S. tokodaii* strain7, and *S. metallicus* DSM 6482). The strategies for sulfur metabolism are most diverse and least understood. *S. tokodaii* encodes sulfur oxygenase/reductase (SOR), whereas both *S. islandicus* and *S. solfataricus* contain genes for sulfur reductase (SRE). However, neither SOR nor SRE genes exist in the genome of strain A20, raising the possibility that an unknown pathway for the utilization of elemental sulfur may be present in the strain. The ability of *Sulfolobus* to utilize nitrate

or sulfur is encoded by a gene cluster flanked by IS elements or their remnants. These clusters appear to have become fixed at a specific genomic site in some strains and lost in other strains during the course of evolution. The versatility in nitrogen and sulfur metabolism may represent adaptation of *Sulfolobus* to thriving in different habitats.

Keywords: *Sulfolobus*, strain A20, genome sequencing, comparative genomics, carbon metabolism, nitrogen metabolism, sulfur metabolism

INTRODUCTION

Archaea of genus *Sulfolobus* are widespread in solfataric fields around the globe. Known *Sulfolobus* species were mostly isolated from the Northern hemisphere (Brock et al., 1972; Grogan et al., 1990; Huber and Stetter, 1991; Jan et al., 1999; Suzuki et al., 2002; Xiang et al., 2003; Guo et al., 2011; Mao and Grogan, 2012; Zuo et al., 2015). These *Sulfolobus* isolates have been classified into nine species. Since *Sulfolobus* is readily grown and manipulated under laboratory conditions (Grogan, 1989), it has been used as a model for the study of Archaea. *Sulfolobus* also serves as a model for the study of eukaryotic genetic mechanisms because of the striking resemblance between Archaea and Eukarya in the flow of genetic information (Bell et al., 2002). In addition, *Sulfolobus* has been used as a host for the study of an increasing number of archaeal viruses and plasmids (Arnold et al., 2000; Rice et al., 2001; Xiang et al., 2003; Guo et al., 2011; Wang et al., 2015).

The complete genomes of 17 *Sulfolobus* strains belonging to four species have so far been deposited in GenBank. These include a *Sulfolobus tokodaii* strain (str.7) (Kawarabayasi et al., 2001), three *Sulfolobus solfataricus* strains (She et al., 2001; McCarthy et al., 2015), four *Sulfolobus acidocaldarius* strains (Chen et al., 2005; Mao and Grogan, 2012), and nine *Sulfolobus islandicus* strains (Reno et al., 2009; Guo et al., 2011; Zhang et al., 2013). Genomic comparisons show that *Sulfolobus* species are genetically diverged in relation to their geographic distance (Whitaker et al., 2003; Reno et al., 2009). Discontinuous and distantly separated habitats seem to be geographic barriers limiting gene flow among *Sulfolobus* populations. The variation in gene content among geographically diverse isolates is consistent with an isolation-by-distance model of diversification (Whitaker et al., 2003; Grogan et al., 2008; Reno et al., 2009). Apparently, genomic analyses of more geographically separated isolates would help shed more light on the genetic diversity and phylogenetic relationships of *Sulfolobus* strains.

All species of *Sulfolobus* are aerobic sulfur oxidizers, and many of them are initially described as autotrophs or mixotrophs (Brock et al., 1972). Two autotrophic carbon fixation cycles have been described in Crenarchaeota, i.e., the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and the dicarboxylate/4-hydroxybutyrate (DC/HB) cycle (Berg et al., 2007, 2010; Huber et al., 2008; Ramos-Vera et al., 2011). The HP/HB cycle was confirmed by biochemical assays in Sulfolobales including *Sulfolobus*, *Acidianus*, and *Metallosphaera* (Berg et al., 2007; Teufel et al., 2009; Estelmann et al., 2011; Demmer et al., 2013). H₂, hydrogen sulfide, sulfur, tetrathionate, and pyrite have been described as electron donors for autotrophically-grown *Sulfolobus* (Brock et al., 1972; Wood

et al., 1987; Huber and Stetter, 1991; Huber et al., 1992). For the heterotrophical growth of *Sulfolobus*, the conversion of glucose to pyruvate was thought to rely on a non-phosphorylative Entner-Doudoroff (ED) pathway, as shown in *S. solfataricus* and *S. acidocaldarius* (Siebers et al., 1997). However, extensive *in vivo* and *in vitro* assays later indicated that both the semi-phosphorylative and the non-phosphorylative ED pathways might operate in *S. solfataricus* (Ahmed et al., 2005; Ettema et al., 2008). Genomic analyses of the metabolic pathways have been reported for several *Sulfolobus* strains (Sensen et al., 1998; Kawarabayasi et al., 2001; She et al., 2001; Chen et al., 2005; Guo et al., 2011; Jaubert et al., 2013). A further genomic comparison of metabolic pathways in various *Sulfolobus* strains will be of significance to the understanding of the strategies of the organisms to adapt to thriving in their environments. In the present study, we isolated a novel *Sulfolobus* species, denoted strain *Sulfolobus* sp. A20, from an acidic hot spring in Laguna Fumarólica, Costa Rica, and sequenced the genome of the strain. The 16S rRNA gene of strain A20 exhibits the highest sequence identity (~95.6%) to those of *S. islandicus* and *S. solfataricus* isolates, but the significant differences suggest that strain A20 represents an independent *Sulfolobus* species. The genome of strain A20 was compared with all other available *Sulfolobus* genomes, and analyses of the pathways of carbon, nitrogen and sulfur metabolism in various *Sulfolobus* strains were performed.

MATERIALS AND METHODS

Isolation of Strain A20

A water sample FL1010-1 was collected in October 2010 from a hot spring, known as Laguna Fumarólica (10°46,365' N and 85°20,646' W, ~85°C, pH 3–4), in the Las Pailas hydrothermal field (Las Pailas sector), which is located in the southwest flank of the Rincón de la Vieja volcano crater. Rincón de la Vieja volcano (10°49' N, 85°19' W), an andesitic volcano in northwestern Costa Rica, belongs to the Circum Pacific Ring of Fire, which is a geothermal belt different from its nearest neighbors, the Yellowstone National Park and the Lassen Volcanic National Park. The sample was concentrated by tangential flow ultrafiltration through a hollow fiber membrane with a molecular mass cutoff of 6 kDa (Tianjin MOTIMO Membrane Technology, China). An enrichment culture was established by inoculating the concentrate in Zillig's medium (Zillig et al., 1994), which contained 0.3% (NH₄)₂SO₄, 0.05% KH₂PO₄·3H₂O, 0.05% MgSO₄·7H₂O, 0.01% KCl, 0.001% Ca(NO₃)₂·4H₂O, 0.07% Glycine, 0.05% yeast extract, 0.2% sucrose, and 0.2% of a trace element solution (0.09% MnCl₂·4H₂O, 0.225% Na₂B₄O₇·10H₂O, 0.011% ZnSO₄·7H₂O, 0.0025% CuCl₂·2H₂O,

0.0015% $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$, 0.0005% $\text{CoSO}_4 \cdot 7\text{H}_2\text{O}$). After incubation for 7–10 days at 75°C with shaking at 150 rpm, samples of the grown culture were spread on Zillig's medium plates solidified with 0.8% gelrite. The plates were incubated for 7 days at 75°C. Colonies were picked and purified by re-plating. Observation of the cells of strain A20 was carried out under a transmission electron microscope (JEM-1400, Jeol Ltd., Tokyo, Japan) at 80 kV by negatively staining with 2% uranyl acetate.

Genome Sequencing and Annotation

The genomic DNA of strain A20 was isolated and purified, as described (Chong, 2001), and sequenced on the Pacific Biosciences (PacBio) RS II and Illumina Hiseq 2000 systems at AnnoGenne, Beijing, China. The genome was assembled with SMRT analysis v2.3.0 and RS_HGAP_Assembly.3, and the genome assembly was improved by using the software Pilon (Walker et al., 2014). Identification of protein-coding open reading frames (ORFs) and annotation of the ORFs were performed by NCBI using the NCBI Prokaryotic Genome Annotation Pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Genes were functionally annotated by BLAST search in COG, KEGG, Nr, and Pfam Databases (Camacho et al., 2009; Finn et al., 2011). Putative insertion sequence (IS) elements were identified by BLASTn search against the IS finder Database (<http://www-is.biotoul.fr>).

Comparative Genomics Analysis

The nucleotide sequences of all genome-sequenced *Sulfolobus* strains and the corresponding amino acid sequences were retrieved from the GenBank database and the NCBI Reference Sequence database (RefSeq) (Table 1). The dot plots of any two genomes for their genomic synteny were profiled with Mummer (Kurtz et al., 2004), and DNA-DNA hybridization (DDH) values *in silico* were computed using the Genome-to-Genome Distance Calculator (GGDC) version 2.0 (Meier-Kolthoff et al., 2013) by submitting the genome sequences to DSMZ (<http://ggdc.dsmz.de>) (Auch et al., 2010). All protein sequences derived from the *Sulfolobus* genomes were compared using all-by-all BLASTp with a threshold *E*-value 10^{-10} , and grouped into orthologous gene families by OrthoMCL (Li et al., 2003). Gene groups consisting of orthologous genes present in all genomes, in more than two but not all genomes or in only one genome were defined as core, variable, or individual gene groups, respectively. A Venn diagram of the orthologous analysis of gene families was built with R version 3.0.2.

Phylogenetic Analysis

The 16S rRNA gene sequences of *Sulfolobus* species were extracted from the genome sequences and aligned using the CLUSTAL X program (Thompson et al., 1997). Phylogenetic trees were constructed using the neighbor-joining, maximum-parsimony, and maximum-likelihood methods implemented in the software package MEGA version 5.0 (Tamura et al., 2011). Evolutionary distances were calculated using Kimura's two-parameter model. The resulting tree topologies were evaluated by bootstrap analysis with 1000 re-samplings.

Metabolic Pathway Assignments

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Ogata et al., 1999; Kanehisa and Goto, 2000) was used in the analysis of the metabolic pathways of *Sulfolobus* species. All amino acid sequences derived from the genomes of *Sulfolobus* were submitted to the KEGG database, and the metabolic functions of these sequences were annotated by kass (Moriya et al., 2007). The KO (KEGG Orthology) term and corresponding KEGG pathway for each ORF were automatically generated and provided.

Sequencing Data Accession Number

The genome data of *Sulfolobus* sp. A20 have been deposited in the Genbank database under accession number CP017006.

RESULTS

General Features of *Sulfolobus* sp. A20

Sulfolobus sp. A20 was isolated from a hot spring in Costa Rica. The cells of strain A20 were irregular cocci (0.8–1.0 μm in diameter) with flagella (Figure 1). Growth occurred at temperatures between 65 and 85°C, and pH between 2.0 and 4.5. The strain grew optimally at 75–85°C and pH 4.0. The doubling time of the strain was ~14.3 h under the optimal growth conditions.

The Genome of *Sulfolobus* sp. A20

The genome of strain A20 was sequenced using a combination of PacBio RS II and Illumina Hiseq 2000 sequencing technologies with a 2×100 bp mode at a 150-fold and a 700-fold coverage, respectively. The genome consists of a single circular chromosome of 2,688,317 bp with 2591 ORFs, a single 16–23S rRNA cluster, a 5S rRNA gene, 45 tRNA genes and 5 miscellaneous RNA genes (misc RNAs). The average size of an ORF is ~291 amino acids. No extra-chromosomal genetic elements were detected in the strain. The G+C content of the genome is 34.78%. BLASTp searches identified matches in the protein database at GenBank for ~97.22% of the total ORFs of strain A20 (2519 ORFs). Among these ORFs, 2223 (~85.80% of total ORFs) are most closely related to those from the genus *Sulfolobus*, and 227 are closely related to those from other genera of the Sulfolobales. The general features of the strain A20 genome are compared with those of the other sequenced *Sulfolobus* genomes in Table 1.

Strain A20 encodes a complete set of enzymes and proteins involved in DNA transactions, including DNA replication, DNA repair and recombination, and RNA transcription. These proteins are highly conserved among the *Sulfolobus* strains, whose genomes have been sequenced, and share the highest sequence identity with those from *S. islandicus* or *S. solfataricus*. For example, DNA replication proteins, including ORC1-type DNA replication proteins (BFU36_RS04705, BFU36_RS02195, and BFU36_RS09865), mini-chromosome maintenance protein (MCM, BFU36_RS02210), primase subunits (BFU36_RS01270, BFU36_RS03220, and BFU36_RS03380), proliferating cell nuclear antigen subunits (PCNA, BFU36_RS01275, BFU36_RS03780, and BFU36_RS03820), replication factor

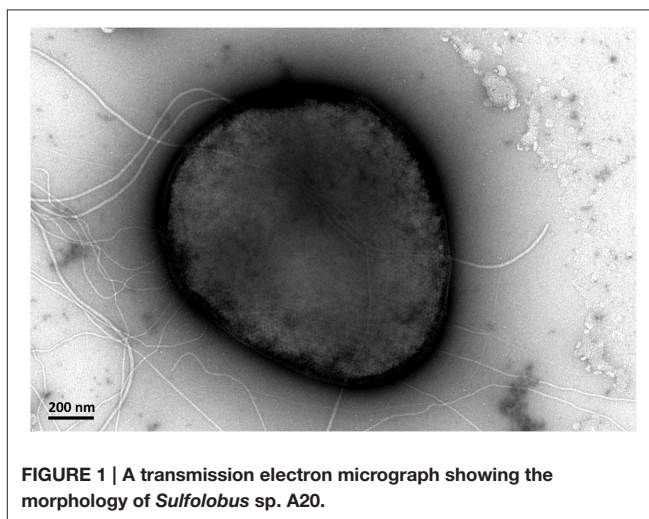
TABLE 1 | *Sulfolobus* strains with complete genome sequences.

| Strains | GenBank accession no. | NCBI RefSeq no. | Genome size (Mb) | No. of ORFs | No. of rRNAs | No. of tRNAs | GC% | Habitat |
|---------------------------------|-----------------------|-----------------|------------------|-------------|--------------|--------------|------|------------------------|
| <i>Sulfolobus</i> sp. A20 | CP017006 | NZ_CP017006 | 2.69 | 2591 | 3 | 45 | 34.8 | Las Palias, Costa Rica |
| <i>S. solfataricus</i> P2 | AE006641 | NC_002754 | 2.99 | 2896 | 3 | 45 | 35.8 | Naples, Italy |
| <i>S. solfataricus</i> P1 | LT549890 | NZ_LT549890 | 3.03 | 2967 | 3 | 45 | 35.8 | Naples, Italy |
| <i>S. solfataricus</i> 98/2 | CP001800 | NC_017274 | 2.67 | 2605 | 3 | 45 | 35.8 | Yellowstone, USA |
| <i>S. islandicus</i> REY15A | CP002425 | NC_017276 | 2.52 | 2535 | 3 | 46 | 35.3 | Reykjanes, Iceland |
| <i>S. islandicus</i> HVE10/4 | CP002426 | NC_017275 | 2.66 | 2692 | 3 | 44 | 35.1 | Hvergaardi, Iceland |
| <i>S. islandicus</i> LAL14/1 | CP003928 | NC_021058 | 2.47 | 2505 | 3 | 45 | 35.1 | Iceland |
| <i>S. islandicus</i> L.S.2.15 | CP001399 | NC_012589 | 2.74 | 2767 | 3 | 45 | 35.1 | Lassen, USA |
| <i>S. islandicus</i> 14.25 | CP001400 | NC_012588 | 2.61 | 2682 | 3 | 45 | 35.1 | Kamchatka, Russia |
| <i>S. islandicus</i> M16.4 | CP001402 | NC_012726 | 2.59 | 2678 | 3 | 45 | 35.0 | Kamchatka, Russia |
| <i>S. islandicus</i> M16.27 | CP001401 | NC_012632 | 2.69 | 2766 | 3 | 45 | 35.0 | Kamchatka, Russia |
| <i>S. islandicus</i> Y57.14 | CP001403 | NC_012622 | 2.7 | 2708 | 3 | 48 | 35.4 | Yellowstone, USA |
| <i>S. islandicus</i> YN15.51 | CP001404 | NC_012623 | 2.81 | 2791 | 3 | 46 | 35.3 | Yellowstone, USA |
| <i>S. acidocaldarius</i> DSM639 | CP000077 | NC_007181 | 2.23 | 2224 | 3 | 48 | 36.7 | Yellowstone, USA |
| <i>S. acidocaldarius</i> N8 | CP002817 | NC_020246 | 2.18 | 2188 | 3 | 48 | 36.7 | Hokkaido, Japan |
| <i>S. acidocaldarius</i> Ron121 | CP002818 | NC_020247 | 2.22 | 2227 | 3 | 30 | 36.7 | Ronneburg, Germany |
| <i>S. acidocaldarius</i> SUSAZ | CP006977 | NC_023069 | 2.06 | 2038 | 3 | 46 | 36.3 | Los Azufres, Mexico |
| <i>S. tokodaii</i> str.7 | BA000023 | NC_003106 | 2.69 | 2764 | 3 | 46 | 32.8 | Kyushu, Japan |

C (RFC, BFU36_RS02175, and BFU36_RS02180) and DNA polymerases (BFU36_RS05445, BFU36_RS13105, and BFU36_RS03245), from strain A20 closely resemble their homologs at the amino acid sequence level from the other *Sulfolobus* strains. Strain A20 also encodes small, basic and nucleic acid-binding proteins, i.e., Cren7 (BFU36_RS01545), two Sul7d proteins (BFU36_RS09545 and BFU36_RS11200), and two members of the Sac10b family (BFU36_RS01605 and BFU36_RS01615).

Like other *Sulfolobus* strains, strain A20 carries integrative elements, CRISPR-based immune systems and antitoxin/toxin systems (Guo et al., 2011). About 13 ORFs are annotated as the homologs of transposase, and nine copies of putative insertion sequence (IS) elements are found. Among these IS elements, eight belong to the IS200/605 family and one to the IS607 family. Six CRISPR loci of the two subtypes (I-A and III-B) and cmr1-6 proteins are identified (Grissa et al., 2007). No apparent sequence homology was detected between the spacers and the known sequences of *Sulfolobus/Acidianus* viruses. Five copies of family II (VapBC) antitoxin-toxin gene pairs are found in the strain A20 genome.

Dot plot analysis reveals no genomic synteny between strain A20 and any of the genome-sequenced *Sulfolobus* strains. Pairwise DNA-DNA hybridization (DDH) *in silico* between strain A20 and one of the tested *Sulfolobus* strains, including *S. tokodaii* str.7, *S. acidocaldarius* DSM 639, three *S. solfataricus* strains, and four *S. islandicus* strains, produces DDH values between 16.7 and 23.1% (Table 2), which are far below the 70% threshold proposed

**FIGURE 1 | A transmission electron micrograph showing the morphology of *Sulfolobus* sp. A20.**

for species definition (Tindall et al., 2009). These results suggest that strain A20 represents a novel *Sulfolobus* species.

Phylogenetic Analysis of *Sulfolobus* Strains

The 16S rRNA gene sequence of strain A20 was retrieved from the genome sequence of the strain. BLAST searches show that it is most similar (~95.6% identity) to those from several isolates of *S. islandicus* and *S. solfataricus*. The known *Sulfolobus* species appear to group into two main clades, as indicated by the phylogenetic analysis based on the 16S rRNA gene

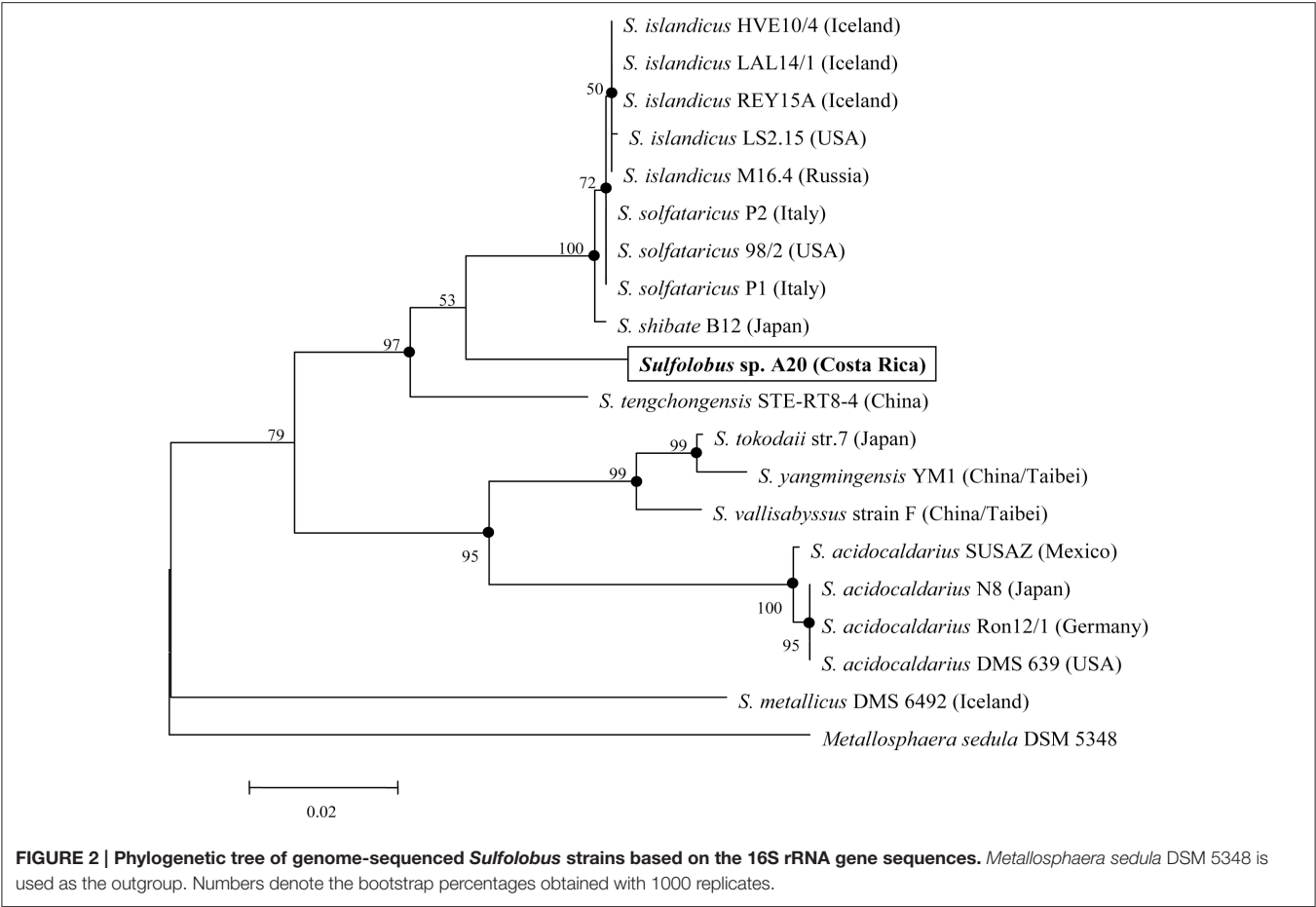


TABLE 2 | *In silico* DNA-DNA hybridization (DDH) values (%) between *Sulfolobus* strains^a.

| Sample | Strain | SSO | | | | SIS | | | | SAC | STO |
|-------------|--------|-----|-------|-------|-------|--------|--------|---------|---------|--------|-------|
| | | A20 | P1 | P2 | 98/2 | LS2.15 | REY15A | HVE10/4 | LAL14/1 | DSM639 | str7 |
| Strain A20 | – | | 16.80 | 16.80 | 16.80 | 16.70 | 16.70 | 16.70 | 16.70 | 23.10 | 19.60 |
| SSO-P1 | | | – | 94.80 | 91.50 | 38.00 | 37.50 | 37.30 | 37.40 | 18.80 | 24.00 |
| SSO-P2 | | | | – | 91.10 | 40.00 | 39.30 | 38.90 | 39.20 | 18.20 | 23.00 |
| SSO-98/2 | | | | | – | 37.20 | 37.00 | 36.90 | 36.90 | 18.90 | 24.30 |
| SIS-LS2.15 | | | | | | – | 85.80 | 87.80 | 81.70 | 18.40 | 20.00 |
| SIS-REY15A | | | | | | | – | 94.90 | 94.10 | 18.10 | 21.10 |
| SIS-HVE10/4 | | | | | | | | – | 94.00 | 18.20 | 21.70 |
| SIS-LAL14/1 | | | | | | | | | – | 18.10 | 21.20 |
| SAC-DSM639 | | | | | | | | | | – | 15.70 |
| STO-str7 | | | | | | | | | | | – |

^aSSO, *S. solfataricus*; SAC, *S. acidocaldarius*; SIS, *S. islandicus*; STO, *S. tokodaii*.

sequences (Figure 2). Strain A20, together with *S. islandicus*, *S. solfataricus*, *S. shibatae*, and *S. tengchongensis*, comprise one clade, while *S. acidocaldarius*, *S. tokodaii*, *S. vallisabyssus*, and *S. yangmingensis* make up the other. *S. metallicus* DSM6482, a strictly chemolithoautotrophic and ore-leaching *Sulfolobus* species, appears to be phylogenetically distant from the two main clades.

Core, Variable, and Individual Genes

A total of 18 *Sulfolobus* genomes, including the strain A20 genome, have been completely sequenced so far. To gain insight into the similarities and differences of the genomes from various *Sulfolobus* species, we compared the genome sequences available for the type strains of four *Sulfolobus* species, i.e., *S. acidocaldarius* DSM 639, *S. islandicus* REY15A, *S. solfataricus*

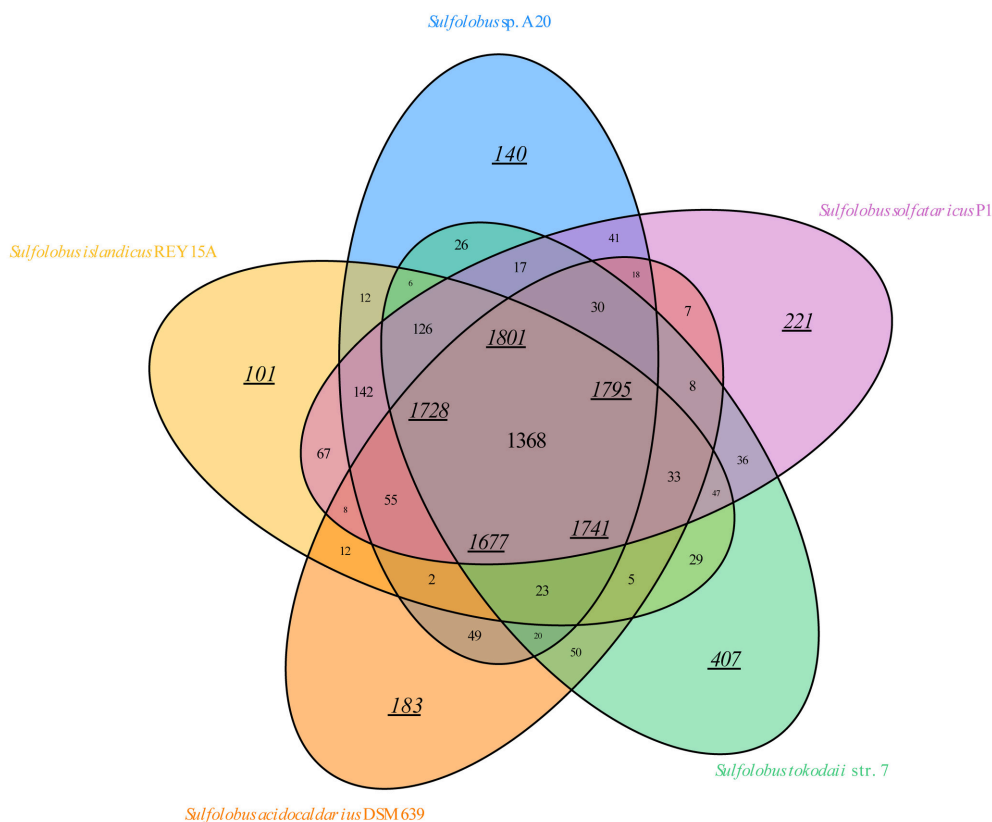


FIGURE 3 | Venn diagram of the conservation of protein-coding ORFs of the genome-sequenced type strains of *Sulfolobus* species. The overlaps between the ellipses show the gene groups shared by different strains with the number of shared gene groups indicated. The number of genes for each strain in a section of the diagram is shown by an underlined number in italics. Each underlined number in the middle of the diagram indicates the number of core genes for a strain analyzed.

P1, and *S. tokodaii* str.7 as well as strain A20. The numbers of predicted ORFs for the five genomes are 2663 ± 439 . The ORFs from these genomes are grouped into homologous groups. A total of 1368 gene groups form the core gene groups of the genus *Sulfolobus* (Figure 3). This number corresponds to 1801 genes (~69.51% of the total genes) in strain A20 (Table S1). Notably, the difference between these two numbers (i.e., 1368 gene groups vs. 1801 genes) is greater in strain A20 than in other *Sulfolobus* strains analyzed in this study, suggesting greater gene redundancy in A20 than in the other strains. Eight hundred and sixty nine gene groups are found in more than one, but not all, of the five genomes. These groups may constitute the variable parts of the *Sulfolobus* genomes. Strain A20 shares most gene groups with *S. solfataricus* P1 (1797), in agreement with their closest phylogenetic relationship. Moreover, the tested *Sulfolobus* genomes contain variable numbers of individual gene groups. In strain A20, 140 genes (~5.40% of the total ORFs) are not found in other four *Sulfolobus* strains. By comparison, *S. tokodaii* str.7 has the most individual genes (407, or ~14.72% of the total ORFs), whereas *S. islandicus* REY15A has the fewest individual genes (101, or ~3.98% of the total ORFs). Notably, the majority (>80%) of the individual genes encode hypothetical proteins. Conceivably, the exact numbers of core, variable and individual

genes in *Sulfolobus* strains will change as the sample size increases but the general pattern of the distribution of these three groups of genes will likely remain.

Metabolic Pathways

KEGG analyses reveal that the genome of strain A20 contains 84, 3 and 10 genes encoding functions in central carbon metabolism, nitrogen metabolism and sulfur metabolism, respectively. As compared to other known *Sulfolobus* genomes, the A20 genome appears to have similar numbers of the genes encoding proteins or protein subunits involved in carbon and sulfur metabolism but fewer genes for nitrogen metabolism. In addition, a total of 15 different ATP-binding cassette (ABC) transporters are identified in the strain A20 genome. By comparison, the numbers of ABC transporters are 10–14 in various *S. islandicus* strains (Guo et al., 2011), 11 in *S. solfataricus* P2 (She et al., 2001), 6 in *S. tokodaii* str.7 (Kawarabayasi et al., 2001), and 3 in *S. acidocaldarius* DSM639 (Chen et al., 2005). The ABC transporters in strain A20 include those for the transportation of trehalose (BFU36_RS00560–BFU36_RS00575, 4 ORFs in all), arabinogalactan oligomer/maltooligosaccharide (BFU36_RS00855–BFU36_RS00870, 4 ORFs), and glucose/arabinose (BFU36_RS07440–BFU36_RS07455, 4 ORFs, and

TABLE 3 | Enzymes involved in the Entner-Doudoroff pathway in strain A20.

| KO term | ORF(BFU36_RS) | Enzyme | Pathway ^a |
|---------|---------------|--|----------------------|
| K18125 | 06060 | Glucose dehydrogenase | ED |
| K05308 | 06085 | Gluconate dehydratase | ED |
| K18126 | 06095 | 2-keto-3-deoxygluconate kinase | sp ED |
| K11395 | 06090 | 2-keto-3-deoxy-6-phosphogluconate aldolase | sp ED |
| K18978 | 06100 | Glyceraldehyde-3-phosphate dehydrogenase | sp ED |
| K15634 | 10260 | | ED |
| K15635 | 03725 | Phosphoglycerate mutase | |
| K01689 | 02015 | Enoase | ED |
| K00873 | 01505 | Pyruvate kinase | ED |
| K11395 | 06090 | 2-keto-3-deoxygluconate aldolase | np ED |
| K18020 | 09145 | | np ED |
| K18021 | 09155 | Glyceraldehyde dehydrogenase | |
| K18022 | 09150 | | |
| K11529 | 02730 | Glycerate kinase | np ED |

^asp, semi-phosphorylative pathway; np, non-phosphorylative pathway.

BFU36_RS08120–BFU36_RS08130, 3 ORFs), suggesting the potential ability of strain A20 to utilize a wide range of sugars. There are 16 ORFs belonging to eight glycoside hydrolase (GHs) families, supporting the possibility that strain A20 uses a number of disaccharides and polysaccharides, e.g., cellobiose, maltotriose, mannan, and starch, for growth. A gene (BFU36_RS09315) encoding a putative trehalose glycosyl-transferring synthase (TreT) exists in the genome of strain A20. TreT from *Thermoproteus tenax* has been shown to catalyze trehalose synthesis from NDP-glucose or glucose (Kouril et al., 2008). Therefore, it is possible that strain A20 is capable of trehalose synthesis. There is also a cluster of four putative carotenoid biosynthetic genes (BFU36_RS07010–BFU36_RS07025), encoding homologs of lycopene cyclase, phytoene synthase, beta-carotene hydroxylase and phytoene desaturase, respectively, in the strain A20 genome, and these genes are arranged in the same manner as those in *S. solfataricus* (Hemmi et al., 2003) (Table S2).

Central Carbon Metabolism

As revealed by the genome analysis of *S. solfataricus* P2, strain A20 lacks the classical Embden–Meyerhof–Parnas (EMP) and pentose phosphate pathways, since the genes encoding the homologs of the key enzymes in these pathways, i.e., phosphofructokinase in the former and glucose-6-phosphate dehydrogenase, 6-phosphogluconolactonase and 6-phosphogluconate dehydrogenase in the latter, are missing from the genomes (She et al., 2001; Ulas et al., 2012). Like other genome-sequenced *Sulfolobus* strains, strain A20 may utilize glucose through either the semi-phosphorylative or the non-phosphorylative–Entner–Doudoroff (ED) pathway, or both (Table 3). Like all other *Sulfolobus* species, strain A20 contains all genes involved in the tricarboxylic acid (TCA) cycle, except

for those encoding the alpha-ketoglutarate dehydrogenase complex. The genes for the alpha-ketoglutarate dehydrogenase complex are replaced by those encoding the two subunits of 2-oxoacid:ferredoxin oxidoreductase, an enzyme catalyzing coenzyme A-dependent oxidative decarboxylation of 2-oxoacids (Zillig, 1991; Nishizawa et al., 2005). Intriguingly, the copy number of the genes for 2-oxoacid:ferredoxin oxidoreductase varies among *Sulfolobus* species. A single copy of the genes are present in strain A20, *S. solfataricus* and *S. islandicus*, whereas two copies of the genes are found in *S. acidocaldarius* and *S. tokodaii*, in apparent agreement with the phylogenetic relationship among these species (Figure 2).

All tested *Sulfolobus* strains are mixotrophs capable of growing chemolithotrophically on CO₂ with inorganic sulfur compounds (RISCs) as an energy source or heterotrophically on organic compounds (Brock et al., 1972; Keeling et al., 1998; Jan et al., 1999). Two CO₂ fixation pathways, i.e., the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and the dicarboxylate/4-hydroxybutyrate (DC/HB) cycle, have been reported to exist in (hyper)thermophilic autotrophic Crenarchaeota (Berg et al., 2010). Like the other 17 *Sulfolobus* genomes, the strain A20 genome contains all of the genes encoding homologs of the enzymes of the two cycles (Figure 4).

Nitrogen Metabolism

Like all other *Sulfolobus* genomes, the A20 genome contains genes encoding putative glutamate dehydrogenase (BFU36_RS08195), glutamine synthetase (BFU36_RS04000, BFU36_RS09525, and BFU36_RS10890) and the two subunits of carbamoylphosphate synthase (BFU36_RS02825 and BFU36_RS02830) (Table 4). It seems that all *Sulfolobus* strains employ a common strategy in the utilization of ammonium as a universal nitrogen source for the synthesis of glutamate, glutamine and carbamoylphosphate.

It is worth noting that four of the *S. islandicus* strains (i.e., REY15A, LAL14/1, M14.25, and M16.27) isolated from Iceland and Russia carry the *narGHJI* operon encoding a nitrate reductase and a nitrate transporter (*narK*) (Table 4), and, therefore, are potentially capable of utilizing nitrate. An operon encoding the subunits of urease (UreAB and UreC) and its accessory proteins (UreE, UreF, and UreG) is found in the genomes of *S. islandicus* HEV10/4, *S. tokodaii* str.7 and *S. metallicus* DSM 6482, suggesting that these strains are probably able to hydrolyze urea. Besides, genes for a putative cyanate lyase and a formamidase are found in the genomes of *S. tokodaii* str.7 and *S. islandicus* HEV10/4, respectively, suggesting a broader spectrum of nitrogen sources for these *Sulfolobus* strains.

Sulfur Metabolism

All sequenced *Sulfolobus* genomes contain a gene cluster (BFU36_RS07995–BFU36_RS08005 in strain A20) coding for sulfite reductase, phosphoadenosine phosphosulfate reductase, and sulfate adenylyltransferase (Tables 4, 5). These enzymes probably catalyze the conversion of hydrogen sulfide into

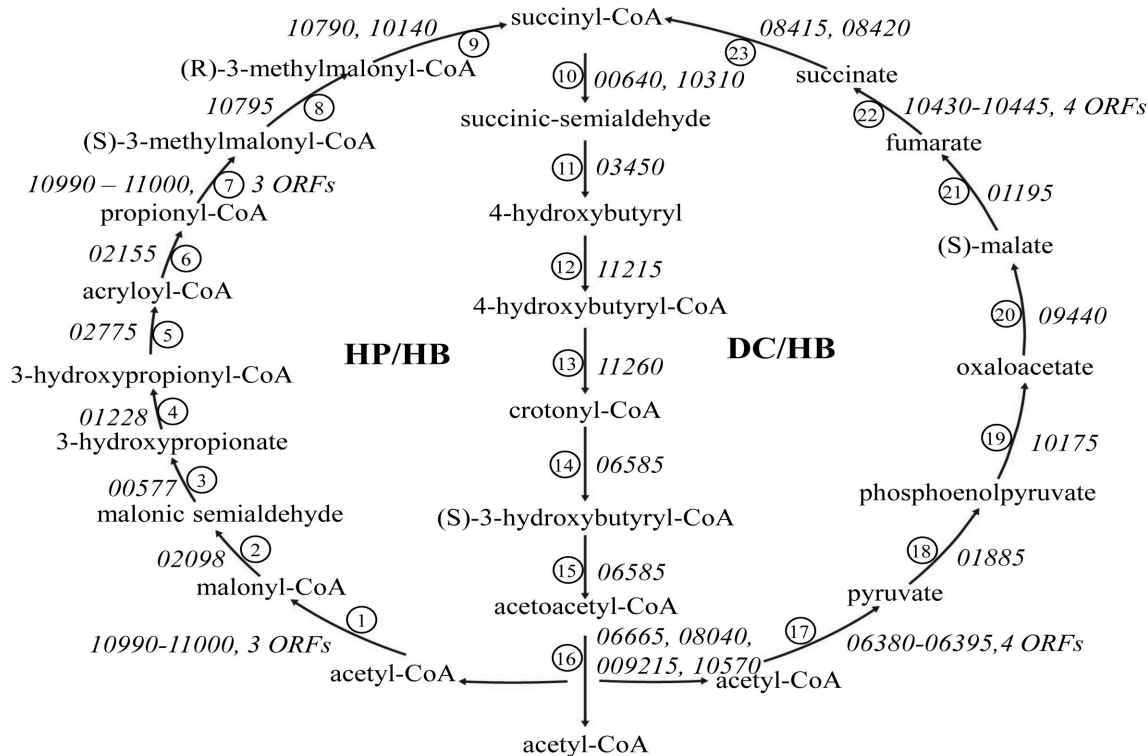


FIGURE 4 | The 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and the dicarboxylate/4-hydroxybutyrate (DC/HB) cycle in *Sulfolobus*.

Homologs of the enzymes in the two pathways in strain A20 are indicated by ORF numbers for the strain: ① acetyl-CoA carboxylase, ② malonyl-CoA reductase (NADPH), ③ malonate semialdehyde reductase (NADPH), ④ 3-hydroxypropionate-CoA ligase (AMP-forming), ⑤ 3-hydroxypropionyl-CoA dehydratase, ⑥ acryloyl-CoA reductase (NADPH), ⑦ propionyl-CoA carboxylase, ⑧ methylmalonyl-CoA epimerase, ⑨ methylmalonyl-CoA mutase, ⑩ succinyl-CoA reductase, ⑪ succinic semialdehyde reductase (NADPH), ⑫ 4-hydroxybutyrate-CoA ligase (AMP-forming), ⑬ 4-hydroxybutyryl-CoA dehydratase, ⑭ crotonyl-CoA hydratase, ⑮ (S)-3-hydroxybutyryl-CoA dehydrogenase (NAD⁺), ⑯ acetoacetyl-CoA beta-ketothiolase, ⑰ pyruvate synthase, ⑱ pyruvate:water dikinase, ⑲ PEP carboxylase, ⑳ malate dehydrogenase (NAD), ㉑ fumarate hydratase, ㉒ fumarate reductase, ㉓ succinyl-CoA synthetase (ADP-forming).

sulfite, and the subsequent transformation of sulfite into sulfate, with concomitant generation of ATP through substrate level phosphorylation (Kappler and Dahl, 2001; Rohwerder and Sand, 2007). A sulfide:quinone oxidoreductase (SQR) gene also exists in all *Sulfolobus* genomes (BFU36_RS09190 in strain A20). SQR may catalyze the oxidation of hydrogen sulfide into polysulfide (Rohwerder and Sand, 2007; Brito et al., 2009). Intriguingly, no homologs of sulfur oxygenase/reductase (SOR), a key enzyme for archaeal sulfur oxidation (Kletzin, 1992; Urich et al., 2006), are found in the genomes of *Sulfolobus* except for that of *S. tokodaii* str.7 (Kawarabayashi et al., 2001). The mechanism of elemental sulfur oxidation in *Sulfolobus* strains lacking SOR remains unknown. Putative genes for sulfur reductase (SRE) and thiosulfate:quinone oxidoreductase (TQO), which serve key roles in the reduction of elemental sulfur into hydrogen sulfide and the transformation of thiosulfate into tetrathionate, respectively (Laska et al., 2003; Guiral et al., 2005; Liu et al., 2012), are also found variably in *Sulfolobus* genomes (Tables 4, 5). Strain A20 and *S. tokodaii* str.7 carry *doxDA* (BFU36_RS07850–BFU36_RS07855), which encode a TQO homolog. *S. islandicus* and *S. solfataricus* have an SRE-encoding gene cluster (*sreABC*) and *doxDA*. *S. acidocaldarius* contains neither of the genes.

DISCUSSION

Sulfolobus sp. A20 was isolated from a hot spring in Costa Rica and the genomic DNA of the strain was completely sequenced. The addition of strain A20 to the growing list of the members of the genus *Sulfolobus* would aid further biogeographic comparison and evolutionary studies of this interesting group of archaea.

Sequence analysis indicates that strain A20 might be a mixotroph. The strain appears to be able to fix CO₂ via the HP/HB cycle. It is also capable of metabolizing glucose through a branched-ED pathway and the TCA cycle, as are other *Sulfolobus* strains. In general, genes involved in central carbon metabolism are conserved in all sequenced *Sulfolobus* genomes. Some of the genes may exist in different numbers of copies and/or be arranged differently among different species, and the differences are in apparent agreement with the phylogenetic relationship rather than the geographical separation of the species (Figure 2). It is of interest that genes encoding enzymes for CO₂ fixation through both HP/HB and DC/HB cycles are found in strain A20 and other sequenced *Sulfolobus* genomes. A similar finding has been reported for the genome of *Acidianus hospitalis* W1, a

TABLE 4 | Patterns of the distribution of genes encoding putative enzymes in nitrogen and sulfur metabolism in various *Sulfolobus* strains^a.

| KO terms | <i>S. solfataricus</i> | | | | <i>S. islandicus</i> | | | | | | | | <i>S. acidocaldarius</i> | | | | <i>S. tokodaii</i> str.7 | |
|----------------------------------|------------------------|----|----|------|----------------------|--------|---------|-------|--------|--------|---------|---------|--------------------------|--------|----|--------|--------------------------|-------|
| | A20 | P2 | P1 | 98/2 | HEV10/4 | REY15A | LAL14.1 | M16.4 | M14.25 | M16.27 | LS2.1.5 | YG57.14 | YN15.51 | DSM639 | N8 | Ron121 | | SUSAZ |
| NITROGEN METABOLISM ^b | | | | | | | | | | | | | | | | | | |
| K00261 (<i>gdhA</i>) | | | | | | | | | | | | | | | | | | |
| K01915 (<i>glnA</i>) | | | | | | | | | | | | | | | | | | |
| K01955 (<i>carB</i>) | | | | | | | | | | | | | | | | | | |
| K01956 (<i>carA</i>) | | | | | | | | | | | | | | | | | | |
| K00370 (<i>narG</i>) | | | | | | | | | | | | | | | | | | |
| K00371 (<i>narH</i>) | | | | | | | | | | | | | | | | | | |
| K00373 (<i>narJ</i>) | | | | | | | | | | | | | | | | | | |
| K00374 (<i>narI</i>) | | | | | | | | | | | | | | | | | | |
| K02575 (<i>narK</i>) | | | | | | | | | | | | | | | | | | |
| K01725 (<i>cynS</i>) | | | | | | | | | | | | | | | | | | |
| K01455 (<i>for</i>) | | | | | | | | | | | | | | | | | | |
| SULFUR METABOLISM ^c | | | | | | | | | | | | | | | | | | |
| K01011 (<i>tst</i>) | | | | | | | | | | | | | | | | | | |
| K00381 (<i>cysI</i>) | | | | | | | | | | | | | | | | | | |
| K00390 (<i>cysH</i>) | | | | | | | | | | | | | | | | | | |
| K00958 (<i>sat</i>) | | | | | | | | | | | | | | | | | | |
| K01738 (<i>cysK</i>) | | | | | | | | | | | | | | | | | | |
| K01739 (<i>metB</i>) | | | | | | | | | | | | | | | | | | |
| K17218 (<i>sqr</i>) | | | | | | | | | | | | | | | | | | |
| K16936 (<i>doxA</i>) | | | | | | | | | | | | | | | | | | |
| K16937 (<i>doxD</i>) | | | | | | | | | | | | | | | | | | |
| K16952 (<i>sor</i>) | | | | | | | | | | | | | | | | | | |
| K17219 (<i>sreA</i>) | | | | | | | | | | | | | | | | | | |
| K17220 (<i>sreB</i>) | | | | | | | | | | | | | | | | | | |
| K17221 (<i>sreC</i>) | | | | | | | | | | | | | | | | | | |

^a The presence of KO terms in nitrogen and sulfur metabolism is shown in gray.

^b Nitrogen metabolism: *gdhA*, glutamate dehydrogenase [NAD(P)⁺] [EC 1.4.1.3]; *glnA*, glutamine synthetase [EC 6.3.1.2]; *carB*, carbamoyl-phosphate synthase large subunit [EC 6.3.5.5]; *carA*, carbamoyl-phosphate synthase small subunit [EC 6.3.5.5]; *narG*, nitrate reductase/nitrite oxidoreductase, alpha subunit [EC 1.7.5.1; 1.7.99.4]; *narH*, nitrate reductase/nitrite oxidoreductase, beta subunit [EC 1.7.5.1; 1.7.99.4]; *narJ*, nitrate reductase delta subunit; *narI*, nitrate reductase gamma subunit [EC 1.7.5.1; 1.7.99.4]; *narK*, nitrate/nitrite transporter; *cynS*, cyanate lyase [EC 4.2.1.104]; *for*, formamidase [EC 3.5.1.49].

^c Sulfur metabolism: *tst*, thiosulfate sulfurtransferase [EC 2.8.1.1]; *cysI*, sulfite reductase (NADPH) hemoprotein [EC 1.8.1.2]; *cysH*, phosphoadenosine phosphosulfate reductase [EC 1.8.4.8; 1.8.4.10]; *sat*, sulfate adenylyltransferase [EC 2.7.7.4]; *cysK*, cysteine synthase A [EC 2.5.1.47]; *metB*, cystathionine gamma-synthase [EC 2.5.1.48]; *sqr*, sulfide:quinone oxidoreductase [EC 1.8.5.4]; *doxA*, thiosulfate dehydrogenase [quinone] small subunit [EC 1.8.5.2]; *doxD*, thiosulfate dehydrogenase [quinone] large subunit [EC 1.8.5.2]; *sor*, sulfur oxygenase/reductase [EC 1.13.11.55]; *sreA*, sulfur reductase molybdopterin subunit; *sreB*, sulfur reductase FeS subunit; *sreC*, sulfur reductase membrane anchor.

facultative anaerobe of the *Sulfolobales* (You et al., 2014). The two pathways differ in their sensitivity to oxygen, although they share many enzymes and intermediates in common (Ramos-Vera et al., 2011). The HP/HB cycle is more oxygen-tolerant than the DC/HB cycle since pyruvate synthase, one of key enzymes in the latter cycle, is oxygen sensitive (Jahn et al., 2007; Huber et al., 2008). As aerobes or microaerobes, members of the *Sulfolobales* have been shown to fix CO₂ through the HP/HB cycle. However, genes coding for putative pyruvate synthase, pyruvate:water dikinase and PEP carboxylase in the DC/HB cycle were found to be expressed, although at a low level, in *Metallosphaera sedula*, an aerobe closely related to *Sulfolobus*

strains (Berg et al., 2010). Therefore, we infer that the DC/HB pathway may also be employed by *Sulfolobus* to fix CO₂ under certain conditions.

Similarly, genes involved in the two ED pathways, i.e., the semi-phosphorylated pathway and the non-phosphorylated pathway, are also conserved in all the sequenced *Sulfolobus* genomes. The two ED pathways were named as the archaeal branched ED pathway (Sato and Atomi, 2011), and their functions were verified in *S. solfataricus* (Ahmed et al., 2005). The redundancy of the pathways for central carbon metabolism in *Sulfolobus* may contribute to the adaption of the organisms to thriving in the extreme and oligotrophic habitats.

TABLE 5 | Predicted reactions in sulfur metabolism in *Sulfolobus*^a.

| Strains | A20 | <i>S. acidocaldarius</i> | <i>S. solfataricus</i> | <i>S. islandicus</i> | <i>S. tokodaii</i> |
|--|-----|--------------------------|------------------------|----------------------|--------------------|
| thiosulfate + cyanide \rightleftharpoons sulfite + thiocyanate (<i>tst</i>) | + | + | + | + | + |
| hydrogen sulfide + 3NADP ⁺ + 3H ₂ O \rightleftharpoons sulfite + 3NADPH + 3H ⁺ (<i>sqr</i>) | + | + | + | + | + |
| PAP + sulfite + thioredoxin disulfide \rightleftharpoons PAPS + thioredoxin (<i>cysH</i>) | + | + | + | + | + |
| APS + diphosphate \rightleftharpoons sulfate + ATP (<i>sat</i>) | + | + | + | + | + |
| Sulfur \rightarrow hydrogen sulfide + thiosulfate (<i>sor</i>) | – | – | – | – | + |
| Thiosulfate + 6-decylubiquinone \rightleftharpoons tetrathionate + 6-decylubiquinol (<i>doxAD</i>) | + | – | + | + | + |
| Sulfur + hydrogen \rightleftharpoons hydrogen sulfide (<i>sre</i>) | – | – | + | + | – |

^aAPS, adenylyl sulfate; PAPS, 3'-phosphoadenylyl sulfate; PAP, adenosine 3', 5'-bisphosphate.

All *Sulfolobus* genomes contained a complete pathway for ammonium assimilation, which is similar to that found in heterotrophic bacteria (Zalkin, 1993; Guo et al., 2011; Wang et al., 2016), suggesting that *Sulfolobus* prefers to use ammonia as the nitrogen source. Strain A20 is probably unable to use other inorganic nitrogen sources for growth, while several of the *S. islandicus* strains and *S. tokodaii* str.7 might be able to use nitrate, urea, cyanate or formamide as their nitrogen source. These results point to the diversity of nitrogen utilization by *Sulfolobus*. It remains to be determined if the difference in the ability of *Sulfolobus* strains to use inorganic nitrogen compounds correlates with the availability of the nitrogen sources in the habitats of the strains.

Genomic analyses reveal the presence of transposase genes and repeating sequences near the *nar* gene cluster, suggesting the potential mobility of the cluster. The *nar* cluster was found at either of the two genomic sites in four *S. islandicus* strains containing the cluster. In the two *S. islandicus* strains from Iceland (i.e., REY15A and LAL14/1), the *nar* cluster resides on the complementary strand downstream of a sequence encoding a GntR family transcriptional regulator, a CoA ester lyase and an esterase (SIRE_RS02235–SIRE_RS02245 in REY15A and SIL_RS02325–SIL_RS02335 in LAL14/1). This site of potential *nar* insertion is termed insertion site A. On the other hand, in the two strains from Kamchatka (i.e., M16.27 and M14.25), the cluster is located downstream of a sequence encoding a 3-hydroxyacyl-CoA dehydrogenase, an AMP-dependence synthetase and an acetyl-CoA synthetase (M1627_RS04095–M1627_RS04105 in M16.27 and M1425_RS04080–M1425_RS04090 in M14.25). We denote this potential location for the insertion of the *nar* cluster insertion site B. Although only two strains were found to contain the *nar* cluster at insertion site A, this insertion site is present in all *Sulfolobus* strains analyzed in this study. Variation occurs downstream of the site. There are seven types of gene organization downstream of insertion site A in the 18 strains (Tables S3, S4). The tandem array of the three genes at insertion site B is found only in *S. islandicus* strains isolated from Kamchatka, Yellowstone National Park (YNP), and Lassen in USA (Tables S3, S4). Three general patterns of gene arrangement were identified at insertion site B. The two *S. islandicus* strains from USA (i.e., L.S 2.15 and Y57.14) are of one type, and the two Kamchatka *S. islandicus* strains (i.e., M16.27

and M14.25) belong to the other type. Remarkable variation in gene arrangement indicates that the two sites are where active transposition has taken place. The biogeographical difference in genomic location of the *nar* gene cluster presumably resulted from the transposition of the cluster. Since the presence of the *nar* cluster is restricted to *S. islandicus* and some of the strains in this species lack the gene cluster, we hypothesize that the species originally carried the cluster. When it spread to various geographical locations, loss or transposition of the gene cluster occurred, producing variants that thrive in various parts of the globe today. Whether the *nar* cluster was originally acquired through horizontal gene transfer is unclear. However, no significant difference in GC content between the gene cluster and the genome was detected.

Elemental sulfur metabolism is complex in *Sulfolobus*, and relatively low conservation in sulfur metabolism exists among the sequenced genomes. Strain A20 is likely capable of utilizing hydrogen sulfide because of the presence in its genome a conserved gene cluster for sulfur metabolism (Kawarabayasi et al., 2001; Chen et al., 2005). Although most *Sulfolobus* strains have been described as sulfur-oxidizing microbes (Brock et al., 1972), the biochemical process of elemental sulfur oxidation has yet to be fully understood. The *sor* gene encoding the classical sulfur oxygenase/reductase required for the initial step in the archaeal sulfur oxidation pathway (Urich et al., 2006) is present in none of the sequenced *Sulfolobus* genomes except for the genome of *S. tokodaii* str.7 (Kawarabayasi et al., 2001; She et al., 2001; Chen et al., 2005; Guo et al., 2011; Jaubert et al., 2013). Instead, there is a gene cluster encoding sulfur reductase (SRE), which reduces S⁰ with the help of a hydrogenase in anaerobically grown *Acidianus ambivalens* (Laska et al., 2003), in the genomes of *S. solfataricus* and *S. islandicus*. However, no hydrogenase genes have been identified in the two species. So, whether and how the sulfur reductase catalyzes sulfur reduction in the absence of a hydrogenase under aerobic conditions remains to be determined. It has been reported that *Sulfolobus tokodaii* str.7 grows poorly in the presence of elemental sulfur under the facultatively chemolithotrophic conditions (Suzuki et al., 2002), although it encodes a homolog of the classical sulfur oxygenase/reductase. However, the strain was able to oxidize hydrogen sulfide into sulfate (Kawarabayasi et al., 2001), suggesting the possibility of functional divergence of the homologs of sulfur oxygenase/reductase in *Sulfolobus*. Therefore,

further investigation is needed to understand the mechanisms underlining elemental sulfur metabolism in *Sulfolobus*.

The *sre* gene cluster is flanked upstream by a hypothetical protein and a 4Fe-4S ferredoxin and downstream by another 4Fe-4S ferredoxin and two hypothetical proteins (Tables S3, S4). This entire sequence is located downstream of a *cupin* gene. Based on the presence of genes between *cupin* and the *sre* cluster, three types of gene arrangement were identified at this site. A transposase gene is located between *cupin* and the *sre* cluster in *S. solfataricus* strains P1 and P2, both of which were isolated from Naples, Italy. However, no transposase gene at this site was found in *S. solfataricus* strain 98/2 or *S. islandicus* strains from YNP. Instead, a gene for the large subunit of nitric oxide reductase is present at this site in these strains. By comparison, a pseudogene is in the place of the transposase gene in *S. islandicus* strain 14.25 from Kamchatka. The two other *S. islandicus* strains (i.e., M16.4 and M16.27) from Kamchatka contain multiple transposase genes as well as hypothetical proteins at the site. Patterns of gene arrangement upstream of the *sre* gene cluster appear to carry distinct geographical markers, since they exhibit similarity among closely located strains of the same species. Whether the function of the *sre* gene cluster is affected by its genomic environment is unclear.

A putative *tusA-dsrE2-dsrE3A* gene cluster is linked to the *hdr* cluster (*hdrC1-hdrB1A-hyp-hdrC2-hdrB2*) in all *Sulfolobus* genomes. The *hdr* cluster encodes a heterodisulfide-reductase complex, which may be involved in sulfur transfer and reversible reduction of the disulfide bond X-S-S-X in *Acidithiobacillus ferrooxidans* (Quatrini et al., 2009; Liu et al., 2014), while the *tusA-dsrE2-dsrE3A* gene cluster may encode functions in the transformation of tetrathionate into thiosulfate in *Metallosphaera cuprina* (Liu et al., 2014). How the two genomically linked gene clusters function in sulfur metabolism remains to be understood.

Taken together, our genomic analyses reveal that these *Sulfolobus* species are conserved in central carbon metabolism,

but differ in the ability to use inorganic nitrogen and sulfur sources. The ability of *Sulfolobus* to utilize nitrate or sulfur is encoded by a gene cluster flanked by IS elements or their remnants. These clusters appear to have become fixed at a specific genomic site in some strains and lost in other strains during the course of evolution.

AUTHOR CONTRIBUTIONS

XD and LH designed the project. XD and ZZ analyzed the data. HW, LW, YZ, ZD, MM-L, and WH-A collected sample, purified the strain and prepared the genomic DNA for sequencing. KL and XZ performed bioinformatic analysis of the genome sequences. CJ and CL analyzed the pathways of sulfur metabolism. LH, XD and ZZ wrote the manuscript.

ACKNOWLEDGMENTS

We thank Drs. Hailiang Dong and Yong Tao for their valuable comments. This work was supported by National Natural Science Foundation of China grant 31130003. Sampling was partially supported by Grant VI 801-B0-530 from Vicerrectoría de Investigación, Universidad de Costa Rica (San José, Costa Rica). Access to the site and collecting permits were respectively granted by Biodiversity Institutional Commission (University of Costa Rica) (Resolution No. 011, 2010) and Guanacaste Conservation Area (Resolution No. ACG-PI-018-2012), Ministry of Environment, Energy and Telecommunications, Costa Rica.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01902/full#supplementary-material>

REFERENCES

- Ahmed, H., Ettema, T. J., Tjaden, B., Geerling, A. C., van der Oost, J., and Siebers, B. (2005). The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem. J.* 390(Pt 2), 529–540. doi: 10.1042/BJ20041711
- Arnold, H. P., Ziese, U., and Zillig, W. (2000). SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. *Virology* 272, 409–416. doi: 10.1006/viro.2000.0375
- Auch, A. F., von Jan, M., Klenk, H. P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120
- Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P., and White, M. F. (2002). The interaction of Alba, a conserved archaeal, chromatin protein, with Sir2 and its regulation by acetylation. *Science* 296, 148–151. doi: 10.1126/science.1070506
- Berg, I. A., Kockelkorn, D., Buckel, W., and Fuchs, G. (2007). A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in *Archaea*. *Science* 318, 1782–1786. doi: 10.1126/science.1149976
- Berg, I. A., Ramos-Vera, W. H., Petri, A., Huber, H., and Fuchs, G. (2010). Study of the distribution of autotrophic CO₂ fixation cycles in *Crenarchaeota*. *Microbiology* 156, 256–269. doi: 10.1099/mic.0.034298-0
- Brito, J. A., Sousa, F. L., Stelter, M., Bandejas, T. M., Vonrhein, C., Teixeira, M., et al. (2009). Structural and functional insights into sulfide:quinone oxidoreductase. *Biochemistry* 48, 5613–5622. doi: 10.1021/bi9003827
- Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972). *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* 84, 54–68. doi: 10.1007/BF00408082
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST plus: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Chen, L. M., Brügger, K., Skovgaard, M., Redder, P., She, Q. X., Torarinsson, E., et al. (2005). The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota*. *J. Bacteriol.* 187, 4992–4999. doi: 10.1128/JB.187.14.4992-4999.2005
- Chong, L. (2001). Molecular cloning – A laboratory manual, 3rd edition. *Science* 292, 446–446. doi: 10.1126/science.1060677
- Demmer, U., Warkentin, E., Srivastava, A., Kockelkorn, D., Pötter, M., Marx, A., et al. (2013). Structural basis for a bispecific NADP⁺ and CoA binding site in an archaeal malonyl-coenzyme A reductase. *J. Biol. Chem.* 288, 6363–6370. doi: 10.1074/jbc.M112.421263
- Estelmann, S., Hügler, M., Eisenreich, W., Werner, K., Berg, I. A., Ramos-Vera, W. H., et al. (2011). Labeling and enzyme studies of the central carbon metabolism in *Metallosphaera sedula*. *J. Bacteriol.* 193, 1191–1200. doi: 10.1128/JB.01155-10

- Ettema, T. J., Ahmed, H., Geerling, A. C., van der Oost, J., and Siebers, B. (2008). The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) of *Sulfolobus solfataricus*: a key-enzyme of the semi-phosphorylative branch of the Entner-Doudoroff pathway. *Extremophiles* 12, 75–88. doi: 10.1007/s00792-007-0082-1
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360
- Grogan, D., Palm, P., and Zillig, W. (1990). Isolate B12, which harbors a virus-like element, represents a new species of the archaeobacterial genus *Sulfolobus*, *Sulfolobus shibatae*, sp. nov. *Arch. Microbiol.* 154, 594–599. doi: 10.1007/BF00248842
- Grogan, D. W. (1989). Phenotypic characterization of the archaeobacterial genus *Sulfolobus*: comparison of five wild-type strains. *J. Bacteriol.* 171, 6710–6719. doi: 10.1128/jb.171.12.6710-6719.1989
- Grogan, D. W., Ozarzak, M. A., and Bernander, R. (2008). Variation in gene content among geographically diverse *Sulfolobus* isolates. *Environ. Microbiol.* 10, 137–146. doi: 10.1111/j.1462-2920.2007.01439.x
- Guiral, M., Tron, P., Aubert, C., Gloter, A., Iobbi-Nivol, C., and Giudici-Ortoni, M. T. (2005). A membrane-bound multienzyme, hydrogen-oxidizing, and sulfur-reducing complex from the hyperthermophilic bacterium *Aquifex aeolicus*. *J. Biol. Chem.* 280, 42004–42015. doi: 10.1074/jbc.M508034200
- Guo, L., Brügger, K., Liu, C., Shah, S. A., Zheng, H. J., Zhu, Y. Q., et al. (2011). Genome analyses of Icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus-host interaction studies. *J. Bacteriol.* 193, 1672–1680. doi: 10.1128/JB.01487-10
- Hemmi, H., Ikejiri, S., Nakayama, T., and Nishino, T. (2003). Fusion-type lycopene beta-cyclase from a thermoacidophilic archaeon *Sulfolobus solfataricus*. *Biochem. Biophys. Res. Commun.* 305, 586–591. doi: 10.1016/S0006-291X(03)00798-8
- Huber, G., Drobner, E., Huber, H., and Stetter, K. O. (1992). Growth by aerobic oxidation of molecular hydrogen in *Archaea*—a metabolic property so far unknown for this domain. *Syst. Appl. Microbiol.* 15, 502–504. doi: 10.1016/S0723-2020(11)80108-6
- Huber, G., and Stetter, K. O. (1991). *Sulfolobus metallicus*, sp. nov., a novel strictly chemolithoautotrophic thermophilic archaeal species of metal-mobilizers. *Syst. Appl. Microbiol.* 14, 372–378. doi: 10.1016/S0723-2020(11)80312-7
- Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I. A., Kockelkorn, D., et al. (2008). A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7851–7856. doi: 10.1073/pnas.0801043105
- Jahn, U., Huber, H., Eisenreich, W., Hügl, M., and Fuchs, G. (2007). Insights into the autotrophic CO₂ fixation pathway of the archaeon *Ignicoccus hospitalis*: comprehensive analysis of the central carbon metabolism. *J. Bacteriol.* 189, 4108–4119. doi: 10.1128/JB.00047-07
- Jan, R. L., Wu, J., Chaw, S. M., Tsai, C. W., and Tsen, S. D. (1999). A novel species of thermoacidophilic archaeon, *Sulfolobus yangmingensis* sp. nov. *Int. J. Syst. Bacteriol.* 49(Pt 4), 1809–1816. doi: 10.1099/00207713-49-4-1809
- Jaubert, C., Danioux, C., Oberto, J., Cortez, D., Bize, A., Krupovic, M., et al. (2013). Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. *Open Biol.* 3:130010. doi: 10.1098/rsob.130010
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kappler, U., and Dahl, C. (2001). Enzymology and molecular biology of prokaryotic sulfite oxidation. *FEMS Microbiol. Lett.* 203, 1–9. doi: 10.1111/j.1574-6968.2001.tb10813.x
- Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., et al. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res.* 8, 123–140. doi: 10.1093/dnares/8.4.123
- Keeling, P. J., Klenk, H. P., Singh, R. K., Schenk, M. E., Sensen, C. W., Zillig, W., et al. (1998). *Sulfolobus islandicus* plasmids pRN1 and pRN2 share distant but common evolutionary ancestry. *Extremophiles* 2, 391–393. doi: 10.1007/s007920050083
- Kletzin, A. (1992). Molecular characterization of the sor gene, which encodes the sulfur oxygenase reductase of the thermoacidophilic archaeum *Desulfurolobus ambivalens*. *J. Bacteriol.* 174, 5854–5859. doi: 10.1128/jb.174.18.5854-5859.1992
- Kouril, T., Zaparty, M., Marrero, J., Brinkmann, H., and Siebers, B. (2008). A novel trehalose synthesizing pathway in the hyperthermophilic crenarchaeon *Thermoproteus tenax*: the unidirectional TreT pathway. *Arch. Microbiol.* 190, 355–369. doi: 10.1007/s00203-008-0377-3
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Laska, S., Lottspeich, F., and Kletzin, A. (2003). Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon *Acidianus ambivalens*. *Microbiology* 149, 2357–2371. doi: 10.1099/mic.0.26455-0
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Liu, L. J., Stockdreher, Y., Koch, T., Sun, S. T., Fan, Z., Josten, M., et al. (2014). Thiosulfate transfer mediated by DsrE/TusA homologs from acidothermophilic sulfur-oxidizing archaeon *Metallosphaera cuprina*. *J. Biol. Chem.* 289, 26949–26959. doi: 10.1074/jbc.M114.591669
- Liu, Y., Beer, L. L., and Whitman, W. B. (2012). Sulfur metabolism in archaea reveals novel processes. *Environ. Microbiol.* 14, 2632–2644. doi: 10.1111/j.1462-2920.2012.02783.x
- Mao, D., and Grogan, D. (2012). Genomic evidence of rapid, global-scale gene flow in a *Sulfolobus* species. *ISME J.* 6, 1613–1616. doi: 10.1038/ismej.2012.20
- McCarthy, S., Gradnig, J., Johnson, T., Payne, S., Lipzen, A., Martin, J., et al. (2015). Complete genome sequence of *Sulfolobus solfataricus* strain 98/2 and evolved derivatives. *Genome Announc.* 3:e00549-15. doi: 10.1128/genomeA.00549-15
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Nishizawa, Y., Yabuki, T., Fukuda, E., and Wakagi, T. (2005). Gene expression and characterization of two 2-oxoacid:ferredoxin oxidoreductases from *Aeropyrum pernix* K1. *FEBS Lett.* 579, 2319–2322. doi: 10.1016/j.febslet.2004.11.117
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/27.1.29
- Quatrini, R., Appia-Ayme, C., Denis, Y., Jedlicki, E., Holmes, D. S., and Bonnefoy, V. (2009). Extending the models for iron and sulfur oxidation in the extreme Acidophile *Acidithiobacillus ferrooxidans*. *BMC Genomics* 10:394. doi: 10.1186/1471-2164-10-394
- Ramos-Vera, W. H., Weiss, M., Strittmatter, E., Kockelkorn, D., and Fuchs, G. (2011). Identification of missing genes and enzymes for autotrophic carbon fixation in *Crenarchaeota*. *J. Bacteriol.* 193, 1201–1211. doi: 10.1128/JB.01156-10
- Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V., and Whitaker, R. J. (2009). Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8605–8610. doi: 10.1073/pnas.0808945106
- Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., et al. (2001). Viruses from extreme thermal environments. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13341–13345. doi: 10.1073/pnas.231170198
- Rohwerder, T., and Sand, W. (2007). Oxidation of inorganic sulfur compounds in acidophilic prokaryotes. *Eng. Life Sci.* 7, 301–309. doi: 10.1002/elsc.200720204
- Sato, T., and Atomi, H. (2011). Novel metabolic pathways in *Archaea*. *Curr. Opin. Microbiol.* 14, 307–314. doi: 10.1016/j.mib.2011.04.014
- Sensen, C. W., Charlebois, R. L., Chow, C., Clausen, I. G., Curtis, B., Doolittle, W. F., et al. (1998). Completing the sequence of the *Sulfolobus solfataricus* P2 genome. *Extremophiles* 2, 305–312. doi: 10.1007/s007920050073
- She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., et al. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7835–7840. doi: 10.1073/pnas.141222098
- Siebers, B., Wendisch, V. F., and Hensel, R. (1997). Carbohydrate metabolism in *Thermoproteus tenax*: in vivo utilization of the non-phosphorylative

- Entner-Doudoroff pathway and characterization of its first enzyme, glucose dehydrogenase. *Arch. Microbiol.* 168, 120–127. doi: 10.1007/s002030050477
- Suzuki, T., Iwasaki, T., Uzawa, T., Hara, K., Nemoto, N., Kon, T., et al. (2002). *Sulfolobus tokodaii* sp. nov. (f. *Sulfolobus* sp. strain 7), a new member of the genus *Sulfolobus* isolated from Beppu Hot Springs, Japan. *Extremophiles* 6, 39–44. doi: 10.1007/s007920100221
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Teufel, R., Kung, J. W., Kockelkorn, D., Alber, B. E., and Fuchs, G. (2009). 3-hydroxypropionyl-coenzyme A dehydratase and acryloyl-coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in the Sulfolobales. *J. Bacteriol.* 191, 4572–4581. doi: 10.1128/JB.00068-09
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882. doi: 10.1093/nar/25.24.4876
- Tindall, B. J., Rosselló-Móra, R., Busse, H. J., Ludwig, W., and Kampfer, P. (2009). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. doi: 10.1099/ijs.0.016949-0
- Ulas, T., Riemer, S. A., Zaparty, M., Siebers, B., and Schomburg, D. (2012). Genome-scale reconstruction and analysis of the metabolic network in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *PLoS ONE* 7:e43401. doi: 10.1371/journal.pone.0043401
- Urich, T., Gomes, C. M., Kletzin, A., and Frazao, C. (2006). X-ray structure of a self-compartmentalizing sulfur cycle metalloenzyme. *Science* 311, 996–1000. doi: 10.1126/science.1120306
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wang, H., Peng, N., Shah, S. A., Huang, L., and She, Q. X. (2015). Archaeal extrachromosomal genetic elements. *Microbiol. Mol. Biol. Rev.* 79, 117–152. doi: 10.1128/MMBR.00042-14
- Wang, J., Yan, D., Dixon, R., and Wang, Y. P. (2016). Deciphering the principles of bacterial nitrogen dietary preferences: a strategy for nutrient containment. *MBio* 7:e00792-16. doi: 10.1128/mBio.00792-16
- Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301, 976–978. doi: 10.1126/science.1086909
- Wood, A. P., Kelly, D. P., and Norris, P. R. (1987). Autotrophic growth of four *Sulfolobus* strains on tetrathionate and the effect of organic nutrients. *Arch. Microbiol.* 146, 382–389. doi: 10.1007/BF00410940
- Xiang, X., Dong, X., and Huang, L. (2003). *Sulfolobus tengchongensis* sp. nov., a novel thermoacidophilic archaeon isolated from a hot spring in Tengchong, China. *Extremophiles* 7, 493–498. doi: 10.1007/s00792-003-0355-2
- You, X., Zhang, Q., Zheng, H., and Jiang, C. (2014). [Adaptation of *Acidianus hospitalis* W1 to oligotrophic and acidic hot spring environments]. *Wei Sheng Wu Xue Bao* 54, 1193–1203. doi: 10.13343/j.cnki.wsxb.2014.10.012
- Zalkin, H. (1993). The amidotransferases. *Adv. Enzymol. Relat. Areas Mol. Biol.* 66, 203–309. doi: 10.1002/9780470123126.ch5
- Zhang, C., Krause, D. J., and Whitaker, R. J. (2013). *Sulfolobus islandicus*: a model system for evolutionary genomics. *Biochem. Soc. Trans.* 41, 458–462. doi: 10.1042/BST20120338
- Zillig, W. (1991). Comparative biochemistry of *Archaea* and *Bacteria*. *Curr. Opin. Genet. Dev.* 1, 544–551. doi: 10.1016/S0959-437X(05)80206-0
- Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., et al. (1994). Screening for Sulfolobales, their plasmids and their viruses in icelandic solfataras. *Syst. Appl. Microbiol.* 16, 609–628. doi: 10.1016/S0723-2020(11)80333-4
- Zuo, G., Xu, Z., and Hao, B. (2015). Phylogeny and taxonomy of *Archaea*: a comparison of the whole-genome-based CVTree approach with 16S rRNA sequence analysis. *Life (Basel)* 5, 949–968. doi: 10.3390/life5010949

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Dai, Wang, Zhang, Li, Zhang, Mora-López, Jiang, Liu, Wang, Zhu, Hernández-Ascencio, Dong and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome-Based Genetic Tool Development for *Bacillus methanolicus*: Theta- and Rolling Circle-Replicating Plasmids for Inducible Gene Expression and Application to Methanol-Based Cadaverine Production

Marta Irla¹, Tonje M. B. Heggeset², Ingemar Nærdal², Lidia Paul¹, Tone Haugen², Simone B. Le², Trygve Brautaset^{2,3} and Volker F. Wendisch^{1*}

OPEN ACCESS

Edited by:

Kian Mau Goh,
Universiti Teknologi Malaysia,
Malaysia

Reviewed by:

María Sofía Urbieto,
Consejo Nacional de Investigaciones
Científicas y Técnicas (CONICET),
Argentina
Kheng Oon Low,
Malaysia Genome Institute, Malaysia

*Correspondence:

Volker F. Wendisch
volker.wendisch@uni-bielefeld.de

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 13 May 2016

Accepted: 06 September 2016

Published: 22 September 2016

Citation:

Irla M, Heggeset TM, Nærdal I,
Paul L, Haugen T, Le SB,
Brautaset T and Wendisch VF (2016)
Genome-Based Genetic Tool
Development for *Bacillus*
methanolicus: Theta- and Rolling
Circle-Replicating Plasmids
for Inducible Gene Expression
and Application to Methanol-Based
Cadaverine Production.
Front. Microbiol. 7:1481.
doi: 10.3389/fmicb.2016.01481

¹ Genetics of Prokaryotes, Faculty of Biology and CeBiTec, Bielefeld University, Bielefeld, Germany, ² SINTEF Materials and Chemistry, Department of Biotechnology and Nanomedicine, Trondheim, Norway, ³ Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway

Bacillus methanolicus is a thermophilic methylotroph able to overproduce amino acids from methanol, a substrate not used for human or animal nutrition. Based on our previous RNA-seq analysis a mannitol inducible promoter and a putative mannitol activator gene *mtlR* were identified. The mannitol inducible promoter was applied for controlled gene expression using fluorescent reporter proteins and a flow cytometry analysis, and improved by changing the -35 promoter region and by co-expression of the *mtlR* regulator gene. For independent complementary gene expression control, the heterologous xylose-inducible system from *B. megaterium* was employed and a two-plasmid gene expression system was developed. Four different replicons for expression vectors were compared with respect to their copy number and stability. As an application example, methanol-based production of cadaverine was shown to be improved from 6.5 to 10.2 g/L when a heterologous lysine decarboxylase gene *cadA* was expressed from a theta-replicating rather than a rolling-circle replicating vector. The current work on inducible promoter systems and compatible theta- or rolling circle-replicating vectors is an important extension of the poorly developed *B. methanolicus* genetic toolbox, valuable for genetic engineering and further exploration of this bacterium.

Keywords: *Bacillus methanolicus*, thermophile, methylotroph, genetic tool box, theta-replicating plasmids, gene expression

INTRODUCTION

Bacillus methanolicus is a thermophilic bacterium, able to grow on methanol as a sole carbon and energy source (Schendel et al., 1990; Arfman et al., 1992). The growth of *B. methanolicus* occurs in a wide temperature range between 37 and 60°C, with an optimum at 50°C. It was, however, observed that a rapid change of growth temperature from 50 to 37°C leads to the initiation of sporulation

processes in the wild type strain MGA3, specifically the upregulation of stage VI sporulation protein D, the anti-sigma F factor antagonist SpoIIAA, and the stage IV sporulation protein A and the downregulation of two proteins which belong to the flagellar apparatus (Schendel et al., 1990; Müller et al., 2014). *B. methanolicus* MGA3 produces 60 g/L of L-glutamate in methanol-controlled high cell density fed-batch fermentations (Schendel et al., 2000; Heggeset et al., 2012). Currently, *Corynebacterium glutamicum* is typically used for the industrial production of L-glutamate in fermentative processes with the most common carbon sources being molasses and sugar cane, and the global annual consumption reaching 3.2 million tons (IHS Chemical, 2016). The second largest product of the amino acid market is L-lysine, a feed additive with an annual demand exceeding 2 million tons (Zahoor et al., 2012). *B. methanolicus* does not naturally overproduce this amino acid, however, during the last two decades several strategies have been employed to generate L-lysine producing strains. To date, the classical mutants of *B. methanolicus* produce up to 65 g/L of L-lysine in high cell density methanol fed-batch fermentations (Brautaset et al., 2010). Furthermore, it was shown both in wild type and lysine producing strains that heterologous expression of a lysine decarboxylase enables the synthesis of cadaverine (Nærdal et al., 2015). Cadaverine, also known as 1,5-diaminopentane, is a five-carbon linear aliphatic diamine (Schneider and Wendisch, 2011; Shimizu, 2013), that finds applications in the (bio)plastics industry since polycondensation of cadaverine with dicarboxylic acids yields polyamides or nylons of the AA,BB-type (Shimizu et al., 2003; Wendisch, 2014). The most significant advantage of *B. methanolicus* for the use in the amino acid industry is its ability to utilize methanol as a carbon source in combination with a high growth temperature, which leads to a reduced need for cooling. Methanol is a cheap, non-food alternative to raw materials commonly used in the biotechnological processes (Müller et al., 2015a). In the recent years, considerable progress has been made in the elucidation of the methanol utilization pathway starting from sequencing of the full genome (Heggeset et al., 2012; Irla et al., 2014), characterization of the enzymes involved in the methanol oxidation and the ribulose monophosphate (RuMP) pathway (Krog et al., 2013; Stolzenberger et al., 2013a,b; Markert et al., 2014; Ochsner et al., 2014; Wu et al., 2016), unraveling of the transcriptome by the means of microarray analysis (Heggeset et al., 2012) and RNA-seq (Irla et al., 2015), of the proteome (Müller et al., 2014), and the metabolome (Kiefer et al., 2015; Müller et al., 2015b). These findings enabled a better understanding of the metabolic processes taking place during growth on methanol, but also on the limited number of alternative C-sources for this facultative methylotroph, in particular on mannitol.

The obvious suitability of *B. methanolicus* for industrial application has been the main motivation behind the extensive work on the development of metabolic engineering tools. The first attempts included random mutagenesis approaches towards increased L-lysine production (Hanson et al., 1996; Brautaset et al., 2010). Furthermore, the protocol for the protoplast transformation with plasmid DNA was developed and several different origins of replication were tested for their

transformation efficiency and stability (Cue et al., 1997). The protoplast-based transformation protocols are known to be laborious and difficult to perform, for this reason a more versatile electroporation procedure was developed, and for the first time the *mdh* promoter (*mp*) was used to establish plasmid based gene expression (Jakobsen et al., 2006). The only alternative vector that has been used for the heterologous gene expression thus far is pNW33N with a *gfp* gene cloned under control of the *mdh* promoter (Nilasari et al., 2012).

Despite the fact that some progress has been made in genetic manipulation of *B. methanolicus*, and that L-lysine and cadaverine producing strains have been created by the plasmid-based gene expression, the available toolbox is a limiting factor for the development of industrially relevant *B. methanolicus* strains. Here, we present the expansion of the metabolic engineering tool box by the addition of two new expression vectors and the establishment and development of xylose- and mannitol-inducible promoter systems.

MATERIALS AND METHODS

Strains, Plasmids, and Primers

All strains, plasmids, and primers constructed and used in this study are listed in the Supplementary Tables. *B. methanolicus* MGA3 was used as the expression host, *Escherichia coli* strain DH5 α (Stratagene) was used as the general cloning host.

Molecular Cloning

All standard recombinant DNA procedures were performed as described by Sambrook and Russell (2001). Plasmid DNA was introduced into chemically competent *E. coli* cells (Higa and Mandel, 1970; Hanahan, 1983). Total DNA was isolated from *B. methanolicus* using the MasterPureTM Gram Positive DNA Purification Kit (Epicenter) or as previously described (Eikmanns et al., 1994). The NucleoSpin[®] Gel and PCR Cleanup kit (Machery–Nagel) and the Qiaquick PCR Purification and Gel Extraction kits (Qiagen) were used for PCR purification and gel extraction. Plasmids were isolated using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) or the Wizard[®] Plus SV Minipreps (Promega). Plasmid backbones were amplified with PfuTurbo DNA polymerase (Agilent), inserts with ALLinTM HiFi DNA Polymerase (highQ) or the ExpandTM High Fidelity PCR System (Roche). Dephosphorylation of plasmid DNA was performed using Antarctic Phosphatase or Calf Intestinal Alkaline Phosphatase (New England Biolabs). The DNA fragments were joined either with Rapid DNA Ligation Kit (Roche), T4 DNA ligase (New England Biolabs) or by the means the isothermal DNA assembly (Gibson et al., 2009). For colony PCR the Taq polymerase (New England Biolabs) was used. Site-directed mutagenesis was performed essentially as described by Liu and Naismith (2008) using Pfu polymerase (Agilent). All cloned DNA fragments and introduced mutations were verified by sequencing. *B. methanolicus* competent cells were prepared according to Jakobsen et al. (2006). SOBsuc plates [1% (w/v) agar] supplemented with suitable antibiotics were used instead of regeneration plates. SOBsuc medium is SOB medium

(Difco) supplemented with 0.25 M sucrose. Electroporation was performed as previously described (Jakobsen et al., 2006).

Media and Cultivation Conditions

Escherichia coli strains were cultivated at 37°C in Lysogeny Broth (LB) or on LB-agar plates supplemented with antibiotics (ampicillin 200 µg/mL, chloramphenicol 30 µg/mL, kanamycin 50 µg/mL) when relevant. Unless otherwise stated, *B. methanolicus* strains were cultured at 50°C in MVcMY minimal medium with 200 mM methanol as previously described (Brautaset et al., 2004). When appropriate, media were supplemented with kanamycin 50 µg/mL (or 10 µg/mL) and/or chloramphenicol 5 µg/mL. Inducers were used at the following concentrations: mannitol [2.5, 5.0, 12.5, 25, 50, and 55 mM (1%)], arabitol (50 mM), ribitol (50 mM), xylitol (50 mM), xylose [0.01, 0.05, 0.1, 0.5, or 1% (w/v)], or CuSO₄ (10, 20, 50, 100, and 200 µM). All experiments were performed in triplicates.

β-Galactosidase (LacZ) Activity Assay

For LacZ enzymatic assays, overnight cultures of *B. methanolicus* strains MGA3 (pTH1mp-lacZ), MGA3 (pTH1xp-lacZ), MGA3 (pTH1cup-lacZ), MGA3 (pTH1mtAp-lacZ), or MGA3 (pHP13), were diluted to OD₆₀₀ 0.2 in fresh medium with appropriate antibiotics. When the cultures reached OD₆₀₀ = 0.5, they were split in two equal halves. Inducer (50 µM CuSO₄, 1% (w/v) xylose, or 1% (w/v) mannitol) was added to one of the two and growth was continued until OD₆₀₀ 1–1.5. Cells were harvested by centrifugation (5000 g, 10 min, 4°C) and the pellets were stored at –80°C. Cells were thawed, resuspended in potassium phosphate buffer (100 mM, pH 7.0) (10% of the original volume) and sonicated on ice/water for 10–15 min (Branson Sonifier 250, output control = 3 and duty cycle = 30%). Cellular debris was removed by centrifugation (10000 g, 45 min, 4°C) followed by filtration through a 0.2 µm sterile filter. Enzymatic activities were measured by monitoring the liberation of *o*-nitrophenol from *o*-nitrophenyl β-D-galactopyranoside (ONPG) at 410 nm. 100 mM potassium phosphate buffer pH 7.0 (910 µl), 68 mM ONPG (30 µl), and 30 mM MgCl₂ (30 µl) were mixed and the catalysis started by the addition of cell extract (30 µl). The molar extinction coefficient used for *o*-nitrophenyl at 410 nm, pH 7.0 used for calculation is 3500 M^{–1} cm^{–1} and the light path 1 cm. One unit (U) is defined as the amount of enzyme able to convert 1.0 µmol of ONPG per min.

Flow Cytometry

For the fluorescent activated cell scanning analysis, overnight cultures were diluted to an initial OD₆₀₀ of 0.15 and cultivated for 6 h at 50°C prior to incubation at 37°C for two hours. Samples were centrifuged at 13,000 g, 5 min, 4°C, washed twice with cold phosphate-buffered saline (PBS) and resuspended therein to a final OD₆₀₀ of 0.3. The fluorescence was determined in a flow cytometer (Becton Dickinson) using the Kaluza for Gallios Acquisition Software 1.0. The fluorescence emission signal was collected with a 450/50 BP bandpass filter (FL9) for GFPuv and with a 620/30 BP, bandpass (FL3) for mCherry. The following data analysis was performed using Kaluza Analysis Software 1.3.

Plasmid Stability

To test for stability of plasmid segregation, overnight cultures were diluted in 50 mL fresh medium with and without relevant antibiotics to an initial OD₆₀₀ of 0.05, grown for 12 h (six generations) and then diluted again into fresh medium to an initial OD₆₀₀ of 0.05. This was repeated over the course of the whole experiment. After 6 h from inoculation, 10 mL of the cultures were aliquoted to 100 mL shaking flasks and incubated for 2 h at 37°C, 200 rpm, after which the flow cytometry analysis was carried out. This procedure was repeated every 24 h (every 12 generations) for a total of 5 days (60 generations). The stability is presented as the ratio of cells fluorescent in absence of antibiotics to the cells fluorescent in presence of antibiotics.

Estimation of Copy Number by the Means of Droplet Digital PCR

Overnight cultures of MGA3 (pHCMC04), MGA3 (pHP13), MGA3 (pNW33Nkan), or MGA3 (pUB110Smp-lacZ) were diluted to 2% in fresh medium (supplemented with 5 µg/ml chloramphenicol or 10 µg/ml kanamycin) and cultivated until the mid-exponential growth phase (OD₆₀₀ 2–4). Cell pellets were harvested from 5 ml cultures by centrifugation and total DNA was extracted using the MasterPure™ Gram Positive DNA Purification Kit (Epicenter), followed by an additional purification step using the Agencourt® AMPure XP system (Beckman Coulter). DNA concentrations were determined on a Qubit® 2.0 Fluorometer using the Qubit® dsDNA BR Assay Kit (ThermoFischer Scientific). Twenty microliter ddPCR reaction mixtures containing EvaGreen Supermix (Bio-Rad), primers (0.2 µM) and gDNA template (8 or 20 pg) were prepared according to the manufacturer's instructions and used for droplet generation (QX200 droplet generator, Bio-Rad). Forty microliter of sample was manually transferred to a 96-well plate and heat-sealed prior to amplification initiated by enzyme activation at 95°C for 5 min, followed by 40 cycles of amplification (95°C for 30 s, 60°C 1 min) and signal stabilization (4°C 5 min, 90°C 5 min), temperature ramp 2.5°C/s. Following amplification, fluorescence intensity was measured in a QX200 Droplet Reader (Bio-Rad) and the signal data were analyzed with QuantaSoft, Version 1.5.38 (Bio-Rad). Primer sequences are listed in the Supplementary Material.

High Cell Density Fed-Batch Methanol Fermentation

Fed-batch fermentation was performed at 50°C in UMN1 medium using Applikon 3 L fermenters with an initial volume of 0.75 L medium essentially as previously described (Jakobsen et al., 2009; Brautaset et al., 2010). Kanamycin (50 µg/mL) or chloramphenicol (5 µg/mL) was added to the initial batch growth medium, the pH was maintained at 6.5 by automatic addition of 12.5% (w/v) NH₃ solution, and the dissolved oxygen level was maintained at 30% saturation by increasing the agitation speed and using enriched air (up to 60% O₂). The methanol concentration in the fermenter was monitored by

online analysis of the headspace gas with a mass spectrometer (Balzers Omnistar GSD 300 02). The headspace gas was transferred from the fermenters to the mass spectrometer in insulated heated (60°C) stainless steel tubing. The methanol concentration in the medium was maintained at a set point of 150 mM by automatic addition of methanol feed solution containing methanol, trace metals and antifoam 204 (Sigma), as previously described (Brautaset et al., 2010). All fermentations were run until the carbon dioxide content of the exhaust gas was close to zero (no cell respiration). Bacterial growth was monitored by measuring OD₆₀₀. Dry cell weight was calculated using a conversion factor of one OD₆₀₀ unit corresponding to 0.24 g dry cell weight per liter (Jakobsen et al., 2009). Due to significant increase in the culture volume throughout the fermentation, the biomass, cadaverine, and amino acid concentrations were corrected for the increase in volume and subsequent dilution. A volume correction factor of 1.8 was used for values presented in **Table 2**. The actual concentrations measured in the bioreactors were therefore accordingly lower as described previously (Jakobsen et al., 2009). Samples for determination of volumetric cadaverine and amino acid yields were collected from early exponential phase and throughout the cultivation (10–47 h).

Measurement of Cadaverine and Amino Acids

Samples were analyzed by RP-HPLC as described previously by Skjerdal et al. (1996) using pre-column derivatization with *o*-phthalaldehyde and a buffer containing 0.02 M sodium acetate +2% tetrahydrofuran at pH 5.9.

Detection of α -Amylase Activity

For the detection of α -amylase activity, overnight cultures were diluted to an initial OD₆₀₀ of 0.15 and cultivated for 6–8 h at 50°C. The cultures were diluted to OD₆₀₀ of 1 and 15 μ L of the diluted cultures were placed on the appropriate plates in the form of a drop. The plates were incubated for 12 h at 50°C to allow the cell growth and then placed at 37°C for next 24 h for in order to support activity of the heterologous α -amylase. Ten milliliter of iodine solution were placed on the plate in order to visualize the formation of the halo in the starch.

RESULTS

Comparison of Different Replicons for Plasmid-Based Gene Expression in *B. methanolicus* MGA3

Genetic engineering of *B. methanolicus* has until now relied on only two plasmids, pNW33N and pHP13. Therefore, we decided to analyze a range of different replicons with regard to their applicability for gene overexpression in *B. methanolicus* MGA3. We compared four different plasmids that were able to replicate: pTH1mp (derived from pHP13), pUB110Smp, pNW33Nmp, and pBV2mp (derived from pHCMC04). As shown in **Table 1**, we have chosen plasmids differing in the copy number, original host organism and the replication mechanism. All rolling circle (RC) plasmids used belong to the pC194/pUB110 family, which is characterized by similarity in Rep protein and the sequences of sites involved in the replication with pNW33N and pUB110 sharing identical Rep protein sequences. The pUB110 plasmid is reported to be a high copy number plasmid in *B. subtilis*, pNW33N – medium, pHP13 and pBV2mp – low copy number, both of the low copy number plasmids originate from *B. subtilis*.

Our initial goal was to characterize the copy number, expression levels and stability of the chosen plasmids in *B. methanolicus* MGA3. To analyze the expression levels, *gfpuv* (Chalfie et al., 1994; Crameri et al., 1996) was used as a reporter controlled by the *mdh* promoter from *B. methanolicus* MGA3. Fluorescence intensity was evaluated during growth in methanol minimal medium by flow cytometry. Using the ddPCR, plasmid copy numbers were estimated for the selected plasmids and, in comparison, for the native MGA3 plasmids pBM19 and pBM69 (**Table 1**). The plasmid pUB110Smp-*gfpuv* showed the highest fluorescence levels among the plasmids tested (**Table 1**), followed by pNW33Nmp-*gfpuv*, pTH1mp-*gfpuv*, and pBV2mp-*gfpuv*, respectively, which was in accordance to the plasmid copy number results (**Table 1**).

Next, we compared the plasmid stability for *gfpuv*-expressing RC plasmids transferred to *B. methanolicus* MGA3. The strains were grown for 60 generations in media with and without antibiotic selection and plasmid-containing cells emitting a fluorescence signal were counted every 12 generations. As shown in **Figure 1** only the pTH1mp plasmid was lost at a significant level over the course of the experiment.

TABLE 1 | Comparison of properties and *gfpUV* expression from methanol-inducible promoter P_{mp} of plasmids used in the study.

| Plasmid | Parental plasmid | Original host organism | Replication model | Copy number in <i>Bacillus subtilis</i> | Copy number in <i>B. methanolicus</i> [#] | Median fluorescence intensity of GFPuv [a.u.] |
|-------------------------|------------------------------|------------------------|-------------------|---|--|---|
| pTH1mp- <i>gfpuv</i> | pHP13 (pTA1060) ^a | <i>B. subtilis</i> | Rolling circle | 5–6 ^a | 5 ± 1 ¹ | 0.9 ± 0.1 |
| pNW33Nmp- <i>gfpuv</i> | pNW33N (pC194) ^b | <i>S. aureus</i> | Rolling circle | 15 ^{†e} | 19 ± 2 ² | 2.5 ± 0.1 |
| pUB110Smp- <i>gfpuv</i> | pUB110 ^c | <i>S. aureus</i> | Rolling circle | 30–50 ^c | 25 ± 1 ³ | 3.9 ± 0.0 |
| pBV2mp- <i>gfpuv</i> | pHCMC04 (pBS72) ^d | <i>B. subtilis</i> | Theta-replication | 6 ^{*f} | 3 ± 1 ⁴ | 0.2 ± 0.0 |

^aHaima et al., 1987; ^bRhee et al., 2007; ^cGryczan et al., 1978, ^dNguyen et al., 2005; ^ete Riele et al., 1986; ^fTitok et al., 2003; [†]Plasmid copy number for parental plasmid pC194; *Plasmid copy number for derivative of parental plasmid pBS72; ¹pHP13; ²pNW33Nkan; ³pUB110Smp-lacZ; ⁴pHCMC04. [#]For comparison, the copy numbers of the native *B. methanolicus* plasmids pBM19 and pBM69 were 3.4 ± 0.7 and 1.3 ± 0.2, respectively. *B. methanolicus* without vector showed a mean GFPuv fluorescence of 0.1 ± 0.0. Mean values and standard deviations of triplicate shake flask cultures are presented.

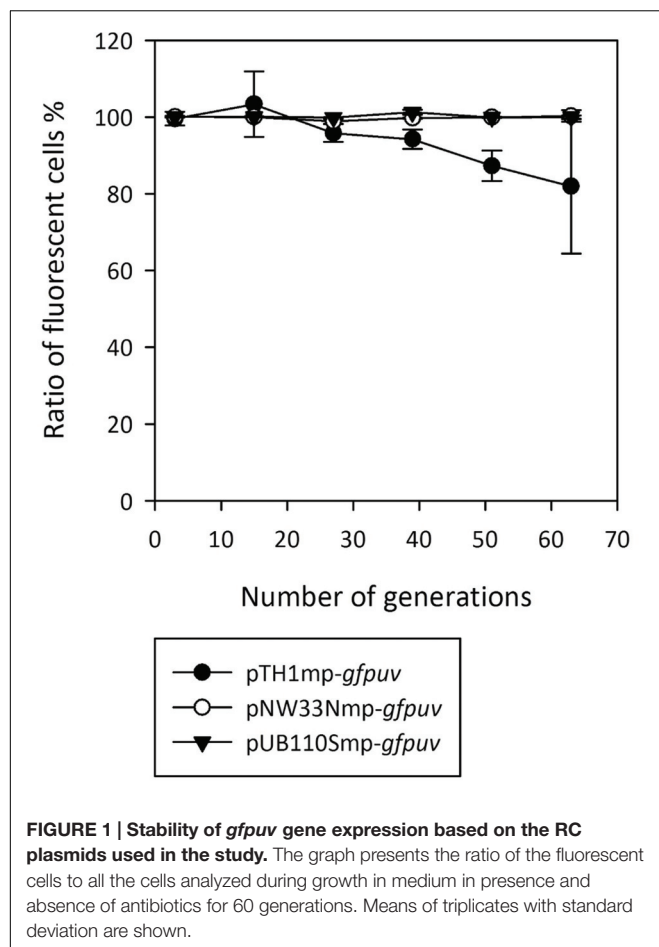


FIGURE 1 | Stability of *gfpuv* gene expression based on the RC plasmids used in the study. The graph presents the ratio of the fluorescent cells to all the cells analyzed during growth in medium in presence and absence of antibiotics for 60 generations. Means of triplicates with standard deviation are shown.

Cadaverine Production from Methanol by Expression of a Heterologous Lysine Decarboxylase Gene from a Theta-Replicating Plasmid

The plasmids pTH1mp and pBV2mp, containing the *mdh* promoter were used to study cadaverine production in *B. methanolicus* during fed-batch methanol fermentation. We have previously reported a methanol-based cadaverine production titer of 6.5 g/L by *B. methanolicus* MGA3 (pTH1mp-*cadA*), a strain overexpressing the lysine decarboxylase *cadA* gene from *E. coli* (corrigendum to Nærdal et al., 2015). We compared cadaverine production in the strain overexpressing *cadA* from a theta-replicating plasmid during high cell density fed-batch fermentation. The *B. methanolicus* strain MGA3 (pBV2mp-*cadA*) was tested in duplicates under comparable fermentation conditions. Samples for cadaverine and amino acid analysis, cell dry weight and OD₆₀₀ were taken throughout the cultivation. As presented in Table 2, we obtained a cadaverine production titer of 10.2 g/L based on the alternative theta-replicating pBV2mp plasmid. A substantial 55% production increase compared to the previously reported (pTH1mp-*cadA*)-based strain was observed. While biomass and by-product levels were similar between the two strains, the specific growth rate of MGA3

TABLE 2 | Fed-batch methanol fermentation production data of strains MGA3 (pBV2mp-*cadA*) and MGA3 (pTH1mp-*cadA*).

| Strain | CDW ^a | μ ^b | Asp ^c | Glu ^c | Ala ^c | Lys ^c | Cad ^c |
|-----------------------------|------------------|--------------------|------------------|------------------|------------------|------------------|------------------|
| | g/L | h ⁻¹ | g/L | g/L | g/L | g/L | g/L |
| MGA3 (pBV2mp- <i>cadA</i>) | 60.9 | 0.38 | 1.6 | 72.2 | 9.2 | 0.5 | 10.2 |
| MGA3 (pTH1mp- <i>cadA</i>) | 65.5 | 0.45 | 1.5 | 71.8 | 10.2 | 0.0 | 6.5 |

Mean values of duplicate cultures for *B. methanolicus* MGA3 (pBV2mp-*cadA*) are shown. Deviation did not exceed 10%. The MGA3 (pTH1mp-*cadA*) data was imported from Nærdal et al. (2015). CDW, cell dry weight; μ , specific growth rate; Asp, L-aspartate; Glu, L-glutamate; Ala, L-alanine; Lys, L-lysine; Cad, cadaverine.

^aBiomass concentrations are maximum values from the stationary growth phase.

^bSpecific growth rates are maximum values calculated from the exponential growth period. ^cCadaverine and amino acid concentrations are maximum values and volume corrected.

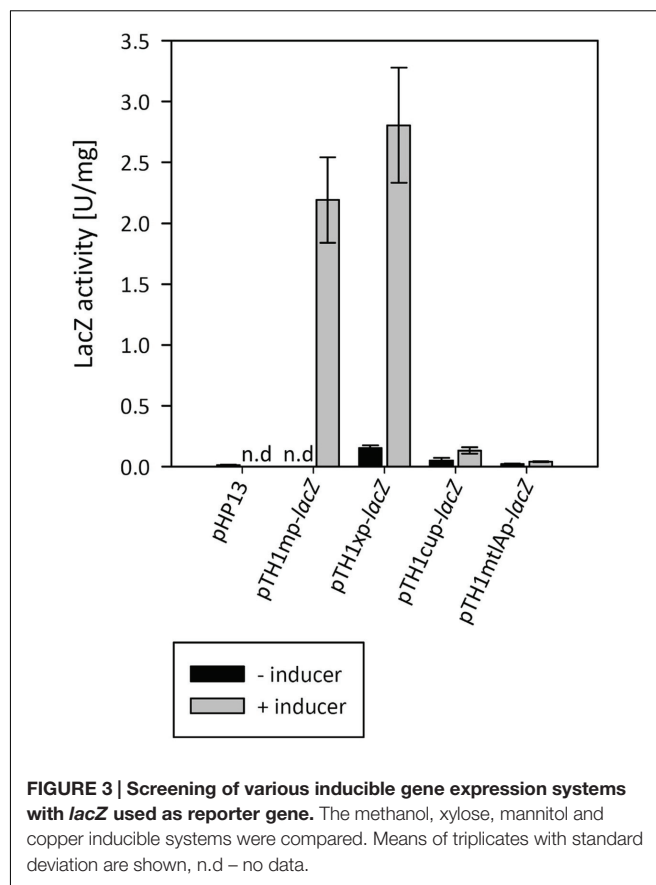
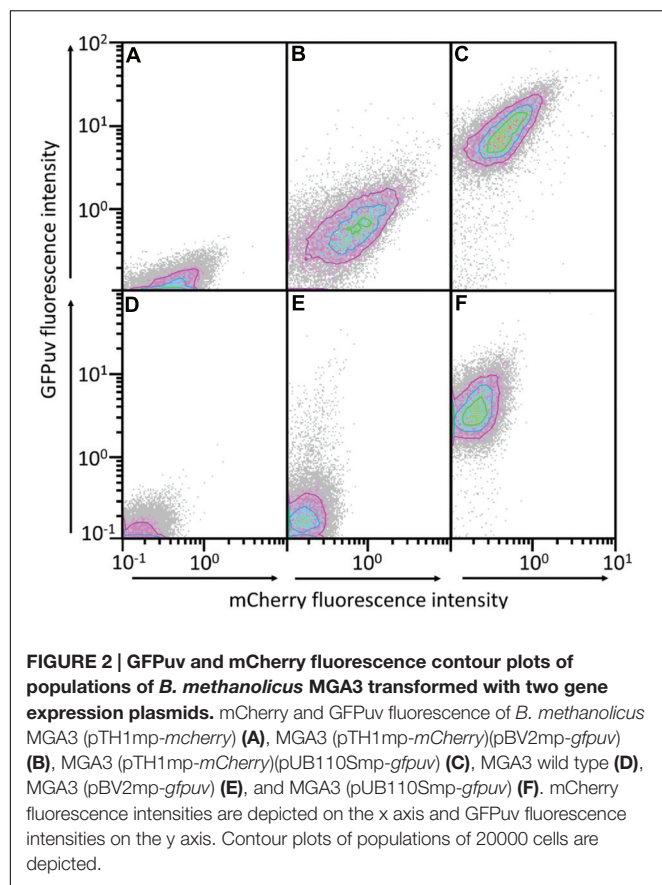
(pBV2mp-*cadA*) was lower than that of MGA3 (pTH1mp-*cadA*) (Table 2).

Plasmid Compatibility

In order to establish a two plasmid-based gene expression system, we analyzed the compatibility of the chosen RC (pTH1mp and pUB110Smp)- and theta (pBV2mp)-replicating plasmids in *B. methanolicus*. pTH1mp and pUB110Smp share high identity (42%) of their replication protein Rep and of the origin of replication sequence (95%) and for this reason it was not clear whether they can coexist in the same cell. Similarly, we did not analyze the pUB110mp/pNW33N plasmid pair which display 100% identity of Rep protein sequences. Plasmids for expression of either *gfpuv* or *mcherry* (Shaner et al., 2004) were constructed to simultaneously analyze gene expression from two vectors. The following plasmid combinations were applied: pTH1mp-*mcherry* with pUB110Smp-*gfpuv* or pTH1mp-*mcherry* with pBV2mp-*gfpuv*. Overexpression of *mcherry* from pTH1mp led to red fluorescence (depicted on the y-axis in Figure 2). Similarly, overexpression of *gfpuv* from pUB110Smp or from pBV2mp yielded green-fluorescent cells (x-axis of Figure 2). Cells transformed with pTH1mp-*mcherry* and pUB110Smp-*gfpuv* or with pTH1mp-*mcherry* and pBV2mp-*gfpuv* showed simultaneous red and green fluorescence (Figure 2) providing evidence for two plasmid-based gene expression in *B. methanolicus*.

Construction of Mannitol Inducible Gene Expression System

In order to choose a suitable system for inducible gene expression we screened several inducible promoter systems using the thermostable LacZ from *B. coagulans* as a reporter (Kovács et al., 2010). We have tested the *B. megaterium* xylose inducible system from plasmid pHCMC04 (Nguyen et al., 2005), a native mannitol inducible promoter from MGA3, and a copper inducible promoter from *Lactobacillus sakei* (Crutz-Le Coq and Zagorec, 2008). As shown in Figure 3, the xylose inducible promoter system was functional in *B. methanolicus* MGA3 and, when fully induced, yielded higher expression levels than the hitherto used *mdh* promoter. Very low expression was observed from both the mannitol-inducible promoter present in the



upstream region of the *mtlA* gene of *B. methanolicus* MGA3, and the copper inducible promoter. The copper-inducible promoter showed a dose-response where the activity in cultures induced by 100 μ M CuSO_4 was approximately threefold higher than in cultures induced by 50 μ M CuSO_4 (data not shown). The inducer, however, had a toxic effect on the cells, reducing the growth rate considerably at concentrations above 50 μ M CuSO_4 (data not shown), making it not suitable for industrial applications.

Since only very low expression was observed from the mannitol-inducible *mtlA* promoter, we used DNA microarray data (Heggeset et al., 2012) and RNA-seq data (Irla et al., 2015) to identify other mannitol-inducible genes. **Figure 4** presents the genomic and transcriptomic organization of four genes which belong to the mannitol utilization pathway: *mtlA* coding for PTS system mannitol-specific EIICB component, *mtlR* encoding a transcriptional regulator, *mtlF* coding for mannitol-specific phosphotransferase enzyme IIA and *mtlD* encoding mannitol-1-phosphate 5-dehydrogenase. Genes *mtlF* and *mtlD* are co-expressed as an operon. Transcription start sites (TSSs) were not detected either for *mtlA* or for *mtlF-mtlD*; however, a TSS was found for *mtlR* (Irla et al., 2015). This 5' untranslated region (5' UTR) of *mtlR* is 80 nt in length and its upstream sequence contains conserved -10 and -35 regions (bold): 5'-TTGTATTAAAGGGATATAAACGTTTTATGATAAATATG-3', furthermore the putative ribosome binding site (RBS) sequence

is AGTGGAG, which differs in two positions from the *B. methanolicus* consensus RBS motif AGGAGG (Irla et al., 2015). We cloned the upstream sequence of the *mtlR* gene into the plasmid pTH1 containing the *gfpuv* gene and exchanged the RBS sequence to the consensus motif, which resulted in plasmid pTH1m2p-*gfpuv*.

At first, several sugar alcohols were tested as potential inducers (**Figure 5A**). Supplementation with 50 mM mannitol induced *gfpuv* expression; however, neither arabinol, ribitol, nor xylitol induced reporter gene expression (**Figure 5A**). Subsequently, a titration experiment with different concentrations of mannitol was performed. While the addition of 5 mM mannitol did not increase reporter gene expression, high GFPuv fluorescence intensities were observed upon addition of 12.5, 25, and 50 mM mannitol (**Figure 5B**). It has to be noted that GFPuv fluorescence intensities in the presence of 12.5, 25, and 50 mM mannitol were comparable suggesting that full induction has been achieved (**Figure 5B**). The expression level of the mannitol-inducible promoter in the presence of 50 mM mannitol was similar to that obtained with the conventionally used *mdh* promoter. To test whether higher gene expression is possible in the mannitol inducible system, we decided to exchange the sequences of the -10 and/or the -35 region for the previously described consensus sequences (Irla et al., 2015). As shown in **Table 3**, the exchange of the -35 region or the -35 region together with the -10 region led to higher fluorescence levels in comparison

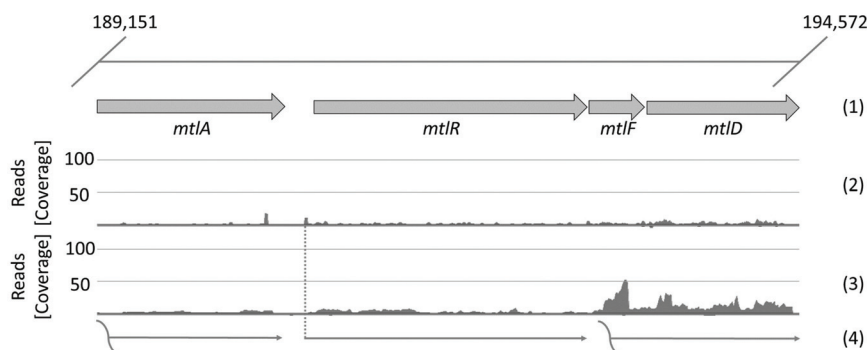


FIGURE 4 | Transcriptional organization of mannitol utilization pathway genes of *B. methanolicus* MGA3. The first line (1) represents the genomic organization of the genes in the cluster, the next one (2) the mapped reads of 5'-ends of primary transcripts, the third (3) depicts the mapped reads of whole primary transcripts and the last (4): the putative corresponding transcripts. The level of transcription is visualized with ReadXplorer (Hilker et al., 2014) and given as absolute reads (coverage) at the corresponding genomic positions. Data are based on the RNA-seq analysis of Irla et al. (2015).

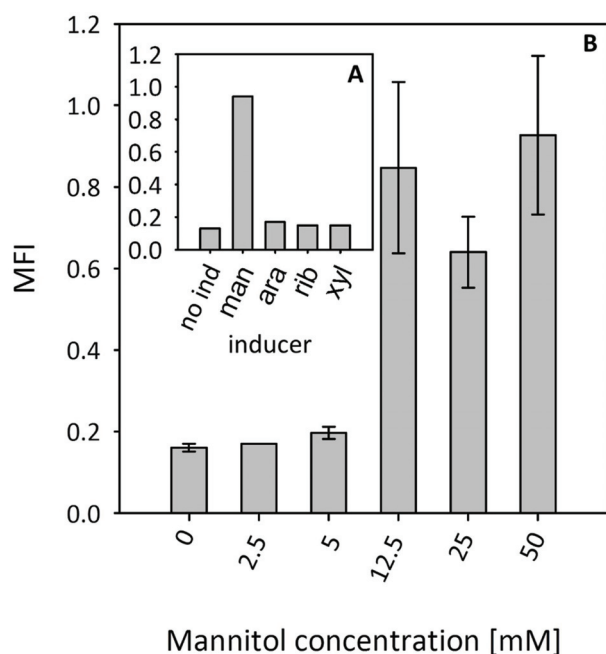


FIGURE 5 | Specificity and dynamic range of induction of the *B. methanolicus* *mtlR* promoter. (A) Screening of sugar alcohols as potential inducers. Single replicates of shake flask cultures are presented. (B) Expression titration with different mannitol concentrations. Abbreviations: no ind, no inducers; man, mannitol; ara, arabinose; rib, ribitol; xyl, xylitol; MFI, median fluorescence intensity of GFPuv given in a.u. GFPuv fluorescence of exponentially growing cells was measured after 2 h incubation at 37°C and 200 rpm. Mean values and standard deviations of triplicate shake flask cultures are presented.

to the native promoter. However, the double exchange caused a 3.5-folds increased background expression.

MtlR has been characterized as a mannitol-dependent transcriptional activator in several species (Joyet et al., 2015). The alignment of the *B. methanolicus* MtlR protein sequence with

TABLE 3 | Reporter gene expression from the *Bacillus methanolicus* mannitol inducible *mtlR* promoter with changed –35 and –10 region sequences.

| Promoter name | –35 region | –10 region | Median fluorescence intensity of GFPuv [a.u.] | |
|---------------|------------|------------|---|----------------|
| | | | 0 mM mannitol | 50 mM mannitol |
| m2p | TTGTAT | TATGAT | 0.2 ± 0.0 | 1.1 ± 0.2 |
| m21p | ---A-A | ----- | 0.2 ± 0.0 | 2.9 ± 0.1 |
| m22p | ---A-A | ---A-- | 0.6 ± 0.1 | 3.8 ± 0.3 |
| m23p | ----- | ---A-- | 0.2 ± 0.0 | 0.9 ± 0.4 |

GFPuv fluorescence of exponentially growing cells was measured after 2 h incubation at 37°C and 200 rpm. Mean values and standard deviations of triplicate shake flask cultures are given.

the sequences of characterized regulators from *L. casei* BL23, *B. subtilis* ssp. *subtilis* str. 168 and *Geobacillus stearothermophilus* ATCC 7954 (Figure 6) revealed the conserved residues important for the regulatory activity of MtlR. The high similarity to the characterized proteins suggested that MtlR of *B. methanolicus* most probably serves as a transcriptional activator. For this reason, we decided to test whether the plasmid-borne overexpression of *mtlR* increased *mtlR* promoter activity. As shown in Figure 7, the overexpression of this gene increased the reporter gene expression from the mannitol inducible promoter m2p by more than 2.5-fold while maintaining a low level of background expression in the absence of mannitol. Taken together, evidence is provided for a versatile mannitol inducible system for the thermophilic *B. methanolicus* on the basis on the previously obtained RNA-seq data.

Xylose Inducible Gene Expression in *B. methanolicus*

Based on our screening experiments (Figure 3), we decided to further develop the xylose inducible system for gene expression in *B. methanolicus*. We have subcloned the *xylR* regulator gene together with the promoter and the RBS sequence of

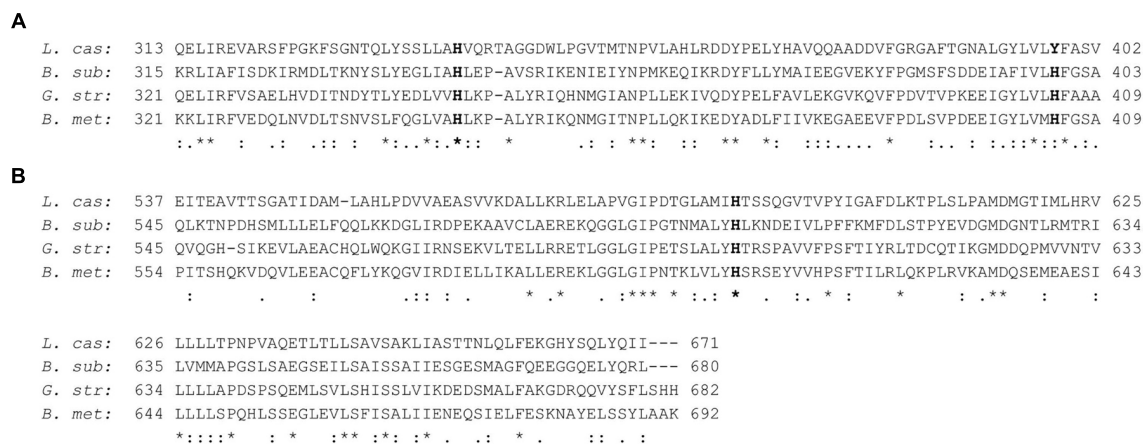


FIGURE 6 | Amino acid sequence alignment of the PRD2 (A) and EII^{MII} (B) domains of various MtlR proteins. The known regulatory sites are in boldface, conserved sequence (*), conservative mutations (:), semi-conservative mutations (.), and non-conservative mutation (). The alignment was performed with T-Coffee (Notredame et al., 2000). The GenBank accession numbers of the sequences are as follows: *L. cas*: *Lactobacillus casei* BL23, FM177140.1; *B. sub*: *B. subtilis* ssp. *subtilis* str. 168, CP010052.1; *G. str*: *Geobacillus stearothermophilus* ATCC 7954, U18943.1; *B. met*: *B. methanolicus* MGA3, CP007739.1.

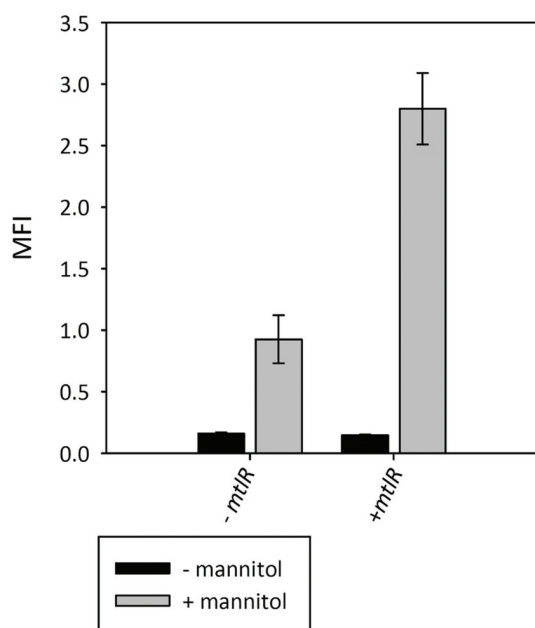


FIGURE 7 | Effect of plasmid-borne overexpression of *mtlR* on mannitol-inducible *gfpuv* reporter gene expression from *mtlR* promoter variant m2p. MFI — median fluorescence intensity of GFPuv given in a.u. GFPuv fluorescence of exponentially growing cells was measured after 2 h incubation at 37°C and 200 rpm. Mean values and standard deviations of triplicate shake flask cultures are given.

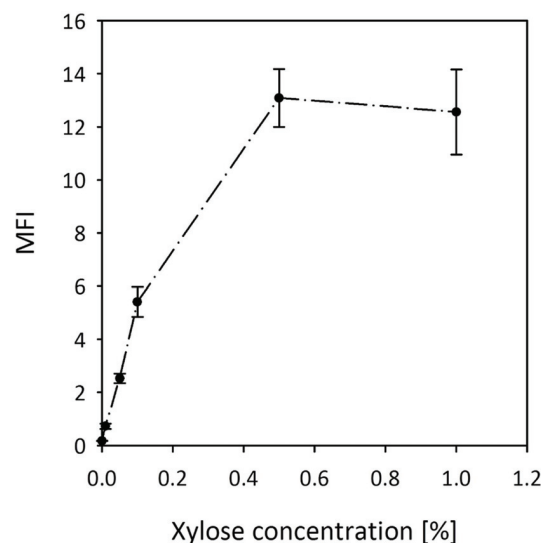


FIGURE 8 | Induction of reporter gene expression from the xylose inducible promoter of pTH1xpx-*gfpuv*. MFI — median fluorescence intensity of GFPuv given in a.u. GFPuv fluorescence of exponentially growing cells was measured after 2 h incubation at 37°C and 200 rpm. Mean values and standard deviations of triplicate shake flask cultures are presented.

the *B. megaterium xylA* gene into pTH1 to drive expression of *gfpuv*. The resulting plasmid was named pTH1xpx-*gfpuv*. **Figure 8** shows the expression levels of *gfpuv* transcribed from the xylose inducible promoter in media with different xylose concentrations. Reporter gene expression increased linearly in the concentration range between 0.01% (w/v) and 0.1%

(w/v) and reached a plateau at 0.5% (w/v). The fluorescence from fully induced xpx promoter is around 15-fold higher in comparison to the conventionally used *mdh* promoter (**Table 1**). Furthermore, the background gene expression with uninduced MGA3 (pTH1xpx-*gfpuv*) was very low (0.17 ± 0.01 a.u.) as compared to the background fluorescence (0.12 ± 0.00 a.u.) obtained for wild type *B. methanolicus* MGA3. Notably, mannitol did not induce expression of *gfpuv* from the xylose inducible promoter (data not shown).

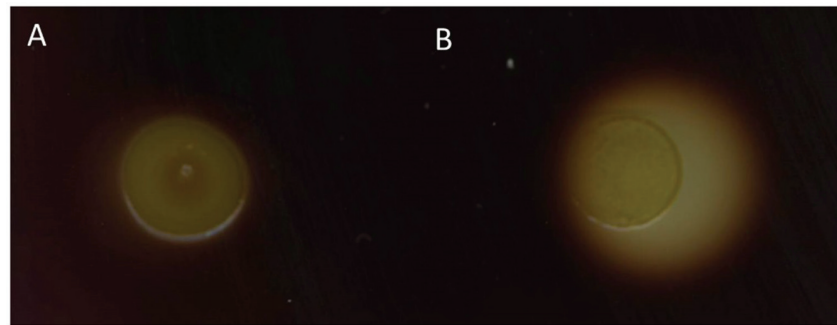


FIGURE 9 | Starch degradation by *B. methanolicus* strains. The strains were grown on LB plates supplemented with 0.5% starch and 0.05% xylose and iodine added to detect starch degradation. The dark area in the plate indicates presence of starch and the colorless halo around cells indicates starch degradation. (A) *B. methanolicus* MGA3 (pTH1mp) and (B) *B. methanolicus* MGA3 (pTH1xpx-amy).

Introduction of Heterologous Starch Degradation Pathway in *B. methanolicus* MGA3 by Heterologous Overexpression of α -Amylase Gene from *Streptomyces griseus* IMRU3570

α -Amylases degrade starch to glucose and expression of heterologous α -amylase genes in glucose-positive, but starch-negative species enabled starch utilization as for example shown for *C. glutamicum* expressing α -amylase gene (*amy*) from *Streptomyces griseus* IMRU3570 (Seibold et al., 2006). A BLAST search of the *Bacillus methanolicus* genome revealed two genes putatively encoding α -amylases (BMMGA3_04340, BMMGA3_04345) and one coding for an α -glucosidase (Heggeset et al., 2012; Irla et al., 2014). For heterologous expression of *amy* from *S. griseus* plasmid pTH1xpx was used for xylose inducible expression in *B. methanolicus*. Starch degradation by the control strain *B. methanolicus* MGA3 (pTH1mp) on LB agar plates supplemented with 0.5% soluble starch and 0.05% xylose at 37°C was not observed (Figure 9A). By contrast *B. methanolicus* MGA3 (pTH1xpx-amy) showed a halo on starch LB plates containing xylose as an inducer and incubated at 37°C (Figure 9B) indicating that expression of *amy* from *S. griseus* plasmid allowed for starch degradation by recombinant *B. methanolicus*.

DISCUSSION

In this study we have developed a versatile toolbox for inducible gene expression in *B. methanolicus* from RC- and theta-replicating plasmids. As a test case, we have applied a theta-replicating plasmid for heterologous expression of the lysine decarboxylase gene from *E. coli* and have shown improved cadaverine (1,5-diaminopentane) production in methanol-controlled fed-batch fermentations.

Strain development for *B. methanolicus* until recently relied on (over-)expression of genes or operons from a single plasmid despite the need for gene co-expression from two different plasmids and for inducible gene expression (Brautaset et al.,

2010; Nærdal et al., 2015). To that end, we have extended the existing portfolio of available expression vectors (based on pHP13 and pNW33N) with the two additional replicons pUB110 and pHCMC04 (Gryczan et al., 1978; Cue et al., 1997; Nguyen et al., 2005; Nilasari et al., 2012). Plasmids pUB110, pHP13 and pNW33N replicate via a RC mechanism (Khan, 1997) and belong to the same plasmid family. This family is named pC194/pUB110 and is characterized by a similar *ori* sequence CTT(G)TTCTTTCTTATCTTGATA. However, they are known to have different copy numbers in *B. subtilis*. Typically, RC plasmids are known to replicate in thermophilic bacteria (Soutsehek-Bauer et al., 1987; Cue et al., 1997; Rhee et al., 2007). This, to the best of our knowledge, was not known for theta-replicating plasmids and we show here for the first time that the theta-replicating plasmid pHCMC04 replicates stably in the thermophilic *B. methanolicus*.

Cadaverine production by recombinant *B. methanolicus* expressing the *E. coli* lysine decarboxylase gene *cadA* was superior when using the theta-replicating plasmid pBV2mp-*cadA* (pHCM04 replicon) as compared to the rolling circle-replicating plasmid pTH1mp-*cadA* (pHP13 replicon) (Table 2). Despite the low copy number of the pHCMC04 replicon (approximately half of that for replicon pHP13) confirmed both by ddPCR and GFPuv fluorescence measurement (Table 1), cadaverine production by *B. methanolicus* MGA3 (pBV2mp-*cadA*) was about 55% higher (Table 2) than by MGA3 (pTH1mp-*cadA*). This observation results most probably from two factors: loss of the pHP13 replicon over cultivation time and high stability of the pHCMC04 replicon. The loss of the pHP13 replicon was somewhat surprising as it was reported to be stable in *B. methanolicus* (Cue et al., 1997). Plasmid pHP13 contains the *ori* sequence from parental plasmid pTA1060 (Haima et al., 1987), but is however lacking a 167-bp fragment outside of the *ori* sequence from pTA1060, which has been shown to improve stable plasmid segregation (Bron et al., 1987; Chang et al., 1987; Haima et al., 1987). By contrast, the theta-replicating plasmid pHCMC04 showed the expected high stability typically observed for this type of plasmids (Bruand et al., 1991; Titok et al., 2003; Nguyen et al., 2005). High plasmid stability is important in large-scale industrial processes requiring long seed trains or in fed-batch

and continuous cultivations since subpopulations of cells which have lost plasmids due to low segregational stability usually lead to significant productivity losses (Friebs, 2004). With respect to methanol utilization in the fed-batch cultivations, it was observed that the overall carbon consumption (from methanol) of strains MGA3 (pTH1mp-*cadA*) and MGA3 (pBV2mp-*cadA*) differed by less than 5% (data not shown). Thus, the finding that cadaverine production increased by 6.2 g/L and biomass formation was reduced by 4.6 g/L (Table 2), suggested a reallocation of carbon source utilization from biomass to product formation.

Here we have characterized the plasmid pUB110 as a very feasible choice for gene expression in thermophilic *B. methanolicus* for two reasons: it showed the highest copy number among the tested replicons (Table 1) and showed high segregational stability (Figure 1). In *B. subtilis*, the pUB110 plasmid is known as a high copy number plasmid (Gryczan et al., 1978), whereas segregational stability seems to be a more complex issue. It was shown in several studies that the wild type plasmid is stable over multiple generations in different *Bacillus* spp. including *B. subtilis* (Polak and Novic, 1982; Alonso et al., 1987; Shoham and Demain, 1990), *B. thuringiensis* (Naglich and Andrews, 1988) and *B. sphaericus* (Seyler et al., 1991). Nonetheless, molecular modifications may lead to decreased stability of pUB110 for several different reasons. The segregational instability may be the function of the insert size (Bron et al., 1988; Zaghloul et al., 1994) or the high expression level of the cloned gene (Vehmaanperii and Korhola, 1986). Moreover, the lack of the so-called BA3 and BA4 regions has been described to destabilize the plasmid (Tanaka and Sueoka, 1983; Bron and Luxen, 1985; Shoham and Demain, 1990). Despite the fact that the pUB110-derived plasmid (pUB110Smp) used in this study did not contain BA3 and BA4 sequences and contained a 2.4 kbp insert, it was stable over 60 generations in *B. methanolicus* (Figure 1). Taken together, in thermophilic *B. methanolicus* pUB110Smp seems to be more feasible for molecular cloning than the hitherto used pTH1mp replicon.

Additionally, pUB110Smp as well as the theta-replicating pBV2mp were shown to be compatible with pTH1mp and could be used for independent expression of two genes in a two-plasmid approach.

Bacillus methanolicus not only grows with mannitol as the carbon source, but also shows mannitol dependent induction of at least two promoters P_{mtlA} and P_{mtlR} as revealed by transcriptome and proteome analyses (Heggeset et al., 2012; Müller et al., 2014; Irla et al., 2015). As compared to growth on methanol, mannitol-grown cells showed about 20-fold higher abundances of the proteins involved in mannitol utilization, i.e., EIIA and EIIBC components of the mannitol-specific PTS and mannitol-1-phosphate 5-dehydrogenase (Müller et al., 2014). The genomic organization suggests monocistronic transcription of *mtlA* and co-transcription of *mtlRFD* although RNA-seq data also suggest co-transcription of *mtlFD* without *mtlR* (see Figure 2). Reporter gene expression from P_{mtlR} was higher than from P_{mtlA} . Background expression from P_{mtlR} was low and the induction when grown in the presence of mannitol was 6.5-fold for the native promoter and 13-fold for the improved version. These values are lower compared to mannitol inducible

promoters in *B. subtilis* and *Pseudomonas putida* which show induction rates of about 20 for the native promoters and up to 176 for modified versions (Heravi et al., 2011; Hoffmann and Altenbuchner, 2015). The threshold concentration of mannitol required for induction of the promoter *mtlR* in the vector pTH1m2p was about 12.5 mM which is higher than the Monod constant, i.e., the concentration supporting growth with mannitol with a half-maximal growth rate (about 0.5 mM). This different threshold may reflect basal expression of mannitol utilization genes (e.g., *mtlA*, *mtlFD* operon). Moreover, we used the promoter of the regulatory gene *mtlR* rather than a promoter of a structural gene (e.g., *mtlA*, *mtlFD* operon) and dose dependency of induction of *mtlA* or the *mtlFD* operon might differ from dose dependency of induction of *mtlR*. Thus, it is conceivable that when mannitol is present in limited concentrations in the environment the background expression of the mannitol utilization genes is sufficient for initial mannitol utilization, whereas only higher mannitol concentrations lead to autoinduction of *mtlR* expression.

Induction of P_{mtlR} was specific to mannitol, while similar sugar alcohols did not affect transcription. The fact that mannitol is one of the few carbon sources of *B. methanolicus* precludes its use as gratuitous inducers similar to IPTG in the *E. coli lac* system or the xylose system applied to *B. methanolicus* (see below).

The xylose inducible system originating from *B. megaterium* was previously successfully used in several bacterial species, including *B. megaterium* (Rygu and Hillen, 1991), *B. subtilis* (Kim et al., 1996), *Staphylococcus aureus* (Zhang et al., 2000), and *Brevibacillus choshinensis* (D'Urzo et al., 2013). Here we show that this system also works in *B. methanolicus*. Since xylose is not metabolized by *B. methanolicus* it serves as a gratuitous inducer in this bacterium. In fact, the xylose inducible system turned out to have multiple advantages, including very low background expression in the uninduced state, titratable induction, and a 75-fold induction window between the uninduced and the fully induced state. Similar high dynamic ranges of xylose induction have been reported for other *Bacillus* spp. (Kim et al., 1996; Zhang et al., 2000; Bhavsar et al., 2001). However, catabolite repression of the xylose inducible promoter in multiple *Bacillus* spp. is disadvantageous for biotechnological applications. Catabolite repression is due to the *cis*-acting catabolite responsive element (*cre*), which is a binding site of the catabolite repressor protein CcpA (Jacob et al., 1991; Lokman et al., 1994; Kim et al., 1996; Schmiedel and Hillen, 1996; Chaillou et al., 1998; Bhavsar et al., 2001; Miyoshi et al., 2004). To avoid this phenomenon, the *cre* sequence TGAAAGCGCAAACA of the *xyl* operon in *B. megaterium*, which is located within the *xylA* gene (Schmiedel and Hillen, 1996), is not present in the plasmids used here. The absence of the *cre* sequence from the plasmids may be relevant since the genome of *B. methanolicus* encodes a homolog of CcpA (BMMGA3_13325). A BLAST search did not indicate that the *cre* sequence TGAAAGCGCAAACA is present upstream of the genes for carbon source utilization (only methanol, glucose, or mannitol are known carbon sources) of *B. methanolicus*. A *cre* sequence may be present upstream of the putative glucosamine-6-phosphate synthetase encoding gene *glmS* (BMMGA3_01020). In *B. subtilis*, *glmS* mRNA acts as a metabolite-responsive

ribozyme (Winkler et al., 2004) and glucose-repressive *glmS* transcription is at least partially under CcpA-independent control (Yoshida et al., 2001). However, the regulatory mechanism of the homolog of CcpA (BMMGA3_13325) of *B. methanolicus* and its target genes need still to be defined. Plasmid pTH1xpx for xylose inducible gene expression was applied for heterologous expression of *Streptomyces griseus*-derived α -amylase gene in *B. methanolicus*. As shown before for mesophilic *Corynebacterium glutamicum* (Seibold et al., 2006), heterologous expression of the α -amylase gene from *S. griseus* supported starch degradation by recombinant *B. methanolicus* assayed at 37°C. While α -amylase from *S. griseus* was an obvious choice it has its limitations in thermophiles since the enzyme is known to exhibit maximal activity at 30°C with 92% of the remaining activity at 40°C, but only trace activity being observable at 50°C (Simpson and McCoy, 1953). Nonetheless, heterologous expression of *amy* from *S. griseus* serves as an example that the gene expression tools described here are suitable for pathway engineering of *B. methanolicus*.

Taken together, a series of plasmids for stable replication in the thermophilic *B. methanolicus* was developed for xylose as well as mannitol inducible gene expression. Thus, an important step for further advancing this thermophilic bacterium as a very promising candidate for industrial production of amino acids and their derivatives has been reached. Improved production of cadaverine using a theta-replicating plasmid for heterologous expression of the lysine decarboxylase gene from *E. coli* in methanol-controlled fed-batch fermentations was demonstrated as a first application example, starch degradation by recombinant *B. methanolicus* carrying xylose inducible expression plasmid pTH1xpx with the gene for α -amylase from *S. griseus* as a second example.

REFERENCES

- Alonso, J. C., Viret, J., and Tailor, R. H. (1987). Plasmid maintenance in *Bacillus subtilis* recombination-deficient mutants. *Mol. Gen. Genet.* 208, 349–352. doi: 10.1007/BF00330464
- Arfman, N., Dijkhuizen, L., Kirchhof, G., Ludwig, W., Schleifer, K. H., Bulygina, E. S., et al. (1992). *Bacillus methanolicus* sp. nov., a new species of thermotolerant, methanol-utilizing, endospore-forming bacteria. *Int. J. Syst. Bacteriol.* 42, 439–445. doi: 10.1099/00207713-42-3-439
- Bhavsar, A. P., Zhao, X., and Brown, E. D. (2001). Development and characterization of a xylose-dependent system for expression of cloned genes in *Bacillus subtilis*: conditional complementation of a teichoic acid mutant development and characterization of a xylose-dependent system for expression of. *Appl. Environ. Microbiol.* 67, 1–3. doi: 10.1128/AEM.67.1.403
- Brautaset, T., Jakobsen, Ø. M., Degnes, K. F., Netzer, R., Nærdal, I., Krog, A., et al. (2010). *Bacillus methanolicus* pyruvate carboxylase and homoserine dehydrogenase I and II and their roles for L-lysine production from methanol at 50°C. *Appl. Microbiol. Biotechnol.* 87, 951–964. doi: 10.1007/s00253-010-2559-6
- Brautaset, T., Jakobsen, Ø. M. M., Flickinger, M. C., Valla, S., and Ellingsen, T. E. E. (2004). Plasmid-dependent methylotrophy in thermotolerant *Bacillus methanolicus*. *J. Bacteriol.* 186, 1229–1238. doi: 10.1128/JB.186.5.1229
- Bron, S., Bosma, P., van Belkum, M., and Luxen, E. (1987). Stability function in the *Bacillus subtilis* plasmid pTA1060 since natural plasmids in *Bacillus subtilis*. *Plasmid* 18, 8–15. doi: 10.1016/0147-619X(87)90073-4
- Bron, S., and Luxen, E. (1985). Segregational instability of pUB110-derived recombinant plasmids in *Bacillus subtilis*. *Plasmid* 14, 235–244. doi: 10.1016/0147-619X(85)90007-1
- Bron, S., Luxen, E., and Swart, P. (1988). Instability of recombinant pUB110 plasmids in *Bacillus subtilis*: plasmid-encoded stability function and effects of DNA inserts. *Plasmid* 19, 231–241. doi: 10.1016/0147-619X(88)90041-8
- Bruand, C., Ehrlich, S. D., and Janniere, L. (1991). Unidirectional theta replication of the structurally stable *Enterococcus faecalis* plasmid pAM beta 1. *EMBO J.* 10, 2171–2177.
- Chaillou, P., Bor, Y., Batt, C. A., and Pouwels, P. H. (1998). Molecular cloning and functional expression in *Lactobacillus plantarum* 80 of xylT, encoding the D-xylose-H⁺ symporter of *Lactobacillus brevis*. *Appl. Environ. Microbiol.* 64, 4720–4728.
- Chalfie, M., Tu, Y., Ward, W. W., Euskirchen, G., and Prasher, D. (1994). Green fluorescent protein as a marker for gene expression. *Science* 263, 1258–1262. doi: 10.1126/science.8303295
- Chang, S., Chang, S.-Y., and Gray, O. (1987). Structural and genetic analyses of a par locus that regulates plasmid partition in *Bacillus subtilis*. *J. Bacteriol.* 169, 3952–3962.
- Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. C. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* 14, 315–319. doi: 10.1038/nbt0396-315
- Crutz-Le Coq, A. M., and Zagorec, M. (2008). Vectors for lactobacilli and other Gram-positive bacteria based on the minimal replicon of pRV500

AUTHOR CONTRIBUTIONS

MI, TH, IN, LP, TH, SL carried out the experimental procedure and the data analysis of the present study. MI prepared a draft of the manuscript. MI, TH, IN, SL, TB, and VW finalized the manuscript. TB and VW coordinated the study. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

This work was supported by the EU7 FWP project PROMYSE and the ERASysAPP project MetApp. MI acknowledges support from the CLIB Graduate Cluster Industrial Biotechnology at Bielefeld University, Germany which is financed by a grant from the Federal Ministry of Innovation, Science and Research (MIWF) of the federal state North Rhine-Westphalia, Germany. We acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University. We thank Dr. Oskar Zelder and Dr. Rober Thummer for providing the pUB110 plasmid and for scientific discussion, Dr. Oscar P. Kuipers for providing the pNZlacZ-plasmid and Dr. Anne-Marie Crutz-Le Coq for providing the pRV613 plasmid. Per O. Hansen, Elisabeth Elgsæter, Nils Kirschnick, Bin Liu, and Julia Koch are thanked for technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01481>

- from *Lactobacillus sakei*. *Plasmid* 60, 212–220. doi: 10.1016/j.plasmid.2008.08.002
- Cue, D., Lam, H., Dillingham, R. L., Hanson, R. S., and Flickinger, M. C. (1997). Genetic manipulation of *Bacillus methanolicus*, a gram-positive, thermotolerant methylotroph. *Appl. Environ. Microbiol.* 63, 1406–1420.
- D'Urzo, N., Martinelli, M., Nenci, C., Brettoni, C., Telford, J. L., and Maione, D. (2013). High-level intracellular expression of heterologous proteins in *Brevibacillus choshinensis* SP3 under the control of a xylose inducible promoter. *Microb. Cell Fact.* 12:12. doi: 10.1186/1475-2859-12-12
- Eikmanns, B. J., Thum-schmitz, N., Eggeling, L., Ludtke, K., and Sahm, H. (1994). Nucleotide sequence, expression and transcriptional analysis of the *Corynebacterium glutamicum* gltA gene encoding citrate synthase. *Microbiology* 140, 1817–1828. doi: 10.1099/13500872-140-8-1817
- Friebs, K. (2004). Plasmid copy number and plasmid stability. *Adv. Biochem. Eng. Biotechnol.* 86, 47–82. doi: 10.1007/b12440
- Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C. A. III, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 12–16. doi: 10.1038/NMETH.1318
- Gryczan, T. J., Contente, S., and Dubnau, D. (1978). Characterization of *Staphylococcus aureus* plasmids introduced by transformation into *Bacillus subtilis*. *J. Bacteriol.* 134, 318–329.
- Haima, P., Bron, S., and Venema, G. (1987). The effect of restriction on shotgun cloning and plasmid stability in *Bacillus subtilis* Marburg. *Mol. Gen. Genet.* 209, 335–342. doi: 10.1007/BF00329663
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580. doi: 10.1016/S0022-2836(83)80284-8
- Hanson, R. S., Dillingham, R. D., Olson, P., Lee, G. H., Cue, D., Schendel, F. J., et al. (1996). "Production of L-lysine and some other amino acids by mutants of *B. methanolicus*," in *Microbial Growth on C₁ Compounds*, Vol. 3, eds M. E. Lidstrom and F. Robert Tabita (Amsterdam: Springer), 227–236. doi: 10.1007/978-94-009-0213-8_31
- Heggeset, T. M., Krog, A., Balzer, S., Wentzel, A., Ellingsen, T. E., and Brautaset, T. (2012). Genome sequence of the thermotolerant *Bacillus methanolicus*: features and regulation related to methylotrophy and production of L-lysine and L-glutamate from methanol. *Appl. Environ. Microbiol.* 78, 5170–5181. doi: 10.1128/AEM.00703-12
- Heravi, K. M., Wenzel, M., and Altenbuchner, J. (2011). Regulation of mtl operon promoter of *Bacillus subtilis*: requirements of its use in expression vectors. *Microb. Cell Fact.* 10, 1–19. doi: 10.1186/1475-2859-10-83
- Higa, A., and Mandel, M. (1970). Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* 53, 159–162. doi: 10.1016/0022-2836(70)90051-3
- Hilker, R., Stadermann, K. B., Doppmeier, D., Kalinowski, J., Stoye, J., Straube, J., et al. (2014). ReadXplorer—visualization and analysis of mapped sequences. *Bioinformatics* 30, 2247–2254. doi: 10.1093/bioinformatics/btu205
- Hoffmann, J., and Altenbuchner, J. (2015). Functional characterization of the mannitol promoter of *Pseudomonas fluorescens* DSM 50106 and its application for a mannitol-inducible expression system for *Pseudomonas putida* KT2440. *PLoS ONE* 10:e0133248. doi: 10.1371/journal.pone.0133248
- IHS Chemical (2016). *IHS Chemical*. Available at: <http://www.businesswire.com/news/home/20151006005707/en/Global-Demand-Flavor-Enhancer-MSG-Grows-Incomes> [accessed April 28, 2016].
- Irla, M., Neshat, A., Brautaset, T., Rückert, C., Kalinowski, J., and Wendisch, V. F. (2015). Transcriptome analysis of thermophilic methylotrophic *Bacillus methanolicus* MGA3 using RNA-sequencing provides detailed insights into its previously uncharted transcriptional landscape. *BMC Genomics* 16:73. doi: 10.1186/s12864-015-1239-4
- Irla, M., Neshat, A., Winkler, A., Albersmeier, A., Heggeset, T. M., Brautaset, T., et al. (2014). Complete genome sequence of *Bacillus methanolicus* MGA3, a thermotolerant amino acid producing methylotroph. *J. Biotechnol.* 188C, 110–111. doi: 10.1016/j.jbiotec.2014.08.013
- Jacob, S., Allmansberger, R., Giirtner, D., and Hillen, W. (1991). Catabolite repression of the operon for xylose utilization from *Bacillus subtilis* W23 is mediated at the level of transcription and depends on a cis site in the xylA reading frame. *Mol. Gen. Genet.* 229, 189–196. doi: 10.1007/BF00272155
- Jakobsen, Ø. M., Benichou, A., Flickinger, M. C., Valla, S., Ellingsen, T. E., and Brautaset, T. (2006). Upregulated transcription of plasmid and chromosomal ribulose monophosphate pathway genes is critical for methanol assimilation rate and methanol tolerance in the methylotrophic bacterium *Bacillus methanolicus*. *J. Bacteriol.* 188, 3063–3072. doi: 10.1128/JB.188.8.3063
- Jakobsen, Ø. M., Brautaset, T., Degnes, K. F., Heggeset, T. M., Balzer, S., Flickinger, M. C., et al. (2009). Overexpression of wild-type aspartokinase increases L-lysine production in the thermotolerant methylotrophic bacterium *Bacillus methanolicus*. *Appl. Environ. Microbiol.* 75, 652–661. doi: 10.1128/AEM.01176-08
- Joyet, P., Derkaoui, M., Bouraoui, H., and Deutscher, J. (2015). PTS-mediated regulation of the transcription activator MtlR from different species: surprising differences despite strong sequence conservation. *J. Mol. Microbiol. Biotechnol.* 25, 94–105. doi: 10.1159/000369619
- Khan, S. A. (1997). Rolling-circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.* 61, 442–455.
- Kiefer, P., Schmitt, U., Müller, J. E. N., Hartl, J., Meyer, F., Ryffel, F., et al. (2015). DynaMet: a fully automated pipeline for dynamic LC-MS data. *Anal. Chem.* 87, 9679–9686. doi: 10.1021/acs.analchem.5b01660
- Kim, L., Mogk, A., and Schumann, W. (1996). A xylose-inducible *Bacillus subtilis* integration vector and its application. *Gene* 181, 71–76. doi: 10.1016/S0378-1119(96)00466-0
- Kovács, A. T., van Hartskamp, M., Kuipers, O. P., and van Kranenburg, R. (2010). Genetic tool development for a new host for biotechnology, the thermotolerant bacterium *Bacillus coagulans*. *Appl. Environ. Microbiol.* 76, 4085–4088. doi: 10.1128/AEM.03060-09
- Krog, A., Heggeset, T. M., Müller, J. E. N., Kupper, C. E., Schneider, O., Vorholt, J. A., et al. (2013). Methylotrophic *Bacillus methanolicus* encodes two chromosomal and one plasmid born NAD⁺ dependent methanol dehydrogenase paralogs with different catalytic and biochemical properties. *PLoS ONE* 8:e59188. doi: 10.1371/journal.pone.0059188
- Liu, H., and Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* 8:91. doi: 10.1186/1472-6750-8-91
- Lokman, B. C., Leer, R. J., van Sorge, R., and Pouwels, P. H. (1994). Promoter analysis and transcriptional regulation of *Lactobacillus pentosus* genes involved in xylose catabolism. *Mol. Gen. Genet.* 245, 117–125. doi: 10.1007/BF00279757
- Markert, B., Stolzenberger, J., Brautaset, T., and Wendisch, V. F. (2014). Characterization of two transketolases encoded on the chromosome and the plasmid pBM19 of the facultative ribulose monophosphate cycle methylotroph *Bacillus methanolicus*. *BMC Microbiol.* 14:7. doi: 10.1186/1471-2180-14-7
- Miyoshi, A., Jamet, E., Commissaire, J., Renault, P., Langella, P., and Azevedo, V. (2004). A xylose-inducible expression system for *Lactococcus lactis*. *FEMS Microbiol. Lett.* 239, 205–212. doi: 10.1016/j.femsle.2004.08.018
- Müller, J. E. N., Heggeset, T. M., Wendisch, V. F., Vorholt, J. A., and Brautaset, T. (2015a). Methylotrophy in the thermophilic *Bacillus methanolicus*, basic insights and application for commodity production from methanol. *Appl. Microbiol. Biotechnol.* 99, 535–551. doi: 10.1007/s00253-014-6224-3
- Müller, J. E. N., Meyer, F., Litsanov, B., Kiefer, P., and Vorholt, J. A. (2015b). Core pathways operating during methylotrophy of *Bacillus methanolicus* MGA3 and induction of a bacillithiol-dependent detoxification pathway upon formaldehyde stress. *Mol. Microbiol.* 98, 1089–1100. doi: 10.1111/mmi.13200
- Müller, J. E. N., Litsanov, B., Bortfeld-Miller, M., Trachsel, C., Grossmann, J., Brautaset, T., et al. (2014). Proteomic analysis of the thermophilic methylotroph *Bacillus methanolicus* MGA3. *Proteomics* 14, 725–737. doi: 10.1002/pmic.201300515
- Naglich, J. G., and Andrews, R. E. J. (1988). Tn916-dependent conjugal transfer of pC194 and pUB110 from *Bacillus subtilis* into *Bacillus thuringiensis* subsp. israelensis. *Plasmid* 20, 113–126. doi: 10.1016/0147-619X(88)90014-5
- Nærdal, I., Pfeifenschneider, J., Brautaset, T., and Wendisch, V. F. (2015). Methanol-based cadaverine production by genetically engineered *Bacillus methanolicus* strains. *Microb. Biotechnol.* 8, 342–350. doi: 10.1111/1751-7915.12257. *Microb. Biotechnol.* 2019, 12, 182–183.
- Nguyen, H. D., Nguyen, Q. A., Ferreira, R. C., Ferreira, L. C. S., Tran, L. T., and Schumann, W. (2005). Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability. *Plasmid* 54, 241–248. doi: 10.1016/j.plasmid.2005.05.001
- Nilasari, D., Rech, S., and Komives, C. (2012). Expression of recombinant green fluorescent protein in *Bacillus methanolicus*. *Biotechnol. Prog.* 28, 662–668. doi: 10.1002/btpr.1522

- Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. doi: 10.1006/jmbi.2000.4042
- Ochsner, A. M., Müller, J. E. N., Mora, C. A., and Vorholt, J. A. (2014). In vitro activation of NAD-dependent alcohol dehydrogenases by Nudix hydrolases is more widespread than assumed. *FEBS Lett.* 588, 2993–2999. doi: 10.1016/j.febslet.2014.06.008
- Polak, J., and Novic, R. P. (1982). Closely related plasmids from *Staphylococcus aureus* and soil bacilli. *Plasmid* 7, 152–162. doi: 10.1016/0147-619X(82)90074-9
- Rhee, M. S., Kim, J., Qian, Y., Ingram, L., and Shanmugam, K. (2007). Development of plasmid vector and electroporation condition for gene transfer in sporogenic lactic acid bacterium, *Bacillus coagulans*. *Plasmid* 58, 13–22. doi: 10.1016/j.plasmid.2006.11.006
- Rygas, T., and Hillen, W. (1991). Inducible high-level expression of heterologous genes in *Bacillus megaterium* using the regulatory elements of the xylose-utilization operon. *Appl. Microbiol. Biotechnol.* 35, 594–599. doi: 10.1007/BF00169622
- Sambrook, J., and Russell, D. (2001). *Molecular Cloning: A Laboratory Manual*, 3 Edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Schendel, F. J., Bremmon, C. E., Flickinger, M. C., and Guettler, M. (1990). L-lysine production at 50 degrees C by mutants of a newly isolated and characterized methylotrophic *Bacillus* sp. *Appl. Environ. Microbiol.* 56, 963–970.
- Schendel, F. J., Dillingham, R., Hanson, R. S., Sano, K., and Matsui, K. (2000). Production of glutamate using wild type bacillus methanolicus. US 6083728.
- Schmiedel, D., and Hillen, W. (1996). Contributions of XylR CcpA and cre to diauxic growth of *Bacillus megaterium* and to xylose isomerase expression in the presence of glucose and xylose. *Mol. Gen. Genet.* 250, 259–266. doi: 10.1007/s004380050075
- Schneider, J., and Wendisch, V. F. (2011). Biotechnological production of polyamines by bacteria: recent achievements and future perspectives. *Appl. Microbiol. Biotechnol.* 91, 17–30. doi: 10.1007/s00253-011-3252-0
- Seibold, G., Auchter, M., Berens, S., Kalinowski, J., and Eikmanns, B. J. (2006). Utilization of soluble starch by a recombinant *Corynebacterium glutamicum* strain: growth and lysine production. *J. Biotechnol.* 124, 381–391. doi: 10.1016/j.jbiotec.2005.12.027
- Seyler, R. W. J., Yousten, A. A., and Burke, W. F. Jr. (1991). Plasmid stability in *Bacillus sphaericus* 2362 during recycling in mosquito larvae. *J. Invertebr. Pathol.* 58, 362–366. doi: 10.1016/0022-2011(91)90181-O
- Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., and Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* 22, 1567–1572. doi: 10.1038/nbt1037
- Shimizu, H., Tanaka, H., Nakato, A., Nagahisa, K., Kimura, E., and Shioya, S. (2003). Effects of the changes in enzyme activities on metabolic flux redistribution around the 2-oxoglutarate branch in glutamate production by *Corynebacterium glutamicum*. *Bioprocess Biosyst. Eng.* 25, 291–298. doi: 10.1007/s00449-002-0307-8
- Shimizu, K. (2013). Metabolic regulation of a bacterial cell system with emphasis on *Escherichia coli* metabolism. *ISRN Biochem.* 2013:645983. doi: 10.1155/2013/645983
- Shoham, Y., and Demain, A. L. (1990). Effect of medium composition on the maintenance of a recombinant plasmid in *Bacillus subtilis*. *Enzym. Microb. Technol.* 12, 330–336. doi: 10.1016/0141-0229(90)90160-R
- Simpson, F. J., and McCoy, E. (1953). The amylases of five streptomycetes. *Appl. Microbiol.* 1, 228–236.
- Skjerdal, O. T., Sletta, H., Flenstad, S. G., Josefsen, K. D., Levine, D. W., and Ellingsen, T. E. (1996). Changes in intracellular composition in response to hyperosmotic stress of NaCl, sucrose or glutamic acid in *Brevibacterium lactofermentum* and *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 44, 635–642. doi: 10.1007/BF00172497
- Soutsehek-Bauer, E., Scholz, W., Grill, E., and Staudenbauer, W. L. (1987). Thermostability and superhelicity of plasmid DNA in *Bacillus stearothermophilus*. *Mol. Gen. Genet.* 209, 575–579. doi: 10.1007/BF00331165
- Stolzenberger, J., Lindner, S. N., Persicke, M., Brautaset, T., and Wendisch, V. F. (2013a). Characterization of fructose 1,6-bisphosphatase and sedoheptulose 1,7-bisphosphatase from the facultative ribulose monophosphate cycle methylotroph *Bacillus methanolicus*. *J. Bacteriol.* 195, 5112–5122. doi: 10.1128/JB.00672-13
- Stolzenberger, J., Lindner, S. N., and Wendisch, V. F. (2013b). The methylotrophic *Bacillus methanolicus* MGA3 possesses two distinct fructose 1,6-bisphosphatase aldolases. *Microbiology* 159, 1770–1781. doi: 10.1099/mic.0.067314-0
- Tanaka, T., and Sueoka, N. (1983). Site-specific in vitro binding of plasmid pUB110 to *Bacillus subtilis* membrane fraction. *J. Bacteriol.* 154, 1184–1194.
- te Riele, H., Michel, B., and Ehrlich, S. D. (1986). Single-stranded plasmid DNA in *Bacillus subtilis* and *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A.* 83, 2541–2545. doi: 10.1073/pnas.83.8.2541
- Titok, M. A., Chapuis, J., Selezneva, Y. V., Lagodich, A. V., Prokulevich, V. A., Ehrlich, S. D., et al. (2003). *Bacillus subtilis* soil isolates: plasmid replicon analysis and construction of a new theta-replicating vector. *Plasmid* 49, 53–62. doi: 10.1016/S0147-619X(02)00109-9
- Vehmaanperii, J. O., and Korhola, M. P. (1986). Stability of the recombinant plasmid carrying the *Bacillus amyloliquefaciens* a-amylase gene in *B. subtilis*. *Appl. Microbiol. Biotechnol.* 23, 456–461. doi: 10.1007/BF02346060
- Wendisch, V. F. (2014). Microbial production of amino acids and derived chemicals: synthetic biology approaches to strain development. *Curr. Opin. Biotechnol.* 30, 51–58. doi: 10.1016/j.copbio.2014.05.004
- Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428, 281–286. doi: 10.1038/nature02362
- Wu, T.-Y., Chen, C.-T., Liu, J. T.-J., Bogorad, I. W., Damoiseaux, R., and Liao, J. C. (2016). Characterization and evolution of an activator-independent methanol dehydrogenase from *Cupriavidus necator* N-1. *Appl. Microbiol. Biotechnol.* doi: 10.1007/s00253-016-7320-3 [Epub ahead of print].
- Yoshida, K., Kobayashi, K., Miwa, Y., Kang, C., Matsunaga, M., Yamaguchi, H., et al. (2001). Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis*. *Nucleic Acids Res.* 29, 683–692. doi: 10.1093/nar/29.3.683
- Zaghloul, T. I., Abdelaziz, A., and Mostafa, M. H. (1994). High level of expression and stability of the cloned alkaline protease (aprA) gene in *Bacillus subtilis*. *Enzyme Microb. Technol.* 16, 534–537. doi: 10.1016/0141-0229(94)90026-4
- Zahoor, A., Lindner, S. N., and Wendisch, V. F. (2012). Metabolic engineering of *Corynebacterium glutamicum* aimed at alternative carbon sources and new products. *Comput. Struct. Biotechnol. J.* 3:e201210004. doi: 10.5936/csbj.201210004
- Zhang, L., Fan, F., Palmer, L. M., Lonetto, M. A., Petit, C., Voelker, L. L., et al. (2000). Regulated gene expression in *Staphylococcus aureus* for identifying conditional lethal phenotypes and antibiotic mode of action. *Gene* 255, 297–305. doi: 10.1016/S0378-1119(00)00325-5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Irla, Heggeset, Nærdal, Paul, Haugen, Le, Brautaset and Wendisch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Complete Genome Sequence of Hyperthermophile *Dictyoglomus turgidum* DSM 6724™ Reveals a Specialized Carbohydrate Fermentor

Phillip J. Brumm^{1,2*}, Krishne Gowda^{2,3}, Frank T. Robb⁴ and David A. Mead^{2,5}

¹ C5-6 Technologies LLC, Fitchburg, WI, USA, ² DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA, ³ Lucigen Corporation, Middleton, WI, USA, ⁴ Department of Microbiology and Immunology, Institute of Marine and Environmental Technology, University of Maryland, Baltimore, MD, USA, ⁵ Varigen Biosciences Corporation, Madison, WI, USA

Here we report the complete genome sequence of the chemoorganotrophic, extremely thermophilic bacterium, *Dictyoglomus turgidum*, which is a Gram negative, strictly anaerobic bacterium. *D. turgidum* and *D. thermophilum* together form the *Dictyoglomi* phylum. The two *Dictyoglomus* genomes are highly syntenic, and both are distantly related to *Caldicellulosiruptor* spp. *D. turgidum* is able to grow on a wide variety of polysaccharide substrates due to significant genomic commitment to glycosyl hydrolases, 16 of which were cloned and expressed in our study. The GH5, GH10, and GH42 enzymes characterized in this study suggest that *D. turgidum* can utilize most plant-based polysaccharides except crystalline cellulose. The DNA polymerase I enzyme was also expressed and characterized. The pure enzyme showed improved amplification of long PCR targets compared to Taq polymerase. The genome contains a full complement of DNA modifying enzymes, and an unusually high copy number (4) of a new, ancestral family of polB type nucleotidyltransferases designated as MNT (minimal nucleotidyltransferases). Considering its optimal growth at 72°C, *D. turgidum* has an anomalously low G+C content of 39.9% that may account for the presence of reverse gyrase, usually associated with hyperthermophiles.

Keywords: *Dictyoglomus turgidum*, thermophile, biomass degradation, phage, *Dictyoglomi*, DNA polymerase, glucanase, reverse gyrase

OPEN ACCESS

Edited by:

Kian Mau Goh,
Universiti Teknologi Malaysia, Malaysia

Reviewed by:

Biswarup Mukhopadhyay,
Virginia Tech, USA
Ida Helene Steen,
University of Bergen, Norway

*Correspondence:

Phillip J. Brumm
pbrumm@c56technologies.com

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 28 July 2016

Accepted: 25 November 2016

Published: 20 December 2016

Citation:

Brumm PJ, Gowda K, Robb FT and
Mead DA (2016) The Complete
Genome Sequence of
Hyperthermophile *Dictyoglomus*
turgidum DSM 6724™ Reveals a
Specialized Carbohydrate Fermentor.
Front. Microbiol. 7:1979.
doi: 10.3389/fmicb.2016.01979

INTRODUCTION

Dictyoglomus species are genetically distinct and divergent from known taxa, and have been assigned to their own phylum, *Dictyoglomi* (Saiki et al., 1985; Euzéby, 2012). They have been cultivated from or detected in anaerobic, hyperthermophilic hot spring environments (Patel et al., 1987; Svetlichny and Svetlichnaya, 1988; Mathrani and Ahring, 1991; Kublanov et al., 2009; Gumerov et al., 2011; Kochetkova et al., 2011; Burgess et al., 2012; Sahm et al., 2013; Coil et al., 2014; Menzel et al., 2015) or isolated from paper-pulp factory effluent (Mathrani and Ahring, 1992), but only two *Dictyoglomus* species have been validly described in the literature (Saiki et al., 1985; Svetlichny and Svetlichnaya, 1988). Both strains grow up to 80°C, are Gram negative, and exhibit unusual morphologies consisting of filaments, bundles, and spherical bodies. The first described *Dictyoglomus* species, *Dictyoglomus thermophilum* was isolated from Tsuetate Hot Spring in Kumamoto Prefecture, Japan (Saiki et al., 1985). The genome of *D. thermophilum* has been

sequenced (Coil et al., 2014), and a number of potentially useful enzymes including amylase (Fukusumi et al., 1988; Horinouchi et al., 1988), xylanases (Gibbs et al., 1995; Morris et al., 1998), a mannanase (Gibbs et al., 1999) and an endoglucanase (Shi et al., 2013) have been cloned and characterized. The second described species, *Dictyoglomus turgidus*, was isolated from a hot spring in the Uzon Caldera, in eastern Kamchatka, Russia (Svetlichny and Svetlichnaya, 1988). The name *Dictyoglomus turgidus* was subsequently corrected to *Dictyoglomus turgidum* (Euzéby, 1998). Unlike *D. thermophilum*, *D. turgidum* was reported to grow on a wide range of substrates including starch, cellulose, pectin, carboxymethylcellulose, lignin, and humic acids, but not on pentose sugars such as xylose and arabinose (Svetlichny and Svetlichnaya, 1988). Because of the wide range of substrates utilized, *D. turgidum* was selected for enzyme library construction and carbohydrase screening (Brumm et al., 2011) as well as whole genome sequencing. Here we describe the complete genome sequence of *D. turgidum*, bioinformatic analysis of the metabolism of this unusual organism, and comparative analysis with the genome of *D. thermophilum*. We also present functional analysis of its DNA Pol I gene and a number of novel carbohydrases.

MATERIALS AND METHODS

D. turgidum strain 6724^T was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). 10G electrocompetent *E. coli* cells, pEZSeq (a lac promoter vector), Taq DNA polymerase and OmniAmp DNA polymerase were obtained from Lucigen, Middleton, WI. Azurine cross-linked-labeled polysaccharides were obtained from Megazyme International (Wicklow, Ireland). 4-methylumbelliferyl- β -D-cellobioside (MUC), 4-methylumbelliferyl- β -D-xylopyranoside (MUX), and 4-methylumbelliferyl- β -D-glucopyranoside (MUG) were obtained from Research Products International Corp. (Mt. Prospect, IL). CellLytic IIB reagent, pNP- β -glucoside, pNP- β -cellobioside, 4-methylumbelliferyl- α -D-arabinofuranoside (MUA), 4-methylumbelliferyl- β -D-lactopyranoside (MUL), 5-Bromo-4-chloro-3-indolyl α -D-galactopyranoside (X- α -Gal, XAG), and 5-Bromo-4-chloro-3-indolyl β -D-galactopyranoside (X-gal, XG) were purchased from Sigma-Aldrich (St. Louis, MO). All other chemicals were of analytical grade.

D. turgidum DSM 6724TM was obtained from the DSMZ culture collection and maintained on DSM Medium 516 reduced with Na₂S and N₂ at 75°C in Balch tubes with a headspace of N₂. Cultures grown in 1 L stoppered flasks were harvested for DNA preparation. YT plate media (16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl and 16 g/l agar) was used in all molecular biology screening experiments. Terrific Broth (12 g/l tryptone, 24 g/l yeast extract, 9.4 g/l K₂HPO₄, 2.2 g/l KH₂PO₄, and 4.0 g/l glycerol added after autoclaving) was used for liquid cultures.

A cell concentrate of *D. turgidum* strain 6724TM was lysed using a combination of SDS and proteinase (Sambrook et al., 1989) and genomic DNA was purified using phenol/chloroform extraction. The genomic DNA was precipitated, treated

with RNase to remove residual contaminating RNA, and fragmented by hydrodynamic shearing (HydroShear apparatus, GeneMachines, San Carlos, CA) to generate fragments of 2–4 kb. The fragments were purified on an agarose gel, end-repaired, and ligated into pEZSeq (Lucigen Corp., Middleton WI). The recombinant plasmids were then used to transform electrocompetent cells. A copy of the library containing the *Dictyoglomus turgidum* genomic DNA was submitted to the Joint Genome Institute of the Department of Energy for whole genome sequencing; a second copy of the library was used for carbohydrase screening experiments.

The genome of *D. turgidum* DSM 6724TM was sequenced at the Joint Genome Institute (JGI) using a combination of 3 and 8 kb DNA libraries. In addition to 20x Sanger sequencing, 454 pyrosequencing was done to a depth of 20x coverage. Draft assemblies were based on 32,817 total reads. The Phred/Phrap/Consed software package was used for sequence assembly and quality assessment (Ewing and Green, 1998; Gordon et al., 1998). After the shotgun stage, reads were assembled with parallel phrap. Possible mis-assemblies were corrected with Dupfinisher or transposon bombing of bridging clones. Gaps between contigs were closed by editing in Consed, custom primer walking or PCR amplification. A total of 80 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The completed genome sequence of *D. turgidum* DSM 6724TM contains 34,756 reads, achieving an average of 17.3x coverage. The Accession number for the complete genome is NC_011661.

Genes were identified using Prodigal (Hyatt et al., 2010) as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to assert a product description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE (Lowe and Eddy, 1997), RNAMmer (Lagesen et al., 2007), Rfam (Griffiths-Jones et al., 2003), TMHMM (Krogh et al., 2001), CRISPRFinder (Grissa et al., 2007), and signalP (Krogh et al., 2001). RAST annotations (Aziz et al., 2008) of *D. turgidum* and *D. thermophilum* were carried out in parallel to further clarify genomic relationships using SEED genome comparison tools (Overbeek et al., 2005).

The phylogeny of *D. turgidum* was determined using its 16S ribosomal RNA (rRNA) gene sequence as well as those of the most closely related 16S rRNA sequences identified by BLASTn. 16S rRNA gene sequences were aligned using MUSCLE (Edgar, 2004), pairwise distances were estimated using the maximum composite likelihood (MCL) approach, and initial trees for heuristic search were obtained automatically by applying the neighbor-joining method in MEGA7 (Kumar et al., 2016). The alignment and heuristic trees were then used to infer the phylogeny using the maximum likelihood method based on Tamura-Nei (Tamura and Nei, 1993; Tamura et al., 2011). The phylogeny of the reverse gyrase protein sequence was inferred

using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.99686421 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 7 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 3230 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

The *endo*-glucanase specificity of enzymes was determined in 0.50 ml of 50 mM acetate buffer, pH 5.8, containing 0.2% azurine cross-linked-labeled (AZCL) insoluble substrates and 50 μ l of clarified lysate. Each purified enzyme was evaluated for *endo*-activities using the following set of substrates: AZCL-arabinan (AR), AZCL-arabinoxylan (AX), AZCL- β -glucan (BG), AZCL-curdlan (CU), AZCL-galactan (GL), AZCL-galactomannan (GM), AZCL-hydroxyethyl cellulose (HEC), AZCL-pullulan (PUL), AZCL-rhamnogalacturonan (RH), and AZCL-xyloglucan (XG). Assays were performed at 70°C, with shaking at 1000 rpm, for 60 min in a Thermomixer R (Eppendorf, Hamburg, Germany). Tubes were clarified by centrifugation and absorbance values at 600 nm determined using a Bio-Tek EL₈800 plate reader. The *exo*-glucanase specificity of enzymes was determined by spotting 2.0 μ l of clarified lysate directly on agar plates containing 10 mM 4-methylumbelliferyl substrate. Plates were placed in a 70°C incubator for 60 min and then examined using a hand-held UV lamp and compared to negative and positive controls for fluorescence.

Amplification efficacy was compared between Dtur, Taq and OmniAmp DNA polymerases (DNAP) in side by side PCR reactions using four different sized amplicons (0.9, 2.8, 5.0, and 10.0 Kb). PCR reaction conditions contained 1–20 ng of template DNA, 2.5U of Taq DNAP or 5U Dtur or OmniAmp DNAP (Lucigen Corp.), 200 μ M dNTPs, and 0.5 μ M primers in a 50 μ l reaction. DNAP buffer (1X) contained 10 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM NH₂SO₄, 2 mM MgSO₄, 0.1% tritonX-100, and 15% sucrose. Cycling conditions were 94°C 2 min and 30 cycles of 94°C for 15 s, 60°C for 30 s, and 72°C for 1 min per kb. The templates and PCR primers are as follows: pUC19 0.9 kb amplicon primers (CCC CTA TTT GTT TAT TTT TCT AAA ATT CAA TAT GTA TCC GCT and TTA CCA ATG CTT AAT CAG TGA GGC ACC TAT CT), *E. coli* 2.8 kb amplicon primers (TAC TGT CTG CCA TGG TTC AGA TCC CCC AAA ATC CAC TTA TCC TTG TAG A and TTA TCT GTG GTC GAC TTA GTG CGC CTG ATC CCA GTT TTC GCC ACT CCC CA), *E. coli* 5 kb amplicon primers (TCT CTC CGA CCA AAG AGT TG and GAA ACA TTG AGC GAA GAG GA), and *E. coli* 10 kb amplicon primers (CTA TGA TTA TCT AGG CTT AGG GTC AC and CAG TGT AGA GAG ATA GTC AGG AGT TA).

Functional screening for active carbohydrase enzymes involved plating transformed *E. coli* cells containing 2–4 kb Dtur genomic DNA inserts in the pEZSeq vector on YT agar

containing IPTG (for *lacZ* promoter induction) and one of the fluorescent substrates MUC, MUG or MUX. A long wavelength UV lamp was used to locate colonies that were fluorescent, which were sequenced by Sanger chemistry to identify the gene. Genes identified in the functional screen as well as additional genes of interest from the completed genome were amplified without their respective signal sequence, ligated into pET28A, and transformed into BL21(DE3) *E. coli* competent cells. Recombinant clones were cultured overnight at 37°C, 100 rpm, in 100 ml Luria Broth containing 50 mg/l kanamycin. Expression was induced using 1 mM IPTG, and cultures were harvested 18 h after induction. Cells were pelleted by centrifugation, and the pellets were lysed using Cellytic B reagent. Proteins were purified using standard methods for His-tagged proteins (Spriestersbach et al., 2015), and their purity and identity verified by SDS PAGE.

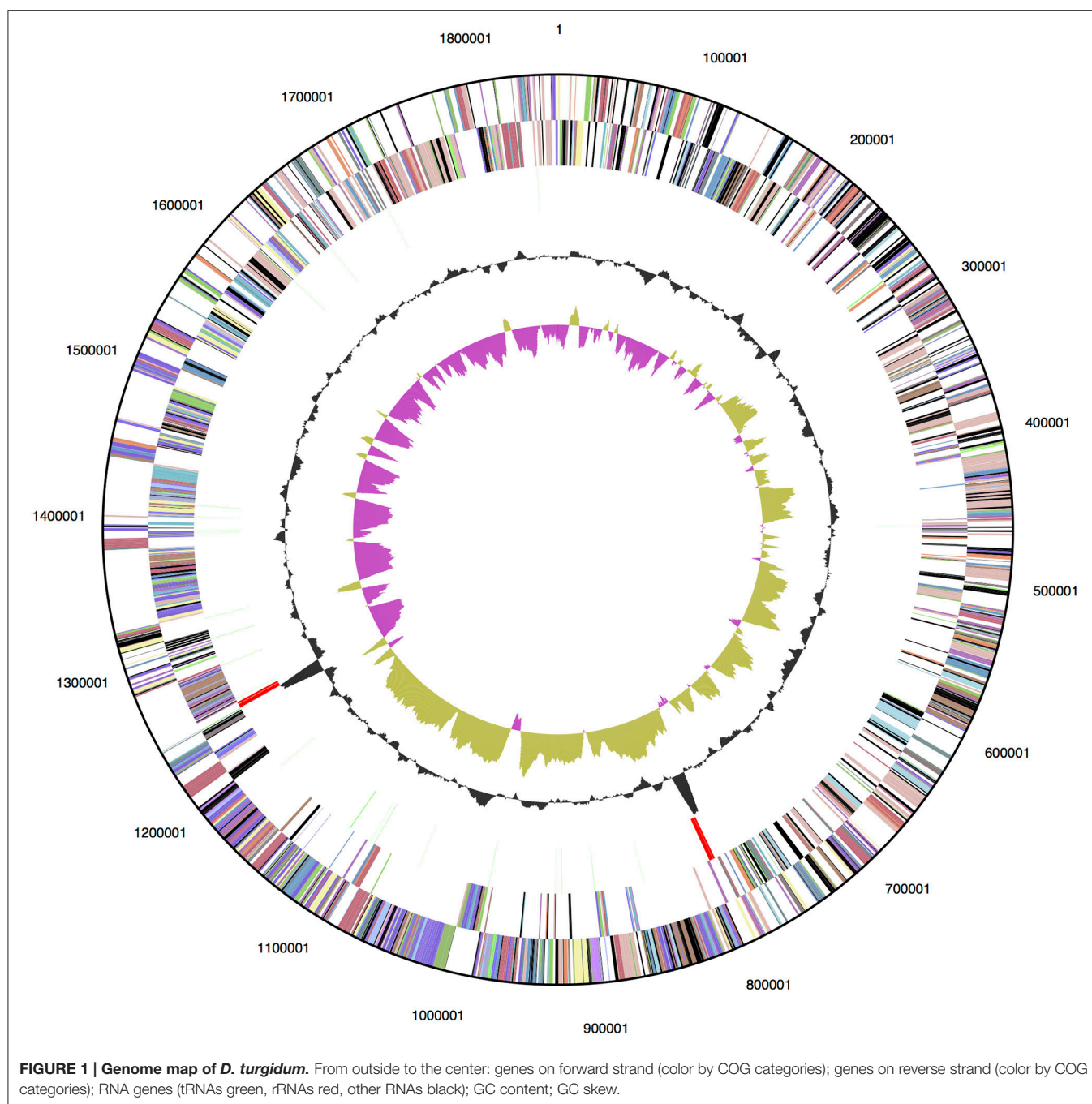
D. turgidum DNA polymerase I (Dtur DNAP) was cloned by PCR amplification using the proofreading enzyme Phusion (NEB, Waltham MA) and forward and reverse 24 base oligonucleotides that spanned the start and stop codons. The amplified DNA was inserted into the rhamnose promoter vector pRham containing an N terminal histidine tag and transformed into 10G competent *E. coli* cells (Lucigen Corp.). Recombinant Dtur DNAP production was induced by rhamnose and the enzyme was purified using standard methods for His-tagged proteins (Spriestersbach et al., 2015).

RESULTS

Genome of *D. turgidum*

The genome of *D. turgidum* DSM 6724TM consists of a single chromosome of 1,855,560 bp and no plasmids or extrachromosomal elements. The GC content of the chromosome is 33.96% based on the genome sequence, slightly higher than the reported value of 32.5% (Svetlichny and Svetlichnaya, 1988) and is predicted to contain 1813 protein-coding genes and 52 RNA genes (**Figure 1**). The completed genome sequence is available from GenBank (GenBank: CP001251.1). Based on 16S rRNA gene sequence analysis, *D. turgidum* DSM 6724 and *D. thermophilum* are separate species. This is confirmed by average nucleotide analysis (ANI), where *D. turgidum* and *D. thermophilum* are calculated to have 82.4% average nucleotide identity, below the threshold for members of the same species.

Of the 1813 protein-coding genes, 1354 genes (72.6%) were assigned to COGs categories (**Table 1**). The fraction of the genes annotated as members of COG class G, carbohydrate transport and metabolism (highlighted in bold), 13.4%, is greater than the fraction observed for 95% of genomes in the MicrobesOnline database (Dehal et al., 2010). This represents the lower limit of proteins involved in carbohydrate metabolism, because it does not include any proteins in categories R, S or not in COGS that were not identified by the algorithm as being involved in carbohydrate metabolism. A number of pectate lyases, for example, are not identified as members of COGs class G. No other COGs category had a significantly higher than average number of members, and no COGs category had a significantly lower than average percentage of members.



Genomic Insights into the Relationship of *D. turgidum* to *D. thermophilum* and Other Organisms

While being separate species, an in-depth comparison of the two *Dictyoglomi* genomes shows that *D. turgidum* is closely related to *D. thermophilum* on a number of levels. The genomes are similar in size, with *D. turgidum* being slightly smaller than the genome of *D. thermophilum* (1,855,560 bp vs. 1,959,987 bp) and containing approximately 100 fewer protein coding genes (1813

vs. 1912). The two organisms have a highly conserved set of genes present in their genomes. Over 95% of the proteins present in *D. turgidum* have orthologs in *D. thermophilum*. There are only 43 proteins of greater than 100 amino acids present in *D. turgidum* without orthologs in *D. thermophilum*, and there are only 109 proteins of greater than 100 amino acids present in *D. thermophilum* without orthologs in *D. turgidum*. Of the proteins with orthologs in both species, there are 614 proteins with >90% sequence identity.

TABLE 1 | Number of genes associated with general COG functional categories.

| Code | Value | Percentage | Description |
|----------|------------|--------------|--|
| J | 168 | 11.0% | Translation, ribosomal structure and biogenesis |
| K | 76 | 5.0% | Transcription |
| L | 61 | 4.0% | Replication, recombination and repair |
| B | 1 | 0.1% | Chromatin structure and dynamics |
| D | 19 | 1.2% | Cell cycle control, Cell division, chromosome partitioning |
| V | 40 | 2.6% | Defense mechanisms |
| T | 48 | 3.1% | Signal transduction mechanisms |
| M | 87 | 5.7% | Cell wall/membrane biogenesis |
| N | 20 | 1.3% | Cell motility |
| U | 18 | 1.2% | Intracellular trafficking and secretion |
| O | 61 | 4.0% | Posttranslational modification, protein turnover, chaperones |
| C | 79 | 5.2% | Energy production and conversion |
| G | 205 | 13.4% | Carbohydrate transport and metabolism |
| E | 170 | 11.1% | Amino acid transport and metabolism |
| F | 60 | 3.9% | Nucleotide transport and metabolism |
| H | 73 | 4.8% | Coenzyme transport and metabolism |
| I | 44 | 2.9% | Lipid transport and metabolism |
| P | 77 | 5.0% | Inorganic ion transport and metabolism |
| Q | 18 | 1.2% | Secondary metabolites biosynthesis, transport and catabolism |
| R | 130 | 8.5% | General function prediction only |
| S | 58 | 3.8% | Function unknown |
| – | 511 | 27.4% | Not in COGs |

Highlighted in bold, COG class G. The fraction of the genes annotated as members of this class is greater than the fraction observed for 95% of genomes in the MicrobesOnline database.

Synteny plots were generated using both RAST and IMG annotation methods. The two annotation methods gave essentially identical plots, as did plots based on DNA or protein sequences. The plots show the genomes of *D. turgidum* and *D. thermophilum* have highly conserved large and small-scale organization (**Figure 2A**). This conserved organization appears to be an unusual phenomenon. Two sets of thermophilic organisms with similar ANI values, *T. thermophilum* and *T. aquaticus* (84.3% ANI, **Figure 2B**) and *C. bescii* and *C. saccharolyticus* (82.0% ANI, **Figure 2C**) show only limited short-range synteny and no extensive long-range synteny. It is unclear if this conserved genomic organization is limited to these two species, or is present in all *Dictyoglomi* genomes.

The relationship of these two *Dictyoglomus* species to other organisms appears significantly more complicated, depending on the type of analysis and interpretation (Love et al., 1993; Rees et al., 1997; Takai et al., 1999; Ding et al., 2000; Wagner and Wiegel, 2008). Phylogenetic analysis using 16S rRNA shows the two *Dictyoglomus* species appear most closely related to *Thermotoga* species before bootstrapping (data not shown). After bootstrapping, the relationship shifts dramatically, with the two *Dictyoglomus* species becoming most closely related to *Caldicellulosiruptor* species (**Figure 3**). Previous work using average nucleotide identity (ANI) calculations (Nishida

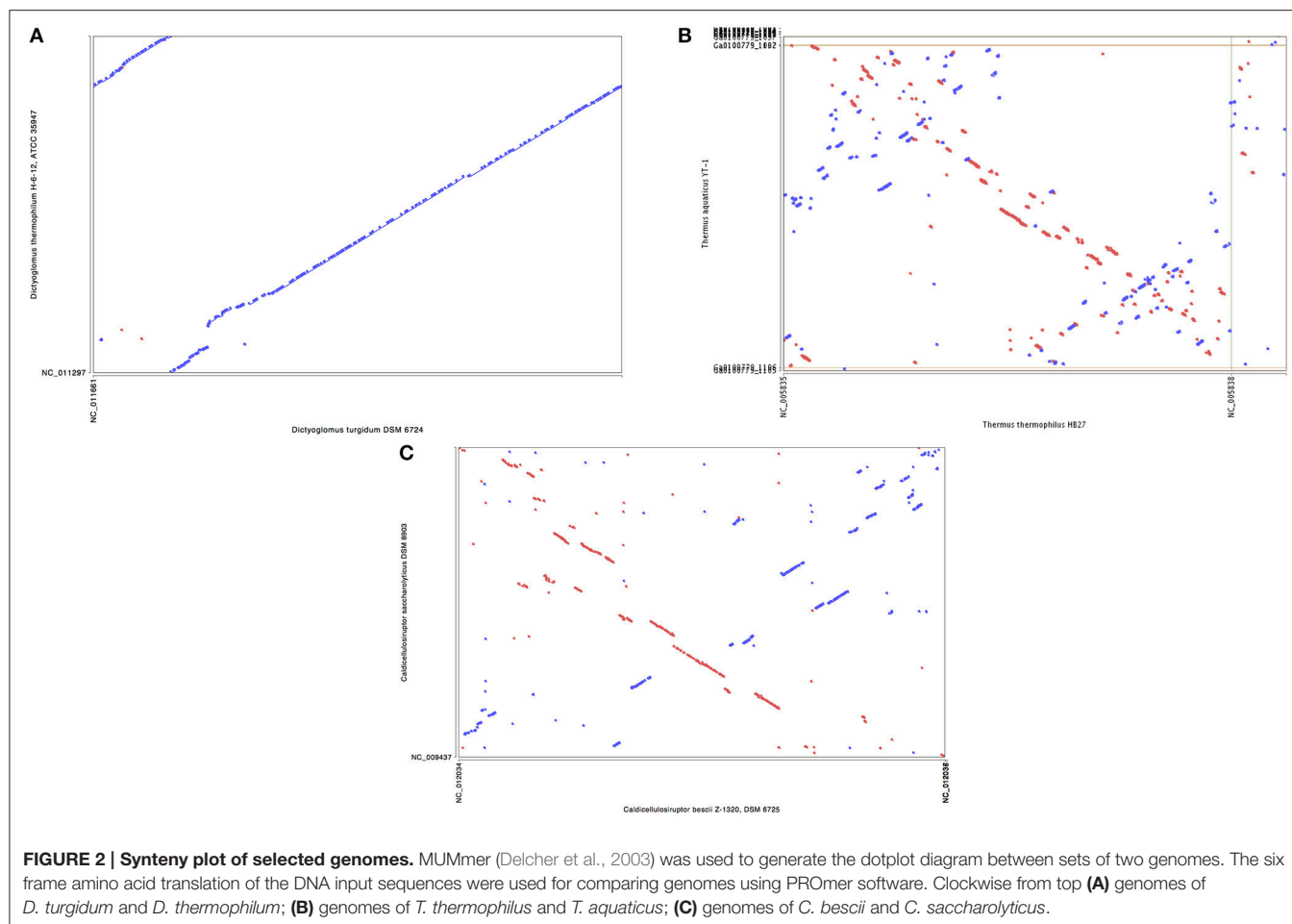
et al., 2011) identified *Thermotoga* species as the closest relatives to *Dictyoglomus*. ANI values were generated using the *D. thermophilum* genome, eight finished, closed *Thermotoga* genomes and three finished, closed *Caldicellulosiruptor* genomes. ANI values (Kim et al., 2014) were computed as pairwise bidirectional best nSimScan hits of genes having 70% or more identity and at least 70% coverage of the shorter gene. ANI calculations performed as described above yielded 82.4% identity between the genomes of *D. turgidum* and *D. thermophilum*, based on 1584 proteins (87% of the genome) that met the criteria. The value of 82.4% is well below the cut-off value of 98% for strains of the same species, and confirms that *D. turgidum* and *D. thermophilum* are separate species. The ANI calculations found 67–68% identity between *D. turgidum* and the three *Caldicellulosiruptor* species, based on 124–129 proteins per genome that met the criteria for the calculation (approximately 7% of the genome). ANI calculations found 66–68% identity between *D. turgidum* and the eight *Thermotoga* species, based on the 36–64 proteins per genome that met the criteria (approximately 2–4% of the genome). Rather than identifying relationships among these organisms, the low number of proteins in *D. turgidum* with at least 70% identity to the proteins in these 11 strains (on which these ANI values are calculated) further demonstrates the uniqueness of this organism.

Protein and Amino Acid Metabolism

Based on the MEROPS database (Rawlings et al., 2014), the *D. turgidum* genome codes for 55 potential peptidases. This value is within the range of peptidases reported in the database for *Thermotoga* species (52–67) and *Caldicellulosiruptor* species (54–74). Of the 55 potential peptidases, only a single peptidase, Dtur_0603, possesses an annotated signal sequence and is predicted to be secreted. While possessing only a single secreted peptidase to generate amino acids and peptides, *D. turgidum* possesses nine potential membrane transporter systems to transport amino acids and peptides into the cell. These nine transporters include seven annotated oligopeptide/dipeptide ABC transporter systems (Dtur_0082 through Dtur_0086; Dtur_0158 through Dtur_0162; Dtur_0214 through Dtur_0217; Dtur_0664 through Dtur_0668; Dtur_1061 through Dtur_1064; Dtur_1704 and Dtur_1707; Dtur_1719 through Dtur_1722) as well as two amino acid ABC transporter systems (Dtur_1051 through Dtur_1053 and Dtur_0932 through Dtur_0936).

D. turgidum appears to utilize the amino acids and peptides taken up for protein synthesis, but it is unable to metabolize most amino acids as an energy or carbon source. Based on the BioCyc (Karp et al., 2005; Caspi et al., 2014) and SEED (Devoid et al., 2013) metabolic reconstructions from the genome sequence, *D. turgidum* is lacking degradation pathways for the following 13 amino acids: aspartate, asparagine, cysteine, histidine, isoleucine, leucine, lysine, phenylalanine, proline, serine, tryptophan, tyrosine, and valine. Arginine is not metabolized, but may be converted to putrescine.

Only four amino acids appear to be metabolized by *D. turgidum*. Glutamate is converted to methyl aspartate using glutamate mutase (Dtur_1345 through Dtur_1347) and then to pyruvate and acetate. Threonine can be degraded to glycine and acetaldehyde via threonine aldolase (Dtur_0449), and the



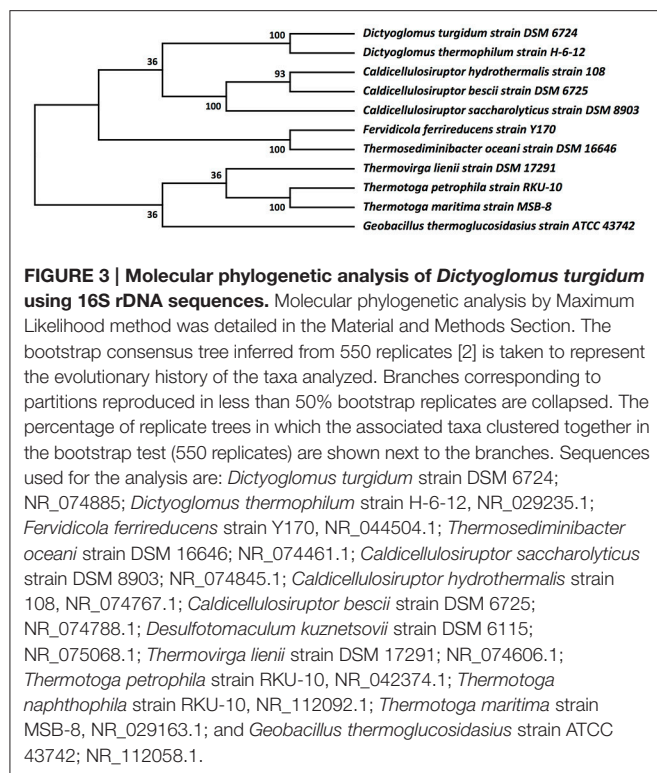
acetaldehyde generated is then converted to acetyl-CoenzymeA (acetyl-CoA) via aldehyde dehydrogenase (Dtur_0484). Alanine can be converted to pyruvate by alanine dehydrogenase (Dtur_1049), and glycine can be converted to ammonium 5,10-methylenetetrahydrofolate via glycine dehydrogenase and glycine cleavage system T protein (Dtur_1515 through Dtur_1518). The ability to utilize these four amino acids may be responsible for the observation of growth by *D. turgidum* on yeast extract, peptone, and casamino acids (Svetlichny and Svetlichnaya, 1988).

Monosaccharide Metabolism

Based on the genomic reconstruction of Dtur, the organism is able to metabolize most five and six carbon sugars, and the following pathways are predicted. Arabinose is utilized via isomerization to L-ribulose (Dtur_0379, or other isomerase), phosphorylation by L-ribulose kinase (Dtur_1748) to L-ribulose-5-phosphate, and isomerization by L-ribulose-5-phosphate-4-epimerase (Dtur_1734) to D-xylulose-5-phosphate, which is then metabolized via the pentose phosphate pathway. Rhamnose is utilized via isomerization by L-rhamnose isomerase to L-rhamulose (Dtur_0427), phosphorylation by L-rhamulose kinase (Dtur_1748) to L-rhamulose-1-phosphate, and cleavage into

dihydroxyacetone phosphate and L-lactaldehyde. Xylose is utilized via isomerization by xylose isomerase (Dtur_0036 or other sugar isomerase) to xylulose, and the xylulose is phosphorylated by xylulose kinase to (Dtur_0920) to D-xylulose-5-phosphate, which is then metabolized via the pentose phosphate pathway.

Fucose is utilized via isomerization by L-fucose isomerase to L-fuculose (Dtur_0410), phosphorylation by L-fuculokinase (Dtur_0920) to L-fuculose-1-phosphate, and cleavage into dihydroxyacetone phosphate and L-lactaldehyde. Galactose is phosphorylated by galactose kinase (Dtur_1195) to galactose-1-phosphate, which is converted to UDP-galactose by galactose-1-phosphate uridyl transferase (Dtur_1196), isomerized by UDP-glucose-4-epimerase (Dtur_1352) to UDP-glucose, and finally to glucose-1-phosphate by UTP-glucose-1-phosphate uridylyltransferase (Dtur_1627). Mannose is phosphorylated by mannose kinase (Dtur_0176; Dtur_0716 or other annotated sugar kinase) to generate mannose-1-phosphate. The mannose-1-phosphate is isomerized to mannose-6-phosphate by phosphomannomutase/phosphoglucosmutase (Dtur_0067) and then to fructose-6-phosphate by phosphoglucose/phosphomannose isomerase (Dtur_1271). UDP-glucose is either isomerized to fructose, or oxidized



to UDP-glucuronate using either Dtur_575 or Dtur_718. The UDP-glucuronate can then be further oxidized to ribulose-5-phosphate by 6-phosphogluconate dehydrogenase (Dtur_0197).

Galacturonate generated by pectin degradation may be epimerized by one of the six UDP sugar epimerase genes found in the genome. Rarely-encountered sugars may be handled by any of a number of sugar isomerases. Dtur rhamnose isomerase (Dtur_0427) isomerizes seven monosaccharides: L-rhamnose, L-lyxose, L-mannose, L-xylulose, L-fructose, D-allose, and D-ribose (Kim et al., 2013). The Dtur fucose isomerase (Dtur_0410) isomerizes L-fucose, D-arabinose, D-altrose, and L-galactose (Hong et al., 2012). Dtur also possesses a cellobiose 2-epimerase that may isomerize non-metabolized disaccharides into easily-degradable ones (Kim et al., 2012).

Polysaccharide Degradation and Transport

Polysaccharide degradation by *D. turgidum* is of interest for a number of reasons. Analysis of the *D. turgidum* genome shows an enrichment in COGS family members annotated as involved in carbohydrate transport and metabolism (Table 1). *D. turgidum* is reported to utilize polysaccharides such as starch, cellulose, pectin, glycogen, and carboxymethyl cellulose (Svetlichny and Svetlichnaya, 1988) while *D. thermophilum* is reported to utilize starch, but not cellulose. Finally, a number of carbohydrates with potential industrial applications have been identified in the two *Dictyoglomus* species including amylases and xylanases. A combination of genomic and enzymatic analyses was carried out to clarify the polysaccharide degradation potential of *D. turgidum*.

TABLE 2 | Annotated secreted polysaccharide-degrading enzymes.

| Gene | GH family | Annotated activity | Nearest ortholog | Identity |
|-----------|-----------|----------------------|------------------|----------|
| Dtur_0097 | GH 44 | β -mannanase | Calkro_0851 | 70.1% |
| Dtur_0172 | GH 28 | pectinase | Cphy_3310 | 47.1% |
| Dtur_0243 | GH 11 | xylanase | Calkro_0081 | 83.7% |
| Dtur_0276 | GH 5 | cellulase | Mahau_0466 | 59.9% |
| Dtur_0277 | GH 26 | β -mannanase | BG52_11385 | 52.7% |
| Dtur_0430 | PL 1 | pectate lyase | SNOD_03765 | 42.1% |
| Dtur_0431 | PL 1 | pectate lyase | M769_0111315 | 60.1% |
| Dtur_0432 | PLNC | pectate lyase | CSE_02370 | 57.3% |
| Dtur_0433 | CE 8 | pectin esterase | Calkro_0154 | 56.0% |
| Dtur_0628 | GH 12 | curdlanase | CTN_1107 | 48.4% |
| Dtur_0669 | GH 5 | cellulase | Mahau_0466 | 54.7% |
| Dtur_0675 | GH 57 | α -amylase | ANT_11030 | 41.3% |
| Dtur_0676 | CBM9 | α -amylase | COCOR_00322 | 39.7% |
| Dtur_0857 | GH 53 | β -galactanase | TRQ7_08325 | 56.5% |
| Dtur_1586 | GH 5 | cellulase | BSONL12_10711 | 41.5% |
| Dtur_1675 | GH 13 | α -amylase | CAAU_0986 | 51.6% |
| Dtur_1715 | GH 10 | xylanase | Pmob_0231 | 46.9% |
| Dtur_1729 | GH 43 | β -xylosidase | Csac_1560 | 67.9% |
| Dtur_1739 | GH 51 | β -xylosidase | Calhy_1625 | 58.9% |
| Dtur_1740 | GH 39 | β -xylosidase | TRQ7_03440 | 38.3% |

Analysis of the *D. turgidum* genome reveals a wide range of genes coding for annotated extracellular and intracellular polysaccharide degrading enzymes. The CAZy database (Lombard et al., 2014) identifies 57 glycosyl hydrolases (GH), 3 polysaccharide lyases (PL) and 6 carbohydrate esterases (CE) in the Dtur genome. Based on signal sequence predictions (Petersen et al., 2011), 20 of the polysaccharide-degrading enzymes are secreted into the medium (Table 2), where they degrade polysaccharides into oligosaccharides and monosaccharides. After polysaccharide degradation, 18 annotated three-component ABC carbohydrate transporters are predicted to transport monosaccharides and oligosaccharides into the cell. *D. turgidum* is reported to utilize fructose, glucose, rhamnose, inositol, mannitol, and sorbitol (Svetlichny and Svetlichnaya, 1988), indicating ABC carbohydrate transporters exist for these monosaccharides and sugar alcohols. *D. turgidum* cannot utilize arabinose, fucose, galactose, mannose, or xylose, indicating a lack of dedicated transport systems for these monosaccharides. These sugars may be transported into the cell as oligosaccharides by the oligosaccharide transporters and degraded to monosaccharides in the cytoplasm. Once inside the cell, oligosaccharides are degraded into monosaccharides by a combination of 46 *exo*-acting and *endo*-acting enzymes (Table 3). Working together, these 46 enzymes appear capable of degrading oligosaccharides from most plant-based polysaccharides to monosaccharides.

BLAST analysis was used to determine the closest orthologs of the 66 Dtur CAZymes. Of these 66 enzymes, 56 have their closest orthologs in *D. thermophilum*, with 80–90% amino acid identity. The remaining 10 enzymes have no orthologs in *D. thermophilum*. Seven of the ten unique enzymes in

TABLE 3 | Annotated intracellular polysaccharide-degrading enzymes.

| Gene | GH family | Annotated activity | Nearest ortholog | Identity |
|-----------|-----------|---------------------------------|-------------------------------------|----------|
| Dtur_0081 | GH 2 | β -galactosidase | Calhy_1828 | 60.9% |
| Dtur_0157 | GH 4 | α -glucosidase | Mc24_02443 | 47.5% |
| Dtur_0171 | GH 31 | α -glucosidase | A500_11654 | 44.9% |
| Dtur_0219 | GH 3 | β -glucosidase | <i>D. tunisiensis</i> bglB3 | 67.3% |
| Dtur_0222 | GH 20 | β -hexosaminidase | CDSM653_01797 | 67.2% |
| Dtur_0242 | CE NC | feruloyl esterase | TM_0033 | 55.1% |
| Dtur_0265 | CE 7 | acetyl xylan esterase | Tmari_0074 | 66.4% |
| Dtur_0289 | GH 3 | β -glucosidase | Cst_c03130 | 66.8% |
| Dtur_0315 | GH 29 | α -fucosidase | Tthe_0662 | 60.7% |
| Dtur_0320 | GH 31 | α -glucosidase | Csac_1354 | 65.9% |
| Dtur_0321 | GH 3 | β -glucosidase | Cst_c12090 | 50.1% |
| Dtur_0384 | GH 4 | α -glucosidase | CTER_5006 | 48.4% |
| Dtur_0435 | PL 1 | pectate lyase | MB27_42800 | 36.0% |
| Dtur_0440 | GH 4 | α -galacturonidase | BTS2_1711 | 61.6% |
| Dtur_0450 | CE 4 | deacetylase | Tnap_0743 | 67.4% |
| Dtur_0451 | GH 16 | curdlanase | TRQ7_04835 | 50.9% |
| Dtur_0462 | GH 1 | β -glucosidase | CLDAP_02840 | 48.5% |
| Dtur_0490 | GH 31 | α -glucosidase | Tbis_2416 | 45.4% |
| Dtur_0502 | GH 127 | β -L-arabinofuranosidase | CTN_0404 | 56.3% |
| Dtur_0505 | GH 42 | β -galactosidase | Mahau_1293 | 59.2% |
| Dtur_0523 | GH 18 | chitinase | Bccel_2454 | 50.1% |
| Dtur_0551 | GH 32 | invertase | Calhy_2186 | 47.6% |
| Dtur_0629 | GH 26 | β -mannanase | Calkro_1144 | 54.5% |
| Dtur_0650 | GH 31 | α -glucosidase | TheetDRAFT_1156 | 45.2% |
| Dtur_0658 | GH 130 | α -D-mannosyltransferase | X274_02975 | 41.2% |
| Dtur_0670 | GH 5 | cellulase | Mahau_0466 | 61.7% |
| Dtur_0671 | GH 5 | cellulase | TM_1752 | 58.7% |
| Dtur_0770 | GH 57 | α -amylase | BROSI_A0626 | 37.3% |
| Dtur_0794 | GH 13 | α -amylase | AC812_10325 | 35.3% |
| Dtur_0852 | GH 3 | β -glucosidase | M164_2324 | 58.2% |
| Dtur_0895 | GH 57 | α -amylase | TSIB_1115 | 46.5% |
| Dtur_0896 | GH 57 | α -amylase | Calab_2422 | 40.9% |
| Dtur_1539 | GH 2 | β -glucuronidase | Calkro_0120 | 60.7% |
| Dtur_1647 | GH 10 | xylanase | PaelaDRAFT_3013 | 51.2% |
| Dtur_1670 | GH 36 | α -galactosidase | Calla_1244 | 77.7% |
| Dtur_1677 | GH 4 | β -glucosidase | L21TH_1859 | 47.0% |
| Dtur_1714 | GH 67 | α -glucuronidase | Mc24_01903 | 69.4% |
| Dtur_1723 | GH 3 | β -glucosidase | <i>C. polysaccharolyticus</i> Xyl3A | 46.7% |
| Dtur_1735 | GH 51 | β -xylosidase | COB47_1422 | 70.2% |
| Dtur_1749 | GH 4 | α -glucosidase | TRQ7_00895 | 68.7% |
| Dtur_1758 | GH 38 | α -mannosidase | CTN_0786 | 41.3% |
| Dtur_1799 | GH 1 | β -glucosidase | Hore_15280 | 57.7% |
| Dtur_1800 | GH 43 | β -xylosidase | Athe_2555 | 82.9% |
| Dtur_1802 | GH 2 | β -galactosidase | Thewi_0408 | 42.2% |

D. turgidum are secreted enzymes, including three of the four predicted pectin-degrading enzymes, three of the four predicted xylan-degrading enzymes and the predicted *endo*-arabinase. Only two Dtur enzymes, Dtur_0243 and Dtur_1800, have non-*Dictyoglomus* orthologs with over 80% identity (Tables 2, 3). The nearest non-*Dictyoglomus* orthologs of most of the 66 have

<60% identity, showing the uniqueness of the Dtur enzymes. The wide range of organisms these orthologs are found in further demonstrates the uniqueness of this organism. Of the 66 enzymes, 13 have nearest orthologs in *Caldicellulosiruptor* species and 13 have nearest orthologs in *Thermotoga* species. Five orthologs are found in mesophilic *Clostridia* species,

four in mesophilic *Mahella* species, and three in thermophilic *Thermoanaerobacter* species. The remaining 28 orthologs are spread over a wide range of mesophilic and thermophilic organisms.

Degradation of Polymeric Substrates

Substrates Reported to Be Degraded for Which Genomic and Enzymatic Support Exists

Both *D. turgidum* and *D. thermophilum* are reported to utilize starch, and a number of α -amylases have been cloned and characterized from *D. thermophilum*. The genome of *D. turgidum* codes for three annotated extracellular α -amylases (Dtur_0675; Dtur_0676, and Dtur_1675) as well as four annotated intracellular α -amylases (Dtur_0770; Dtur_0794; Dtur_0895, and Dtur_0896) and six annotated β -glucosidases (Dtur_0157; Dtur_0171; Dtur_0320; Dtur_0384; Dtur_0490, and Dtur_1749). These intracellular enzymes may function in both degradation of starch oligosaccharides transported into the cell as well as degradation of glycogen stored in the cell.

D. turgidum is reported to utilize pectin, while no data on pectin utilization was reported for *D. thermophilum*. The genome of *D. turgidum* possesses three annotated secreted pectin lyases (Dtur_0430, Dtur_0431, and Dtur_0432), one secreted pectin esterase (Dtur_0433), an annotated intracellular pectin lyase (Dtur_0435), and an annotated intracellular α -galacturonidase (Dtur_0440).

D. turgidum is reported to utilize carboxymethyl cellulose. Because carboxymethyl cellulose is a man-made chemically-modified derivative of cellulose, there are no specific annotated carboxymethyl cellulases present in nature. Unlike cellulose which is crystalline and insoluble in water, carboxymethyl cellulose is an amorphous polymer that is soluble in aqueous solutions. As a result of this solubility, carboxymethyl cellulose is used as a substrate in the assay of a number of enzyme families including xylanases, cellulases, and β -glucanases. The genome of *D. turgidum* has two annotated, secreted xylanases (Dtur_0243 and Dtur_1715) and three secreted annotated cellulases (Dtur_0276; Dtur_0669, and Dtur_1586). The organism also possesses one annotated intracellular xylanase (Dtur_1647), two intracellular cellulases (Dtur_0670 and Dtur_0671) as well as six annotated β -glucosidases (Dtur_0219; Dtur_0289; Dtur_0321; Dtur_0462; Dtur_1723, and Dtur_1799). Assay of the enzymes expressed and purified in this work showed three cellulases (Dtur_0276; Dtur_0670 and Dtur_0671) and two xylanases (Dtur_1647 and Dtur_1715) utilized carboxymethyl cellulose as substrate, producing high levels of reducing sugars from a carboxymethyl cellulose solution (data not shown). These enzymatic assay results confirm the genomic analyses indicating that *D. turgidum* can utilize carboxymethyl cellulose.

Substrates Reported to Be Degraded for Which Genomic and Enzymatic Support Does Not Exist

D. turgidum is reported to utilize crystalline cellulose (Svetlichny and Svetlichnaya, 1988), though the authors report “the organism grew markedly less readily on microcrystalline cellulose, lignin and humic acids.” This is in contrast to

D. thermophilum, which is reported unable to utilize cellulose. Microbial degradation of crystalline cellulose requires the expression and secretion of multiple cellulases and accessory proteins to decrystallize the cellulose chains and generate soluble, low molecular weight cellooligosaccharides (Brumm, 2013). These cellooligosaccharides are then taken up via membrane transporters and further degraded into glucose monomers in the cytoplasm. Genomic and enzymatic analysis of *D. turgidum* indicate the organism is most likely unable to degrade crystalline cellulose. Comparison of the two genomes indicates *D. turgidum* contains no additional annotated cellulases not found in *D. thermophilum*. All five of the *D. turgidum* cellulases (Dtur_0276; Dtur_0669; Dtur_0670; Dtur_0671 and Dtur_1586) have orthologs in *D. thermophilum* (Dich_0008; Dich_0505; Dich_0506; Dich_0508 and Dich_1476, respectively). Analysis of the genome shows a lack of GH9, GH6, GH8, GH12, or GH48 cellulases found in truly cellulytic organisms (Brumm, 2013). Close examination of the genome reveals no cellulases containing CBM2 or CBM3 modules present in cellulose-degrading *Caldicellulosiruptor* species or cellosomal structures present in cellulose-degrading *C. thermocellum* species within the genome. Assay of the enzymes expressed and purified in this work showed three intracellular enzymes (Dtur_1647; Dtur_0670 and Dtur_0671) produced low levels of reducing sugars from crystalline cellulose (data not shown). The two secreted xylanases (Dtur_0276 and Dtur_1715) produced no reducing sugar from the crystalline cellulose. The lack of activity by the secreted enzymes confirms the genomic analyses indicating that *D. turgidum* most likely cannot utilize crystalline cellulose as a growth substrate. The microcrystalline cellulose preparation used in the original study may have contained glucan, xylan or mannan, resulting in the observed weak growth of the organism on cellulose used in the experiments (Svetlichny and Svetlichnaya, 1988).

The ability of *D. turgidum* to utilize lignin and humic acids is questionable for many of the reasons described above. The authors do not describe the source, purification, and analysis of the lignin and humic acids used in the growth experiments. Depending on the method of purification, lignin is often contaminated with mannan, cellulose and hemicellulose. Humic acids and lignin also contain sugars chemically bonded via ester linkages. Utilization of these sugars may be responsible for the low-level growth seen with these substrates. Thermophilic organisms capable of degrading aliphatic and aromatic organic compounds such as *Geobacillus* species, contain clearly identifiable extended gene clusters with these functions. For example, *Geobacillus* species Y41MC52 possesses three clusters annotated for degradation of aromatic acid molecules, GYMC52_1956 through GYMC52_1962; GYM C52_1990 through GYMC52_2001, and GYMC52_3134 through GYMC52_3141. Three similar clusters are found in the related strain *Geobacillus* species Y41MC61. Manual annotation of the *D. turgidum* genome failed to identify orthologs of any of the genes present in the three clusters, confirming that *D. turgidum* cannot utilize the aromatic ring structures found in lignin and humic acids.

Substrates Not Reported to Be Degraded for Which Genomic and Enzymatic Support Exists

No data was reported on xylan utilization by *D. turgidum*, however a xylanase has been cloned and expressed from *D. thermophilum*. The genome of *D. turgidum* has two annotated, secreted xylanases (Dtur_0243 and Dtur_1715) and three secreted β -xylosidases (Dtur_1729; Dtur_1739, and Dtur_1740). Genes for intracellular enzymes annotated as feruloyl esterase (Dtur_0242), acetyl xylan esterase (Dtur_0265), xylanase (Dtur_1647), α -glucuronidase (Dtur_1714), and two β -xylosidases (Dtur_1735 and Dtur_1800) may be involved in degradation of oligosaccharides derived from xylan.

Mannans and glucans comprise a diverse group of plant-based polysaccharides that share a β -linked hexose backbone. Among the members of these two groups are mannan, glucomannan, galactomannan, galactoglucomannan, β -glucan, curdlan, and xyloglucan. No data was reported on mannan or glucan utilization by either *D. turgidum* or *D. thermophilum*. The genome of *D. turgidum* codes for two annotated, secreted β -mannanases (Dtur_0097, and Dtur_0277) and one intracellular β -mannanase (Dtur_0629), three secreted annotated cellulases (Dtur_0276; Dtur_0669, and Dtur_1586) and two intracellular cellulases (Dtur_0670 and Dtur_0671), as well as six annotated β -glucosidases (Dtur_0219; Dtur_0289; Dtur_0321; Dtur_0462; Dtur_1723, and Dtur_1799).

Arabinogalactan is a polysaccharide found in many plants, with the highest concentration in larch wood. No data was reported on arabinogalactan utilization by either *D. turgidum* or *D. thermophilum*. The genome of *D. turgidum* codes for a secreted annotated β -galactanase (Dtur_0857) as well as four cytoplasmic β -galactosidases (Dtur_0081; Dtur_0081; Dtur_0505, and Dtur_1802) and one cytoplasmic α -galactosidase (Dtur_1670). Together these enzymes may be adequate for degradation of arabinogalactan as well as galactose-containing oligosaccharides. Annotated genes also code for intracellular fucosidase (Dtur_0315), invertase (Dtur_0551), galacturonidase (Dtur_0440) and β -glucuronidase (Dtur_1539), pectate lyase (Dtur_0435), and chitinase (Dtur_0523).

To verify the activities of some of these enzymes, cloning, expression, and purification was attempted for 30 of the annotated carbohydrase genes. Of these thirty, 16 Dtur enzymes were successfully expressed, purified, and characterized (Table 4). The 16 included two each of GH1 and GH3, four GH5, two GH10, and one each of GH36, GH42, GH43, GH53, GH57, and GH67. The remaining genes either failed to give amplicons of the correct size, or failed to express a soluble protein of the correct molecular weight (Brumm et al., 2011).

The two GH1 family members, Dtur_0462 and Dtur_1799, are annotated as β -glucosidases. These two cloned enzymes possess not only the predicted β -glucosidase activity, but also possess β -cellobiosidase, β -galactosidase, β -xylosidase and β -arabinofuranosidase activities (Table 4). GH 3 family members Dtur_0852 and Dtur_1723, also annotated as β -glucosidases, show β -glucosidase, β -xylosidase and β -arabinofuranosidase activity. Dtur_0852 also possesses β -cellobiosidase activity, which is absent in Dtur_1723 (Table 4).

TABLE 4 | Enzymatic activity of cloned gene products.

| Gene | GH family | Annotated activity | Substrates hydrolyzed |
|-----------|-----------|-------------------------|-----------------------------|
| Dtur_0462 | GH 1 | β -glucosidase | MUA, MUC, MUG, MUX, XG |
| Dtur_1799 | GH 1 | β -glucosidase | MUA, MUC, MUG, MUX, XG |
| Dtur_0852 | GH 3 | β -glucosidase | MUA, MUC, MUG, MUX |
| Dtur_1723 | GH 3 | β -glucosidase | MUA, MUG, MUX |
| Dtur_0276 | GH 5 | Cellulase | BG |
| Dtur_0669 | GH 5 | Cellulase | BG, GM |
| Dtur_0670 | GH 5 | Cellulase | AX, BG, GM, HEC, XG |
| Dtur_0671 | GH 5 | Cellulase | GM |
| Dtur_1647 | GH 10 | Xylanase | AX, ARA, BG, HEC, MUG, MUX |
| Dtur_1715 | GH 10 | Xylanase | AX |
| Dtur_1670 | GH 36 | α -galactosidase | XAG |
| Dtur_0505 | GH 42 | β -galactosidase | MUA, MUX |
| Dtur_1729 | GH 43 | α -arabinase | ARA |
| Dtur_0857 | GH 53 | β -galactanase | MUA, XAG |
| Dtur_0675 | GH 57 | α -amylase | PUL |
| Dtur_1714 | GH 67 | α -glucuronidase | Xylan, xylooligosaccharides |

Legend: MUC, 4-methylumbelliferyl- β -D-cellobioside; MUX, 4-methylumbelliferyl- β -D-xylopyranoside; MUG, 4-methylumbelliferyl- β -D-glucopyranoside; MUA, 4-methylumbelliferyl- α -D-arabinofuranoside; XAG, 5-Bromo-4-chloro-3-indolyl α -D-galactopyranoside; XG, 5-Bromo-4-chloro-3-indolyl β -D-galactopyranoside; AR, AZCL-arabinan; AX, AZCL-arabinoxylan; BG, AZCL- β -glucan; GL, AZCL-galactan; GM, AZCL-galactomannan; HEC, AZCL-hydroxyethyl cellulose; PUL, AZCL-pullulan; XG, AZCL-xyloglucan.

The four GH5 family annotated cellulases show a wide range of activities. Three of the GH5 family members hydrolyze a limited number of substrates. Dtur_0276 possesses only β -glucanase activity, while Dtur_0671 possesses only β -mannanase activity. Dtur_0669 possesses both β -mannanase and β -glucanase activities. None of the three possess β -glucosidase, β -cellobiosidase, β -galactosidase, or β -xylosidase activity. In contrast to these three enzymes, Dtur_0670 (Dtur CelA) possesses both endo-activity and exo-activity on a wide range of substrates (Table 4). Dtur_0670 possesses endoglucanase activity on a number of insoluble chromogenic substrates including AZCL-HE cellulose, AZCL- β -glucan, and AZCL-xyloglucan, endomannanase activity on AZCL-glucomannan, endoxylanase activity on AZCL-arabinoxylan, as well as β -glucosidase and β -cellobiosidase activity (Brumm et al., 2011). None of the GH5 family members released physiologically relevant amounts of sugar from crystalline cellulose even under prolonged incubation.

The two GH10 family xylanases show significantly different activities. Dtur_1715 possesses only endoxylanase activity, with no other detectable endo- or exo-activities. Dtur_1647 (XynA) displays endo-activity on β -(1,4)-linked pentose substrates such as xylan, arabinoxylan, and linear arabinan and β -(1,4)-linked hexose substrates such as β -glucan and hydroxyethyl cellulose. XynA also possesses β -glucosidase, β -xylosidase and β -cellobiosidase activity (Table 4).

The GH42 family member, Dtur_0857 predicted to be a β -galactanase, possesses no endo-activity, but instead possesses

β -galactosidase and β -arabinofuranosidase activities. The GH43 family member, Dtur_1729, does not possess β -xylosidase and β -arabinofuranosidase activity as expected from the annotation, but instead possesses only *endo*- β -arabinase activity. The GH57 family member, Dtur_0675, possesses α -amylase activity as predicted. The GH67 family member, Dtur_1714, shows strong α -glucuronidase activity on xylan and xylan oligosaccharides as predicted (Gao et al., 2011). Comparison to structural orthologs indicate that all 16 enzymes possess a single active site, with the differences in substrate range being a function of active site accessibility for each enzyme. There is no evidence of multiple active sites in any of the enzymes examined. The properties of these 16 enzymes show *D. turgidum* possesses the enzymes capable of hydrolyzing arabinoxylan, arabinan and arabinogalactan, β -glucan, mannan, and glucomannan to usable sugars. Failure to obtain active pectinase clones prevented us from confirming the ability of the organism to utilize pectin.

Energy Generation

Dtur is predicted to utilize the Embden–Meyerhof–Parnas pathway to produce ATP, reducing equivalents, and fermentation products from monosaccharides. Predicted products from pyruvate include lactate (Dtur_0700), acetate via acetyl-CoenzymeA (acetyl-CoA) (Dtur_0260; Dtur_0261, and Dtur_0262), and ethanol via acetyl CoA (Dtur_0260; Dtur_0261, and Dtur_0262) and acetaldehyde (Dtur_0484 and Dtur_1632). Hydrogen production is predicted by the presence of three hydrogenase gene clusters. The annotation reveals a partial *hypA* operon (Dtur_0074 through Dtur_0080) upstream of a hydrogenase assembly cluster (Dtur_0086 through Dtur_0090). Additional hydrogenase genes are located in a downstream cluster (Dtur_0556 through Dtur_0561). These metabolic predictions are in agreement with the published microbiological studies (Svetlichny and Svetlichnaya, 1988) showing the organism produces acetate, lactate, ethanol, CO₂, and H₂ during fermentation on sugars.

Based on the genome annotation, Dtur possesses an incomplete, reductive TCA cycle. This cycle allows the organism to convert acetate to pyruvate, oxaloacetate, and eventually to α -ketoglutarate. The α -ketoglutarate generated in this pathway can then be utilized for production of glutamate and other amino acids. Based on the annotation, the organism is unable to synthesize citrate from oxaloacetate and acetyl CoA.

Dtur has an extremely simple respiratory system. The genome codes for no respiratory cytochromes. The pathways for production of aminolevulinic acid from either glycine and succinyl-CoA or glutamate are both absent in Dtur. No tetrapyrroles (hemes, sirohemes, or corrinoids) are synthesized by Dtur, and the organism has no ABC-type heme transporters to utilize exogenous heme. The organism is also lacking the pathway for ubiquinone biosynthesis, indicating either ubiquinone is scavenged from the environment, or an alternate electron acceptor is utilized, like ferridoxin (Dtur_0730) or ferredoxin-like proteins (Dtur_0076; Dtur_0457; Dtur_0556; Dtur_0730; Dtur_0774, and Dtur_1717). The proton gradient needed for ATP generation is produced by NADH oxidoreductase (Dtur_0558; Dtur_0559; Dtur_0916; Dtur_0919,

and Dtur_1091), and succinate dehydrogenase (Dtur_0445). ATP is generated by proton translocation via an F₀F₁-type ATP synthase (Dtur_129 through Dtur_135). *D. turgidum* also possesses a V-type ATP synthase (Dtur_1499 through Dtur_1506), the function of which is unclear. The V-type ATP synthase may also be used for ATP generation, or may it may hydrolyze ATP to generate proton or ion gradients for transport.

Other Metabolic Pathways

As expected from its small genome size, *D. turgidum* does not possess a full set of biosynthetic and metabolic capabilities. *D. turgidum* appears able to synthesize all 20 amino acids from carbohydrate precursors, but as mentioned previously, is unable to metabolize the majority back to carbohydrates. Like the amino acid situation, *D. turgidum* is able to synthesize fatty acids, but not degrade them. The organism appears to synthesize folate, pyridoxal 5'-phosphate, thiamine, NAD from aspartate, and ascorbate from glucose or galactose, but is lacking pathways for biosynthesis of biotin, pantothenate or flavins.

Carbon monoxide is utilized by strict anaerobes via the Wood–Ljungdahl pathway (Techtmann et al., 2009), using the anaerobic CO dehydrogenase/acetyl CoA synthase complex. This complex catalyzes the complex multistep anaerobic reactions that include oxidizing CO to CO₂, formation of H₂, and biosynthesis of acetyl CoA. This pathway is found in thermophilic anaerobes such as *T. tengcongensis* and *M. thermoacetica* as well as in two *G. thermoglucosidasius* species. Manual curation of the genome indicates that *D. turgidum* does not possess the Wood–Ljungdahl pathway and is unable to utilize carbon monoxide as a carbon and energy source.

DNA Replication, Recombination, and Repair

IMG (DOE JGI) annotation methods identified 61 COG functional category L members (Table 1) for *D. turdigim*. Manual reannotation of each L category gene uncovered six mis-annotated genes. Five genes annotated as excinuclease ATPase subunits (Dtur_0247, 1011, 1053, 1153, 1667) are more likely ABC type transporters and an endonuclease (Dtur_0036) has supporting evidence to be annotated as a xylose isomerase. These genes were removed from the compilation shown in Table 5. A complete review of the annotated genes for *D. turdigim* identified a number of missed and overlooked genes that properly belong in the L COG family which totals 85 members in our revised tabulation of the genome (Table 5). Even though the 16S rRNA genes of the only two described species of *Dictyoglomus*, *D. turgidum* (CP001251) and *D. thermophilum* (CP001146), share 99% sequence identity, the divergence of their orthologous replication proteins is significant, as described below.

The genome of *D. turdigim* possesses 85 annotated genes for DNA replication, recombination and repair (Table 5), 78 of which have their closest ortholog to Diclth genes, whereas 6 have no orthologs in Diclth (Dtur_0102, 0317, 0545, 1284, 1294, 1297). The last four genes are part of a prophage that is not present in *D. thermophilum*. There are two genes in this category that are present in *D. thermophilum* but not *turdigim*, including a deoxyribodipyrimidine photolyase (Diclth_0072)

TABLE 5 | *D. turgidum* annotated DNA replication, recombination, and repair enzymes.

| Dtur gene | Annotation | Nearest neighbor/ % identity |
|-----------|---|--|
| 0745 | ATP-dependent DNA helicase | <i>Thermoanaerobaculum aquaticum</i> /43 |
| 1397 | ATP-dependent DNA helicase PcrA | <i>Caloranaerobacter azorensis</i> /49 |
| 1140 | ATP-dependent DNA helicase RecG | <i>Halothermothrix orenii</i> /52 |
| 0780 | ATP-dependent DNA ligase | uncultured <i>Acidobacteria</i> /56 |
| 1514 | Bacterial nucleoid DNA-binding protein | <i>Symbiobacterium thermophilum</i> /66 |
| 1151 | Cell division protein FtsK | <i>Thermoanaerobacterium aotearoense</i> /41 |
| 0001 | Chromosomal replication initiator protein DnaA | <i>Pelotomaculum thermopropioni</i> /59 |
| 1468 | Chromosome segregation and condensation protein ScpA | <i>Caldsalinibacter Kiritimati</i> /28 |
| 1467 | Chromosome segregation and condensation protein ScpB | <i>Dethiobacter alkaliphilus</i> /45 |
| 0947 | Chromosome segregation protein SMC | <i>Caloranaerobacter azorensis</i> /27 |
| 1104 | Competence protein ComE | <i>Clostridium purinilyticum</i> /29 |
| 1103 | Competence protein ComEA | <i>Clostridium</i> sp. CAG:273/55 |
| 1614 | Crossover junction endodeoxyribonuclease RuvC | <i>Alkaliphilus oremlandii</i> /50 |
| 0771 | Deoxyinosine 3' endonuclease (endonuclease V) | bacterium JGI-24/54 |
| 1268 | DNA and RNA Helicase | candidate division <i>Zixibacteria</i> /57 |
| 1547 | DNA gyrase B subunit | <i>Thermoanaerobacter</i> sp. YS13/66 |
| 1263 | DNA gyrase, A subunit | <i>Mahella australiensis</i> /54 |
| 1478 | DNA integrity scanning protein DisA | <i>Bacillus</i> sp. CMAA 1185/52 |
| 1078 | DNA mismatch repair protein MutL | <i>Caldanaerobacter subterraneus</i> /42 |
| 1077 | DNA mismatch repair protein MutS | <i>Clostridium thermocellum</i> /46 |
| 0884 | DNA or RNA helicase of superfamily II (UvrB) | <i>Caldanaerobacter subterraneus</i> /68 |
| 0104 | DNA polymerase beta domain-containing protein | <i>Syntrophaceticus schinkii</i> /57 |
| 0317 | DNA polymerase beta domain-containing protein | marine sediment metagenome/38 |
| 0545 | DNA polymerase beta domain-containing protein | <i>Microgenomates</i> bacterium/39 |
| 1295 | DNA polymerase beta domain-containing protein* | <i>Chloracidobacterium thermophilum</i> /60 |
| 0882 | DNA polymerase I 3'-5' exonuclease & polymerase domains | <i>Tepidanaerobacter acetatoydans</i> /44 |
| 1391 | DNA polymerase III alpha subunit | <i>Caloranaerobacter</i> sp. TR13/52 |
| 1105 | DNA polymerase III delta subunit | <i>Alkaliphilus oremlandii</i> /29 |
| 0257 | DNA polymerase III gamma/tau subunits | <i>Thermincola potens</i> /41 |
| 0789 | DNA polymerase III gamma/tau subunits | marine sediment metagenome/39 |

(Continued)

TABLE 5 | Continued

| Dtur gene | Annotation | Nearest neighbor/ % identity |
|-----------|---|--|
| 1551 | DNA polymerase III sliding clamp subunit beta | <i>Clostridium acidurici</i> /37 |
| 1600 | DNA Polymerase X | <i>Caldisericum exile</i> /55 |
| 1316 | DNA primase N | <i>Lachnospiraceae</i> bacterium/38 |
| 1527 | DNA protecting protein DprA | <i>Caldanaerobacter subterraneus</i> /43 |
| 1549 | DNA recombination protein RecF | <i>Clostridium acetici</i> /36 |
| 1625 | DNA repair exonuclease | <i>Thermobaculum terrenum</i> /33 |
| 0327 | DNA repair photolyase | <i>Parcubacteria</i> bacterium/41 |
| 0463 | DNA repair photolyase | <i>Caldicellulosiruptor owensensis</i> /66 |
| 1479 | DNA repair protein RadA | <i>Thermoanaerobacterium xylanolyticum</i> /48 |
| 0881 | DNA repair protein RADC | <i>Paenibacillus</i> sp./50 |
| 1047 | DNA repair protein RecN | <i>Caloramator australicus</i> /42 |
| 1321 | DNA repair protein RecO | <i>Microgenomates</i> bacterium/29 |
| 0015 | DNA replication and repair protein RecF | <i>Fervidobacterium nodosum</i> /80 |
| 1526 | DNA topoisomerase type I | <i>Carboxydotherrmus hydrogenofomans</i> /54 |
| 1522 | Double-stranded DNA repair protein Rad50 | <i>Thermoanaerobacter siderophilus</i> /26 |
| 1626 | Double-stranded DNA repair protein Rad50 | <i>Thermofilum</i> sp./25 |
| 0264 | Endonuclease IV | <i>Thermincola potens</i> /36 |
| 1485 | Excinuclease ABC subunit C | <i>Acetohalobium arabaticum</i> /45 |
| 0885 | Excinuclease ATPase subunit (ABC-ATPase UvrA) | <i>Thermosediminibacter oceani</i> /64 |
| 1613 | Holliday junction DNA helicase RuvA | <i>Thermodesulfobivrio yellowstonii</i> /34 |
| 1612 | Holliday junction DNA helicase RuvB | <i>Caldicellulosiruptor obsidiansis</i> /54 |
| 0792 | Holliday junction resolvosome helicase subunit | <i>Carboxydotherrmus hydrogenofomans</i> /54 |
| 1284 | Integrase* | <i>Acetothermus autotrophicum</i> /35 |
| 0886 | Methylated DNA-protein cysteine methyltransferase | <i>Anaerococcus lactolyticus</i> /40 |
| 1227 | Mg-dependent Dnase—deoxyribonuclease TatD | <i>Bacillus</i> sp. SA1-12/47 |
| 1308 | Mismatch repair ATPase (MutS family) | <i>Peptococcaceae</i> bacterium/40 |
| 1294 | Modification methylase, type III R/M system* | <i>Sulfurihydrogenibium yellowstonense</i> /65 |
| 0341 | N6-adenine-specific methylase | <i>Caloranaerobacter azorensis</i> /65 |
| 1141 | N6-adenine-specific methylase | <i>Thermoanaerobacter ethanolicus</i> /44 |
| 1497 | Poly(A) polymerase | <i>Peptococcaceae</i> bacterium/36 |
| 0683 | Predicted EndoIII-related endonuclease | <i>Thermodesulfatator indicus</i> /58 |
| 0846 | Predicted EndoIII-related endonuclease | <i>Candidatus Methanoperedens</i> sp./32 |
| 1024 | Predicted endonuclease involved in recombination | <i>Veillonella atypica</i> /44 |

(Continued)

TABLE 5 | Continued

| Dtur gene | Annotation | Nearest neighbor/ % identity |
|-----------|---|---|
| 1530 | Predicted endonuclease related to Holliday junction resolvase | <i>Caldicellulosiruptor saccharolyticus</i> /42 |
| 1709 | Predicted exonuclease | marine sediment metagenome/61 |
| 0102 | Predicted nucleic acid-binding protein contains PIN domain | <i>Leptospira wolbachii</i> /47 |
| 1441 | Primosomal protein N' (replication factor Y) | human gut metagenome/29 |
| 1507 | Putative chromosome partitioning protein | <i>Plasmodium knowlesi</i> /38 |
| 1162 | RecA/RadA recombinase | <i>Thermosediminibacter oceani</i> /73 |
| 1524 | Recombinase XerD | <i>Symbiobacterium thermophilum</i> /47 |
| 0259 | Recombination protein RecR | <i>Oxobacter pfennigii</i> /58 |
| 1358 | Replicative DNA helicase | <i>Acetohalobium arabaticum</i> /56 |
| 0014 | Reverse gyrase | <i>Fervidobacterium nodosum</i> /79 |
| 0708 | Ribonuclease HII | <i>Kosmotoga pacifica</i> /35 |
| 1531 | Ribonuclease HII | <i>Clostridium</i> sp./59 |
| 0118 | Single stranded nucleic acid binding protein (R3H domain) | <i>Butyrivibrio</i> sp./45 |
| 1602 | Single-stranded DNA exonuclease RecJ | <i>Clostridium thermocellum</i> /36 |
| 1362 | Single-stranded DNA-binding protein | <i>Syntrophothermus lipocalidus</i> /52 |
| 0786 | Single-stranded nucleic acid binding protein (R3H domain) | <i>Syntrophobacter fumaroxidans</i> /59 |
| 0880 | Smc1 chromosome segregation protein, putative | marine sediment metagenome/28 |
| 0202 | Thermostable 8-oxoguanine DNA glycosylase | <i>Thermotoga maritima</i> /59 |
| 1511 | Transcription-repair coupling factor | marine sediment metagenome/40 |
| 1297 | Type III restriction endonuclease subunit R* | bacterium JGI-6/67 |
| 1473 | Tyrosine recombinase XerD | <i>Tepidanaerobacter acetatoxydans</i> /51 |
| 1393 | Uracil-DNA glycosylase | <i>Nitrospira moscoviensis</i> /54 |

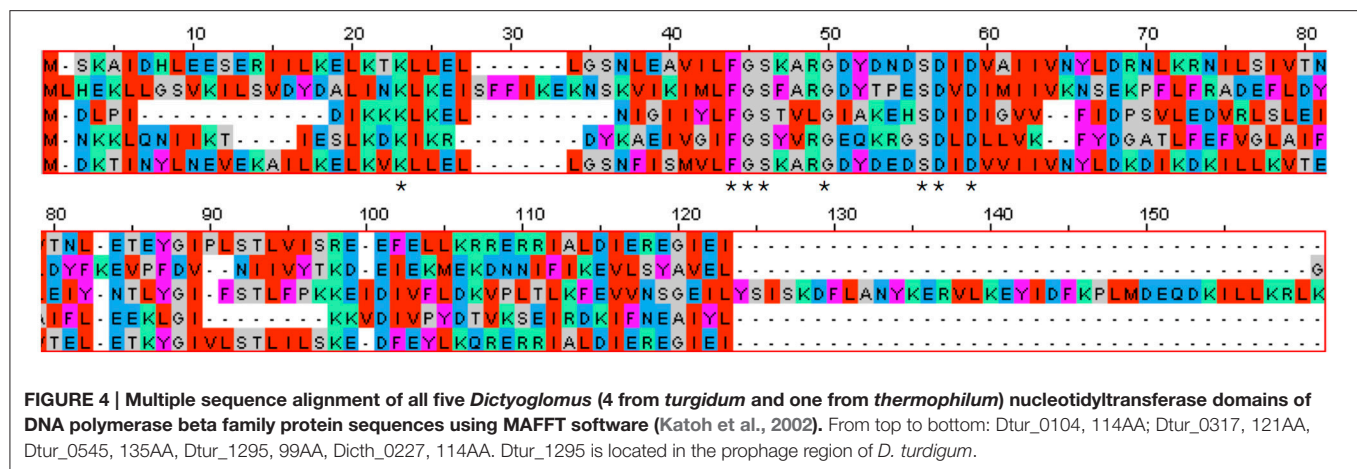
Prophage related genes are marked*.

and a DNA modification methylase (Dichth_0253). Only one gene (Dtur_1514), annotated as a bacterial nucleoid DNA-binding protein has 100% identity to its Dichth ortholog. On average the genes in this COG category share 85% identity with their Dichth counterparts, with the lowest at 66% identity (Dtur_1626, double-stranded DNA repair protein Rad50). Based on blastP analysis the most striking feature of this class of genes is how dissimilar they are to other sequenced genes and genomes in the database, other than *D. thermophilum*. This is apparent when the next nearest neighbors to *D. turgidum* DNA replication proteins are tabulated (Table 5). On average the genes in this COG category share 48% amino acid identity to nearest neighbor genes (25–80% range) and the cross section of homology is widespread among taxa that are primarily

anaerobic, thermophilic, or halophilic. Only two of the genes share homology to Thermotogae (Dtur_0202 and 0708) and 3 to *Caldicellulosiruptor* (Dtur_0463, 1530, and 1612). The greatest frequency of nearest neighbor orthologs after *D. thermophilum* are to clostridial (9%) and *Thermoanaerobacterium* (8%) genus members, followed by unknown metagenomic genes (6%). The low homology of *Dictyoglomus* replication proteins to orthologs in other organisms is another testament to how phylogenetical unique the genus is.

In spite of its preferred growth temperature of 72°C (Svetlichny and Svetlichnaya, 1988), *D. turgidum* has an extremely low (33.96 mol%) G+C content, which seems counterintuitive to genome stability and repair (Ishino and Narumi, 2015). Dtur does possess a reverse gyrase (Dtur_0014), a hallmark enzyme that is systematically present in all hyperthermophiles (Brochier-Armanet and Forterre, 2007), which introduces positive supercoils in DNA and thereby protects it from unwinding. Dtur and Dichth do not appear to contain genes for exonuclease III, a DNA-repair enzyme that hydrolyzes the phosphodiester bond 5' to an abasic site in DNA, which is commonly induced by heat. However, they both possess endonuclease IV, which has been shown to perform a similar abasic site processing function in *Thermotoga maritima* (Haas et al., 1999).

D. turgidum possesses 7 DNA replication and repair genes annotated to contain a nucleotidyltransferase (NT) domain, a superfamily that includes DNA polymerase beta domain-containing proteins (NT_Pol-beta), family X and poly-A DNA polymerases, as well as other proteins (Aravind and Koonin, 1999). The majority of the NTs are characterized by a distinct amino acid residue pattern, namely hG[GS]x(9,13)Dh[DE]h (x indicates any amino acid and h indicates a hydrophobic amino acid) that are essential for catalysis, which is true for all 7 NT domain containing genes in *D. turgidum*. Three of the *D. turgidum* NT domain-containing DNA repair proteins are larger than 450 amino acids (Dtur_0257, 1497, and 1600), whereas four members further annotated as belonging to the Pol-beta subfamily only encode 99–135 amino acids (Figure 4). DNA polymerase B is a proofreading-proficient enzyme thought to be involved with DNA repair activities in eubacteria (Wijffels et al., 2005) and replication in archaea (Kelman and Kelman, 2014), however there are no known thermophilic bacterial DNA polymerase B genes. A new, ancestral family of polB type nucleotidyltransferases designated as MNT (minimal nucleotidyltransferases) has been described (Aravind and Koonin, 1999) that are one-half to one-third the size of the larger orthologs. They are not uncommon as 258 cases can be found in the protein NCBI database “dna polymerase beta domain-containing protein” in 129 different microbes as of January 2016. However, there are no known biochemical studies showing whether these diminutive NTs are catalytically functional monomeric enzymes or whether they are part of a larger multimeric complex. The four NT_Pol-beta genes found in *D. turgidum*, one of which is associated with the prophage element discussed below, is an unsolved mystery as to the function these diminutive proteins might play,



particularly with regard to the lack of a PolB enzyme in this hyperthermophile.

Functional Analysis of *D. turgidum* DNA Polymerase I

D. turgidum possesses four sets of DNA polymerases: Pol X (Dtur_1600), Poly(A) polymerase (Dtur_1497), DNA polymerase I (Dtur_0882), and a minimal DNA polymerase III set of subunits (alpha, Dtur_1391; beta, Dtur_1551; delta, Dtur_1105, and gamma/tau, Dtur_0257 and _0789). In a survey looking for thermostable reverse transcriptase (RT) activity, the DNA polymerase I (PolI) from *Dictyoglomus thermophilum* strain Rt46B.1 has been cloned and expressed (Shandilya et al., 2004). While the enzyme did not exhibit RT activity, it did show significant thermal stability at 85°C compared to eight other enzymes being studied. Presumably its ortholog behaves the same, as Dcth_0729 shares 90% identity (770/856) and 96% positives (826/856) at the amino acid level with Dtur_0882.

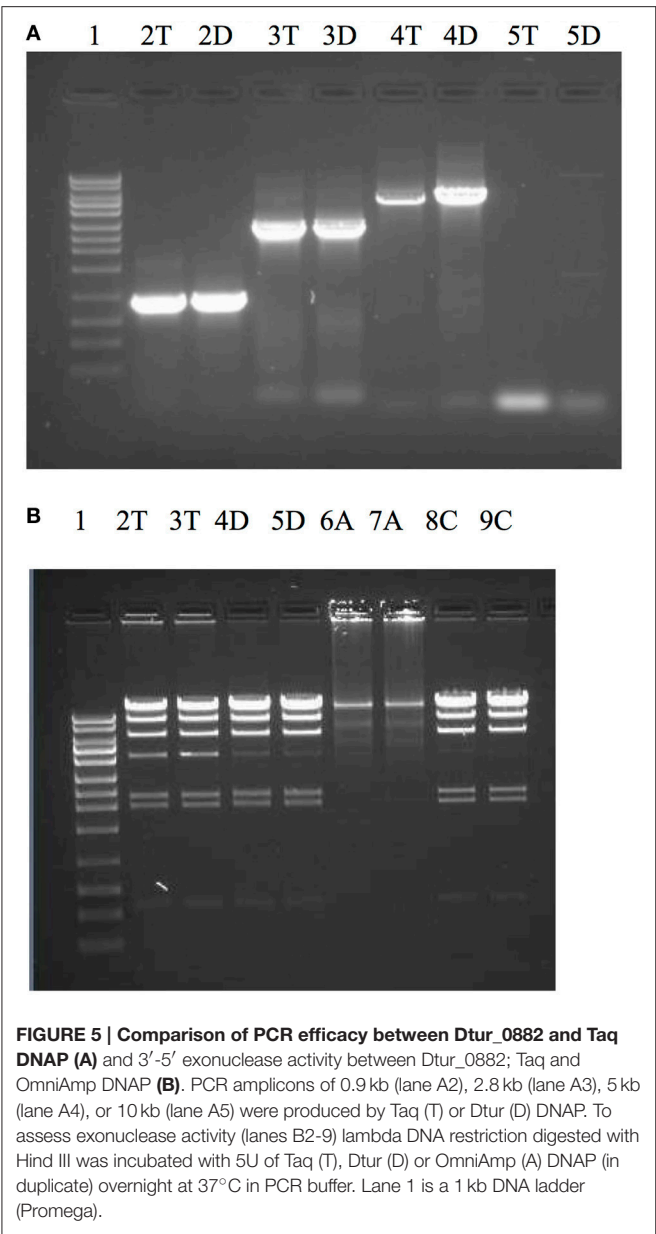
Dtur_0882 is a PolA type polymerase annotated to contain a 5'-3' and 3'-5' exonuclease domain in addition to the DNA-directed DNA polymerase domain. With the exception of *Rhodothermus marinus* PolA (PMID:11483153) and the enzyme OmniAmp polymerase (Chander et al., 2014) every thermostable PolA enzyme characterized thus far lacks a functional 3'-5' exonuclease domain associated with proofreading and enzyme fidelity. Because the high temperature growth conditions for *D. turgidum* are very similar to *Thermus aquaticus* (Taq), but the amino acid identities are so different between the two PolI enzymes (41%, 351/847), the utility of Dtur_0882 as a PCR enzyme was evaluated and compared with Taq DNAP (Figure 5). The 3'-5' exonuclease activity of both enzymes were also compared with the thermostable proof reading PolA enzyme called OmniAmp polymerase (Chander et al., 2014). Dtur_0882 produced the same yield of amplicon for the 0.9 and 2.8 kb reactions as Taq DNAP (Figure 5A lanes 2, 3), but was more efficient at amplifying the 5 and 10 kb primer templates (Figure 5A lanes 4, 5). As with Taq DNAP, Dtur_0882 does not appear to have any measurable 3'-5' exonuclease activity (Figure 5B lanes 2/3 compared to 4/5) as opposed to a strong

exonuclease activity demonstrated by the proofreader OmniAmp DNAP (Figure 5 lanes 6/7).

Prophage and CRISPR Elements

D. turgidum possesses two regions containing CRISPR repeats, suggesting the previous exposure to phage(s). The first CRISPR region, located between 470870 and 474424 nucleotides in the genome codes for 54 repeats. The repeat sequence is 30 nucleotides long, and the spacer average length is 36 nucleotides long. The second CRISPR region, located between nucleotides 615130 and 617311 in the genome codes for 33 repeats. The repeat sequence is the same as the first CRISPR region, and the spacer average length is 36 nucleotides long. No CRISPR-associated proteins are in the vicinity of the first CRISPR region. Upstream of the second CRISPR region are eight CRISPR-associated proteins (Table 5), while downstream of the second CRISPR region is a four-gene insert coding for biosynthetic enzymes (Dtur_0614–Dtur_0617) and eight additional CRISPR-associated proteins. *D. thermophilum* shows a similar organization of its CRISPR-associated proteins (Table 6), with a larger, 15-gene insert coding for biosynthetic enzymes (not shown).

An incomplete prophage sequence was identified by PHAST (Hubisz et al., 2011; Zhou et al., 2011) as a 17,407 base insert from 1,301,419 to 1,318,825 (Dtur_1284-1300) and confirmed as foreign DNA by IslandViewer Software (Hsiao et al., 2003) (1,298,422 to 1,317,048 containing genes Dtur_1282 through Dtur_1297). The prophage contains an integrase (Dtur_1284) followed by four annotated putative lipoproteins that potentially form part of a *beta*-barrel assembly machinery (Dtur_1285; Dtur_1286; Dtur_1288, and Dtur_1289) and a Type III restriction system (Dtur_1294 and Dtur_1297). The four annotated lipoprotein genes are closely related, as they share 79–88% amino acid identity. Inspection of the amino acid sequence reveals a unique periodicity of aspartic (D) and glutamic (E) residues to hydrophobic residues. The reason for this novel periodicity is due to the seven back to back, nearly perfect 49–52 amino acid tandem repeats found in these proteins (data not shown). Tandem repeat proteins are ubiquitous (Jernigan and



Bordenstein, 2015), but the unique signature sequence found here is only partially common to a handful of hypothetical proteins found in bacteria. Additional work is needed to clarify the function of these four repeated proteins with seven tandem internal repeats. This genomic island is unique to *D. turgidum*. *D. thermophilum* possesses an ortholog to only one of the four lipoproteins and no Type III restriction system proteins.

Thermophily, Stress Responses, and Heat Shock Proteins

D. turgidum in common with *D. thermophilum* has a complement of heat shock proteins typical of thermophilic bacteria, including a single GroEL/ES locus encoding Hsp60 and a DnaK/DnaJ locus encoding Hsp70 and cochaperones GrpE. Interestingly, in both

TABLE 6 | *D. turgidum* CRISPR-associated proteins.

| Gene | Annotation | Orthologs |
|-----------|---|------------|
| Dtur_0606 | CRISPR-associated protein, TM1812 family | Dicth_0458 |
| Dtur_0607 | CRISPR-associated protein DxTHG motif protein | Dicth_0187 |
| Dtur_0608 | CRISPR-associated RAMP protein, Csm5 family | Dicth_0186 |
| Dtur_0609 | CRISPR-associated RAMP protein, Csm4 family | Dicth_0185 |
| Dtur_0610 | CRISPR-associated RAMP protein, Csm3 family | Dicth_0184 |
| Dtur_0611 | CRISPR-associated RAMP protein, Csm2 family | Dicth_0183 |
| Dtur_0612 | CRISPR-associated RAMP protein, Csm1 family | Dicth_0182 |
| Dtur_0613 | CRISPR-associated protein, Csx3 family | Dicth_0181 |
| Dtur_0618 | CRISPR-associated protein, Cas2 family | Dicth_0165 |
| Dtur_0619 | CRISPR-associated protein, Cas1 family | Dicth_0164 |
| Dtur_0620 | CRISPR-associated endonuclease, Cas4 family | Dicth_0163 |
| Dtur_0621 | CRISPR-associated helicase, Cas3 family | Dicth_0162 |
| Dtur_0622 | CRISPR-associated protein, Cas5h family | Dicth_0161 |
| Dtur_0623 | CRISPR-associated protein, Csh2 family | Dicth_0160 |
| Dtur_0624 | CRISPR-associated protein, Csh1 family | Dicth_0159 |
| Dtur_0625 | CRISPR-associated protein, Cas6 family | Dicth_0158 |

D. turgidum and *D. thermophilum*, the reverse gyrase is encoded in a gene cluster shared with *recJ* and *revG*, closely linked to the *DnaK/DnaJ* operon. Heat shock regulation is enigmatic since the genome lacks both CIRCE elements and sigma32 SOS regulation. It is tempting to speculate that conditional expression of the reverse gyrase under high temperature growth conditions might be a mechanism for regulating DNA positive supercoiling in concert with the heat shock response. The phylogeny of the reverse gyrase is extraordinary. The reverse gyrase is most closely related to orthologs in *Fervidobacterium* species as shown in Figure 6. This phylogenetic position of the reverse gyrase does not conform to the 16S rRNA phylogeny (Figure 3) where *D. turgidum* is most closely related to *Caldicellulosiruptor* species. This suggests that lateral gene transfer of the reverse gyrase may have taken place.

Morphological Characteristics

Microbiological testing of *D. turgidum* indicate the organism stains Gram-negative (Svetlichny and Svetlichnaya, 1988). The *D. turgidum* genome contains a cluster of 10 genes potentially coding for outer membrane proteins characteristic of Gram-negative organisms and outer membrane lipid biosynthesis (Dtur_0814 through Dtur_0824). This cluster contains genes coding for a TamB (Dtur_0814), two BamA orthologs (Dtur_0815 and Dtur_0816), and two outer membrane chaperone Skp (OmpH) orthologs (Dtur_0817 and Dtur_0818), all potentially involved in outer membrane transport and assembly. This cluster of genes provides genomic support for the observed Gram-negative membrane structure observed in electron micrographs (Svetlichny and Svetlichnaya, 1988). Following these five genes, the *D. turgidum* genome contains a lipid biosynthesis cluster coding for the first four enzymes of Lipid A biosynthesis, LpxD (Dtur_0819), LpxC (Dtur_0820), LpxA (Dtur_0821), and (Dtur_0823) and an ortholog of FabZ (beta-hydroxyacyl-(acyl-carrier-protein) dehydratase,

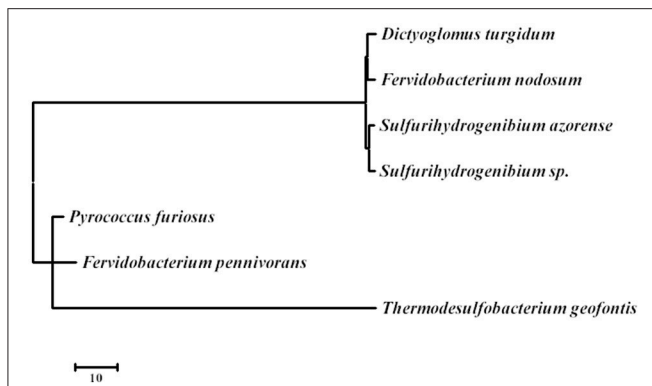


FIGURE 6 | Evolutionary relationships of reverse gyrase. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Tamura et al., 2011). The tree with the highest log likelihood (−22290.6302) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 7 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 1122 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). UniProt sequences used for the analysis were: B8DYH3, *Dictyoglomus turgidum* strain DSM 6724; A7HMS7, *Fervidobacterium nodosum* strain DSM 5306; H9UDK4, *Fervidobacterium pennivorans* strain DSM 9078; C1DT23, *Sulfurihydrogenibium azorense* strain DSM 15241; B2V6S9, *Sulfurihydrogenibium* sp. strain YO3AOP; F8C2X1, *Thermodesulfobacterium geofontis* strain OPF15, and P95479; *Pyrococcus furiosus* strain DSM 3638.

Dtur_0821). The final gene in the cluster (Dtur_0824) is a hypothetical protein related to LpxB. This 10-gene cluster has the identical organization in *D. thermophilum*, with individual genes averaging 80% identity to its Dtur ortholog. The cluster appears to be unique to *Dictyoglomus* species, as no similar cluster is found in any other sequenced organisms. The individual genes also show little homology to orthologs in other species, with observed amino acid identities being $\leq 30\%$ for all genes in the cluster.

Downstream of this ten-gene cluster is an annotated cluster of six proteins potentially involved in Gram-negative outer membrane efflux including orthologs of an ABC transporter ATP-binding protein (Dtur_0835), two of TolC (Dtur_0836 and Dtur_0837), HlyD (Dtur_0838) an ABC transporter permease protein (Dtur_0839) and a predicted transmembrane protein (Dtur_0840). An identical cluster is found in *D. thermophilum* (DICTH_0678 through DICTH_0683). A similar cluster is found in *Meiothermus taiwanensis* DSM 14542 as well as other *Meiothermus* and *Thermus* species. No orthologs of this cluster are found in any sequenced Thermotogales or Firmicutes species.

Many thermophilic bacteria possess complex morphologies, with varying shapes seen under different growth conditions. Examples of this include “rotund bodies” in *Thermus aquaticus* (Brumm P. J. et al., 2015), the outer membrane “toga” of *Thermotoga maritima* (Huber et al., 1986) and the multicellular spheres of *D. turgidum* (Svetlichny and Svetlichnaya, 1988). Regulation of these morphologies may be controlled by the action

of SpoVS (Brumm P. J. et al., 2015). *D. turgidum* possesses a gene coding for SpoVS (Rigden and Galperin, 2008) (Dtur_0800) similar to SpoVS proteins found in sporulating *Firmicutes* species as well as in the non-sporulating *Thermus-Deinococcus* and *Thermotoga* groups, but not in non-sporulating *Firmicutes* species. Phylogenetic reconstruction indicates that *D. turgidum* SpoVS is most closely related to the *D. thermophilum* SpoVS, (Figure 7) followed by the SpoVS of *Thermosediminibacter oceani* DSM 16646. These three SpoVS molecules form a separate clade from the SpoVS of the *Firmicutes* and *Deinococcus/Thermus* species. The SpoVS orthologs show much higher homology than orthologs of other *Dictyoglomus* proteins, suggesting an important conserved function. *Thermotoga* SpoVS orthologs have 53–78% identity, a *Clostridium thermocellum* ortholog has 74% identity, *Caldicellulosiruptor* orthologs have 65–71% identity, and *Thermus* orthologs have 56% identity to Dtur_0800. SpoVS may be an important regulator of cell morphology and differentiation in both sporulating thermophiles where it regulates the transition from vegetative growth to spore formation as well as the non-sporulating thermophiles where it regulates the transition from vegetative growth to formation of multiple morphologies (Brumm P. J. et al., 2015).

DISCUSSION

We report here the genome sequence, sequence analysis, and cloning of key enzymes of *D. turgidum*, an anaerobic, thermophile reported to degrade a wide range of biomass components including starch, cellulose, pectin and lignin [14]. This hyperthermophile has a small 1.8 M bp genome with a G+C content of the 34%. COGS analysis shows the organism is enriched in genes coding for carbohydrate transport and metabolism. While *Dictyoglomus* make up 25% of the species identified by 16S rRNA sequencing in some environments currently only two species, *D. thermophilum* (Saiki et al., 1985) and *D. turgidus* (Svetlichny and Svetlichnaya, 1988), corrected to *D. turgidum*., *D. turgidum*, and *D. thermophilum*, have been sequenced and annotated. This first comparison of the two genomes shows that, while the two organisms are unique species, they show extremely high levels of orthologous genes, average nucleotide identity, and synteny. No unique metabolic pathways are present in either organism. Approximately 1/3 of the proteins present in Dtur have orthologs in *D. thermophilum* with over 90% amino acid identity, and less than 10% of the proteins present in either genome have no ortholog in the other genome. The two organisms show extensive short-range and long-range synteny. Genome sequences of additional *Dictyoglomus* species are needed to determine if this is coincidence or a conserved feature of the *Dictyoglomi*. Additional work is also needed to confirm that the differences in synteny observed between the two genomes are real and are not artifacts of the assembly of the genomes.

The genome of *D. turgidum* provides insights into an organism that is strangely foreign and vaguely familiar at the same time. At first glance, the genome is remarkably unremarkable, containing no novel pathways or secondary products. In fact, the organism is lacking many of the pathways

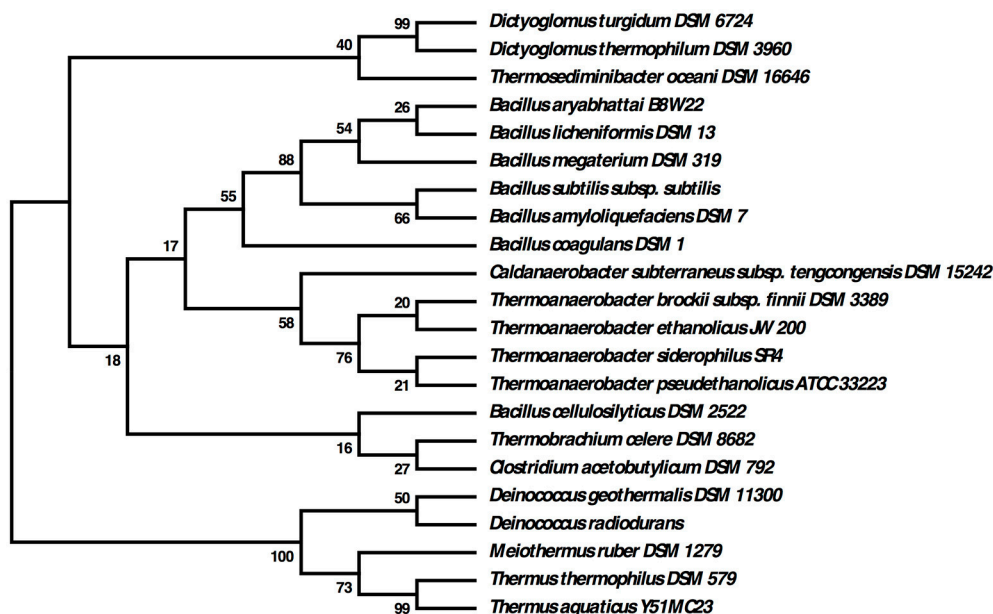


FIGURE 7 | Evolutionary relationships of SpoVS proteins. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The bootstrap consensus tree inferred from 550 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (550 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 22 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 86 positions in the final dataset.

normally associated with microbes, including amino acid and fatty acid degradation pathways and energy harvesting via proteins containing hemes, sirohemes, or quinones and appears to be the genome of a strict carbohydrate fermentor. Yet, at the same time, *D. turgidum* cannot be identified as similar to any one organism or Phylum. Results presented here and elsewhere (Nishida et al., 2011; Vesth et al., 2013) show the chameleon-like nature of the organism. Changing the method of comparison radically changes the resulting relationships between *D. turgidum* and other organisms.

The metabolic reconstruction based on the *D. turgidum* genome reveals two unusual features. Most evident is the importance of carbohydrate metabolism for the organism, because *D. turgidum* lacks the ability to metabolize fatty acids and most amino acids. The genome, while lacking in enzymes to degrade crystalline cellulose, possesses genes coding for utilization of most other biomass-derived polymers including xylans, glucans, pectins, arabinans and galactans. Utilization of these polysaccharides appears to involve secretion of enzymes that degrade the polysaccharides to oligosaccharides, transport of the oligosaccharides into the cytoplasm, and degradation of the oligosaccharides to monosaccharides in the cytoplasm. A similar strategy is utilized by thermophilic *Geobacillus* species (Brumm P. et al., 2015). Genes for utilization of carbohydrates are distributed randomly throughout the *D. turgidum* genome, unlike the *Geobacillus* genomes, where genes for individual polysaccharide degradation pathways are organized into distinct operons. The second feature is the obligate fermentative

nature of *D. turgidum*. Unlike many other thermophilic anaerobes including *Thermus*, *Geobacillus*, *Caldicellulosiruptor*, and *Thermotoga* species, *D. turgidum* possesses no genes for production or utilization of either cytochromes or quinones. Dtur is predicted to utilize the EMP pathway, to produce ATP, reducing equivalents, and fermentation products from monosaccharides. The predicted fermentation products, lactate, acetate, ethanol and hydrogen are in agreement with the published microbiological studies (Svetlichny and Svetlichnaya, 1988) showing the organism produces these four products during fermentation on sugars. The proton gradient needed for ATP generation is produced by NADH oxidoreductase and succinate dehydrogenase, and the ATP is generated by an F₀F₁-type and a V-type ATP synthases.

Sixteen *D. turgidum* carbohydrases were cloned, expressed and characterized to better understand their function in the metabolism of the organism. The 16 included two each of GH1 and GH3, four GH5, two GH10, and one each of GH36, GH42, GH43, GH53, GH57, and GH67. Based on the proposed mechanism for polysaccharide utilization, *D. turgidum* produces oligosaccharides using secreted enzymes, and degrades the oligosaccharides using intracellular enzymes. The secreted enzymes would be expected to have high substrate specificity to generate oligosaccharides recognized by the transporter systems. The cloned enzymes predicted to be secreted showed activity only on one or two substrates, showing activity on xylan, arabinan, *beta*-glucan, starch, or mannan. Conversely, the intracellular enzymes would be expected to have low specificity, allowing

them to degrade multiple substrates and linkages efficiently. The cloned intracellular enzymes typically showed a broader range of activities. GH1 and GH3 enzymes possess *exo*-activity on four or five different carbohydrate substrates. The cloned intracellular xylanase and cellulase both possess both *exo*-activity and *endo*-activity, as well as activity on multiple substrates.

Replication, recombination, and repair enzymes are critical to the genome maintenance and integrity of all cells. Many proteins from this COG functional category are expected to share conserved domains and motifs that could in theory be used to understand the phylogenetic relationship of *D. turgidum* to other organisms. The 16S rRNA genes of *D. turgidum* and *D. thermophilum* share 99% sequence identity. The fraction of replication proteins having 100–90% similarity between the two species is 35%, with 42% sharing 89–80% similarity, 17% sharing 79–70% similarity and 5% with similarity below 69%. The similarity of *D. turgidum* replication proteins to other taxa drops off considerably from that with *D. thermophilum*. The fraction of replication proteins having 100–90% similarity between *D. turgidum* and the next nearest neighbor species is 0%, with 1% sharing 89–80% similarity, 3% sharing 79–70% similarity, 13% sharing 69–60 similarity, 32% sharing 59–50% similarity, 28% sharing 49–40 similarity, and 32% similarity below 40%. This informal comparison again demonstrates how unique *Dictyoglomi* are compared to other species. The number and type of replication proteins found in *D. turgidum* is similar to those found in other hyperthermophilic bacteria using IMG tools

at the Joint Genome Institute (data not shown). The phylogenetic position of the reverse gyrase does not conform to the 16S rRNA phylogeny suggesting that lateral gene transfer may have taken place.

AUTHOR CONTRIBUTIONS

FR analyzed data and contributed to manuscript preparation. PB wrote the manuscript, produced and purified Dtur proteins, and performed carbohydrate analyses. KG produced and purified Dtur proteins. DM managed genome sequencing and analysis, performed DNAP analyses, and contributed to manuscript preparation.

FUNDING

This work was completely funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494 and DOE OBP Office of Energy Efficiency and Renewable Energy DE-AC05-76RL01830). FR acknowledges support from the NASA Exobiology Program.

ACKNOWLEDGMENTS

The authors would like to thank Robb lab and Elizabeth O'Connor for assistance in the growth and preparation of the cells used for genomic DNA isolation.

REFERENCES

- Aravind, L., and Koonin, E. V. (1999). DNA polymerase β -like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res.* 27, 1609–1618. doi: 10.1093/nar/27.7.1609
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Brochier-Armanet, C., and Forterre, P. (2007). Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* 2, 83–93. doi: 10.1155/2006/582916
- Brumm, P. J., De Maayer, P., Cowan, D. A., and Mead, D. A. (2015). Genomic analysis of six new Geobacillus strains reveals highly conserved carbohydrate degradation architectures and strategies. *Front. Microbiol.* 6:430. doi: 10.3389/fmicb.2015.00430
- Brumm, P. J. (2013). Bacterial genomes: what they teach us about cellulose degradation. *Biofuels* 4, 669–681. doi: 10.4155/bfs.13.44
- Brumm, P. J., Hermanson, S., Luedtke, J., and Mead, D. A. (2011). Identification, cloning and characterization of *Dictyoglomus turgidum* CelA, an endoglucanase with cellulase and mannanase activity. *J. Life Sci.* 5, 488–496.
- Brumm, P. J., Monsma, S., Keough, B., Jasnovica, S., Ferguson, E., Schoenfeld, T., et al. (2015). Complete genome sequence of thermus aquaticus Y51MC23. *PLoS ONE* 10:e0138674. doi: 10.1371/journal.pone.0138674
- Burgess, E. A., Unrine, J. M., Mills, G. L., Romanek, C. S., and Wiegel, J. (2012). Comparative geochemical and microbiological characterization of two thermal pools in the Uzon Caldera, Kamchatka, Russia. *Microb. Ecol.* 63, 471–489. doi: 10.1007/s00248-011-9979-4
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 42, D459–D471. doi: 10.1093/nar/gkt1103
- Chander, Y., Koelbl, J., Puckett, J., Moser, M. J., Klingele, A. J., Liles, M. R., et al. (2014). A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). *Front. Microbiol.* 5:395. doi: 10.3389/fmicb.2014.00395
- Coil, D. A., Badger, J. H., Forberger, H. C., Riggs, F., Madupu, R., Fedorova, N., et al. (2014). Complete genome sequence of the extreme thermophile *Dictyoglomus thermophilum* H-6-12. *Genome Announc.* 2:e00109-14. doi: 10.1128/genomeA.00109-14
- Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., et al. (2010). MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* 38, D396–D400. doi: 10.1093/nar/gkp919
- Delcher, A. L., Salzberg, S. L., and Phillippy, A. M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinformatics* Chapter 10, Unit 10.13. doi: 10.1002/0471250953.bi1003s00
- Devoid, S., Overbeek, R., DeJongh, M., Vonstein, V., Best, A. A., and Henry, C. (2013). Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. *Methods Mol. Biol.* 985, 17–45. doi: 10.1007/978-1-62703-299-5_2
- Ding, Y. H., Ronimus, R. S., and Morgan, H. W. (2000). Sequencing, cloning, and high-level expression of the pfp gene, encoding a PP(i)-dependent phosphofructokinase from the extremely thermophilic eubacterium *Dictyoglomus thermophilum*. *J. Bacteriol.* 182, 4661–4666. doi: 10.1128/JB.182.16.4661-4666.2000
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Euzéby, J. (2012). List of new names and new combinations previously effectively, but not validly, published. *Int. J. Syst. Evol. Microbiol.* 62, 1–4. doi: 10.1099/ijs.0.039487-0

- Euzéby, J. P. (1998). Taxonomic note: necessary correction of specific and subspecific epithets according to rules 12c and 13b of the international code of nomenclature of bacteria (1990 Revision). *Int. J. Syst. Evol. Microbiol.* 48, 1073–1075. doi: 10.1099/00207713-48-3-1073
- Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186–194. doi: 10.1101/gr.8.3.186
- Fukusumi, S., Kamizono, A., Horinouchi, S., and Beppu, T. (1988). Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic thermophile, *Dictyoglomus thermophilum*. *Eur. J. Biochem.* 174, 15–21. doi: 10.1111/j.1432-1033.1988.tb14056.x
- Gao, D., Uppugundla, N., Chundawat, S. P., Yu, X., Hermanson, S., Gowda, K., et al. (2011). Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnol. Biofuels* 4:5. doi: 10.1186/1754-6834-4-5
- Gibbs, M. D., Reeves, R. A., and Bergquist, P. L. (1995). Cloning, sequencing, and expression of a xylanase gene from the extreme thermophile *Dictyoglomus thermophilum* Rt46B.1 and activity of the enzyme on fiber-bound substrate. *Appl. Environ. Microbiol.* 61, 4403–4408.
- Gibbs, M. D., Reeves, R. A., Sunna, A., and Bergquist, P. L. (1999). Sequencing and expression of a β -mannanase gene from the extreme thermophile *Dictyoglomus thermophilum* Rt46B.1, and characteristics of the recombinant enzyme. *Curr. Microbiol.* 39, 351–0357.
- Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195–202. doi: 10.1101/gr.8.3.195
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Res.* 31, 439–441. doi: 10.1093/nar/gkg006
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360
- Gumerov, V. M., Mardanov, A. V., Beletskii, A. V., Bonch-Osmolovskaia, E. A., and Ravin, N. V. (2011). [Molecular analysis of microbial diversity in the Zavarzin Spring, the Uzon caldera]. *Mikrobiologiya* 80, 258–265. doi: 10.1134/s002626171102007x
- Haas, B. J., Sandigursky, M., Tainer, J. A., Franklin, W. A., and Cunningham, R. P. (1999). Purification and characterization of *Thermotoga maritima* endonuclease IV, a thermostable apurinic/aprimidinic endonuclease and 3'-repair diesterase. *J. Bacteriol.* 181, 2834–2839.
- Hong, S. H., Lim, Y. R., Kim, Y. S., and Oh, D. K. (2012). Molecular characterization of a thermostable L-fucose isomerase from *Dictyoglomus turgidum* that isomerizes L-fucose and D-arabinose. *Biochimie* 94, 1926–1934. doi: 10.1016/j.biochi.2012.05.009
- Horinouchi, S., Fukusumi, S., Ohshima, T., and Beppu, T. (1988). Cloning and expression in *Escherichia coli* of two additional amylase genes of a strictly anaerobic thermophile, *Dictyoglomus thermophilum*, and their nucleotide sequences with extremely low guanine-plus-cytosine contents. *Eur. J. Biochem.* 176, 243–253. doi: 10.1111/j.1432-1033.1988.tb14275.x
- Hsiao, W., Wan, L., Jones, S. J., and Brinkman, F. S. (2003). IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19, 418–420. doi: 10.1093/bioinformatics/btg004
- Huber, R., Langworthy, T. A., König, H., Thomm, M., Woese, C. R., Sleytr, U. B., et al. (1986). *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* 144, 324–333.
- Hubisz, M. J., Pollard, K. S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinformatics* 12, 41–51. doi: 10.1093/bib/bbq072
- Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal prokaryotic dynamic programming gene-finding algorithm. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Ishino, Y., and Narumi, I. (2015). DNA repair in hyperthermophilic and hyperthermoresistant microorganisms. *Curr. Opin. Microbiol.* 25, 103–112. doi: 10.1016/j.mib.2015.05.010
- Jernigan, K. K., and Bordenstein, S. R. (2015). Tandem-repeat protein domains across the tree of life. *Peer J.* 3:e732. doi: 10.7717/peerj.732
- Karp, P. D., Ouzounis, C. A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahrén, D., et al. (2005). Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.* 33, 6083–6089. doi: 10.1093/nar/gki892
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kelman, L. M., and Kelman, Z. (2014). Archaeal DNA replication. *Annu. Rev. Genet.* 48, 71–97. doi: 10.1146/annurev-genet-120213-092148
- Kim, J. E., Kim, Y. S., Kang, L. W., and Oh, D. K. (2012). Characterization of a recombinant cellobiose 2-epimerase from *Dictyoglomus turgidum* that epimerizes and isomerizes β -1,4- and α -1,4-glucosyl oligosaccharides. *Biotechnol. Lett.* 34, 2061–2068. doi: 10.1007/s10529-012-0999-z
- Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64(Pt 2), 346–351. doi: 10.1099/ijs.0.059774-0
- Kim, Y. S., Shin, K. C., Lim, Y. R., and Oh, D. K. (2013). Characterization of a recombinant L-rhamnose isomerase from *Dictyoglomus turgidum* and its application for L-rhamnulose production. *Biotechnol. Lett.* 35, 259–264. doi: 10.1007/s10529-012-1069-2
- Kochetkova, T. V., Rusanov, I. I., Pimenov, N. V., Kolganova, T. V., Lebedinsky, A. V., Bonch-Osmolovskaya, E. A., et al. (2011). Anaerobic transformation of carbon monoxide by microbial communities of Kamchatka hot springs. *Extremophiles* 15, 319–325. doi: 10.1007/s00792-011-0362-7
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Kublanov, I. V., Perevalova, A. A., Slobodkina, G. B., Lebedinsky, A. V., Bidzhieva, S. K., Kolganova, T. V., et al. (2009). Biodiversity of thermophilic prokaryotes with hydrolytic activities in hot springs of Uzon Caldera, Kamchatka (Russia). *Appl. Environ. Microbiol.* 75, 286–291. doi: 10.1128/AEM.00607-08
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178
- Love, C. A., Patel, B. K. C., Ludwig, W., and Stackebrandt, E. (1993). The phylogenetic position of *Dictyoglomus thermophilum* based on 16S rRNA sequence analysis. *FEMS Microbiol. Lett.* 107, 317–320. doi: 10.1111/j.1574-6968.1993.tb06050.x
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955
- Mathrani, I., and Ahring, B. (1991). Isolation and characterization of a strictly xylan-degrading *Dictyoglomus* from a man-made, thermophilic anaerobic environment. *Arch. Microbiol.* 157, 13–17. doi: 10.1007/BF00245328
- Mathrani, I., and Ahring, B. (1992). Thermophilic and alkalophilic xylanases from several *Dictyoglomus* isolates. *Appl. Microbiol. Biotechnol.* 38, 23–27. doi: 10.1007/BF00169413
- Menzel, P., Gudbergssdóttir, S. R., Rike, A. G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* 70, 411–424. doi: 10.1007/s00248-015-0576-9
- Morris, D. D., Gibbs, M. D., Chin, C. W., Koh, M. H., Wong, K. K., Allison, R. W., et al. (1998). Cloning of the xynB gene from *Dictyoglomus thermophilum* Rt46B.1 and action of the gene product on kraft pulp. *Appl. Environ. Microbiol.* 64, 1759–1765.
- Nishida, H., Beppu, T., and Ueda, K. (2011). Whole-genome comparison clarifies close phylogenetic relationships between the phyla Dictyoglomi and Thermotogae. *Genomics* 98, 370–375. doi: 10.1016/j.ygeno.2011.08.001
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use

- in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi: 10.1093/nar/gki866
- Patel, B. K., Morgan, H. W., Wiegel, J., and Daniel, R. M. (1987). Isolation of an extremely thermophilic chemoorganotrophic anaerobe similar to *Dictyoglomus thermophilum* from New Zealand hot springs. *Arch. Microbiol.* 147, 21–24. doi: 10.1007/BF00492899
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Rawlings, N. D., Barrett, A. J., and Bateman, A. (2014). Using the MEROPS database for proteolytic enzymes and their inhibitors and substrates. *Curr. Protoc. Bioinformatics* 48, 1.25.1–33. doi: 10.1002/0471250953.bi0125s48
- Rees, G. N., Patel, B. K., Grassia, G. S., and Sheehy, A. J. (1997). *Anaerobaculum thermoterrenum* gen. nov., sp. nov., a novel, thermophilic bacterium which ferments citrate. *Int. J. Syst. Bacteriol.* 47, 150–154. doi: 10.1099/00207713-47-1-150
- Rigden, D. J., and Galperin, M. Y. (2008). Sequence analysis of GerM and SpoVS, uncharacterized bacterial ‘sporulation’ proteins with widespread phylogenetic distribution. *Bioinformatics* 24, 1793–1797. doi: 10.1093/bioinformatics/btn314
- Sahm, K., John, P., Nacke, H., Wemheuer, B., Grote, R., Daniel, R., et al. (2013). High abundance of heterotrophic prokaryotes in hydrothermal springs of the azores as revealed by a network of 16S rRNA gene-based methods. *Extremophiles* 17, 649–662. doi: 10.1007/s00792-013-0548-2
- Saiki, T., Kobayashi, Y., Kawagoe, K., and Beppu, T. (1985). *Dictyoglomus thermophilum* gen. nov., sp. nov., a Chemoorganotrophic, Anaerobic, Thermophilic Bacterium. *Int. J. Syst. Evol. Microbiol.* 35, 253–259.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Shandilya, H., Griffiths, K., Flynn, E. K., Astatke, M., Shih, P. J., Lee, J. E., et al. (2004). Thermophilic bacterial DNA polymerases with reverse-transcriptase activity. *Extremophiles* 8, 243–251. doi: 10.1007/s00792-004-0384-5
- Shi, R., Li, Z., Ye, Q., Xu, J., and Liu, Y. (2013). Heterologous expression and characterization of a novel thermo-halotolerant endoglucanase Cel5H from *Dictyoglomus thermophilum*. *Bioresour. Technol.* 142, 338–344. doi: 10.1016/j.biortech.2013.05.037
- Spriestersbach, A., Kubicek, J., Schäfer, F., Block, H., and Maertens, B. (2015). Purification of his-tagged proteins. *Meth. Enzymol.* 559, 1–15. doi: 10.1016/bs.mie.2014.11.003
- Svetlichny, V. A., and Svetlichnaya, T. P. (1988). *Dictyoglomus turgidus* sp. nov., a new extremely thermophilic eubacterium isolated from hot springs of the Uzon volcano caldera. *Mikrobiologiya* 57, 435–441.
- Takai, K., Inoue, A., and Horikoshi, K. (1999). *Thermaerobacter marianensis* gen. nov., sp. nov., an aerobic extremely thermophilic marine bacterium from the 11,000 m deep Mariana Trench. *Int. J. Syst. Bacteriol.* 49(Pt 2), 619–628. doi: 10.1099/00207713-49-2-619
- Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Teichtmann, S. M., Colman, A. S., and Robb, F. T. (2009). That which does not kill us only makes us stronger: the role of carbon monoxide in thermophilic microbial consortia. *Environ. Microbiol.* 11, 1027–1037. doi: 10.1111/j.1462-2920.2009.01865.x
- Vesth, T., Ozen, A., Andersen, S. C., Kaas, R. S., Lukjancenko, O., Bohlin, J., et al. (2013). Veillonella, firmicutes: microbes disguised as Gram negatives. *Stand. Genomic Sci.* 9, 431–448. doi: 10.4056/sigs.2981345
- Wagner, I. D., and Wiegel, J. (2008). Diversity of thermophilic anaerobes. *Ann. N.Y. Acad. Sci.* 1125, 1–43. doi: 10.1196/annals.1419.029
- Wijffels, G., Dalrymple, B., Kongsuwan, K., and Dixon, N. E. (2005). Conservation of eubacterial replicases. *IUBMB Life* 57, 413–419. doi: 10.1080/15216540500138246
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucl. Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

Conflict of Interest Statement: At the time this work was performed, the authors PB and DM were employees and shareholders of C5-6 Technologies Inc. (WI, USA), a company that created bio-based solutions to efficiently convert biomass into five and six carbon sugars. The company ceased operation in December of 2014. PB has since purchased the assets of the company and started C5-6 Technologies LLC (WI, USA), a company focused on supplying reagent enzymes for carbohydrase research. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Brumm, Gowda, Robb and Mead. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership