Searching for novel thermostable β-galactosidases

Mª Eugenia de Castro de Antonio

PhD Thesis 2020

Directors: Dra. Mª Isabel González-Siso and Dra. Mª Esther Rodríguez Belmonte

Programa Oficial de Doutoramento en Bioloxía Celular e Molecular

UNIVERSIDADE DA CORUÑA

El presente trabajo, **Searching for novel thermostable β-galactosidases**, presentado por Doña Mª Eugenia de Castro de Antonio para aspirar al grado de doctor en Biología, ha sido realizado bajo nuestra dirección en el Departamento de Biología, de la Universidad de A Coruña.

Revisado el texto. Estamos conformes con su presentación para ser juzgado.

A Coruña, 30 de noviembre de 2020

VºBº Dra. Mª Isabel González-Siso

Catedrática de Bioquímica y Biología Molecular

VºBº Dra. Mª Esther Rodríguez-Belmonte

Profesora de Bioquímica y Biología Molecular

A Inés y Álvaro

"Un día supe que buscar era lo que me mantenía despierto"

(Manolo García, Un año y otro año)

En primer lugar, me gustaria dar las gracias a mis dos directoras de Tesis, Isabel y Esther, sin las que hoy no estaría aquí. Gracias por brindarme la oportunidad de aprender haciendo lo que me gusta, investigar, y por la inmensa paciencia que habéis demostrado tener en estos años.

Por supuesto también quiero agradecer a todos los profesores del departamento de Bioquímica de la UDC: Esperanza, Manuel, Mónica, Ana y Marian, porque todos ellos han contribuido con su ayuda y su atención, tanto a nivel académico como a nivel personal, a que hoy, por fin, haya llegado hasta aquí.

Cómo olvidarme de los grandes amigos que he conocido en el laboratorio y que me han acompañado durante este camino. Supongo que no hay un GRACIAS lo suficientemente grande para todos vosotros, porque vuestra calidad humana está, si cabe, por encima de vuestras grandes aptitudes como investigadores. Vosotros habéis sido los responsables de que recuerde con una sonrisa todos los momentos de esta etapa en el labo (como solemos llamarlo). Siempre dispuestos a celebrar los logros (científicos y personales), pero también ofreciendo una sonrisa y una posible solución cuando las cosas no salían tan bien. Olalla, gracias porque desde el principio te he sentido como una amiga, cuando yo no conocía a nadie en Coruña, ahí estabas tú con esos planes de costura y manualidades que tanto me gustaban. Espero y creo que Alvarito y Leo serán dos grandes amigos. Aida, gracias por tu amistad incondicional, por todos los paseos, los momentos de risas y los bailes descoordinados al son de "Madre tierra" al salir del labo y, por supuesto, por cuidar como nadie de Alvarito los días que no encontraba quién me echara una mano. Mari, mil gracias por esos ratos de confidencias a solas en las comidas, por todos los buenos consejos que me has dado y por ser el mejor ejemplo de fuerza y superación que he conocido. Mónica, no puedes faltar aquí, gracias por la atención y la ilusión que pones en cada cosa que haces, y por tu capacidad de organización, de la que dependemos (y en la que confiamos) todos. María "pequeña", cuando llegaste con tu espíritu fresco y alegre, parecía que la feria de Sevilla se había colado en el labo, gracias por poner esa chispa a nuestro día a día. Gracias Ángel por haberme enseñado tantas cosas, dispuesto en todo momento a ayudar, eres una pieza clave en el labo y creo que sin tu esfuerzo y dedicación no sería lo que es. Agustín, muchas

gracias por tu serenidad, por compatir conmigo sin recelo todos tus conocimientos acerca de las β-galactosidasas, sin duda gracias a ello he podido llegar hasta aquí. Juanjo, gracias por tu bondad, por tantos y tantos favores, como dejarme tu abrigo aquella famosa noche en Cambridge o ayudarme a cargar bidones y bidones de agua termal hasta el coche, porque nada habría sido lo mismo sin ti.

También quisiera dar las gracias a todas aquellas personas con las que de un modo u otro he coincidido en el laboratorio como Flor, Kamila, Ángel Pereira, Rafa, Maya, Natalí, Rosa, Quique, Silvia, María, y otros muchos que probablemente me estaré dejando en el tintero.

Me gustaría agradecer a la doctora Elizabeth Dinsdale por haberme acogido en su laboratorio durante los tres meses de mi estancia en San Diego. Y a todos los que allí me hicieron sentir como en casa. En especial a Geni, por ayudarme desde el primer momento con todo y enseñarme lo que significa "pwd", "mkdir" o "ls". También quisiera agradecer a Kevin, por su alegría contagiosa y por amenizar mis días en San Diego con sus listas musicales. A Mike, por su ayuda con la secuenciación y por darme la oportunidad de navegar por el Pacífico en busca de bosques de kelp. Sin olvidar a Dough, a Matt, a Fabyano, a Valeria, a André, a Carmen y a Bobby.

Quisiera recordar en estos agradecimientos a María Rubira, la profesora que, ya en el colegio, hizo que me apasionase la Biología y encendió en mí la chispa de la "curiosidad científica" que me ha llevado hasta aquí.

Pero, cuando se trata de agradecer, no puedo olvidarme de mi familia. En primer lugar me gustaría dar las gracias a Álvaro, porque sin duda es quién más ha vivido conmigo todos los momentos buenos de estos años y sufrido los no tan buenos. Gracias por estar siempre a mi lado, intentando comprender y ayudar, aunque la mera palabra "galactosidasa" te sonara a chino. A Alvarito e Inés, porque ellos hacen que quiera ser mejor cada día y han llenado de amor estos últimos años. Espero que, si algún día leéis esto, os sintáis orgullosos de vuestra madre.

Sin duda alguna si hay alguien a quien tengo que dar las gracias es a mis padres, que siempre han confiado en mí, manteniéndose a mi lado en todo momento, sin

importar que la decisión que había tomado fuese acertada o no. A ellos les debo todo lo que tengo y todo lo que soy. Gracias por confiar en mí, por darme la oportunidad de estudiar y de ser lo que me gusta y por haber sido el mejor ejemplo a seguir.

Por último quisiera dar un gracias generalizado a todos los que de alguna manera me han ayudado en este camino, como mi abuela, mis hermanos, sobrinos, amigos y tantas y tantas personas que seguramente esté olvidando. GRACIAS A TODOS.

Contents

Short abstracts

ABSTRACT

β-galactosidases are biotechnologically interesting enzymes for the hydrolysis of lactose and GOS synthesis. The main objective of this work was to analyze the microbial community structure and the functional potential of As Burgas hot spring, focusing on the discovery of novel thermostable β-galactosidases in this thermal ecosystem.

A novel thermostable β-galactosidase, named BWbg1, was discovered through functional screening of a metagenomic library from As Burgas water. The high thermal stability displayed by the enzyme, with an optimum temperature of 80 °C and optimum pH close to the pH of milk, coupled with its high GOS yield production, make BWbg1 a very suitable catalyst for the dairy industry.

Proteobacteria are the main inhabitants of As Burgas hot spring. The functional analysis reveals the importance and correlation between carbon, nitrogen, sulfur, and hydrogen cycles in As Burgas. From the two β-galactosidases found by sequence annotation, only pTsbg showed hydrolytic activity towards ONPG, but was unable to hydrolyze lactose.

The significant differences between As Burgas population and the nearby Muiño da Veiga hot spring are mostly associated with their variability in the geochemical water composition. Temperature and pH are two important factors shaping hot springs' microbial community, as was determined by comparative analysis with other thermal springs.

RESUMEN

Las β-galactosidasas son enzimas biotecnológicamente interesantes para la hidrólisis de lactosa y la síntesis de GOS. El objetivo principal de este trabajo fue analizar la estructura de la comunidad microbiana y el potencial funcional de las aguas termales de As Burgas, centrándose en el descubrimiento de nuevas β-galactosidasas termoestables en este ecosistema termal.

Short abstracts

Mediante el cribado funcional de una metagenoteca construida a partir de agua de As Burgas se ha descubierto una nueva β-galactosidasa termoestable, denominada BWbg1. La alta estabilidad térmica que muestra la enzima, con una temperatura óptima de 80 °C y un pH óptimo cercano al pH de la leche, junto con su alto rendimiento en la producción de GOS, hacen de BWbg1 un catalizador muy adecuado para la industria láctea.

Las proteobacterias son los principales habitantes de las aguas termales de As Burgas. El análisis funcional revela la importancia y correlación entre los ciclos del carbono, nitrógeno, azufre e hidrógeno en As Burgas. De las dos β-galactosidasas detectadas por la anotación de secuencias, sólo pTsbg mostró actividad hidrolítica frente a ONPG, pero fue incapaz de hidrolizar la lactosa.

Las diferencias significativas entre la población de As Burgas y la cercana fuente termal de Muiño da Veiga, se asocian principalmente a su variabilidad en la composición geoquímica del agua. La temperatura y el pH son dos factores importantes que perfilan la comunidad microbiana de las aguas termales, como se determinó mediante análisis comparativo con otras fuentes termales.

RESUMO

As β-galactosidasas son encimas biotecnolóxicamente interesantes para a hidrólise da lactosa e a síntese de GOS. O principal obxectivo deste traballo foi analizar a estrutura da comunidade microbiana e o potencial funcional das augas termais das Burgas, centrándose no descubrimento de novas β-galactosidasas termoestables neste ecosistema termal.

Mediante o cribado funcional dunha metaxenoteca de auga de As Burgas descubriuse unha nova β-galactosidasa termoestable, chamada BWbg1. A alta estabilidade térmica mostrada polo encima, cunha temperatura óptima de 80 °C e un pH óptimo próximo ao pH do leite, xunto co seu alto rendemento na produción de GOS, fan do BWbg1 un catalizador moi axeitado para a industria láctea.

As proteobacterias son os principais habitantes das augas termais das Burgas. O analise funcional revela a importancia e correlación entre os ciclos de carbono, nitróxeno, xofre e hidróxeno nas Burgas. Das dúas β-galactosidasas detectadas por anotación de secuencia de contigs, só pTsbg mostrou actividade hidrolítica contra ONPG, pero non foi capaz de hidrolizar a lactosa.

As diferenzas significativas entre a poboación das Burgas e a fonte termal próxima de Muíño da Veiga están asociadas principalmente á súa variabilidade na composición xeoquímica da auga. A temperatura e o pH son dous factores importantes que configuran a comunidade microbiana das augas termais, como se determinou mediante análise comparativa con outras augas termais.

Introduction

β-galactosidases (EC.3.2.1.23; β-D-galactoside-galacto-hydrolases or lactases) catalyze the hydrolysis of the β-1,4-D-glycosidic bond of lactose, releasing glucose and galactose. Besides the hydrolysis, some β-galactosidases can perform transgalactosylation reactions, in which the galactosyl moiety is transferred to lactose or another carbohydrate acceptor, producing galacto-oligosaccharides (GOS) (Fig 1).
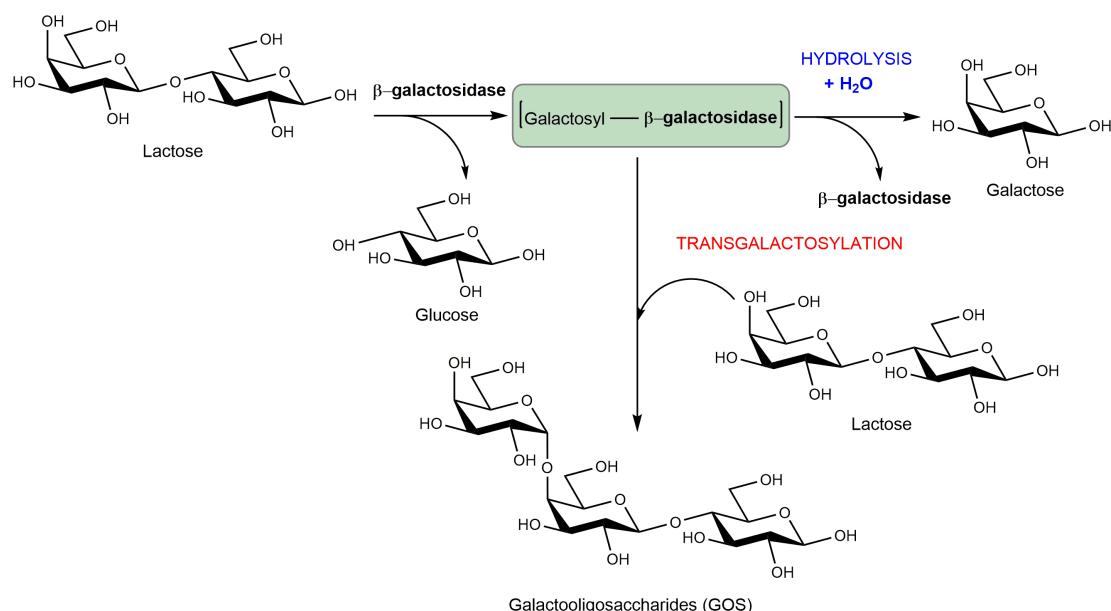
Figure 1. Schematic representation of the enzymatic reactions catalyzed by β-galactosidases. Extracted from DeCastro et al., 2018.

With a characteristic TIM barrel in their quaternary structure (Saqib et al., 2017), β-galactosidases are included within the glycoside hydrolases (GHs) in the Carbohydrate-Active enZYmes Database (CAZY, http://www.cazy.org), having representatives in GH1, GH2, GH35, GH42, GH59 and GH147 families (Lu et al., 2020). However, this grouping is constantly changing, since the discovery of novel β-galactosidases belonging to other families of glycoside hydrolases may occur. For example, a novel multifunctional GH43 enzyme showing β-galactosidase activity was found by functional metagenomic analysis of cow rumen (Ferrer et al., 2012).

The hydrolytic activity of β-galactosidases has been intensively used in the dairy industry for the production of low-lactose milk and milk derivatives, suitable for lactose intolerant people (Xavier et al., 2018). Additionally, lactose hydrolysis can be applied to improve the properties of some dairy products, increasing their sweetness

and creaminess. From an ecological point of view, β-galactosidases play an important role in the revalorization of whey, a highly polluting byproduct of the dairy industry that must be eliminated.

The transgalactosylation potential of β-galactosidases is mainly used for the obtention of GOS, non-digestible carbohydrates able to induce the growth of beneficial bifidobacteria such as *Bifidobacterium* and *Lactobacillus* (Monteagudo-Mera et al., 2016; Thongaram et al., 2017). These prebiotics can help in the prevention of colorectal cancer (Bruno-Barcena and Azcarate-Peril, 2015), activation of the immune system (Shokryazdan et al., 2017), and the enhancement of intestinal mineral absorption (Whisner and Castillo, 2018; Seijo et al., 2019), and thus they are frequently added to infant milk formulas, dairy products, and pet food, among others.

Other applications for β-galactosidases include the production of astragalin galactosides through transgalactosylation (Han et al., 2017), their use as biosensors for lactose determination in milk (Sharma and Leblanc, 2017), or the transformation of stevioside into rubusoside (Chen et al., 2014).

Microorganisms from thermophilic origin are not only capable of resisting high temperatures, but many of them can also tolerate other adverse conditions such as high conductivity, exposure to heavy metals, or radiation (Ranawat and Rawat, 2017; Gallo et al., 2018). Thermostable β-galactosidases such as those from *Sulfolobus solfataricus, Dictyoglomus turgidum,* and *Bacillus stearothermophilus* could be used in the industry together with high temperatures to enhance initial productivity, prevent microbial contamination, or increase the substrates solubilization (Pisani et al. 1990, Zolnere and Ciprovica, 2017). Therefore, hot environments are continuously bio-prospected in order to find novel β-galactosidases with these desirable conditions. A more extensive review of thermophilic β-galactosidases and their sources can be found in our previous publication (DeCastro et al., 2018).

Among the high-temperature environments, geothermal springs have been intensively studied, revealing the vast diversity of thermophiles inhabiting these

habitats, initially thought to be lifeless. Moreover, the study of these ecosystems and their microorganisms can shed light on the earliest life forms, as a high-temperature origin of life on Earth has been proposed (Damer and Deamer 2019; McClendon, 1999) or can increase our knowledge of the possible forms of extraterrestrial life (Cavicchioli, 2002). Additionally, due to the ease of access and sampling in most hot springs, these biomes have become one of the main sources for the bioprospecting of novel thermozymes of biotechnological interest, including a wide number of β-galactosidases (Chackraborti et al., 2003; Rani et al., 2019).

The irreproducibility of the special conditions that take place in geothermal springs has been one of the main difficulties for the study of hot springs thermophiles and their enzymatic potential, and has been overcome with the development of metagenomics. This approach, based on the study of the whole community DNA (metagenome) from an environment, can be addressed in two different ways: functional metagenomics and sequence metagenomics.

Functional metagenomics depends on the extraction, fragmentation, and cloning of the metagenome followed by the functional screening of the clones. When searching for β-galactosidases, the substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) is the most frequently used for the functional screening of the clones. As a result of the insoluble blue compound released by the active β-galactosidases after the hydrolysis of X-gal, those clones harboring β-galactosidase activity develop a characteristic blue color and can be selected for sequencing, sub-cloning, expression, and characterization of the enzyme responsible for this activity (Fig. 2). The main advantage of functional metagenomics is its potential to detect novel functional β-galactosidases that wouldn´t be predicted by their DNA sequence, (Cheng et al., 2017). Although a relatively high number of β-galactosidases have been found through functional metagenomics from several environments such as soil (Zhang et al., 2013; Wang et al., 2014; Cheng et al., 2017) or wheat straw (Maruthamuthu et al., 2016), there is only one reported thermostable β-galactosidase isolated from a hot spring following this approach (Gupta et al., 2012).

Sequence metagenomics requires the extraction, sequencing, and analysis of the environmental DNA. The gene prediction and annotation of the metagenomic reads, based on a reference sequence database, enables the identification of the microorganisms inhabiting the hot springs and facilitates the determination of the functions they perform in the ecosystem. Additionally, the prediction and annotation of genes in the assembled metagenomic sequences can be used to identify, amplify, and clone enzymes of biotechnological interest, like β-galactosidases, from the metagenome (Fig 2). So far, only one thermostable β-galactosidase from geothermal origin has been obtained through sequence-based metagenomics (Liu et al., 2015). The main disadvantage of this approach is its dependence on the sequence, which precludes finding novel enzymes with the desired activity but lacking sequence homology with others already described. Moreover, even when an enzyme is detected by sequence metagenomics and cloned, it may not have activity, since protein functionality does not rely solely on its sequence. And thus, the presence of certain conserved sequences or domains does not guarantee its enzymatic activity. A detailed description of metagenomics of thermophiles and its applications can be found in our previous publication (DeCastro et al., 2016).

With its 66 °C, As Burgas is among the hottest hot springs in Ourense (Northwestern Spain) and it is one of the most frequented by locals and tourists. In the present study, we have used the two described metagenomic approaches to explore the taxonomical and functional profile of this ecosystem and for the bioprospecting of novel thermostable β-galactosidases.

In the first chapter, we present the construction of a plasmid metagenomic library from As Burgas hot spring water. The functional screening of the library led to the discovery of a hitherto unknown β-galactosidase, named BWbg1, which was cloned and characterized. The enzyme, belonging to GH35 family, displayed a great activity towards o-Nitrophenyl-β-D-galactopyranoside (ONPG) and lactose and showed an interesting transgalactosylation rate, with an elevated production of GOS at high temperatures.
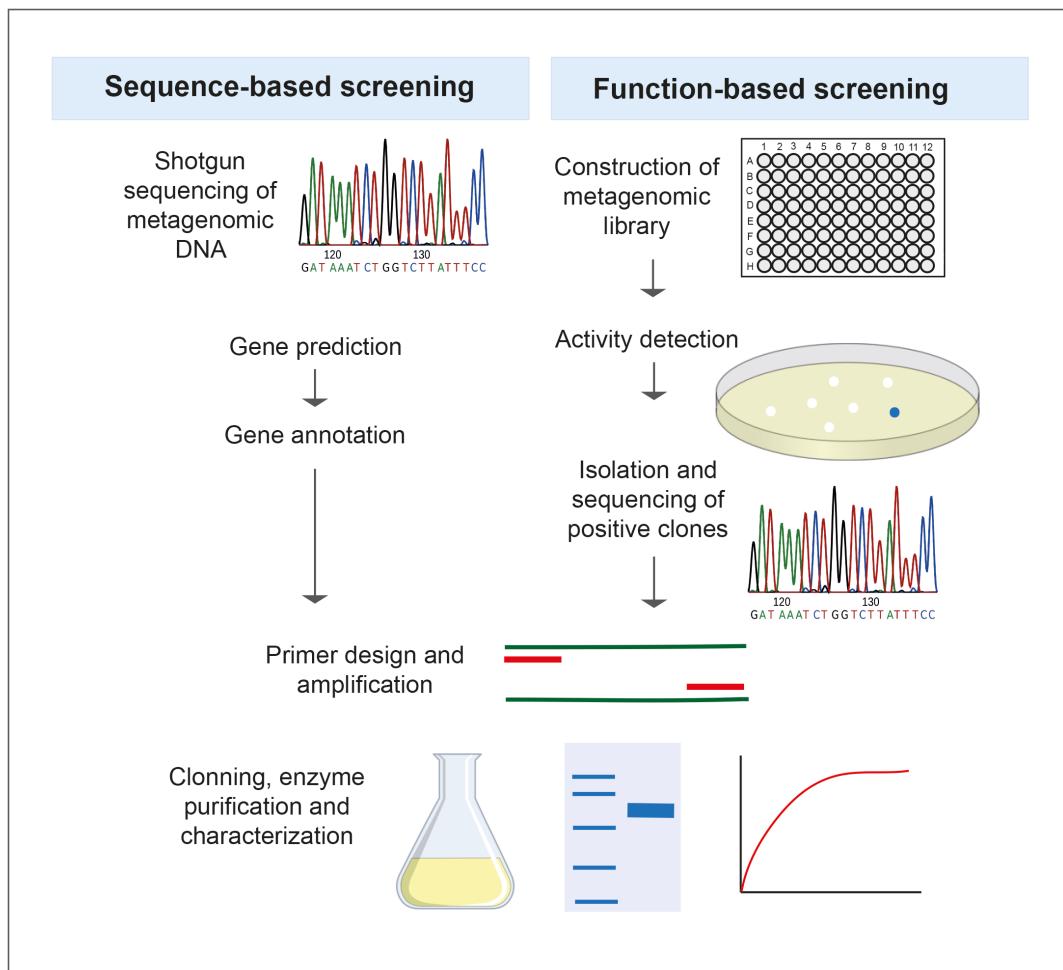
Figure 2. The two main strategies used for screening metagenomes in search of novel thermostable β-galactosidases. Taken from DeCastro et al., 2016.

In the second chapter, we analyze the taxonomical and functional profile of As Burgas through the shotgun sequencing of its water metagenome, followed by the annotation of the reads. The results reflect a microbial community dominated by Proteobacteria, in which carbon, sulfur, and nitrogen cycles play a crucial role. After the metagenomic assembly, the prediction and annotation of sequences with homology to β-galactosidases allowed us to amplify and clone two potential β-galactosidases, previously uncharacterized (Tsbg and pTsbg) and presumably belonging to *Thermus scotoductus* SA-01, as was revealed by sequence alignment. Unfortunately, only pTsbg was able to hydrolyze ONPG and none of them showed activity towards lactose, manifesting the drawbacks of using sequence metagenomics to find novel active enzymes.

In the third chapter, we used comparative metagenomics to explore the differences between As Burgas and a nearby geothermal spring: Muiño da Veiga. The statistical analysis revealed significant differences in the microbial communities inhabiting both ecosystems that might be related to the dissimilarities in water chemistry such as ammonia and sulfate concentration. Moreover, the comparison with other remote geothermal springs unveils a clear influence of factors such as pH and temperature on hot springs populations.

REFERENCES

Bruno-Barcena, J. M., and Azcarate-Peril, M. A. (2015). Galacto-oligosaccharides and colorectal cancer: Feeding our intestinal probiome. *J. Funct. Foods* 12, 92–108. doi:10.1016/j.jff.2014.10.029.

Cavicchioli, R. (2002). Extremophiles and the search for extraterrestrial life. *Astrobiology* 2, 281–292. doi:10.1089/153110702762027862.

Chackraborti, S., Sani, R. K., Shoo, D. K., Banerjee, U. ., and Sobti, R. C. (2003). Production and partial characterization of a novel, beta-galactosidase from a newly isolated *Bacillus Polymyxa*. *Scientia Iranica* 10 (3), 279-286.

Chen, J. M., Xia, Y. M., Wan, H. Da, Wang, H. J., and Liu, X. (2014). A complete specific cleavage of glucosyl and ester linkages of stevioside for preparing steviol with a β-galactosidase from *Sulfolobus solfataricus*. *J. Mol. Catal. B Enzym.* 105, 126–131. doi:10.1016/j.molcatb.2014.03.011.

Cheng, J., Romantsov, T., Engel, K., Doxey, A. C., Rose, D. R., Neufeld, J. D., et al. (2017). Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. *PLoS One* 12, e0172545. doi:10.1371/journal.pone.0172545.

Damer, B., and Deamer, D. (2019). The hot spring hypothesis for an origin of life. *Astrobiology* 20, 429–452. doi:10.1089/ast.2019.2045.

DeCastro, M.E., Escuder-Rodríguez, J.J., Cerdán, M.E., Becerra, M., Rodríguez-Belmonte, E., and González-Siso, M.I. (2018). Heat-loving β-galactosidases from cultured and uncultured microorganisms. *Curr. Protein Pept. Sci.* 19, 1224–1234. doi:10.2174/1389203719666180809111659.

DeCastro, M. E., Rodríguez-Belmonte, E., and González-Siso, M. I. (2016). Metagenomics of thermophiles with a focus on discovery of novel thermozymes. *Front. Microbiol.* 7, 1521. doi:10.3389/fmicb.2016.01521.

Ferrer, M., Ghazi, A., Beloqui, A., Vieites, J. M., López-Cortés, N., Marín-Navarro, J., et al. (2012). Functional metagenomics unveils a multifunctional glycosyl

hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS One* 7, e38134. doi:10.1371/journal.pone.0038134.

Gallo, G., Puopolo, R., Limauro, D., Bartolucci, S., and Fiorentino, G. (2018). Metal-tolerant thermophiles: From the analysis of resistance mechanisms to their biotechnological exploitation. *Open Biochem. J.* 12, 149–160. doi:10.2174/1874091x01812010149.

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.

Han, S., Hanh Nguyen, T. T., Hur, J., Kim, N. M., Kim, S.-B., Hwang, K.-H., et al. (2017). Synthesis and characterization of novel astragalin galactosides using β-galactosidase from *Bacillus circulans*. *Enzyme Microb. Technol.* 103, 59–67. doi:10.1016/J.ENZMICTEC.2017.05.003.

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.* 2818. doi:10.1080/13102818.2015.1015244.

Lu, L., Guo, L., Wang, K., Liu, Y., and Xiao, M. (2020). β-Galactosidases: A great tool for synthesizing galactose-containing carbohydrates. *Biotechnol. Adv.* 39, 107465. doi:10.1016/j.biotechadv.2019.107465.

Maruthamuthu, M., Jiménez, D. J., Stevens, P., and Dirk Van Elsas, J. (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.

McClendon, J. H. (1999). The origin of life. *Earth Sci. Rev.* 47, 71–93. doi:10.1016/S0012-8252(99)00015-X.

Monteagudo-Mera, A., Arthur, J. C., Jobin, C., Keku, T., Bruno-Barcena, J. M., Azcarate-Peril, M. A., et al. (2016). High purity galacto-oligosaccharides (GOS) enhance specific Bifidobacterium species and their metabolic activity in the mouse gut microbiome. *Benef Microbes* 7, 247–264. doi:10.3920/BM2015.0114.

Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., et al. (1990). Thermostable beta-galactosidase from the archaebacterium *Sulfolobus solfataricus*: Purification and properties. *Eur. J. Biochem.* 187, 321–328. doi:10.1111/j.1432-1033.1990.tb15308.x.

Ranawat, P., and Rawat, S. (2017). Radiation resistance in thermophiles: mechanisms and applications. *World J. Microbiol. Biotechnol.* 33, 1–22. doi:10.1007/s11274-

017-2279-5.

Rani, V., Sharma, P., and Dev, K. (2019). Characterization of thermally stable β-galactosidase from *Anoxybacillus flavithermus* and *Bacillus licheniformis* isolated from Tattapani hot spring of North Western Himalayas. *India Artic. Int. J. Curr. Microbiol. Appl. Sci.* 8(1), 2517-2542. doi:10.20546/ijcmas.2019.801.266.

Saqib, S., Akram, A., Halim, S. A., and Tassaduq, R. (2017). Sources of β-galactosidase and its applications in food industry. *3 Biotech* 7, 79. doi:10.1007/s13205-017-0645-5.

Seijo, M., Bryk, G., Zeni Coronel, M., Bonanno, M., Río, M. E., Pita Martín de Portela, M. L., et al. (2019). Effect of adding a galacto-oligosaccharides/fructo-oligosaccharides (GOS/FOS®) mixture to a normal and low calcium diet, on calcium absorption and bone health in ovariectomy-induced osteopenic rats. *Calcif. Tissue Int.* 104, 301–312. doi:10.1007/s00223-018-0490-5.

Sharma, S. K., and Leblanc, R. M. (2017). Biosensors based on β-galactosidase enzyme: Recent advances and perspectives. *Anal. Biochem.* 535, 1–11. doi:10.1016/j.ab.2017.07.019.

Shokryazdan, P., Faseleh Jahromi, M., Navidshad, B., and Liang, J. B. (2017). Effects of prebiotics on immune system and cytokine expression. *Med. Microbiol. Immunol.* 206, 1–9. doi:10.1007/s00430-016-0481-y.

Thongaram, T., Hoeflinger, J. L., Chow, J., and Miller, M. J. (2017). Prebiotic galactooligosaccharide metabolism by probiotic *Lactobacilli* and *Bifidobacteria*. 65, 4184–4192 doi:10.1021/acs.jafc.7b00851.

Wang, S., Guo, G., Li, L., Cao, L., Tong, L., Ren, G., 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.

Whisner, C. M., and Castillo, L. F. (2018). Prebiotics, bone and mineral metabolism. *Calcif. Tissue Int.* 102, 443–479. doi:10.1007/s00223-017-0339-3.

Xavier, J. R., Ramana, K. V., and Sharma, R. K. (2018). β-galactosidase: Biotechnological applications in food processing. *J. Food Biochem.* 42, e12564. doi:10.1111/jfbc.12564.

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.

Zolnere, K., and Ciprovica, I. (2017). The comparison of commercially available β-galactosidases for dairy industry : review. Research for rural Development, Annual 23rd International Scientific Conference Proceedings, Latvia, 1.

Objectives

The main objective of the research work planned in this doctoral thesis is to analyze As Burgas geothermal spring from a metagenomic perspective, especially focusing on its biotechnological potential as a reservoir of new thermostable enzymes such as β-galactosidases. Thermostable β-galactosidases have significant advantages, both for obtaining low-lactose dairy products and for the production of GOS, compared to thermolabile enzymes. Therefore, thermal stability is a desirable quality for the use of β-galactosidases in industrial, biotechnological, and pharmaceutical applications.

The specific objectives in this work are

1. To find and characterize novel thermostable β-galactosidases from As Burgas geothermal spring through functional metagenomics.

2. To analyze the taxonomical diversity and functional potential of the microbial community inhabiting As Burgas water and to discover and characterize new β-galactosidases using sequence metagenomics.

3. To compare the biodiversity and community composition of As Burgas with a nearby hot spring (Muiño Da Veiga) and other geographically distant thermal springs.

Chapter 1

Identification and characterization of a novel thermostable β-galactosidase from As Burgas water discovered through functional metagenomics

The content of this chapter is under patent secret

Chapter 2

Exploring the taxonomical and functional profile of As Burgas hot spring: Characterization of a thermostable β-galactosidase found through sequence metagenomics

ABSTRACT

In the present study, we investigate the microbial community inhabiting As Burgas geothermal spring, located in Ourense (Galicia, Spain). The approximately 23 Gbp of Illumina sequences generated for each replicate revealed a complex microbial community dominated by Bacteria in which Proteobacteria and Aquificae were the two prevalent phyla. An association between the two most abundant genera, *Thermus* and *Hydrogenobacter*, was suggested by the relationship of their metabolism. The high relative amount of sequences involved in the Calvin-Benson cycle and the reductive TCA cycle unveils the dominance of an autotrophic population. Important pathways from the nitrogen and sulfur cycle are potentially taking place in As Burgas hot spring. In the assembled reads, two complete ORFs matching GH2 β-galactosidases were found. To assess their functional characterization, the two ORFs were cloned and overexpressed in *E. coli*. The pTsbg enzyme had activity towards o-Nitrophenyl-β-D-galactopyranoside (ONPG) and p-Nitrophenyl-β-D-fucopyranoside, with high thermal stability and showing maximal activity at 85 °C and pH 6, nevertheless the enzyme failed to hydrolyze lactose. The other enzyme, Tsbg, was unable to hydrolyze even ONPG or lactose. This finding highlights the challenge of finding novel active enzymes based only on their sequence.

INTRODUCTION

Thermophiles, growing optimally at temperatures over 55 °C, are found in hot environments such as fumaroles, hydrothermal vents, hot springs, or deserts (Takai et al., 2004; Neveu et al., 2011; Amin et al., 2017; Nagata et al., 2017). Apart from high temperatures, these habitats usually show other harsh conditions like extreme pH or high salt concentration. Therefore, the study of microorganisms inhabiting hot environments and their enzymes has drawn considerable interest from a biotechnological point of view, as these extremophiles have features suitable for industrial processes, in which high stability and activity at elevated temperatures, as well as high tolerance toward various reagents and solvents, are required.

The potential of thermal water as a source of novel thermostable biocatalysts has been demonstrated since a considerable number of thermozymes such as lipases (López-

López et al., 2015; Kaur et al., 2016), polymerases (Suharti et al., 2015), or cellulases (Zarafeta et al., 2016), among others, have been isolated from hot springs. In recent years, metagenomics has become a powerful tool to explore the microbiological community composition and activity of extreme environments, like hot springs, whose conditions are difficult to reproduce in a lab-bench. The metagenomic approach is based on the study of the whole environmental microbial DNA (metagenome) that is directly sequenced, in what is called sequence metagenomics, or ligated into a vector and transformed to generate a metagenomic library, in what is known as functional metagenomics. Sequence metagenomics has enabled the study of a large number of hot springs extended all over the world like Tuwa, Lasundra, and Unkeshwar hot springs in India (Mangrola et al., 2015a, 2015b; Mehetre et al., 2016), a hot spring in Kamchatka, Russia   (Eme et al., 2013), Sungai Klah hot spring in Malaysia (Chan et al., 2015), or several hot springs in Yellowstone National Park USA (Inskeep et al., 2010; Klatt et al., 2013; Colman et al., 2016).

β-galactosidases catalyze the hydrolysis of lactose to glucose and galactose and they have drawn considerable interest from the biotechnological industry for the production of low-lactose milk and the revalorization of whey. Furthermore, some β-galactosidases can transfer the galactosyl residue of lactose carrying transgalactosylations reactions, which are frequently used for the synthesis of galacto-oligosaccharides (GOS), attractive prebiotics (Panesar et al., 2018), and to synthesize other galactosylated products (Wojciechowska et al., 2018). Metagenomics has contributed to the exploration of heated habitats such as hot springs, either for ecological study or for bioprospection of novel enzymes. Two thermal enzymes with β-galactosidase activity have been isolated from hot springs using functional metagenomics (Gupta et al., 2012; Schröder et al., 2014), but there is only one reported study of thermostable β-galactosidases found in hot springs through sequence metagenomics (Liu et al., 2015).

In Ourense, there are at least 13 geothermal springs widespread across the region. Because of its accessibility and its historical importance, in this study, we have focused on As Burgas hot spring. Although some authors have previously investigated its water composition (González-Barreiro et al., 2009), or its culturable microorganisms (Leira et

al., 2017), the present is the first reported metagenomic study of this hot spring. From the unassembled reads obtained through shotgun metagenomic DNA sequencing, we have assessed taxonomical and functional characteristics of As Burgas water population. Then, metagenomic sequences were assembled and annotated, finding two potential β-galactosidases that have been cloned, purified, and characterized.

MATERIALS AND METHODS

Sampling

Thermal water, with temperature 66.3 °C and pH 7.56 (González-Barreiro et al., 2009) was collected from As Burgas hot spring (GPS 42.334626, -7.865332), in Ourense (Galicia, Spain), following the same procedure described in chapter 1. Briefly, two samples (BW1 and BW2) of 50 L of water were collected into bottles which were previously prewashed with 70 % ethanol and rinsed with thermal water. The water samples were stored at room temperature until the next day when water was filtered through a nitrocellulose filter of 0.2 μm. Filters were preserved at -20 °C until metagenomic DNA extraction.

DNA extraction and shotgun sequencing

Total DNA was isolated from the filters using the Metagenomic DNA Isolation Kit for Water (Epicentre Biotechnologies), according to the manufacturer's protocol. Metagenomic DNA of both replicates was quantified using Qubit dsDNA HS Assay kit (Invitrogen) and prepared for Next Generation Sequencing using the Accel-NGS® 2S Plus DNA Library Kit (Swift Biosciences). The amplified libraries were checked with a Bioanalyzer 2100 (Agilent Technologies) and concentrations were quantified by Qubit dsDNA HS Assay kit (Invitrogen). Paired-end sequencing of the metagenomic DNA libraries was performed with 2 x 300 bp using the MiSeq sequencer (Illumina, San Diego, CA, USA) at San Diego State University.

Taxonomic and functional assignment of metagenomic sequences

Illumina reads were treated with PRINSEQ software (Schmieder et al., 2011) for quality control, removing all artificial duplicate reads and reads shorter than 60 base-pairs.

High-quality unassembled reads of both replicates were uploaded into the Metagenomics Rapid Annotation using the Subsystem Technology (MG- RAST) v4.0.3 server (Meyer et al., 2008) and are available under the accession numbers mgm4709017.3 (BW1) and mgm4709018.3 (BW2). MG-RAST is an automated annotation pipeline in which taxonomic assignment is done with BLAT comparisons (Wilke et al., 2012) to the NCBI and gene functional potential with BLAT comparisons to the SEED protein database (Meyer et al., 2008). Sequence annotations were performed using the following parameters: cut off e-value 10^{-5}, minimum 60 % identity, and >15 bp alignment length.

To reduce the differences related to library size, relative abundance was calculated as the percentage of reads assigned to a taxon or gene function in proportion to the total number of annotated reads.

Sequence assembly and screening for sequences annotated as β-galactosidase

Paired-end unassembled high-quality reads were merged using PEAR (Zhang et al., 2014) and assembled with the SPAdes pipeline (Bankevich et al., 2012). Then, assembled reads were uploaded to MG-RAST for functional annotation with the SEED subsystem database (cut off e-value 10^{-5}, minimum 60 % identity, and >15 bp alignment length). The contigs that contained β-galactosidases sequences were downloaded and analyzed for all possible open reading frames (ORFs) using NCBI ORF finder (Wheeler et al., 2003). The ORFs and the deduced amino acid sequence were compared with other known sequences using nucleotide-nucleotide and protein-protein basic local alignment search tool (BLASTN and BLASTP) search (Altschul et al., 1990). The Pfam 32.0 web server, based on Pfam family database (El-Gebali et al., 2019) was used to infer the conserved domains within the amino acidic sequences.

Cloning, expression, and purification of *T.scotoductus* β-galactosidases

T.scotoductus β-galactosidase (*Tsbg*) and putative β-galactosidase (*pTsbg*) ORFs were amplified directly from the metagenomic DNA with the primers listed in table 1 and both were cloned in the pDONR211 vector using the Invitrogen Gateway Technology (Invitrogen). From the gateway vector, the gene was shuttled into the his-tagged

expression vector pDEST-527, using the Gateway LR recombination reaction (Invitrogen). The constructions were transformed and expressed in T7 Express (C2566) *E.coli* (NEB). Induction was done with 0.4 mM IPTG for 2 hours at 37 °C. Cells were collected by centrifugation (5000 rpm for 15 min 4 °C) and resuspended in 20 mM sodium phosphate buffer 500 mM NaCl (pH 7.2) and Complete Mini protease inhibitor cocktail (Roche), following the manufacturer instructions. Cell disruption was done by sonication on ice using Vibra Cell sonicator (100 W, 5 min 2" ON/8" OFF), (Sonics & Materials). The resulting crude extract was preheated at 70 °C for 10 min to denature *E.coli* proteins, as suggested by Pessela et al., 2004. Then, the clear lysate obtained after centrifugation (14000 rpm for 20 min) was passed through a HisTrap^{TM} HP column (GEHealthcare), following the manufacturers' protocol and using an ÄKTA chromatography system (GEHealthcare). Briefly, the column was equilibrated with 20 mM sodium phosphate buffer 500 mM NaCl and 20 mM imidazole (pH 7.2) and the elution of the bound Tsbg and pTsbg His-tagged fusion proteins was done with a 20 mM sodium phosphate buffer 500 mM NaCl and 500 mM imidazole (pH 7.2). The selected fractions were concentrated and dialyzed using an Amicon Ultra-15 30,000 MWCO column (Millipore). Purified protein concentration was quantified according to the Bio-Rad Protein Assay (Bio-Rad), employing bovine serum albumin as a standard. Protein samples of the different stages of the purification were run in a 10 % SDS-PAGE gel for its molecular weight determination. NZYcolour Protein Marker II (Nzytech) was used as molecular weight standard and proteins were detected by staining with Coomassie Brilliant Blue.

Table 1. Primers used for the amplification of *T.scotoductus* β-galactosidase and putative β-galactosidase ORFs

ORF amplified	Name	Sequence
T.scotoductus β-gal	MECA01f	5´ GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAAGCTGGACCCCAACCATCCC 3´
	MECA02r	5´ GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTCCCAAAGCACCCGCCT 3´
T.scotoductus putative β-gal	MECA03f	5´ GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGGTGGGAAAGAGCTTGGTTTTTGG 3´
	MECA04r	5´ GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACCAGGCCACCCCCAGG 3

Determination of β-galactosidase activity

Enzymatic activity was measured using ortho-nitrophenyl-β-D-galactopyranoside (ONPG). Purified protein preparations were diluted in 150 µL Z buffer (100 mM

Na$_2$HPO$_4$, 40 mM NaH$_2$PO$_4$, 10 mM KCl, 1.6 mM MgSO$_4$, pH 7). After incubation for 5 min at 85 °C, the reaction was started by adding 150 μL of a solution of 4 mg mL^{-1} ONPG in Z buffer to the enzyme preparation. Aliquots (100 μL) of the reaction mixture were stopped by adding 100 μL 1 M Na$_2$CO$_3$. Released o-nitrophenol was measured by UV absorbance at 420 nm. β-galactosidase activity is expressed in enzymatic units (U), defined as the amount of enzyme capable of releasing one μmol of the product (o-nitrophenol) per min (μmol·min^{-1}·mL^{-1}) under the experimental conditions. All measurements were determined in triplicate.

Effect of pH and temperature on activity and stability of recombinant pTsbg

To estimate the effect of pH on enzyme activity, the relative activities against ONPG (4 mg mL^{-1}) were measured in the range of pH 5.0 – 8.5 using 20 mM Britton–Robinson buffer (Britton and Robinson, 1931). The influence of temperature was determined by measuring relative enzyme activities at 55 – 90 °C with ONPG (4 mg mL^{-1}) in Z buffer. The thermal stability of the protein was assessed by pre-incubation of the enzyme in Z buffer at a range of 55 – 85 °C for different times followed by an activity assay against ONPG at 85 °C. Graphics were created using Prism 6.00 for Windows (GraphPad Software Inc.).

Determination of substrate specificity and GOS production

The substrate specificity of the purified pTsbg was determined at 85 °C using 4 mg mL^{-1} solutions of the following chromogenic substrates in Z buffer (pH 7): ONPG, p-Nitrophenyl-β-D-fucopyranoside, p-Nitrophenyl-β-D-mannoside, p-Nitrophenyl-α-D-mannoside, p-Nitrophenyl-β-D-glucoside, p-Nitrophenyl- α-D-glucoside, p-Nitrophenyl-β-D-xyloside, and p-Nitrophenyl-α-D-xyloside.

GOS and lactose concentrations were determined by HPLC (HPLC Waters Breeze I), using a Waters Sugar-Pak column eluted at 90 °C with 0.1 M EDTA disodium salt in Milli-Q water at a flow rate of 0.5 mL min^{-1}, and a Waters 2414 refractive-index detector. Purified protein was incubated at 70 °C and 650 rpm in phosphate buffer 0.1 M (pH 6.8), supplemented with 40 % lactose. Samples were taken at 0, 0.5, 1, 2, 4, 6, and 24 h and immediately transferred to 99 °C for 5 min to inactivate the enzyme and

stored at -20 °C for subsequent analysis. Carbohydrates were quantified by external calibration, using standard solutions of galactose, glucose, lactose, raffinose, and stachyose.

RESULTS AND DISCUSSION

DNA extraction and shotgun sequencing

After DNA extraction from the two water samples, a total of 50 µl of high-quality metagenomic DNA were obtained with concentrations of 42.8 ng µL^{-1} for BW1 and 38.2 ng µL^{-1} for BW2. After sequencing, a total of 867,096 and 873,846 of paired-end reads were obtained for BW1 and BW2 respectively. More features of the raw sequences are collected in table 2.

Table 2. Characteristics of the paired-end raw sequences obtained after Illumina MiSeq sequencing of As Burgas water before and after quality control (QC) with PRINSEQ. Read 1 and read 2 correspond to the paired reads.

		BW1		BW2	
		Read 1	Read 2	Read 1	Read 2
Before PRINSEQ QC	**Number sequences**	867,096	867,096	873,846	873,846
	Total bases	227,341,174	232,496,706	230,953,710	235,685,441
	Seq. length (bp)	262.19 ± 46.53	268.13 ± 47.31	264.30 ± 44.08	269.71 ± 45.04
	Mean GC content (%)	54.95 ± 11.23	55.32 ± 11.61	54.09 ± 11.76	54.46 ± 12.20
	Number of pairs	867,096 (100 % sequences)		873,846 (100 % sequences)	
After PRINSEQ QC	**Number sequences**	747,684	747,684	761,635	761,635
	Total bases	193,410,210	192,903,192	199,007,260	198,412,539
	Seq. length (bp)	258.68 ± 46.62	258.00 ± 45.55	261.29 ± 43.84	260.51 ± 42.86
	Mean GC content (%)	54.22 ± 11.41	54.51 ± 11.63	53.31 ± 11.87	53.60 ± 12.11
	Number of pairs	747,684 (100 % sequences)		761,635 (100.00 % sequences)	

Taxonomic and functional assignment of metagenomic sequences

From a total of 867,096 and 873,846 paired-end sequences obtained, 747,684 (86.23 %) and 761,635 (87.15 %) metagenomic sequences were retained for BW1 and BW2 samples respectively after quality assessment with PRINSEQ, as reflected in table 2. These high-quality raw reads were uploaded to MG-RAST, where paired reads were automatically joined on the overlapping ends, and taxonomical and functional annotation of the sequences was done. As there were no significant differences between samples BW1 and BW2 (data not shown), the relative abundances of assigned reads to each taxon or function were expressed as an average between both samples.

The taxonomical community analysis revealed a predominance of Bacteria (93.11 ± 1.86 %), followed by Archaea (6.18 ± 1.84 %), Eukaryota (0.67 ± 0.009 %), and Viruses (0.02 ± 0.03 %) (Fig 1). From the 27 bacterial phyla detected, the most abundant were Proteobacteria (68.25 ± 3.59 %), Aquificae (11.24 ± 1.15 %), Deinococcus-Thermus (5.26 ± 1.01 %), Firmicutes (4.29 ± 0.53 %) and Bacteroidetes (1.95 ± 0.19 %) (Fig 2). More detailed information on the community structure is provided in the annex to chapter 2 (Tables S1 and S2).
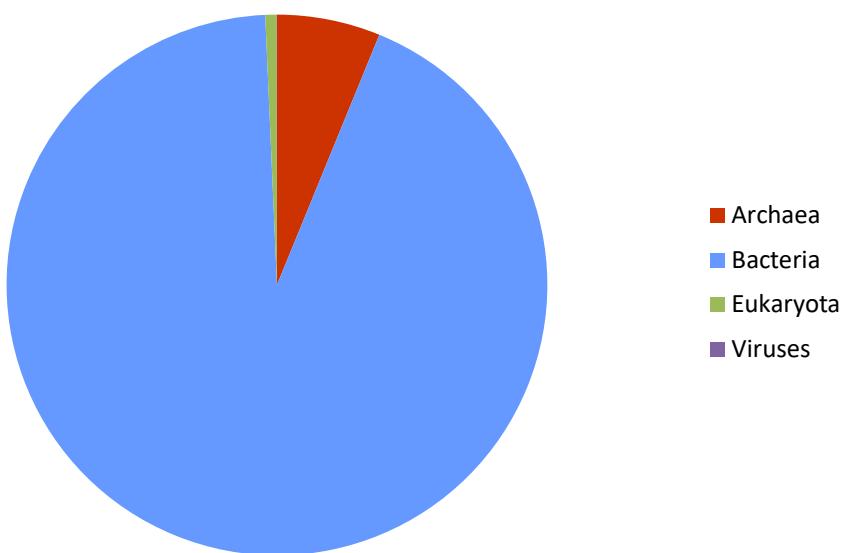
Figure 1. Taxonomic assignment of the reads at domain level. The chart represents the percentage of reads assigned to each domain (relative abundance expressed as a percentage from the total assigned reads).

The predominance of Bacteria followed by Archaea was also found in the soil and the water of the Lobios hot spring, located in the same Galician region (López-López et al., 2015; Knapik et al., 2019). Nevertheless, in contrast with the significant relative abundance of Proteobacteria found in As Burgas water, Acidobacteria was the major phylum in the Lobios sediment while Deinococcus-Thermus dominated the Lobios water. These differences might be due to the influence of physicochemical parameters, such as pH and temperature, on the microbial community composition. In fact, As Burgas water has a lower temperature (66.3 °C) and pH (7.56) (González-Barreiro et al., 2009) than Lobios water (76 °C, pH= 8.2) (López-López et al., 2015). It is also important to consider that taxonomical assignment in the study of Lobios water was done using

assembled reads rather than the unassembled reads and thus, real phyla abundance might be lost (Ju and Zhang, 2015).

Temperature has been reported as a key factor in the prevalence of Proteobacteria. Dominance of this phylum has been found in geographically distant but moderate-temperature (29 – 65 °C) geothermal springs like Deulajhari and Tattapani in India (Mohanrao Mahajan et al., 2016; Singh and Subudhi, 2016), Aguas Calientes in the Amazon rainforest of Perú (Paul et al., 2016), Chiraleu, Ciocaia, and Mihai Bravu in Romania (Chiriac et al., 2017) or El Coquito in the Colombian Andes (Bohorquez et al., 2012). Moreover, Power et al., 2018 found that phyla Proteobacteria and Aquificae dominated in 925 geothermal springs in New Zealand (65.2% total average relative abundance across all springs), especially in hot springs with temperatures below 50 °C, where Proteobacteria were the most abundant phylum. Similar results were found by Najar et al., 2018 that studied the microbial diversity of Polok (75 – 77 °C) and Borong (50 – 52 °C) hot springs in India finding that the dominance of the Phylum Proteobacteria was more pronounced in Borong hot spring, which had a lower temperature. Another distinctive aspect of Proteobacteria is that they are known to tolerate a higher concentration of sulfur and use reduced compounds of this element as an electron donor during their physiological processes (Najar et al., 2018).

Aquificae is the second most abundant phylum in As Burgas ecosystem consisting of 11.24 ± 1.15 % of the metagenome. This phylum encompasses strictly thermophilic bacteria with an optimum growth temperature above 65 °C (Griffiths and Gupta, 2006). The high relative abundance of Aquificae occurs in other hot springs with a broad range of pH and temperatures, including six geothermal springs in the Philippines (60 – 92 °C, pH 3.72 – 6.58) (Huang et al., 2013), the Mihai Bravu in Romania (Chiriac et al., 2017) and the Ganzi Prefecture hot springs in China (Tang et al., 2018). Members of this phylum dominate in environments with limited biomass and low ion concentrations, such as the King-Yu, Nono-Yu Koya, Yamanojo, and Jinata Onsen hot springs in Japan (Nishiyama et al., 2018; Ward et al., 2019), among others. Most Aquificae representatives are hydrogen-oxidizing bacteria that use hydrogen as electron donor, carbon dioxide as carbon source, and oxygen as the final electron acceptor. Alternatively, some species can oxidize thiosulfate or sulfur as energy

sources (Griffiths and Gupta, 2006). Compared with other geothermal springs worldwide, the community structure of As Burgas is very similar to the Mihai-Bravu spring in Romania, which has similar temperature and pH (65 °C, pH 7.91) (Chiriac et al., 2017), as both springs were dominated by phyla Proteobacteria, Aquificae, and Deinococcus-Thermus. This result suggests that chemolitotrophy by oxidation of H_2 and reduced sulfur compounds are important metabolic processes in these springs and that the members of phylum Aquificae play a main role in primary productivity in this community.
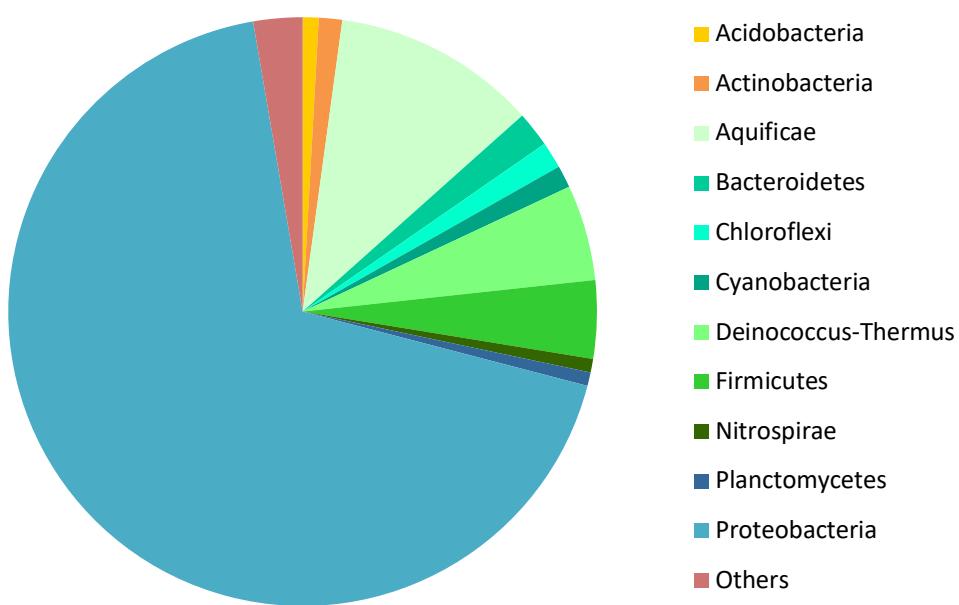
Figure 2. Taxonomic assignment of sequences within Bacteria domain. Percentage of reads annotated at phylum level is represented. Others include those phyla with less than 0,7 % sequences assigned (Candidatus Poribacteria, Chlamydiae, Chlorobi, Chrysiogenetes, Deferribacteres, Dictyoglomi, Elusimicrobia, Fibrobacteres, Fusobacteria, Gemmatimonadetes, Lentisphaerae, Spirochaetes, Synergistetes, Tenericutes, Thermotogae, unclassified (derived from Bacteria) and Verrucomicrobia).

Focusing on the genus level, the three most abundant genera in As Burgas water were *Thermus* (21,221 sequences (15.77 %)), *Hydrogenobacter* (11,517 sequences (8.56 %)) and *Thiobacillus* (5,659 sequences (4.20 %)). *Thermus* spp. has been traditionally described as heterotrophic thermophilic Gram-negative aerobic bacteria; although most are facultative anaerobes in the absence of oxygen and presence of nitrate (Cava et al., 2009), and some species from the genera have the ability to grow mixotrophically (Skirnisdottir et al., 2001; Bjornsdottir et al., 2009). The dominance of *Thermus* in As Burgas water is consistent with this genus optimal growth temperature

(62 – 75 °C) (Cava et al., 2009), in fact, members of this genus are commonly found in other thermal springs with temperatures above 60 °C. For example, in the hot springs of Heart Lake Geyser Basin in Yellowstone National Park, a shift in the microbial population was detected from several cyanobacterial genera at 44 °C to the observation of *Thermus* members at 63 °C and finally a predominance of this genus in the 75 °C geysers (Bowen De León et al., 2013). *Thermus* genus was also dominant in the 65 °C Mihai-Bravu spring in Romania (Chiriac et al., 2017) and the Rupi Basin geothermal spring in Bulgaria (Tomova et al., 2010). This genus also dominates the water of the geographically close Lobios hot spring in Ourense (López-López et al., 2015).

Hydrogenobacter was the second most abundant genus in As Burgas. These extremely thermophilic representatives of phylum Aquificae are obligate chemolithotrophic organisms with anaerobic anabolism but aerobic catabolism (Pitulle et al., 1994). High relative abundance and co-existence of *Hydrogenobacter* with *Thermus* genera was found in Lobios (López-López et al., 2015), Rupi Basin (Tomova et al., 2010), Elegedi (Ghilamicael et al., 2017) and in Niujie hot springs (Bai and Peng, 2019). The reported association between hydrogen-oxidizing *Hydrogenobacter* with hydrogen-producing *Thermus* in these hot springs suggests hydrogen metabolism as an essential component of these ecosystems.

Table 3. MG-RAST resume of the two replicates of As Burgas water metagenome (BW1 and BW2 samples MG-RAST Ids mgm4709017.3 and mgm4709018.3 respectively)

		BW1	BW2
Processed	Predicted Protein Features	347,814	368,188
	Predicted rRNA Features	45,681	47,293
Alignment	Identified Protein Features	181,371	194,410
	Identified rRNA Features	452	519
Annotation	Identified Functional Categories	152,744	163,890

In addition to the community analysis, functional analysis was performed with MG-RAST. The sequences that passed MG-RAST quality control produced 347,814 and 368,188 predicted protein-coding features for BW1 and BW2, respectively. From these, 52.1 % (181,371 sequences) for BW1 and 52.8 % (194,410 sequences) for sample BW2, were assigned annotation by MG-RAST to SEED functional categories (Subsystems) (Table 3). Among the functional categories at Level 1 identified by the

SEED subsystems annotation, the four most dominant were the clustering-based subsystems (functional coupling evidence but unknown function; 13.44 ± 0.55 %), protein metabolism (10.77 ± 0.17 %), carbohydrates (9.55 ± 0.11 %) and miscellaneous (6.42 ± 0.24 %), based in the relative abundance of assigned reads (Fig 3). Similar results were found in Lobios hot spring water where the clustering-based subsystems were found as the largest category followed by miscellaneous, carbohydrates, and protein metabolism (López-López et al., 2015). The predominance of the clustering-based subsystems in both metagenomes shows how limited our knowledge is regarding the functional annotation of the microbial proteome, as the precise functions of most proteins in metabolic pathways are yet to be revealed. Thus, the strategy of discovering new activities by a functional-driven metagenomic approach rises as a valid alternative to overcome such challenges.
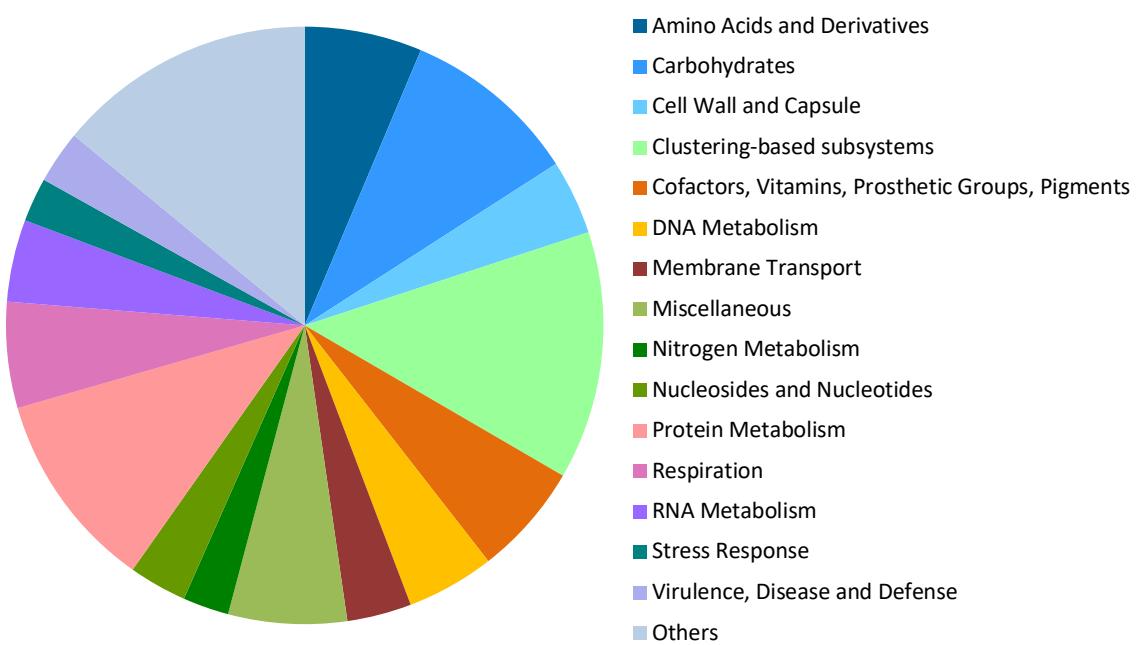
Figure 3. Functional profile of As Burgas hot spring at SEED subsystems level 1. The percentage of reads assigned to each function is represented. Others include those functions with less than 2.11 % reads assigned (Cell Division and Cell Cycle; Dormancy and Sporulation; Fatty Acids, Lipids, and Isoprenoids; Iron acquisition and metabolism; Metabolism of Aromatic Compounds; Phages, Prophages, Transposable elements, Plasmids; Phosphorus Metabolism; Photosynthesis; Potassium metabolism; Regulation and Cell signaling; Secondary Metabolism; Sulfur Metabolism; Motility and Chemotaxis).

Since O_2 concentration is reduced in hot springs due to lower oxygen solubility in heated water, other electron acceptors are important, such as nitrate, elemental S, sulfate, or CO_2. Thus, an overrepresentation of sequences related to nitrogen and

sulfur metabolism could be expected in these kinds of habitats. Consequently, in this study, we specially review those pathways involved in nitrogen and sulfur metabolism.

Analysis of the nitrogen metabolism at subsystem level 3 revealed a high abundance of sequences involved in nitrate and nitrite ammonification, also known as dissimilatory nitrate reduction to ammonium (DNRA) (Table 4). DNRA is the result of anaerobic respiration by chemoorganoheterotrophic microorganisms using nitrate (NO_3^-) as a final electron acceptor, producing ammonia (NH_4^+). This metabolic pathway results in nitrogen (N) conservation in the ecosystems and is favored in habitats where NO_3^- is limiting in relation to organic carbon (Kraft et al., 2011). Therefore, the low NO_3^- content found in As Burgas water in comparison to other proximal geothermal springs such as Outariz, Tinteiro, and Chavasqueira (González-Barreiro et al., 2009) might be promoting the prevalence of DNRA bacteria like Proteobacteria (Otte et al., 1999; Mohan et al., 2004; Giacomucci et al., 2012). This result is in accordance with the dominance of phylum Proteobacteria found in the taxonomical analysis of As Burgas metagenomic sequences. Nevertheless, it is important to remark that the presence or relative abundance of a gene in a metagenome does not mean that it is active. Metatranscriptomic studies are necessary to determine if DNRA is an important pathway in this ecosystem. In this aspect, other studies have reported the occurrence of an active DNRA pathway in some hot springs (Dodsworth et al., 2011; Tripathy et al., 2016; Alcamán-Arias et al., 2018).

A high number of reads with similarity to ammonia assimilation were found in As Burgas water metagenome (Table 4). The abundance of sequences annotated as glutamine synthetase and glutamate synthase, key enzymes in this metabolic pathway, were already expected as they are widely distributed among microorganisms, playing an important role in nitrogen metabolism (Nagatani et al., 1971).

Reads annotated as Nitrogenase (*Nif*) genes, for nitrogen fixation were also abundant in the metagenome. Although the distribution of these genes seems to be widespread in nature, as they have been described in different environments (Dos Santos et al., 2012) including hot springs (Klatt et al., 2011; Jiménez et al., 2012; Badhai et al., 2015), active nitrogen fixation has been reported in several thermophilic organisms (Wahlund

and Madigan, 1993; Mehta and Baross, 2006). Nitrogen fixation could be important in As Burgas as this ecosystem harbors phyla with known diazotrophic representatives like Proteobacteria and the phylum Aquificae in which some members of *Hydrogenobacter* were recently described as nitrogen-fixing bacteria (Nishihara et al., 2018a). Furthermore, nitrogen fixation has been demonstrated in other geothermal springs such as several hot springs from Yellowstone National Park (Hamilton et al., 2011; Loiacono et al., 2012) and Nakabusa hot springs in Japan (Nishihara et al., 2018b), among others.

Nitrification might also take place in As Burgas ecosystem. Sequences matching the ammonia monooxigenase (AMO) enzyme were detected in the two metagenomes. This enzyme catalyzes the oxidation of ammonia to hydroxylamine and it is essential for chemolithotrophic ammonia-oxidizing bacteria. The oxidation of ammonia to nitrite in As Burgas hot spring water could be associated with the abundant Proteobacteria, since several members of this phylum have been described as autotrophic nitrifiers (Rotthauwe et al., 1997; Stein and Nicol, 2018).

Another important component in the nitrogen cycle is denitrification, which competes with DNRA, due to the dependence of both metabolic pathways on NO_3^-. Members of the genus *Thermus* can perform facultative anaerobic respiration using NO_3^- as the final electron acceptor, producing N_2 or nitrous oxide (N_2O) (Cava et al., 2009). In addition, representatives from another abundant genus in As Burgas, *Thiobacillus*, also perform denitrification processes (Wood and Kelly, 1988; Yu et al., 2015). Unexpectedly, not many sequences related to denitrification were annotated in the metagenome at level 3 (771 sequences in BW1 and 692 in BW2), even though these potential denitrifiers were two of the most abundant genera found in As Burgas. At function level, sequences related to denitrification, such as nitrite reductase (*nir*), nitric-oxide reductase (*nor*), and nitrous-oxide reductase (*nos*), were present in both metagenomes, but not in high abundance.

Functions involved in sulfur oxidation were also abundant in As Burgas water (Table 4). The high abundance of these sequences can be attributed to the prevalence of Proteobacteria in the microbial community since these microorganisms are important

sulfur-oxidizing phylum (Shao et al., 2010; Watanabe et al., 2019). Numerous members of the abundant phylum Aquificae and Deinococcus-Thermus can oxidize thiosulphate or sulfur as an energy source and thus harbor *sox* genes (Skirnisdottir et al., 2001; Bjornsdottir et al., 2009; Sano et al., 2010). Moreover, some sulfur-oxidizing bacterial species of the genus *Thermus* and *Thiobacillus* are also nitrate-reducing bacteria that accept electrons from the oxidation of reduced inorganic sulfur compounds and have been frequently identified in a diverse range of geothermal springs (Wood and Kelly, 1988; Skirnisdottir et al., 2001; Bjornsdottir et al., 2009). Therefore, sulfur oxidation coupled with denitrification could be an important source of energy for carbon fixation in this hot spring, like was previously described for other hot springs (Merkel et al., 2017) and diverse heated habitats, including hydrothermal vents (Li et al., 2018).

The analysis of carbon-fixation metabolism revealed a high abundance of sequences associated with the reductive pentose phosphate cycle (Calvin-Benson cycle) (Table 4). This cycle has been described as the principal pathway of carbon fixation in Cyanobacteria and Proteobacteria (Kusian and Bowien, 1997) and some studies have reported the presence of genes related to this cycle in several *Thermus* strains (Müller et al., 2016).

The number of sequences affiliated to the tricarboxylic acid (TCA) cycle was also representative (1,742 sequences in BW1 and 1,775 in BW2), but slightly lower than those for the Calvin-Benson cycle. Most enzymes involved in the TCA cycle function in an oxidative way (releasing stored energy through the oxidation of acetyl-CoA into ATP and CO_2), but they can be used by some microorganisms in a reductive TCA cycle that is essentially the oxidative TCA cycle running in reverse, leading to the fixation of two molecules of CO_2 and the production of one molecule of acetyl-CoA (Hügler et al., 2005). Reverse TCA is suggested to be the more ancient pathway for carbon fixation (Ragsdale, 2018) and has been described as the main route for primary production at high temperatures (above 70 °C) (Hügler et al., 2007). The ability to perform the reverse TCA cycle is typical of bacteria from the phylum Aquificae such as *Hydrogenobacter* (Shima and Suzuki, 1993; Yoon et al., 1997; Ishii et al., 1998; Hügler et al., 2007; Chernyha et al., 2017) and was confirmed in a variety of anaerobic and microaerobic bacteria, including several proteobacteria (Hügler et al., 2005).

Moreover, reads annotated as pyruvate:ferredoxin oxidoreductases (POR) were found in the two metagenomes. POR enzyme decarboxylates pyruvate to form acetyl-CoA and is crucial for the reverse TCA cycle, as it is able to act as pyruvate synthase catalyzing the reverse reaction (Furdui and Ragsdale, 2000; Ikeda et al., 2010). The high abundance of sequences involved in the Calvin-Benson and reverse TCA cycles reveals that autotrophy is an important source of energy of the ecosystem, as was expected, in accordance with the low organic content of this kind of thermal habitats.

A high relative abundance of reads associated with one-carbon metabolism such as YgfZ, a folate-binding regulatory protein (Teplyakov et al., 2004), and sequences related to the serine-glyoxylate cycle (Table 4) was identified. Serine-glyoxylate cycle is a carbon assimilation pathway found in aerobic methanotrophs belonging to the classes Alpha-, Gammaproteobacteria, and the phylum Verrucomicrobia (But et al., 2019). Sequences annotated as crucial enzymes for methanotrophic metabolism such as methane monooxygenase, methanol dehydrogenase or hydroxypyruvate reductase (Hanson and Hanson, 1996; Baik et al., 2003) were present in the two replicates of As Burgas metagenome. A similar result was previously reported for the nearby Lobios hot spring, in which a high abundance of sequences associated with YgfZ and the serine-glyoxylate cycle was also detected. However, Lobios metagenome lacks the methane monooxygenase and methanol dehydrogenase encoding genes (López-López et al., 2015). The methanogenic microorganisms frequently found in hot springs microbial mats (Karnauchow et al., 1992; Hedlund et al., 2013; Merkel et al., 2015) would be the methane producers for methanotrophs in As Burgas. In fact, sequences annotated to the methanogenic orders Methanobacteriales, Methanocellales, Methanomicrobiales, Methanosarcinales, and Methanopyrales were found among the archaeal reads in the taxonomical analysis of As Burgas. Moreover, sequences matching several proteins involved in methanogenesis such as heterodisulfite reductase, formate dehydrogenase, and carbon monoxide dehydrogenase were found in the metagenome. Nevertheless, the presence of methyl-coenzyme M reductase gene, a key enzyme in methanogenesis (Lyu et al., 2018), was not detected in the metagenome.

Table 4. Analysis of Subsystems at level 3. From the 28 subsystems at level 3 registered by MG-RAST, only those subsystems with more than 2,000 reads assigned were collected in the table.

SEED Subsystems			No. of reads	
Level 1	Level 2	Level 3	BW1	BW2
Amino Acids and Derivatives	Branched-chain amino acids	Branched-Chain_Amino_Acid_Biosynthesis	2,799	2,930
	Lysine, threonine, methionine, and cysteine	Methionine_Biosynthesis	3,119	3,134
Carbohydrates	CO_2 fixation	Calvin-Benson_cycle	2,204	2,351
	One-carbon Metabolism	Serine-glyoxylate_cycle	3,311	3,219
Cell Wall and Capsule	NULL	Peptidoglycan_Biosynthesis	2,039	2,106
Clustering-based subsystems	NULL	Bacterial_Cell_Division	2,293	2,119
Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	Heme_and_Siroheme_Biosynthesis	2,244	2,205
	Folate and pterines	YgfZ	2,527	2,741
DNA Metabolism	DNA replication	DNA-replication	2,446	2,332
	DNA repair	DNA_repair,_UvrABC_system	2,038	2,019
	DNA repair	DNA_repair,_bacterial	2,532	2,661
Fatty Acids, Lipids, and Isoprenoids	Fatty acids	Fatty_Acid_Biosynthesis_FASII	2,365	2,459
Membrane Transport	ABC transporters	ABC_transporter_branched-chain_amino_acid_(TC_3.A.1.4.1)	2,598	2,380
Motility and Chemotaxis	Flagellar motility in Prokaryota	Flagellum	3,061	2,874
Nitrogen Metabolism	NULL	Ammonia_assimilation	2,028	2,279
	NULL	Nitrate_and_nitrite_ammonification	4,965	4,169
Nucleosides and Nucleotides	Purines	De_Novo_Purine_Biosynthesis	2,868	3,365
	Purines	Purine_conversions	2,695	2,672
Phosphorus Metabolism	NULL	Phosphate_metabolism	3,204	3,396
Protein Metabolism	Protein folding	Protein_chaperones	2,584	2,551
	Protein degradation	Proteolysis_in_bacteria,_ATP-dependent	2,054	1,933
	Protein biosynthesis	Ribosome_LSU_bacterial	4,896	4,258
	Protein biosynthesis	Ribosome_SSU_bacterial	3,084	2,958
	Protein biosynthesis	Universal_GTPases	2,110	2,012
Respiration	Electron donating reactions	Respiratory_Complex_I	5,524	5,578
	Electron accepting reactions	Terminal_cytochrome_C_oxidases	3,054	2,826
RNA Metabolism	Transcription	RNA_polymerase_bacterial	3,475	3,402
	RNA processing and modification	tRNA_modification_Archaea	1,801	2,035
Sulfur Metabolism	NULL	Sulfur_oxidation	3,112	2,665

Sequence assembly and screening for sequences annotated as β-galactosidase

From the 873,846 quality paired-end BW2 raw reads, a total of 28,296 contigs with a maximum length of 263,962 and an average length of 932pb (26,379,150 bp) were obtained using SPADes. From these, 26,417 sequences (93.36 %) were annotated to the functional level with the MG-RAST. A search for β-galactosidase sequences with

this tool resulted in only 2 sequences that harbor complete coding ORFs that were chosen for further study. Both selected ORFs belong to *Thermus scotoductus* SA-01, as their nucleotidic sequence had 100 % alignment with the *T. scotoductus* SA-01 complete genome, deposited in the GenBank by Gounder et al. (2011) under the accession number CP001962.1. This result is consistent with the dominance of *Thermus* genera reported in the taxonomical analysis. The deduced protein sequence of Tsbg and pTsbg consisted of 574 and 690 residues, respectively, and showed 100 % homology with two different β-galactosidases from *T.scotoductus* with GeneBank accession number WP_015717803.1 and WP_015717801.1 for Tsbg and pTsbg respectively. The two proteins have been registered in GeneBank as part of a whole shotgun genome sequencing and annotation, but their cloning and expression have never been reported, therefore we selected both ORFs for further study and characterization. Both protein sequences contain a Glycosyl hydrolases family 2 (GH2) TIM barrel Domain (PF02836) according to Pfam protein database (El-Gebali et al., 2019). Therefore they are included within the GH2 superfamily, in agreement with other thermostable microbial β-galactosidases like those from *Thermotoga maritima* (Talens-Perales et al., 2016) or *Streptococcus thermophilus* (Geiger et al., 2016).

Cloning, expression, and purification of *T.scotoductus* β-galactosidases

Both sequences were efficiently amplified, cloned in pDEST-527 vector, and overexpressed in T7 Express *E.coli*. As no activity towards ONPG or lactose was detected for Tsbg, the gene was cloned in pDEST-527 without the histidine tag, in an attempt to discard the possibility of an incorrect folding or blocking of the active site due to the tag. Nevertheless, purified Tsbg protein without tag did not show activity using both lactose and ONPG as substrates. The lack of β-galactosidase activity in Tsbg is similar to the results obtained for its close relative *T. scotoductus* DSM 8553, as no β-galactosidase activity was detected in this strain (Yu et al., 2013; Ullah Khan et al., 2017). Therefore, the successive characterization steps were only performed with the pTsbg.

Effect of pH and temperature on activity and stability of recombinant pTsbg

pTsbg showed maximal activity at pH 6.0 in Britton-Robinson buffer using ONPG as substrate (Fig 4). This result is slightly lower than the optimum pH reported for other bacteria from *Thermus* genera like *T.thermophilus* HB8 (Macluńska et al., 1998), *T.thermophilus* HB27 (Li et al., 2010) and it is comparable to the optimal pH reported for other thermostable β-galactosidases such as those from *Bacillus licheniformis* (Jin and Yoon, 2014), *Caldicellulosiruptor saccharolyticus* (Park and Oh, 2010), *Marinomonas* sp. BSi20414 (Ding et al., 2017) and much lower than the pH 7.8 reported for *T. oshimai* DSM 12092 β-galactosidase (Gezgin et al., 2013).
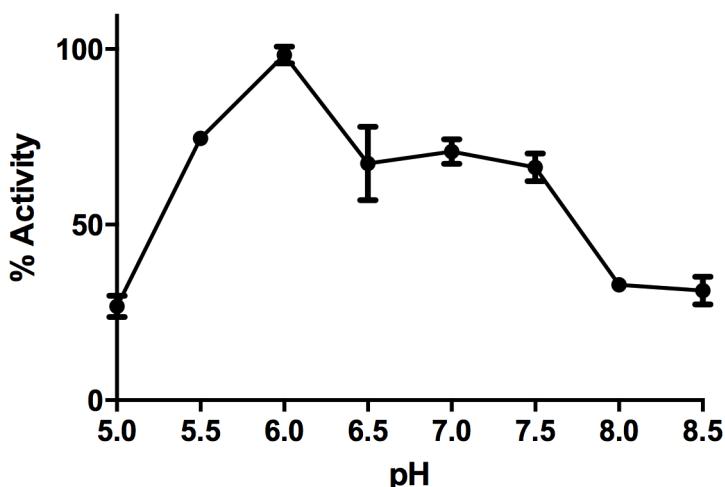
Figure 4. Effect of pH on the activity of pTsbg in Z Buffer using ONPG (4 mg mL^{-1}) as substrate.

As shown in figure 5, maximal pTsbg β-galactosidase activity towards ONPG was found at 85 °C. This is higher than the optimal temperature described using the same substrate for other counterparts of the genus *Thermus* such as *T.thermophilus* HB8 (Macluńska et al., 1998), *T.thermophilus* HB27 (Li et al., 2010), *T. aquaticus* YT-1 (Berger et al., 1997), *T. oshimai* DSM 12092 (Gezgin et al., 2013) and is the same reported as optimal to *T.thermophilus* KNOUC114 β-galactosidase (Sook Nam et al., 2012). When compared to other genera of thermophilic bacteria β-galactosidases, pTsbg showed higher optimal temperature than documented for the extremely thermophilic *C. saccharolyticus* and *Marinomonas* sp. BSi20414, which showed an optimum temperature at 80 °C and 60 °C respectively (Park and Oh, 2010; Ding et al.,

2017). Nevertheless, the optimal temperature described for *Thermotoga naphthophila* RUK-10 β-galactosidase is higher (Kong et al., 2014).
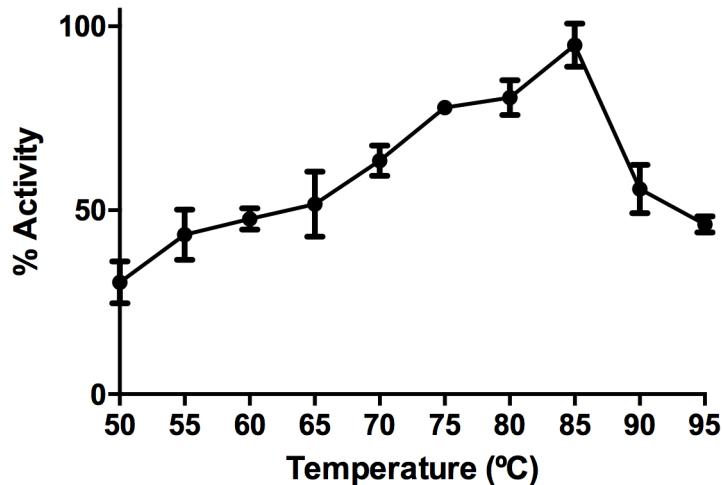
Figure 5. Effect of temperature on the activity of pTsbg in Z Buffer using ONPG (4 mg mL^{-1}) as substrate.

In relation to the thermal stability, pTsbg was able to retain up to 60 % of its maximal activity towards ONPG after 24 hours of incubation at 75 °C (Fig 6).
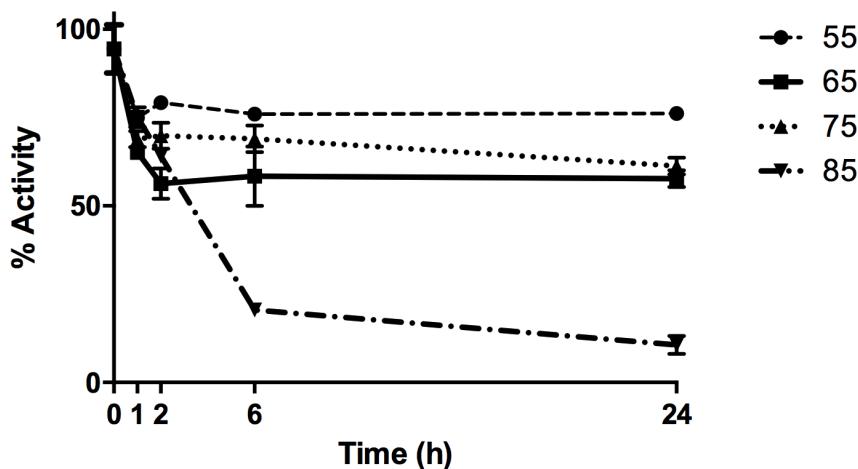
Figure 6. Effect of temperature on the stability of purified pTsbg.

Determination of substrate specificity of pTsbg

Although substrate specificity of the enzyme was studied using various chromogenic substrates, pTsbg was only active towards ONPG and p-Nitrophenyl-β-D-fucopyranoside (Table 4). Moreover, the enzyme was unable to hydrolyze lactose and no transgalactosylation was observed in the presence of this substrate, as was determined by HPLC after carrying the reaction with 40 % lactose at 70 °C and using a

mix of galactose, glucose, lactose, raffinose, and stachyose as standard (data not shown).

The preference for β-linked galactosidic substrates such as ONPG or p-Nitrophenyl-β-D-galactopyranoside over lactose has been frequently described in the characterization of β-galactosidases (Hung and Lee, 2002; Iqbal et al., 2011; Lee et al., 2011; Kong et al., 2014). Similar to our results with pTsbg, other studies have reported β-galactosidases with activity towards ONPG but unable to hydrolyze their natural substrate, lactose, *in vitro* such as YesZ β-galactosidase from *Bacillus subtilis* (Carneiro et al., 2018), or the β-Gal II from *Bifidobacterium adolescentis* DSM 20083 (Van Laere et al., 2000). The lack of β-galactosidase activity towards lactose reduces considerably the biotechnological potential of pTsbg, as it could not be applied to produce GOS from lactose and to generate lactose-free dairy products. Nevertheless, more studies focused on the fucosidase activity should be conducted, since pTsbg showed high activity with p-Nitrophenyl-β-D-fucopyranoside and may harbor fucosyltransferase activity that could be used for the synthesis of fucosylated oligosaccharides (FUCOS) with biological interest (Guzmán-Rodríguez et al., 2019) such as those from human milk.

Table 4. Activity of purified pTsbg enzyme with several nitrophenyl-derived chromogenic substrates.

Substrate (4 mg/mL)	Activity
o-Nitrophenyl-β-D-galactopyranoside (ONPG)	+
p-Nitrophenyl-β-D-fucopyranoside	+
p-Nitrophenyl-β-D-mannoside	-
p-Nitrophenyl-α-D-mannoside	-
p-Nitrophenyl-β-D-glucoside	-
p-Nitrophenyl- α-D-glucoside	-
p-Nitrophenyl-β-D-xyloside	-
p-Nitrophenyl-α-D-xyloside	-

CONCLUSIONS

The taxonomical analysis of As Burgas hot spring metagenome reveals a microbial community dominated by Bacteria in which Proteobacteria (68.25 ± 3.59 %) and Aquificae (11.24 ± 1.15 %) are the most abundant phyla. The prevalence of the genera *Thermus* (15.77 %) and *Hydrogenobacter* (8.56 %) and the relation of their metabolism suggests an association between these two genera.

Moreover, the high relative abundance of sequences involved in the Calvin-Benson cycle and sequences annotated as key for the reductive TCA cycle unveils the dominance of an autotrophic population. Important pathways from the nitrogen and sulfur cycle such as DNRA, nitrification, or sulfur oxidation are potentially taking place in As Burgas hot spring, as was determined by the functional annotation of the metagenomic reads and in accordance with the microbial composition of the ecosystem.

After assembling the metagenomic reads, two complete ORFs annotated as β-galactosidases were found. Both of them showed 100 % homology with *T.scotoductus* SA-01 and were cloned and overexpressed in *E.coli*. The enzyme Tsbg lacked β-galactosidase activity using ONPG and lactose as substrates. On the contrary, pTsbg showed β-galactosidase activity towards ONPG, but was not able to hydrolyze lactose; it showed β-fucosidase activity on the substrate p-Nitrophenyl-β-D-fucopyranoside, which suggests *a priori* unexpected biotechnological application. Once more this result reveals that the presence of a gene in a metagenome does not mean that it is active in the way predicted from the sequence, and highlights the importance of combining both functional and sequence metagenomics to find novel enzymes from metagenomes.

Our culture-independent study has provided an insight into the diversity of the microorganisms that inhabit As Burgas thermal environment, in an attempt to find novel β-galactosidases. Future research should be directed to characterize new environments, which will lead to a better understanding of their ecological differences, and to find new enzymes of interest.

REFERENCES

Alcamán-Arias, M. E., Pedrós-Alió, C., Tamames, J., Fernández, C., Pérez-Pantoja, D., Vásquez, M., et al. (2018). Diurnal changes in active carbon and nitrogen pathways along the temperature gradient in porcelana hot spring microbial mat. *Front. Microbiol.* 9. doi:10.3389/fmicb.2018.02353.

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.

Amin, A., Ahmed, I., Salam, N., Kim, B.-Y., Singh, D., Zhi, X.-Y., et al. (2017). Diversity and distribution of thermophilic bacteria in hot springs of Pakistan. *Microb. Ecol.* 74, 116–127. doi:10.1007/s00248-017-0930-1.

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.

Bai, S., and Peng, X. (2019). Distinct microbial composition and functions in an underground high-temperature hot spring at different depths. *Biogeosciences Discuss.*, 1–31. doi:10.5194/bg-2019-406.

Baik, M. H., Newcomb, M., Friesner, R. A., and Lippard, S. J. (2003). Mechanistic studies on the hydroxylation of methane by methane monooxygenase. *Chem. Rev.* 103, 2385–2419. doi:10.1021/cr950244f.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–77. doi:10.1089/cmb.2012.0021.

Berger, J.-L., Lee, B. H., and Lacroix, C. (1997). Purification, properties and characterization of a high-molecular-mass β-galactosidase isoenzyme from *Thermus aquaticus* YT-I. *Biotechnol. Appl. Biochem.* 25, 29–41. doi:10.1111/j.1470-8744.1997.tb00411.x.

Bjornsdottir, S. H., Petursdottir, S. K., Hreggvidsson, G. O., Skirnisdottir, S., Hjorleifsdottir, S., Arnfinnsson, J., et al. (2009). *Thermus islandicus* sp. nov., a mixotrophic sulfur-oxidizing bacterium isolated from the Torfajokull geothermal area. *Int. J. Syst. Evol. Microbiol.* 59, 2962–2966. doi:10.1099/ijs.0.007013-0.

Bohorquez, L. C., Delgado-Serrano, L., López, G., Osorio-Forero, C., Klepac-Ceraj, V., Kolter, R., et al. (2012). In-depth Characterization via complementing culture-independent approaches of the microbial community in an acidic hot spring of the Colombian Andes. *Microb. Ecol.* 63, 103–115. doi:10.1007/s00248-011-9943-3.

Bowen De León, 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, 1–10. doi:10.3389/fmicb.2013.00330.

Britton, H. T. S., and Robinson, R. A. (1931). CXCVIII.—Universal buffer solutions and the dissociation constant of veronal. *J. Chem. Soc.* 0, 1456–1462. doi:10.1039/JR9310001456.

But, S. Y., Egorova, S. V., Khmelenina, V. N., and Trotsenko, Y. A. (2019). Serine-glyoxylate aminotranferases from methanotrophs using different C1-assimilation pathways. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* 112, 741–751. doi:10.1007/s10482-018-1208-4.

Carneiro, L. A. B. C., Yu, L., Dupree, P., and Ward, R. J. (2018). Characterization of a β-galactosidase from *Bacillus subtilis* with transgalactosylation activity. *Int. J. Biol. Macromol.* 120, 279–287. doi:10.1016/j.ijbiomac.2018.07.116.

Cava, F., Hidalgo, A., and Berenguer, J. (2009). *Thermus thermophilus* as biological model. *Extremophiles* 13, 213–231. doi:10.1007/s00792-009-0226-6.

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.

Chernyha, N. A., Kublanova, I. V., Prokof'evaa, M. I., Pimenova, N. V., Frolova, E. N., Mardanovb, A. V., et al. (2017). Production of organic matter and diversity of the Ribulose Bisphosphate Carboxylase genes in sediments of the Solnechny spring, Uzon Caldera, Kamchatka. 86, 666–669. doi:10.1134/S0026261717050071.

Chiriac, C. M., Szekeres, E., Rudi, K., Baricz, A., Hegedus, A., Dragoş, N., et al. (2017). Differences in temperature and water chemistry shape distinct diversity patterns in thermophilic microbial communities. *Appl. Environ. Microbiol.* 83. doi:10.1128/AEM.01363-17.

Colman, D. R., Jay, Z. J., Inskeep, W. P., Jennings, R. de M., 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.

Ding, H., Zeng, Q., Zhou, L., Yu, Y., and Chen, B. (2017). Biochemical and structural insights into a novel thermostable β-1,3-galactosidase from *Marinomonas* sp. BSi20414. *Mar. Drugs* 15. doi:10.3390/md15010013.

Dodsworth, J. A., Hungate, B. A., and Hedlund, B. P. (2011). Ammonia oxidation, denitrification and dissimilatory nitrate reduction to ammonium in two US Great Basin hot springs with abundant ammonia-oxidizing archaea. *Environ. Microbiol.* 13, 2371–2386. doi:10.1111/j.1462-2920.2011.02508.x.

Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C., and Dixon, R. (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* 13, 1–12. doi:10.1186/1471-2164-13-162.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi:10.1093/nar/gky995.

Eme, L., Reigstad, L. J., Spang, A., Lanzén, A., Weinmaier, T., Rattei, T., et al. (2013). Metagenomics of Kamchatkan hot spring filaments reveal two new major (hyper)thermophilic lineages related to Thaumarchaeota. *Res. Microbiol.* 164, 425–438. doi:10.1016/j.resmic.2013.02.006.

Furdui, C., and Ragsdale, S. W. (2000). The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. *J. Biol. Chem.* 275, 28494–28499. doi:10.1074/jbc.M003291200.

Geiger, B., Nguyen, H. M., Wenig, S., Nguyen, H. A., Lorenz, C., Kittl, R., et al. (2016). From by-product to valuable components: Efficient enzymatic conversion of lactose in whey using β-galactosidase from *Streptococcus thermophilus*. *Biochem. Eng. J.* 116, 45–53. doi:10.1016/j.bej.2016.04.003.

Gezgin, Y., Tanyolac, B., and Eltem, R. (2013). Some characteristics and isolation of novel thermostable β-galactosidase from *Thermus oshimai* DSM 12092. *Food Sci. Biotechnol.* 22, 63–70. doi:10.1007/s10068-013-0009-9.

Ghilamicael, A. M., Budambula, N. L. M., Anami, S. E., Mehari, T., and Boga, H. I. (2017). Evaluation of prokaryotic diversity of five hot springs in Eritrea. *BMC Microbiol.* 17, 203. doi:10.1186/s12866-017-1113-4.

Giacomucci, L., Purdy, K. J., Zanardini, E., Polo, A., and Cappitelli, F. (2012). A new non-degenerate primer pair for the specific detection of the Nitrite Reductase Gene *nrfA* in the genus *Desulfovibrio*. *J. Mol. Microbiol. Biotechnol.* 22, 345–351. doi:10.1159/000345768.

González-Barreiro, C., Cancho-Grande, B., Araujo-Nespereira, P., Cid-Fernández, J. A., and Simal-Gándara, J. (2009). Occurrence of soluble organic compounds in thermal waters by ion trap mass detection. *Chemosphere* 75, 34–47. doi:10.1016/J.CHEMOSPHERE.2008.11.067.

Gounder, K., Brzuszkiewicz, E., Liesegang, H., Wollherr, A., Daniel, R., Gottschalk, G., et al. (2011). Sequence of the hyperplastic genome of the naturally competent *Thermus scotoductus* SA-01. *BMC Genomics* 12, 577. doi:10.1186/1471-2164-12-577.

Griffiths, E., and Gupta, R. S. (2006). Molecular signatures in protein sequences that are characteristics of the phylum Aquificae. *Int. J. Syst. Evol. Microbiol.* 56, 99–107. doi:10.1099/ijs.0.63927-0.

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.

Guzmán-Rodríguez, F., Alatorre-Santamaría, S., Gómez-Ruiz, L., Rodríguez-Serrano, G., García-Garibay, M., and Cruz-Guerrero, A. (2019). Employment of fucosidases for the synthesis of fucosylated oligosaccharides with biological potential. *Biotechnol. Appl. Biochem.* 66, 172–191. doi:10.1002/bab.1714.

Hamilton, T. L., Lange, R. K., Boyd, E. S., and Peters, J. W. (2011). Biological nitrogen fixation in acidic high-temperature geothermal springs in Yellowstone National Park, Wyoming. *Environ. Microbiol.* 13, 2204–2215. doi:10.1111/j.1462-2920.2011.02475.x.

Hanson, R. S., and Hanson, T. E. (1996). Methanotrophic bacteria. *Microbiol. Mol. Biol. Rev.* 60, 439–471.

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, Int. J. Gen. Mol. Microbiol.* 104, 71–82. doi:10.1007/s10482-013-9927-z.

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.

Hügler, M., Huber, H., Molyneaux, S. J., Vetriani, C., and Sievert, S. M. (2007). Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. *Environ. Microbiol.* 9, 81–92. doi:10.1111/j.1462-2920.2006.01118.x.

Hügler, M., Wirsen, C. O., Fuchs, G., Taylor, C. D., and Sievert, S. M. (2005). Evidence for autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle by members of the ε subdivision of proteobacteria. *J. Bacteriol.* 187, 3020–3027. doi:10.1128/JB.187.9.3020-3027.2005.

Hung, M. N., and Lee, B. (2002). Purification and characterization of a recombinant β-galactosidase with transgalactosylation activity from *Bifidobacterium infantis* HL96. *Appl. Microbiol. Biotechnol.* 58, 439–445. doi:10.1007/s00253-001-0911-6.

Ikeda, T., Yamamoto, M., Arai, H., Ohmori, D., Ishii, M., and Igarashi, Y. (2010). Enzymatic and electron paramagnetic resonance studies of anabolic pyruvate synthesis by pyruvate: ferredoxin oxidoreductase from *Hydrogenobacter thermophilus*. *FEBS J.* 277, 501–510. doi:10.1111/j.1742-4658.2009.07506.x.

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.

Iqbal, S., Nguyen, T. H., Nguyen, H. A., Nguyen, T. T., Maischberger, T., Kittl, R., et al. (2011). Characterization of a heterodimeric GH2 β-galactosidase from *Lactobacillus sakei* Lb790 and formation of prebiotic galacto-oligosaccharides. *J. Agric. Food Chem.* 59, 3803–3811. doi:10.1021/jf103832q.

Ishii, M., Yoon, K. S., Ueda, Y., Ochiai, T., Yun, N., Takishita, S., et al. (1998). Reductive TCA cycle in an aerobic bacterium, *Hydrogenobacter thermophilus* strain TK-6. *Stud. Surf. Sci. Catal.* 114, 613–616. doi:10.1016/s0167-2991(98)80834-3.

Jiménez, D. J., Andreote, F. D., Chaves, D., Montaña, 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, 1–15. doi:10.1371/journal.pone.0052069.

Jin, H. K., and Yoon, K. H. (2014). *Bacillus licheniformis* β-galactosidase. *Korean J. Microbiol. Biotechnol.* 42, 339–346. doi:10.4014/kjmb.1410.10004.

Ju, F., and Zhang, T. (2015). Experimental design and bioinformatics analysis for the application of metagenomics in environmental sciences and biotechnology. Environ. Sci. Technol. 49, 12628–12640. doi: 10.1021/acs.est.5b03719.n

Karnauchow, T. M., Koval, S. F., and Jarrell, K. F. (1992). Isolation and characterization of three thermophilic anaerobes from a St. Lucia hot spring. *Syst. Appl. Microbiol.* 15, 296–310. doi:10.1016/S0723-2020(11)80104-9.

Kaur, G., Singh, A., Sharma, R., Sharma, V., Verma, S., and Sharma, P. K. (2016). Cloning, expression, purification and characterization of lipase from *Bacillus licheniformis*, isolated from hot spring of Himachal Pradesh, India. *3 Biotech* 6, 49. doi:10.1007/s13205-016-0369-y.

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.

Knapik, K., Becerra, M., and González-Siso, M.I. (2019). Microbial diversity analysis and screening for novel xylanase enzymes from the sediment of the Lobios Hot Spring in Spain. *Sci. Rep.* 9. doi:10.1038/s41598-019-47637-z.

Kong, F., Wang, Y., Cao, S., Gao, R., and Xie, G. (2014). Cloning, purification and characterization of a thermostable β-galactosidase from *Thermotoga naphthophila* RUK-10. *Process Biochem.* 49, 775–782. doi:10.1016/j.procbio.2014.02.008.

Kraft, B., Strous, M., and Tegetmeyer, H. E. (2011). Microbial nitrate respiration - Genes, enzymes and environmental distribution. *J. Biotechnol.* 155, 104–117. doi:10.1016/j.jbiotec.2010.12.025.

Kusian, B., and Bowien, B. (1997). Organization and regulation of cbb CO_2 assimilation genes in autotrophic bacteria. *FEMS Microbiol. Rev.* 21, 135–155. doi:10.1111/j.1574-6976.1997.tb00348.x.

Lee, J. H., Kim, Y. S., Yeom, S. J., and Oh, D. K. (2011). Characterization of a glycoside hydrolase family 42 β-galactosidase from *Deinococcus geothermalis*. *Biotechnol. Lett.* 33, 577–583. doi:10.1007/s10529-010-0459-6.

Leira, M., Meijide-Failde, R., and Torres, E. (2017). Diatom communities in thermo-mineral springs of Galicia (NW Spain). *Diatom Res.* 32, 29–42. doi:10.1080/0269249X.2017.1286266.

Li, Y., Tang, K., Zhang, L., Zhao, Z., Xie, X., Chen, C.-T. A., et al. (2018). Coupled Carbon, Sulfur, and Nitrogen cycles mediated by microorganisms in the water column of a shallow-water hydrothermal ecosystem. *Front. Microbiol.* 9, 2718. doi:10.3389/fmicb.2018.02718.

Li, Y., Yao, T., Wei, M., Yan, M., Hao, N., and Xu, L. (2010). Study on the characterization of a potential thermostable β-galactosidase from *Thermus thermophilus HB27*. in *Proceedings - 2010 3rd International Conference on Biomedical Engineering and Informatics, BMEI 2010*, 2118–2121. doi:10.1109/BMEI.2010.5639979.

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.* 2818. doi:10.1080/13102818.2015.1015244.

Loiacono, S. T., Meyer-Dombard, D. R., Havig, J. R., Poret-Peterson, A. T., Hartnett, H. E., and Shock, E. L. (2012). Evidence for high-temperature in situ nifH transcription in an alkaline hot spring of Lower Geyser Basin, Yellowstone National Park.

Environ. Microbiol. 14, 1272–1283. doi:10.1111/j.1462-2920.2012.02710.x.

López-López, O., Knapik, K., Cerdán, M. E., and González-Siso, M. I. (2015). 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.

Lyu, Z., Shao, N., Akinyemi, T., and Whitman, W. B. (2018). Methanogenesis. *Curr. Biol.* 28, R727–R732. doi:10.1016/j.cub.2018.05.021.

MacIuńska, J., Czyz, B., and Synowiecki, J. (1998). Isolation and some properties of β-galactosidase from the thermophilic bacterium *Thermus thermophilus*. *Food Chem.* 63, 441–445. doi:10.1016/S0308-8146(98)00069-7.

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.

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.

Mehta, M. P., and Baross, J. A. (2006). Nitrogen fixation at 92°C by a hydrothermal vent archaeon. *Science.* 314, 1783–1786. doi:10.1126/science.1134772.

Merkel, A. Y., Pimenov, N. V., Rusanov, I. I., Slobodkin, A. I., Slobodkina, G. B., Tarnovetckii, I. Y., et al. (2017). Microbial diversity and autotrophic activity in Kamchatka hot springs. *Extremophiles* 21, 307–317. doi:10.1007/s00792-016-0903-1.

Merkel, A. Y., Podosokorskaya, O. A., Chernyh, N. A., and Bonch-Osmolovskaya, E. A. (2015). Occurrence, diversity, and abundance of methanogenic archaea in terrestrial hot springs of Kamchatka and Saõ Miguel Island. *Microbiology* 84, 577–583. doi:10.1134/S002626171504013X.

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 phylo- genetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386. doi:10.1186/1471-2105-9-386.

Mohan, S. B., Schmid, M., Jetten, M., and Cole, J. (2004). Detection and widespread distribution of the nrfA gene encoding nitrite reduction to ammonia, a short circuit in the biological nitrogen cycle that competes with denitrification. *FEMS Microbiol. Ecol.* 49, 433–443. doi:10.1016/j.femsec.2004.04.012.

Mohanrao Mahajan, M., pratap Singh, D., Kumar, K., Goyal, E., Mahesh Mohanrao, M., Pratap Singh, D., et al. (2016). Deciphering the microbial diversity of Tattapani hot

water spring Using Metagenomic Approach. *Int. J. Agric. Sci. Res.* 6, 371–382. Available at: www.tjprc.org [Accessed June 5, 2020].

Müller, W. J., Tlalajoe, N., Cason, E. D., Litthauer, D., Reva, O., Brzuszkiewicz, E., et al. (2016). Whole genome comparison of *Thermus* sp. NMX2.A1 reveals principal carbon metabolism differences with closest relation *Thermus scotoductus* SA-01. *G3 Genes, Genomes, Genet.* 6, 2791–2797. doi:10.1534/g3.116.032953.

Nagata, R., Takaki, Y., Tame, A., Nunoura, T., Muto, H., Mino, S., et al. (2017). *Lebetimonas natsushimae* sp. nov., a novel strictly anaerobic, moderately thermophilic chemoautotroph isolated from a deep-sea hydrothermal vent polychaete nest in the Mid-Okinawa Trough. *Syst. Appl. Microbiol.* 40, 352–356. doi:10.1016/J.SYAPM.2017.06.002.

Nagatani, H., Shimizu, M., and Valentine, R. C. (1971). The mechanism of ammonia assimilation in nitrogen fixing bacteria. *Arch. Mikrobiol.* 79, 164–175. doi:10.1007/BF00424923.

Najar, I. N., Sherpa, M. T., Das, S., Das, S., and Thakur, N. (2018). Microbial ecology of two hot springs of Sikkim: Predominate population and geochemistry. *Sci. Total Environ.* 637–638, 730–745. doi:10.1016/J.SCITOTENV.2018.05.037.

Neveu, J., Regeard, 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.

Nishihara, A., Matsuura, K., Tank, M., McGlynn, S. E., Thiel, V., and Haruta, S. (2018a). Nitrogenase activity in thermophilic chemolithoautotrophic bacteria in the phylum Aquificae isolated under nitrogen-fixing conditions from Nakabusa hot springs. *Microbes Environ.* 33, 394–401. doi:10.1264/jsme2.ME18041.

Nishihara, A., Thiel, V., Matsuura, K., McGlynn, S. E., and Haruta, S. (2018b). Phylogenetic diversity of Nitrogenase Reductase genes and possible nitrogen-fixing Bacteria in thermophilic chemosynthetic microbial communities in Nakabusa hot springs. *Microbes Environ.* 33, 357–365. doi:10.1264/jsme2.ME18030.

Nishiyama, E., Higashi, K., Mori, H., Suda, K., Nakamura, H., Omori, S., et al. (2018). The relationship between microbial community structures and environmental parameters revealed by metagenomic analysis of hot spring water in the Kirishima area, Japan. *Front. Bioeng. Biotechnol.* 6, 202. doi:10.3389/fbioe.2018.00202.

Otte, S., Kuenen, J. G., Nielsen, L. P., Paerl, H. W., Zopfi, J., Schulz, H. N., et al. (1999). Nitrogen, carbon, and sulfur metabolism in natural *Thioploca* samples. *Appl. Environ. Microbiol.* 65, 3148–3157.

Panesar, P. S., Kaur, R., Singh, R. S., and Kennedy, J. F. (2018). Biocatalytic strategies in the production of galacto-oligosaccharides and its global status. *Int. J. Biol. Macromol.* 111, 667–679. doi:10.1016/J.IJBIOMAC.2018.01.062.

Park, A. R., and Oh, D. K. (2010). Effects of galactose and glucose on the hydrolysis reaction of a thermostable β-galactosidase from *Caldicellulosiruptor saccharolyticus*. *Appl. Microbiol. Biotechnol.* 85, 1427–1435. doi:10.1007/s00253-

009-2165-7.

Paul, S., Cortez, Y., Vera, N., Villena, G. K., and Gutiérrez-Correa, M. (2016). Metagenomic analysis of microbial community of an Amazonian geothermal spring in Peru. *Genomics Data* 9, 63–66. doi:10.1016/j.gdata.2016.06.013.

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.

Pitulle, C., Yang, Y., Marchiani, M., Moore, E. R. B., Siefert, J. L., Aragno, M., et al. (1994). Phylogenetic position of the genus *Hydrogenobacter*. *Int. J. Syst. Bacteriol.* 44, 620–626. doi:10.1099/00207713-44-4-620.

Power, J. F., Carere, C. R., Lee, C. K., Wakerley, G. L. J., Evans, D. W., Button, M., et al. (2018). Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* 9, 2876. doi:10.1038/s41467-018-05020-y.

Ragsdale, S. W. (2018). Stealth reactions driving carbon fixation New twists to bacterial metabolic pathways that contribute to the global carbon cycle. *Science (80-.).* 359, 517–518. doi:10.1126/science.aar6329.

Rotthauwe, J. H., Witzel, K. P., and Liesack, W. (1997). The ammonia monooxygenase structural gene amoa as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations. *Appl. Environ. Microbiol.* 63, 4704–4712. doi:10.1128/aem.63.12.4704-4712.1997.

Sano, R., Kameya, M., Wakai, S., Arai, H., Igarashi, Y., Ishii, M., et al. (2010). Thiosulfate oxidation by a thermo-neutrophilic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. *Biosci. Biotechnol. Biochem.* 74, 892–894. doi:10.1271/bbb.90948.

Schmieder, R., Edwards, R., and Bateman, A. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinforma. Appl. NOTE* 27, 863–864. doi:10.1093/bioinformatics/btr026.

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.

Shao, M. F., Zhang, T., and Fang, H. H. P. (2010). Sulfur-driven autotrophic denitrification: Diversity, biochemistry, and engineering applications. *Appl. Microbiol. Biotechnol.* 88, 1027–1042. doi:10.1007/s00253-010-2847-1.

Shima, S., and Suzuki, K. I. (1993). *Hydrogenobacter acidophilus* sp. nov., a thermoacidophilic, aerobic, hydrogen-oxidizing bacterium requiring elemental sulfur for growth. *Int. J. Syst. Bacteriol.* 43, 703–708. doi:10.1099/00207713-43-4-703.

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.

Skirnisdottir, S., Hreggvidsson, G. O., Holst, O., and Kristiansson, J. K. (2001). Isolation and characterization of a mixotrophic sulfur-oxidizing Thermus scotoductus. *Extremophiles* 5, 45–51. doi:10.1007/s007920000172.

Sook Nam, E., Bong Choi, H., Hyun Lim, J., and Jin Park, H. (2012). β-Galactosidase-producing thermophilic bacterium, *Thermus thermophilus* KNOUC114: Identification of the bacterium, gene and properties of β-galactosidase. *Int. J. Biol.* 4. doi:10.5539/ijb.v4n1p57.

Stein, L. Y., and Nicol, G. W. (2018). "Nitrification," in *eLS* (Chichester, UK: John Wiley & Sons, Ltd), 1–9. doi:10.1002/9780470015902.a0021154.pub2.

Suharti, S., Hertadi, R., Warganegara, F. M., Nurbaiti, S., and Akhmaloka, A. (2015). Novel archaeal DNA Polymerase B from Domas hot spring West Java. Proceeding of 5th International Seminar on New Paradigm and Innovation on Natural Sciences and Its Application (5th ISNPINSA), 2015, Semarang, Indonesia. ISSN: 978-602-71169-7-9.

Takai, K., Hirayama, H., Nakagawa, T., Suzuki, Y., Nealson, K. H., and Horikoshi, K. (2004). *Thiomicrospira thermophila* sp. nov., a novel microaerobic, thermotolerant, sulfur-oxidizing chemolithomixotroph isolated from a deep-sea hydrothermal fumarole in the TOTO caldera, Mariana Arc, Western Pacific. *Int. J. Syst. Evol. Microbiol.* 54, 2325–2333. doi:10.1099/ijs.0.63284-0.

Talens-Perales, D., Polaina, J., and Marín-Navarro, J. (2016). Structural dissection of the active site of *Thermotoga maritima* β-galactosidase identifies key residues for transglycosylating activity. doi:10.1021/acs.jafc.6b00222.

Tang, J., Liang, Y., Jiang, D., Li, L., Luo, Y., Shah, M. M. R., et al. (2018). Temperature-controlled thermophilic bacterial communities in hot springs of western Sichuan, China. *BMC Microbiol.* 18, 134. doi:10.1186/s12866-018-1271-z.

Teplyakov, A., Obmolova, G., Sarikaya, E., Pullalarevu, S., Krajewski, W., Galkin, A., et al. (2004). Crystal structure of the YgfZ protein from Escherichia coli suggests a folate-dependent regulatory role in one-carbon metabolism. *J. Bacteriol.* 186, 7134–7140. doi:10.1128/JB.186.21.7134-7140.2004.

Tomova, I., Stoilova-Disheva, M., Lyutskanova, D., Pascual, J., Petrov, P., and Kambourova, M. (2010). Phylogenetic analysis of the bacterial community in a geothermal spring, Rupi Basin, Bulgaria. *World J. Microbiol. Biotechnol.* 26, 2019–2028. doi:10.1007/s11274-010-0386-7.

Tripathy, S., Padhi, S. K., Mohanty, S., Samanta, M., and Maiti, N. K. (2016). Analysis of the metatranscriptome of microbial communities of an alkaline hot sulfur spring revealed different gene encoding pathway enzymes associated with energy metabolism. *Extremophiles* 20, 525–536. doi:10.1007/s00792-016-0846-6.

Ullah Khan, I., Habib, N., Hussain, F., Xian, W.-D., Amin, A., Zhou, E.-M., et al. (2017). *Thermus caldifontis* sp. nov. a thermophilic bacterium isolated from a hot spring. *Int. J. Sytematic Evol. Microbiol.* 67, 2868–2872. doi:10.1099/ijsem.0.002037.

Van Laere, K. M. J., Abee, T., Schols, H. A., Beldman, G., and Voragen, A. G. J. (2000). Characterization of a novel β-galactosidase from *Bifidobacterium adolescentis* DSM 20083 active towards transgalactooligosaccharides. *Appl. Environ. Microbiol.*

66, 1379–1384. doi:10.1128/AEM.66.4.1379-1384.2000.

Wahlund, T. M., and Madigan, M. T. (1993). Nitrogen fixation by the thermophilic green sulfur bacterium *Chlorobium tepidum*. *J. Bacteriol.* 175, 474–478. doi:10.1128/jb.175.2.474-478.1993.

Ward, L. M., Idei, A., Nakagawa, M., Ueno, Y., Fischer, W. W., and McGlynn, S. E. (2019). Geochemical and metagenomic characterization of Jinata Onsen, a Proterozoic-analog hot spring, reveals novel microbial diversity including iron-tolerant phototrophs and thermophilic lithotrophs. *bioRxiv*, 428698. doi:10.1101/428698.

Watanabe, T., Kojima, H., Umezawa, K., Hori, C., Takasuka, T. E., Kato, Y., et al. (2019). Genomes of neutrophilic sulfur-oxidizing chemolithoautotrophs representing 9 proteobacterial species from 8 genera. *Front. Microbiol.* 10. doi:10.3389/fmicb.2019.00316.

Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., et al. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* 31, 28–33. doi:10.1093/nar/gkg033.

Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E. M., Kyrpides, N., et al. (2012). The M5nr: A novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* 13, 1–5. doi:10.1186/1471-2105-13-141.

Wojciechowska, A., Klewicki, R., Sójka, M., and Grzelak-Błaszczyk, K. (2018). Application of transgalactosylation activity of β-galactosidase from *Kluyveromyces lactis* for the synthesis of ascorbic acid galactoside. *Appl. Biochem. Biotechnol.* 184, 386–400. doi:10.1007/s12010-017-2551-z.

Wood, A. P., and Kelly, D. P. (1988). Isolation and physiological characterisation of Thiobacillus aquaesulis sp. nov., a novel facultatively autotrophic moderate thermophile. *Arch. Microbiol.* 149, 339–343. doi:10.1007/BF00411653.

Yoon, K. S., Ishii, M., Kodama, T., and Igarashi, Y. (1997). Purification and characterization of pyruvate:ferredoxin oxidoreductase from *Hydrogenobacter thermophilus* TK-6. *Arch. Microbiol.* 167, 275–279. doi:10.1007/s002030050443.

Yu, L., Yuan, Y., Chen, S., Zhuang, L., and Zhou, S. (2015). Direct uptake of electrode electrons for autotrophic denitrification by *Thiobacillus denitrificans*. *Electrochem. commun.* 60, 126–130. doi:10.1016/j.elecom.2015.08.025.

Yu, T.-T., Yao, J.-C., Ming, H., Yin, Y.-R., Zhou, E.-M., Liu, M.-J., et al. (2013). *Thermus tengchongensis* sp. nov., isolated from a geothermally heated soil sample in Tengchong, Yunnan, south-west China. *Antonie Van Leeuwenhoek* 103, 513–518. doi:10.1007/s10482-012-9833-9.

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.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate

Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620. doi:10.1093/bioinformatics/btt593.

ANNEX TO CHAPTER 2

Table S1. Taxonomic assignment of the reads at domain level. The relative abundance of sequences assigned to each domain is expressed as a percentage from the total assigned reads

	BW1	BW2	Average	SD
Archaea	4.88	7.49	6.18	1.84
Bacteria	94.43	91.79	93.11	1.86
Eukaryota	0.66	0.67	0.67	0.01
Viruses	0.00	0.05	0.02	0.03

Table S2. Taxonomic assignment of the sequences within Bacteria domain. The relative abundance of sequences assigned to each phyla is expressed as a percentage from the total assigned reads. Those phyla with less than 0.7 % sequences assigned are highlighted in grey

	BW1	BW2	Average	SD
Acidobacteria	0.79	0.96	0.88	0.12
Actinobacteria	1.20	1.36	1.28	0.12
Aquificae	10.43	12.06	11.25	1.15
Bacteroidetes	1.81	2.08	1.95	0.19
Chloroflexi	1.34	1.57	1.46	0.16
Cyanobacteria	1.15	1.30	1.22	0.10
Deinococcus-Thermus	4.54	5.97	5.26	1.01
Firmicutes	3.91	4.67	4.29	0.53
Nitrospirae	0.75	0.74	0.74	0.01
Planctomycetes	0.73	0.75	0.74	0.01
Proteobacteria	70.78	65.71	68.25	3.59
Candidatus Poribacteria	0.02	0.03	0.03	0.00
Chlamydiae	0.05	0.05	0.05	0.01
Chlorobi	0.66	0.69	0.68	0.02
Chrysiogenetes	0.04	0.03	0.04	0.00
Deferribacteres	0.13	0.14	0.13	0.01
Dictyoglomi	0.18	0.21	0.19	0.02
Elusimicrobia	0.02	0.02	0.02	0.00
Fibrobacteres	0.01	0.01	0.01	0.00
Fusobacteria	0.06	0.08	0.07	0.02
Gemmatimonadetes	0.05	0.06	0.05	0.01
Lentisphaerae	0.04	0.05	0.05	0.01
Spirochaetes	0.21	0.25	0.23	0.02
Synergistetes	0.15	0.15	0.15	0.00
Tenericutes	0.02	0.02	0.02	0.00
Thermotogae	0.39	0.47	0.43	0.06
unclassified (derived from Bacteria)	0.15	0.18	0.17	0.02
Verrucomicrobia	0.37	0.38	0.38	0.01

Table S3. Functional profile of As Burgas hot spring at SEED subsystems level 1 expressed as a percentage of reads assigned to each function. Functions with less than 2.11 % of reads assigned are highlighted in grey.

	BW1	BW2	Average	SD
Amino Acids and Derivatives	6.24	6.48	6.36	0.17
Carbohydrates	9.63	9.47	9.55	0.11
Cell Wall and Capsule	3.86	4.22	4.04	0.25
Clustering-based subsystems	13.82	13.05	13.44	0.55
Cofactors, Vitamins, Prosthetic Groups, Pigments	5.95	6.21	6.08	0.19
DNA Metabolism	4.84	4.69	4.76	0.10
Membrane Transport	3.70	3.28	3.49	0.30
Miscellaneous	6.24	6.59	6.42	0.24
Nitrogen Metabolism	2.61	2.37	2.49	0.17
Nucleosides and Nucleotides	3.15	3.14	3.15	0.01
Protein Metabolism	10.65	10.89	10.77	0.17
Respiration	5.82	5.65	5.73	0.12
RNA Metabolism	4.34	4.56	4.45	0.16
Stress Response	2.24	2.53	2.39	0.20
Virulence, Disease and Defense	2.74	2.98	2.86	0.18
Cell Division and Cell Cycle	1.16	0.88	1.02	0.20
Dormancy and Sporulation	0.24	0.22	0.23	0.01
Fatty Acids, Lipids, and Isoprenoids	2.10	1.89	2.00	0.15
Iron acquisition and metabolism	0.45	0.52	0.49	0.05
Metabolism of Aromatic Compounds	1.75	1.71	1.73	0.02
Phages, Prophages, Transposable elements, Plasmids	1.63	1.56	1.59	0.05
Phosphorus Metabolism	1.84	1.52	1.68	0.22
Photosynthesis	1.46	0.06	0.76	0.99
Potassium metabolism	0.01	0.48	0.25	0.33
Regulation and Cell signaling	0.30	1.54	0.92	0.87
Secondary Metabolism	1.55	0.26	0.90	0.91
Sulfur Metabolism	0.24	1.59	0.92	0.95
Motility and Chemotaxis	1.44	1.67	1.55	0.16

Chapter 3

Comparative metagenomic analysis of two hot springs

from Ourense, Northwestern Spain

ABSTRACT

With their circumneutral pH and their moderate temperature (66 °C and 68 °C respectively), As Burgas and Muiño da Veiga are two important human-use hot springs, previously studied with traditional culture methods, but never explored with a metagenomic approach. In the present study, we have performed metagenomic sequence-based analyses to compare the taxonomic composition and functional potential of these hot springs. Proteobacteria, Deinococcus-Thermus, Firmicutes, Nitrospirae, and Aquificae are the dominant phyla in both geothermal springs but there is a significant difference in the abundance of these phyla between As Burgas and Muiño da Veiga. Phylum Proteobacteria dominates As Burgas ecosystem while Aquificae is the most abundant phylum in Muiño da Veiga. Taxonomic and functional analyses reveal that the variability in water geochemistry might be shaping the differences in the microbial communities inhabiting these geothermal springs. The content in organic compounds of As Burgas water promotes the presence of heterotrophic populations of the genus *Thermus*, whereas the sulfate-rich water of Muiño da Veiga favors the co-dominance of genera *Sulfurihydrogenibium* and *Thermodesulfovibrio*. Differences in ammonia concentration determine the abundance of ammonia oxidizing Archaea from the genus *Nitrosopumilus* in As Burgas, and exert a selective pressure towards the growth of nitrogen-fixing bacteria such as *Thermodesulfovibrio* in Muiño da Veiga. Temperature and pH are two important factors shaping hot springs microbial community, as was determined by comparative analysis with other thermal springs.

INTRODUCTION

Metagenomics has revolutionized microbial ecology, overcoming, and complementing the traditional time-consuming culture methods. This approach has widely contributed to reveal the great microbial diversity of ecosystems previously thought to be lifeless and scarcely studied due to their irreproducible environmental conditions, such as hot springs. Nowadays hot springs are considered ideal sites for studying the early life forms, or even extraterrestrial conditions (Konhauser et al., 2003; Pirajno, 2020). Furthermore, metabolically diverse microbial communities have been described in hot

springs with very different physicochemical parameters (Sahay et al., 2017; Saxena et al., 2017) and some of their thermostable enzymes have been characterized as valuable biocatalysts with potential use in biotechnology and industry (Li and Liu, 2017; Ferrandi et al., 2018).

Previous studies have utilized comparative metagenomics to unveil the influence of physicochemical parameters in the diversity of thermophilic microbial communities inhabiting thermal springs (Menzel et al., 2015; Chiriac et al., 2017; Hussein et al., 2017; Sahoo et al., 2017; Mehetre et al., 2018). Among those, temperature and pH have been frequently identified as important factors shaping hot springs microbial populations (Chaudhuri et al., 2017; Mahato et al., 2019; Podar et al., 2020). Numerous investigations have reported a decrease in hot spring biodiversity with increasing temperature (Sharp et al., 2014; Li et al., 2015; Lavrentyeva et al., 2018). Chan et al. (2017) found that species richness and evenness in Malaysian hot springs were negatively correlated with temperature, and thus the composition of the microbial community was determined by this parameter. A significant role of temperature regulating the distribution of hot spring microbial communities was also reported in the Tibetan Plateau Geothermal Belt (China) (Guo et al., 2020) and in Odisha (India) (Badhai et al., 2015). Contrariwise, Power et al. (2018) studied 925 geothermal springs across New Zealand and determined that pH was the main factor influencing hot springs diversity at temperatures below 70 °C, while temperature had a significant effect on the microbial distribution at those hot springs with water temperature above 70 °C.

From all the thermal spots found across Ourense (Galicia, Northwestern Spain), As Burgas and Muiño da Veiga hot springs are located just scarce 5 kilometers away but their waters show very different chemical and mineral composition due to the high geologic variability of the region (González-Barreiro et al., 2009; López et al., 2019). The diatom communities and the lipolytic enzyme-producing thermophiles inhabiting As Burgas, have been previously investigated by traditional methods (Deive et al., 2013; Leira et al., 2017), nevertheless, the sequencing of the whole environmental DNA, known as shotgun metagenomics, has never been applied to study of As Burgas or Muiño da Veiga microbiome.

In this work, we have used metagenomics in conjunction with statistical tools to compare the microorganisms and community structure of As Burgas and Muiño da Veiga. Moreover, in an attempt to ascertain whether environmental conditions such as pH or temperature determine the microbial diversity and function of these nearby ecosystems, we have analyzed the functional and taxonomical profile of these geothermal springs and other geographically distant hot springs that encompass different pH and temperatures.

MATERIALS AND METHODS

Sample collection

As Burgas (BW) water was sampled as described in chapter 2. Muiño da Veiga (MDV) water, with temperature 68 °C and pH 7, was collected from Muiño da Veiga hot spring (GPS 42.352397, -7.909714), in Ourense (Galicia, Spain) following the same procedure. 50 L of water were collected into water bottles, which were prewashed with 70 % ethanol. The sample was stored at room temperature until the next day when it was filtered through a nitrocellulose filter of 0.2 µm. Filters were preserved at -20°C until metagenomic DNA extraction.

DNA extraction and sequencing

Total DNA from BW was isolated, quantified, and sequenced following the procedure detailed in chapter 2. Similarly, metagenomic DNA from MDV was isolated from the filters by Juan José Escuder, using the Metagenomic DNA Isolation Kit for Water (Epicentre Biotechnologies), according to the manufacturer's protocol. MDV metagenomic DNA was sequenced using the Illumina Hi-seq 1500 platform with 2 x 100 base read length by the sequencing services of Health In Code (A Coruña, Spain).

Taxonomic distribution and statistical analysis

Illumina reads were treated with PRINSEQ software (Schmieder et al., 2011) for quality control, removing all artificial duplicate reads and reads shorter than 60 base-pairs. High-quality unassembled reads of both samples were uploaded into the Metagenomics Rapid Annotation using the Subsystem Technology (MG- RAST) v4.0.3

server (Meyer et al., 2008). Only BW metagenome is publicly available under the accession number mgm4709018.3, since MDV metagenome is uploaded as private in the MG-RAST until its publication. MG-RAST server was used to assign the taxonomic profile of the metagenomic reads, with a maximum e-value of $1e^{-05}$, a minimum identity of 60 %, and a minimum alignment length of 15 bp. The statistical analysis of the different taxonomic levels generated on MG-RAST was performed using the Statistical Analyses of Metagenomic Profiles (STAMP) (Parks et al., 2014) software. The significance of differences between proportions in the taxonomic distribution of BW and MDV samples was performed using the two-sided Fishers exact test, with Newcombe–Wilson confidence interval method. Because P-values were not uniformly distributed, Benjamin–Hochberg false discovery rate was applied for correction. Results with q < 0.05 were considered significant. A difference of at least 1 % and a twofold ratio between proportions was applied.

Functional analyses

Functional analyses were performed using the SEED subsystems annotation in the MG-RAST, with a maximum e-value of $1e^{-05}$, a minimum identity of 60 %, and a minimum alignment length of 15 bp. The functional profiles generated by MG-RAST were statistically analyzed in the STAMP, with the same procedure and parameters described previously for the taxonomic analyses.

Metagenome sample selection for comparative metagenomics

Apart from As Burgas hot spring water metagenome whose taxonomic and functional profile was studied in the previous chapter, other 6 hot spring metagenomes were selected for the comparative analysis. To reduce the "type of sample" influences, this study included only hot spring water samples. All of the samples selected are listed in table 1 and publicly available on MG-RAST, excepting MDV. This metagenomic sample is part of an unpublished study from our group and was selected for its proximity to As Burgas hot spring, as both of them are located only a few kilometers away in the same Galician region (Ourense, Spain).

Taxonomic and functional profiles of the different samples were extracted from the MG-RAST. All of them are hot spring water samples that contain unassembled raw metagenomic reads in order to obtain information about the abundance of the sequences and, therefore, to be able to properly compare between samples. For each metagenome, sequence counts on MG-RAST were standardized against the total number of hits to remove bias in different sequencing efforts and library size, as described in Chapter 2.

Table 1. Characteristics of the metagenomic data selected for this study

Sample id	MG-RAST id	Hot spring	Location	pH	Temperature (°C)	Size (Mbp)	Sequencing method	Reference
BW	mgm4709018.3	As Burgas	Ourense, Spain	7.6	66	235.68	Illumina MiSeq	This study
MDV	-	Muiño da Veiga	Ourense, Spain	7.0	68	10,408.06	Illumina Hi-seq 1500	Unpublished
Coamo	mgm4726046.3	Coamo	Coamo, Puerto Rico	8.2	47	33.10	Illumina MiSeq	Padilla-Del Valle et al. 2017
AT-4	mgm4555635.3	Atri		7.4	58	22.6		
HT-1	mgm4555636.3	Athamallik	Odisha, India	7.4	56	11.96	Roche 454 GS-FLX	Badhai et al. 2015
TB-3	mgm4555637.3	Tarabalo		7.3	57	22.79		
TP-2	mgm4555638.3	Taptapani		7.2	42	13.81		
Coquito	mgm4449206.3	El Coquito	Los Nevados, Colombia	2.7	29	53	Roche 454 GS-FLX Titanium	Jiménez et al. 2012

RESULTS AND DISCUSSION

DNA extraction and sequencing

After the quality control, 893,557 and 27,113,937 sequences with a total of 253,083,221 and 3,968,584,153 bp, an average length of 283 ± 71 and 146 ± 24 bp, and a GC content of 54 ± 12 % and 44 ± 13 % were uploaded to MG-RAST for BW and MDV respectively. With this pipeline, 368,188 proteins were predicted for BW sample and 194,410 were identified as protein features. For MDV 6,422,118 proteins were predicted and 2,985,268 were identified as protein features.

Comparative analysis of microbial diversity among the hot springs

At domain level, the taxonomical profile generated by MG-RAST was similar in the BW and MDV samples. Highest representation of Bacterial sequences was found in both metagenomes (94.42 % and 97.43 %), followed by Archaea (4.88 % and 1.96 %),

Eukaryota (0.66 % and 0.54 %), and Viruses (0.03 % and 0.06 %) (Fig 1). Proteobacteria (70.78 % and 27.81 %), Deinococcus-Thermus (4.54 % and 6.55 %), Firmicutes (3.91 % and 7.89 %), Nitrospirae (0.75 % and 9.22 %), and Aquificae (10.46 % and 36.54 %) were the more abundant phyla in the two samples (Fig 2). However, significant differences were found between both metagenomes, as Proteobacteria were predominant in BW while Aquificae, Firmicutes, and Nitrospirae were significantly more abundant in MDV. Classes Betaproteobacteria (42.8 % and 9.88 %), and Gammaproteobacteria (10.43 % and 4.94 %) were overrepresented in BW, in contrast with the prevalence of Aquificae (11.07 % and 35.60 %), Nitrospira (0.68 % and 8.99 %) and Deltaproteobacteria (3.43 % and 6.84 %) in MDV (Fig 3).
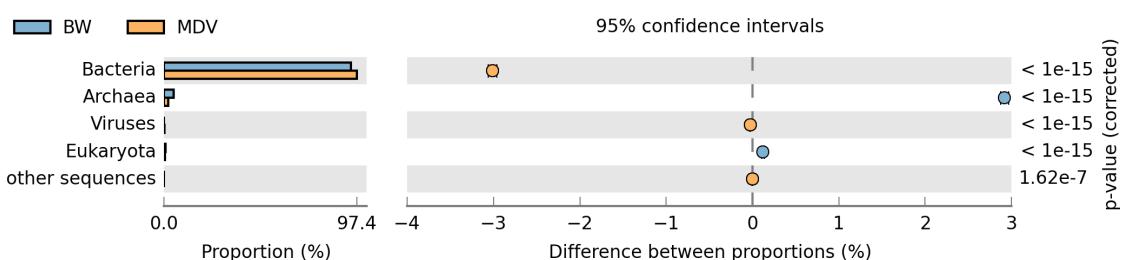

Figure 1. Comparative taxonomic profile of BW and MDV at domain level.

Proteobacteria, Thermus, Aquificae, and Firmicutes were also among the most abundant phyla found in the water of the Lobios hot spring (76 °C, pH= 8.2) located nearby in Ourense (Spain) (López-López et al., 2015), and are frequently found in hot springs (Huang et al., 2013; Chan et al., 2015; Paul et al., 2016).

The high relative abundance of phyla Proteobacteria and Firmicutes in the studied metagenomes is consistent with the results reported for other neutral springs, with different combinations of extreme environmental conditions, where these phyla were abundantly found (Filippidou et al., 2016; Chan et al., 2017). Nevertheless, the higher relative abundance of Firmicutes in MDV sample could be related to the temperature, similar to the results found in Bakreshwar (India) in which the hot spring with higher temperature showed more abundance of this phylum (Chaudhuri et al., 2017). A positive correlation between the abundance of phylum Firmicutes and temperature was also found in the hot springs of the Tibetan Plateau (Zhang et al., 2018).
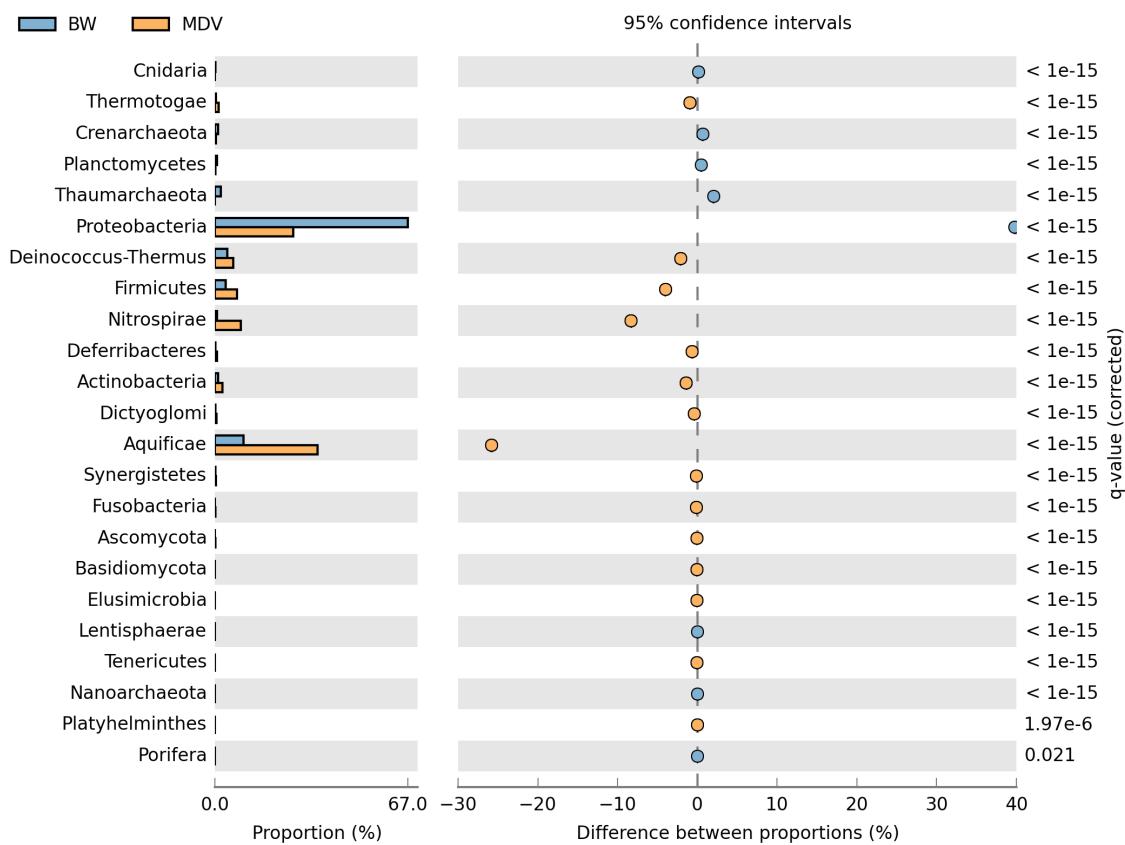
Figure 2. Comparative taxonomic profile of BW and MDV at phylum level. Only phylum with significant biological differences (P<0.05, difference between proportions >1 % and twofold of ratio between proportions, STAMP) were represented.

Compared to BW, MDV water showed a higher abundance of phylum Nitrospirae (0.75 % and 9.2 %). Phylum Nitrospirae is composed mainly of aerobic chemolithotrophs, including microorganisms able to perform nitrification and sulfate-reducing bacteria (Garrity et al., 2001). Within phylum Nitrospirae, most species of the genus *Thermodesulfovibrio* have been isolated from thermal springs (Henry et al., 1994; Sonne-Hansen and Ahring, 1999; Haouari et al., 2008) and are characterized as obligately anaerobic thermophilic bacteria able to reduce sulfate and other sulfur compounds (Matsuura et al., 2016). *Thermodesulfovibrio* is the second most abundant genus in MDV while it is not abundant in BW (Table 2). This finding could be related to the relatively higher sulfate concentration of this hot spring compared to BW (López et al., 2019). *Thermodesulfovibrio* was also abundantly found in several hot springs from Odisha (India), especially in AT-4 hot spring (Badhai et al., 2015) and in Borong hot spring (Najar et al., 2018).
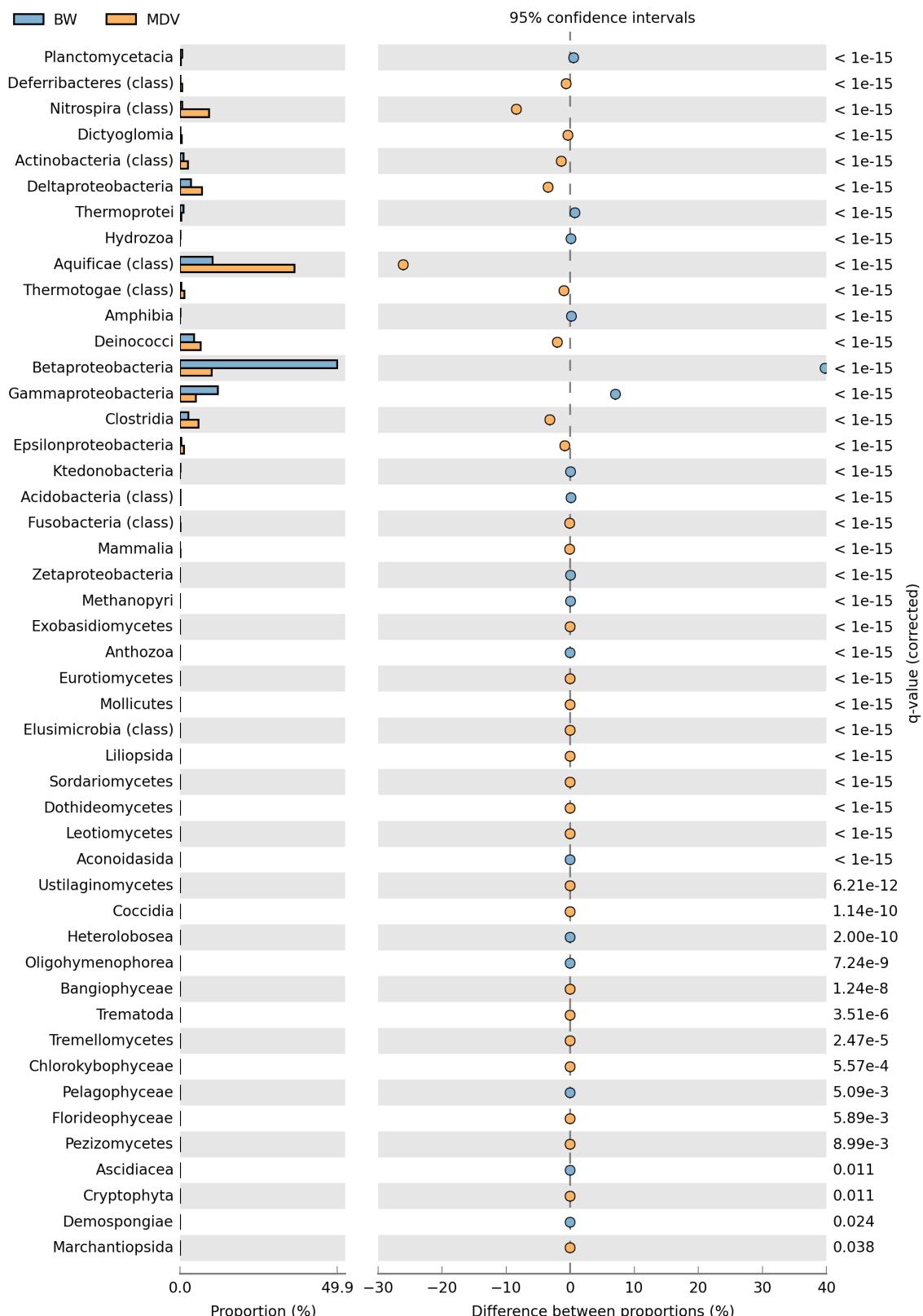
Figure 3. Comparative taxonomic profile of BW and MDV at class level. Only classes with significant biological differences (P<0.05, difference between proportions >1 % and twofold of ratio between proportions, STAMP) were represented.

Table 2. List of the 14 most abundant genera in BW and MDV metagenomes. Genera abundantly found in both metagenomes are highlighted in bold. Percentages were generated by MG-RAST using data from the M5NR database.

BW		MDV	
	% sequences		% sequences
Thermus	15.77	*Sulfurihydrogenibium*	30.19
Hydrogenobacter	8.56	*Thermodesulfovibrio*	20.49
Thiobacillus	4.20	*Thermus*	9.66
Thermocrinis	2.42	*Hydrogenobacter*	6.81
Nitrosopumilus	2.31	*Meiothermus*	2.96
Acidithiobacillus	2.25	*Thiomonas*	2.23
Burkholderia	2.25	*Thermocrinis*	1.82
Sulfurihydrogenibium	2.24	*Fervidobacterium*	1.15
Unclassified (derived from Thaumarchaeota)	1.74	*Geobacter*	1.01
Thiomonas	1.72	*Anaerolinea*	0.86
Cenarchaeum	1.71	*Calditerrivibrio*	0.78
Acidovorax	1.43	*Dictyoglomus*	0.74
Thermodesulfovibrio	1.40	*Hydrogenivirga*	0.67
Pseudomonas	1.39	*Caldicellulosiruptor*	0.59

With a 30.19 % of sequences assigned, *Sulfurihydrogenibium* (phylum Aquificae) is the predominant genus in MDV (Table 2). Bacteria belonging to this genus are neutrophilic and thermophilic, microaerobic or facultatively anaerobic, chemolithoautotrophic, or facultatively heterotrophic microorganisms (O'Neill et al., 2008). These sulfur-oxidizing bacteria can use sulfur and thiosulfate as electron donors and molecular oxygen as the electron acceptor. *Sulfurihydrogenibium* is also among the 14 most abundant genera found in BW and this finding is consistent with previous reports, as this genus has been frequently found to be the most prevalent in circumneutral hot springs with temperatures below 75 °C (O'Neill et al., 2008). For example, in Kamchatka hot springs it was found that lithoautotrophic bacteria from the genus *Sulfurihydrogenibium* were the most abundant in those springs with near-neutral pH (Merkel et al., 2017). Similar results were found in the analysis of 6 geothermal springs from Yellowstone National Park, in which *Sulfurihydrogenibium* sp. dominated in neutral sulfide-rich areas (Takacs-Vesbach et al., 2013).

The co-dominance of genera *Sulfurihydrogenibium* and *Thermodesulfovibrio* in MDV reveals the importance of sulfur metabolism in this hot spring and suggests the existence of a sulfur cycle in MDV geothermal spring between the two dominant genera, in which *Thermodesulfovibrio* would be reducing sulfate to sulfide through anaerobic sulfate respiration, sulfide is then abiotically oxidized to thiosulfate in the

presence of oxygen (Chen and Morris, 1972) and thiosulfate is used by *Sulfurihydrogenibium* as electron donor, producing sulfate.

On the contrary, dominance in BW is more distributed among different genera (Table 2). *Thermus* (phylum Deinococcus-Thermus) and *Hydrogenobacter* (phylum Aquificae) are the most abundant genera in this ecosystem. *Thermus* species are generally thermophilic heterotrophs able to grow using different organic sources, while several can grow mixotrophically with inorganic electron donors (thiosulfate, elemental sulfur) for respiration (Skirnisdottir et al., 2001; Bjornsdottir et al., 2009; Murugapiran et al., 2013). Although most of them are aerobic bacteria, some members of this genus are facultative anaerobic, using NO^{3-}, Fe^{3+}, or S^0 as terminal electron acceptors (Alvarez et al., 2014). Its respiratory flexibility suggests an important role of the genus *Thermus* in sulfur, metal, and nitrogen biochemical cycles in terrestrial geothermal springs (Zhou et al., 2020). With a 9.66 %, *Thermus* sp. is also abundant in MDV water. This finding agrees with the features described for this genus, with an optimum pH of 7.0 − 8.5 and optimum temperature of 65 − 75 °C (Massello et al., 2020) thus, *Thermus* sp. has been commonly found to be abundant in circumneutral thermal springs with temperatures between 60 − 80 °C worldwide (Chan et al., 2017; Power et al., 2018; Kaushal et al., 2018).

Hydrogenobacter sp., mainly constituted by chemolithoautotrophic hydrogen-oxidizing bacteria that use the reductive tricarboxylic acid cycle to fix CO_2 (Aoshima et al., 2004), is the second most abundant genus in BW and the third most abundant genus in MDV (Table 2). These results point to the members of genera *Hydrogenobacter* and *Sulfurihydrogenibium* as the main responsible for carbon fixation in both ecosystems. The abundance of *Hydrogenobacter* sp. in BW and MDV could be attributed to pH, as this genus has been described as neutrophilic (Bonjour and Aragno, 1986) and has been predominantly found in neutral geothermal springs in which genus *Thermus* is frequently present (Massello et al., 2020). As described in Chapter 2, abundance and co-occurrence of *Thermus* and *Hydrogenobacter* was previously reported for other hot springs, suggesting a possible metabolic association between both genera.

Among the phylum Proteobacteria, members of the genus *Thiobacillus* are characterized by their diversity of metabolism, with a predominance of chemolitoautotrophy (Lavrentyeva et al., 2018). This genus includes sulfur-oxidizing species able to use elemental sulfur, sulfide, and thiosulfate as electron donors, such as *Thiobacillus ferrooxidans* and *T. thiooxidans*. Some species can use NO_3^- as terminal electron acceptor like *Thiobacillus denitrificans* (Kumar et al., 2018), moreover, *T. ferrooxidans* has been described as potential nitrogen-fixing bacteria (Mackintosh, 1978; Rawlings and Kusano, 1994). Genus *Thiobacillus* is present with a 4.30 % in BW, while it is not among the most abundant genera found in MDV. This genus was also detected in high abundance in Yumthang hot springs (India) (Panda et al., 2016), Shi-Huang-Ping acidic hot spring (Taiwan) (Lin et al., 2015), and Tsenkher Spring (Lavrentyeva et al., 2018).

Ammonia-oxidizing archaea from the genus *Nitrosopumilus* (phylum Thaumarchaeota) were abundantly found in BW, while this is not among the 14 most abundant genera in MDV (Table 2). This finding might be related to the ammonia concentration of both springs, as this is considered to be an important factor influencing the potential activity of ammonia oxidizers and their community structure (Li et al., 2011). Several studies have shown that Archaea rather than Bacteria are the main microorganisms driving ammonia oxidation in high-temperature hot springs environments (Chen et al., 2016). In As Burgas hot spring, the ammonia concentration of 0.94 mg L^{-1} is higher compared to Muiño da Veiga hot spring (0.58 mg L^{-1}) (Delgado-Outeiriño et al., 2009) and these differences may be responsible for the abundance of genus *Nitrosopumilus* in BW and its absence in MDV.

When compared to the other studied hot springs, at domain level, there is a predominance of Bacterial sequences followed by Archaea in the eight samples, excepting El Coquito and Coamo, in which eukaryotic sequences and viruses are more abundant than archaeal sequences, respectively (Fig 4). The high abundance of Eukaryota in El Coquito metagenome might be related to its relatively low temperature (29 °C) compared to the others. At this temperature, eukaryotic micro-algae can proliferate (Varshney et al., 2015) as important primary producers of the ecosystem taking advantage of the solar energy at the surface.

The high abundance of bacterial sequences in all the studied metagenomes, with temperatures ranging from 29 to 68 °C, is comparable to other previously analyzed hot springs such as Lobios in Ourense (López-López et al., 2015), Comano in Italy (Pedron et al., 2019), Ma'in and Afra hot springs in Jordania (Hussein et al., 2017) or the hot springs from Bakreshwar (India) (Chaudhuri et al., 2017).
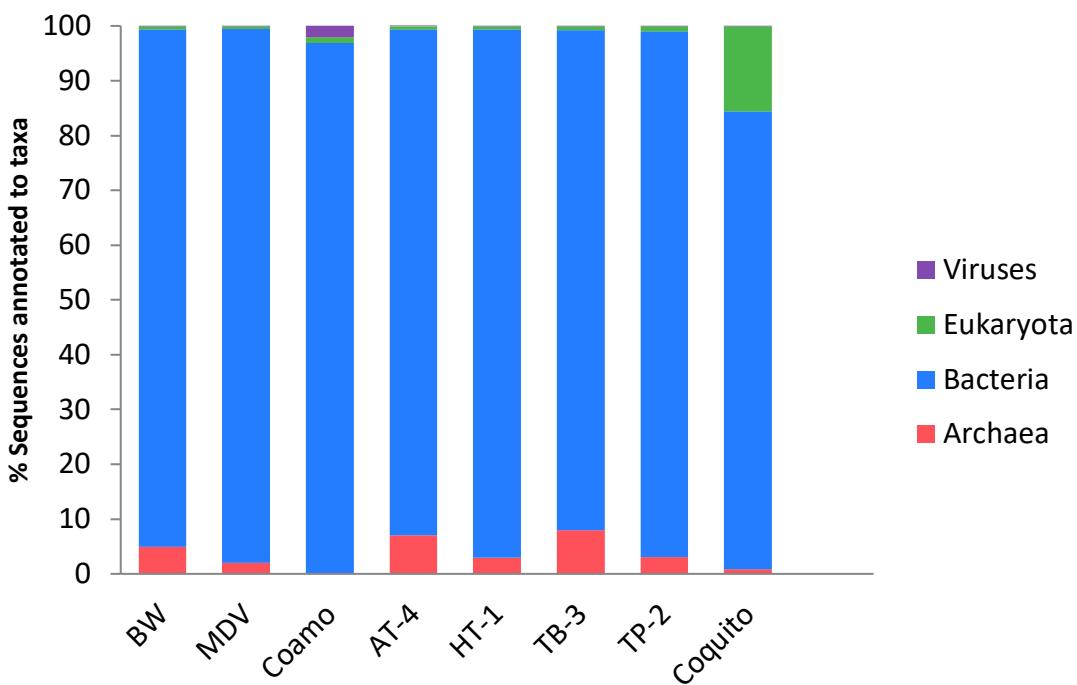

Figure 4. Comparative microbial diversity at different hot springs at domain level. Chart was generated using microbial abundance data. Each chart represents the percentage of abundance of each microbial group in a specific spring. Abundance values are generated from normalized data retrieved from MG-RAST.

Focusing on the Bacteria domain, at phylum level, there is a clear predominance of Proteobacteria in BW, Coamo, HT-1, TB-3, TP-2, and Coquito metagenomes (Fig 5). The 95.03 % of Proteobacterial sequences in Coamo hot spring metagenome is the highest of the eight metagenomes and could be a product of the library construction, as in the pCC1FOS system utilized by Padilla-Del Valle et al. (2017) *Escherichia coli* Epi300-T1[R] was used as host, and the taxonomic assignment in MG-RAST shows that the more abundant sequences in Coamo thermal spring at genus level are annotated as *Escherichia* (26 %), a genus that is not frequently found as predominant in this kind of thermal environments. Moreover, although the study mentions the removal of vector pCC1FOS sequences, there is no evidence in the material and methods section of the extraction of *E.coli* host sequences, which is an important step, as was reflected in

other taxonomical studies that generated their sequences from a metagenomic library constructed in pCC1FOS system (López-López et al., 2015; Rabelo-Fernandez et al., 2018; Soriano et al., 2018). On the contrary, this result might be a real reflection of Coamo microbial population, since this is a moderate temperature hot spring (47 °C) and it has been demonstrated that several *E.coli* strains are thermotolerant bacteria that can evolve to grow at higher temperatures (Rudolph et al., 2010). Furthermore, *E.coli* has been detected in a thermal spring in Unkeswar (India) (Rekadwad et al. 2016) and in five hot springs of Eritrea (Ghilamicael et al., 2018). Additionally, in Coamo thermal spring the genus *Microvirus*, constituted by single-stranded DNA viruses that infect Enterobacteria such as *Escherichia* (Krupovic and Forterre, 2011), is also abundantly found after taxonomic assignment by MG-RAST.
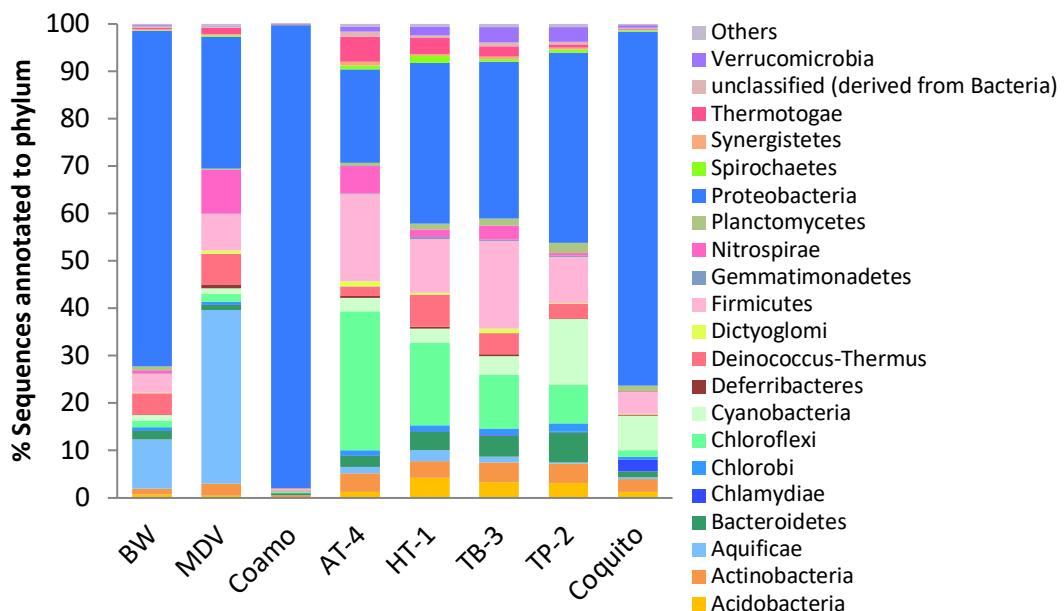

Figure 5. Comparative microbial diversity within Bacteria domain at phylum level. Others include those phyla with less than 0.3 % sequences annotated in the 8 metagenomes: Candidatus Poribacteria, Chrysiogenetes, Elusimicrobia, Fibrobacteres, Lentisphaerae, Tenericutes, and Fusobacteria.

The high abundance of proteobacterial sequences in 6 of the 8 studied metagenomes might be related to temperature, according to previous studies that have reported the predominance of phylum Proteobacteria in geographically distant moderate-temperature thermal springs such as Siloam in South Africa (Tekere et al., 2011), Attri and Yumthang in India (Panda et al., 2016; Tripathy et al., 2016), Ayer Hangat, Sungai Serai and Dusun Tua in Malaysia (Chan et al., 2017) or Tatta Pani in Pakistan (Amin et al., 2017). Guo et al. (2020) suggested temperature as the main factor shaping the

microbial community of 16 hot springs from the Tibetan Plateau as they found that hot springs with lower temperature (< 45 °C) showed a dominance of Proteobacteria and Nitrospirae, while in moderate (55 – 70 °C) to high-temperature (> 70 °C) geothermal springs Aquificae, Deinococcus-Thermus, Thermodesulfobacteria, Thermotogae, and Cyanobacteria were the most abundant phyla. In agreement with Guo et al. (2020), from the 8 hot springs studied here, proteobacterial sequences were the most abundant in those with lower temperatures (Coquito, Coamo, and TP-2) but, on the contrary, phylum Nitrospirae was not abundant. Moreover, the dominant phyla described for the moderate to high-temperature Tibetan Plateau hot springs are relatively more abundant in the moderate to high-temperature hot springs studied here, but Proteobacteria phylum is also abundant.

On the other hand, in all the Indian samples (AT-4, TP-2, HB-1, TB-3) Firmicutes is a very abundant phylum with a percentage ranging from 9.7 to 18.6 %, as reflected in figure 5. Based on its abundance, Firmicutes is considered a signature phylum for circumneutral hot springs in several studies (Chan et al., 2017). Nevertheless, high abundance of Firmicutes has been reported in other Indian hot springs with alkaline pH and temperatures between 42 to 65 °C such as Tuwa, Lasundra, Tulsi Shyam, and Bakreshwar (Ghelani et al., 2015; Mangrola et al., 2015a, 2015b; Chaudhuri et al., 2017), suggesting that Firmicutes abundance in these samples might be due to other parameters related with the geographical proximity. Also, in all the Indian springs studied here, there is a clear abundance of Chloroflexi when compared with the rest of samples. Bacteria from phylum Chloroflexi show a great variation in their metabolisms, with autotrophic, heterotrophic, and mixotrophic members (Bennett et al., 2020). The presence of Chloroflexi in the 4 Indian hot springs investigated by Badhai et al. in 2015 and selected for this study, is clearly correlated with temperature and with the distribution of phylum Cyanobacteria, as the abundance of phylum Chloroflexi increases with temperature while the abundance of cyanobacterial sequences decreases (Fig 5). Similarly, Wang et al. (2013) found that in Tibetan hot springs with moderate temperatures, abundances of phyla Cyanobacteria and Chloroflexi were negatively correlated. Another recent study also reported a decrease of the

phototrophic Cyanobacteria with increasing temperature and maintenance of phototrophic Chloroflexi populations (Bennett et al., 2020).

Coquito and TP-2 are the hot springs with higher representation of Cyanobacteria in their microbial communities with 7.3 % and 13.1 % of cyanobacterial sequences, respectively. This finding agrees with their lower temperature (29 °C in Coquito and 42 °C in TP-2) that favors the survival of bacteria belonging to this phylum. Other studies have reported higher abundance of Cyanobacteria in lower temperature hot springs (Sompong et al., 2005; Wang et al., 2013a; Chan et al., 2017; Singh et al., 2018).

The presence of phylum Aquificae is higher in the two thermal springs from Ourense (Spain) with abundances of 10.43 % in BW and 36.46 % in MDV, in which it is the dominant phylum. This result might be related to their higher temperature, as some studies reported that the presence of Aquificae is positively correlated with temperature (Wang et al., 2013). Guo et al. (2020) also described higher abundance of phylum Aquificae in moderate to high-temperature hot springs of the Tibetan Plateau than in those with lower temperatures. Other thermal springs with dominance of phylum Aquificae are Malangto (75.8 °C, pH 5.08) and Balasbas (60.5 °C, pH 5.20) in Philippines (Huang et al., 2013).

Actinobacteria are present in all the studied metagenomes with percentages between 4.2 % (TB-3) and 0.5 % (Coamo). This phylum, first considered as characteristic from soil, nowadays has been reported as ubiquitous in hot springs as it has been found in several hot springs with very diverse pH and temperatures (Jiang et al., 2012; Valverde et al., 2012; Liu et al., 2016).

Phylum Deinococcus-Thermus is present in all the samples excepting Coamo and Coquito metagenomes in which this phylum is almost absent. The absence of this phylum in Coquito and Coamo hot springs is related to their low temperatures, as all the members of the order Thermales require temperatures higher than 60 °C to grow (Banerjee et al., 2014).

Comparative analysis of functional features among the hot springs

The functional profile generated with MG-RAST showed significant differences in proportion for subsystems at level 1 between BW and MDV metagenomes (considering $P < 0.05$ and differences of at least 1 % and a twofold ratio between proportions) (Fig 6). Nevertheless, the clustering-based subsystems (11.68 % and 12.48 %), protein metabolism (11.29 % and 11.12 %), and amino acid and derivatives (9.12 % and 10.06 %) subsystems were the most abundant functional categories at level 1 in BW and MDV. The clustering-based subsystems groups protein families with functional coupling evidence but unknown function, and was also the most abundant subsystem found in the water of the Lobios hot spring in Ourense (López-López et al., 2015).

Interestingly, sulfur and nitrogen metabolism subsystems are overrepresented in BW. Focusing on these subsystems, there is a clear difference in the percentage of sequences assigned to each functional category at level 3 between both metagenomes (Table 3). In the sulfur metabolism, there is a predominance of sulfur oxidation affiliated sequences in BW with respect to MDV (Table 3). The relative abundance of sequences annotated as sulfur oxidation might be related to the richness of *Thermus*, *Thiobacillus*, *Acidithiobacillus*, *Thiomonas,* and *Sulfurihydrogenibium* (Table 2) in BW. On the contrary, the abundance of the sulfur-reducing genus *Thermodesulfovibrio* in MDV could be correlated with the higher proportion of sequences annotated as sulfate reduction-associated complexes in MDV when compared to BW.

The abundance of sequences related to nitrate and nitrite ammonification in BW, also known as dissimilatory nitrate reduction to ammonium (DNRA), could be associated with the predominance of genera *Thermus* and *Thiobacillus* in this metagenome, which can use NO^{3-} as the final electron acceptor, producing ammonia. Therefore, these genera might also be responsible for the relatively higher concentration of ammonia in BW when compared to MDV, stimulating the occurrence of ammonia-oxidizers, such as those from the genus *Nitrosopumilus*, in the ecosystem.
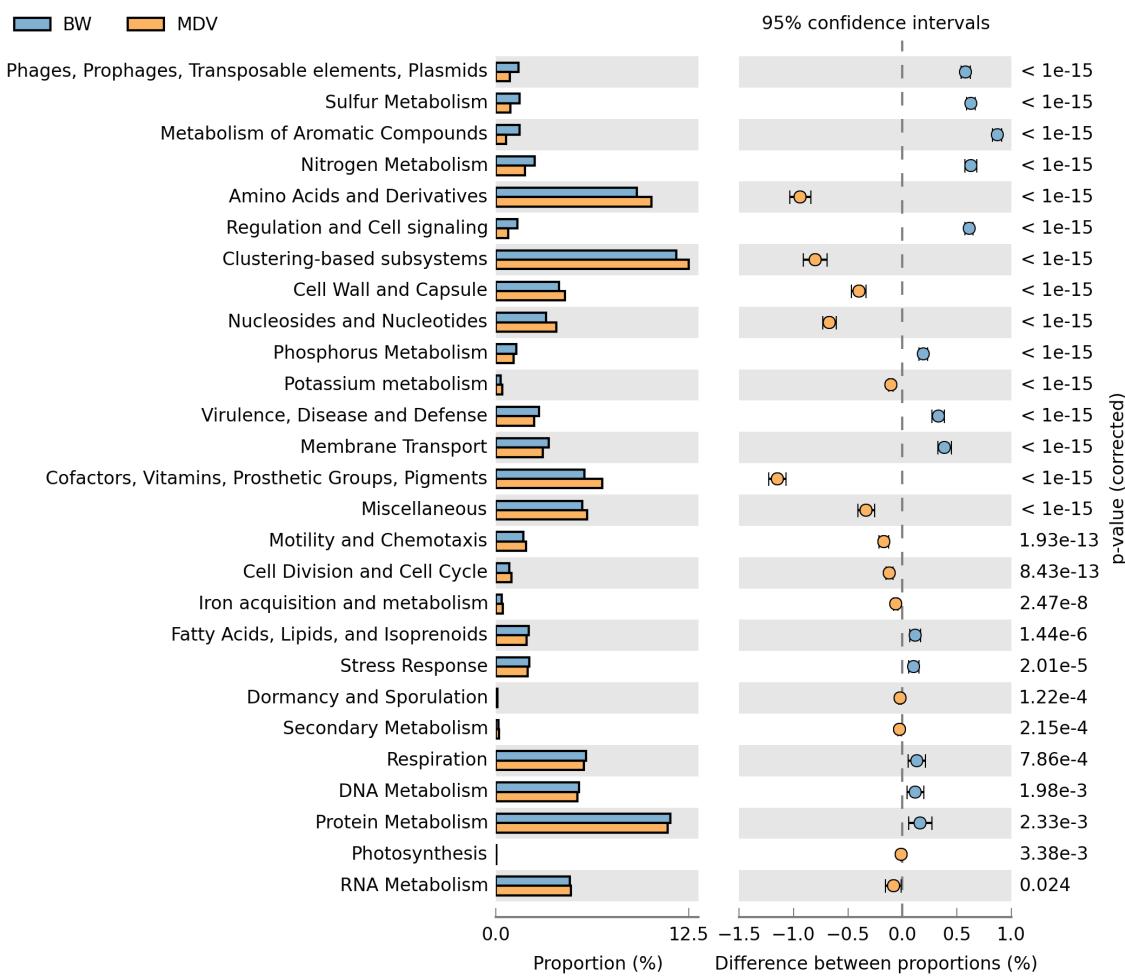
Figure 6. Comparative functional profile of BW and MDV at level 1 of the SEED subsystems. Only subsystems with significant biological differences (P<0,05, difference between proportions >1% and twofold ratio between proportions, STAMP) were represented.

The percentages of sequences assigned to denitrification, nitrogen fixation, and nitrosative stress functions are relatively higher in MDV than in BW (Table 3). Members of genus *Thermodesulfovibrio*, the second most abundant in MDV (Table 2), might be related to the higher proportion of sequences involved in nitrogen fixation. Recent studies have suggested an association between nitrogen fixation and sulfate reduction in hot springs, proposing chemolithoautotrophic sulfate-reducing bacteria as the main responsible for nitrogen fixation (Nishihara et al., 2018b). Moreover, all the genes necessary for nitrogen fixation have been found in several *Thermodesulfovibrio* species (Frank et al., 2016) and members of this genus have been pointed as possible nitrogen-fixing bacteria in Nakabusa hot springs in Japan (Nishihara et al., 2018a) and Mushroom spring in Yellowstone National Park (Thiel et al., 2017), although diazotrophic growth has not been demonstrated yet in the laboratory for any

Thermodesulfovibrio species. In addition, the relatively low ammonia concentration of MDV (Delgado-Outeiriño et al., 2009) might be promoting a selective pressure for the growth of diazotrophic bacteria able to fix N_2, as has been generally reported in geothermal springs (Hamilton et al., 2014).

Table 3. Percentages of sequences assigned to each functional category in Sulfur Metabolism, Nitrogen metabolism, and CO_2 fixation at level 3 in BW and MDV metagenomes.

		% sequences	
		BW	MDV
Sulfur Metabolism	Inorganic Sulfur Assimilation	0.374	0.212
	Galactosylceramide and Sulfatide metabolism	0.002	0.009
	Release of Dimethyl Sulfide from Dimethylsulfoniopropionate	0	8.03E-05
	Sulfate reduction-associated complexes	0.156	0.164
	Sulfur oxidation	0.731	0.364
	Thioredoxin-disulfide reductase	0.100	0.112
	Alkanesulfonate assimilation	0.023	0.033
	Alkanesulfonates Utilization	0.002	0.004
	DMSP breakdown	2.742e-04	2.892e-04
	L-Cystine Uptake and Metabolism	0.003	0.007
	Taurine Utilization	0.003	0.004
	Utilization of_glutathione as a sulphur source	0.018	0.016
Nitrogen Metabolism	Allantoin Utilization	0.013	0.013
	Amidase clustered with urea and nitrile hydratase functions	0	2.580E-04
	Ammonia assimilation	0.625	0.586
	Cyanate hydrolysis	0.012	0.011
	Denitrification	0.190	0.218
	Dissimilatory nitrite reductase	0.153	0.132
	Nitrate and nitrite ammonification	1.144	0.578
	Nitric oxide synthase	0.011	0.019
	Nitrilase	0	6.427E-05
	Nirogen fixation	0.164	0.271
	Nitrosative stress	0.009	0.071
CO_2 fixation	TCA cycle	0.487	0.311
	Calvin-Benson cycle	0.645	0.398
	Carboxysome	0.065	0.049
	CO_2 uptake, carboxysome	0.295	0.163
	Photorespiration (oxidative C2 cycle)	0.266	0.297

The elevated percentage of sequences involved in denitrification annotated in MDV could be associated with the dominance of genus *Sulfurihydrogenibium*, since the ability to denitrify completely to N_2 has been described in several *Sulfurihydrogenibium*

species (Takai et al., 2003). Diverse members of *Thermus* and *Hydrogenobacter* genera can also be performing denitrification in both hot springs, apart from DNRA (Dodsworth et al., 2011).

These results suggest that the two main genera in MDV (*Sulfurihydrogenibium* and *Thermodesulfovibrio*) are playing an important role, not only in the sulfur cycle, but also in the nitrogen cycle. Furthermore, *Sulfurihydrogenibium* can grow autotrophically using elemental sulfur as an electron donor and nitrate as a final electron acceptor, producing sulfate, liberating N_2, and coupling both sulfur and nitrogen cycles.

Focusing on the carbon fixation, an abundance of sequences affiliated to the Calvin-Benson Cycle is found in BW when compared to MDV. This might be related to the significantly higher proportion of Betaproteobacteria and Gammaproteobacteria (Fig 3) in BW hot spring, as the autotrophic members of these taxonomic classes use the reductive pentose phosphate (Calvin-Benson) cycle to fix carbon (Hügler and Sievert, 2011). Sequences annotated within the tricarboxylic acid (TCA) cycle can be associated with the catabolism but also to the carbon fixation via the reductive TCA cycle performed by representatives of the genera *Hydrogenobacter*, *Thermocrinis,* and *Sulfurihydrogenibium*, among others (Hügler et al., 2007).

In relation to the functional profile of the studied metagenomes, the relative abundance of the 23 functions assigned by subsystems at level 1 is very similar in all the metagenomes with the exception of Coamo hot spring in which a higher proportion of sequences annotated as phages, prophages, transposable elements, plasmids was detected (Fig 6). This result was expectable due to the abundance of viral sequences (Fig 4) reported in Coamo metagenome compared to the other hot springs. On the contrary, sequences related to protein metabolism and cofactors, vitamins, prosthetic groups, pigments were significantly lower in Coamo than in the rest of metagenomes.
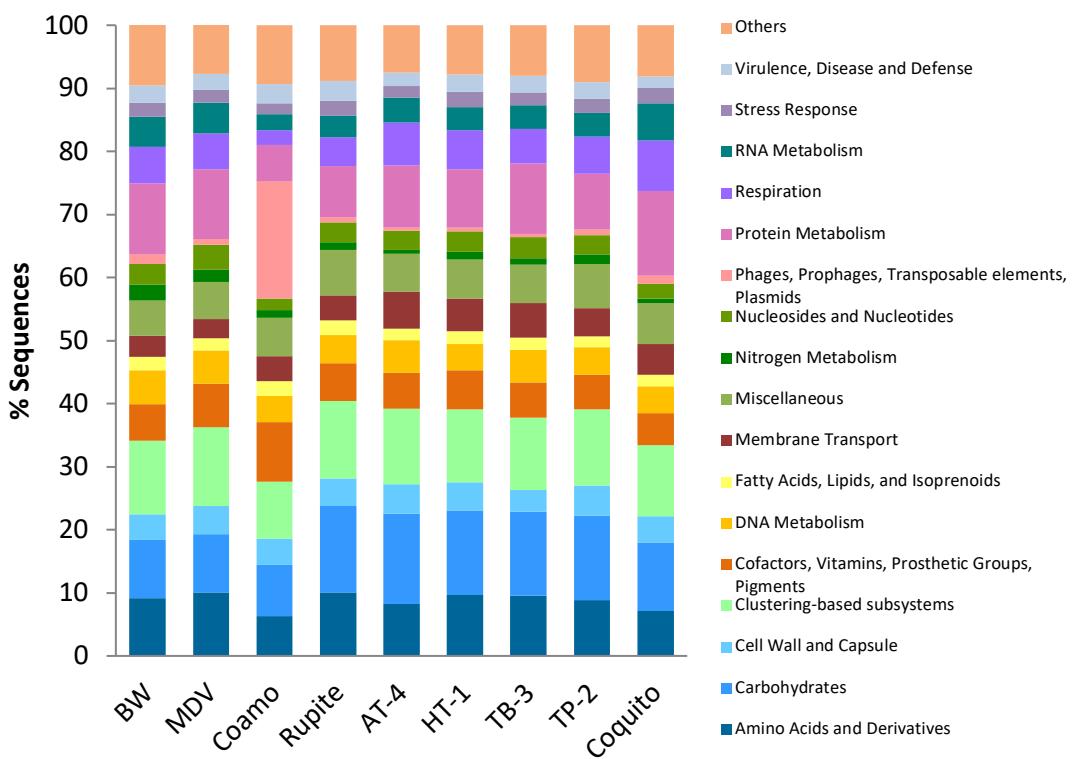
Figure 6. Comparative functional diversity at level 1. Others include those subsystems with less than 2 % sequences annotated in the 8 metagenomes: Cell division and cell cycle, dormancy and sporulation, iron acquisition and metabolism, metabolism of aromatic compounds, motility and chemotaxis, phosphorus metabolism, photosynthesis, potassium metabolism, regulation and cell signaling, secondary metabolism, and sulfur metabolism.

CONCLUSIONS

As the two hot springs from Ourense showed small differences in temperature and pH, the differences in bacterial community between BW and MDV are due to the previously described differences in the geochemical composition of their waters (González-Barreiro et al., 2009). The dominance of a heterotrophic population of the genus *Thermus* in BW is related to the high abundance of organic compounds detected in this geothermal spring (González-Barreiro et al., 2009), while in MDV there is a predominance of chemolithoautotrophy performed by the genus *Sulfurihydrogenibium.*

Taxonomic and functional analyses showed that primary production in both hot springs is mainly driven by members of the genera *Sulfurihydrogenibium*, *Hydrogenobacter,* and *Thermocrinis,* which are sulfur and hydrogen oxidizers that can fix carbon using the reverse tricarboxylic acid pathway (rTCA). However, the

differences in the abundance of these genera between the two metagenomes suggest that genera *Sulfurihydrogenibium* and *Hydrogenobacter* are the main responsible for carbon fixation in MDV and BW, respectively.

The higher concentration of sulfate in MDV might be behind the abundance of genus *Thermodesulfovibrio* in this hot spring in which the existence of a sulfur cycle between the two dominant genera (*Sulfurihydrogenibium* and *Thermodesulfovibrio*) is taking place.

In BW ammonia oxidation driven by Archaea of the genus *Nitrosopumilus* can occur, while this genus is not abundant in MDV. This finding is associated with the relatively higher concentration of ammonia in BW, while the lower concentration of NH^{4+} in the MDV ecosystem could be driving a selection to nitrogen fixation, performed by members of the genus *Thermodesulfovibrio*, among others.

From a functional point of view, the results suggest a clear interaction between nitrogen and sulfur cycles in both metagenomes as some members of the genera *Thermodesulfovibrio*, *Sulfurihydrogenibium,* and *Thermus* have been described as important players in both biogeochemical cycles and are abundantly found in BW and MDV hot springs, as well as the sequences related with metabolic pathways involved in nitrogen and sulfur cycles.

When compared to other geographically distant hot springs metagenomes, a clear effect of pH and temperature determining the taxonomy and function of hot springs microbial community can be detected. Phylum Cyanobacteria dominates in low-temperature hot springs, Proteobacteria in moderate-temperature, and Aquificae and Deninococcus-Thermus are more abundant in the high-temperature hot springs.

REFERENCES

Alvarez, L., Bricio, C., Blesa, A., Hidalgo, A., and Berenguer, J. (2014). Transferable denitrification capability of thermus thermophilus. *Appl. Environ. Microbiol.* 80, 19–28. doi:10.1128/AEM.02594-13.

Amin, A., Ahmed, I., Salam, N., Kim, B.-Y., Singh, D., Zhi, X.-Y., et al. (2017). Diversity and distribution of thermophilic bacteria in hot springs of Pakistan. *Microb. Ecol.* 74, 116–127. doi:10.1007/s00248-017-0930-1.

Aoshima, M., Ishii, M., and Igarashi, Y. (2004). A novel enzyme, citryl-CoA lyase, catalysing the second step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. *Mol. Microbiol.* 52, 763–770. doi:10.1111/j.1365-2958.2004.04010.x.

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.

Banerjee, R., Roy, A., and Mukhopadhyay, S. (2014). Genomic and proteomic signatures of radiation and thermophilic adaptation in the *Deinococcus-Thermus* genomes. Int. J. Pharm. Pharm. Sci. 6, 287–300.

Bennett, A. C., Murugapiran, S. K., and Hamilton, T. L. (2020). Temperature impacts community structure and function of phototrophic Chloroflexi and Cyanobacteria in two alkaline hot springs in Yellowstone National Park. *Environ. Microbiol. Rep.* 12, 503–513. doi:10.1111/1758-2229.12863.

Bjornsdottir, S. H., Petursdottir, S. K., Hreggvidsson, G. O., Skirnisdottir, S., Hjorleifsdottir, S., Arnfinnsson, J., et al. (2009). *Thermus islandicus* sp. nov., a mixotrophic sulfur-oxidizing bacterium isolated from the Torfajokull geothermal area. *Int. J. Syst. Evol. Microbiol.* 59, 2962–2966. doi:10.1099/ijs.0.007013-0.

Bonjour, F., and Aragno, M. (1986). Growth of thermophilic, obligatorily chemolithoautotrophic hydrogen-oxidizing bacteria related to *Hydrogenobacter* with thiosulfate and elemental sulfur as electron and energy source. *FEMS Microbiol. Lett.* 35, 11–15. doi:10.1111/j.1574-6968.1986.tb01490.x.

Chan, C. S., Chan, K.-G., Ee, R., Hong, K.-W., Urbieta, M. S., Donati, E. R., et al. (2017). Effects of physiochemical factors on prokaryotic biodiversity in Malaysian circumneutral hot springs. *Front. Microbiol.* 8, 1252. doi:10.3389/fmicb.2017.01252.

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.

Chaudhuri, B., Chowdhury, T., and Chattopadhyay, B. (2017). Comparative analysis of microbial diversity in two hot springs of Bakreshwar, West Bengal, India. *Genomics Data* 12, 122–129. doi:10.1016/j.gdata.2017.04.001.

Chen, K. Y., and Morris, J. C. (1972). Kinetics of Oxidation of Aqueous Sulfide by O_2. *Environ. Sci. Technol.* 6, 529–537. doi:10.1021/es60065a008.

Chen, S., Peng, X., Xu, H., and Ta, K. (2016). Nitrification of archaeal ammonia oxidizers in a high-temperature hot spring. *Biogeosciences* 13. doi:10.5194/bg-13-2051-2016.

Chiriac, C. M., Szekeres, E., Rudi, K., Baricz, A., Hegedus, A., Dragoş, N., et al. (2017). Differences in temperature and water chemistry shape distinct diversity patterns in thermophilic microbial communities. *Appl. Environ. Microbiol.* 83. doi:10.1128/AEM.01363-17.

Deive, F. J., Álvarez, M. S., Sanromán, M. A., and Longo, M. A. (2013). North Western Spain hot springs are a source of lipolytic enzyme-producing thermophilic microorganisms. *Bioprocess Biosyst. Eng.* 36, 239–250. doi:10.1007/s00449-012-0780-7.

Delgado-Outeiriño, I., Araujo-Nespereira, P., Cid-Fernández, J. A., Mejuto, J. C., Martínez-Carballo, E., and Simal-Gándara, J. (2009). Behaviour of thermal waters through granite rocks based on residence time and inorganic pattern. *J. Hydrol.* 373, 329–336. doi:10.1016/j.jhydrol.2009.04.028.

Dodsworth, J. A., Hungate, B. A., and Hedlund, B. P. (2011). Ammonia oxidation, denitrification and dissimilatory nitrate reduction to ammonium in two US Great Basin hot springs with abundant ammonia-oxidizing archaea. *Environ. Microbiol.* 13, 2371–2386. doi:10.1111/j.1462-2920.2011.02508.x.

Ferrandi, E. E., Sayer, C., De Rose, S. A., Guazzelli, E., Marchesi, C., Saneei, V., et al. (2018). New thermophilic α/β class epoxide hydrolases found in metagenomes from hot environments. *Front. Bioeng. Biotechnol.* 6, 144. doi:10.3389/fbioe.2018.00144.

Filippidou, S., Wunderlin, T., Junier, T., Jeanneret, N., Dorador, C., Molina, V., et al. (2016). A Combination of extreme environmental conditions favor the prevalence of endospore-forming Firmicutes. *Front. Microbiol.* 7, 1707. doi:10.3389/fmicb.2016.01707.

Frank, Y. A., Kadnikov, V. V., Lukina, A. P., Banks, D., Beletsky, A. V., Mardanov, A. V., et al. (2016). Characterization and Genome Analysis of the First Facultatively Alkaliphilic Thermodesulfovibrio Isolated from the Deep Terrestrial Subsurface. *Front. Microbiol.* 7, 2000. doi:10.3389/fmicb.2016.02000.

Garrity, G. M., Holt, J. G., Spieck, E., Bock, E., Johnson, D. B., Spring, S., et al. (2001). "Phylum BVIII. Nitrospirae phy. nov.," in *Bergey's Manual® of Systematic Bacteriology* (Springer New York), 451–464. doi:10.1007/978-0-387-21609-6_25.

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.

Ghilamicael, A. M., Boga, H. I., Anami, S. E., Mehari, T., and Budambula, N. L. M. (2018). Potential human pathogenic bacteria in five hot springs in Eritrea revealed by next generation sequencing. *PLoS One* 13, e0194554. doi:10.1371/journal.pone.0194554.

González-Barreiro, C., Cancho-Grande, B., Araujo-Nespereira, P., Cid-Fernández, J. A., and Simal-Gándara, J. (2009). Occurrence of soluble organic compounds in thermal waters by ion trap mass detection. *Chemosphere* 75, 34–47. doi:10.1016/J.CHEMOSPHERE.2008.11.067.

Guo, L., Wang, G., Sheng, Y., Sun, X., Shi, Z., Xu, Q., et al. (2020). Temperature governs the distribution of hot spring microbial community in three hydrothermal fields, Eastern Tibetan Plateau Geothermal Belt, Western China. *Sci. Total Environ.* 720, 137574. doi:10.1016/j.scitotenv.2020.137574.

Hamilton, T. L., Koonce, E., Howells, A., Havig, J. R., Jewell, T., de la Torre, J. R., et al.

(2014). Competition for ammonia influences the structure of chemotrophic communities in geothermal springs. *Appl. Environ. Microbiol.* 80, 653–661. doi:10.1128/AEM.02577-13.

Haouari, O., Fardeau, M. L., Cayol, J. L., Fauque, G., Casiot, C., Elbaz-Poulichet, F., et al. (2008). Thermodesulfovibrio hydrogeniphilus sp. nov., a new thermophilic sulphate-reducing bacterium isolated from a Tunisian hot spring. *Syst. Appl. Microbiol.* 31, 38–42. doi:10.1016/j.syapm.2007.12.002.

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 yellowstonii, gen. nov. and sp. nov.: its phylogenetic relationship to Thermodesulfobacterium commune and their origins deep within the bacterial domain. *Arch. Microbiol.* 161, 62–69. doi:10.1007/BF00248894.

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.

Hügler, M., Huber, H., Molyneaux, S. J., Vetriani, C., and Sievert, S. M. (2007). Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. *Environ. Microbiol.* 9, 81–92. doi:10.1111/j.1462-2920.2006.01118.x.

Hügler, M., and Sievert, S. M. (2011). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Ann. Rev. Mar. Sci.* 3, 261–289. doi:10.1146/annurev-marine-120709-142712.

Hussein, E. I., Jacob, J. H., Shakhatreh, M. A. K., Abd Al-razaq, M. A., Juhmani, A. F., and Cornelison, C. T. (2017). Exploring the microbial diversity in Jordanian hot springs by comparative metagenomic analysis. *Microbiologyopen* 6, e00521. doi:10.1002/mbo3.521.

Jiang, H., Dong, C. Z., Huang, Q., Wang, G., Fang, B., Zhang, C., et al. (2012). Actinobacterial diversity in microbial mats of five hot springs in Central and Central-Eastern Tibet, China. *Geomicrobiol. J.* 29, 520–527. doi:10.1080/01490451.2011.590872.

Jiménez, D. J., Andreote, F. D., Chaves, D., Montaña, 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, 1–15. doi:10.1371/journal.pone.0052069.

Kaushal, G., Kumar, J., Sangwan, R. S., and Singh, S. P. (2018). Metagenomic analysis of geothermal water reservoir sites exploring carbohydrate-related thermozymes. *Int. J. Biol. Macromol.* 119, 882–895. doi:10.1016/j.ijbiomac.2018.07.196.

Konhauser, K. O., Jones, B., Reysenbach, A. L., and Renaut, R. W. (2003). Hot spring sinters: Keys to understanding Earth's earliest life forms. *Can. J. Earth Sci.* 40, 1713–1724. doi:10.1139/e03-059.

Krupovic, M., and Forterre, P. (2011). *Microviridae* goes temperate: Microvirus-related Proviruses reside in the genomes of Bacteroidetes. *PLoS One* 6, e19893. doi:10.1371/journal.pone.0019893.

Kumar, U., Panneerselvam, P., Gupta, V. V. S. R., Manjunath, M., Priyadarshinee, P., Sahoo, A., et al. (2018). Diversity of sulfur-oxidizing and sulfur-reducing microbes in diverse ecosystems. In Advances in Soil Microbiology: Recent Trends and Future Prospects. *Microorganisms for Sustainability* 3, 65-89. doi:10.1007/978-981-10-6178-3_4.

Lavrentyeva, E. V., Radnagurueva, A. A., Barkhutova, D. D., Belkova, N. L., Zaitseva, S. V., Namsaraev, Z. B., et al. (2018). Bacterial diversity and functional activity of microbial communities in hot springs of the Baikal Rift zone. *Microbiol. (Russian Fed.* 87, 272–281. doi:10.1134/S0026261718020078.

Leira, M., Meijide-Failde, R., and Torres, E. (2017). Diatom communities in thermo-mineral springs of Galicia (NW Spain). *Diatom Res.* 32, 29–42. doi:10.1080/0269249X.2017.1286266.

Li, H., Yang, Q., Li, J., Gao, H., Li, P., and Zhou, H. (2015). The impact of temperature on microbial diversity and AOA activity in the Tengchong Geothermal Field, China. *Sci. Rep.* 5, 1–12. doi:10.1038/srep17056.

Li, J., and Liu, X. (2017). Identification and Characterization of a novel thermophilic, organic solvent sable lpase of *Bacillus* from a hot spring. *Lipids* 52, 619–627. doi:10.1007/s11745-017-4265-y.

Li, M., Cao, H., Hong, Y., and Gu, J. D. (2011). Spatial distribution and abundances of ammonia-oxidizing archaea (AOA) and ammonia-oxidizing bacteria (AOB) in mangrove sediments. *Appl. Microbiol. Biotechnol.* 89, 1243–1254. doi:10.1007/s00253-010-2929-0.

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, L., Salam, N., Jiao, J. Y., Jiang, H. C., Zhou, E. M., Yin, Y. R., et al. (2016). Diversity of culturable thermophilic Actinobacteria in hot springs in Tengchong, China and studies of their biosynthetic gene profiles. *Microb. Ecol.* 72, 150–162. doi:10.1007/s00248-016-0756-2.

López-López, O., Knapik, K., Cerdán, M. E., and González-Siso, M. I. (2015). 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.

López, D. L., Araujo, P. A., Outeiriño, I. D., Cid, J. A., and Astray, G. (2019). Geochemical signatures of the groundwaters from Ourense thermal springs, Galicia, Spain. *Sustain. Water Resour. Manag.* 5, 103–116. doi:10.1007/s40899-018-0239-3.

Mackintosh, M. E. (1978). Nitrogen fixation by *Thiobacillus ferrooxidans. J. Gen. Microbiol.* 105, 215–218. doi:10.1099/00221287-105-2-215.

Mahato, N. K., Sharma, A., Singh, Y., and Lal, R. (2019). Comparative metagenomic analyses of a high-altitude Himalayan geothermal spring revealed temperature-constrained habitat-specific microbial community and metabolic dynamics. *Arch. Microbiol.* 201, 377–388. doi:10.1007/s00203-018-01616-6.

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.

Massello, F. L., Chan, C. S., Chan, K.-G., Goh, K. M., Donati, E., and Urbieta, M. S. (2020). Meta-analysis of microbial communities in hot springs: Recurrent taxa and complex shaping factors beyond pH and temperature. *Microorganisms* 8, 906. doi:10.3390/microorganisms8060906.

Matsuura, N., Ohashi, A., Tourlousse, D. M., and Sekiguchi, Y. (2016). Draft genome sequence of *Thermodesulfovibrio aggregans* TGE-P1T, an obligately anaerobic, thermophilic, sulfate-reducing bacterium in the phylum Nitrospirae. *Genome Announc.* 4, 89–105. doi:10.1128/genomeA.00089-16.

Mehetre, G., Shah, M., Dastager, S. G., and Dharne, M. S. (2018). Untapped bacterial diversity and metabolic potential within Unkeshwar hot springs, India. *Arch. Microbiol.* 200, 753–770. doi:10.1007/s00203-018-1484-4.

Menzel, P., Gudbergsdóttir, 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.

Merkel, A. Y., Pimenov, N. V., Rusanov, I. I., Slobodkin, A. I., Slobodkina, G. B., Tarnovetckii, I. Y., et al. (2017). Microbial diversity and autotrophic activity in Kamchatka hot springs. *Extremophiles* 21, 307–317. doi:10.1007/s00792-016-0903-1.

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 phylo- genetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386. doi:10.1186/1471-2105-9-386.

Murugapiran, S. K., Huntemann, M., Wei, C.-L., Han, J., Detter, J. C., Han, C., et al. (2013). *Thermus oshimai* JL-2 and T. thermophilus JL-18 genome analysis illuminates pathways for carbon, nitrogen, and sulfur cycling. *Stand. Genomic Sci.* 7, 449–468. doi:10.4056/sigs.3667269.

Najar, I. N., Sherpa, M. T., Das, S., Das, S., and Thakur, N. (2018). Microbial ecology of two hot springs of Sikkim: Predominate population and geochemistry. *Sci. Total Environ.* 637–638, 730–745. doi:10.1016/J.SCITOTENV.2018.05.037.

Nishihara, A., Haruta, S., McGlynn, S. E., Thiel, V., and Matsuura, K. (2018a). Nitrogen fixation in thermophilic chemosynthetic microbial communities depending on hydrogen, sulfate, and carbon dioxide. *Microbes Environ.* 33, 10–18. doi:10.1264/jsme2.ME17134.

Nishihara, A., Matsuura, K., Tank, M., McGlynn, S. E., Thiel, V., and Haruta, S. (2018b). Nitrogenase activity in thermophilic chemolithoautotrophic bacteria in the

phylum Aquificae; Isolated under nitrogen-fixing conditions from Nakabusa hot springs. *Microbes Environ.* 33, 394–401. doi:10.1264/jsme2.ME18041.

O'Neill, A. H., Liu, Y., Ferrera, I., Beveridge, T. J., and Reysenbach, A. L. (2008). *Sulfurihydrogenibium rodmanii* sp. nov., a sulfur-oxidizing chemolithoautotroph from the Uzon Caldera, Kamchatka Peninsula, Russia, and emended description of the genus Sulfurihydrogenibium. *Int. J. Syst. Evol. Microbiol.* 58, 1147–1152. doi:10.1099/ijs.0.65431-0.

Padilla-Del Valle, R., Morales-Vale, L. R., and Ríos-Velázquez, C. (2017). Unraveling the microbial and functional diversity of Coamo thermal spring in Puerto Rico using metagenomic library generation and shotgun sequencing. *Genomics Data* 11, 98–101. doi:10.1016/J.GDATA.2016.12.010.

Panda, A. K., Bisht, S. S., De Mandal, S., and Kumar, N. S. (2016). Bacterial and archeal community composition in hot springs from Indo-Burma region, North-east India. *AMB Express* 6, 111. doi:10.1186/s13568-016-0284-y.

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.

Paul, S., Cortez, Y., Vera, N., Villena, G. K., and Gutiérrez-Correa, M. (2016). Metagenomic analysis of microbial community of an Amazonian geothermal spring in Peru. *Genomics Data* 9, 63–66. doi:10.1016/j.gdata.2016.06.013.

Pedron, R., Esposito, A., Bianconi, I., Pasolli, E., Tett, A., Asnicar, F., et al. (2019). Genomic and metagenomic insights into the microbial community of a thermal spring. *Microbiome* 7, 8. doi:10.1186/s40168-019-0625-6.

Pirajno, F. (2020). Subaerial hot springs and near-surface hydrothermal mineral systems past and present, and possible extraterrestrial analogues. *Geosci. Front.* 11, 1549–1569. doi:10.1016/j.gsf.2020.04.001.

Podar, P. T., Yang, Z., Björnsdóttir, S. H., and Podar, M. (2020). Comparative analysis of microbial diversity across temperature gradients in hot Springsfrom Yellowstone and Iceland. *Front. Microbiol.* 11, 1625. doi:10.3389/fmicb.2020.01625.

Power, J. F., Carere, C. R., Lee, C. K., Wakerley, G. L. J., Evans, D. W., Button, M., et al. (2018). Microbial biogeography of 1,000 geothermal springs in New Zealand. *bioRxiv*, 247759. doi:10.1101/247759.

Rabelo-Fernandez, R. J., Santiago-Morales, K., Morales-Vale, L., and Rios-Velazquez, C. (2018). The metagenome of *Caracolus marginell*a gut microbiome using culture independent approaches and shotgun sequencing. *Data Br.* 16, 501–505. doi:10.1016/j.dib.2017.11.043.

Rawlings, D. E., and Kusano, T. (1994). Molecular genetics of *Thiobacillus ferrooxidans*. *Microbiol. Mol. Biol. Rev.* 58.

Rekadwad, B., and Khobragade C.N. (2016). Bioinformatics data supporting revelatory diversity of cultivable thermophiles isolated and identified from two terrestrial hot springs, Unkeshwar, India. *Data in Brief.* 7, 1511– 1514. doi:10.1016/j.dib.2016.04.038.

Rudolph, B., Gebendorfer, K. M., Buchner, J., and Winter, J. (2010). Evolution of *Escherichia coli* for growth at high temperatures. *J. Biol. Chem.* 285, 19029–19034. doi:10.1074/jbc.M110.103374.

Sahay, H., Yadav, A. N., Singh, A. K., Singh, S., Kaushik, R., and Saxena, A. K. (2017). Hot springs of Indian Himalayas: potential sources of microbial diversity and thermostable hydrolytic enzymes. *3 Biotech* 7, 118. doi:10.1007/s13205-017-0762-1.

Sahoo, R. K., Kumar, M., Sukla, L. B., and Subudhi, E. (2017). Bioprospecting hot spring metagenome: lipase for the production of biodiesel. *Environ. Sci. Pollut. Res.* 24, 3802–3809. doi:10.1007/s11356-016-8118-7.

Saxena, R., Dhakan, D. B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A., et al. (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.

Schmieder, R., Edwards, R., and Bateman, A. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinforma. Appl. NOTE* 27, 863–864. doi:10.1093/bioinformatics/btr026.

Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B., and Dunfield, P. F. (2014). Humboldt's spa: Microbial diversity is controlled by temperature in geothermal environments. *ISME J.* 8, 1166–1174. doi:10.1038/ismej.2013.237.

Singh, Y., Gulati, A., Singh, D. P., and Khattar, J. I. S. (2018). Cyanobacterial community structure in hot water springs of Indian North-Western Himalayas: A morphological, molecular and ecological approach. *Algal Res.* 29, 179–192. doi:10.1016/J.ALGAL.2017.11.023.

Skirnisdottir, S., Hreggvidsson, G. O., Holst, O., and Kristiansson, J. K. (2001). Isolation and characterization of a mixotrophic sulfur-oxidizing *Thermus scotoductus*. *Extremophiles* 5, 45–51. doi:10.1007/s007920000172.

Sompong, U., Hawkins, P. R., Besley, C., and Peerapornpisal, Y. (2005). The distribution of Cyanobacteria across physical and chemical gradients in hot springs in Northern Thailand. *FEMS Microbiol. Ecol.* 52, 365–376. doi:10.1016/j.femsec.2004.12.007.

Sonne-Hansen, J., and Ahring, B. K. (1999). *Thermodesulfobacterium hveragerdense* sp. nov., and *Thermodesulfovibrio islandicus* sp. nov., two thermophilic sulfate reducing bacteria isolated from a Icelandic hot spring. *Syst. Appl. Microbiol.* 22, 559–564. doi:10.1016/S0723-2020(99)80009-5.

Soriano, B. M., Del Valle-Perez, L. M., Morales-Vale, L., and Rios-Velazquez, C. (2018). Datasets generated by shotgun sequencing of metagenomic libraries of the Guajataca water reservoir. *Data Br.* 21, 2531–2535. doi:10.1016/j.dib.2018.11.114.

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

doi:10.3389/fmicb.2013.00084.

Takai, K., Kobayashi, H., Nealson, K. H., and Horikoshi, K. (2003). *Sulfurihydrogenibium subterraneum* gen. nov., sp. nov., from a subsurface hot aquifer. *Int. J. Syst. Evol. Microbiol.* 53, 823–827. doi:10.1099/ijs.0.02506-0.

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.

Thiel, V., Hügler, M., Ward, D. M., and Bryant, D. A. (2017). The dark side of the Mushroom Spring microbial mat: Life in the shadow of chlorophototrophs. II. Metabolic functions of abundant community members predicted from metagenomic analyses. *Front. Microbiol.* 8, 943. doi:10.3389/fmicb.2017.00943.

Tripathy, S., Padhi, S. K., Mohanty, S., Samanta, M., and Maiti, N. K. (2016). Analysis of the metatranscriptome of microbial communities of an alkaline hot sulfur spring revealed different gene encoding pathway enzymes associated with energy metabolism. *Extremophiles* 20, 525–536. doi:10.1007/s00792-016-0846-6.

Valverde, A., Tuffin, M., and Cowan, D. A. (2012). Biogeography of bacterial communities in hot springs: A focus on the actinobacteria. *Extremophiles* 16, 669–679. doi:10.1007/s00792-012-0465-9.

Varshney, P., Mikulic, P., Vonshak, A., Beardall, J., and Wangikar, P. P. (2015). Extremophilic micro-algae and their potential contribution in biotechnology. *Bioresour. Technol.* 184, 363–372. doi:10.1016/j.biortech.2014.11.040.

Wang, S., Hou, W., Dong, H., Jiang, H., Huang, L., Wu, G., et al. (2013a). Control of Temperature on Microbial Community Structure in Hot Springs of the Tibetan Plateau. *PLoS One* 8, e62901. doi:10.1371/journal.pone.0062901.

Wang, S., Wang, K., Li, L., and Liu, Y. (2013b). 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.

Zhang, Y., Wu, G., Jiang, H., Yang, J., She, W., Khan, I., et al. (2018). Abundant and rare microbial biospheres respond differently to environmental and spatial factors in tibetan hot springs. *Front. Microbiol.* 9, 2096. doi:10.3389/fmicb.2018.02096.

Zhou, E. M., Adegboruwa, A. L., Mefferd, C. C., Bhute, S. S., Murugapiran, S. K., Dodsworth, J. A., et al. (2020). Diverse respiratory capacity among *Thermus* strains from US Great Basin hot springs. *Extremophiles* 24, 71–80. doi:10.1007/s00792-019-01131-6.

Concluding Remarks

CONCLUDING REMARKS

The taxonomical and functional diversity of the microbial population inhabiting As Burgas hot spring have been analyzed in this work by metagenomic approaches, emphasizing the search for new thermophilic β-galactosidases in the metagenome.

The construction of a plasmid metagenomic library from As Burgas water and the functional screening for β-galactosidases led to the detection of an active β-galactosidase (BWbg1), previously uncharacterized, that was successfully cloned, expressed, and purified. The analysis of its enzymatic properties allows to conclude:

1. BWbg1 is the first described thermostable β-galactosidase from family GH35, showing an optimal temperature at 80ºC, higher than the reported for other β-galactosidases from thermophilic origin.

2. BWbg1 developed maximal activity at pH 7, near to the natural milk pH, revealing its suitability for its use in the dairy industry for the preparation of low-lactose products.

3. The high yield of GOS produced by the enzyme, its thermal stability, and heat activation after 2 hours of incubation at 65°C are also advantageous for the industrial application of BWbg1, especially for GOS production.

The study of the community composition and the functional profile of As Burgas through sequence metagenomics allow to conclude:

4. Bacteria dominate As Burgas microbial community, with Proteobacteria and Aquificae being the most abundant phyla. The prevalence of the genera *Thermus* and *Hydrogenobacter* and the relation of their metabolism suggest an association between these two genera.

5. Both heterotrophic and autotrophic populations are detected in the metagenome. Important pathways from the nitrogen and sulfur cycle such as dissimilatory nitrate reduction to ammonium, nitrification, or sulfur oxidation are potentially taking place in As Burgas hot spring, as was determined by the functional annotation of the metagenomic reads.

6. Two complete ORFs annotated as β-galactosidases were found in the metagenome. Tsbg and pTsbg belonged to *T.scotoductus* SA-01 and were cloned and overexpressed in *E.coli*. The enzyme Tsbg lacked β-galactosidase activity using ONPG and lactose as substrates. On the contrary, pTsbg showed β-galactosidase activity towards ONPG but was unable to hydrolyze lactose.

The comparisons between As Burgas and Muiño da Veiga metagenomes as well as with other metagenomes from distant geothermal springs allow to conclude:

7. The differences in bacterial communities between As Burgas and Muiño da Veiga may be due to the variability in their water geochemistry. Heterotrophic members of the genera *Thermus* dominate As Burgas while chemolitoautotrophic bacteria from the genus *Sulfurihydrogenibium* and anaerobic sulfate-reducers from the genus *Thermodesulfovibrio* dominate in the sulfate-rich water from Muiño da Veiga.

8. Primary production in both hot springs is mainly driven by sulfur and hydrogen oxidizers that can fix carbon using the reverse tricarboxylic acid pathway (rTCA) (*Sulfurihydrogenibium*, *Hydrogenobacter*, and *Thermocrinis*).

9. The abundance of the ammonia-oxidizing genus *Nitrospira* in As Burgas might be related to its higher ammonia concentration, while the lower ammonia concentration in Muiño da Veiga is driving a selective pressure to nitrogen fixation, presumably carried out by members of *Sulfurihydrogenibium*.

10. The higher concentration of sulfate in Muiño da Veiga might be determining the abundance of genus *Thermodesulfovibrio* and the existence of a sulfur cycle between the two dominant genera (*Sulfurihydrogenibium* and *Thermodesulfovibrio*) in this hot spring.

11. When compared to other geographically distant hot springs metagenomes, the influence of pH and temperature in the taxonomy and function of hot springs microbial communities can be clearly detected. The phylum Cyanobacteria is more abundant in hot springs with lower temperatures while Proteobacteria dominates in moderate

temperature hot springs and Aquificae and Deninococcus-Thermus are the predominant phyla in those springs with higher temperatures.

Appendix
Resumen

Introducción

Las β-galactosidasas son enzimas capaces de hidrolizar la lactosa liberando glucosa y galactosa. Además, algunas β-galactosidasas pueden transferir el grupo galactosil a otro carbohidrato, en lo que se conoce como reacciones de transglicosilación, produciendo galacto-oligosacáridos (GOS).

La actividad hidrolítica de las β-galactosidasas se emplea habitualmente en la industria, no sólo para la producción de leche y derivados lácteos sin lactosa, si no también para incrementar la cremosidad y el dulzor de algunos productos, así como para la revalorización del suero, un subproducto contaminante con alto contenido en materia orgánica, que debe ser eliminado.

Por otro lado, el potencial de transglicosilación de las β-galactosidasas se utiliza principalmente para la obtención de GOS, carbohidratos no digeribles capaces de inducir el crecimiento de bifidobacterias beneficiosas como *Bifidobacterium* y *Lactobacillus* (Monteagudo-Mera et al., 2016; Thongaram et al., 2017). Estos prebióticos pueden ayudar en la prevención del cáncer colorrectal (Bruno-Barcena y Azcarate-Peril, 2015), la activación del sistema inmunológico (Shokryazdan et al., 2017) y la mejora de la absorción intestinal de minerales (Whisner y Castillo, 2018; Seijo et al., 2019), por lo que se agregan frecuentemente a fórmulas lácteas infantiles, productos lácteos y alimentos para mascotas.

Las β-galactosidasas de origen termofílico presentan ventajas para su aplicación industrial, ya que pueden emplearse en combinación con altas temperaturas, para mejorar la productividad inicial, prevenir la contaminación microbiana o aumentar la solubilización de los sustratos (Pisani et al. 1990, Zolnere y Ciprovica, 2017). Los manantiales geotermales son uno de los ambientes de alta temperatura más estudiados, lo que ha permitido descubrir la gran biodiversidad que albergan estos hábitats y, por tanto, su potencial como reservorios de enzimas termoestables de interés biotecnológico. Asimismo, el estudio de los microorganismos que habitan las aguas termales puede arrojar luz sobre las primeras formas de vida de nuestro

planeta, que parecen ser de origen termal (Damer y Deamer 2019; McClendon, 1999).

La irreproducibilidad de las condiciones que se dan en las fuentes geotermales ha sido una de las principales dificultades para el estudio de su biodiversidad y su potencial enzimático, que se ha superado con el desarrollo de la metagenómica. Este enfoque, basado en el estudio de todo el ADN de la comunidad (metagenoma) de un entorno, puede abordarse de dos formas diferentes: metagenómica funcional y metagenómica de secuencia.

La metagenómica funcional depende de la extracción, fragmentación y clonado del metagenoma, seguida del cribado funcional de los clones. Esta estrategia permite detectar nuevas β-galactosidasas funcionales que no podrían ser predichas por su secuencia de ADN (Cheng et al., 2017). Aunque se han encontrado β-galactosidasas a través de la metagenómica funcional de varios ambientes como el suelo (Zhang et al., 2013; Wang et al., 2014; Cheng et al., 2017) o la paja de trigo (Maruthamuthu et al., 2016), sólo se ha descrito una β-galactosidasa termoestable aislada de una fuente termal siguiendo este enfoque (Gupta et al., 2012).

La metagenómica de secuencias requiere la extracción, secuenciación y análisis del ADN ambiental. La predicción de genes y la anotación de las lecturas metagenómicas, basada en una base de datos de secuencias de referencia, permite identificar los microorganismos que habitan las aguas termales y facilita la determinación de las funciones que realizan en el ecosistema. Además, la predicción y la anotación de genes en las secuencias metagenómicas ensambladas se pueden utilizar para identificar, amplificar y clonar enzimas de interés biotecnológico, como las β-galactosidasas, del metagenoma. Hasta ahora, solo se ha obtenido una β-galactosidasa termoestable de origen geotérmico mediante metagenómica basada en secuencia (Liu et al., 2015).

La metagenómica funcional presenta dificultades ya que debe conseguirse una expresión adecuada, así como un número representativo de clones, mientras que la metagenómica de secuencia impide encontrar nuevas enzimas que posean la

actividad deseada pero carezcan de homología de secuencia con otras ya descritas. Por ello, la combinación de ambas aproximaciones podría ser el mejor modo de abordar un ecosistema. En el presente estudio, hemos utilizado los dos enfoques metagenómicos descritos para explorar el perfil taxonómico y funcional de la fuente geotermal de As Burgas, así como para la bioprospección de nuevas β-galactosidasas termoestables.

Objetivos

El principal objetivo del trabajo de investigación previsto en esta tesis doctoral es analizar el manantial geotérmico de As Burgas desde una perspectiva metagenómica, centrándose especialmente en su potencial biotecnológico como reservorio de nuevas enzimas termoestables como las β-galactosidasas. En comparación con las β-galactosidasas termolábiles, las β-galactosidasas termoestables presentan importantes ventajas, tanto para la obtención de productos lácteos bajos en lactosa como para la producción de GOS. Por tanto, la estabilidad térmica es una cualidad deseable para el uso de β-galactosidasas en aplicaciones industriales, biotecnológicas y farmacéuticas.

Los objetivos específicos de este trabajo son

1. Encontrar y caracterizar nuevas β-galactosidasas termoestables del manantial geotérmico de As Burgas mediante metagenómica funcional.

2. Analizar la diversidad taxonómica y el potencial funcional de la comunidad microbiana que habita el agua de As Burgas y descubrir y caracterizar nuevas β-galactosidasas mediante metagenómica de secuencia.

3. Comparar la biodiversidad y composición de la comunidad de As Burgas con una fuente termal cercana (Muiño Da Veiga) y otras fuentes termales geográficamente distantes.

Capítulo 1. Identificación y caracterización de una nueva β-galactosidasa termoestable descubierta en las aguas de As Burgas mediante metagenómica funcional

El contenido del capítulo 1 se ha eliminado por encontrarse bajo secreto de patente.

Capítulo 2. Explorando el perfil taxonómico y funcional de las fuente termal de As Burgas: Caracterización de una β-galactosidasa termoestable encontrada mediante metagenómica de secuencia

El estudio mediante metagenómica de secuencia de las comunidades microbianas que habitan los manantiales geotermales puede dar información, no solo de quiénes son los microorganismos que habitan estos ecositemas, si no también de las funciones que en él desempeñan. En este capítulo se ha empleado esta herramienta para el análisis taxonómico y funcional del metagenoma de As Burgas, así como para la búsqueda de secuencias con homología con β-galactosidasas.

El análisis taxonómico de las aguas termales de As Burgas reveló la existencia de una comunidad microbiana dominada por Bacterias en la que Proteobacteria (68.25 ± 3.59 %) y Aquificae (11.24 ± 1.15 %) son los filos más abundantes. Además, la prevalencia de los géneros *Thermus* (15.77 %) e *Hydrogenobacter* (8.56 %) y la relación de su metabolismo sugiere una asociación entre ambos.

En relación al análisis funcional, la alta abundancia relativa de secuencias involucradas en el ciclo de Calvin-Benson y de secuencias anotadas como proteínas clave para el ciclo de los ácidos tricarboxílicos (TCA) revela el dominio de una población autótrofa. Asimismo, la anotación funcional de las secuencias metagenómicas determinó que diversas vías importantes del ciclo del nitrógeno y del azufre como el DNRA, la nitrificación o la oxidación del azufre están potencialmente teniendo lugar en las aguas termales de As Burgas, en concordancia con la composición microbiana del ecosistema.

Tras el ensamblaje de las lecturas metagenómicas, se encontraron dos marcos abiertos de lectura completos anotados como β-galactosidasas, denominados *Tsbg* y *pTsbg*. Ambas secuencias nucleotídicas mostraron un 100% de homología con *T. scotoductus* SA-01 y fueron clonadas y sobreexpresadas en *E. coli*, ya que no se encontraron evidencias de que las enzimas para las que codificaban hubieran sido previamente caracterizadas. La enzima Tsbg resultó carecer de actividad β-

galactosidasa empleando ONPG y lactosa como sustratos. Por el contrario, pTsbg mostró actividad β-galactosidasa hacia ONPG, con una temperature óptima de 85 °C y un pH óptimo de 6.0. Otra característica importate de la enzima es su alta termoestabilidad, dado que fue capaz de retener hasta el 60 % de su actividad frente a ONPG tras 24 horas de incubación a 75 °C. Pese a que no fue capaz de hidrolizar la lactosa; pTsbg mostró actividad β-fucosidasa sobre el sustrato p-nitrofenil-β-D-fucopiranósido, lo que sugiere a priori la posibilidad de una aplicación biotecnológica inesperada. Una vez más, este resultado revela que la presencia de un gen en un metagenoma no garantiza que codifique para una enzima activa en la forma prevista a partir de la secuencia, y destaca la importancia de combinar la metagenómica funcional y de secuencia para encontrar nuevas enzimas a partir de metagenomas.

Este estudio independiente de cultivo ha proporcionado una idea de la diversidad de los microorganismos que habitan el entorno térmico de As Burgas y de las actividades que estos desempeñan, en un intento por encontrar nuevas β-galactosidasas. La investigación futura deberá estar dirigida a caracterizar nuevos entornos, lo que conducirá a una mejor comprensión de sus diferencias ecológicas y a encontrar nuevas enzimas de interés.

Capítulo 3. Análisis metagenómico comparativo de dos fuentes termales de Ourense, Noroeste de España

Pese a la proximidad geográfica de las termas de As Burgas y Muiño da Veiga, ambas presentan grandes diferencias en la composición química y mineral de sus aguas, debido a la gran variabilidad geológica de la región (López et al. 2019). En este trabajo, se ha utilizado una aproximación metagenómica combinada con herramientas estadísticas para comparar la composición y las características funcionales de las poblaciones microbianas que habitan las aguas termales de As Burgas (BW) y Muiño da Veiga (MDV). Además, en un intento por determinar si las condiciones ambientales como el pH o la temperatura determinan la diversidad microbiana y la función de estos ecosistemas cercanos, hemos analizado el perfil funcional y taxonómico de estas fuentes geotermales y otras geográficamente distantes que presentan diferentes pH y temperaturas.

Dado que las aguas termales de BW y MDV muestran pequeñas diferencias en cuanto a temperatura y pH, las disimilitudes en sus comunidades microbianas se deben a las diferencias descritas en la composición geoquímica de sus aguas (González-Barreiro et al., 2009). El predominio de una población heterótrofa del género *Thermus* en las aguas de As Burgas está relacionado con la alta abundancia de compuestos orgánicos detectados en este manantial geotérmico (González-Barreiro et al., 2009), mientras que en MDV domina la quimiosíntesis, realizada por el género *Sulfurihydrogenibium*.

El análisis taxonómico y funcional reveló que la producción primaria en ambos ecosistemas termales es llevada a cabo principalmente por miembros de los géneros *Sulfurihydrogenibium*, *Hydrogenobacter* y *Thermocrinis*, bacterias oxidantes de azufre e hidrógeno que pueden fijar carbono utilizando el ciclo del ácido tricarboxílico inverso (rTCA). Sin embargo, las diferencias en la abundancia de estos géneros entre los dos metagenomas sugieren que los géneros *Sulfurihydrogenibium* e *Hydrogenobacter* son los principales responsables de la fijación de carbono en MDV y BW, respectivamente.

La mayor concentración de sulfato en MDV podría estar detrás de la abundancia del género *Thermodesulfovibrio* en esta fuente termal en la que existe un ciclo de azufre entre los dos géneros dominantes en el que *Sulfurihydrogenibium* oxida los compuestos del azufre reducidos por *Thermodesulfovibrio* y viceversa.

En BW las archaeas del género *Nitrosopumilus* pueden llevar a cabo la oxidación del amonio, mientras que este género no es abundante en MDV. Este hallazgo está asociado con la concentración relativamente más alta de amonio en BW, mientras que la menor concentración de NH^{4+} en el ecosistema de MDV podría estar impulsando una selección hacia la fijación de nitrógeno, en la que participan miembros del abundante género *Thermodesulfovibrio*, entre otros.

Desde un punto de vista funcional, los resultados sugieren una clara interacción entre los ciclos de nitrógeno y azufre en ambos metagenomas, ya que algunos miembros de los géneros *Thermodesulfovibrio*, *Sulfurihydrogenibium* y *Thermus* han

sido descritos como actores importantes en ambos ciclos biogeoquímicos y se encuentran en abundancia en las aguas termales de BW y de MDV, así como las secuencias relacionadas con las vías metabólicas implicadas en ambos ciclos biogeoquímicos.

En comparación con otros metagenomas de fuentes termales geográficamente distantes, se puede detectar un efecto claro del pH y la temperatura que determinan la taxonomía y la función de la comunidad microbiana de las fuentes termales. Como resultado de este análisis podría afirmarse que el filo Cyanobacteria domina en las aguas termales de baja temperatura, Proteobacteria en las de temperatura moderada y Aquificae y Deninococcus-Thermus son los más abundantes en las aguas termales de alta temperatura.

Conclusiones

La diversidad taxonómica y funcional de la población microbiana que habita las aguas termales de As Burgas ha sido analizada en este trabajo mediante enfoques metagenómicos, enfatizando en la búsqueda de nuevas β-galactosidasas termofílicas en el metagenoma.

La construcción de una metagenoteca en plásmidos a partir de agua de As Burgas y la búsqueda funcional de β-galactosidasas condujeron a la detección de una β-galactosidasa activa (BWbg1), previamente no caracterizada, que se clonó, expresó y purificó con éxito. El análisis de sus propiedades enzimáticas permite concluir:

1. BWbg1 es la primera β-galactosidasa termoestable descrita de la familia GH35, presenta una temperatura óptima de 80ºC, superior a la reportada para otras β-galactosidasas de origen termófilo.

2. BWbg1 mostró su actividad máxima a pH 7, cercano al pH natural de la leche, lo que la hace idónea para su uso en la industria láctea en la preparación de productos bajos en lactosa.

3. El alto rendimiento de GOS producido por la enzima, su estabilidad térmica y la activación por calor después de 2 horas de incubación a 65 ° C también son

ventajosos para la aplicación industrial de BWbg1, especialmente para la producción de GOS.

El estudio de la composición de la comunidad y el perfil funcional de As Burgas a través de la metagenómica de secuencias permite concluir:

4. Las bacterias dominan la comunidad microbiana de As Burgas, siendo Proteobacteria y Aquificae los filos más abundantes. La prevalencia de los géneros *Thermus* e *Hydrogenobacter* y la relación de su metabolismo sugieren una asociación entre ambos.

5. En el metagenoma se detectan poblaciones heterótrofas y autótrofas. Rutas importantes del ciclo del nitrógeno y del azufre, como la reducción disimilatoria del nitrato a amonio, la nitrificación o la oxidación del azufre, están teniendo lugar potencialmente en las aguas termales de As Burgas, como se determinó mediante la anotación funcional de las secuencias metagenómicas.

6. Se encontraron dos marcos abiertos de lectura (ORFs) completos anotados como β-galactosidasas en el metagenoma. *Tsbg* y *pTsbg* pertenecían a *T. scotoductus* SA-01 y se clonaron y sobreexpresaron en *E. coli*. La enzima Tsbg carecía de actividad β-galactosidasa usando ONPG y lactosa como sustratos. Por el contrario, pTsbg mostró actividad β-galactosidasa hacia ONPG pero fue incapaz de hidrolizar la lactosa.

Las comparaciones entre los metagenomas de As Burgas y Muiño da Veiga así como con otros metagenomas de fuentes geotermales distantes permiten concluir:

7. Las diferencias en las comunidades bacterianas entre As Burgas y Muiño da Veiga pueden deberse a la variabilidad en la geoquímica del agua. Los miembros heterótrofos del género *Thermus* dominan As Burgas, mientras que las bacterias quimioautótrofas del género *Sulfurihydrogenibium* y los reductores de sulfato anaerobios del género *Thermodesulfovibrio* dominan en el agua rica en sulfatos de Muiño da Veiga.

8. La producción primaria en ambas aguas termales es llevada a cabo principalmente por oxidantes del azufre e hidrógeno que pueden fijar carbono utilizando la vía del

ácido tricarboxílico inverso (rTCA) (*Sulfurihydrogenibium*, *Hydrogenobacter* y *Thermocrinis*).

9. La abundancia de bacterias oxidantes de amonio del género *Nitrospira* en As Burgas podría estar relacionada con su mayor concentración de amonio, mientras que la menor concentración de dicho compuesto en Muiño da Veiga está impulsando una presión selectiva hacia la fijación de nitrógeno, presumiblemente llevada a cabo por miembros del género *Sulfurihydrogenibium*.

10. La mayor concentración de sulfato en Muiño da Veiga podría estar determinando la abundancia del género *Thermodesulfovibrio* y la existencia de un ciclo de azufre entre los dos géneros dominantes (*Sulfurihydrogenibium* y *Thermodesulfovibrio*) en esta fuente termal.

11. En comparación con otros metagenomas de fuentes termales distantes geográficamente, se puede detectar claramente la influencia del pH y la temperatura en la taxonomía y función de las comunidades microbianas de las fuentes termales. El filo Cianobacteria es más abundante en manantiales de aguas termales con temperaturas más bajas mientras que las Proteobacterias dominan en manantiales de temperatura moderada y Aquificae y Deninococcus-Thermus son los filos predominantes en aquellos manantiales con temperaturas más altas.

Referencias

Bruno-Barcena, J. M., and Azcarate-Peril, M. A. (2015). Galacto-oligosaccharides and colorectal cancer: Feeding our intestinal probiome. *J. Funct. Foods* 12, 92–108. doi:10.1016/j.jff.2014.10.029.

Cheng, J., Romantsov, T., Engel, K., Doxey, A. C., Rose, D. R., Neufeld, J. D., et al. (2017). Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. *PLoS One* 12, e0172545. doi:10.1371/journal.pone.0172545.

Damer, B., and Deamer, D. (2019). The hot spring hypothesis for an origin of life. *Astrobiology* 20, 429–452. doi:10.1089/ast.2019.2045.

González-Barreiro, C., Cancho-Grande, B., Araujo-Nespereira, P., Cid-Fernández, J. A., and Simal-Gándara, J. (2009). Occurrence of soluble organic compounds in thermal waters by ion trap mass detection. *Chemosphere* 75, 34–47.

doi:10.1016/J.CHEMOSPHERE.2008.11.067.

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.

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.* 2818. doi:10.1080/13102818.2015.1015244.

López, D. L., Araujo, P. A., Outeiriño, I. D., Cid, J. A., and Astray, G. (2019). Geochemical signatures of the groundwaters from Ourense thermal springs, Galicia, Spain. *Sustain. Water Resour. Manag.* 5, 103–116. doi:10.1007/s40899-018-0239-3.

Maruthamuthu, M., Jiménez, D. J., Stevens, P., and Dirk Van Elsas, J. (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.

McClendon, J. H. (1999). The origin of life. *Earth Sci. Rev.* 47, 71–93. doi:10.1016/S0012-8252(99)00015-X.

Monteagudo-Mera, A., Arthur, J. C., Jobin, C., Keku, T., Bruno-Barcena, J. M., Azcarate-Peril, M. A., et al. (2016). High purity galacto-oligosaccharides (GOS) enhance specific Bifidobacterium species and their metabolic activity in the mouse gut microbiome. *Benef Microbes* 7, 247–264. doi:10.3920/BM2015.0114.

Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., et al. (1990). Thermostable beta-galactosidase from the archaebacterium *Sulfolobus solfataricus*: Purification and properties. *Eur. J. Biochem.* 187, 321–328. doi:10.1111/j.1432-1033.1990.tb15308.x.

Shokryazdan, P., Faseleh Jahromi, M., Navidshad, B., and Liang, J. B. (2017). Effects of prebiotics on immune system and cytokine expression. *Med. Microbiol. Immunol.* 206, 1–9. doi:10.1007/s00430-016-0481-y.

Thongaram, T., Hoeflinger, J. L., Chow, J., and Miller, M. J. (2017). Prebiotic galactooligosaccharide metabolism by probiotic *Lactobacilli* and *Bifidobacteria*. 65, 4184–4192 doi:10.1021/acs.jafc.7b00851.

Wang, S., Guo, G., Li, L., Cao, L., Tong, L., Ren, G., 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.

Whisner, C. M., and Castillo, L. F. (2018). Prebiotics, bone and mineral metabolism. *Calcif. Tissue Int.* 102, 443–479. doi:10.1007/s00223-017-0339-3.

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.

Zolnere, K., and Ciprovica, I. (2017). The comparison of commercially available β-galactosidases for dairy industry : review. Research for rural Development, Annual 23rd International Scientific Conference Proceedings, Latvia, 1. doi:10.22616/rrd.23.2017.032.