THÈSE DE DOCTORAT D'AIX-MARSEILLE UNIVERSITÉ

École doctorale nº 251 : Sciences de l'Environnement
MIO -Institut Méditerranéen d'Océanologie

Présentée et soutenue publiquement par

Eléonore FROUIN

En vue de l'obtention du grade de docteur d'Aix-Marseille Université

Discipline : BioInformatique
Spécialité : Océanographie

Taxonomic and functional exploration of the biosphere of serpentinizing hydrothermal systems by metagenomics

---

Soutenue le 17/12/2018, devant le jury composé de :

Mme LÓPEZ-GARCÍA,	DR, Université Paris-Sud XI	Rapportrice
Mme PASCAL,	IR, INRA Occitanie-Toulouse	Rapportrice
Mme MÉNEZ,	PR, Université Paris Diderot, IPGP	Examinatrice
M. BRAZELTON,	Assistant Prof., University of Utah	Examinateur
M. ERAUSO,	PR, Univeristé Aix-Marseille	Directeur de thèse
M. ARMOUGOM,	IR, Univeristé Aix-Marseille	Co-directeur de thèse

Remerciements

Ce travail de thèse a été réalisé à l'Institut Méditerranéen d'Océanologie (MIO) avec le soutien financier du Ministère de l'Enseignement Supérieur et de la Recherche. Je remercie M. Richard Sempéré de m'avoir accueillie au sein du MIO et Mme Catherine Keller pour son suivi au sein de l'école doctorale ED251. Je remercie Mme Patricia Bonin, responsable de l'équipe Microbiologie Environnementale et Biotechnologie (MEB), pour ses conseils et son soutien amical.

Je remercie les rapporteurs, Mme Purificación López-García et Mme Géraldine Pascal, pour le temps qu'elles ont passé à évaluer mon travail de thèse et Mme Bénédicte Ménez et M. William Brazelton, d'avoir accepté de faire partie du jury.

Merci à mes deux directeurs de thèse M. Gaël Erauso et M. Fabrice Armougom pour m'avoir associée à leur projet de recherche sur les extrêmophiles et m'avoir fait confiance en me laissant une grande liberté dans mes choix d'analyses. Ce fut une expérience très enrichissante que de pouvoir participer à une campagne de terrain et de réaliser quelques expériences de biologie moléculaire en parallèle à l'analyse des données. Merci de m'avoir fait découvrir le monde de la recherche et de m'avoir encouragée à présenter mes résultats dans différentes conférences.

Merci à tous les membres de la thématique Prony, à Anne Postec et à Marianne Quéméneur pour m'avoir associée à leurs travaux. Merci à Bernard Ollivier pour nos discussions sur les métabolismes. Merci à Matthew Schrenk pour m'avoir apporté son expertise sur les milieux serpentinisés et son aide pour la rédaction. Merci à Méline et Nan pour tous les travaux qu'elles ont menés en amont de ma thèse et pour tous les conseils prodigués pour mener à bien ce doctorat.

Merci à Maurice Libes pour les nombreuses réinstallations du serveur de calcul bioinfo qui a connu quelques déconvenues au cours des trois dernières années. Merci également à Didier Laborie de m'avoir accordé des accès aux clusters de Genotoul qui m'ont permis de réaliser les assemblages de données métagénomiques.

Merci à Pascal Hingamp pour les discussions relatives aux approches métagénomiques et pour avoir soutenu ma candidature pour le monitorat que j'ai pu réaliser lors de ma dernière année de thèse. Ce fut une expérience très enrichissante.

Merci aux membres de l'équipe MEB Marie, Léa S., Cécile, Laurie, Gwenola, Corinne, Sophie, Manon, Léa C., pour leurs discussions joyeuses, les trajets de covoiturage et leur bonne humeur. Merci à Pierre, Pierre-Pol, et Sylvain pour leur amitié et leurs discussions et à Émilie P. pour sa gaité contagieuse et son optimisme.

Merci à tous mes collègues de bureau : à Hassiba toujours prête à partager ses dernières réussites culinaires, à Émilie V. pour son enthousiasme et ses conseils et enfin à Romain, toujours à l'écoute et très positif, pour sa gentillesse, son aide pour Docker et tous nos échanges sur les derniers softwares/packages en vogue.

Une mention spéciale pour les membres de la Dream Team, Lisa et Guillaume, mes co-thésards. Merci pour votre présence et votre patience pour discuter des dernières expériences, des résultats insolites ou des difficultés de rédaction, sans oublier bien sûr votre amitié pendant ces trois ans à Marseille incluant les nombreuses sessions d'escalade, randonnée, canoë, ciné, bar... Merci pour tout.

Un grand merci à Alaïs, Julie, Claire, Pierre, Matthieu, Yasmin pour votre soutien, les WEs détentes et nos échanges sur nos débuts dans la vie professionnelle. Votre amitié m'est très précieuse.

Merci à mes deux sœurs, Bérengère pour son soutien sans faille et ses paroles réconfortantes et Mélisande pour m'avoir montré combien on peut/doit rester serein même sous pression. Merci à mes parents qui ont toujours patiemment répondu au raz-de-marée de mes questions. Merci de m'avoir encouragée à cultiver ma curiosité scientifique et de m'avoir montré la voie pour "chercher par moi-même". Un grand merci à ma famille et belle-famille pour les conseils, les relectures et les encouragements. Et enfin une reconnaissance infinie à toi, Béranger, pour ton soutien scientifique et moral, ta confiance et ta présence à mes côtés.

Contents

List of Figures

List of Tables

Introduction (French part)

Les systèmes serpentinisés proviennent de l'hydratation des roches ultramafiques, roches caractéristiques du manteau lithosphérique. Sur Terre, le phénomène de serpentinisation s'observe à proximité des dorsales océaniques "lentes" ou dans des structures ophiolitiques. Les fluides hydrothermaux générés dans ces systèmes sont fortement alcalins et enrichis en hydrogène, méthane et autres petites molécules organiques (acétate, formiate, petits hydrocarbures, etc.). Les environnement serpentinisés suscitent un intérêt particulier dans la communauté scientifique car ils sont considérés comme de bons analogues aux environnements ayant permis l'émergence de la vie sur la Terre primitive il y a quelques 3,8 milliards d'années. Les systèmes serpentinisés sont colonisés par des micro-organismes extrêmophiles qui se développent grâce à leurs interactions avec des composants issus de processus abiotiques. Les écosystèmes serpentinisés sont souvent décrits comme des "fenêtres ouvertes" sur la biosphère des environnement de subsurface, la circulation des fluides entrainant les déplacements des communautés microbiennes. Les environnements serpentinisés présentent des conditions de vie extrêmes, en raison des pH très élevés, de la disponibilité limitée en accepteurs d'électrons, et pour certains, de fortes température et pression. La biodiversité observée met en lumière les capacités d'adaptation uniques des micro-organismes indigènes.

Les relations entre communautés microbiennes et processus abiotiques sont centrales dans les écosystèmes serpentinisés. Ainsi, il a été proposé que les composés organiques issus de la serpentinisation permettent le développement de micro-organismes chimioorganotrophes, qui seraient les producteurs majeurs de ces écosystèmes (Schrenk et al., 2013). Cependant la grande diversité géochimique de ces systèmes fait émerger des interrogations : les micro-organismes colonisateurs présentent-ils une unité taxonomique ou fonctionnelle dans différents milieux serpentinisés ? Le processus de serpentinisation impacte-t-il davantage les communautés microbiennes que d'autres facteurs tels que la pression hydrostatique ou la géochimie des fluides hydrothermaux ?

L'objectif principal de ces travaux de thèse est de mieux comprendre le fonctionnement écologique de ces milieux extrêmes. L'exploration de la diversité taxonomique et fonctionnelle de la biosphère des milieux serpentinisés a été réalisée selon deux axes de recherche. Le premier axe était majoritairement centré sur le site de Prony, un écosystème serpentinisé côtier de Nouvelle Calédonie. Le système hydrothermal de Prony est situé à faible profondeur (à environ 50m sous le niveau de la mer) et constitue un observatoire de la serpentinisation facilement

accessible. De plus, comme ce site se situe à la transition entre les compartiments terrestres et marins, il semble idéal pour faire le lien entre sites hydrothermaux sous-marins et sites ophiolitiques. Le second axe repose sur une comparaison de différents écosystèmes serpentinisés pour faire émerger leurs ressemblances et tenter d'établir des signatures biologiques associées aux phénomènes de serpentinisation.

Ce mémoire de thèse s'organise en six chapitres. Les deux premiers chapitres présentent le contexte du travail. Le premier chapitre présente un état de l'art sur la biosphère microbienne qui colonise les milieux serpentinisés. Le second chapitre présente les différentes approches OMICS utilisées en écologie microbienne en général et dans l'étude des systèmes serpentinisés en particulier. En particulier, ce chapitre décrit les outils utilisés et les analyses effectuées pour répondre aux objectifs de cette thèse. Les trois chapitres suivants présentent les résultats obtenus sous forme de publications (une publiée, une soumise et une dernière en préparation). Ainsi le chapitre 3 démontre l'importance des phylotypes en relative faible abondance qui constituent la "biosphère rare". Dans le chapitre 4, une approche métagénomique permet de comparer les communautés microbiennes vivant dans plusieurs systèmes hydrothermaux serpentinisés (marins et continentaux) et d'autres systèmes hydrothermaux (sources chaudes et cheminées hébergées dans des roches basaltiques). Cette étude permet la mise en valeur de caractéristiques métaboliques communes à tous les écosystèmes serpentinisés. Le chapitre 5, quant à lui, présente quelques caractéristiques génomiques de 82 génomes procaryotes, reconstruits à partir des métagénomes issus de cheminées hydrothermales du site de Prony. Un génome, affilié au genre *Serpentinomonas*, y est étudié plus en détail. Pour finir, le dernier chapitre propose une discussion sur les méthodes utilisées et les résultats obtenus au cours de cette thèse et donne quelques pistes de réflexion pour les perspectives de ce projet.

Chapter 1

Microbial diversity of serpentinizing systems

“ Study hard what interests you the most, in the most undisciplined, irreverent, and original manner possible. ”

Richard Feynman

Sommaire

1.1 Serpentinization and associated reactions

The lithosphere is the rigid outermost shell of Earth (Figure 1.1). It is composed of the oceanic and continental crusts and a portion of the upper mantle. Earth's lithosphere is subdivided into tectonic plates, delimited by convergent, divergent and transform boundaries. The lithosphere is underlain by the asthenosphere (hotter and less rigid part of the upper mantle), allowing tectonic plates to move relative to each other. The oceanic lithosphere is essentially composed of basalts and gabbros in the oceanic crust and ultramafic rocks in the upper mantle (Ciazela et al., 2015). New oceanic crust is continuously being produced at mid-ocean ridges and is recycled back to the mantle at subduction zones (Figure 1.1). However, parts of the oceanic lithosphere are sometimes carried over continental lithosphere during an obduction. These lithosphere fragments are referred as ophiolites.

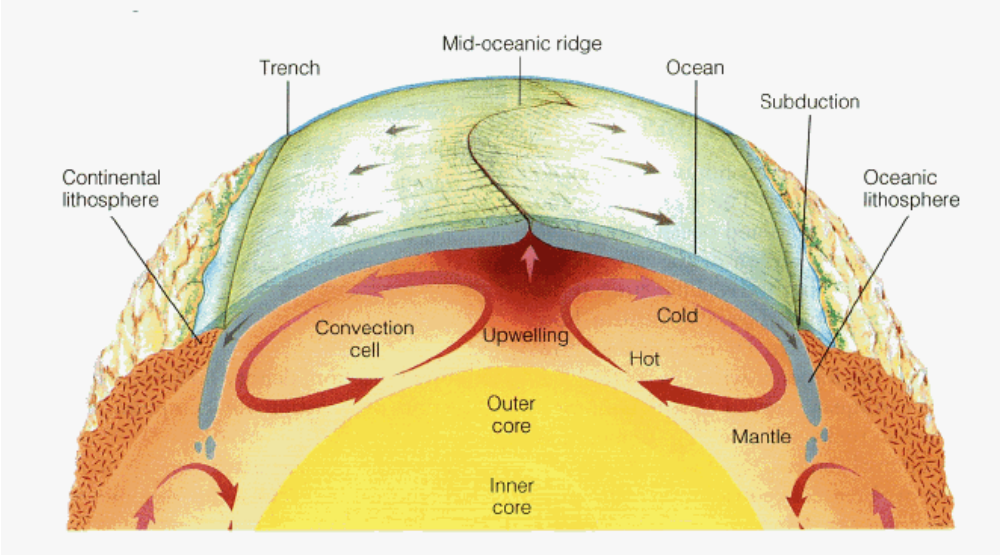


Figure 1.1 – Illustration showing the mantle convection and its relation to the movements of tectonic plates, from the Byrd Polar Research Center at Ohio State University.

Serpentinization is a geological metamorphic process in which water interacts with ultramafic rocks (Figure 1.2). Ultramafic rocks have relatively low silica contents (less than 45% by weight) but a high content of ferromagnesian minerals like olivine and pyroxene (Holm et al., 2006). The hydration of ultramafic rocks, mainly comprising peridotites, leads to the production of new sets of minerals such as serpentine [$(Fe,Mg)_3Si_2O_5(OH)_4$], brucite [$Mg(OH)_2$] and magnetite [Fe_3O_4]. The process of serpentinization can be portrayed by the general reaction 1.1 (McCollom and Seewald, 2013). The exothermic reaction of serpentinization results in the release of abundant molecular hydrogen (H_2) (McCollom and Seewald, 2013). The generation of H_2 is a consequence of the oxidation of ferrous iron (Fe^{II}) from the reactant minerals to ferric iron (Fe^{III}), which precipitates in mineral products (reaction 1.2).

$$olivine + pyroxene + H_2O \rightarrow serpentine + brucite + magnetite + H_2 \tag{1.1}$$

$$2(Fe^{II}O)_{rock} + H_2O \rightarrow (Fe_2^{III}O_3)_{rock} + H_2 \tag{1.2}$$

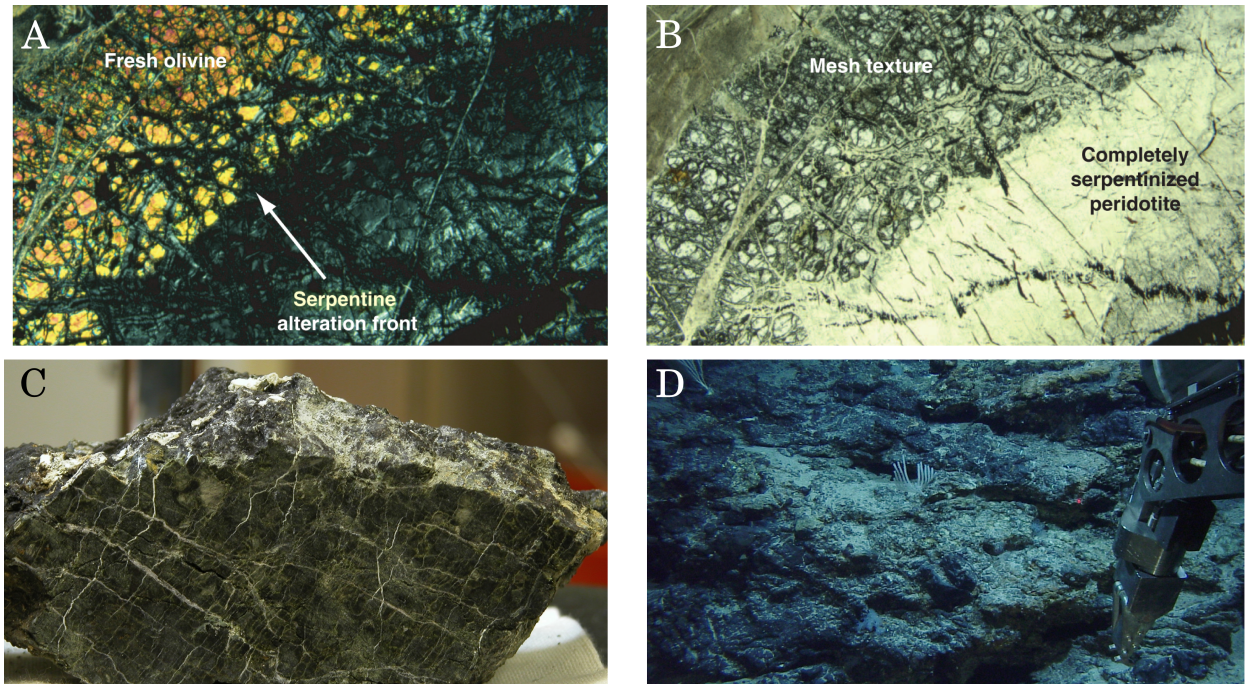


Figure 1.2 – **A**, **B**. Photomicrographs from Kelemen et al. (2004) showing serpentinization of olivine in rocks from Mid-Atlantic Ridge with cross-polarized light and plane-polarized light, respectively. **C**. Sample of a serpentinite recovered from the Atlantis Massif. Thin fractures are filled with calcium carbonate (Credit: G. Früh-Green). **D**. Serpentinized peridotites that occur at the top of the Atlantis Massif. The arm of the submersible Alvin is visible in the right portion of the image (Credit: Lost City 2003).

In hydrothermal systems, H_2 is a feedstock for a wide range of chemical reactions (Bradley, 2016), such as Fischer-Tropsch type (FTT) synthesis (equation 1.3) or Sabatier reaction (equation 1.4), also involving mantle CO_2 or other inorganic carbon sources, such as CO, to produce abiogenic hydrocarbons (Mccollom and Seewald, 2007, Proskurowski et al., 2008).

$$(2n+1)H_2 + nCO \rightarrow C_nH_{(2n+2)} + nH_2O \tag{1.3}$$

$$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O \tag{1.4}$$

It should be noted that FTT or Sabatier reactions are often considered at temperatures above 200°C, despite the existence of many low-temperature serpentinizing systems. In 2011, a study reported abiotic production of methane (CH_4) from abiotic reduction of inorganic carbon during experimental hydration of olivine at low temperatures, between 30°C and 70°C (Neubeck et al., 2011). However, McCollom argued that the actual origin of the methane remained uncertain because Neubeck's study lacked methods to identify the source of carbon (McCollom, 2016). In low-temperature serpentinizing environments, McCollom instead suggested that the potential for abiotic synthesis of CH_4 by reduction of dissolved inorganic carbon would be much more limited than previously inferred and could rather be generated by an unrecognized background source. According to the author, the abiotic reduction of dissolved inorganic carbon to methane in low-temperature serpentinizing environments indeed requires very specific conditions to occur, such as much longer time of fluid residence than those of the previous experiments, the presence of catalytic minerals (e.g., NiFe alloys), or the presence of H_2-rich vapor

phase (McCollom, 2016). Parallel to this work, Etiope and Ionescu focused on the ruthenium (Ru) as a catalyst in serpentinizing systems to produce methane at temperatures below 100°C (Etiope and Ionescu, 2015). Further studies will be needed to demonstrate that R*u* minerals are indeed effective catalysts in natural environments.

At temperatures below 150°C, the serpentinization reactions result in high pH, typically above 10 (Schrenk et al., 2013). The main reason of this hyperalkalinity is that hydroxide ions (OH^-) produced by hydration of olivine and pyroxene are not counterbalanced by OH^- consumption by precipitation of serpentine minerals except at very high pH (Palandri and Reed, 2004). The instability of clinopyroxenes (and more specifically Ca-silicate minerals, such as diopside) during serpentinization leads to an increase the calcium ions (Ca^{2+}) in the fluid (Morrill et al., 2013). The Ca^{2+} ions, react with carbonate ions (CO_3^{2-}) at high pH to induce calcium carbonate precipitation (Schrenk et al., 2013). Calcium carbonates ($CaCO_3$) are often detected in these environments as fracture infillings in the host rock (Figure 1.2C) or as large travertines in continental margins context or chimneys (up to several tens of meters high) in submarine systems (Figure 1.4).

1.2 Distribution and geochemistry of serpentinizing systems

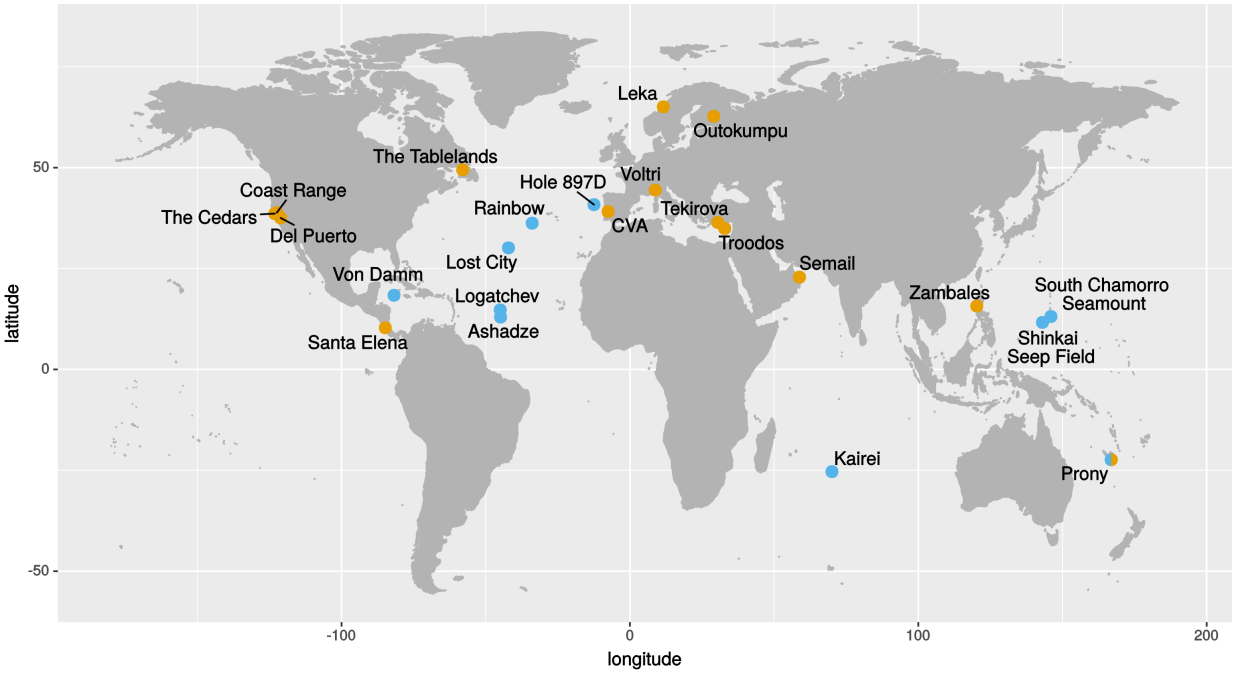


Figure 1.3 – Global distribution of serpentinizing systems mentioned in this manuscript. Continental sites are indicated by orange circles, while submarine systems are indicated by blue circles.

Serpentinizing systems are ubiquitous throughout the world (Figure 1.3). The serpentinization occurs in a variety of tectonic settings where mantle rocks are exposed to aqueous fluids, including subduction zones, slow expansion mid-ocean ridges, and ophiolites (Figure 1.4)

(Schrenk et al., 2013). Moreover, the reactions of serpentinization do not only occur on Earth (Figure 1.4). The presence of serpentinites has indeed been demonstrated on Mars (Ehlmann et al., 2010) and the serpentinization is even supposed to drive hydrothermal circulation in planetary bodies such as Europa, one of the moons of Jupiter (Vance et al., 2007).

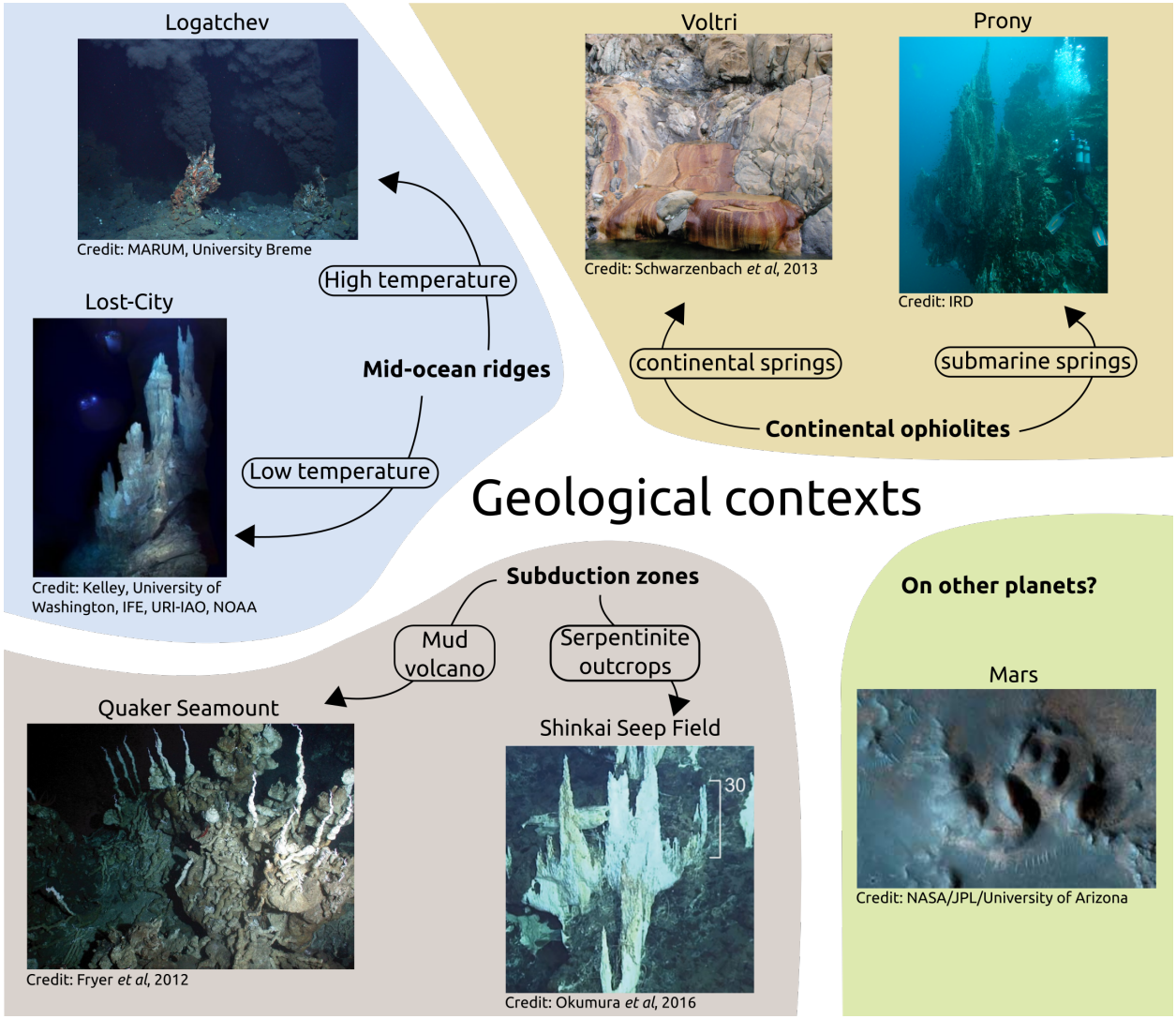


Figure 1.4 – A few examples of serpentinite-hosted environments from different geological settings

1.2.1 Submarine serpentinite-hosted systems

Discovery of the Lost-City Hydrothermal Field

A hydrothermal vent is a fissure from which geothermally heated water is emitted. The best-known submarine hydrothermal fields are located along mid-ocean ridges and are created by the cooling of hot basaltic material (Kelley et al., 2001). They were first discovered off the coast of the Galapagos, in 1977 (Corliss et al., 1979). Named "black smokers", these vents emit very hot (200-400°C), acidic (pH 2 to 4) and highly reduced fluids. These hydrothermal fluids are enriched in metals, such as Fe^{2+}, Cu^{2+}, Zn^{2+}, Mn^{2+} and commonly contain high concentration of

dissolved gasses, such as H_2S, CO_2 and CH_4 (Martin et al., 2008). When the fluids reach the surface of the oceanic crust, the precipitation of many dissolved minerals occurs upon mixing with cold oxygenated seawater and forms sulfide chimneys. Unexpectedly, these ecosystems harbor dense biological communities. These observations have expanded our knowledge of life under extreme conditions such as hydrostatic pressure, high temperature and lack of photosynthetic production (Gaill, 1993).

The Lost-City Hydrothermal Field (LCHF), discovered in 2000, represents a new class of submarine hydrothermal ecosystems that differs greatly from black smokers. Hosted on ultramafic oceanic crust, the LCHF is located on the Atlantis massif at 30°N, 15km from the Mid-Atlantic Ridge and at a water depth of 800m (Kelley et al., 2001). The heat produced by subseafloor serpentinization reactions may drive hydrothermal flow within this system (Kelley et al., 2001). On the contrary to typical basalt-hosted vents, at LCHF the hydrothermal fluids have moderate temperatures (40-90°C), high pH (9-11) and low metal content. The LCHF is outstanding in that it hosts many active and inactive hydrothermal edifices, the largest of which reaches 60m in height. These hydrothermal structures are composed primarily of calcium carbonate ($CaCO_3$: aragonite, calcite) and magnesium hydroxide ($Mg(OH)_2$: brucite) minerals, making them very different from black smoker chimneys (Früh-Green et al., 2003, Kelley et al., 2001).

Table 1.1 – Geochemical characteristics of serpentinizing hydrothermal vents, modified from Schrenk et al. (2013)

Sites	Seawater[α]	Lost-City[β]	Rainbow[γ]	Logatchev[α]	Ashadze[α]	South Chamorro Seamount[δ]	Kairei[ε]
Pays		Mid-Atlantic Ridge	Mid-Atlantic Ridge	Mid-Atlantic Ridge	Mid-Atlantic Ridge	Mariana Forearc	Central Indian Ridge
PHYSICOCHEMICAL PARAMETERS							
Depth (m)		700 to 800	2300	2700 to 3000	3263 to 4088	2960	2422 to 2452
pH	7.8	8.7 to 11.0	2.8	4.1 to 4.2	3.1 to 4.1	12.5	3.4 to 3.6
Temperature (°C)	2	24 to 91	360 to 365	320 to 352	269 to 355		315 to 365
GEOCHEMICAL PARAMETERS							
Na^+ (mM)	464	479 to 485	553 to 573	112 to 430	298 to 505	610 ± 10	484 to 528
K^+ (mM)	9.8		18.6 to 20.4	5.4 to 27.6	8.7 to 29.4	19 ± 1	12.5 to 15.2
Ca^{2+} (mM)	10.2	21.0 to 23.3	66.6 to 77.5	10.7 to 29.6	15.3 to 42.8	0.3 ± 0.1	25.6 to 31.3
Mg^{2+} (mM)		9 to 19	0	0	0	<0.01	0.2 to 0.5
Cl^- (mM)	546	546 to 549	750 to 796	127 to 515	326 to 614	510 ± 5	571 to 623
SO_4^{2-} (mM)	28.2	5.9 to 12.9	0	0	0		bdl to 1.56
SiO_2 (mM)			6.9	8.2 to 11.5	6.6 to 7.3	28 ± 1	
DOC (µM)		58 to 106					
δ^{13}C DOC		-21 to -10.5					
acetate (µM)		bdl to 35				0 to 208	
formate (µM)		bdl to 158				0 to 2272	
GASEOUS GEOCHEMISTRY AND ISOTOPIC COMPOSITION							
CO_2 (mM)	2.3	0.0001 to 0.026	5 to 16	6.2 to 10.1	3.7		8.57 to 10.1
CH_4 (mM)	0.0003	0.28 to 1.98	1.6 to 2.5	1.2 to 2.6	0.5 to 1.2	2 ± 1	0.12 to 0.20
δ^{13}C-CH_4 (‰)		-13.6 to 9.5	-17.7 to -15.8	-10.3 to -6.1	-14.1 to -8.7		-18.0 to 9.8
C_2H_6 (µM)		0.25 to 1.8	0.8	0.19 to 0.8	0.1 to 5.7	2	
H_2 (mM)	0.0004	0.48 to 14.38	13 to 16	9 to 12.5	8 to 26		2.48 to 8.19
δ^2H-H_2 (‰)			-379 to -356	-350 to -231	-343 to -270		

[α] Charlou et al. (2010) [β] Kelley et al. (2001, 2005), Lang et al. (2010) [γ] Charlou et al. (2002) [δ] Haggerty and Fisher (1992), Mottl et al. (2003) [ε] Gallant and Von Damm (2006), Kumagai et al. (2008), Takai et al. (2004)

Mid-oceanic ridge systems

Serpentinization reactions may also occurred in magmatic-hydrothermal systems from slow-spreading ridges having ultramafic basements, such as Rainbow, Logatchev, Ashadze located on the Mid-Atlantic Ridge (MAR). Such geochemical settings were also detected at Kairei on the Central Indian Ridge and at Von Damm on the Mid-Cayman Rise (Caribbean Sea). These high-temperature hydrothermal systems are characterized by a high concentration of methane and hydrogen (Table 1.1), in contrast to purely basalt-hosted systems. However, compared to the Lost-City Hydrothermal Field, these hydrothermal systems have higher temperatures (up to 365°C), more acidic pH (between 2.8 and 4.2) and higher concentration of CO_2 (Table 1.1).

Subduction zones and mud volcano

Serpentinizing environments also occurred in subduction zones. The most studied site is located in the Mariana Forearc, off Japan, where the Pacific plate subducts under the Philippine Sea plate. During the subduction of an oceanic lithosphere, high temperature leads to the dehydration of marine sediments and basement rocks of the descending slab (Mottl et al., 2003). The released fluids hydrate the overlying peridotite mantle wedge and enable serpentinite muds (which has low density) to rise along faults to the seafloor (Fryer, 2012). These muds reach the seafloor forming edifices of mud volcanoes. Alkaline fluids (pH up to 12.5) emitted from mud volcanoes are enriched with H_2, CH_4, formate and acetate (Haggerty and Fisher, 1992, Mottl et al., 2003) (Table 1.1). There, temperatures are low in the hills (5 to 20°C) but high at depth (100 to 250°C where the serpentinization occurs).

Recently, Ohara and collaborators reported a serpentinite-hosted hydrothermal vent within a subduction zone (southern Mariana forearc), consisting of serpentinized peridotite outcrops instead of mud volcanoes. This site, named Shinkai Seep Field (SSF), is the deepest known serpentinite-hosted system in the world, located at ~ 5700m depth (Ohara et al., 2012). The chimneys observed at the SSF mainly consist of brucite, calcite and aragonite with various proportions depending on the chimneys (Okumura et al., 2016). This site showed that serpentinization can also occur on the forearc slopes where peridotite is exposed and suggests that there are more SSF-like systems to be discovered along the subduction zones.

1.2.2 Terrestrial serpentinizing systems, the ophiolites

Ophiolites are sections of the oceanic lithosphere carried over a continent during an obduction phenomenon. The term ophiolite comes from *ophio* (Greek for snake), and *litos* (stone) and refers to the appearance of green-color of many serpentinites (Juteau, 2003). In continental ophiolites, the alteration of olivine and pyroxene to serpentine may have occurred through a multi-stage history during interactions with seawater in oceanic settings, or with meteoric water on the continent (Barnes et al., 1967). The hydrothermal fluids from ophiolites generally lack dissolved ions because these fluids are of meteoric origin (Table 1.2). In some cases

ions, such as sulfate, may result from former oceanic reactions of serpentinization (Schrenk et al., 2013, Schwarzenbach et al., 2016). The fluids discharged from these hydrothermal sites generally have high Ca^{2+} concentrations. Consequently, these springs are often surrounded by mineral deposits of calcite due to precipitation of Ca^{2+} ions in contact with atmospheric CO_2.

The Coast Range ophiolite in the western United States was amongst the first active serpentinizing sites to be intensively studied (Barnes et al., 1967). The first microbiological analyses (enrichment cultures) were performed on samples from the Semail ophiolite in Oman (Bath et al., 1987). This site is one of the most extensive known continental ophiolites that are >350 km long, ~40 km wide and have an average thickness of ~5 km (Schrenk et al., 2013). Many ophiolites were subsequently subjected to geochemical and microbiological analyses, such as Del Puerto Ophiolite and The Cedars in California, The Tablelands Ophiolite in Canada, Leka Ophiolite in Norway, Troodos ophiolitic complex in Cyprus and Tekirova Ophiolite in Turkey (Figure 1.3, Table 1.2 and references herein).

The study of serpentinizing seeps in Santa Elena and Zambales ophiolites (in Costa Rica and Philippines respectively; Figure 1.3) was unique because these sites are located in tropical regions (Sánchez-Murillo et al., 2014, Woycheese et al., 2015). Indeed, the investigations of geobiology in continental serpentinizing sites have mainly been performed in temperate and subtropical regions. In these regions, precipitation is either intermittent with relatively low intensities in the temperate regions or rare with limited infiltration due to high evaporation in arid areas. In tropical environments, however, rainfall amounts and intensities are typically higher, which may facilitate water-rock interaction (Sánchez-Murillo et al., 2014). This type of hydrogeological system is of peculiar interest to microbial life, given the potential for high influxes of nutrients or dissolved carbon via meteoric water infiltration (Woycheese et al., 2015). The concentration of CH_4 and H_2 in spring fluids from these ophiolites are in the same range as the highest concentrations described at other continental serpentinizing sites (Table 1.2).

Serpentinizing subsurface aquifers have been monitored *in-situ* through drilled wells. A microbiological study of such boreholes was carried out at the Cabeço de Vide in Portugal (at 130m depth). Other boreholes were located in the Coast Range Ophiolite Microbial Observatory (CROMO, in California, USA), in Semail Ophiolite (in Oman) or in Outokumpu Ophiolite (in Finland) (Table 1.2, Itävaara et al. (2011)) .

Table 1.2 – Geochemical characteristics of continental serpentinizing systems, modified from Schrenk et al. (2013) and updated.

Sites	Semail[α]	Coast Range[β]	Del Puerto[γ]	The Cedars[δ]	Tablelands[ε]	Cabeço de Vide Aquifer[ζ]	Voltri massif[η]	Zambales[θ]	Chimaera[ι]	Leka ophiolite[κ]	Santa Elena ophiolite[λ]	Troodos ophiolite[μ]
Pays	Oman	USA	USA	USA	Canada	Protugal	Italy	Philippines	Turkey	Norway	Costa Rica	Chypre
PHYSICOCHEMICAL PARAMETERS												
pH	8.3 to 11.4	9.8 to 12.4	8.7	11.5 to 11.9	11.8-12.3	7.37 to 8.04	9.5 to 11.7	7 to 11.3	9.4 to 11.95	8.80 to 9.56	11.5 to 11.6	9.25 to 11.71
Temp. (°C)	30.1 to 35.8	14.7 to 18.2	17.8	17.1 to 17.4		19.1 to 19.2	14.0 to 24.7	26.8 to 34.4	18.5 to 19.1	8.2 to 9.4	26.1 to 29.2	
Eh (mV)	- 546 to 180	-65 to -293		-585 to -656	-609 to 186	33 to 108	-202 to -60	-703 to 15			-348 to -251	-158 to 158
Cond. (µS/cm)	297 to 3199	2900 to 11200		740 to 3010		678 to 741		170 to 606	1357 to 1359		466 to 542	
GEOCHEMICAL PARAMETERS												
Na^+ (mM)	0.591 to 12.1		0.235	0.96 to 14.69		0.42 to 1.21	0.28 to 1.27	0.06 to 5.45	0.522 to 0.557	0.429 to 0.545		61.3 to 74.4
K^+ (mM)			0.008	0.01 to 0.13		0.01 to 0.21	0.02 to 0.13	0.003 to 0.033	0.072 to 0.176	0.012		1.74 to 2.10
NH_4^+ (mM)		0.22 to 2.07		bdl				bdl to 0.05	0.0 to 0.032	bdl to 0.0005	0.0006 to 0.0017	
Ca^{2+} (mM)	0.115 to 3.82		0.087	0.94 to 1.30	0.22 to 1.59	0,41 to 1.04	0.07 to 0.86	0.027 to 1.32	0.274 to 3.462	0.031 to 0.057		0.050 to 0.968
Mg^{2+} (mM)	bdl to 1.89		4.527	0.004 to 0.036	0.002 to 0.312	2.48 to 3.06	bdl to 0.251	0 to 0.959	0.028 to 2.82	0.279 to 0.420		0 to 0.021
Cl^- (mM)	0.75 to 16.6		0.135	0.94 to 8.73	1.3 to 13.5	0.26 to 0.57	0.22 to 0.65	<0.14 to 0.53	0.515 to 0.685	0.141 to 0.543		52.96 to 62.82
SO_4^{2-} (mM)	0.01 to 0.50		0.167	bdl to 0.001		0.15 to 0.19	0.0005 to 0.0328	0 to 0.93	0.083 to 0.303	0.027 o 0.038		0.56 to 1.35
NO_3^- (mM)	bdl to 0.15			bdl		0.146 to 0.289	0.0 to 3.5	0.015 to 0.081	0.0008 to 0.0015	0.013 to 0.019		bdl to <0.002
NO_2^- (µM)	bdl to 11	bdl to 0.09						bdl	0		0	
PO_4^{3-} (µM)		0.11 to 0.89		bdl				bdl	0	0.4 to 0.36	0.1 to 0.3	
SiO_2 (µM)	5 to 247		93			715 to 1211	2.9 to 78.4	bdl to 6710				
DIC (µM)	190 to 2500	21-210		6 to 70 (TIC)			7.8 to 29	<167 to 3217	bdl to 4225	506 to 792	126.3 to 254.3	
δ^{13}C DIC				-32.6 to bdl					-11.8 to bdl			
DOC (µM)		27 to 1163		20 to 170				8 to 883	bdl to 433		5.6 to 73.2	
δ^{13}C DOC				-21.9 to bdl					-22.75 to bdl			
acetate (µM)	0.47 to 4.4	<0.41 to 57.8										
formate (µM)	1.2 to 1.7	<0.65 to 19.8										
GASEOUS GEOCHEMISTRY AND ISOTOPIC COMPOSITION												
CH_4 (µM)	bdl to 170	98 to 1983			bdl to 23.7		155 to 733	0 to 400.0		detected	145.0-912.3	
δ^{13}C-CH_4 (‰)					-28.5 to bdl						-44.0 to -0.9	
C_2H_6 (µM)					bdl to 1.33							
H_2 (µM)	bdl to 2900	<0.003 to 1.48			29.7 to 584.2		0.5 to 26.8	0.0 to 495.5		0.585	10.9-53.1	
WATER ISOTOPES												
δ^{18}O (‰)			-7.9		-12.3 to -9.5				-4.29 to -2.89		-7.18 to -7.22	
δ^2H (‰)			-57		-83 to -61.5				-23.05 to -17.71		-50.4 to -50.0	

[α] Rempfert et al. (2017) [β] Crespo-Medina et al. (2014) [γ] Blank et al. (2009) [δ] Morrill et al. (2014) [ε] Szponar et al. (2013) [ζ] Marques et al. (2008) [η] Brazelton et al. (2016), Chavagnac et al. (2013)
[θ] Cardace et al. (2015) [ι] Meyer-Dombard et al. (2015) [κ] Okland et al. (2012) [λ] Crespo-Medina et al. (2017) [μ] Rizoulis et al. (2014)

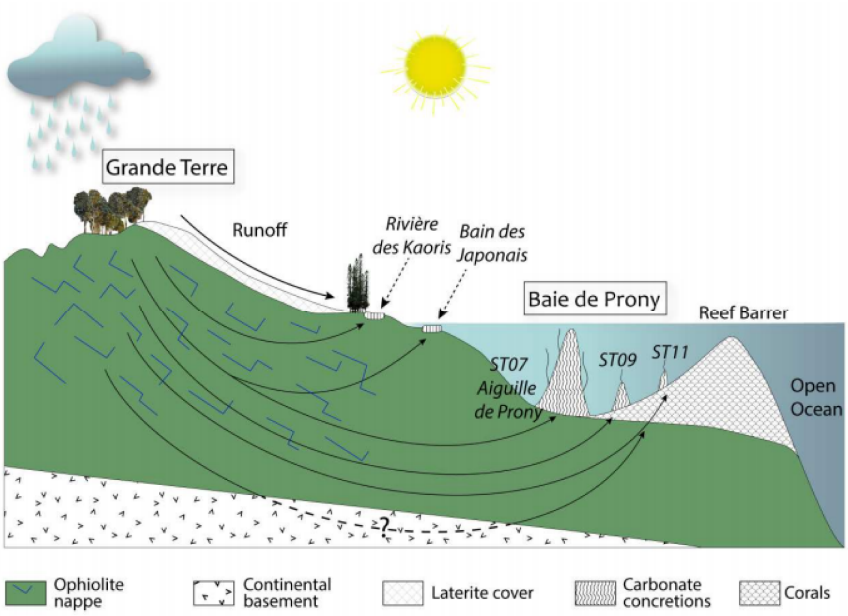


Figure 1.5 – Sketch of the geological setting of the Prony Hydrothermal Field, from Monnin et al. (2014) (not to scale). The arrows indicate meteoric waters that percolate through the ophiolite nappe and discharge at springs in the lagoon

1.2.3 Intermediate geochemical settings in the Prony Hydrothermal Field

An original kind of hydrothermalism takes place in the bay of Prony in New Caledonia. In the southern lagoon, the **P**rony **H**ydrothermal **F**ield (abbreviated PHF) consists of several submarine springs (with a depth < 50 meters below sea level) and some intertidal springs. This hydrothermalism, due to serpentinization, generates alkaline and anoxic fluids containing large quantities of hydrogen and methane. The singularity of the Prony site comes from its coastal location, which results in the discharge of hydrothermal fluids in a marine lagoon. At PHF, the fluids are meteoric waters that percolated through the ophiolitic ultramafic substratum (Figure 1.5). The fluids emitted into the seawater cause the precipitation of elements contained in the fluids to form carbonate chimneys. These chimneys can reach several meters high, similar to those of LCHF. The Prony site is, therefore, an ideal observatory to study an ecosystem located at the transition between marine and terrestrial serpentinizing sites.

Several hydrothermal chimneys were sampled during oceanographic cruises (in 2011 and 2014). In total, seven sites were samples in the Prony Bay (Figure 1.6) in a 3-km radius. The site "Rivière des Kaoris" is located above sea level, while the "Bain des Japonais" is uncovered at low tide. The five submarine sites (ST07, ST08, ST09, ST11 and ST12) are found down to 50 meters deep. Most submarine chimneys were colonized by various organisms such as ascidia, sponges and corals (Figure 1.6A, B and E), which attest for their long-lasting existence and may reduce their permeability to the surrounding seawater (Postec et al., 2015, Quéméneur et al., 2014). Fragile crumbly white needles are observed at the fluid discharge. The youngest structures were observed at the ST09 site (Figure 1.6F), where the chimneys were not colonized by marine macrofauna (Pisapia et al., 2017).

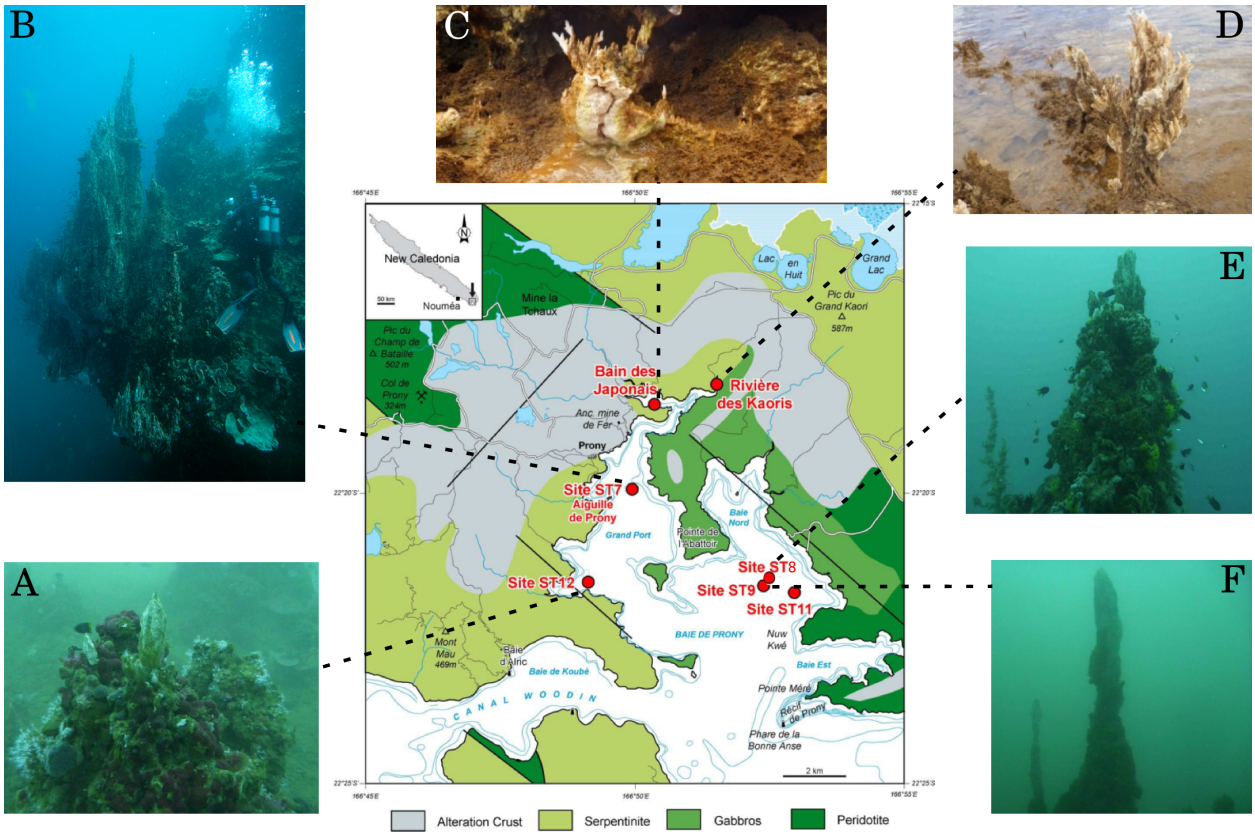


Figure 1.6 – Geological map of the Prony Bay, with the locations (modified from Maurizot and Vendé-Leclerc (2009), Monnin et al. (2014)) and pictures of sampling sites (modified from Mei et al. (2016b)).

Table 1.3 – Geochemical characteristics of fluids collected in 6 sites of the Prony Bay, summarized from Monnin et al. (2014)

Type of sites	Chimneys, covered at high tide		Submarine chimneys			
Sites	RdK	BdJ	ST07	ST09	ST11	ST12
PHYSICOCHEMICAL PARAMETERS						
Depth (m)	0	0	16.5	45		38.5
pH	10.8 to 10.9	8.30 to 11.08	9.61 to 10.14	9.18 to 10.62	8.76 to 10.64	8.15 to 11.0
Temp. (°C)	31.5 to 31.0	28.2 to 37.6				
Eh (mV)	-195	-352 to 33				
GEOCHEMICAL PARAMETERS						
Na^+ (mM)	0.58 to 0.65	1.28 to 392.8	97.32 to 346.46	6.46 to 337.76	40.52 to 386.58	40.05 to 475.50
K^+ (mM)	0.08	0.30 to 8.52	2.00 to 7.33	0.98 to 7.34	0 to 7.36	0.14 o 9.28
Ca^{2+} (mM)	0.36 o 0.38	0.21 to 9.16	3.49 to 7.16	2.76 to 8.56	2.28 to 9.45	0.55 to 10.28
Mg^{2+} (mM)	bdl	bdl to 45.16	12.43 to 41.14	2.77 to 38.77	2.52 to 45.44	bdl to 54.65
Cl^- (mM)	0.19 to 0.23	0.41 to 322.0	163.85 to 400.87	50.58 to 380.77	45.78 to 451.96	41.74 to 515.39
SO_4^{2-} (mM)	bdl	0.01 to 21.45	7.53 to 19.95	1.72 to 18.85	1.54 to 22.56	0.38 to 25.89
DIC (µM)	0 to 144	bdl to 1788	448 to 769	398 to 1688	538 to 1845	382 to 1885
GASEOUS GEOCHEMISTRY AND ISOTOPIC COMPOSITION						
CH_4 (volumetric %)	6 ± 1	13 ± 4				
H_2 (volumetric %)	24 11±	19 ± 11				
N_2 (volumetric %)	69 ± 11	67 ± 11				

1.3 Microbial life in serpentinizing ecosystems

1.3.1 Serpentinization and the origins of life hypotheses

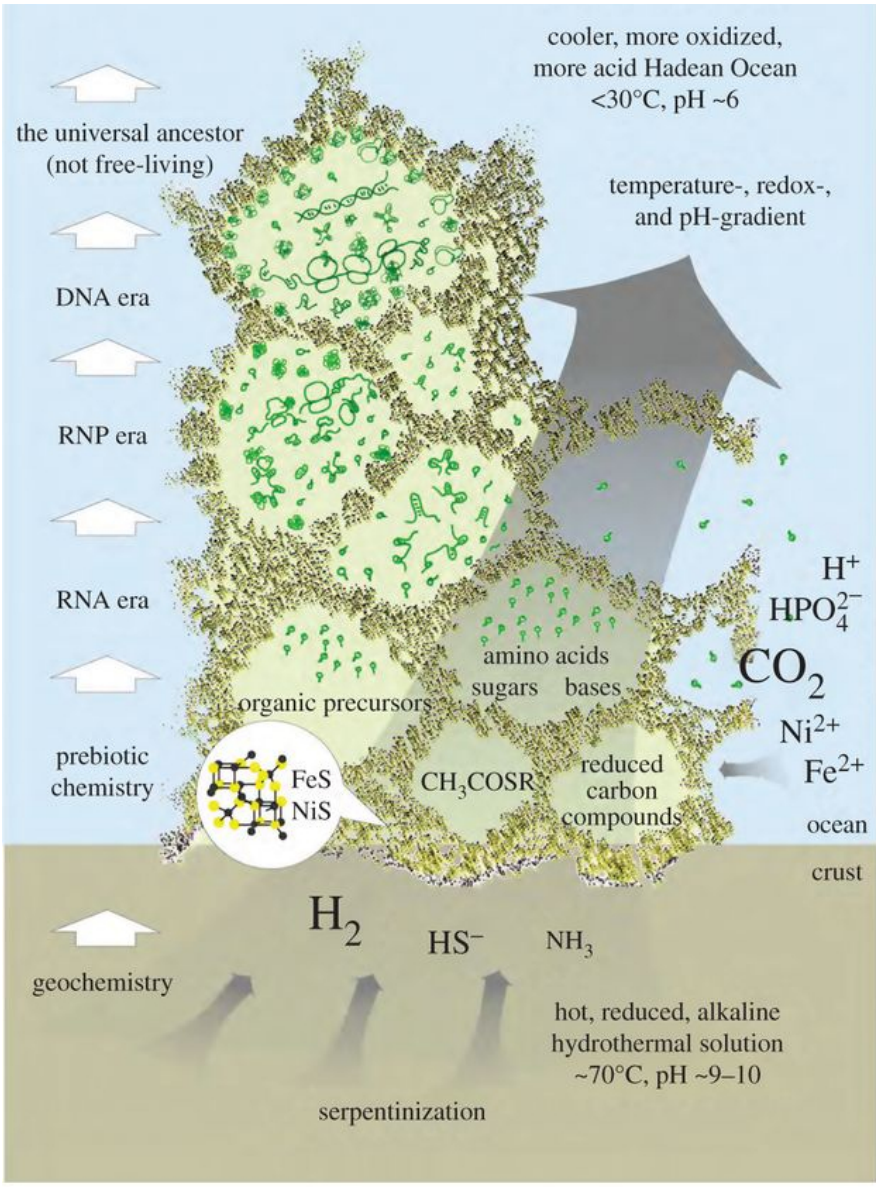


Figure 1.7 – Hypothetical origin of cells in serpentinizing environments, from Koonin and Martin (2005), Sousa et al. (2013).

Among the many scenarios of life's origin, a frequently advanced hypothesis is that living systems emerged inside hydrothermal vents (Baross and Hoffman, 1985). The microcompartments, delimited by iron monosulphide (FeS) walls and formed naturally in the hydrothermal vents (Koonin and Martin, 2005, Martin et al., 2003), have been suggested as precursors of the lipid membrane of the modern cells (Figure 1.7). In this theory, the walls of these compartments combine two important properties for early chemical evolution: catalysis (with metals) and compartmentation (to concentrate the chemical reaction products and substrates) (Sousa et al., 2013). Black smokers were quickly considered too hot and acidic to be associated with the

origin of life, but the discovery of the serpentinite-hosted LCHF, (alkaline vents with more moderate temperatures) brings out new scenarios for the origins of life (Martin and Russell, 2007). The presence of redox, pH and temperature gradients at the alkaline fluids-seawater mixing zones would be key in the earliest stages of biochemical evolution in microcompartments. The strong gradients would bring favorable thermodynamic conditions to generate new products. (Schrenk et al., 2013). Today, the ultramafic rocks that undergo serpentinization correspond only to sections of the mantle that have been uplifted and exposed to the circulation of seawater or meteoric water. However, it is assumed that serpentinization was more widespread on early Earth (during Hadean and Archean Eons) than it is today (Sleep et al., 2011). The chemistry in the Hadean Ocean would have been slightly different, with abundant Fe^{2+} and far more CO_2 in the ocean (Sousa et al., 2013).

Moreover, the serpentinizing environments, enriched in H_2, are potential locations for early metabolic evolution because hydrogen use is at the heart of modern reduction-oxidation reactions. Thus, it has been suggested that the first metabolic pathways were fueled by H_2. Interestingly, catalytic sites of enzymes that catalyze H_2 oxidation or production contain iron and/or nickel, these elements being typically enriched in serpentinites (Schrenk et al., 2013). Therefore, the characteristics of modern days serpentinizing habitats have risen in attention on these systems as possible sites for the origin and early evolution of life (Sleep et al., 2011). If life on Earth really emerged at a serpentinite-hosted system, it may be possible that life exists in other serpentinizing environments beyond Earth (McCollom and Seewald, 2013). Thus, the geochemical and microbiological studies of present-day serpentinizing systems on Earth could help us better understand the potential for life to emerge elsewhere.

1.3.2 Challenging environments and adaptation of microorganisms

The habitability of an environment can be modelled by comparing the balance of energy produced by an environment and the energy required for the growth and functioning of life forms (Hoehler, 2007). Thermodynamic models have shown that serpentinizing environments produce enough energy in the form of dihydrogen and methane to allow microbial growth (Mccollom and Seewald, 2007). Beyond the supply of energy required for life, there are many other constraints associated with serpentinizing environments.

Hyperalkaline environments

Relatively few alkaliphiles were isolated from serpentinizing environments compared to other natural alkaline environments such as soda lakes, which combine high pH and high salinity, and thus host haloalkalophilic prokaryotes (Sorokin et al., 2014). Until the start of the PHF microbiological studies, all isolates were aerobic bacteria, representing cosmopolitan genera but very few new species (Tiago et al., 2004). Among them however, three strains, isolated from The Cedars ground waters, are of peculiar interest as they have an optimal growth at pH 11, making them the most alkaliphilic organism known so far. They are also capable of autotrophic

growth using H_2 as energy source with O_2 as electron acceptor and, surprisingly $Ca CO_3$ as carbon source. These strains have been described as representing the new genus *Serpentinomonas* (Suzuki et al., 2014). Although a range of anaerobic microorganisms have been evidenced by molecular methods, and microcosms experiments previously reported (Crespo-Medina et al., 2014, Morrill et al., 2014) in serpentinizing environments, very few strains have been isolated and even less described as new species. The first anaerobes isolated from serpentinizing systems come from PHF, such as *Acetoanaerobium pronyense* ST07-YE (Bes et al., 2015), *Clostridium* sp. PROH2 (Mei et al., 2014) or *Serpentiniticella alkaliphila* 3b (Mei et al., 2016a).

In serpentinizing systems, the hydrothermal fluids commonly have a pH values greater than 10 and can even reach 12 at the discharge points (Tables 1.1, 1.2 and 1.3). As a consequence, microorganisms living therein must have specific adaptations to cope in these harsh conditions. For example, they encode enzymes or protein complexes tolerant to a high pH if they are exposed to the outside medium (Preiss et al., 2015). Additionally, cells must maintain a cytoplasmic pH that is significantly lower than the pH of the environment to protect cytoplasmic components. To overcome this difficulty, alkaliphiles possess proteins to maintain pH homeostasis such as cation/H^+ antiporter systems (Horikoshi, 1999). Moreover, the pH gradient across the cytoplasmic membrane reduces the total proton motive force yet required to synthesize ATP by conventional mechanisms (Preiss et al., 2015). They may also use other metabolic processes such as substrate-level phosphorylation gained by fermentation to obtain ATP, although the latter process has a lower ATP yield.

In addition to these challenges, high pH values decrease the solubility of certain mineral nutrients that co-precipitate with byproducts of serpentinization, and then have low bioavailability to the cells. The carbon source is often a limiting resource in such environments. Indeed, under alkaline conditions, DIC is essentially found in the form of CO_3^{2-} which precipitates into mineral calcium carbonate ($Ca CO_3$) (Schrenk et al., 2013). Calcium carbonate can only be assimilated by microorganisms that have the peculiar ability to use it as a carbon source (e.g., *Serpentinomonas*, (Suzuki et al., 2014)). Furthermore, under alkaline conditions, phosphorus co-precipitates with brucite (Bradley et al., 2009). This reaction is so effective that it is a standard protocol for quantitative removal of phosphorus in seawater, allowing the determination of nanomolar concentration (Karl and Tien, 1992).

Influence of high pressure and temperature

Depending on the location of hydrothermal ecosystems, habitats provide additional challenges for microbial life. For example, the microorganisms inhabiting deep-sea environments (such as LCHF, SSF, etc.) need to be adapted to high hydrostatic pressure, as it inhibits cellular processes (such as motility, substrate transport, cell division) in non-piezophilic microorganisms (Marietou and Bartlett, 2014). This adaptation to hydrostatic pressure is an interesting feature that helps to remove microbial contaminants by using enrichment cultures to select indigenous microorganisms and isolate novel strains (Quéméneur et al., submitted).

Although the temperature of the fluids at the outlet of serpentinizing ecosystems is often moderate (<50°C) and much lower than in magmatic-hydrothermal systems, some hydrothermal fluids harbor (hyper)thermophilic phylotypes (Brazelton et al., 2006, Postec et al., 2015). The presence of such microorganisms may reflect the existence of a hot subsurface habitat in the shallow mantle rocks beneath the hydrothermal fields. Interestingly, sequences closely related to cultivated members of the hyperthermophilic order *Thermococcales* and to environmental phylotypes associated with hot biotopes, such as the SAGMEG (South African Gold Mine Euryarchaeota Group, recently renamed *Hadesarchaea* (Baker et al., 2016)) were detected in PHF (Postec et al., 2015). High growth temperatures were predicted for these sequences, despite the mild temperature of the fluids venting at PHF (max 42°C). This result suggested that these archaea came from the hot subsurface environments beneath the PHF carbonate chimneys.

Scarce electron acceptors

Energy supplied by serpentinization in the form of electron donor (i.e., H_2 and CH_4), is usable by microbial cells only if electron acceptors are bioavailable in the environment to drive respiratory metabolisms. However, the redox potential (E_h in Tables 1.1, 1.2 and 1.3) is low (even negative) in the serpentinizing systems. Indeed, serpentinizing environments are characterized by limited availability of electron acceptors such as dissolved oxygen, nitrate and sulfate. Thus, the mixing zones between the reduced hydrothermal fluids and the oxygenated seawater or atmosphere provide inputs of potential electron acceptors, allowing a diverse bacterial community to flourish (Brazelton et al., 2006).

Microbial habitats, geochemical gradients and fluctuating conditions

Serpentinizing environments, like many hydrothermal systems, exhibit steep physical and chemical gradients, essential for the development of diverse microbial communities as explained above, but which may also represent challenging conditions for individual cells. Even if serpentinizing fluids are anoxic and reduced, they are discharged in oxic environments such as oceanic settings. For example, in the porous chimneys of LCHF, the microorganisms are affected by the hydrothermal fluids but can also be more or less in contact with oxygenated seawater. Therefore, multiple microorganisms with similar metabolic requirements may occupy slightly different microhabitats along geochemical gradients (McCollom and Seewald, 2013). The microorganisms living in the mixing zone are more likely to be a combination of microbes coming from the anoxic and hot subsurface as well as microbes coming from the discharge media which have been able to take advantage of the by-products of serpentinization and adapt to harsh conditions such as a high pH. Finally, the serpentinizing systems are subject to fluctuating conditions depending on the hydrothermal regime. For example, in terrestrial systems, hydrological processes (water infiltration and flow) depend on rainfall. These variations lead to frequent changes in the volumes of compounds generated by reaction of serpentinization

(Sánchez-Murillo et al., 2014) and do not provide a constant source of energy for microbes.

1.3.3 Microbial diversity in serpentinizing ecosystems

Most abundant Bacteria

The most represented phylum in the serpentinizing environments is the *Proteobacteria*, being consistently reported in all serpentinizing studied sites. *Betaproteobacteria* are particularly abundant in most continental sites: at The Cedars, Leka, Tablelands and Santa Elena (Brazelton et al., 2012, Crespo-Medina et al., 2017, Daae et al., 2013, Suzuki et al., 2013), while *Gammaproteobacteria* seem to prevail in marine sites. *Epsilonproteobacteria* are present only at LCHF and in deep-sea serpentinizing sites influenced by magmatic activities (Brazelton et al., 2010a, Flores et al., 2011). Their presence is most likely linked to the abundance of sulfur compounds in these environments, typically used for energetic metabolism by members of this class of bacteria (Kelley et al., 2001, Konn et al., 2009). Finally, *Deltaproteobacteria* members were detected at PHF, potentially ensuring the dissimilatory sulfate reduction (Quéméneur et al., 2014), and in a few other sites such as The Cedars, CROMO, Chimaera and Zambales ophiolites and LCHF (in varying proportions).

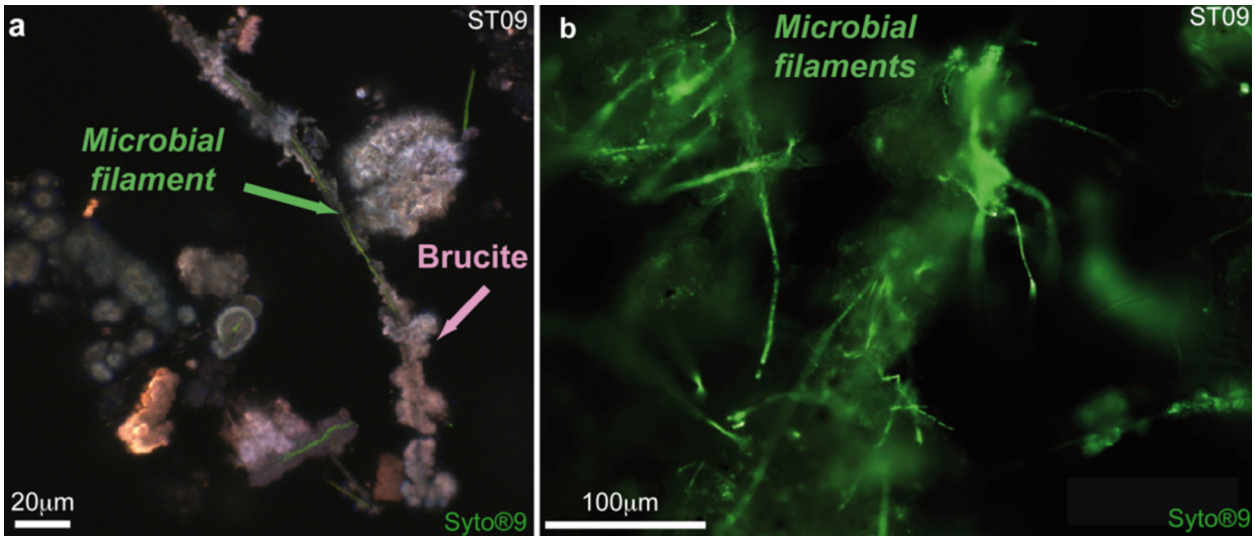


Figure 1.8 – Maximum intensity projection of stacked Confocal Laser Scanning Microscopy images, showing Syto®9-stained cells being mineralized by brucite in juvenile chimney, from Pisapia et al. (2017)

The *Firmicutes* phylum and especially the *Clostridiales* are also commonly detected in serpentinizing systems. These *Clostridiales* include diverse fermentative H_2-producing bacteria (Brazelton et al., 2012, 2016, Mei et al., 2016b). In addition to the few strains isolated from PHF sites, many of the *Firmicutes* detected in serpentinizing environments represent novel uncultivated lineage (probably at least at the family rank). Their metabolisms are therefore only speculative. One phylotype, was found dominating in the conduits of juvenile chimney in PHF (ST09), but was not the most abundant in mature chimneys from the same site (Pisapia et al., 2017). The phylotype and other filamentous microorganisms were mineralized by brucite form-

ing networks that promote the early stages of chimney construction (Figure 1.8). These uncultivated *Firmicutes* were thus proposed as pioneering colonizers of the chimneys most likely carried by hydrothermal fluids from the subseafloor (Pisapia et al., 2017).

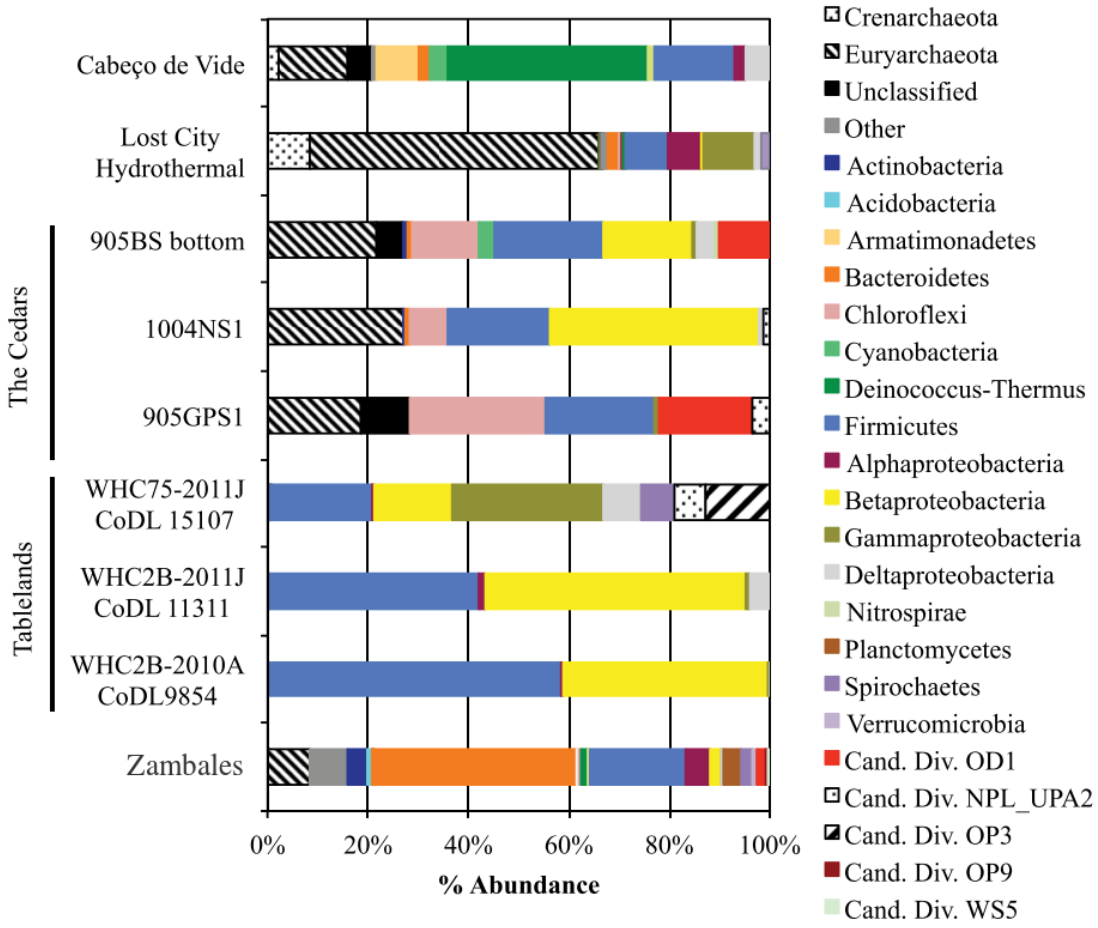


Figure 1.9 – Comparison of microbial communities from five serpentinizing systems, classified at the phylum level (or class in the *Proteobacteria*), modified from Woycheese et al. (2015).

Besides *Proteobacteria* and *Firmicutes* typically detected in serpentinizing systems, other phyla are more site-specific, as it can be seen on the Figure 1.9. In their study, Woycheese and colleagues compared the microbial communities from several serpentinizing systems, based on taxonomic affiliations of 16S rRNA gene sequences (Woycheese et al., 2015). Thus, in the Zambales ophiolite, the *Bacteroidetes* represent more than 40% of the total microbial community. In CVA, such a proportion was affiliated to the *Deinococcus-Thermus*. Finally, *Chloroflexi* were showed abundant in the Cedars. This phylum is typically found in many other serpentinizing ecosystems, such as Voltri (Quéméneur et al., 2015) Chimaera ophiolite (Neubeck et al., 2017), LCHF (Brazelton et al., 2006) and PHF (Postec et al., 2015, Quéméneur et al., 2014). Since the sequences detected correspond to phylotypes distantly related to the few cultivated representatives (mostly known as degraders of halogenate compounds (Biderre-Petit et al., 2016, Kaster et al., 2014)), their actual role in these ecosystems is unknown.

Most abundant Archaea

One striking feature of the microbial communities of serpentinizing systems is the very low diversity of the Archaea. Indeed, the Archaea are composed of a few dominant phylotypes in most of serpentinizing sites (Quéméneur et al., 2014, 2015, Schrenk et al., 2004, Tiago and Veríssimo, 2013), while, in other sites, the Archaea seem to be absent (Twing et al., 2017).

In 2004, Schrenk and collaborators reported the first analysis of the microbial communities inhabiting LCHF chimneys (Schrenk et al., 2004). The low archaeal diversity detected within the active carbonate structures was dominated by a single archaeal phylotype according to 16S rRNA gene sequences. This phylotype was related to the *Methanosarcinales* therefore termed Lost City Methanosarcinales (LCMS). The authors used FISH techniques to target this LCMS in carbonate samples. With this approach, the authors identified the morphology of LCMS phylotype (small cocci, 1-3 µm), and they quantified its proportion in the cell populations (up to 32%) (Schrenk et al., 2004).

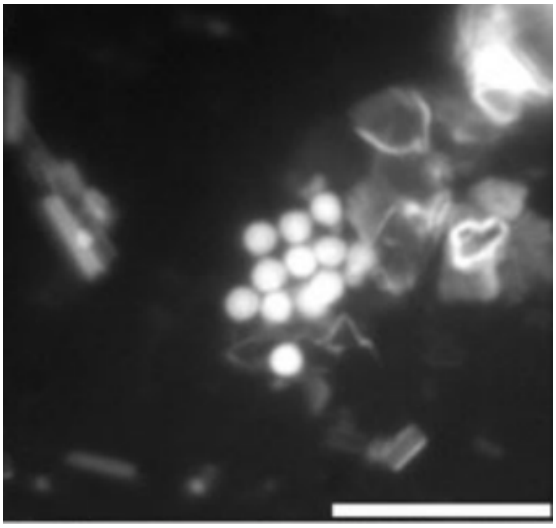


Figure 1.10 – Epifluorescence microscopy of LCMS860-hybridized cells detected in the LCHF carbonates (sample LC1022) from Schrenk et al. (2004). Scale bar is 20 µm.

Furthermore, these abundant LCMS are replaced by distinct phylotypes (Anaerobic Methanotrophs of group 1, ANME-1) in the chimneys where hydrothermal activity ceased (Brazelton et al., 2010a). This observation can suggest the LCMS were more adapted to the conditions of serpentinization (by-products and hydrothermal circulations) than other microorganisms. A few years later, in 2013, the only archaeal phylotype detected in the terrestrial springs of the Cedars was also related to the order *Methanosarcinales*, although quite distant from any clade of known methanogens and methanotrophs or LCMS phylotypes (Suzuki et al., 2013). Uncultivated *Methanosarcinales* phylotypes detected in PHF were similar to those found in LCHF and The Cedars ecosystems (Figure 1.11) (Quéméneur et al., 2014). This result reinforces the possible link between these phylotypes and serpentinizing systems.

The *Thaumarchaeota* phylum was also found dominant in several serpentinizing sites such as Leka Ophiolite (Daae et al., 2013), Santa Elena Ophiolite (Sánchez-Murillo et al., 2014) and Voltri Massif (Quéméneur et al., 2015). These archaea are potentially capable of oxidizing ammonium.

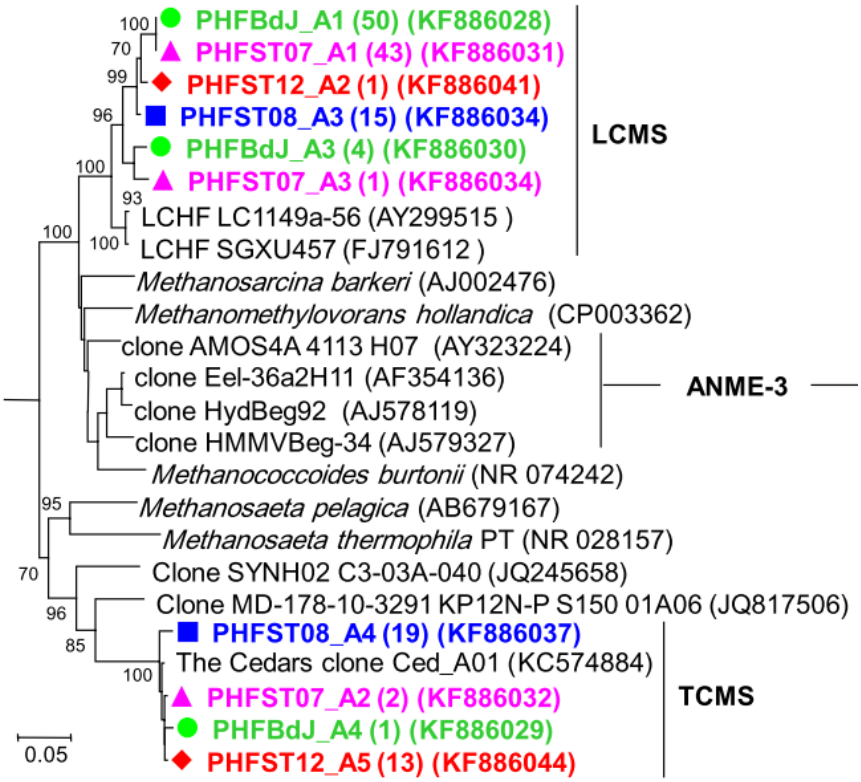


Figure 1.11 – Neighbour-joining phylogenetic tree of Methanosarcinales 16S rRNA gene sequences from four PHF chimneys from Quéméneur et al. (2014). Clusters are abbreviated as follows: LCMS, Lost City Methanosarcinales; ANME-3, anaerobic methanotrophs of group 3; TCMS, The Cedars Methanosarcinales

Candidate phyla

Parcubacteria members (previously known as OD1, (Rinke et al., 2013)) were found in high abundance in the deep groundwater sources at The Cedars (Suzuki et al., 2013), in some springs at Voltri (Brazelton et al., 2013) and at PHF (Postec et al., 2015) but also, to a lesser extent, in the hydrothermal chimneys of LCHF (Gerasimchuk et al., 2010). Precaution should be taken regarding the interpretation of *Parcubacteria* abundance in these environments. Indeed, Suzuki and colleagues reported that the abundance of *Parcubacteria* members (via 16S rRNA gene-based survey) could be biased due to introns in the 16S rRNA gene and potential risk of biases from PCR (Suzuki et al., 2017). New findings have been made on the *Parcubacteria* through draft genome, reconstructed from metagenomes of The Cedars springs (Suzuki et al., 2017). The draft genomes of these *Parcubacteria* were very small, often with no ATP-synthase and lacking one or more key biosynthetic pathways. For these reasons, the authors suggested that these *Parcubacteria* are incapable of independent growth and may be either intercellular symbiont of other microbes or scavengers of dead cells. The possible symbiotic lifestyle of *Parcubacteria* (recovered from oxic groundwater) was previously pointed by Nelson and Stegen (2015).

Other microorganisms affiliated to candidate phyla were detected in serpentinizing sites, but even less information is available since none could be isolated or sequenced. Only partial sequences of their 16S rRNA gene are known. Among the most commonly abundant taxa, one

can mention the phyla *Acetothermia* (previously known as OP1) and *Omnitrophica* (OP3) in PHF (Pisapia et al., 2017, Postec et al., 2015) and Semail Ophiolite (Rempfert et al., 2017). It would be essential to learn more about these abundant microorganisms to better understand the serpentinizing ecosystems.

1.3.4 Endemic or cosmopolitan populations?

Biogeography

Based on 16S rRNA sequence similarity, closely related microorganisms were detected in distant serpentinizing sites (e.g., Mid-Atlantic Ridge, California, New Caledonia). The most striking example concerns the only two *Methanosarcinales* phylotypes that surprisingly were found overdominant (up to 90% of the archaea) in LCHF (the Lost City Methanosarcinales, LCMS) and at The Cedars (TCMS). Both phylotypes were identified in PHF (Postec et al., 2015, Quéméneur et al., 2014). Recently, these LCMS were also found abundant in a newly explored deep-sea site: the Old City hydrothermal field in the South West Indian Ridge, which chimneys and geochemichal context resemble LCHF (Lecoeuvre *et al.*, Goldschmidt2018). Other similar cases have been reported for many microorganisms, such as members of the *Truepera*, *Silanimonas* or *Hydrogenophaga* genera, and phylotypes affiliated to the candidate division SRB2 (*Clostridia*) that were detected at least into two different serpentinizing sites (Quéméneur et al., 2015, Rizoulis et al., 2014, Suzuki et al., 2013, Twing et al., 2017). The detection of closely-related phylotypes in remote locations raises the issue of the biogeography of these microorganisms. Indeed, their presence can be related to either a potential common subsurface origin or to a selection of ubiquitous species (commonly detected in natural environments) by the conditions of serpentinization. To date, these issues have received little attention in serpentinizing systems.

Rare biosphere

Most of the microbial ecosystems are dominated by a few species, while a large number of species are represented by only a few individuals. The detection of wide variety of sequences at low abundance, termed "rare biosphere" was first discovered in water column and sediment samples, using Next Generation Sequencing (Sogin et al., 2006). Brazelton and colleagues were the first to test the validity of this new concept, applied to the LCHF ecosystem (Brazelton et al., 2010a). In particular, the authors sought to determine whether a shift in species composition (rare populations becoming abundant) may be observed after a long period of time. For that purpose, they sequenced archaeal and bacterial communities inhabiting carbonate chimneys of different ages over a 1,200-year period. Thus, they detected ANME-1 sequences which were rare in young chimneys samples, while they dominated the older samples. This result suggested that these ANME-1 organisms may remain rare during the period of hydrothermal activity of chimneys before proliferating in old chimneys that do not vent any more high-temperature and alkaline fluids (Brazelton et al., 2010a).

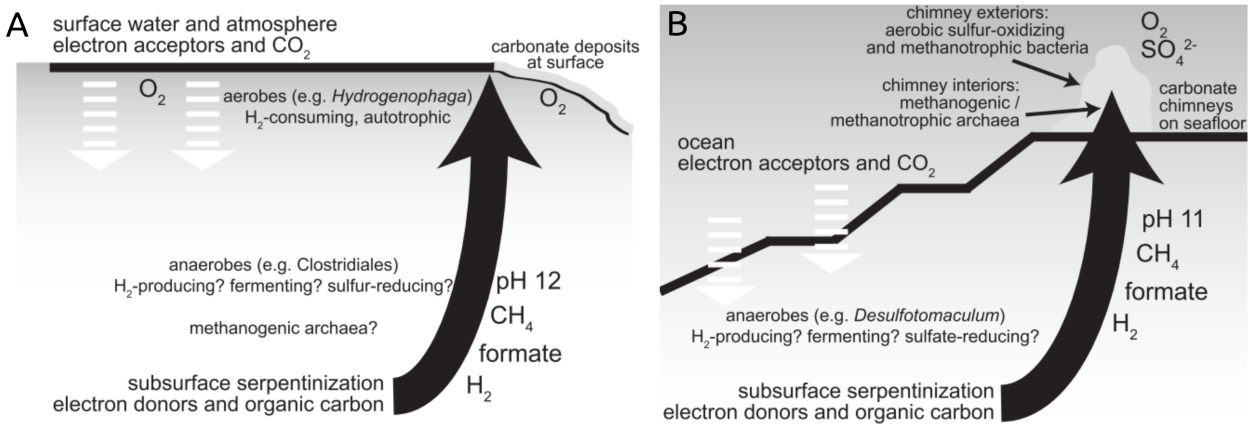


Figure 1.12 – Schematic of putative biogeochemical processes in continental (A) and submarine (B) serpentinizing environments, from Schrenk et al. (2013)

1.4 Microbial metabolisms in serpentinite-hosted systems

The serpentinization-associated reactions impose a unique set of conditions (i.e., a high concentration of H_2, CH_4 and abiotic organic compounds) upon the microorganisms and their metabolisms in serpentinizing ecosystems (Figure 1.12). The main metabolisms potentially used for chemosynthetic microbes in serpentinizing ecosystems are presented in the Table 1.3. Some of these metabolisms may occur only if a circulation of fluids exists that exposes microorganisms to electron acceptors present in the atmosphere or in seawater (McCollom and Seewald, 2013). The thermodynamic modelling provides predictions of metabolisms that are most likely to be active given the energy yields under different settings. More abundant available energies were detected in ultramafic-influenced hydrothermal systems (such as LCHF) compared to the basalt-hosted EPR, Lucky Strike and TAG sites (Figure 1.13) (Reveillaud et al., 2016).

METABOLIC ENERGY SOURCE	OVERALL CHEMICAL REACTION
Hydrogen oxidation	$2H_2 + O_2 \rightarrow 2H_2O$
Methanotrophy	$CH_4 + 2O_2 \rightarrow HCO_3^- + H^+ + H_2O$
Methanogenesis	$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$
Anaerobic methane oxidation (AMO)	$SO_4^{2-} + 2H^+ + CH_4 \rightarrow CO_2 + H_2S + 2H_2O$
Sulfate reduction	$SO_4^{2-} + 2H^+ + 4H_2 \rightarrow H_2S + 4H_2O$
Aerobic sulfide oxidation	$H_2S + 2O_2 \rightarrow SO_4^{2-} + 2H^+$
Aerobic iron oxidation	$4Fe^{2+} + 2O_2 + 10H_2O \rightarrow 4FeOH_3 + 8H^+$
Sulfide supported denitrification	$8NO_3^- + 5H_2S \rightarrow 5SO_4^{2-} + 4N_2 + 2H^+ + 4H_2O$
Hydrogen supported denitrification	$2NO_3^- + 5H_2 + 2H^+ \rightarrow 2N_2 + 6H_2O$

Table 1.4 – Primary sources of chemical energy for chemosynthetic microorganisms in hydrothermal environments, modified from McCollom and Seewald (2013).

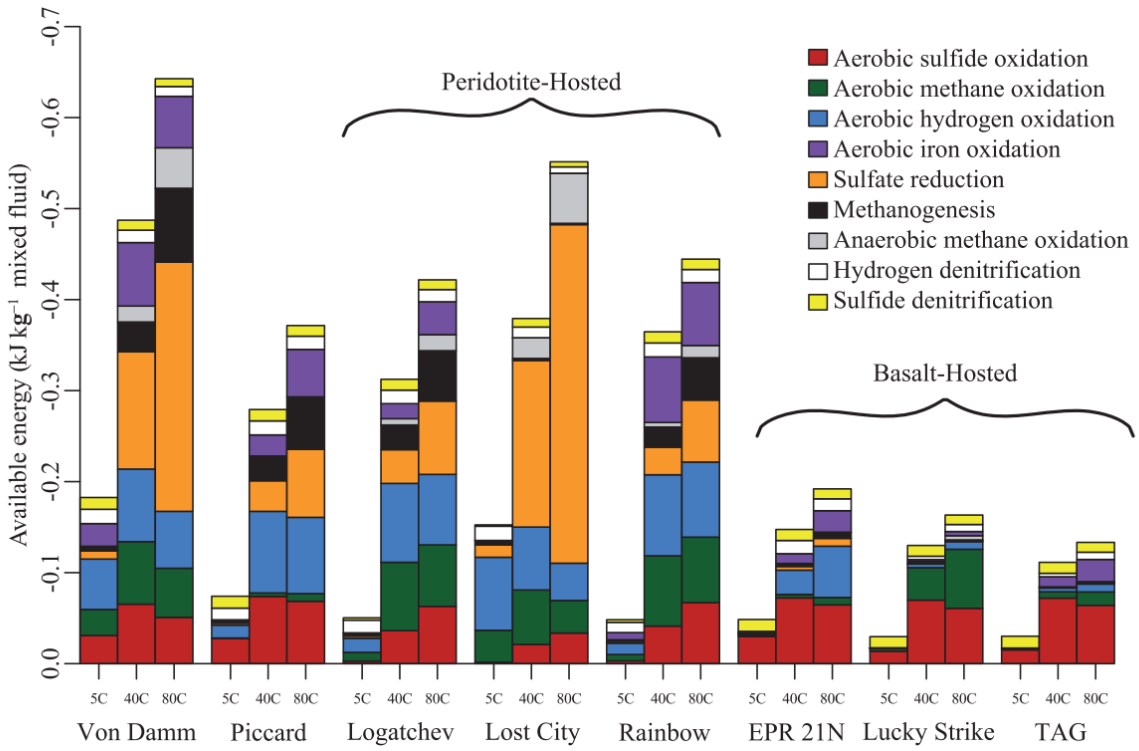


Figure 1.13 – Catabolic energies available from key redox reactions in eight hydrothermal systems at 250 bar and at three different temperatures: 5°C, 40°C and 80°C, from Reveillaud et al. (2016)

1.4.1 Hydrogen cycle and hydrogenases

Hydrogenases catalyze the reversible oxidation of hydrogen gas. The reaction 1.5 is catalyzed by two main classes of hydrogenases: the [NiFe]-hydrogenases and [FeFe]-hydrogenases. Despite their similarity, these two classes are phylogenetically distinct and differ in the composition of their active site metal clusters. While [NiFe]-hydrogenases tend to be involved in H_2 consumption, [FeFe]-hydrogenases are typically involved in H_2 production (Vignais and Colbeau, 2004). However, numerous exceptions are known, such as a periplasmic [FeFe]-hydrogenase of *D. vulgaris* demonstrated to consume H_2 (Pohorelic et al., 2002) or the electron-bifurcating [FeFe]-hydrogenase from *M. thermoacetica* that can both produce and oxidize H_2 (Wang et al., 2013).

$$H_2 \leftrightarrow 2H^+ + 2e^- \tag{1.5}$$

The class of [NiFe]-hydrogenases is widely distributed in both Bacteria and Archaea domains (Figure 1.14B). Absent in archaeal representatives, [FeFe]-hydrogenases are found in anaerobic prokaryotes, such as *Clostridia* and sulfate reducers, but also in some anaerobic eukaryotes such as fungi, ciliates or some green algae (Vignais and Billoud, 2007, and ref herein). There is a great diversity of hydrogenases, with differences in affinity for H_2, sensitivity to O_2, cell location (membrane-bound or cytoplasmic) and the number of subunits composing them, even in a same class.

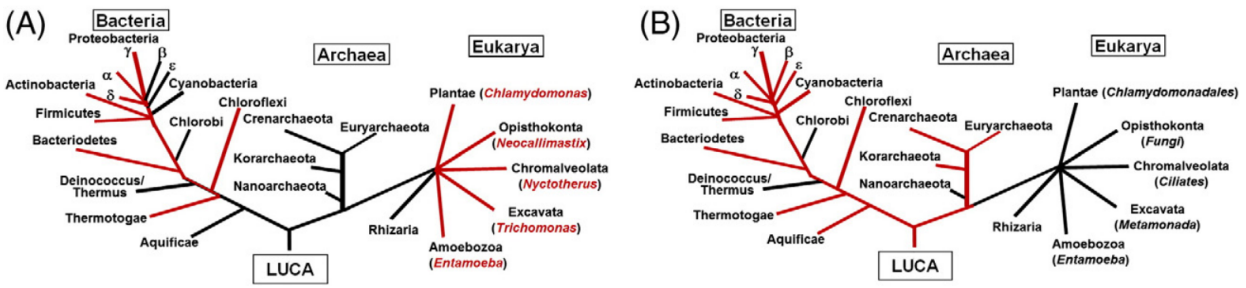


Figure 1.14 – Distribution of (A) [FeFe]-hydrogenases and (B) and [NiFe]-hydrogenases, mapped on a taxonomic tree of life, from Peters et al. (2015).

The genetic potential for H_2 oxidation and production were investigated in many microbial communities from serpentinite-hosted systems. The pioneering study that investigated hydrogen metabolism in a serpentinizing system was a metagenomic study of Brazelton et al. (2012). A high enrichment in [NiFe]-hydrogenases were detected in Lost-City and the Tablelands metagenomes. These hydrogenases were primarily encoded by *Betaproteobacteria* and matched against 3 of the 4 groups of characterized [NiFe]-hydrogenases (Vignais, 2001). The main taxonomic affiliations of theses hydrogenases indicated close relationships with the *Hydrogenophaga* and *Ralstonia* genera. These organisms are aerobic or facultative anaerobic and may inhabit systems with strong oxygen gradient between hydrothermal fluids and oxygenated seawater or atmosphere. The strains belonging to the recently proposed genus "*Serpentinomonas*" by Suzuki et al. (2014), isolated from The Cedars environments, contained at least two [NiFe]-hydrogenases. Microbes related to these strains (with more than 99% identity in 16S rRNA gene sequences) were detected as dominant in many terrestrial serpentinizing sites, such as CVA, The Tablelands, Outokumpu ophiolite (Suzuki et al., 2014) and CROMO (Twing et al., 2017). However, they are not specific of serpentinizing environments as they were also detected in other alkaline environments such as industrial waste and deep groundwater (Suzuki et al., 2014). Moreover, they were not yet reported in ocean settings. Numerous [FeFe]-hydrogenases were mainly affiliated to *Clostridiales* in most serpentinizing sites (Brazelton et al., 2012, Mei et al., 2016b, Twing et al., 2017) but also recently detected in *Bacteroidetes* in Voltri Massif (Brazelton et al., 2016) and in terrestrial sites of PHF (Mei et al., 2016b). A high diversity of [FeFe]-hydrogenase genes was detected in PHF (Mei et al., 2016b). The authors suggested that reflected either numerous fermenting bacteria or individual metabolic flexibility of microorganisms adapted to environmental fluctuations. Overall, the serpentinizing systems harbor microbial community associated with both oxidation and production of hydrogen, due to high frequency of *Betaproteobacteria* and *Firmicutes* in these ecosystems.

1.4.2 Methanogenesis and methanotrophy

The high concentration of hydrogen and reducing conditions provided in the hydrothermal fluids make hydrogenotrophic methanogenesis a thermodynamically favorable pathway in many serpentinizing systems (Amend et al., 2011, Schrenk et al., 2013). The *mcrA* gene, encoding the

alpha subunit of methyl-coenzymeM reductase, is used by all known orders of methanogens to catalyze the last step of methanogenesis (Lloyd et al., 2006). This gene provides a good phylogenetic marker to study methanogen populations (Luton et al., 2002). However, its presence is not a definitive evidence of methanogenesis since *mcr* genes are also identified in archaea, which use methyl-coenzymeM reductase to catalyze Anaerobic Methane Oxidation (AMO), the so-called anaerobic methane oxidizers (ANME) (Lloyd et al., 2006). Amplicon sequencing of the *mcrA* gene permitted the detection of phylotypes affiliated to methanogens and/or ANME in most serpentinizing systems: e.g., in LCHF (Kelley et al., 2005), Rainbow, Ashadze (Rousel et al., 2011), Del Puerto (Blank et al., 2009), in CVA (Tiago and Veríssimo, 2013) and Prony (Quéméneur et al., 2014). These methanogens mainly belonged to *Methanosarcinales*, ANME-1 and ANME-3. In addition, both methanogenesis and methanotrophy were detected within a biofilm of LCMS phylotypes using isotopic ratios (Brazelton et al., 2011).

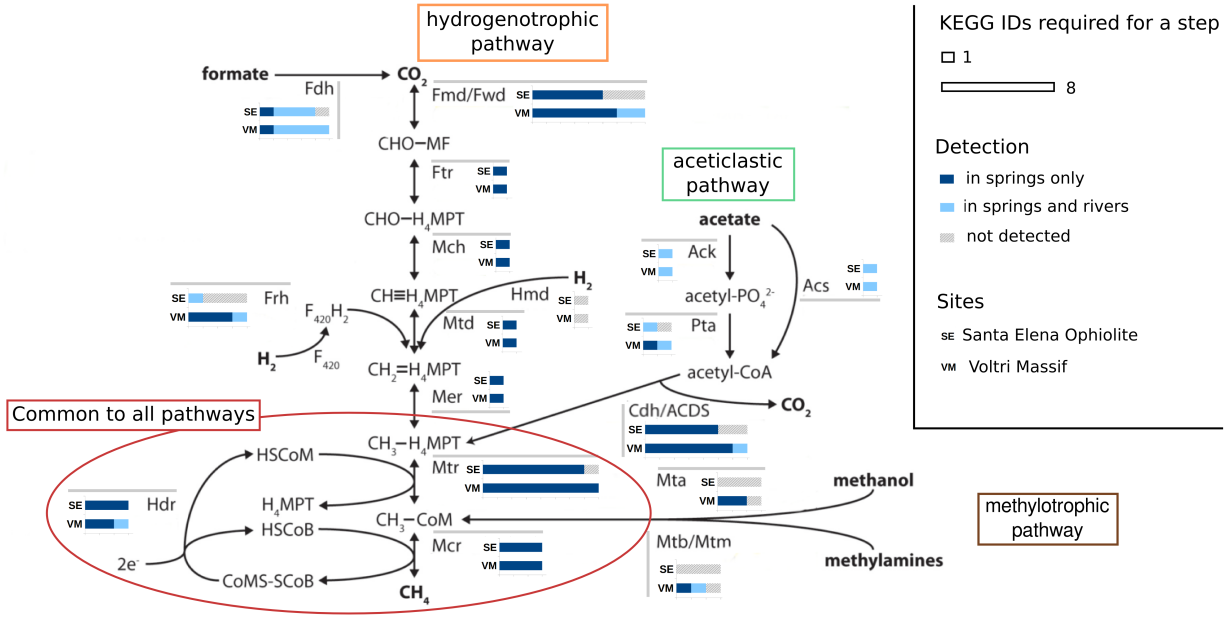


Figure 1.15 – Diagram of methanogenesis pathways from CO_2, formate, acetate, methanol, and methylamines with associated protein homologs. The KEGG IDs associated with each step are counted in Voltri Massif (Italy) and Santa Elena (Costa Rica). Modified from Brazelton et al. (2016) and Crespo-Medina et al. (2017).

The complete methanogenesis pathways were investigated (Figure 1.15) in two continental serpentinizing sites: the Voltri Massif (Brazelton et al., 2016) and the Santa Elena Ophiolite (Crespo-Medina et al., 2017). In both sites, most of the key genes involved in the hydrogenotrophic methanogenesis are specific to springs only. The lack of the gene *hmd* (coding for H_2-forming methylenetetrahydromethanopterin dehydrogenase) is not surprising because this gene is not very widespread among methanogens (Brazelton et al., 2016). The methylotrophic methanogenesis was only detected in the metagenomes from Voltri. Finally, the genes associated with methanogenesis using formate or acetate as substrates are identified in springs and adjacent rivers. Moreover, the presence of formate dehydrogenase (Fdh), acetate kinase (Ack), phosphotransacetylase (Pta), AMP-forming acetyl-CoA synthetase (Acs) and carbon monox-

ide dehydrogenase/ acetyl-CoA decarbonylase synthase (Cdh/ACDS) is more difficult to discuss because these proteins can be found in non-methanogens as well.

Based on metagenomic evidences (a lack of *fdhC* sequences, thought to be necessary for transport of formate into the cell), it is unlikely that in LCHF, methanogenic archaea could directly use formate as a source of carbon (Lang et al., 2018). For this reason and given the low availability of DIC in these environments, the authors suggested that the methanogens live in a syntrophic relationship with microorganisms which metabolism releases CO_2 (directly assimilable by the methanogens). The same metagenomics dataset revealed the presense of bacterial *fdhA* (encoding the key enzyme, formate dehydrogenase), having high similarity with alkaliphilic sulfate-reducing bacteria (SRB) such as *Desulfitibacter alkalitolerans* and *Desulfotomaculum alkaliphilum*. These SRB could thus be the primary producers using formate abiotically formed by serpentinization reaction to convert it in CO_2 and thus be the methanogens partners.

The aerobic methanotrophy metabolism was essentially investigated through marker genes encoding methane monooxygenase (*pmoABC* and *mmoXYBZDC*). In the springs of the Voltri Massif, the aerobic oxidation of methane detected was mainly associated with *Methylococcaceae* family (Brazelton et al., 2016). Considering their spatial distribution (i.e., their absence in the completely surface-exposed habitats), the authors suggested that these microorganisms inhabit mixing zones with low levels of oxygen from the surface and methane from subsurface water.

1.4.3 Autotrophy/Heterotrophy

Carbon fixation

The autotrophs can sustain the primary production in the ecosystems. They use the carbon fixation to convert inorganic carbon (typically CO_2) into organic compounds. The six autotrophic carbon fixations known to date are described below (Berg et al., 2010).

(i) Calvin cycle: This cycle represents the most frequent carbon fixation pathway in nature. The CO_2 reacts with a sugar to produce 3-phosphoglycerate. The key enzymes of this fixation pathway are the ribulose 1,5-bisphosphate carboxylase-oxygenase (RubisCO, encoded by *rbcL* and *rbcS*) and phosphoribulokinase (*prkB*). Furthermore, several studies proposed the Calvin cycle as a late innovation (references in Berg et al. (2010)).

(ii) Reverse tricarboxylic acid cycle (rTCA): This cycle requires much less energy than the Calvin cycle. However, it involves O_2-sensitive enzymes and is therefore found only in anaerobes or in aerobes that live in environments with low oxygen concentration. The key enzymes are the 2-Oxoglutarate ferredoxin oxidoreductase (*korABDG*) and ATP-citrate lyase (*aclAB*).

(iii) Reductive acetyl-coenzyme A pathway: This pathway, also named Wood-Ljungdahl pathway, is the most energetically favorable autotrophic fixation pathway, but it requires the use of coenzymes (tetrahydropterin, cobalamin, etc.) and metals (Fe, Co, Ni, Mo, or W).

Furthermore, this pathway is restricted to strictly anaerobic organisms, since one of its key enzymes, the acetyl-CoA synthase/carbon-monoxide dehydrogenase (*acsB*, *acsA*), is among the most oxygen-sensitive enzymes known.

(iv) 3-hydroxypropionate bicycle: This pathway contains two cycles. The first cycle results in the fixation of bicarbonate and formation of glyoxylate. In the second cycle, glyoxylate and propionyl-CoA are used to produce pyruvate and acetyl-CoA. Until now, this bicycle has appeared to be restricted to *Chloroflexaceae*, a family of green non-sulfur bacteria, although its presence is assumed in some proteobacteria (Hügler and Sievert, 2011). The main key enzymes are malonyl-CoA reductase (*mcr*), acetyl-CoA/propionyl-CoA carboxylase and malyl-CoA lyase (*mcl*).

(v) 3-hydroxypropionate/4-hydroxybutyrate cycle: This cycle was reported to operate in thermoacidophilic members of *Crenarchaea*. In the first part of the pathway, acetyl-CoA and two bicarbonate molecules are transformed in succinyl-CoA, while in the second part the succinyl-CoA is converted into two molecules of acetyl-CoA. The key enzymes of this cycle are the bifunctional acetyl-CoA/propionyl-CoA carboxylase and the 4-hydroxybutyryl-CoA dehydratase (*abfD*).

(vi) Dicarboxylate/4-hydroxybutyrate cycle: Recently detected in anaerobic archaeal orders of the *Thermoproteales* and the *Desulfurococcales*, this metabolic pathway combines several stages of the cycles discussed above (rTCA and 3-hydroxypropionate/4-hydroxybutyrate cycle). As there are no enzymes that are unique to this pathway, the measurement of a combination of enzymes is required to determine if this cycle is operating in a given microorganism (Hügler and Sievert, 2011).

The low concentration of dissolved inorganic carbon (DIC) is the primary challenge for autotrophic microorganisms living in serpentinizing systems. Several studies highlight the importance of the Calvin cycle in many serpentinizing systems. Brazelton and Baross hypothesized that the *Thiomicrospira* organisms from LCHF were adapted to low CO_2 environments (Brazelton and Baross, 2010). They identified the genes for a putative Type-I RubisCO as well as genes encoding for carboxysome shell proteins, adjacent to RubisCO genes. The carboxysomes form microcompartments in which CO_2 is concentrated to optimize carbon fixation (Brazelton and Baross, 2010). Moreover, the comparative genomic study of different *Thiomicrospira* lineages also showed a possible adaptation to low CO_2 concentration. Indeed, the LCHF lineage lacks the RubisCO form II, expressed at high CO_2 concentration, and detected in *Thiomicrospira crunogena* XCL-2 (Brazelton and Baross, 2010). Suzuki and collaborators also searched for genes involved in carbon fixation in the three *Serpentinomonas* strains (Suzuki et al., 2014). They conclude that carbon dioxide fixation is occurring via the Calvin cycle, since only the genes encoding enzymes of the Calvin cycle were detected in all strains. Finally, metagenomic analyses of the highly alkaline springs of the Tablelands also identified genes for a Type I RuBisCO, associated with facultatively anaerobic *Betaproteobacteria* (Brazelton et al., 2012).

Evidences for other carbon fixation pathways are rather scarce in the serpentinizing sys-

tems. The presence of acetyl-CoA synthase (ACS) was detected in the microbiomes of LCHF and the Tablelands, suggesting the potential for carbon fixation via the Wood-Ljungdahl pathway (Brazelton et al., 2012). These ACS homologs are related to *Clostridia* in the Tablelands springs and related to *Methanosarcinales* in LCHF. Moreover, a metagenomic study of microbial community from Von Damm hydrothermal field (Mid-Cayman Rise) showed a high abundance of key genes involved in carbon fixation via the rTCA cycle (Reveillaud et al., 2016).

In the end, it appears that despite the severe limitation in the carbon source, several metabolic strategies exist or co-exist in serpentinizing ecosystems for carbon fixation, carried out by diverse actors, depending on local environmental conditions and biogeography.

Heterotrophy and fermentation

Although serpentinizing fluids have very low concentration of DIC, they contain a variety of organic molecules, such as small hydrocarbons, formate or acetate (Schrenk et al., 2013), which can serve as carbon source for the microbial communities. Many geochemical studies have aimed to quantify these organic molecules, their origin and formation conditions, using isotopic analyses (e.g., Delacour et al., 2008, Lang et al., 2012). The primary production by chimiolithoautotrophs can also be an important source of organic carbon (Schrenk et al., 2013). The biotic organic molecules can feed heterotrophs growing in the close vicinity of the primary producers (biofilms) or dissolved in the fluid, they can sustain distant heterotrophs populations. For example, significant concentrations of acetate (in the μM range) that have been shown to be mainly of biological origin have often been reported in the fluids of several serpentine systems (Lang et al., 2010). This acetate can feed various heterotrophs (including acetoclastic methanogens). Another source or organic carbon for microbial communities is the geological recycling of organic matter.

For heterotrophs, there are two energy-producing alternatives: the respiration and the fermentation. In fermentation, the energy derives from partial oxidation of an organic compound and no outside electron acceptors or electron transportation systems are required. Therefore, the fermentation may be an attractive metabolic strategy in serpentinizing environments (Schrenk et al., 2013). Many alkaliphilic heterotrophs have been detected in serpentinizing sites with culture and isolation methods (Bath et al., 1987, Ben Aissa et al., 2015, Bes et al., 2015, Mei et al., 2016a, Tiago et al., 2004, 2006a, 2005b).

1.4.4 Other metabolisms

Sulfur cycling

The presence of sulfate and abundant hydrogen in serpentinizing systems offers ideal conditions for microbial sulfate reduction (Schrenk et al., 2013). The dissimilatory sulfate reduction was investigated in several serpentinizing sites using the *dsrAB* gene as a marker of sulfate reduction although this gene is also essential for some sulfites and sulfur reducers, being more

widely distributed than previously thought (Müller et al., 2015). At PHF, based on *dsrB* qPCR data, about 5% of the total bacteria were considered as sulfate reducers. Studies on 16S rRNA gene also demonstrated that they belonged to the *Delaproteobacteria* class and were essentially related to the genus *Desulfonatronum* (Postec et al., 2015). There are nine species in this genus, all but one of have been isolated in soda lakes. All are anaerobic, mesophilic and alkaliphilic, requiring sodium and carbonate for their growth (Pérez Bernal et al., 2017, and references therein). All of them use sulfate, sulfite and thiosulfate as terminal electron acceptors to be reduced to sulfide. Most of them are able to growth autotrophically on H_2 and CO_2. Another notable physiological property of Desulfonatronum species is their ability to disproportionate thiosulfate and sulfite to sulfate and sulfide. Recent investigations of alkaline environments (Poser et al., 2013) provided evidence that these oxido-reductive processes were also performed by non-SRB (e.g., *Desulfurivibrio* and *Dethiobacter*) suggesting that they may be of significatnt interest regarding the sulfur cycling in such extreme habitats. On the contrary, in LCHF, this metabolism was not detected as associated to members of the *Deltaproteobactria* but to the *Clostridia* class. Indeed, *dsrB* revealed phylotypes related to the genus *Desulfotomaculum*, which is usually associated with geothermal wells or subsurface and contains autotrophic hydrogenotrophic species (Gerasimchuk et al., 2010). In the Outokumpu deep borehole (Finland), a part of sulfate reducers was also found related to *Desulfotomaculum* sp. and *Desulfosporosinus* sp. (Itävaara et al., 2011).

The sulfur oxidation has only been poorly studied using the genetic markers: Sox proteins and adenosine 5'-phosphosulfate reductase (*aprA*) in metagenomes. The use of sulfur was studied in Lost City *Thiomicrospira* population (Brazelton and Baross, 2010). The authors identified homologs for all Sox genes required for sulfur-oxidation. However, the gene encoding for the Sox pathway does not seem to be organized into a single cluster, which is rare in prokaryotes. An adjacent transposase gene was thought to be involved in lateral transfer of sulfur oxidation genes among bacteria (Brazelton and Baross, 2010). However, these organisms are presumed to be abundant at oxic-anoxic interfaces within the chimney walls. Thus, their potential functions do not give any clues about the metabolisms of sulfur in the subsurface where serpentinization occurs. In CROMO, the detection of *aprA* linked the potential for sulfur oxidation to members of *Clostridia* (Twing et al., 2017).

Nitrogen cycling

The nitrogen cycling was only partly investigated in serpentinizing systems, looking for nitrogen fixation, nitrite/nitrate reduction and nitrite ammonification (Figure 1.16). Nitrogen metabolism was mainly investigated at two serpentinizing sites: the LCHF and the Chimaera ophiolite.

The *nifH* gene, encoding a subunit of a nitrogenase, was used as marker in screening for N_2 fixation in LCHF (Brazelton et al., 2011). Indeed, fixed nitrogen (such as nitrate and ammonium) have relatively low concentrations in the LCHF fluids, suggesting that microbial commu-

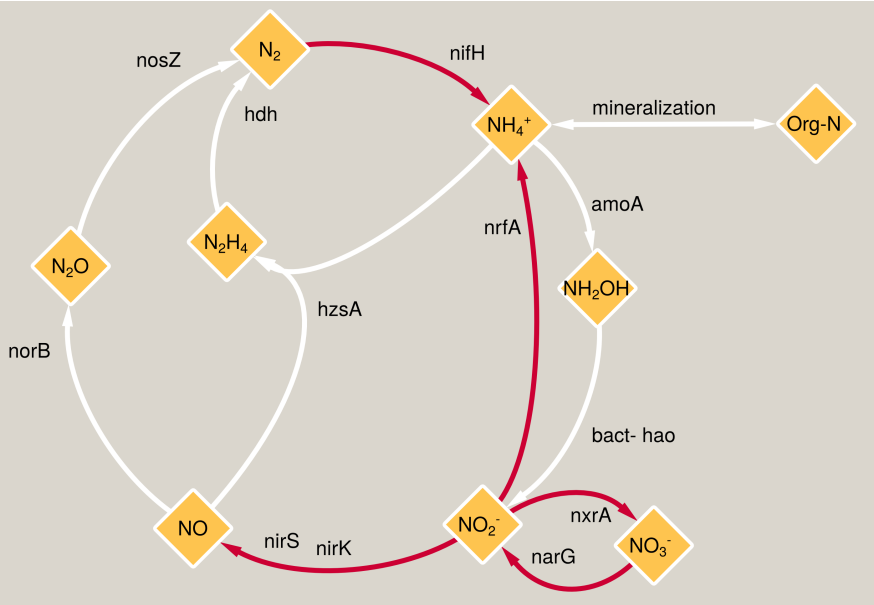


Figure 1.16 – Investigations on the potential of nitrogen cycling in serpentinizing systems. The steps of nitrogen cycling, detected in at least on serpentinizing ecosystems, are indicated by red arrows. Modified from Lüke et al. (2016).

nities use dissolved N_2 instead. In this study, *nifH* genes were mainly affiliated to methanogenic archaea, including the predominant LCMS phylotypes. Furthermore, the N_2 fixation may be more favorable in highly reducing conditions found within Lost City chimneys than in surface environments, since this metabolism is often inhibited by oxygen (Brazelton et al., 2011). In addition, one study reported low isotopic value ($\delta^{15}N$) of organic nitrogen in carbonate samples from LCHF, compatible with biological nitrogen fixation (Lang et al., 2013). In a comparative metagenomic study of microbial communities from deep-sea vents, Xie and collaborators also identified genes involved in nitrite ammonification (*nrfA*) and nitrite reduction (*nirK* and *nirS*) in LCHF chimneys, surprisingly the genes for key enzymes of the denitrification pathway, typically recovered in microbial communities from other basalt-hosted deep-sea vents (Xie et al., 2011).

On the contrary, in Chimaera ophiolite, the nitrogen fixation does not seem to be a dominant metabolism, according to isotopic analyses (Meyer-Dombard et al., 2015). Furthermore, the nitrogenase *nifH* gene has only been detected in two of the five samples and is not necessarily expressed in microbial populations. In this system, the primary process rather appears to be the nitrate/nitrite reduction (or even denitrification) since all samples included such potential capabilities (i.e., the genes *narG*, *nirS*). This result is consistent with the assumption that nitrogen is recycled into the ecosystem, based on isotopic values of nitrogen in biomass (Meyer-Dombard et al., 2015).

Phosphorus cycling

Phosphorus (P) is an essential nutrient owing to its pivotal role in all cellular processes, mainly as part of essential molecules, such as nucleic acids, nucleotidic co-factors and membrane

lipids, thus involved in cell structure, metabolism, genetic expression for all living organisms (White and Metcalf, 2007). The mineral brucite, often formed through the mixing of high pH serpentinite springs with carbonate bearing waters, is effective at scavenging P from solution (Bradley et al., 2009). Additionally, under phosphorus-limited conditions, microorganisms have been shown to substitute sulfur for phosphorus in their membrane lipids (Van Mooy et al., 2006). A similar phosphorus conservation strategy has been proposed for microbial populations inhabiting carbonate chimneys at the LCHF, where glycosyl head groups replace the classical phosphatadyl head groups in bacterial membrane lipids (Bradley et al., 2009).

Moreover, in environments with low inorganic phosphate (P_i, valence of +5) there is a considerable interest in understanding the utilization of other types of P biochemistry (Martinez et al., 2010). The discovery of compounds in which P is at a lower valence in nature and in a wide range of organisms, suggested their key role in the metabolism of P (White and Metcalf, 2007). For example, the dissolved organic phosphorus (DOP) of some marine environments contains abundant phosphonates (Pn) (Martinez et al., 2010). The phosphonate are reduced P compounds (valence of +3) that contain a stable C-P bond. Four distinct degradative pathways are present in various prokaryotic species: three hydrolase pathways (from phosphonopyruvate, phosphonoacetate and phosphonoacetaldehyde) and the carbon-phosphorus lyase (C-P lyase) pathway (White and Metcalf, 2007) to cleave the C-P bond. These degradation pathways are alternatives for microorganisms to obtain phosphorus. However, the microbial capacity to use reduced P compounds have not yet been investigated in any of the serpentinizing environments.

Chapter 2

Omics for environmental microbial studies

“ An expert is a person who has made all the mistakes that can be made in a very narrow field. ”

Niels Bohr

Sommaire

2.1 Brief history of microbial community study

Microbial communities are defined as multi-species assemblages, in which microorganisms coexist in the same space and time, and interact with each other (Konopka, 2009). Environmental microbiology seeks to understand how biological assemblages are structured, what their functional interactions are and how communities structure can change. As reviewed by Escobar-Zepeda et al. (2015), the study of microbial communities has evolved considerably from Leeuwenhoek's first report on the oral microbiota to current molecular techniques (Figure 2.1). From the beginning, the microscope has been an essential tool for studying microbial communities, associated with staining techniques (e.g., Gram, Ziehl-Neelsen). For nearly 300 years, microbiologists, such as Pasteur, Koch and Winogradsky, used culture and isolation to study microorganisms under controlled conditions. At the time, the classification of microorganisms was based on morphology features, growth, respiration processes, and metabolic features (Escobar-Zepeda et al., 2015).

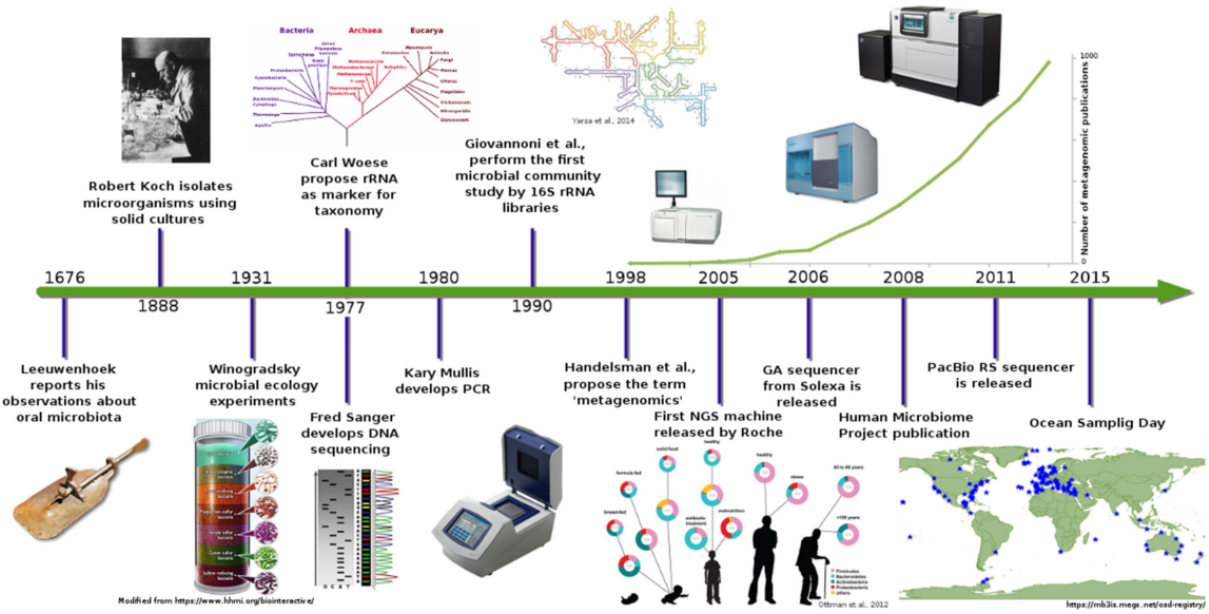


Figure 2.1 – Timeline showing advances in microbial communities studies from Leeuwenhoek to NGS, from Escobar-Zepeda et al. (2015)

In 1977, two major advances revolutionized the microbiology. DNA sequencing technology, developed by Frederick Sanger and known as Sanger sequencing (Sanger et al., 1977), was combined with Carl Woese's idea to use the small subunit (16S/18S) of ribosomal RNA genes as molecular markers to establish a phylogenetic classification of living organisms (Woese and Fox, 1977). However, at that time, DNA sequencing was a long and laborious process (Liu et al., 2012). At that time, many advances in molecular techniques were discovered and applied to microbial studies such as the polymerase chain reaction (PCR), marker genes cloning and sequencing, fluorescent in situ hybridization (FISH), denaturing gradient gel electrophoresis (DGGE and TGGE) (Escobar-Zepeda et al., 2015). In 2001, the Sanger technologies had been improved and highly automatized to detect until 500-600bp and produce up to 115Kbp per day,

allowing the sequencing of human genome (Mardis, 2011). However, the completion of Human Genome Project revealed the need for more efficient sequencing technologies to consider novel projects of the same or greater scale (Metzker, 2010). This awareness stimulated the development of new technologies and it is in this context that the Next Generation Sequencing technologies (NGS) have emerged.

2.2 Recent advances in environmental microbiology

2.2.1 Next generation sequencing technologies for microbial ecology

The main advantage of NGS technologies is their efficiency: millions of DNA molecules can be sequenced in parallel, increasing the speed of analysis and reducing costs and manpower. Among the current NGS approaches, two types of methods can be identified: the short-read and the long-read sequencing (Goodwin et al., 2016). Short-read sequencing includes two approaches: sequencing by ligation (e.g., SOLiD platform) and sequencing by synthesis (e.g., 454 pyrosequencing). Likewise, there are two types of long-read technologies: single-molecule real-time sequencing (e.g., MinION of Oxford Nanopore Technologies) and synthetic approaches based on short-read technologies to construct long reads *in silico* (e.g., 10X Genomics). A comparison of a few platforms is presented in the Table 2.1. However, the technical details of the sequencing platforms are only provided for the 454 pyrosequencer and the Illumina (MiSeq/HiSeq) platforms that were used during this doctoral research project.

Table 2.1 – Summary of several sequencing platforms, modified from Liu et al. (2012) and Goodwin et al. (2016). Abbreviations: AT, adenine and thymine; bp, base pairs; Kb, kilobase pairs; Mb, megabase pairs; Gb, gigabase pairs; d, days; h, hours; Indel, insertions and deletions; PE, paired-end sequencing; SE, single-end sequencing; (*), sequencing platform used during this doctoral research project.

Platform	Generation	Read length	Throughput	Reads	Runtime	Error profile, error rate	Cost per Gb (US$)
Sanger 3730 xl	First	400 - 900 bp	1.9 - 84 Kb	-	0.33 - 3 h	0.001 %	$2,400,000
SOLiD 5500 Wildfire	Second	75 (SE)	120 Gb	~700 M	6 d	AT bias, ≤0.1%	$130
Ion PGM 318	Second	400 (SE)	1 - 2 Gb	4 - 5.5 M	7.3 h	Indel, 1%	$450 - 800
454 GS FLX Titanium (*)	Second	Up to 600 bp (SE, PE)	450 Mb	~ 1 M	10 h	Indel, 1 %	$15,500
Illumina MiSeq v2 (*)	Second	250 bp (PE)	7.5 - 8.5 Gb	24 - 30 M (PE)	39 h	Substitution, 0.1%	$142
Illumina HiSeq3000 (*)	Second	150 bp (PE)	650 - 750 Gb	5,000 M (PE)	1 - 3.5 d	Substitution, 0.1%	$22
Pacific BioSciences RS II	Third	~20 Kb	0.5 - 1 Gb	~ 55,000	4 h	Indel, 13% single pass, ≤1% circular consensus read	$1,000
Oxford Nanopore MK 1 MinION	Third	Up to 200 Kb	Up to 1.5 Gb	>100,000	Up to 48h	Indel, ~ 12 %	$750

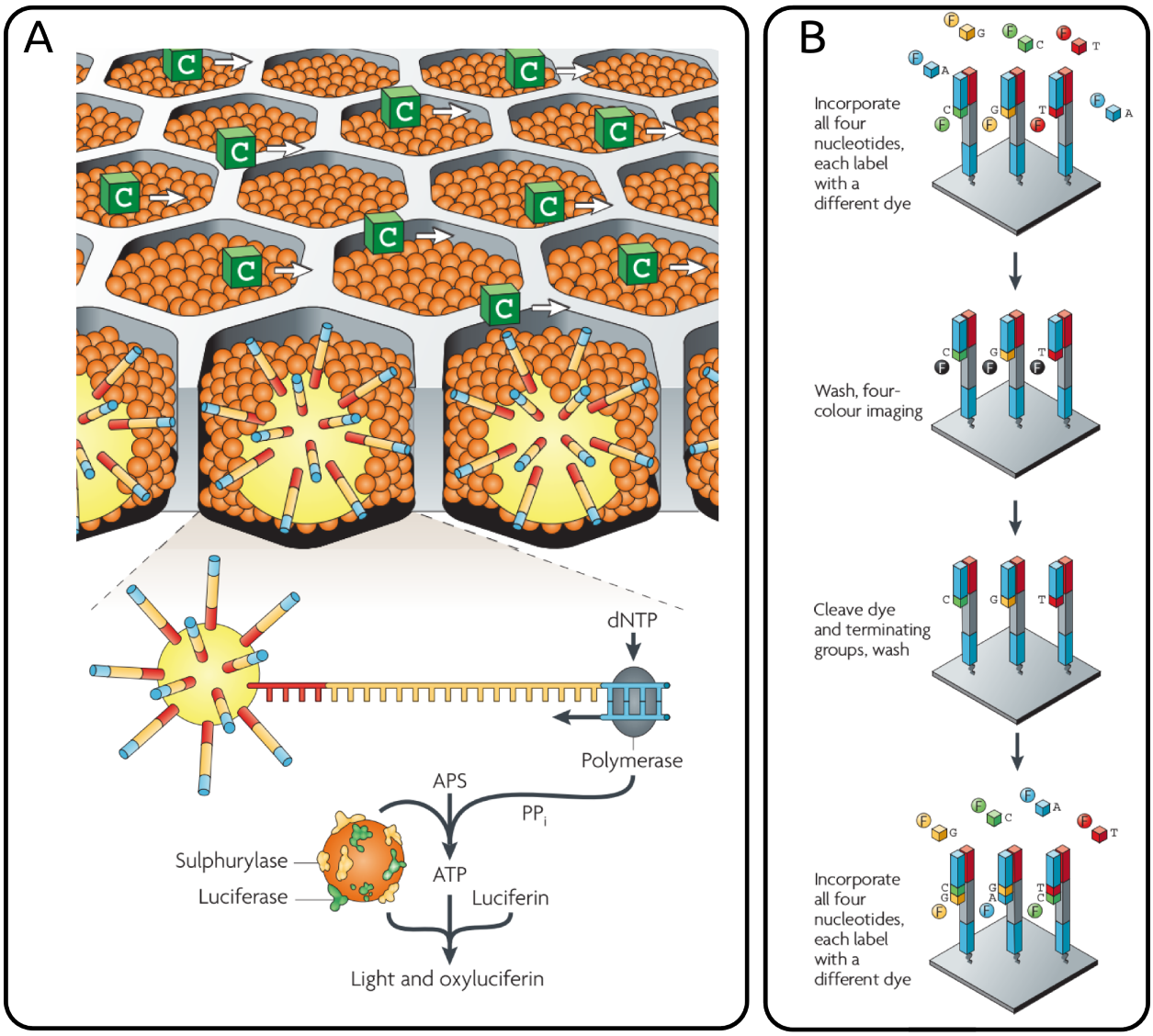


Figure 2.2 – Two sequencing strategies that use clonally amplified template, modified from Metzker (2010). (A) Single-nucleotide addition with 454 technologies (B) Cyclic reversible termination with Illumina technologies.

The first NGS instrument was the 454 sequencing platform integrating pyrosequencing approach (Margulies et al., 2005). The principle of such sequencing in as follows: first, the DNA is fragmented and then two adapters are attached to both ends of each fragment. Each DNA fragment is bound through the adapters to beads under conditions which favor one fragment per bead. The beads are immersed in water-oil emulsion in the presence of PCR mixtures to achieve amplification of each fragment. After amplification, each bead contains millions of copies of the DNA fragment initially fixed. Beads are deposited into wells of a fibre optic slide. The size of the wells ensures the deposition of one bead per well. Next, the enzymes required for pyrophosphate sequencing are added in all wells (DNA polymerase, ATP sulfurylase and luciferase). The sequencing of each fragment is based on the addition of a single kind of dNTP by a DNA polymerase (Figure 2.2A). The DNA polymerase extends the DNA strand, by incorporating complementary dNTP or waits for the next kind of dNTP. A bioluminescence method

measures the release of inorganic pyrophosphate (emitted during the polymerization process) by proportionally converting it into light using several enzymatic reactions. The order and intensity of the light peaks are recorded by CCD sensors and processed to reveal the underlying DNA sequence (Metzker, 2010).

The Illumina sequencing platforms have gradually replaced the 454 pyroesquencing technologies, because of its low cost and high yield of reads (Escobar-Zepeda et al., 2015). This sequencing technology generates reads of 100-300bp (Table 2.1). As for pyrosequencing, the DNA is first fragmented. Generic adapters are added to the fragments and these sequencing libraries are flowed across a flowcell (e.g., a glass slide). Other primers, fixed to the flowcell, provide a complementary end to which DNA fragments can hybridize. Clonally amplified clusters are generated for each fragment through bridge amplification using DNA polymerases. The sequencing of each cluster of DNA is based on a cyclic reversible termination (Figure 2.2B). The flowcell is flooded with fluorescently labelled and 3'-blocked dNTPs and DNA polymerase. Each nucleotide is associated with a unique color to be identified. A terminator ensures that only one dNTP is added at a time. The fluorescent signal is imaged to identify which dNTP was incorporated into each cluster. At this stage, the fluorophore and terminator molecules are removed and an additional washing is performed before starting the next incorporation step. The process is repeated until the full sequences are characterized.

NGS platforms provide large amounts of data, but there are still limitations. For the short-reads approaches, these sequencing methods have higher error rates (0.1-15%) and the lengths of reads are shorter than traditional Sanger sequencing (Table 2.1). On the other hand, the long-read sequencing remains considerably more expensive and has lower throughput than other platforms, limiting a widespread adoption of this technology (Goodwin et al., 2016).

2.3 Omics for microbial communities

2.3.1 Environmental DNA and meta-analyses

Environmental DNA (eDNA) refers to the genetic material that can be extracted from bulk environmental samples such as soil, water, and even air, without first isolating any target organisms (Taberlet et al., 2012). The analysis of eDNA provides access to the genetics of all microorganisms present in a given environment. Microbiologists focus on two main issues: (i) Who is there? -i.e., the identification of microbial taxa and (ii) What are they doing? -i.e., the identification of biological functions via the analysis of coding genes (Mendoza et al., 2014). The first isolation and cloning of eDNA was realized in 1991, when Schmidt *et al.* extracted total DNA from a sample of seawater, cloned it in phage λ vectors and finally selected clones containing 16S rRNA gene to analyze the marine picoplakton community (Schmidt et al., 1991). Sequencing of total eDNA was conducted in 2004 for samples from acid mine drainage (Tyson et al., 2004) and Sargasso Sea water (Venter et al., 2004), using cloning in plasmid vectors and sequencing by the Sanger method. Afterwards, the NGS technologies have been key to general-

ize eDNA studies and enable the study of complex communities found in soils, water, sediment, human microbiome, etc. (Turnbaugh et al., 2007, Vieites et al., 2009).

The next-generation sequencing was then applied in large scale analyses such as metabarcoding (amplicons), metagenomics (total DNA) and metatranscriptomics (total RNA). These meta-analyses provide an overview of the whole community in its native environment. The microbiologists also aspire to reconstruct genomes of unknown microorganisms *in silico*. These analyses can be completed by metaproteomics and metabolomics (Figure 2.3) to get an even more complete view of the biological phenomena in the community.

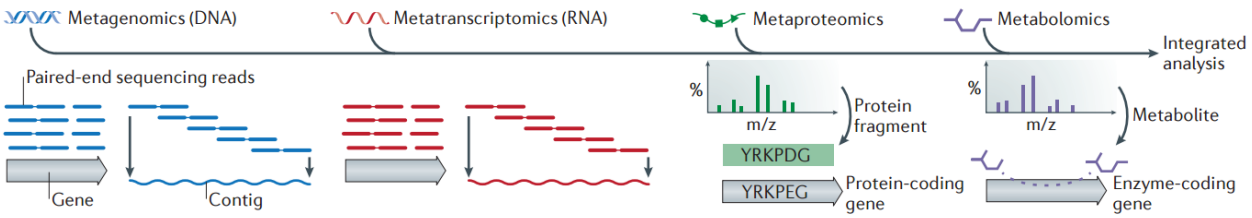


Figure 2.3 – A summary of multi-omics data types from Franzosa et al. (2015)

2.3.2 Metabarcoding

Overview Metabarcoding aims to identify the taxonomic composition of a microbial community (i.e., to answer the question: "Who is there?"). This approach consists in amplifying phylogenetic markers (such as 16S rRNA gene, *dsrAB*, *mcrA*, etc.) by PCR and sequencing the amplicons. Amplicon sequences are either directly matched to reference taxa or are first grouped into Operational Taxonomic Units (OTUs), clusters of sequences that share a given percentage of sequence identity. Several samples can be sequenced on a same run using barcodes, which are unique tags, added to each primer before PCR amplification. Such experimental designs can be used to study shifts in microbial communities in fluctuating environmental conditions.

Processing of 16S rRNA gene The first step is to remove the barcodes and check the base quality of the sequences (base calling accuracy) in order to filter low quality sequences. PCR amplification often produces chimeric sequences that must be removed in order not to be misinterpreted as new species (Edgar et al., 2011). The softwares dedicated to the chimera detection are based on reference databases of chimera-free sequences and on the relative abundance of amplicons in the data set. Once the chimeras have been removed, the reads are typically clustered into OTUs, based on sequence similarity. The largest benefits of OTU clustering is computational, since it reduces the run time of subsequent analysis steps. A similarity threshold of 97% is commonly used for OTU clustering. It comes from an empirical study that showed most strains had 97% of 16S rRNA sequence similarity (Konstantinidis and Tiedje, 2005). Many analytical tools, such as uclust have been developed on this basis (Figure 2.4A). Since the analysis of metabarcoding data from raw sequences to interpretation requires many steps, software

pipelines such as QIIME (Caporaso et al., 2010b) or FROGS (Escudié et al., 2018) have been implemented to facilitate and automate data processing.
Recently, methods based on empirical thresholds have been challenged and have led to the emergence of new tools, such as Swarm (Mahé et al., 2015) that were implemented to avoid arbitrary thresholds and eliminate the effects of input sequence ordering. These methods reveal OTUs' internal structures and allow to redefine them (for example, by breaking OTUs in sub-OTUs, Figure 2.4B). In addition to reducing a large dataset to a smaller collection of representative sequences, OTU-clustering also corrects sequencing errors (Rosen et al., 2012). However, OTUs under-utilize the quality of the data produced by the current sequencing platforms. According to some authors, one solution would be to use model-based approaches to correct amplicon errors without constructing OTUs and thus keeping biological variations (e.g., AmpliconNoise (Quince et al., 2011) or DADA2 (Callahan et al., 2016)). These methods based on exact sequence variants are considered to be the best practices for processing the *de-novo* maker genes (Knight et al., 2018).

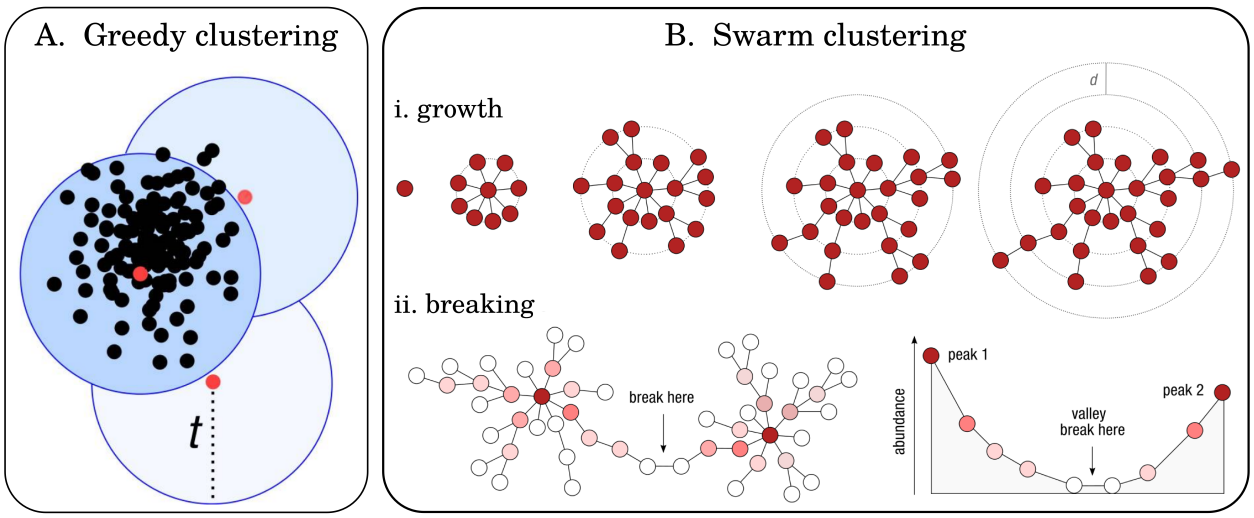


Figure 2.4 – Comparison of two clusterings, modified from Mahé et al. (2014, 2015). (A) Schematic view of a greedy clustering, where the OTU placements are based on centroid selection and a global clustering threshold, t. (B) Schematic view of Swarm's clustering and refinement approach. During the growth phase (i), Swarm clusters sequences using a local threshold, d. Suring the breaking phase (ii), Swarm takes into account the abundance of each amplicon and the internal structure of each OTU to produce clusters with a better resolution, by splitting amplicon chains. On the figure B.ii, the abundance of OTUs is indicated by their color: the darker the red, the higher the abundance.

Advantages & Limitations Metabarcoding is now widely used because of its low cost, speed and ease of preparation and analysis (Knight et al., 2018). This method provides the microbial diversity, even from low biomass, and includes uncultivated species. However, numerous factors are known to bias metabarcoding results (Knight et al., 2018). The best known negative point is due to primers and regions targeted during PCR amplification. They introduce a bias into amplified microorganisms (amplifying more preferentially some, not at all others). Moreover, this approach does not differentiate between live, dead or active populations. Other

limitations are linked to the completeness and accuracy of phylogenetic databases. Thus, many OTUs could be mis-annotated, poorly-annotated (typically limited to the genus level) or even unclassified. Finally, predictions on functional content using metabarcoding data (using PICRUSt (Langille et al., 2013) for example) are often controversial. Indeed, it requires sufficient closely related reference genomes to get accurate results. And even so, predictive functional profiling remains tricky since some microbes have very similar 16S rRNA variable regions, although they are phenotypically divergent (Knight et al., 2018).

2.3.3 Metagenomics and gene-centric analysis

Overview The term "metagenomics" was first coined in 1998 by Jo Handelsman; it refers to the investigation of all genetic material obtained from an environmental sample (Mardanov et al., 2018). Metagenomics studies aim to characterize taxonomic, functional and evolutionary aspects of organisms contained in an environmental sample. After sequencing the eDNA of a sample, sequences are annotated against taxonomic or functional reference databases (e.g., KEGG).

Annotation process There are two ways to annotate a metagenome, according to whether the raw data are assembled or not. The first approach consists in assigning taxonomy or function to short DNA fragments. As taxonomic and functional assignments are based on homology between a short read and a reference, database choice is crucial for the accuracy of results. This approach should be preferred for well-characterized environments to ensure an accurate specificity of the assignment (Knight et al., 2018). The second approach is based on the preliminary assembly of short reads into contigs. The *de-novo* assembly of metagenomes avoids introducing biases associated with fragmented and noisy short reads (Thomas et al., 2012, van der Walt et al., 2017). However, this method has some disadvantages: here, the assembly gets rid of some of the biological information by dropping reads. Most of the tools dedicated to metagenome assembly are based on de Bruijn graph approach. First, metagenomics reads are split into k-mers (sequences of length k) and are stored into a network. Then, paths are traversed iteratively to find the longer contigs. However, even after data filtering, many errors and polymorphisms remain in the data, which greatly increases the size of the de Bruijn graph. Consequently, infrastructures with a large memory size are necessary for this process. For functional annotation, gene predictions are performed on the contigs and thereafter compared to references database, such as KEGG, COG or PFAM. To find the distribution of each annotated gene, the reads needed to be mapped against the contigs.

Metagenomics normalization and comparison The gene-centric approach provides information regarding the whole microbial community. In this type of analysis, the microbial community is considered as a set of genes rather than a set of organisms (Hugenholtz and Tyson, 2008). The difference in the abundance of annotated genes between metagenomes can provide

information highlighting beneficial functions for communities (Figure 2.5). However, metagenomic data often lack replicates and the number of samples is often small (<10) (Jonsson et al., 2016). Additionally, a shallow sequencing depth results in under-sampling of genes that can be represented by only a few or no DNA fragments. Thus, methods that take into account the specificity of metagenomic count data are required to normalize them and identify the truly differentially abundant genes. Jonsson and collaborators showed that DESeq2 (Anders and Huber, 2010)) and edgeR (Robinson et al., 2009) outperform other methods for the statistical analyses in metagenomic gene count data (Jonsson et al., 2016). In order to identify the main differences in functional or taxonomic profiles, other authors recommend relying on data mining approaches, such as random forests (Dinsdale et al., 2013).

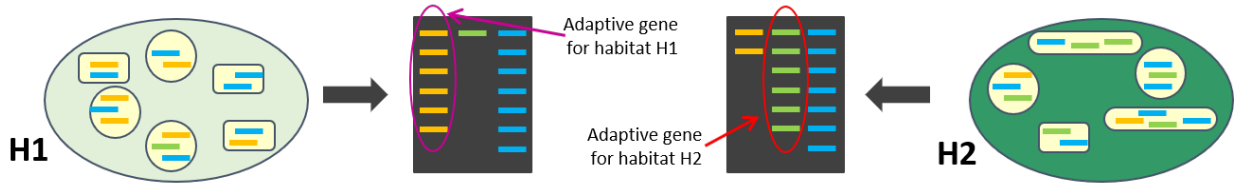


Figure 2.5 – A schematic of the gene-centric analysis, described in Hugenholtz and Tyson (2008). Genes from microbial communities inhabiting the H1 and H2 environments are assigned to a gene family (yellow, green or blue) and counted to highlight functions that are potentially beneficial or having a generalized function. This approach does not require knowing the composition of the microbial community

Advantages & Limitations Metagenomic sequencing provides the relative abundance of microbial functional genes. Unlike marker gene analyses, there is no bias related to PCR. Taxonomic affiliation can be made at the species and strains level for known organisms. Metagenomics analyses capture DNA of all microbial community (prokaryotes but also phages, viruses, microbial eukaryotes, etc.). In parallel, metagenomics provides sequences of unknown functions, which can be mined to find novel gene families (Knight et al., 2018). However, it should be noted that there are some limitations specific to metagenomic studies. Metagenomics analyses are relatively more expensive and require deeper sequencing depths than metabarcoding (Knight et al., 2018). This method requires a large amount of DNA, sometimes laborious to obtain from environmental samples. Again, metagenomics approaches cannot specifically target active processes; only metatranscriptomic methods could produce such results.

2.3.4 Metagenomics and binning

Overview Binning is the process in which DNA sequences (from a metagenomic dataset) are sorted into groups that could represent an individual genome or the genomes of very close organisms (Figure 2.6). These groups are called "MAGs" for metagenome-assembled genomes. Supervised or unsupervised approaches can be used. The supervised binning is based on phylogeny and homology: the unknown DNA fragment may encode for a gene, which have a similarity to known genes in a reference database. It requires a good knowledge of microbial diver-

sity and does not work well on environmental samples including many unknown organisms. In this thesis, only unsupervised approaches have been investigated to reconstruct microbial genomes. Most are based on tetranucleotide composition and differential coverage. The term "coverage" refers here to the number of reads that are aligned against a same contig. This concept is related to species abundance: the more a genome is represented in a sample, the higher the coverage. Binning based on species abundance considers that contigs with the same coverage come from the same genome. However, this method does not differentiate between genomes with similar abundances. The solution to improve this approach is to use other samples containing the same species in different abundance. The binning process is more reliable when microbial diversity is low, and the sequencing depth is high.

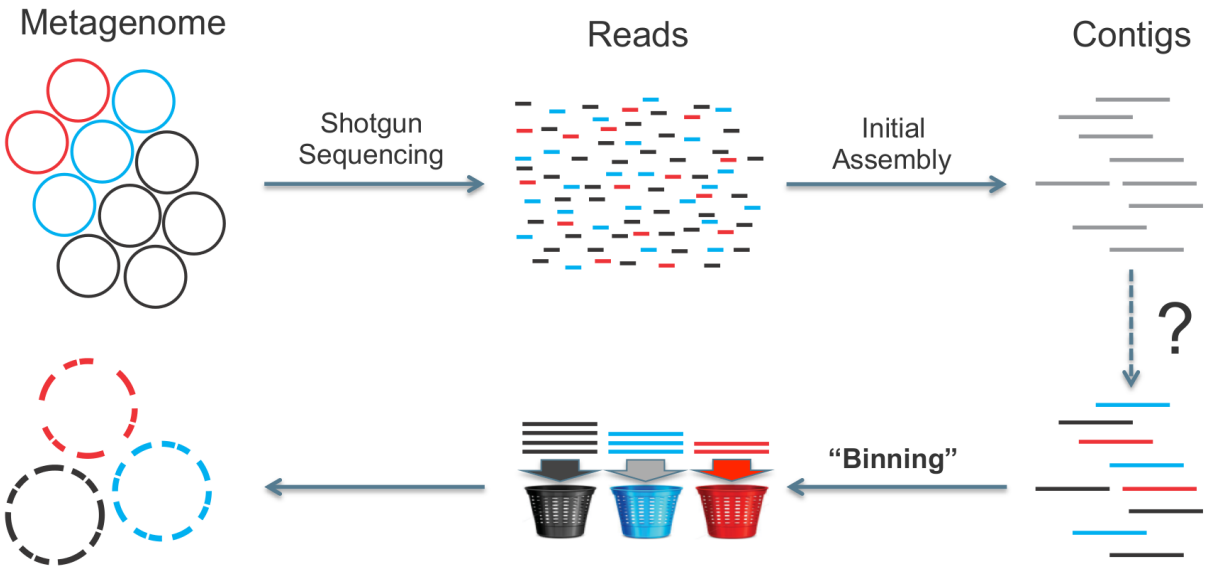


Figure 2.6 – Diagram explaining how the tool MetaBat works to reconstruct MAGs (Kang et al., 2015), from the oral presentation "An Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities", Don Kang Joint Genome Institute.

MAGs and quality evaluation The binning process is performed after the assembly of metagenomic reads into contigs. If the binning software requires differential coverage across related samples, all these metagenomes need to be combined in a same assembly (co-assembly) (Sangwan et al., 2016). Binning tools produce MAGs with different quality levels. Therefore, completeness and contamination of the MAGs need to be assed too avoid the selection of incomplete or chimeric MAGs. According to the literature, the most commonly used tool to date is checkM (Parks et al., 2015). Basically, this software searches the presence of single copy genes, an adaptive set of marker genes depending on the position of each MAG within a reference genome tree.

Advantages & Limitations Thanks to the binning approach, it is now possible to recover genomes *in silico* from metagenomic datasets, and thus to associate sets of metabolic potential to taxonomy (Sangwan et al., 2016). For example, this method was recently applied to recover 3,438

near complete genomes, including the first representatives of several candidate phyla (Parks et al., 2017). However, to obtain high quality MAGs, it requires a high depth of sequencing and many samples that shared some of the same organisms. These requirements are therefore associated with high costs (to produce many deep sequencing), long computational time, and significant IT resources (essentially to achieve the co-assembly step) (Sangwan et al., 2016). For all these reasons, binning approaches are more easily applicable in low diversity environments (Alneberg et al., 2014).

2.4 Computational challenges in the analysis of NGS

2.4.1 Massive data

The emergence of next generation sequencing led to a dramatic change in the scale of data sets, in the mid-2000s (Muir et al., 2016). Efficient storage infrastructures were thus required and databases, such as Sequence Read Archive (SRA) were created to store and organized high-throughput sequencing data. At the end of May 2018, the SRA database contained more than 18 petabases (10^{15} bases), which corresponds to twice as much data as in 2016. (Figure 2.7).

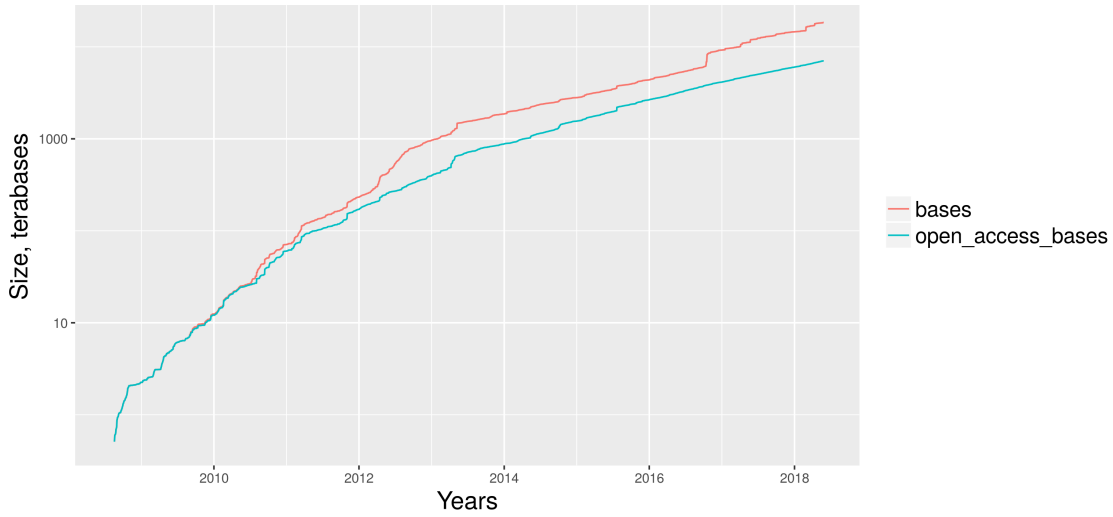


Figure 2.7 – Growth of Sequence Read Archive (SRA) over the last 10 years, download from https://trace.ncbi.nlm.nih.gov/Traces/sra/

Genomics is now considered as a Big Data science, including four critical domains: data acquisition, storage, access, and analysis (Stephens et al., 2015). Genomic data acquisition will continue to grow in parallel with technological advances in sequencing that tend to reduce costs, improve throughput and achieve very high accuracy (Stephens et al., 2015). Defining a standard for data (as well as metadata) collection and formatting is an essential step in making sequences more useful (i.e., to easily compare data sets with each other). Storage can be reduced by compressing and indexing data, but decompression times and fidelity must be prioritized for sequencing data. Various issues such as the detection of novel genes or the identification of expression changes can be addressed using sequences analysis and will require com-

puting infrastructure, machine learning systems that can support flexible and dynamic queries on highly dimensional collections (Stephens et al., 2015).

2.4.2 Computational resources

Several critical steps in NGS processing require significant computational resources. For example, as previously seen in section 2.3.2, the computation time for the processing of metabarcoding data is considerably reduced with the creation of clusters (OTUs). However, IT challenges are even greater in metagenomes processing. For example, the *de-novo* assembly of the metagenome requires up to hundreds GB of memory (RAM) (van der Walt et al., 2017). Other critical steps, such as gene annotation (or proteins research) against reference databases have high computational costs in terms of CPUs. However, to make all these comparisons, many search tools have been developed to improve performance over the gold standard, BLAST (sensitive but slow). Speed gains can be made at the expense of a lower sensitivity (Figure 2.8), but some tools manage to ensure a good sensitivity such as DIAMOND (Buchfink et al., 2014) or MMseqs2 (Steinegger and Söding, 2017).

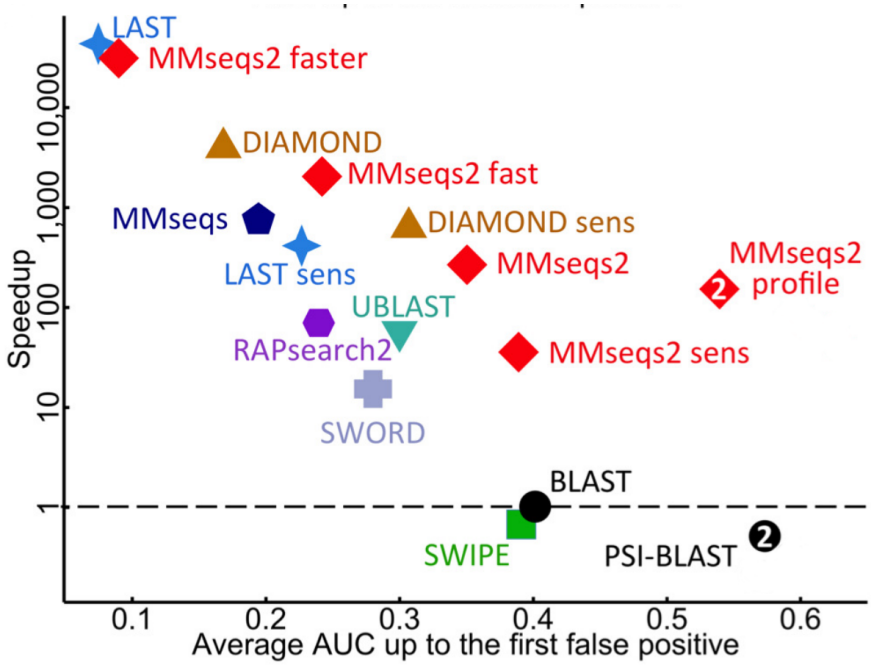


Figure 2.8 – Comparison of recent search tools, showing their trade-off between sensitivity and speed relative to BLAST, modified from Steinegger and Söding (2017). White numbers in plot symbols corresponds to the number of search iterations.

2.4.3 Reproducibility and pipelines

Reproducibility is a fundamental principle of scientific methodology. The underlying concept is to obtain the same results from the same data and softwares. In most cases, there is a significant discrepancy between the apparent methods described in the articles (i.e., above water portion of iceberg, Figure 2.9) and the information required to reproduce the analyses (i.e., full iceberg). Recently, a "culture" of reproducibility has emerged in computational science: it is now recommended to use tools that combine narratives with code (iterate programming notebook such as R Markdown (Hoffman, 2016)) or that recreate the computational environment (e.g., Docker (Beaulieu-Jones and Greene, 2017)).

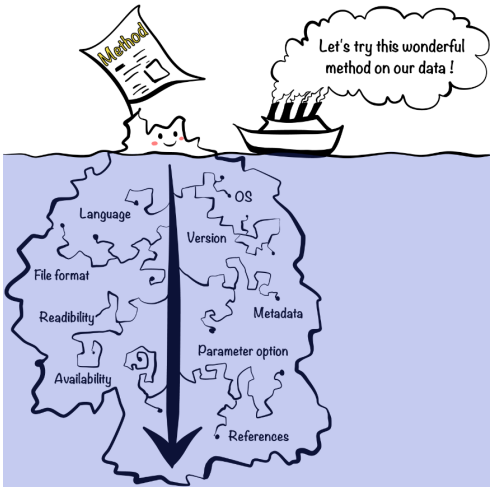
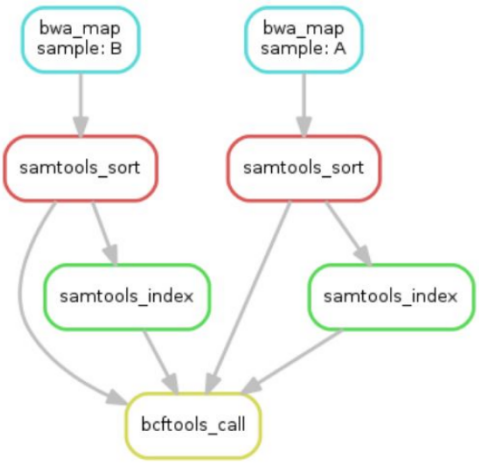


Figure 2.9 – Hidden reproducibility issues compared to underwater iceberg in Kim et al. (2017)

Sequencing data processing usually involves a chained execution of many command lines. To create automated pipelines, it is advisable to use workflow engines. There are a large number of workflow managers, with graphical interfaces (e.g., Galaxy) or not (e.g., Bpipe, Ruffus). They increase code readability, manage resources and ensure reproducibility. All the pipelines designed during this thesis were implemented with Snakemake (Köster and Rahmann, 2012). This workflow manager combines advantage of Python (clear syntax) and GNU Make (parallel execution, resource management and rules system). With Snakemake, the pipeline execution can be resumed in case of errors or changes in parameters for one or more steps. Finally, it has interesting additional features to track versions or parameters, create html reports and view tasks and their dependencies (represented as a graph, Figure 2.10).




```
rule samtools_sort:
    input:
        "mapped_reads/{sample}.bam"
    output:
        "sorted_reads/{sample}.bam"
    shell:
        "samtools sort -T sorted_reads/{wildcards.sample} "
        "-O bam {input} > {output}"

rule samtools_index:
    input:
        "sorted_reads/{sample}.bam"
    output:
        "sorted_reads/{sample}.bam.bai"
    shell:
        "samtools index {input}"

rule bcftools_call:
    input:
        fa="data/genome.fa",
        bam=expand("sorted_reads/{sample}.bam", sample=SAMPLES),
        bai=expand("sorted_reads/{sample}.bam.bai", sample=SAMPLES)
    output:
        "calls/all.vcf"
    shell:
        "samtools mpileup -g -f {input.fa} {input.bam} | "
        "bcftools call -mv - > {output}"
```

Figure 2.10 – Graph showing the link between the rules and the corresponding code in Snakemake

2.5 Overview of omics studies in serpentinizing ecosystems

Extremophile communities, exposed to heat or cold, associated with extreme pH, high salinity, hydrostatic pressure or even buried below the Earth's surface, have a fascinating capacity for survival. The metagenome sequencing, associated with other 'omics' technologies, provides a better understanding of microbial assemblages, their metabolisms and adaptive strategies (Cowan et al., 2015).

Studies in extreme environments have been conducted to provide new biological hypotheses at different scales: ecosystem (e.g., including the host), community or phylotype. For example, a metagenomic study revealed unexpected evolutionary strategies in thermophilic environments of hydrothermal vent systems (Anderson et al., 2014). These hydrothermal systems were characterized by a high abundance of transposases, integrases and lateral gene transfer compared to other ecosystems. It has been hypothesized that viruses and their hosts have a mutualist relationship (instead of parasitic relationship) and they have co-evolved to obtain adaptive advantages to survive in an extreme and dynamic environment. Another example is the reconstruction of genomes from *de-novo* assembly of metagenomic sequences to study extremophiles difficult to grow. Thus, in halophilic habitats, a study provided an overview of the potential capabilities of an uncultured Nanohaloarchaea, including proteins with high flexibility and osmo-resistance (Narasingarao et al., 2012). These are only examples, as it would be difficult to draw up an exhaustive list of all omics studies conducted in extreme environments. Serpentinizing systems, also considered extreme, have been intensively studied through 'omics' since the emergence of NGS technologies (Table 2.2).

2.5.1 Novel tools to explore microbial communities

Evolution of microbial studies

In the late 1980s, microbiology studies of serpentinizing ecosystems were mainly based on culture-dependent methods (Bath et al., 1987, Tiago et al., 2004). They have revealed many alkaliphilic heterotrophs, but only a few strains were fully characterized by Tiago and colleagues (details and references in the Table 2.2). Thus, culture-independent approaches quickly became necessary to access the large fraction of uncultivated microorganisms.

At first, these studies were based on molecular cloning and it was not until 2009 that NGS technologies were introduced to describe the microbial community of serpentinizing ecosystems (Brazelton and Baross, 2009). Since the last few years, 15 studies of metabarcoding and 9 metagenomes have been published on serpentinizing ecosystems (Table 2.2). To date, no other OMIC techniques (such as transcriptomics or proteomics) have been used in serpentinizing systems, leaving open questions about the microbial activities. It is important to note that the emergence of NGS has not hindered efforts to isolate new microorganisms, given the recent characterizations of anaerobic bacteria in PHF (Ben Aissa et al., 2015, Bes et al., 2015, Mei et al., 2016a).

Site	Location	Hydrothermalism origin	Characterization of new species	Count and references of microbial studies in Molecular cloning	MetaBarcoding	Metagenomics
Hyperalkaline Allas Springs	Cyprus	Serpentinization		* Rizoulis et al. (2014)		
Del Puerto Ophiolite	California, USA	Serpentinization		* Blank et al. (2009)		
The Cedars	California, USA	Serpentinization	*** *Serpentinomonas raichei* spp. (Suzuki et al., 2014), *S. mccroryi* sp. (Suzuki et al., 2014)	* Suzuki et al. (2013)		* Suzuki et al. (2017)
Coast Range Ophiolite	California, USA	Serpentinization			* Crespo-Medina et al. (2014)	* Twing et al. (2017)
The Tablelands	Canada	Serpentinization			* Brazelton et al. (2013)	* Brazelton et al. (2012)
The Santa Elena Ophiolite	Costa Rica	Serpentinization			** Crespo-Medina et al. (2017), Sánchez-Murillo et al. (2014)	* Crespo-Medina et al. (2017)
Outokumpu	Finland	Serpentinization		* Itävaara et al. (2011)		
Voltri Massif	Italy	Serpentinization			* Quéméneur et al. (2015)	* Brazelton et al. (2016)
Leka ophiolite	Norway	Serpentinization		* Daae et al. (2013)		
Samail Ophiolite	Oman	Serpentinization			* Rempfert et al. (2017)	
Zambales ophiolite	Philippines	Serpentinization	* *Exiguobacterium* sp. AB2 (Cabria et al., 2014)		* Woycheese et al. (2015)	
Cabeço de Vide Aquifer	Portugal	Serpentinization	***** *Bacillus foraminis* sp. (Tiago et al., 2006c), *Chimaereicella alkaliphila* sp. (Tiago et al., 2006a), *Microcella alkaliphila* sp. (Tiago et al., 2006b), *Microcella putealis* sp. (Tiago et al., 2005b), *Phenylobacterium falsum* sp. (Tiago et al., 2005a)	** Tiago et al. (2004), Tiago and Veríssimo (2013)	* Tiago and Veríssimo (2013)	
Chimaera ophiolite	Turkey	Serpentinization			* Neubeck et al. (2017)	
Prony Hydrothermal Field	New Caledonia	Serpentinization	***** *Acetoanaerobium pronyense* sp. (Bes et al., 2015), *Alkaliphilus hydrothermalis* sp. (Ben Aissa et al., 2015), *Clostridium* sp. (Mei et al., 2014), *Vallitalea pronyensis* sp. (Ben Aissa et al., 2014), *Serpentinicella alkaliphila* sp. (Mei et al., 2016a)	*** Pisapia et al. (2017), Postec et al. (2015), Quéméneur et al. (2014)	** Frouin et al. (2018), Mei et al. (2016b)	* Mei et al. (2016b)
Lost-City Hydrothermal Field	Mid-Atlantic Ridge	Serpentinization		***** Brazelton et al. (2006, 2010b), Gerasimchuk et al. (2010), Roussel et al. (2011), Schrenk et al. (2004)	*** Brazelton and Baross (2010), Brazelton et al. (2010a,b)	*** Brazelton and Baross (2009, 2010), Brazelton et al. (2012)
Rainbow	Mid-Atlantic Ridge	Serpentinization and Magmatic influence		*** Nercessian et al. (2005), Roussel et al. (2011)	* Flores et al. (2011)	
Logatchev	Mid-Atlantic Ridge	Serpentinization and Magmatic influence		** Perner et al. (2007, 2010)		
Ashadze	Mid-Atlantic Ridge	Serpentinization and Magmatic influence	* *Pyrococcus yayanosii* sp. (Birrien et al., 2011)	* Roussel et al. (2011)		
South Chamorro	Marianna Forearc	Serpentinization	* *Marinobacter alkaliphilus* sp. (Takai et al., 2005)	* Curtis et al. (2013)		
Kairei Field	Central Indian Ridge	Serpentinization and Magmatic influence		* Takai et al. (2004)		
site 30I-TVG05	Indian Ocean	Serpentinization	* *Desulfovibrio indicus* (Cao et al., 2016)			

Table 2.2 – Summary of microbial studies in serpentinizing ecosystems

Characterization of uncultivated microorganisms

The majority of microorganisms remain uncultivated and this is why binning approaches are considered a real revolution and a way to infer the metabolisms of unknown organisms. In serpentinizing ecosystems, the binning method was used to reconstruct genomes in samples from Voltri Massif and The Cedars (Brazelton et al., 2016, Suzuki et al., 2017) to better understand the role of some microorganisms. However, as explained earlier in the section 2.3.4, the recovery of high quality MAGs is challenging as exemplified by only five incomplete MAGs recovered from the ecosystem of the Voltri Massif (Brazelton et al., 2016). In their study, Suzuki and colleagues discussed the small genome size of organisms affiliated to candidate division OD1 (Suzuki et al., 2017), but no one can state for sure that this result was not related to the loss of information during data processing (assembly or binning).

At least three possibilities can help to strengthen this kind of result. The first idea is to try to isolate these organisms by using their metabolic potential. Indeed, the presence or absence of metabolic pathways in the target microorganism can help to design a customized culture media (Vartoukian, 2016). A second approach could be to use single-cell genomic dataset to improve metagenomic binning as proposed by Becraft et al. (2016). Finally, to overcome a high microbial diversity that complicates the binning process, a gene capture by hybridization can be applied to specifically enrich a sample in large DNA fragments associated with a given microorganism (Denonfoux et al., 2013).

2.5.2 Integrative approaches

Comparison of microbial communities

Although most of microbes living in serpentinizing systems are not characterized or isolated, it is possible to compare microbial communities with each other, using NGS approaches. Using the numerous analyses based on cloning and metabarcoding methods, the predominant taxa of a serpentinizing site were often also searched in other serpentinizing systems (Mei et al., 2016b, Suzuki et al., 2013), highlighting some similarities. Surprisingly, no global comparison of serpentinizing sites has been made. Such comparison is most likely complicated by the absence of standard to analyze taxonomic composition of different microbial communities (selection of primers, regions 16S rRNA gene, sequencing methods, etc.). Regarding whole metagenome analysis, the functional content was only compared between different samples of a same serpentinizing system (Brazelton et al., 2016, Crespo-Medina et al., 2017, Twing et al., 2017), but never within different serpentinizing systems. To our knowledge, inter-system comparisons involve only one serpentinizing site, which is always LCHF, and other environments, such as black smokers (Cao et al., 2014, He et al., 2013, Oulas et al., 2016, Xie et al., 2011).

Integrative approach of multiple types of data

To better understand how an environment works, we need to collect as many observations as possible. For microbial community studies, DNA sampling is not sufficient. The collection of associated metadata, such as geochemical variables (e.g., pH, temperature, ion concentrations) is crucial. The metadata can then be combined with a sequencing analysis and their correlations can be evaluated by statistical methods. Two examples of integrative approaches are presented in Figure 2.11. However, these methods are only marginal in published studies of serpentinizing ecosystems and should be further explored in future studies.

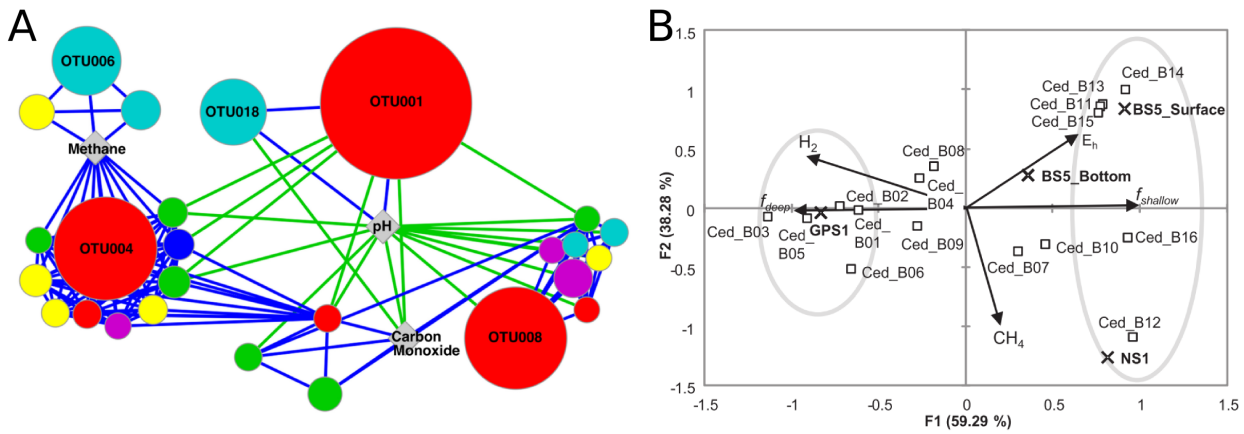


Figure 2.11 – Integration of environmental variables into phylotypes sequence dataset. (A) Network diagram of significant positive (blue) and negative (geen) correlations between phylotypes (circles) and environmental variables (diamonds) from Twing et al. (2017). (B) Canonical correspondence analysis diagram of environmental variables and phylotypes from Suzuki et al. (2013).

The integration of metaproteomic, metatranscriptomic and metagenomic studies offers the opportunity to link potential, transcribed and translated functions of microbial communities. The integration of multi-omic data can provide more in-depth biological knowledge (Franzosa et al., 2015). For example, the combination of metagenomics and metatranscriptomics allows the normalization of transcript abundances and can help to highlight more accurately the differentially expressed functions. Moreover, signals (e.g., a level of abundance) from different studies can provide collective support for a hypothesis, even if they are individually weak (Franzosa et al., 2015). Regarding serpentinizing ecosystems, such approaches could be interesting to evaluate the link between microbial community composition, their metabolism and the concentrations of certain chemical components.

Chapter 3

Rare Biosphere of Prony Hydrothermal Field [1]

Sommaire

[1] The work described in this chapter was published in: Frouin E, Bes M, Ollivier B, Quéméneur M, Postec A, Debroas D, Armougom F and Erauso G (2018) Diversity of Rare and Abundant Prokaryotic Phylotypes in the Prony Hydrothermal Field and Comparison with Other Serpentinite-Hosted Ecosystems. *Frontiers in Microbiology* 9:102. doi: 10.3389/fmicb.2018.00102

Overview (French part)

L'article intitulé "Diversity of rare and abundant prokaryotic phylotypes in the Prony Hydrothermal Field and comparison with other serpentinite-hosted ecosystems" présente de nouvelles données sur les communautés microbiennes du site hydrothermal serpentinisé de Prony. Le site d'étude, situé dans la baie de Prony, en Nouvelle-Calédonie est caractérisé par la présence de cheminées carbonatées (sous-marines ou en zone intertidale) qui libèrent un fluide hydrothermal alcalin enrichi en méthane et hydrogène. Des études de biodiversité, réalisées précédemment et qui avaient ciblé les communautés microbiennes de ces cheminées n'avaient pas permis de couvrir toute la diversité, car elles s'appuyaient sur du clonage, une technique ne permettant pas d'être exhaustif. L'étude présentée dans ce chapitre cherche à établir la distribution des taxons procaryotes en fonction des cheminées échantillonnées en incluant les micro-organismes présents en très faible abondance grâce au pyroséquençage 454 d'amplicons du gène universel de la petite sous-unité de l'ARNr.

Le premier résultat de ce chapitre révèle l'importance de la biosphère rare à Prony, puisque plus de 70% des micro-organismes identifiés ont une abondance inférieure à 0.2%, incluant notamment tous les taxons affiliés aux divisions candidates. Dans un second temps, la distribution différentielle des taxons procaryotes a été étudiée selon la localisation de leur cheminée d'origine. La communauté bactérienne issue du site intertidal se distingue clairement des communautés provenant de cheminées sous-marines, principalement en raison de quelques taxons abondants appartenant aux phyla *Cloroflexi* et *Proteobacteria*. En revanche, la distribution des archées semble moins dépendante de la localisation de la cheminée échantillonnée comme en atteste l'occurrence, dans chaque communauté échantillonnée, des phylotypes de *Methanosarcinales* que l'on pensait plus spécifiques aux sites serpentinisés terrestres (pour les The Cedars Methanosarcinales) ou marins (pour les Lost-City Methanosarcinales).

Ce chapitre présente également la comparaison des phylotypes procaryotes provenant du site hydrothermal de Prony avec ceux de cinq autres sites serpentinisés, éloignés géographiquement et ayant précédemment fait l'objet d'études sur leur diversité microbienne. Bien qu'il n'ait pas été possible d'établir un sous-ensemble de taxons qui seraient partagés par tous les sites étudiés, certains micro-organismes incultivés, notamment ceux appartenant à l'ordre des *Methanosarcinales* et la classe des *Dehalococcoidia* (la division candidate MSBL5) ont été retrouvés exclusivement dans des sites serpentinisés. Enfin des taxons apparentés à l'ordre des *Clostridiales*, *Thermoanaerobacterales*, ou le genre *Hydrogenophaga*, sont abondants dans plusieurs sites. Ces micro-organismes pourraient constituer une signature taxonomique des écosystèmes serpentinisés. Ces résultats permettent d'étendre notre connaissance de la diversité microbienne colonisant les milieux serpentinisés et leur biogéographie.

3.1 Introduction

Serpentinization is the alteration process that abiotically transforms olivine and pyroxene-rich rocks into serpentinites and yields alkaline hot fluids enriched in H_2 and CH_4 (Chavagnac et al., 2013). Active serpentinization generates physico-chemical conditions that might have prevailed in early Earth and could have been fundamental with regard to the origin of life (Russell et al., 2010, Schrenk et al., 2013). Several serpentinizing systems, particularly terrestrial ones, have recently been studied by molecular approaches using next generation sequencing, to reveal the indigenous microbial community. They included the Cabeço de Vide Aquifer (CVA, Portugal) (Tiago and Veríssimo, 2013) and the ophiolites of Tablelands (Canada) (Brazelton et al., 2013), Santa Elena (Costa Rica) (Crespo-Medina et al., 2017, Sánchez-Murillo et al., 2014), Zambales (Philippines) (Woycheese et al., 2015), Samail (Oman) (Rempfert et al., 2017), Chimaera (Turkey) (Neubeck et al., 2017), Voltri (Italy) (Brazelton et al., 2016, Quéméneur et al., 2015), The Cedars (Placer, CA, United States) (Suzuki et al., 2017) and the nearby site of Coast Range Ophiolite Microbial Observatory (CROMO, Burbank, CA, United States) (Twing et al., 2017). In contrast, only a few submarine serpentinizing fields have been investigated using high-throughput sequencing. Their sampling is indeed complicated due to their remote geographic location and great depths. Among them, the emblematic Lost-City Hydrothermal Field (LCHF) was the only serpentinizing ecosystem to be subjected to microbiological study targeting the rare biosphere (Brazelton et al., 2010b).

The recently described Prony Hydrothermal Field (PHF) (Monnin et al., 2014, Pelletier et al., 2006, Postec et al., 2015, Quéméneur et al., 2014) is a shallow (50 m maximum depth) marine alkaline hydrothermal system, located in the Bay of Prony, South of New Caledonia. The run-off waters from the rocky hills surrounding the Bay percolate through the densely fractured peridotite basement. The resulting end-member fluids discharge into the lagoon and, upon mixing with ambient seawater, precipitate, forming large carbonate edifices resembling Lost-City chimneys (Monnin et al., 2014). Similarly, to those of Lost-City, PHF fluids are characterized by high pH values (between 9.0 and 11.5) and elevated N_2, H_2, and CH_4 concentrations (Monnin et al., 2014). However, the PHF fluids are different in terms of temperatures (maximum 40°C) and salinity (less than 0.5 g/L). The microbial community composition of the inner part of PHF chimneys has been previously characterized by culture-dependent (Ben Aissa et al., 2015, 2014, Bes et al., 2015, Mei et al., 2014), molecular-based techniques such as fingerprinting methods (SSCP and DGGE), clonal Sanger sequencing (Postec et al., 2015, Quéméneur et al., 2014), and Fluorescent In Situ Hybridization (FISH) and micro-imaging (Pisapia et al., 2017). Molecular-based studies revealed that bacteria largely outnumbered the archaea communities. The bacterial communities were composed of diverse populations mostly affiliated to the *Chloroflexi* and *Firmicutes* in the submarine sites and to the *Proteobacteria* in the intertidal sites (Quéméneur et al., 2014). The low-diversity archaeal communities were dominated by *Methanosarcinales* with phylotypes similar to those of the Lost-City Methanosarcinales (LCMS) and The Cedars Methanosarcinales (TCMS). While previous studies in PHF have identified the most abundant

phylotypes, a substantial part of the indigenous microbial diversity was assumed to remain missing from the 16S rRNA genes inventories, due to the low-coverage sequencing (Postec et al., 2015, Quéméneur et al., 2014). Indeed, the coverage of the clone libraries was low, especially for the bacteria, with values ranging from 54 to 79%. Since such approaches underestimate the true biodiversity of environments (Curtis and Sloan, 2005), higher resolution techniques such as deep sequencing are therefore required in order to recover the overall microbial diversity of PHF. A preliminary study was performed with next-generation sequencing in PHF sites, but was only focused on the diversity of potentially hydrogen-producing bacteria and their hydrogenases (Mei et al., 2016b).

In this study, we performed 454 pyrosequencing of the V1-V3 regions of the prokaryotic 16S rRNA gene in order to investigate the composition and the diversity of the microbial biosphere inhabiting carbonate chimneys of four sampling sites of the PHF. We showed that the diversity of bacteria (but not archaea) was site-specific and that the overall PHF biosphere was dominated by rare prokaryotic phylotypes. Finally, using publicly available sequences of 16S rRNA genes, we reported the co-occurrence of bacterial and archaeal operational taxonomic units (OTUs) identified in PHF and in other serpentinizing systems.

3.2 Materials and Methods

3.2.1 Sample Collection

Description of the sampling sites, sample processing and storage, and physico-chemical characteristics of PHF have been reported previously(Monnin et al., 2014, Postec et al., 2015, Quéméneur et al., 2014). Briefly, four active carbonate chimneys were sampled at four distinct sites of the Bay of Prony, including the “Bain des Japonais” (BdJ), the sites (ST) ST07 (also called “Aiguille de Prony”), ST09 and ST12. The BdJ site is located on the foreshore of the Carenage Bay and is uncovered at low tide. It consists of many fluid outlets building a carbonate plateau. There, small needle-like brittle structures of 10-20 cm high and 3-6 cm in diameter were collected. The submarine sites ST07, ST12, and ST09, located in the Prony Bay, were sampled at 16 meters below sea level (mbsl), 38 mbsl, and 48 mbsl, respectively. They form large carbonate edifices, carrying several chimneys of up to several meters high. The chimneys are porous structures, mainly composed of brucite [$Mg(OH)_2$], aragonite and calcite ($CaCO_3$) and magnesium carbonates ($MgCO_3$) (Pisapia et al., 2017). The external walls of the mature submarine chimneys are colonized by various invertebrates (e.g., ascidia, sponges, corals). The top of such chimneys was collected by scuba-diving in each submarine site. They were kept at 4-8°C during their transportation (3 h) to the laboratory, where they were immediately processed. Cross sections (4 cm thick) of these chimneys top were realized aseptically; peripheral parts of the sections (2 cm thick) were cautiously removed using a chisel and sub-samples of the inside parts of the chimneys, bathed with the alkaline fluid, were collected and stored at -80°C until DNA extraction, as detailed previously (Postec et al., 2015, Quéméneur et al., 2014).

Table 3.1 – Samples and multiplex identifiers.

Site	Sampling date	MID	Sequence
ST09	2010	MID1	ACG-AGT-GCG-T
ST07	2012	MID2	ACG-CTC-GAC-A
ST12	2011	MID3	AGA-CGC-ACT-C
BdJ	2011	MID4	AGC-ACT-GTA-G

3.2.2 Ribosomal RNA Gene Amplification and 454 Pyrosequencing

DNA extraction and quantification procedures have been described previously (Postec et al., 2015). Briefly, about 0.4 g of chimney slurry was transferred into a sterile 2 mL tube containing glass beads (lysing matrix E from MP BioMedicals) and 0.8 mL of lysis buffer (Tris.HCl 100 mM, NaCl 100 mM, EDTA 50 mM, pH 8). Cell lysis was achieved with chemical reactions (adding Sodium Dodecyl Sulfate and Lauryl Sarcosine) and mechanic disruption using a Fast Prep homogenizer. DNA was extracted using phenol:chloroform:isoamyl alcohol and chloroform:isoamyl alcohol. DNA from each sample or dilution was used as template for 16S rRNA amplification. The hypervariable region V1-V3 of the 16S rRNA gene was amplified using the pair of primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') / 533R (5'-TTACCGCGGCTGCTGGCAC-3') and 21F (5'-TCYGKTTGATCCYGSCRGA-3') / 529R (5'-TCGCGCCTGCTGCRCCCCGT-3') for bacteria and archaea, respectively. The forward and reverse primers were barcoded (Table 3.1). PCR reactions were performed in a 50 µl mixture containing PlatiniumTM Taq High Fidelity (Life Technologies) 1 unit, High Fidelity PCR Buffer 1X, dNTP mixture 0.2 mM each, MgSO4 2 mM, primer mix 0.3 µM each, and 2 µl template DNA. The protocol mentioned below was used for amplification: initial denaturation at 95°C for 2 min, followed by 25 cycles of denaturation 95°C for 45 s, primers annealing at 54°C for 45 s, elongation at 72°C for 1 min 30 s, and a final elongation step at 72°C for 7 min. At least three PCR products from each combination of sample and primer pairs were pooled and then purified with QIAquick Gel Extraction Kit (Qiagen). An equimolar amount of each PCR pool was finally used for library construction (addition of the adaptors by ligation) and pyrosequencing using 454 GS FLX Titanium technology (Beckman Coulter Genomics).

3.2.3 16S rRNA Gene Analysis

Pyrosequencing of the V1-V3 region of the microbial 16S rRNA generated 188,058 sequences. Raw nucleotide sequence data were submitted to the European Nucleotide Archive under project accession number PRJEB21795. These sequences were filtered according the following criteria: read length between 150 and 500 bp, no ambiguous bases, average quality score ≥ 25, and no error allowed in primer and barcode sequences. A total of 74,457 and 95,916 high quality sequences were obtained for the bacteria and archaea, respectively. Alignments and clustering to Operational Taxonomic Units (OTUs) were carried out using the QIIME 1.9.1 workflow

(http://www.qiime.org/) (Caporaso et al., 2010a). Chimera sequences were identified and removed using UCHIME (Edgar et al., 2011). Multiple sequence alignments were performed using PyNAST (Caporaso et al., 2010b). OTUs were defined by clustering sequences using UCLUST (Edgar, 2010) with a pairwise distance threshold value of 3%, a cutoff currently used for species demarcation while reducing the potential inflation of the number of OTUs due to pyrosequencing errors. To limit the impact of spurious OTUs, low abundance OTUs accounting for less than 0.005% of total sequences were removed as recommended by Bokulich et al. (2013). Taxonomic assignment of the filtered sequences was performed using the RDP classifier algorithm (Wang et al., 2007) with a minimum bootstrap confidence of 80% against the SILVA database (Quast et al., 2013) release for QIIME (v.123). The OTU table was rarefied (i.e., downsampled to the sample with the smallest set of sequences) to reduce sequencing depth heterogeneity between samples. There is currently no standardized threshold for the rare biosphere. Here, rare OTUs were defined as OTUs comprising 0.005 to 0.2% of sequences per sample, an empirical threshold previously adopted by Hugoni et al. (2013). OTUs comprising 0.2 to 1% of sequences per sample were considered as intermediate and OTUs representing more than 1% of sequences per sample were considered as abundant. The alpha diversity within the four samples was estimated using the Shannon (Shannon and Weaver, 1949) and Simpson (Simpson, 1949) indices, and the beta diversity among the samples was measured using weighted UniFrac (Lozupone and Knight, 2005) metric. Taxonomic composition, and diversity results were visualized using the R package phyloseq v.1.14.0 (McMurdie and Holmes, 2013). BLASTn (Basic Local Alignment Search Tool) searches against the NCBI (National Center for Biotechnology Information) database of nucleotides (nt) and against the RefSeq database were performed for some OTU sequences. To infer the phylogeny of *Methanosarcinales*, cultivated members of this order and environmental clones were added to the sequences of abundant or ubiquitous PHF OTUs. The V1-V3 regions of these sequences were aligned with MAFFT v7.123b (Katoh et al., 2002) and a maximum likelihood tree was inferred using the software SeaView (Gouy et al., 2010) with 1000 bootstrapped trials.

3.2.4 Taxonomic Comparison of Phylotypes in Serpentinizing Environments

The correctly formatted amplicons and clone libraries of five serpentinizing ecosystems were downloaded from public databases (SRA and Genbank). These data constituted a subset of known serpentinizing systems, including four continental environments: Cabeço de Vide Aquifer (Tiago and Veríssimo, 2013), The Cedars (Suzuki et al., 2013), Voltri ophiolite (Quéméneur et al., 2015), Leka ophiolite (Daae et al., 2013), and one submarine system: the famous Lost-City Hydrothermal Field (Brazelton et al., 2006, Schrenk et al., 2004). A closed-reference OTU picking process was carried out to compare the sequences of amplicons and clone libraries covering different regions of the 16S rRNA gene. The sequences were clustered against a reference sequence database (SILVA database, v123). The OTUs shared among samples were visualized with Anvi'o v2.3.2 (Eren et al., 2015).

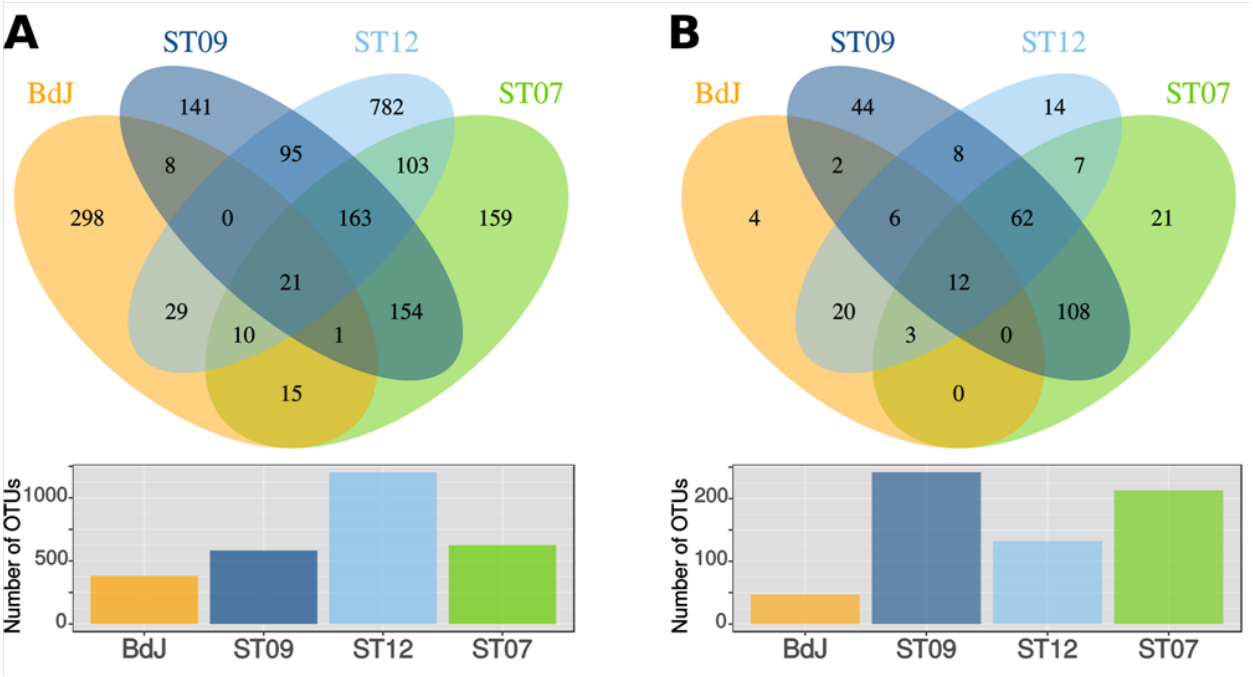


Figure 3.1 – Venn diagrams illustrating the number of shared and specific OTU in the four chimneys of PHF for bacteria (A) and archaea (B). The two histograms show the total number of OTUs per site.

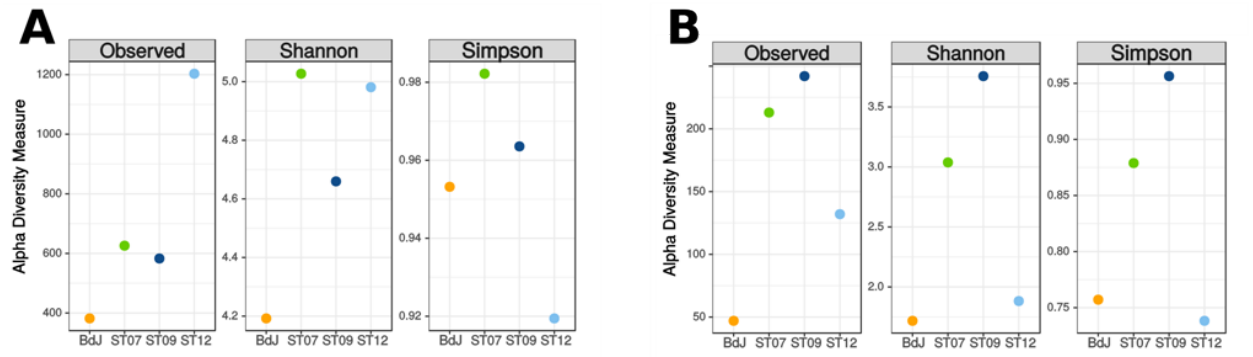


Figure 3.2 – Alpha diversity of bacterial (A) and archaeal (B) comunities of PHF. Three metrics are shown the number of OUs observed, the Shannon and the Simpson Indexes.

3.3 Results and Discussion

3.3.1 Microbial Community Diversity

The bacterial species richness of the sampling sites comprised 1,979 distinct OTUs (Figure 3.1A). However, the species richness distribution varied significantly from one site to another, with at least a twofold decrease in OTU abundance for BdJ, ST07, and ST09 compared to the ST12 site (1203 OTUs, Table 3.2). The presence of more than 780 OTUs, specific to the ST12 site (Figure 3.1A) and comprising mainly of rare taxa, might result from a lower in situ pH condition (around 9.2), in comparison with the other studied PHF sites (pH > 9.8) (Monnin et al., 2014). This trend has previously been observed in the Coast Range Ophiolite serpentinizing site, where circumneutral pH wells contained many rare taxa that were absent in the high pH ones (Twing et al., 2017), most likely because these latter drastic conditions selected the

Table 3.2 – Basic statistics of OTUs distribution and diversity indexes

		Archaea				Bacteria			
		BdJ	**ST07**	**ST09**	**ST12**	**BdJ**	**ST07**	**ST09**	**ST12**
Trimmed sequences		21395	26167	20958	27396	15090	15431	15731	28205
Number of OTUs		47	213	242	132	382	626	583	1203
Diversity indexes	Shannon	1.72	3.04	3.76	1.88	4.19	5.03	4.66	4.98
	Simpson	0.76	0.88	0.96	0.74	0.95	0.98	0.96	0.92
Rare OTUs (<0.2 %)	number of OTUs	34	168	190	115	324	548	513	1146
	% rare OTUs	72.3	78.9	78.5	87.1	84.8	87.5	88	95.3
	% of sequences	1.7	7.4	8.8	4.6	20.3	28	26.8	42.8
Intermediate OTUs (0.2-1%)	number of OTUs	7	34	30	12	42	59	56	54
	% interm. OTUs	14.9	16	12.4	9.1	11	9.4	9.6	4.5
	% of sequences	2.7	16.9	12.7	5.3	18.8	24.8	23.3	19.4
Abundant OTUs (>1 %)	number of OTUs	6	11	22	5	16	19	14	3
	% abundant OTUs	12.8	5.2	9.1	3.8	4.2	3	2.4	0.2
	% of sequences	95.5	75.7	78.5	90.1	60.9	47.2	50	37.7

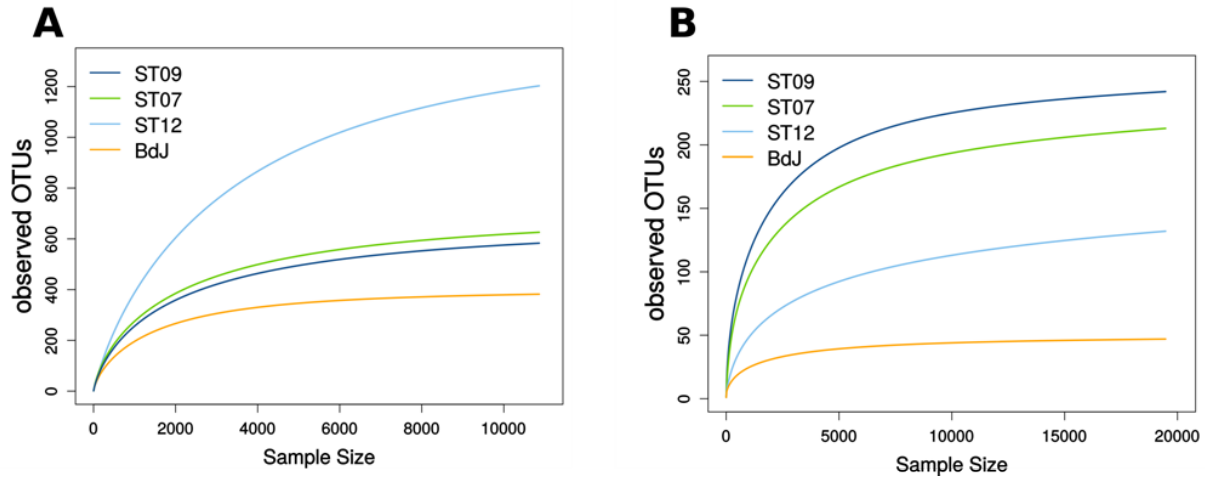


Figure 3.3 – Rarefaction curves, showing the number of OTUs given the number of sequenced 16S rRNA gene amplicons or bacteira (A) and archaea (B).

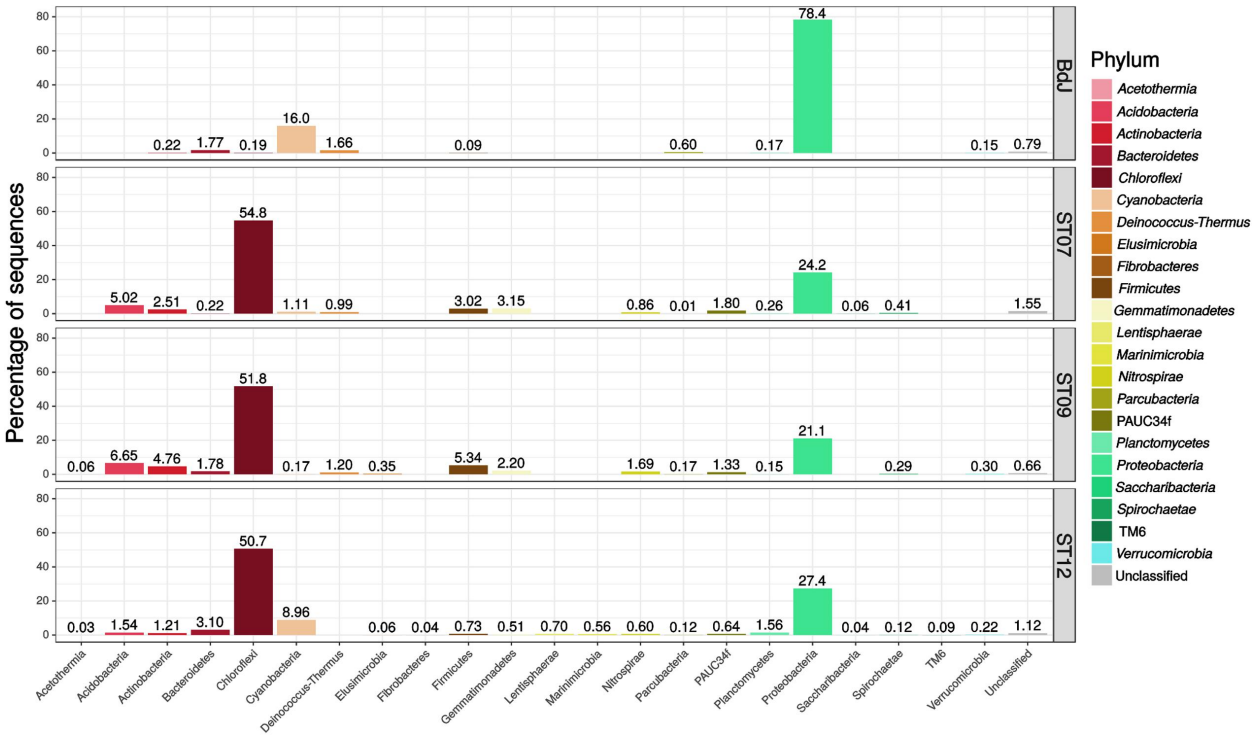


Figure 3.4 – Phylum-level bacterial diversity in Prony Hydrothermal Field (PHF) sites based on 16S rRNA gene sequences (V1-V3 region). The relative abundances were estimated for the four chimneys collected in BdJ, ST07, ST09, and ST12 sites.

few most alkaliphilic members within the prokaryotic community. The overall species richness was considerably lower in the archaea (310 OTUs) than in the bacteria, and also showed an imbalance in OTU distribution across the four sites (Figure 3.1B and Table 3.2). The Shannon and Simpson indexes indicated lower archaeal diversity, as compared to the bacterial one (Figure 3.2 and Table 3.2). This lower richness and diversity of archaea is not specific to PHF as it has been generally observed in various environments globally (Aller and Kemp, 2008), including deep-sea hydrothermal vents (Anderson et al., 2015) and Lost-City (Brazelton et al., 2010b). The rarefaction curves had a near-asymptotic shape (Figure 3.3) suggesting that the sequencing effort was suitable to capture the near-full diversity of this serpentinizing environment.

3.3.2 Bacterial Phylogenetic Composition in PHF

The bacterial community composition of the PHF sites encompassed 22 different phyla (including six candidate phyla) reflecting a high biodiversity. The phylum *Proteobacteria* was preponderant in the intertidal BdJ (78.4% of sequences, 244 OTUs), while the three submarine sites were mainly composed of *Chloroflexi* (50.7-54.8%, 150-161 OTUs) and, to a lesser extent, of *Proteobacteria* members (21.1-27.4%, 250-603 OTUs) (Figure 3.4). Remarkably, all other bacterial phyla shared by all the PHF sites were essentially recovered at much lower abundance. These bacterial phyla comprised *Cyanobacteria* (0.2-16.0%), *Actinobacteria* (0.2-4.8%), *Bacteroidetes* (0.2-3.1%), *Firmicutes* (0.1-5.3%), *Planctomycetes* (0.2-1.6%), and the candidate phylum *Parcubacteria* (formerly candidate division OD1) (0.01-0.6%). By contrast, the *Acetothermia*, *Aci-*

Table 3.3 – Taxonomy details of abundant bacterial OTUs. The representative sequences of each OTU were blasted against RefSeq database to identify their closest relatives.

OTU ID	ST12	BdJ	ST07	ST09	Taxonomy	Closest relatives retrieved from RefSeq database		
						Bacterial strain	Genbank ID	% id
OTU149	17	0	62	249	p_Acidobacteria; c_Acidobacteria; o_Subgroup 11; f_uncultured Acidobacteria bacterium; g_uncultured Acidobacteria bacterium; s_uncultured Acidobacteria bacterium	Silvibacterium bohemicum	NR_135209.1	0.82
OTU167	6	0	33	391	p_Actinobacteria; c_Acidimicrobiia; o_Acidimicrobiales; f_Sva0996 marine group; g_uncultured Acidimicrobiales bacterium; s_uncultured Acidimicrobiales bacterium	Aciditerrimonas ferrireducens JCM 15389	NR_112972.1	0.9
OTU1388	4	0	0	140	p_Bacteroidetes; c_Sphingobacteriia; o_Sphingobacteriales	Crenotalea thermophila	NR_125473.1	0.82
OTU753	0	0	402	0	p_Chloroflexi; c_Anaerolineae; o_Anaerolineales; f_Anaerolineaceae; g_uncultured; s_uncultured Chloroflexus sp.	Saccharomonospora viridis	NR_074713.2	0.81
OTU793	0	0	175	9	p_Chloroflexi; c_Anaerolineae; o_Anaerolineales; f_Anaerolineaceae; g_uncultured; s_uncultured Chloroflexus sp.	Nocardia camponoti	NR_148835.1	0.83
OTU1317	0	0	181	0	p_Chloroflexi; c_Anaerolineae; o_Anaerolineales; f_Anaerolineaceae; g_uncultured; s_uncultured Chloroflexus sp.	Ornatilinea apprima	NR_109544.1	0.8
OTU1604	0	0	162	0	p_Chloroflexi; c_Anaerolineae; o_Anaerolineales; f_Anaerolineaceae; g_uncultured; s_uncultured Chloroflexus sp.	Saccharomonospora viridis	NR_074713.2	0.81
OTU641	11	0	177	153	p_Chloroflexi; c_Caldilineae; o_Caldilineales; f_Caldilineaceae; g_uncultured; s_uncultured Chloroflexus sp.	Litorilinea aerophila	NR_132330.1	0.84
OTU1432	25	0	682	5	p_Chloroflexi; c_Dehalococcoidia; o_GIF9	Dehalogenimonas lykanthroporepellens	NR_074337.1	0.81
OTU1706	891	0	345	25	p_Chloroflexi; c_Dehalococcoidia; o_GIF9	Dehalogenimonas lykanthroporepellens	NR_074337.1	0.81
OTU1947	8	0	557	23	p_Chloroflexi; c_Dehalococcoidia; o_GIF9	Dehalogenimonas alkenigignens	NR_109657.1	0.82
OTU1117	2915	0	504	1122	p_Chloroflexi; c_Dehalococcoidia; o_MSBL5	Lihuaxuella thermophila	NR_126245.2	0.8
OTU679	295	0	14	27	p_Chloroflexi; c_Dehalococcoidia; o_MSBL5; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	Acidimicrobium ferrooxidans	NR_074390.1	0.8
OTU915	15	0	442	1346	p_Chloroflexi; c_Dehalococcoidia; o_MSBL5; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	Acidimicrobium ferrooxidans	NR_074390.1	0.81
OTU1039	25	1	225	696	p_Chloroflexi; c_Dehalococcoidia; o_MSBL5; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	Acidimicrobium ferrooxidans	NR_074390.1	0.81
OTU1050	0	0	28	170	p_Chloroflexi; c_Dehalococcoidia; o_MSBL5; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	Lihuaxuella thermophila	NR_126245.2	0.81
OTU1275	77	0	99	135	p_Chloroflexi; c_Dehalococcoidia; o_MSBL5; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	Acidimicrobium ferrooxidans	NR_074390.1	0.82
OTU1529	5	0	55	222	p_Chloroflexi; c_Dehalococcoidia; o_MSBL5; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	Lihuaxuella thermophila	NR_126245.2	0.81
OTU161	35	0	67	155	p_Chloroflexi; c_SAR202 clade	Roseiflexus castenholzii	NR_074188.1	0.81
OTU300	32	0	333	14	p_Chloroflexi; c_SAR202 clade	Roseiflexus castenholzii	NR_074188.1	0.81
OTU1534	0	153	0	0	p_Cyanobacteria; c_Cyanobacteria; o_SubsectionIII; f_FamilyI	Halomicronema excentricum	NR_114591.1	0.9
OTU1127	0	150	0	0	p_Cyanobacteria; c_Cyanobacteria; o_SubsectionIII; f_FamilyI; g_Phormidium	Synechococcus elongatus	NR_074309.1	0.91
OTU968	0	120	0	0	p_Cyanobacteria; c_Cyanobacteria; o_SubsectionIII; f_FamilyI; g_Prochlorothrix; s_uncultured marine bacterium	Microcystis aeruginosa	NR_074314.1	0.92
OTU1474	0	252	0	0	p_Cyanobacteria; c_Cyanobacteria; o_SubsectionIII; f_FamilyI; g_Prochlorothrix; s_uncultured marine bacterium	Microcystis aeruginosa	NR_074314.1	0.92
OTU1266	0	151	0	0	p_Deinococcus-Thermus; c_Deinococci; o_Thermales; f_Thermaceae; g_Meiothermus; s_uncultured bacterium	Meiothermus hypogaeus NBRC 106114	NR_113226.1	0.95
OTU796	9	0	217	411	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Syntrophomonadaceae; g_uncultured; s_uncultured Firmicutes bacterium	Natranaerobaculum magadiense	NR_135713.1	0.84
OTU135	3	0	144	18	p_Gemmatimonadetes; c_Gemmatimonadetes; o_PAUC43f marine benthic group; f_uncultured sponge symbiont PAUC51f; g_uncultured sponge symbiont PAUC51f; s_uncultured sponge symbiont PAUC51f	Eilatimonas milleporae	NR_109076.1	0.82
OTU234	16	0	115	115	p_Gemmatimonadetes; c_Gemmatimonadetes; o_PAUC43f marine benthic group; f_uncultured sponge symbiont PAUC51f; g_uncultured sponge symbiont PAUC51f; s_uncultured sponge symbiont PAUC51f	Eilatimonas milleporae	NR_109076.1	0.81
OTU266	3	2	118	76	p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Rhodobiaceae; g_Methyloceanibacter; s_uncultured bacterium	Methyloceanibacter caenitepidi	NR_125465.1	0.98
OTU594	1	1	113	61	p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Rhodobiaceae; g_Methyloceanibacter; s_uncultured bacterium	Methyloceanibacter caenitepidi	NR_125465.1	0.98
OTU680	0	114	0	0	p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae	Pseudooceanicola antarcticus	NR_134107.1	0.96
OTU1285	0	271	0	0	p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae	Paracoccus siganidrum	NR_118463.1	0.96
OTU1946	0	217	0	0	p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae	Loktanella litorea	NR_118329.1	0.99
OTU279	30	1340	38	16	p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae; g_Roseibaca	Roseinatronobacter monicus	NR_043914.1	0.96
OTU1806	4	1037	6	8	p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae; g_Roseibaca	Paenirhodobacter enshiensis	NR_125604.1	0.95
OTU153	4	1239	29	0	p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae; g_Roseibaca; s_uncultured bacterium	Roseinatronobacter monicus	NR_043914.1	0.96
OTU1298	2	417	2	0	p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae; g_Roseibaca; s_uncultured bacterium	Roseinatronobacter monicus	NR_043914.1	0.96
OTU1067	0	122	0	0	p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae; g_Tropicimonas; s_uncultured bacterium	Dinoroseobacter shibae	NR_074166.1	0.95
OTU137	0	0	114	0	p_Proteobacteria; c_Alphaproteobacteria; o_Rickettsiales; f_SAR116 clade	Nisaea nitritireducens	NR_043924.1	0.91
OTU429	0	0	122	0	p_Proteobacteria; c_Alphaproteobacteria; o_Rickettsiales; f_SAR116 clade	Nisaea nitritireducens	NR_043924.1	0.91
OTU607	7	654	0	0	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae	Hydrogenophaga defluvii	NR_029024.1	0.98
OTU763	4	230	0	0	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae	Hydrogenophaga defluvii	NR_029024.1	0.98
OTU933	0	0	1	123	p_Proteobacteria; c_Deltaproteobacteria; o_Bdellovibrionales; f_Bdellovibrionaceae; g_Bdellovibrio	Bdellovibrio exovorus	NR_102876.1	0.81
OTU1347	10	155	15	26	p_Proteobacteria; c_Deltaproteobacteria; o_Desulfovibrionales; f_Desulfonatronaceae; g_Desulfonatronum; s_uncultured bacterium	Desulfonatronum alkalitolerans	NR_108631.1	0.97

dobacteria, *Gemmatimonadetes*, and *Nitrospirae* phyla were only retrieved from submarine sites, representing 7.4% of the total sequences. Relative abundances of some bacterial phyla detected in our study were not in agreement with those reported by sequencing clone libraries (Postec et al., 2015, Quéméneur et al., 2014). This was the case for the *Firmicutes* and the *Parcubacteria*, which were found in much lower proportion in this study. Nevertheless, significant shifts in phyla proportions have already been observed in the same chimney edifice over a 6 year survey (Postec et al., 2015). Our pyrosequencing approach also allowed the identification of 6 novel phyla that had not been previously detected. Members of all these phyla, i.e., *Fibrobacteres*, *Lentisphaerae*, *Marinimicrobia*, PAUC34f, *Saccharibacteria*, and *Verrucomicrobia*, were found among the large diversity of the ST12 site. On the contrary, several phyla previously found in PHF: *Atribacteria*, *Microgeomates*, *Omnitrophica*, and candidate division NPL-UPA2 (Mei et al., 2016b, Pisapia et al., 2017, Postec et al., 2015, Quéméneur et al., 2014) were not represented in this study. This discrepancy may be explained by the use of different PCR primers targeting different regions of the 16S rRNA gene of bacteria and archaea. It may also be the consequence of the ongoing evolution of the new generation sequencing technology (i.e., 454 Roche vs. Illumina). That raises a major issue regarding the lack of standards for 16S rRNA gene surveys, which reduces the reproducibility and complicates comparisons between studies (Escobar-Zepeda et al., 2015). Additionally, the sampling area of the chimneys (either central or more peripheral) could also explain these taxonomic discrepancies. Indeed, the highly porous structure of the chimney interiors is exposed to steep physicochemical gradients (in terms of salinity, pH, redox potential, nutrients) along their transverse axis, defining microhabitats colonized by contrasted microbial communities.

3.3.3 Archaeal Phylogenetic Composition in PHF

The archaeal community of the PHF sites mainly comprised potential methanogens and autotrophic ammonia-oxidizer representatives among *Euryarchaeota* and *Thaumarchaeota* phyla, respectively. The BdJ and ST12 sites were dominated by members of *Methanosarcinales* (99.8% of sequences, 38 OTUs and 94.6%, 44 OTUs, respectively) within the class *Methanomicrobia* (Figure 3.5). All abundant and ubiquitous OTUs of *Methanosarcinales* were similar to the sequences from other serpentinite-hosted ecosystems, as shown by OTU-based phylogeny (Figure 3.6). The first set of OTUs constituted a sister group of Lost-City Methanosarcinales (LCMS) (Schrenk et al., 2004) whereas the remaining OTUs were closely associated with the sequence of The Cedars Methanosarcinales (TCMS) (phylotype Ced_A01) (Suzuki et al., 2013). Although the present study reported a more accurate diversity than the previous studies based on clone libraries, no new phylotypes were identified within the *Methanosarcinales* group. Interestingly, although first detected in a terrestrial serpentinizing system (The Cedars), TCMS-like archaea were detected at lower abundance in the intertidal BdJ site than in the submarine sites of PHF. Moreover, here the overall abundance of the LCMS-like was two-fold higher than the abundance of the TCMS-like (Figure 3.6). This result is in opposition with the ratio previ-

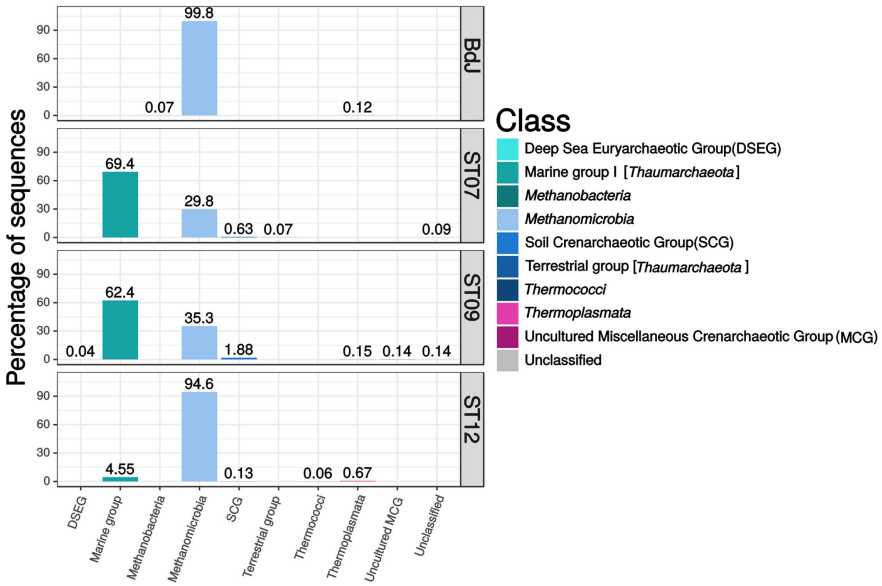


Figure 3.5 – Class-level archaeal diversity in PHF sites based on 16S rRNA gene sequences (V1-V3 region). The relative abundances were estimated for the four chimneys collected in BdJ, ST07, ST09, and ST12 sites.

ously observed in clone libraries and thus contradicts the hypothesis that the predominance of TCMS-like archaea in PHF was due to the low salinity of fluids (of meteoric origin such as in The Cedars) (Postec et al., 2015). Finally, these two specific phylotypes of *Methanosarcinales* were ubiquitous in all PHF sites. In contrast, the ST07 and ST09 sites were dominated by the *Thaumarchaeota* Marine Group I (69.4%, 179 OTUs and 62.4%, 194 OTUs, respectively), which was detected in all submarine sites despite its low abundance in the ST12 site (Figure 3.5). The Marine Group I is frequently observed in the global ocean (Auguet et al., 2010, Fernàndez-Guerra and Casamayor, 2012), which may explain why they are found only in submarine sites of PHF. In addition, the *Thaumarchaeota* Soil Crenarchaeotic Group was also detected in the submarine sites at very low-abundance level (0.1 to 1.9%, 3 to 6 OTUs) and may originate from the surrounding hills having been carried away by run-off waters. Other less represented archaeal classes belonged to the *Thermoplasmata* (mainly Marine Group II) (0.1-0.7%, 7-18 OTUs), *Methanobacteria* (0.1%, 2 OTUs at BdJ), Terrestrial Group of *Thaumarchaeota* (0.1%, 2 OTUs in ST07) and presumed thermophiles of the *Thermococci* class (0.1%, 1 OTU at ST12) (Figure 3.5). Finally, the remaining OTUs were affiliated to the Deep Sea Euryarchaeotic Group within the *Aenigmarchaeota* phylum and to the Miscellaneous Crenarchaeotic Group that were not previously detected in PHF. The archaea of the Miscellaneous Crenarchaeotic Group are widespread in marine sediment and are essentially retrieved from anoxic environments such as deep oceanic subsurface sediments or mud volcanoes (Kubo et al., 2012). Since the temperature of PHF fluids measured at the chimney outlets did not exceed 43°C, the presence of archaea closely related to hyperthermophilic species (e.g., *Thermococcus* spp.) originating from deep-sea environments was surprising. Despite their low abundance, such representatives of the *Thermococcales* were, however, previously detected in PHF, even in clone libraries, and the high GC content of their sequence strongly suggested they were bona fide

hyperthermophiles (Postec et al., 2015). It is thus likely that they originated from subsurface hyperthermophilic microbial communities fed by serpentinization reactions and carried up to the surface by the fluid as previously suggested (Postec et al., 2015).

Table 3.4 – Taxonomy details of abundant archaeal OTUs. The affiliations at the genus level of OTU155, OTU192, OTU294, OTU174, OTU280, OTU303, OTU16, OTU184, OTU193 were automatically generated by the RDP pipeline. They slightly differed from the phylogenetic annotations presented in Figure 3.7 due to some discrepancies of the *Methanosarcinales* classification in the Silva database. The representative sequences of each OTU were blasted against RefSeq database to identify their closest relatives.

OTU ID	ST12	BdJ	ST07	ST09	Taxonomy	Closest relatives retrieved from RefSeq database		
						Bacterial strain	Genbank ID	% id
OTU155	805	32	797	2467	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methanosarcinaceae	Methanolobus psychrophilus R15	NR_102921.1	0.92
OTU192	0	0	141	257	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methanosarcinaceae	Methanolobus psychrophilus R15	NR_115986.1	0.9
OTU294	2237	78	74	68	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methanosarcinaceae; g_ANME-3	Methanolobus psychrophilus R15	NR_115986.1	0.92
OTU174	1105	42	7080	7281	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methanosarcinaceae; g_ANME-3; s_uncultured archaeon	Methanolobus psychrophilus R15	NR_102921.1	0.91
OTU280	0	0	6609	4652	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methanosarcinaceae; g_ANME-3; s_uncultured archaeon	Methanolobus psychrophilus R15	NR_115986.1	0.92
OTU303	634	17	1501	3288	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methanosarcinaceae; g_ANME-3; s_uncultured archaeon	Methanomethylovorans hollandica DSM 15978	NR_102454.1	0.91
OTU16	1428	4926	1565	662	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methermicoccaceae; g_Methermicoccus; s_uncultured bacterium	Methanoculleus sediminis	NR_136474.1	0.88
OTU184	86	266	77	45	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methermicoccaceae; g_Methermicoccus; s_uncultured bacterium	Methanoculleus sediminis	NR_136474.1	0.88
OTU193	153	389	115	102	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methermicoccaceae; g_Methermicoccus; s_uncultured bacterium	Methanosaeta harundinacea	NR_043203.1	0.87
OTU20	238	93	2	0	p_Thaumarchaeota; c_Marine Group I	Nitrosopumilus maritimus SCM1	NR_102913.1	0.92
OTU61	388	169	4	0	p_Thaumarchaeota; c_Marine Group I	Nitrosopumilus maritimus SCM1	NR_102913.1	0.92
OTU78	268	87	4	0	p_Thaumarchaeota; c_Marine Group I	Nitrosopumilus maritimus SCM1	NR_102913.1	0.92
OTU2	228	157	34	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.98
OTU35	232	152	39	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.98
OTU56	432	254	15	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.97
OTU109	1070	594	119	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.98
OTU114	444	445	61	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.98
OTU128	245	86	35	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.98
OTU149	239	133	94	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.99
OTU172	270	103	8	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.95
OTU188	124	335	13	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.97
OTU198	235	48	8	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus	Nitrosopumilus maritimus SCM1	NR_102913.1	0.98
OTU31	1103	3746	79	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus; s_uncultured Nitrosopumilales archaeon	Nitrosopumilus maritimus SCM1	NR_102913.1	0.96
OTU86	817	1388	37	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus; s_uncultured Nitrosopumilales archaeon	Nitrosopumilus maritimus SCM1	NR_102913.1	0.96
OTU113	684	2056	32	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus; s_uncultured Nitrosopumilales archaeon	Nitrosopumilus maritimus SCM1	NR_102913.1	0.96
OTU146	121	351	14	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus; s_uncultured Nitrosopumilales archaeon	Nitrosopumilus maritimus SCM1	NR_102913.1	0.97
OTU194	1244	146	27	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus; s_uncultured Nitrosopumilales archaeon	Nitrosopumilus maritimus SCM1	NR_102913.1	0.95
OTU272	945	165	23	0	p_Thaumarchaeota; c_Marine Group I; o_Unknown Order; f_Unknown Family; g_Candidatus Nitrosopumilus; s_uncultured Nitrosopumilales archaeon	Nitrosopumilus maritimus SCM1	NR_102913.1	0.95

3.3.4 Beta-Diversity and Abundant OTUs across the PHF Sites

The bacteria taxonomic profiles observed for the three submarine sites were very similar, while the one observed for the intertidal site (BdJ) differed significantly (Figure 3.7A). This pattern of distribution, previously observed in a study on PHF (Quéméneur et al., 2014), may be ex-

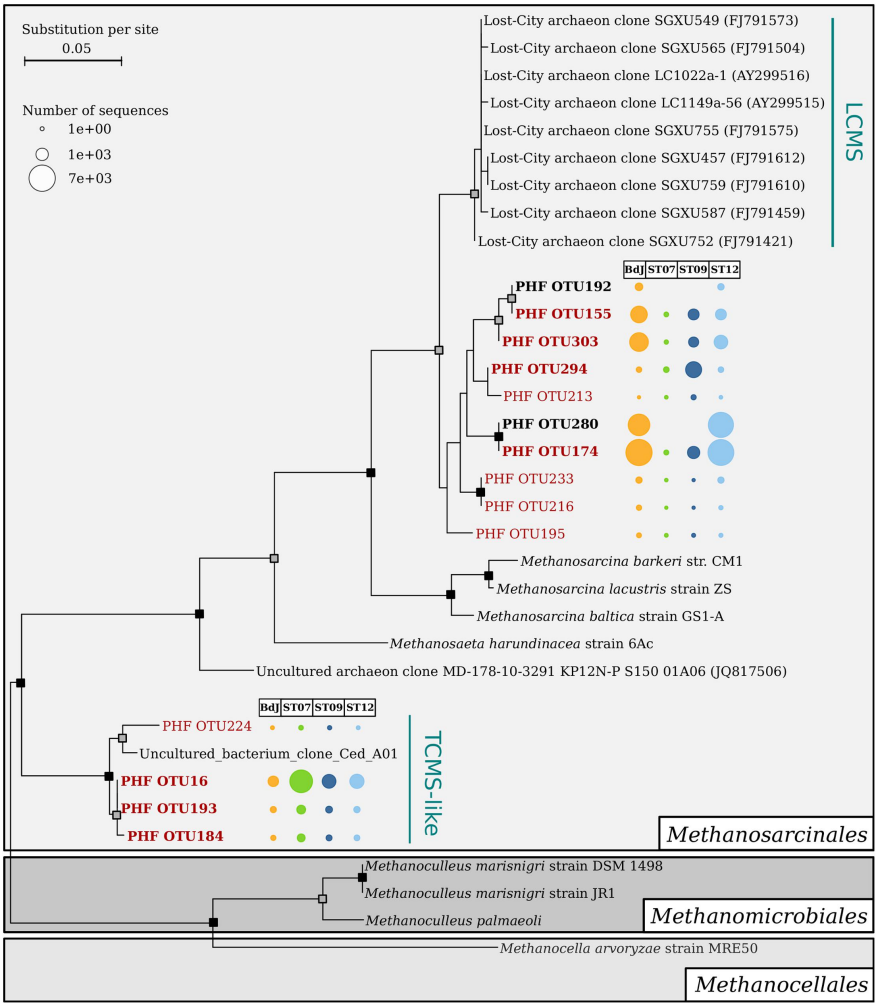


Figure 3.6 – Maximum likelihood phylogenetic tree of V1-V3 archaeal 16S rRNA sequences affiliated to the order *Methanosarcinales*. Bootstrap support values are indicated on nodes with a gray or black square for support greater than 70% or 90%, respectively. Scale bar indicates substitutions per site. The *Methanosarcinales* sequences are marked in bold if they were abundant in at least one site of PHF, and red font color indicates ubiquitous operational taxonomic units (OTUs; found in the four chimneys of PHF). The bubble chart corresponds to the contingency table of OTUs. Uncultured *Methanosarcinales* clusters are abbreviated as follows: LCMS, Lost-City Methanosarcinales, and TCMS, The Cedars Methanosarcinales.

plained by the influence of the marine environment regime (intertidal vs. submarine) variation in the composition of the end-member fluids (Monnin et al., 2014) or the resulting differences observed in the mineralogy of the carbonate chimneys (Pisapia et al., 2017). The relationship between species and sampling sites was mainly driven by a few dominant OTUs in the ordination plot (Figure 3.7A). Among them, five abundant OTUs in submarine sites were assigned to the class *Dehalococcoidia* (DEH, phylum *Chloroflexi*) and were related to clones retrieved from the serpentinizing springs of The Cedars (Suzuki et al., 2013). Two of these dominant OTUs (#915 and #1117) were assigned to the candidate order MSBL5 (Mediterranean Sea Brine Lake group 5) and three (#1432, #1706, and #1947) to the division GIF9 (Groundwater InFlow clone 9) (Table 3.3). The candidate orders MSBL5 and GIF9 were originally detected in a deep-sea anoxic hypersaline lake (Daffonchio et al., 2006) and in a contaminated groundwater-fed

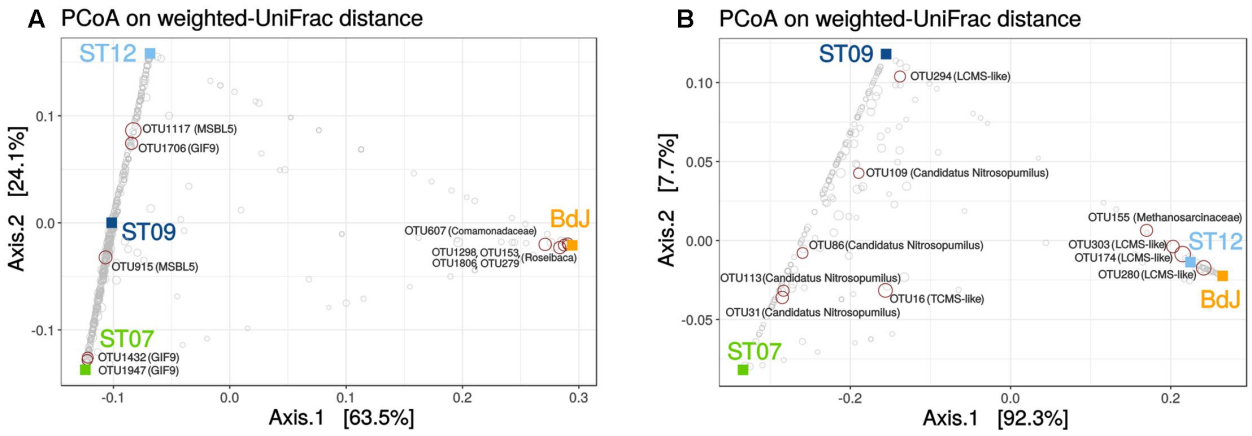


Figure 3.7 – Beta diversity of bacterial (A) and archaeal (B) communities of PHF. PCoA analysis was carried out using the weighted UniFrac metric. Squares indicate PHF sites and circles correspond to OTUs. Taxonomic annotation was added to the top ten OTUs explaining this partition (red circles). Lost-City Methanosarcinales and The Cedars Methanosarcinales are abbreviated as LCMS and TCMS, respectively.

bio reactor (Alfreider et al., 2002), respectively. These five dominant OTUs are quite divergent from any cultivated species (the closest similarity is found for OTU1947 with *Dehalogenimonas alkenigignens*, at 82% identity). The metabolic capabilities of these candidate groups are still largely unknown. Even for the entire class *Dehalococcoidia*, our knowledge of metabolism is essentially limited to that of the few cultivated representatives, which have the ability to oxidize hydrogen and reduce organohalogenic compounds (Biderre-Petit et al., 2016, Kaster et al., 2014). However, a single cell sequencing of a *Dehalococcoidia* phylotype, abundant in marine sediments, suggested that this bacterium may replace organohalides by dimethyl sulfoxide as a terminal electron acceptor (Wasmund et al., 2013). In the BdJ site, 4 OTUs (#1298, #153, #1806, #279) representing more than a third of all sequences (Figure 3.7A) were assigned to the haloalkaliphilic genus *Roseibaca* (*Alphaproteobacteria*), a sister genus of the phototrophic anaerobic sulfur-oxidizing *Rhodobaca*, previously detected as dominant in BdJ (Mei et al., 2016b). Representatives of the *Roseibaca* genus are known to be rose-colored (Labrenz et al., 2009) and could in part explain the dark rose color of the biofilm covering the exteriors of BdJ chimneys. The BdJ site also contained two dominant OTUs (#607 and #763) affiliated to the *Comamonadaceae* family (*Betaproteobacteria*) showing 98% identity with *Hydrogenophaga defluvii* (Table 3.3). Members of the genus *Hydrogenophaga* are chemoorganotrophic or chemolithoautotrophic, thriving by oxidation of hydrogen under microaerophilic conditions (Willems, 2014). Despite the potential involvement of *Firmicutes* in the organomineralization of carbonates chimneys in PHF (Pisapia et al., 2017), only one abundant OTU of this phylum was identified in submarine sites and assigned to uncultured anaerobic *Syntrophomonadaceae*. This OTU, however, differed from the one (KM207235 HPst091-1-1) reported to be dominant in juvenile chimneys of PHF (Pisapia et al., 2017), most likely because of the differences of age of the chimneys studied (i.e. mature vs. juvenile). Members of this family are anaerobes that usually grow in syntrophic association with hydrogenotrophic and formate-utilizing methanogens by thriving energy from oxidation of carboxylic acids (Sobieraj, M., and Boone, 2006). These anaerobic oxidations are

thermodynamically feasible only when the H_2 or formate produced is concomitantly consumed by methanogens or other syntrophic partner. Recently, syntrophic associations were shown to permit the conversion of organic acids and alcohols to methane under extremely haloalkaline conditions (Sorokin et al., 2016). In PHF, such syntrophic associations between uncultivated members of the *Syntrophomonadaceae* and the *Methanosarcinales* could be a relevant strategy to overcome energetic limitations, due to the high pH conditions and the lack of electron acceptors. For archaea, no differential segregation of OTUs between intertidal and submarine sites could be observed, in contrast to the trend previously reported (Quéméneur et al., 2014). Nevertheless, the beta-diversity analysis linked BdJ to ST12 (Figure 3.7B), with both sites harboring abundant OTUs related to the LCMS (their annotation was refined through phylogenetic analysis, Figure 3.6), whereas ST09 clustered with ST07 because they shared abundant OTUs related to Candidatus *Nitrosopumilus* (*Thaumarchaeota*) and numerous OTUs of Marine Group I (Figure 3.7B and Table 3.4). In light of these beta-diversity results, the archaeal community seemed less site-dependent than the bacterial one, as exemplified by the two groups of *Methanosarcinales* OTUs (LCMS and TCMS) found in both intertidal and submarine serpentinizing systems of PHF sites.

3.3.5 Comparison between Phylotypes of Serpentinizing Environments

A comparison of microbial communities across six geographically distant serpentinizing systems was carried out at the OTU level (at 97% identity threshold) used as a proxy of the prokaryotic species that is the most commonly found unit in microbial biogeographic studies. For practical reasons, this comparison was limited to the fraction of known microorganisms (i.e., those which could be assigned to a taxon in the Silva database), thus the results cannot be extrapolated to the whole prokaryotic community. Despite this limitation, some biogeographic patterns could be observed. The general OTU composition appeared to be quite different from one serpentinizing system to another and no OTU was common to all sites, but several related phylotypes were recovered from more than two systems; among them, a few were detected exclusively in serpentinizing systems. The distribution pattern of PHF phylotypes in other geographic sites and its implications is discussed below.

OTUs Belonging to Abundant Taxa Commonly Detected in Serpentinite-Hosted Ecosystems

It mainly concerns the *Betaproteobacteria* and the *Clostridia* (Schrenk et al., 2013). The abundance of *Betaproteobacteria* often exceeded 20% in the terrestrial systems studied herein (Daae et al., 2013, Quéméneur et al., 2015, Suzuki et al., 2013, Tiago and Veríssimo, 2013). Four *Betaproteobacteria* OTUs, found in both PHF sites and continental serpentinizing springs, were assigned to the H_2-consuming genus *Hydrogenophaga* (Figure 3.8). Although not exclusive of serpentinizing systems, these OTUs were otherwise distributed in similar geothermal environments such as deep hydrothermal groundwater and sulphidic springs (Table 3.5). This suggests that similar geochemical settings (presence of H_2 rich fluids and oxic-anoxic interfaces) may

Table 3.5 – Taxonomy details of bacterial OTUs shared at least by one sample from PHF (this study, Pisapia et al., 2017, Postec et al., 2015, Quéméneur et al., 2014) and by one sample from another serpentinizing system among Cabeço de Vide Aquifer (Tiago and Veríssimo, 2013), The Cedars (Suzuki et al., 2013), Voltri ophiolite (Quéméneur et al., 2015), Leka ophiolite (Daae et al., 2013) and Lost-City (Brazelton et al., 2006, Schrenk et al., 2004). If the representative sequence of an OTU was isolated in a serpentinizing ecosystem, the last two columns indicate the closest relative of this sequence, identified in another kind of environment (i.e. not influenced by serpentinization reaction).

OTUs and their distribution in studied serpentinizing systems					Closest relative in ecosystems that are not reported as serpentinizing sites	
GenbankID *	Taxonomy (p_Phylum; c_Class; o_Order; f_Family; g_Genus; s_Species)	PHF sites	Other ecosystems	Representative sequence source	Genbank ID (coverage, id)	Environment
AY741058.1	p_Actinobacteria; c_Actinobacteria; o_Propionibacteriales; f_Propionibacteriaceae; g_Propionibacterium; s_Propionibacterium acnes	ST12	CVA, Lost-City	Pericardial fluid		
AM777981.1	p_Actinobacteria; c_Coriobacteriia; o_Coriobacteriales; f_Coriobacteriaceae; g_uncultured; s_uncultured bacterium	ST07	CVA	Serpentinizing site	AB936639.1 (97%, 95%)	River sediment
GU056099.1	p_Actinobacteria; c_Thermoleophilia; o_Solirubrobacterales; f_480-2; g_uncultured bacterium; s_uncultured bacterium	BdJ	Voltri	Gas and oil field		
KC574890.1	p_Chloroflexi; c_Dehalococcoidia; o_MSBL5; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	ST12, ST09, ST08, ST07	The Cedars	Serpentinizing site	JN207189.1 (100%, 92%)	Sediment of Namako-ike lake
JN123498.1	p_Chloroflexi; c_Dehalococcoidia; o_vadinBA26; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	ST12	The Cedars	Seafloor sediment		
HQ433568.1	p_Cyanobacteria; c_Chloroplast; o_uncultured bacterium; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	ST12	Voltri	Shrimp pond		
AB757744.1	p_Cyanobacteria; c_Cyanobacteria; o_SubsectionI; f_FamilyI; g_Synechococcus; s_uncultured bacterium	BdJ	The Cedars	Hot Spring Water		
KF912961.1	p_Deinococcus-Thermus; c_Deinococci; o_Deinococcales; f_Trueperaceae; g_Truepera; s_uncultured bacterium	ST09, ST07, BdJ	Voltri, CVA	Chromium soil contaminated		
AM777965.1	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Syntrophomonadaceae; g_Dethiobacter; s_uncultured bacterium	ST09	CVA	Serpentinizing site	DQ088769.1 (93%, 98%)	Deep gold mine
DQ088764.1	p_Firmicutes; c_Clostridia; o_D8A-2; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	ST12, ST09, ST08, ST07	CVA	Deep gold mine		
AM778006.1†	p_Firmicutes; c_Clostridia; o_Thermoanaerobacterales; f_SRB2; g_uncultured bacterium; s_uncultured bacterium	ST11	CVA, The Cedars	Serpentinizing site	AB476673.1 (97%, 98%)	Sulfur containing freshwater source
AB476673.1	p_Firmicutes; c_Clostridia; o_Thermoanaerobacterales; f_SRB2; g_uncultured Firmicutes bacterium; s_uncultured Firmicutes bacterium	ST07	CVA	Sulfur containing freshwater source		
LN561460.1	p_Parcubacteria; c_uncultured bacterium; o_uncultured bacterium; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	ST07	Voltri	Refuse dump		
FJ612223.1	p_Proteobacteria; c_Alphaproteobacteria; o_Caulobacterales; f_Hyphomonadaceae; g_Hyphomonas; s_uncultured bacterium	BdJ	Voltri	Lake water		
DQ125521.1	p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Hyphomicrobiaceae; g_uncultured; s_uncultured bacterium	ST07	Voltri	Uranium contaminated soil		
JF119689.1	p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Hyphomicrobiaceae; g_Hyphomicrobium; s_uncultured bacterium	ST09	Voltri	Skin		
KF182250.1	p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Bradyrhizobiaceae; g_Bradyrhizobium; s_uncultured bacterium	ST12	Leka	Coal gangue dump soil		
JN038233.1	p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Erythrobacteraceae; g_Porphyrobacter; s_uncultured Sphingomonadales bacterium	BdJ	Voltri	Petroleum contaminated soil		
JX521390.1	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae; g_Hydrogenophaga; s_uncultured bacterium	BdJ	The Cedars	Terrestrial sulfidic spring		
AF385534.1	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae; g_Schlegelella; s_Leptothrix sp. oral clone AW043	ST09, ST07	CVA	Noma lesions		
KF300927.1	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae; g_Hydrogenophaga; s_uncultured bacterium	BdJ	Voltri, Leka	Yellow River sediment		
DQ256326.1	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae; g_Hydrogenophaga; s_uncultured bacterium	ST07	CVA, The Cedars	Subsurface water of the Kalahari Shield		
KF441648.1	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae; g_Hydrogenophaga; s_Hydrogenophaga sp. 7A-385	ST12, ST07, BdJ	Leka	Urgeirica mine		
EU030486.1	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Burkholderiaceae; g_Ralstonias_uncultured bacterium	ST12, ST09	Lost-City	Antarctic ice sheet		
KF733685.1	p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Burkholderiaceae; g_Burkholderia; s_Burkholderia ambifaria	ST12, ST07	CVA	Oral cavity		
EU470514.1	p_Proteobacteria; c_Gammaproteobacteria; o_B38; f_uncultured bacterium; g_uncultured bacterium; s_uncultured bacterium	ST12	CVA	Red panda feces		
KJ127981.1	p_Proteobacteria; c_Gammaproteobacteria; o_Chromatiales; f_Chromatiaceae; g_Nitrosococcus; s_uncultured Ectothiorhodospiraceae bacterium	BdJ	Voltri	Slime from paper machine		
HQ739644.1	p_Proteobacteria; c_Gammaproteobacteria; o_Enterobacteriales; f_Enterobacteriaceae; g_Escherichia-Shigella; s_uncultured bacterium	ST12	CVA	Gastrointestinal specimens		
KF475873.1	p_Proteobacteria; c_Gammaproteobacteria; o_Pseudomonadales; f_Pseudomonadaceae; g_Pseudomonas; s_Pseudomonas koreensis	ST07	Voltri	Rhizosphere soil		
AJ628163.1	p_Proteobacteria; c_Gammaproteobacteria; o_Pseudomonadales; f_Pseudomonadaceae; g_Pseudomonas; s_Pseudomonas pseudoalcaligenes	ST12, ST07	Voltri, CVA	Guadalquivir River		
AM778014.1	p_Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae; g_Silanimonas; s_uncultured bacterium	BdJ	CVA	Serpentinizing site	KX348537.1 (99%, 98%)	Denitrifying process
KF511881.1	p_Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae; g_Silanimonas; s_uncultured bacterium	BdJ	Voltri, CVA, The Cedars	Chromium contaminated soil		
FJ405365.1	p_Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae; g_Stenotrophomonas; s_Xanthomonas sp. SPf	ST12	CVA	Sweet potato plants		
KC238323.1	p_Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_JTB255 marine benthic group; g_uncultured bacterium; s_uncultured bacterium	ST12	Lost-City	Bio-filtration of seawater		

* Accession number of representative sequences of OTUs, † This OTU was also abundantly detected in CROMO. (Twing et al., 2017)

have selected *Hydrogenophaga* species which metabolic capabilities are well adapted to these particular conditions. Indeed, 16S rRNA gene and, in some cases, related [NiFe]-hydrogenases genes of *Hydrogenophaga* spp. were previously shown to be abundant in the serpentinizing-hosted systems listed in the comparison and others such as the springs of the Tablelands Ophiolite (Brazelton et al., 2012) or the Samail Ophiolite (Rempfert et al., 2017). It is noteworthy that no OTU representative of its sister genus *Serpentinomonas* was identified in PHF. These hyperalkaliphilic hydrogenotrophs isolated from The Cedars (Suzuki et al., 2014) have so far been detected only in terrestrial serpentinizing sites where they can be dominant, e.g., one OTU representing up to 14% of the sequences in Voltri (Quéméneur et al., 2015), or even constituting the unique OTU of *Betaproteobacteria*, accounting for 50% of the sequences, in the most alkaline well CSW1.1 (pH12.2) of CROMO but absent in the other wells with lower pH values (Twing et al., 2017). Otherwise, the submarine PHF sites shared a *Ralstonia* OTU (EU030486.1) with Lost-City, emphasizing the diversity of microorganisms that could oxidize hydrogen in the PHF sites.

In the *Clostridia* class, 4 OTUs representing clone sequences of PHF were also detected at CVA, at The Cedars. However, phylotypes closely related to these four OTUs were more generally distributed in other serpentinizing sites including Voltri and CROMO. As noticed for the *Hydrogenophaga* OTUs, these 4 OTUs were restricted to serpentinizing systems or to groundwater. The OTUs AM778006.1 and AB476673.1 were close to the candidate group SRB2 found in CROMO (Twing et al., 2017), while AM777965.1 was affiliated to the *Dethiobacter* genus, of which the only described species, *D. alkaliphilus*, is strictly anaerobic with a respiration metabolism and able to use acetate and sugars with sulfur, polysulfides, or thiosulfate as an electron acceptor (Sorokin et al., 2008). This bacterium has also been shown to be capable of chemolithoautotrophic growth by disproportion of elemental sulfur (a kind of mineral fermentation), a metabolic process which is more favorable under alkaline conditions (Poser et al., 2013) such those encountered in serpentinizing systems.

Regarding the archaeal community, the present comparison included the uncultivated *Methanosarcinales* lineages only detected in serpentinizing ecosystems (Quéméneur et al., 2014, 2015, Schrenk et al., 2004, Suzuki et al., 2013). The originality of PHF is provided by the simultaneous presence of the two *Methanosarcinales* OTUs: the LCMS and TCMS phylotypes (Figure 3.8 and Table 3.6).

OTUs Belonging to Dominant Taxa in PHF Not Generally Distributed in Other Sites

In the PHF, uncultured *Chloroflexi* and members of the class *Alphaproteobacteria* were the main taxa, discriminating the submarine sites from the intertidal one. The large *Chloroflexi* fraction of the PHF submarine sites was not commonly recovered among other serpentinizing systems, except in less active or extinct chimneys of the Lost-City Hydrothermal Field (Brazelton et al., 2010b) and in deep ground water of The Cedars (Suzuki et al., 2013). Additionally, the ecosys-

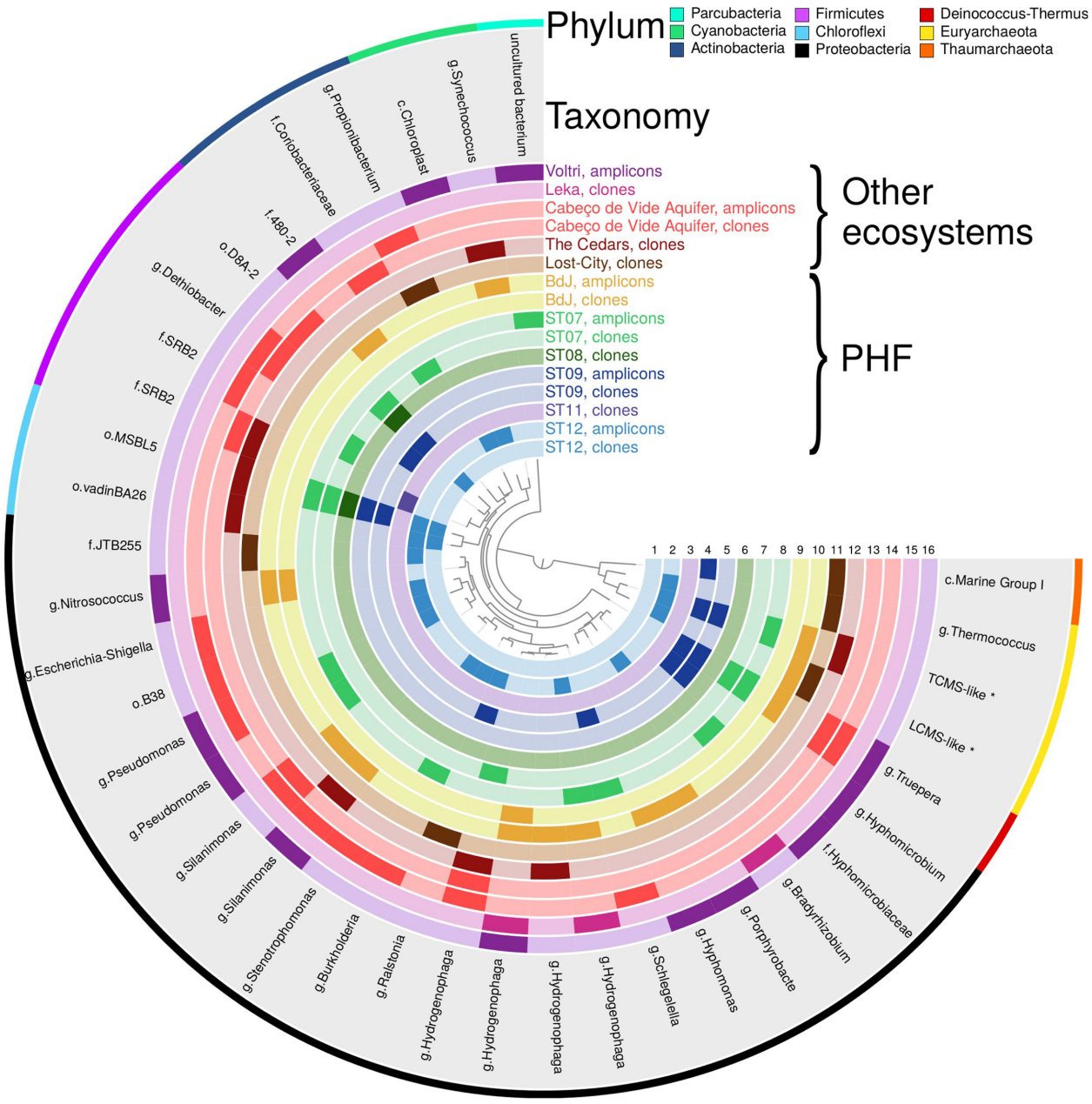


Figure 3.8 – Presence-absence plot of OTUs shared at least by one sample of PHF sites and one sample from another serpentinizing ecosystem. The inner circles, numbered from 1 to 16, correspond to the different samples. The 13 samples of PHF include the circles 2, 5, 8, 10 (this study) and 1, 3, 4, 6, 7, 9 (Pisapia et al., 2017, Postec et al., 2015, Quéméneur et al., 2014). The seven other serpentinizing samples comprise the circles 11 from Lost-City Hydrothermal Field (Brazelton et al., 2006, Schrenk et al., 2004), 12 from The Cedars (Suzuki et al., 2013), 13, 14, from Cabeço de Vide (Tiago and Veríssimo, 2013), 15 from Leka ophiolite (Daae et al., 2013), and 16 from Voltri ophiolite (Quéméneur et al., 2015). Each circle color corresponds to a sampling system and the darker cells indicate the presence of OTUs. The taxonomic affiliation of OTUs is indicated in the gray circle at genus level (or at the lowest taxonomic rank). Letters g, f, and o refer to the ranks genus, family and order, respectively. The outer circle specifies the phylum of each OTU.

Table 3.6 – Taxonomy details of archaeal OTUs shared at least by one sample from PHF (this study, Pisapia et al., 2017, Postec et al., 2015, Quéméneur et al., 2014) and by one sample from another serpentinizing system among: Cabeço de Vide Aquifer (Tiago and Veríssimo, 2013), The Cedars (Suzuki et al., 2013), Voltri ophiolite (Quéméneur et al., 2015), Leka ophiolite (Daae et al., 2013) and Lost-City (Brazelton et al., 2006, Schrenk et al., 2004). If the representative sequence of an OTU was isolated in a serpentinizing ecosystem, the last two columns indicate the closest relative of this sequence, identified in another kind of environment (not influenced by serpentinization reaction).

OTUs and their distribution in studied serpentinizing systems					Closest relative in ecosystems that are not reported as serpentinizing sites	
GenbankID *	Taxonomy (p_Phylum; c_Class; o_Order; f_Family; g_Genus; s_Species)	PHF sites	Other ecosystems	Representative sequence source	Genbank ID (coverage, id)	Environment
FJ791591.1	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methanosarcinaceae; g_ANME-3; s_uncultured archaeon	ST12, BdJ	Lost-City	Serpentinizing site	DQ228615.1 (98%, 95%)	sulfide-microbial incubator
KC574884.1	p_Euryarchaeota; c_Methanomicrobia; o_Methanosarcinales; f_Methermicoccaceae; g_Methermicoccus; s_uncultured bacterium	ST12, ST09, ST07, BdJ	The Cedars	Serpentinizing site	AY835421.2 (99%, 89%)	hydrothermal sediment
JX507272.1	p_Euryarchaeota; c_Thermococci; o_Thermococcales; f_Thermococcaceae; g_Thermococcus; s_uncultured archaeon	ST12	Lost-City	Guaymas Basin Hydrothermal Vent		
AY505046.1	p_Thaumarchaeota; c_Marine Group I; o_uncultured archaeon; f_uncultured archaeon; g_uncultured archaeon; s_uncultured archaeon	ST09	Lost-City	Serpentinizing site	KP204688.1 (100%, 97%)	Orca Basin South hypersaline sediment

* Accession number of representative sequences of OTUs

tem of The Cedars shared two uncultivated *Chloroflexi* OTU with PHF (Figure 3.8). The high-abundance level of *Alphaproteobacteria* (64.7%) in the intertidal BdJ site was also not representative of all serpentinizing systems, as exemplified by their low abundance in CVA and The Cedars ecosystems (Suzuki et al., 2013, Tiago and Veríssimo, 2013). The PHF sites shared, however, five *Alphaproteobacteria* OTUs assigned to N2-fixing *Rhizobiales*, chemoheterotrophic *Caulobacterales* and *Sphingomonadales* with the Leka ophiolite or Voltri. *Thaumarchaeota* of the Marine Group I (MG-I) class (Tables 3.4 and 3.6) also represented a dominant taxon in half of the PHF sites as outlined above. One PHF OTU (AY505046.1) was also detected within a group of new MG-I sequences from Lost-City (Brazelton et al., 2006) (Figure 3.8). These sequences may represent a new ecotype of *Thaumarchaeota* specifically adapted to serpentinizing hydrothermal system conditions and, in particular, to low ammonium concentrations (Russell et al., 2010) and high pH values (>10). Precisely, pH adaptation has shown evidence to be the major driving force in *Thaumarchaeota* evolutionary diversification (Gubry-Rangina et al., 2015). Thus, serpentinite-hosted systems may represent a niche for specialized alkaliphilic *Thaumarchaeota*.

OTUs Belonging to Alkaliphilic Cosmopolitan Taxa Shared with Other Serpentinizing Sites

Two *Gammaproteobacteria* OTUs assigned to the genus *Silanimonas* (AM778014.1 and KF511881.1) were detected in BdJ and CVA and even in The Cedars and in Voltri for KF511881.1. The genus *Silanimonas* contains three species (Chun et al., 2017), two of which are slightly thermophilic alkaliphiles with pH optima for growth around 9-10. Although OTU AM778014.1 seems restricted to serpentinizing sites, OTU KF511881.1 is also commonly detected in various types of aquatic habitat such as lake water, marine sediment or hot springs. Considering the presence of *Silanimonas* in the four serpentinizing systems with a maximum pH around 11.5-12.5 and its absence in the Leka ophiolite (where the maximum in situ pH is 9.6; Daae

et al. (2013), this genus might be an indicator of a highly alkaline environment rather than terrestrial serpentinizing ecosystems. A similar conclusion may be drawn for the *Truepera* OTU (*Deinococcus-Thermus* phylum) shared by the PHF sites with high pH fluid, Voltri and CVA. Members of the genus *Truepera* are known to be chemoorganotrophic alkaliphiles, slightly thermophilic and extremely resistant to ionizing radiation.

OTUs Endemic to the PHF Belonging to Candidate Phyla

The OTUs classified into the six bacterial candidate phyla detected in PHF were not recovered in other serpentinizing systems, with the exception of one *Parcubacteria* OTU (LN561460.1) identified in both the submarine ST07 site and the Voltri terrestrial springs. Members of this candidate phylum could be involved in 'dark' anaerobic fermentation (Castelle et al., 2017). Several metagenomic reconstructed genomes of *Parcubacteria* are very small with reduced metabolic capabilities (e.g., lacking ATP synthase genes), suggesting that they may be intercellular symbionts of other prokaryotes sharing the same habitat (Nelson and Stegen, 2015, Suzuki et al., 2017). Since some *Parcubacteria* members were found in abundance in the deep groundwater of The Cedars (as well as in Lost-City and in PHF) they were thought to be selected by the geochemical conditions associated with serpentinization reactions (Suzuki et al., 2017). However, no *Parcubacteria* OTU was shared by PHF and The Cedars (Figure 3.8) and their closest OTUs indicated less than 90% identity on their overlapping region. Although clone sequencing of The Cedars (Suzuki et al., 2013) may have failed to identify the overall *Parcubacteria* community, PHF most likely hosted distantly related taxa potentially associated with other metabolisms and different microbial consortia.

3.4 Conclusion

In this study, 16S rRNA gene pyrosequencing provided a comprehensive overview of the microbial biosphere of PHF and extended its known diversity by revealing the presence of prokaryotic populations pertaining to six novel phyla not previously detected. Moreover, this study showed that bacterial and archaeal communities in PHF were dominated by rare phylotypes. This study also included the first exhaustive comparison at OTU scale of microbial taxa across six geographically distant serpentinite-hosted ecosystems. Although at this taxonomic resolution, no *sensu stricto* core microbial community could be defined, a few phylotypes were shared by several sites. These phylotypes, belonging to generally dominant taxa in serpentinizing systems, seem to be specifically associated with unique niches resulting from the peculiar geochemical conditions linked to serpentinization. Indeed, OTUs representing uncultivated *Clostridia* of the *Syntrophomonaceae* family (and related to the *Dethiobacter alkaliphilus*) or of the *Thermoanaerobacterales* group SRB2 are hypothesized to live in the anoxic, H_2 rich end-member fluids or inside the porous carbonates structures, alike the *Methanosarcinales* (LCMS and TCMS) (Schrenk et al., 2013). *Hydrogenophaga* OTUs identified in most of the terrestrial

serpentinizing sites probably live at oxic-anoxic interfaces where the mixing of alkaline H_2-rich fluids and oxygenated waters creates microaerophilic conditions. Upon mixing, steep gradients of pH, temperature and electron donors and acceptors, also provide a range of microhabitats favorable for endemic microbial populations in these extreme ecosystems. Additional factors, abiotic (geological and hydrological settings, water depths, light, etc.) or biotic (species interactions, viruses), superimposed to the serpentinization reactions most likely contributed significantly in shaping the microbial assemblages, thus explaining the limited similarities between ecosystems at the OTU level. This apparent divergence of microbiome composition in serpentinizing systems raised the question of the existence of common metabolic capabilities that could explain the functioning of these ecosystems better. These issues could be addressed in the future by using new genomic tools, such as comparative metagenomics.

Chapter 4

Functional comparison of serpentinizing systems[1]

Sommaire

[1]The work described in this chapter was submitted to the ISME Journal for publication:. Frouin E, Armougom F, Schrenk M and Erauso G (in review). Comparative metagenomics highlight a pathway involved in the catabolism of phosphonates in marine and terrestrial serpentinizing ecosystems

Overview (French part)

L'article intitulé "Comparative metagenomics highlight a pathway involved in the catabolism of phosphonates in marine and terrestrial serpentinizing ecosystems" présente la première comparaison exhaustive de métagénomes issus de différents sites serpentinisés. Les conclusions du chapitre précédent, qui montraient une faible conservation taxonomique dans les systèmes serpentinisés ont rapidement fait émerger la question de la redondance fonctionnelle dans les communautés microbiennes. L'objectif du travail présenté dans ce chapitre est de déterminer si les pressions de sélection imposées par les réactions de serpentinisation ont permis de sélectionner des capacités fonctionnelles identiques au sein des communautés microbiennes. Le lien entre la serpentinisation et les micro-organismes associés a été étudié au travers d'une comparaison métagénomique de systèmes serpentinisés et non serpentinisés mais comparables en termes de pH, température ou niche écologique. Des analyses en composantes principales et l'application de méthodes d'apprentissage supervisées ont permis d'établir et de quantifier les ressemblances entre les ecosystèmes serpentinisés.

Dans un premier temps, ce chapitre s'intéresse à la catégorisation des systèmes serpentinisés. Les profils taxonomiques et fonctionnels des métagénomes étudiés ont montré que les microbiomes des deux systèmes serpentinisés sous-marins, PHF et LCHF, étaient plus éloignés qu'attendu d'après les premières analyses de métabarcoding. Plus précisément, la biosphère microbienne de PHF serait plus proche de celle des sites serpentinisés terrestres, tandis que les métagénomes de LCHF ressembleraient davantage à ceux des évents océaniques de sites basaltiques. Une explication possible est que les micro-organismes de LCHF et cheminées hydrothermales situées sur des socles basaltiques sont soumis à de nombreuses contraintes physiques (pression hydrostatique) et chimiques (métaux lourds, radionucléides, etc.), qui sont probablement spécifiques aux cheminées hydrothermales profondes. Ces facteurs peuvent avoir davantage d'importance dans le processus de sélection de la diversité microbienne.

Ce chapitre présente ensuite les ressemblances, en termes de potentiel fonctionnel, entre les micro-organismes qui colonisent différents sites serpentinisés. L'étude a notamment confirmé l'importance des activités métaboliques liées à l'hydrogène mais elle a également révélé un ensemble de gènes enrichis impliqués dans la réponse aux stress environnementaux. L'observation la plus frappante est un enrichissement de tous les gènes d'une voie de dégradation des phosphonates dans les métagénomes des écosystèmes serpentinisés. Il a été estimé que jusqu'à 44 % des micro-organismes de ces écosystèmes posséderaient les gènes codant pour une carbone-phosphore lyase (C-P lyase), enzyme clé de cette voie de dégradation. Ces gènes ont été détectés dans une large gamme de phylotypes bactériens, dont la plupart appartenaient aux *Clostridiales*, α- et β-*proteobacteria*. Ces résultats ont été mis en parallèle avec le fait que les phosphonates peuvent être utilisés par des micro-organismes pour obtenir du phosphore lorsque le phosphate inorganique est rare. De plus, la dégradation des phosphonates via l'activité de la C-P lyase peut, suivant le substrat, conduire à la libération d'hydrocarbures et doit être envisagée comme source potentielle de CH_4 biotique dans les milieux serpentinisés.

4.1 Introduction

Serpentinite-hosted hydrothermal systems are the product of the hydration of ultramafic rocks (peridotite), originating in the Earth's mantle. The serpentinization process produces serpentine minerals and emits ultrabasic fluids that are enriched in hydrogen gas, methane and small organic molecules (formate, acetate, and methanol) (Schrenk et al., 2013). Abiotic products of serpentinization and associated reactions are assumed to feed the microbial inhabitants of these ecosystems. In submarine environments, carbonate minerals precipitate when hydrothermal fluids mix with seawater and form chimneys. The mixing of highly reduced alkaline fluids with seawater also sustains ionic and redox gradients within the porous chimney walls. Similar gradients are thought to have played a role in the development of metabolic processes during the early evolution of life on Earth (Martin et al., 2008, Schrenk et al., 2013). The emblematic Lost City Hydrothermal Field (LCHF, located at ~800m depth, 15 km off the Mid-Atlantic Ridge) has been the only studied submarine system for many years and still represents a reference model for serpentinization-influenced geochemical processes and associated microbial ecosystems (Brazelton et al., 2006, Kelley et al., 2001, Lang et al., 2010, Schrenk et al., 2004). However, the sampling of serpentinizing systems at depth is technically challenging, and therefore other more accessible serpentinizing environments have been investigated such as continental ophiolites, portions of ancient seafloor obducted onto continental margins (Schrenk et al., 2013).

The Prony Hydrothermal Field (PHF) recently re-discovered at shallow depth (< 50 mbsl) in the Bay of Prony, New Caledonia (Pelletier et al., 2006) represents an alternate example of a marine serpentinizing ecosystem. Like the LCHF, tall brucite-rich carbonate chimneys vent highly alkaline fluids enriched in H_2 and CH_4, although at lower temperatures (max 42°C versus 90°C) (Quéméneur et al., 2014). As a consequence of their distinctive geologic settings (mid-ocean ridge versus supra-subduction zone), the two hydrothermal systems differ by the nature of their circulating fluids. PHF is alimented by meteoric waters (instead of seawater as at the LCHF) percolating through the thick peridotite nappe (>1.5 km (Quesnel et al., 2016)) and discharges low-salinity fluids into the lagoon (Monnin et al., 2014). Thus, from a geological point of view, PHF appears as an intermediate between a submarine and terrestrial serpentinizing systems (Monnin et al., 2014). This peculiar situation was also confirmed at a microbial community level, since PHF shared several phylotypes with both continental and submarine serpentinizing ecosystems, such as members of the genus *Hydrogenophaga*, typical of terrestrial systems or two distinct phylotypes of uncultivated *Methanosarcinales* (Postec et al., 2015, Quéméneur et al., 2014), first discovered at LCHF (Schrenk et al., 2004) and The Cedars (Suzuki et al., 2013). Other than these few phylotypes, commonly identified in serpentinizing habitats, the remaining microbial community of PHF showed limited similarity at the species level with microbiomes of other serpentinizing ecosystems (Frouin et al., 2018), suggesting local environmental factors influence the taxonomic distribution beyond the geochemical conditions imposed by serpentinization. In natural ecosystems, many different species possess similar metabolic

capabilities and can potentially play the same ecological roles depending on the biogeography (Louca et al., 2018). Therefore, differences in the microbiome composition of serpentinizing systems from distinct geographic sites do not necessarily reflect a different functional pattern. Harsh conditions imposed by serpentinization: high pH, low concentrations of terminal electron acceptors and dissolved inorganic carbon, presumably exert a strong selective pressure on the associated microbial communities. Our working hypothesis was that to thrive in such challenging conditions, microorganisms of these ecosystems have evolved convergent adaptive strategies that should be observable at the functional level across distant serpentinizing sites, beyond apparent difference at the taxonomic level.

Until now, microbiological studies of serpentinizing ecosystems have focused their attention on the most apparent metabolisms associated with the two dominant sources of energy, hydrogen and methane, in submarine systems such as the LCHF (Brazelton et al., 2012) and the PHF (Mei et al., 2016b), and in continental sites of the Coast Range Ophiolite Microbial Observatory (CROMO, USA), Tablelands Ophiolite (Canada), Voltri Massif (Italy) and Santa Elena Ophiolite (Costa Rica) (Brazelton et al., 2012, 2016, Crespo-Medina et al., 2017, Twing et al., 2017). Other studies have provided new insights into the potential sources of carbon feeding microbial communities, given the near-complete depletion of dissolved inorganic carbon in end-member fluids. It has been suggested that some microorganisms can use $CaCO_3$ of the chimney (Pisapia et al., 2017, Schrenk et al., 2013), a capacity demonstrated only for the hyper-alkaliphilic candidate genus "*Serpentinomonas*" isolated at The Cedars (Suzuki et al., 2014). In the Tablelands and CROMO, the detection of numerous carbon monoxide dehydrogenase genes suggested microbial carbon monoxide assimilation and acetogenesis via the Wood-Ljungdahl pathway (Brazelton et al., 2012, Twing et al., 2017). Organic molecules produced by serpentinization and associated reactions, such as formate (Lang et al., 2018) and aromatic amino acids (Pisapia et al., 2018) could also be a valuable source of dissolved carbon, but direct evidences are missing.

Metabolic strategies for nitrogen or phosphorus acquisition are particularly understudied in serpentinization-influenced ecosystems. Functional gene surveys of LCHF chimneys showed the presence of up to 18 different types of *nifH* gene, all related to methanogenic Archaea (Brazelton et al., 2011). The authors of this study speculated that low concentrations of dissolved nitrogen compounds in end member fluids explain the prevalence of the genes involved in nitrogen fixation. Phosphate (PO_4^{3-}), commonly used as a source of phosphorus (P) in aquatic ecosystems, was reported at very low concentrations (< 1µM) in several geochemical studies of serpentinizing systems (Bradley et al., 2009, Cardace et al., 2015, Morrill et al., 2013), raising questions about the source of P and its mechanism of assimilation in serpentinite-hosted ecosystems. Indeed, phosphorus is an essential nutrient owing to its pivotal role in cell structure, storage, metabolism or gene expression for all living organisms (White and Metcalf, 2007). The major bioavailable form of P in marine ecosystems is the inorganic phosphate (+5 valence), which can be directly assimilated by microorganisms (Karl, 2014). However, under phosphate-limited conditions, a wide range of microorganisms can metabolize reduced P compounds (va-

lence +3) as substrates for growth, including phosphonates (Pn), which are characterized by a chemically stable Carbon-Phosphorus (C-P) bond, and phosphite (PO_3^{3-}) (White and Metcalf, 2007). Among the microbial enzymes of the Pn catabolism, the Carbon-Phosphorus lyase pathway has been extensively studied because its enzymatic activity based on methylphosphonate esters as substrate explains the biological production of methane in oxygenated environments (Repeta et al., 2016, Sosa and Delong, 2017). To our knowledge, issues concerning the metabolism of P acquisition in alkaline serpentinizing systems, have not been addressed previously and deserve investigation.

In this study, we performed a cross-comparison of metagenomes from submarine and terrestrial serpentinizing systems, along with a selection of metagenomes from other hydrothermal environments, a priori not influenced by serpentinization (deep-sea basalt-hosted vents and terrestrial hot springs). We aimed to identify core microbial communities, common functions or metabolic pathways, which could be characteristic of serpentinizing systems and help us understand microbial growth and survival in these challenging environments.

4.2 Material and Methods

4.2.1 Metagenomics data sets

Twenty-one publically available metagenomic data sets from six serpentinizing systems (submarine and continental) and four other hydrothermal systems: hot springs from continental volcanic area and deep-sea hydrothermal vents, were retrieved from SRA and MG-RAST databases and described in the Table 4.1. To limit the bias arising from the use of different sequencing technologies and protocols, we selected only the metagenomic datasets containing Illumina raw paired-end reads, with a partial overlap.

4.2.2 Comparative metagenomics pipeline

With the objective of standardizing data processing, all 21 metagenomes (Table 4.1) were reprocessed using an in-house bioinformatics pipeline for quality control, assembly and annotation against taxonomic and functional databases. This pipeline was implemented in Snakemake (Köster and Rahmann, 2012) and deposited on the Github repository: `https://github.com/elfrouin/MetaGPipeline.git`. The raw paired-end reads of metagenomes were firstly merged using PandaSeq v2.8 (Masella et al., 2012) with default parameters. The merged reads were next processed by Trimommatic v0.32 (Bolger et al., 2014), which trimmed reads once the average quality in a 4bp-window fell below 15, cut bases off the extremities if below a quality score of 20, and removed short reads (< 35bp). The high quality trimmed reads were finally assembled into contigs, longer than 200 bp, with IDBA-UD v1.1.1 (Peng et al., 2012). The quality of assemblies was checked by estimating the rate of reads mapped to the contigs with BWA (Li and Durbin, 2009). The gene prediction was done with Prodigal v47 (Hyatt et al., 2010), us-

ing the "-p meta" flag for metagenomics sequences. The gene coverage was computed with Bedtools v2.25.0 (Quinlan and Hall, 2010). The microbial diversity of each metagenome was inferred by taxonomic annotation of predicted genes. Similarity searches were first performed against the NCBI non-redundant (nr) database with DIAMOND blastp v.0.8.34 (Buchfink et al., 2014) using a maximal E-value of 1e-5. MEGAN6 (Huson et al., 2016) was used to assign the annotated genes to their most confidently predicted taxonomic rank, by applying the Lowest Common Ancestor (LCA) algorithm. A taxonomic contingency table was created by weighting annotated genes by their sequencing coverage. A first functional annotation of genes was performed by RPS-Blast search (E-value cutoff of 1e-5) against the Clusters of Orthologous Group (COG) database (Tatusov et al., 2003). The second functional annotation was performed using the online tool GhostKOALA (the only non-automated step) against the KEGG Orthology (KO) database with a minimum alignment score threshold of 50 (Kanehisa et al., 2016). The coverage of functional genes was normalized between metagenomes using the TMM normalization method (Robinson and Oshlack, 2010), implemented in the R package edgeR v3.14.0 (Robinson et al., 2009).

4.2.3 Statistical analyses

The taxonomic similarity between the 21 metagenomes was determined using relative abundance of archaeal and bacterial genera. Based on the normalized contingency table (exported from MEGAN), a distance matrix was computed with the Jensen-Shannon divergence using the R package phyloseq v1.20.0 (McMurdie and Holmes, 2013). The results were visualized using a principal coordinates analysis (PCoA, Figure 4.3). A non-parametric Kruskal-Wallis test combined with pairwise Wilcoxon tests (adjusted using Bonferroni's correction) were performed to detect genera with significant differential abundance among the subsets of metagenomes identified on PCoA. These specific microbial genera were added on the PCoA biplot. Regarding the functional capabilities of the microbial community, the Pearson correlation and Ward's method were employed for agglomerative hierarchical clustering of metagenomes and annotated genes. The gene coverage per metagenome was visualized using heatmaps, generated with the R package gplots v3.0.1. To discriminate the functional capabilities of serpentinizing versus non-serpentinizing systems, a "top-down" Random Forest analysis was carried out using the R package randomForest v4.6.12 (Liaw and Wiener, 2002) with 3000 independent decision trees. Finally, a heatmap was built with the top 50 genes contributing to segregate the serpentinizing metagenomes from the others (i.e., the genes with the highest value of mean decrease Gini, a measure of variable importance).

Table 4.1 – Details of sampling and sequencing of the 21 metagenomes

Sample names		Sampling			Sequencing			Reference
Short ID	**Long name**	**Date**	**Latitude & longitude**	**Size of filter**	**Accession (SRA or MG-RAST)**	**Strategy**	**# of raw reads**	
QV1.1	QV11A	2012-08-21	38.862 N 122.43 W	0.2 µm	4569550.3	HiSeq 2000, 2*100, overlap 30	21,654,302	Twing et al. (2017)
CSW1.3	CSW13A	2012-08-21	38.862 N 122.414 W	0.2 µm	4569551.3	HiSeq 2000, 2*100, overlap 30	23,167,547	Twing et al. (2017)
CSW1.1	CSW11AC	2012-08-21	38.862 N 122.414 W	0.2 µm	4569549.3	HiSeq 2000, 2*100, overlap 30	14,014,225	Twing et al. (2017)
QV1.2	QV12A	2012-08-21	38.862 N 122.43 W	0.2 µm	4569552.3	HiSeq 2000, 2*100, overlap 30	18,898,568	Twing et al. (2017)
SE.9	Spring9	2014-02-25	10.8939 N 85.7274 W	0.2 µm	SRR4101185	HiSeq 2000, 2*100, overlap 30	60,780,480	Crespo-Medina et al. (2017)
GOR.13	GOR34-spring1-2013	2012-07-01	44.597 N 8.7833 E	0.2 µm	SRR1636513	HiSeq 1000, 2*100, overlap 30	30,005,532	Brazelton et al. (2016)
GOR.12	GOR34-spring3-2012	2012-07-01	44.597 N 8.7833 E	0.2 µm	SRR2058407	HiSeq 1000, 2*100, overlap 30	20,259,258	Brazelton et al. (2016)
BR2.13	BR2-spring-2013	2012-07-05	44.4512 N 8.782 E	0.2 µm	SRR1636510	HiSeq 1000, 2*100, overlap 30	10,501,350	Brazelton et al. (2016)
BR2.12	BR2-spring-2012	2012-07-05	44.597 N 8.7833 E	0.2 µm	SRR2058405	HiSeq 1000, 2*100, overlap 30	15,452,030	Brazelton et al. (2016)
CVA	Cabeco de Vide travertine	2012-11-01	39.23 N 7.97 W	-	SRR1636515	HiSeq 1000, 2*100, overlap 30	7,898,772	Not published
PHF.27	Prony 27	2005-06-14	22.36092 S 166.8800 E	none	SRR1636516	HiSeq 1000, 2*100, overlap 30	5,507,502	Mei et al. (2016b)
PHF.28	Prony 28	2005-06-14	22.36092 S 166.8800 E	none	SRR1636517	HiSeq 1000, 2*100, overlap 30	8,104,211	Mei et al. (2016b)
LC.75	Lost City 75	2003-04-27	30.1240 N 42.1201 W	-	SRR1636508	HiSeq 1000, 2*100, overlap 30	17,173,696	Lang et al. (2018)
LC.81	Lost-City 81	2005-08-01	30.1241 N 42.1201 W	-	SRR1636509	HiSeq 1000, 2*100, overlap 30	16,578,254	Lang et al. (2018)
Mrk.33	Axial Seamount Marker33	2013-09-28	45.9332 N 129.9823 W	0.2 µm	ERR694199-ERR694206	HiSeq 1000, 2*100, overlap 30	27,905,863	Not published
Mrk.113	Axial Seamount Marker113	2013-09-27	45.9227 N 129.9881 W	0.2 µm	ERR694207-ERR694213	HiSeq 1000, 2*100, overlap 30	24,653,254	Fortunato and Huber (2016)
MCR.SG	Shrimp Gulley #2, Piccard	2012-01-23	18.5466 N 81.7177 W	0.2 µm	ERR868087-ERR868115	HiSeq 2000, 2*100, overlap 30	109,104,481	Reveillaud et al. (2016)
TAT.3	Tattapani TAT-3	2016-01-26	23.41 N 83.39 E	0.2 µm	SRR3961742	NextSeq 500, 2*146, overlapped	15,209,380	Saxena et al. (2017)
TAT.4	Tattapani TAT-4	2016-01-26	23.41 N 83.39 E	0.2 µm	SRR3961743	NextSeq 500, 2*146, overlapped	8,839,629	Saxena et al. (2017)
Yell.O	Octopus Spring	2011-10-11	44.5341 N 110.7979 W	-	SRR4030098	HiSeq 2000, 2*150, overlapped	20,590,784	Not published
Yell.C	Conch Spring	2011-10-11	44.5564 N 110.8322 W	-	SRR4030102	HiSeq 2000, 2*150, overlapped	19,085,473	Not published

(-) missing information

4.2.4 Phn operon analysis

Analyses of *phn* genes were carried out to identify their associated taxonomies and genomic organizations. Genes were queried against the NCBI nr database using blastn (with an E-value of 1e-6). The determination of gene taxonomy was next assessed in two ways: (i) via the best BLAST hit and (ii) via the LCA algorithm implemented in MEGAN6. To estimate the percentage of microorganisms possessing a *phn* operon, the occurrences of *phnGHIJKLM* genes were normalized by the abundance of a single-copy gene *recA*, in each sample. This normalization allowed a comparison with a former study that had revealed a high abundance of the *phn* genes in Sargasso Sea metagenomes (Karl et al., 2008). Finally, the structural organization of the operon *phn* was investigated in the metagenomics contigs. All contigs with at least the seven genes of the catalytic unit were reannotated with PROKKA v1.11 (Seemann, 2014) to characterize the genomic context of the operon.

4.3 Results

4.3.1 Dataset and site location

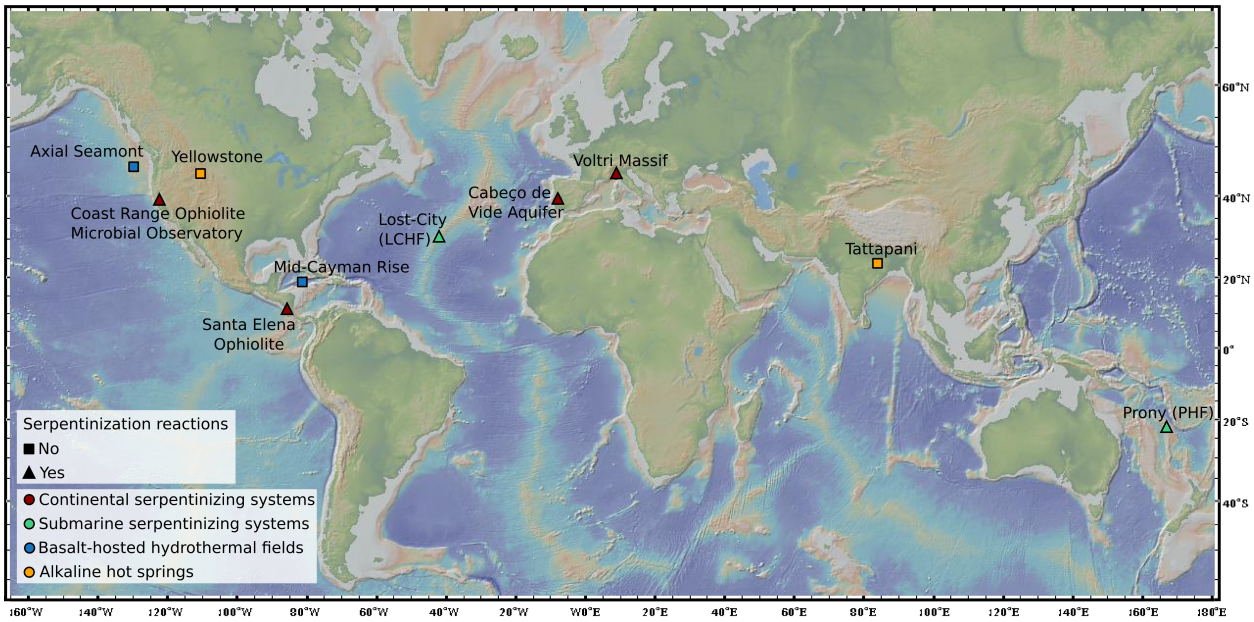


Figure 4.1 – Global distribution of studied samples

The dataset consisted of 21 publically available environmental metagenomes encompassing marine and continental serpentinite-hosted systems (14 metagenomes), hot springs (4) and basalt-hosted hydrothermal vents (3) spread worldwide (Figure 4.1). They originated from ten well-studied sites comprising PHF, LCHF, CROMO, Santa Elena Ophiolite, Voltri Massif, Cabeço de Vide Aquifer (CVA), Tattapani and Yellowstone hot springs, Axial Seamount and the Mid-Cayman Rise. For each metagenome, the main environmental parameters, including type of habitat, depth, temperature, and pH, are reported in Table 4.2. All the serpentinizing systems

here considered emit highly alkaline fluids, with pH ranging between 9 and 12.3, except for the sample QV1.2 in CROMO (pH 7.9). Fluids from other ecosystems have rather circumneutral pH values. Raw metagenomic data were entirely reprocessed to minimize methodological biases (see Materials and Methods). The de-novo assemblies of the 21 metagenomes generated between 10,084 and 401,248 contigs, depending on the sample (Table 4.3). These contrasted counts of contigs resulted from sequencing depth variability between metagenomes. Consequently, the number of predicted genes ranged from 16,780 to 579,028. Their rate of annotation to known functional categories reached on average 55% and 38% based on COG and KEGG databases, respectively.

Table 4.2 – Global distribution of studied samples. (*) indicates data collected in other studies.

Sample ID	Site, Country	Type	Depth	pH	Temperature
QV1.1	Coast Range Ophiolite, USA	groundwater from well	-23 m	11.5	18°C
CSW1.3	Coast Range Ophiolite, USA	groundwater from well	-23 m	10.1	17°C
CSW1.1	Coast Range Ophiolite, USA	groundwater from well	-19 m	12.2	17°C
QV1.2	Coast Range Ophiolite, USA	groundwater from well	-15 m	7.9	18°C
SE.9	Santa Elena Ophiolite, Costa Rica	spring water		11.54	26°C
GOR.13	Voltri Massif, Italy	spring water		12.3	24°C (*)
GOR.12	Voltri Massif, Italy	spring water		11.8	24°C (*)
BR2.12	Voltri Massif, Italy	spring water		12.1	22°C (*)
BR2.13	Voltri Massif, Italy	spring water		12.3	22°C (*)
CVA	Cabeco de Vide, Portugal	groundwater from a borehole	-130 m	11.4	20°C
PHF.27	Bay of Prony, New Caledonia	hydrothermal chimney	-43 m	10.6	<40°C (*)
PHF.28	Bay of Prony, New Caledonia	hydrothermal chimney	-43 m	10.6	<40°C (*)
LC.75	Lost City, Atlantic Ocean	hydrothermal chimney	-733 m	10.2	40-90°C (*)
LC.81	Lost City, Atlantic Ocean	hydrothermal chimney	-767 m	10.2-10.7 (*)	40-90°C (*)
Mrk.33	Axial Seamount, Pacific Ocean	hydrothermal fluid	-1516 m	5.4	28°C
Mrk.113	Axial Seamount, Pacific Ocean	hydrothermal fluid	-1522 m	6.2	24°C
MCR.SG	Mid-Cayman Rise, Caribbean Sea	hydrothermal fluid	-4940 m	6.7	108°C
TAT.3	Tattapani, Chhattisgarh, India	water from hot spring		7	69°C
TAT.4	Tattapani, Chhattisgarh, India	water from hot spring		7.8	67°C
Yell.O	Octopus Spring, USA	water from hot spring		8	84°C
Yell.C	Conch Spring, USA	water from hot spring		8	84°C

4.3.2 Microbial communities composition

The overall microbial diversity, identified in the set of metagenomes was mostly affiliated to 22 phyla (19 bacterial and 3 archaeal phyla, Figure 4.2), based on the taxonomic annotation of all genes. The *Proteobacteria* (up to 96%) and *Firmicutes* (up to 50%) phyla made up the bulk of the microbial composition in most samples, except for the hot springs. Compared with the other environments, the hot springs showed a distinct taxonomic profile, with a large proportion of *Aquificae*, as well as abundant *Thermotogae* and *Crenarchaeota* in Tattapani and Yellowstone hot springs, respectively. The *Deinoccocus-Thermus* clade was prominent only in the metagenome of TAT.3. The strong prevalence of phyla containing mostly thermophilic or hyperthermophilic lineages is typical of continental hot springs (Saxena et al., 2017). Aside from the hot springs samples, the archaeal communities were dominated by the *Euryarchaeota* phylum, which represented up to 20% in hydrothermal fields (LCHF, PHF, and basalt-hosted hy-

Table 4.3 – Summary statistics of metagenomics assembly and annotation

Sample ID	Pairs of raw reads	Without duplicates	Merged reads	Trimmed reads	Mapped reads	# contigs >200 bp	# ORFs	ORFs in COG	ORFs in KEGG
QV1.1	21,654,302	21,070,664	17,868,635	17,571,815	99.23%	17,032	33,211	64.96%	49.39%
CSW1.3	23,167,547	23,160,350	19,486,728	19,154,819	91.07%	79,146	160,978	60.75%	40.66%
CSW1.1	14,014,225	14,004,274	11,156,997	10,948,889	98.92%	10,084	16,78	65.82%	51.75%
QV1.2	18,898,568	18,451,018	15,566,466	15,260,812	88.08%	116,78	223,23	57.94%	37.73%
SE.9	60,780,480	60,678,790	53,515,679	52,409,340	78.85%	401,248	579,028	48.30%	33.43%
GOR.13	30,005,532	29,838,650	29,195,809	28,847,759	87.85%	217,496	321,685	55.15%	40.35%
GOR.12	20,259,258	16,151,215	15,947,160	15,816,431	89.36%	131,011	186,518	54.51%	41.86%
BR2.13	10,501,350	10,405,948	10,043,010	9,952,707	78.27%	65,379	103,002	56.04%	40.02%
BR2.12	15,452,030	14,183,543	14,106,161	14,049,164	96.45%	61,292	87,242	51.58%	34.72%
CVA	7,898,772	7,555,854	7,410,944	7,365,030	85.21%	64,784	102,407	59.85%	40.62%
PHF.27	5,507,502	3,976,463	3,939,864	3,913,625	60.00%	24,869	34,983	58.00%	38.21%
PHF.28	8,104,211	7,579,941	7,494,997	7,455,672	80.56%	49,383	81,067	58.61%	38.73%
LC.75	17,173,696	17,113,803	16,563,824	16,460,777	84.60%	130,986	216,451	53.59%	38.50%
LC.81	16,578,254	16,524,071	16,068,746	15,978,331	76.95%	160,498	254,61	53.01%	37.78%
Mrk.33	27,905,863	26,998,320	26,265,132	26,123,952	48.33%	277,94	384,98	52.71%	36.12%
Mrk.113	24,653,254	24,029,495	23,325,992	23,194,885	32.14%	277,468	352,83	48.29%	31.30%
MCR.SG	109,104,481	84,526,933	84,089,856	83,922,387	94.54%	289,334	398,092	42.72%	28.82%
TAT.3	15,209,380	12,965,983	11,204,846	11,082,171	96.08%	29,494	61,874	61.15%	40.24%
TAT.4	8,839,629	8,605,337	7,070,491	6,997,217	93.64%	28,799	55,628	62.36%	43.23%
Yell.O	20,590,784	20,582,006	19,847,197	19,755,692	98.08%	193,05	234,742	42.09%	24.47%
Yell.C	19,085,473	19,076,435	18,339,772	18,244,121	98.39%	136,376	165,409	47.96%	30.16%

drothermal systems) and continental serpentinizing systems (Voltri Massif and CVA). Whilst being considered as marine, the two metagenomes of PHF harbored taxonomic profiles distinct from those of LCHF or other deep-sea hydrothermal systems. The relatively balanced distribution of the dominant *Firmicutes* and *Proteobacteria* at PHF was more similar to that of terrestrial serpentinizing systems (e.g., CVA). However, the abundance of candidate phylum *Acetothermia* in PHF constituted a singularity as they were scarce in the other serpentinizing environments here compared.

The similarity relationships within metagenomes were further investigated at the genus-level by plotting the distance between their relative taxonomic compositions and revealed a clustering of metagenomes according to the geologic context of the microbiome habitat (Figure 4.3). A first cluster gathered the four hot springs, sharing mainly hyperthermophilic genera (e.g., *Hydrogenobacter*, *Pyrobaculum*), typically associated with hydrogen or iron metabolisms (Feinberg et al., 2008, Kawasumi et al., 1984). The second cluster included five samples from deep-sea hydrothermal fields, encompassing both basaltic and peridotite host rock. Abundant genera found within this cluster included γ- and ε-*proteobacteria* mostly associated with sulfur cycling (e.g., *Thiomicrospira* and *Sulfurospirillum* or *Sulfurovum*, respectively) in hydrothermal environments (Brazelton and Baross, 2010, Crépeau et al., 2011). The last cluster contained all the serpentinizing systems fed by fresh water, i.e., continental serpentinizing systems and the coastal PHF. In this group, the highest abundances were found for members of the *Clostridiales* (e.g., *Youngiibacter*, *Dehalobacter*) and β-*proteobacteria* (e.g., *Hydrogenophaga*). Thus, unexpectedly, this analysis set the LCHF metagenomes apart from the ones of all other serpentinizing systems (Figure 4.3). This result was consistent with their contrasting distribution of phyla.

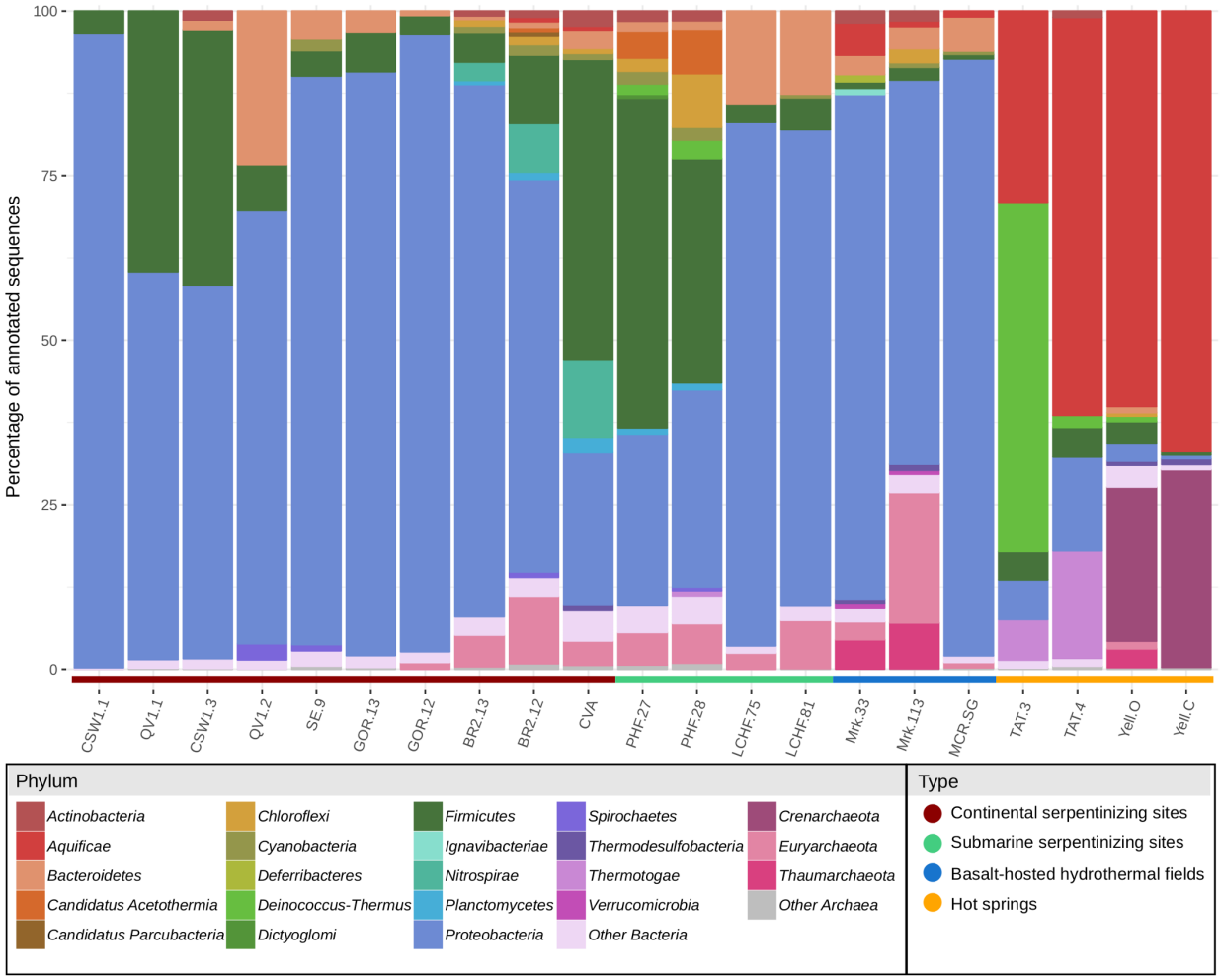


Figure 4.2 – Relative abundance of microbial phyla in the 21 metagenomes compared in this study. The taxonomic distribution was determined using genes successfully assigned to known prokaryotic taxa with MEGAN (using the "Lowest Common Ancestor" algorithm).

4.3.3 Metabolic potential

The metabolic potential of microbial communities was first investigated in the metagenomes by searching a set of key genes, associated with the metabolism of the products of serpentinization (hydrogen, methane and carbon monoxide) or other potential sources of energy (sulfur or nitrogen compounds) and carbon (Figure 4.4). To investigate the potential for hydrogen metabolism, genes encoding [NiFe]- and [FeFe]-hydrogenases were searched in the 21 metagenomes. The bidirectional [NiFe]-hydrogenases of the group 3d (Peters et al., 2015, Vignais, 2001) were prevalent in all serpentinizing ecosystems (*hoxY*, Figure 4.4), compared to other metagenomes. However, this peculiar distribution of gene abundance was not observed for the other genes associated with hydrogen metabolism that even showed significant differences across sites. For example, [NiFe]-hydrogenases with a quinone redox partner were found at LCHF and in basalt-hosted hydrothermal ecosystems, but absent in PHF and other continental serpentinizing sites. In contrast, PHF and continental serpentinizing systems showed an over-representation of [FeFe]-hydrogenases (COG4624, Figure 4.5), commonly involved in fer-

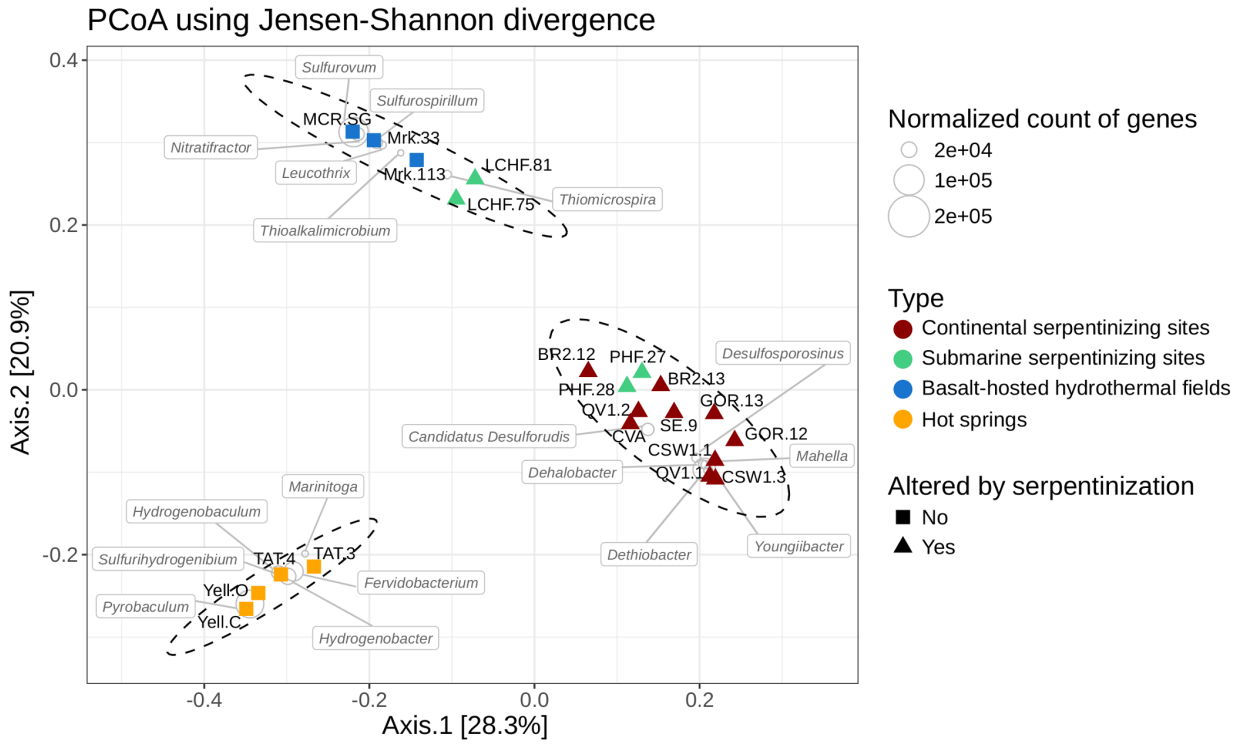


Figure 4.3 – The PCoA plot shows the microbial community composition at the genus level of the 21 metagenomes. Each metagenome is symbolized by a triangle or a square depending on the presence of serpentinization reactions and is colored by ecosystem category. In each of the 3 clusters, the set of genera with statistically significant overabundance were indicated by circles and labeled. Circle size is proportional to the number of genes assigned to a specific genus.

mentative hydrogen production (Mei et al., 2016b).

Methane metabolism was not exclusive to the serpentinizing systems, as biomarkers were also identified in basalt-hosted hydrothermal vents. The *mcrA* gene (encoding the alpha subunit of methyl-coenzyme M reductase), required for the last step of methanogenesis was absent in CROMO and PHF metagenomes. The absence of *mcrA* at PHF could be a consequence of moderate sequencing depth since the detection of *mcrA* was previously reported in PCR-based surveys (Quéméneur et al., 2014). Despite high methane concentrations in serpentinizing sites, the representative gene for bacterial aerobic methane oxidation (*pmoA*) was missing in many metagenomes.

The complete denitrification process (with *nirK* and *norB* as genetic markers) was rarely detected in the serpentinizing metagenomes, while the key genes of the dissimilatory nitrate reduction to ammonium (DNRA: *napA*, *narG*, and *nrfA*) were relatively abundant. The marker for nitrogen fixation *nifH* (encoding the electron transfer subunit of the nitrogenase) was found in all serpentinizing systems, including the terrestrial sites. Additionally, genes for anaerobic ammonium oxidation (Anammox), including *hzsA* (encoding the hydrazine synthase) and *hdh* (encoding the hydrazine dehydrogenase) were not detected among the 21 datasets.

Regarding the potential metabolism of sulfur, genes associated with the dissimilatory sulfate reduction such as *dsrA* (dissimilarity sulfite reductase), *aprA* (adenylysulfate reductase) and *sat* (sulfate adenylytransferase) were detected in most of the metagenomes, but with a relatively

Figure 4.4 – Microbial metabolisms in the 21 metagenomes. Bubble plot of key biomarkers involved in nitrogen (green), hydrogen (turquoise), methane (blue), sulfur (beige) metabolism and in carbon fixation (brown). The size of the circles corresponds to the abundance of normalized KOs.

low abundance at CROMO. Based on the abundance of the *sqr* gene (sulfide:quinone oxidoreductase), the sulfide oxidation seemed a major pathway, especially in basalt-hosted hydrothermal vents. The alternative pathway of sulfide oxidation, involving the flavocytochrome c sulfide dehydrogenase was restricted to deep hydrothermal systems, including LCHF.

Most of the key genes of the bacterial carbon fixation pathways (reviewed by Berg et al. (Berg et al., 2010)) were recovered in the metagenomes except for the *aclA* gene (encoding ATP citrate lyase), only present in LCHF and in the basalt-hosted metagenomes. The complete autotrophic CO_2 fixation via reductive tricarboxylic (rTCA) cycle seemed thereby restricted to these five metagenomes. Other dominant pathways of carbon fixation were differentially represented in the metagenomes of LCHF and PHF. Indeed, genetic markers for 3-hydroxypropionate bicycle and Calvin Benson Bassham cycle were more abundant at LCHF, while at PHF, the dominant markers were those of the reductive acetyl-CoA (Wood–Ljungdahl) pathway, suggesting different strategies for microbial carbon fixation in these two sites.

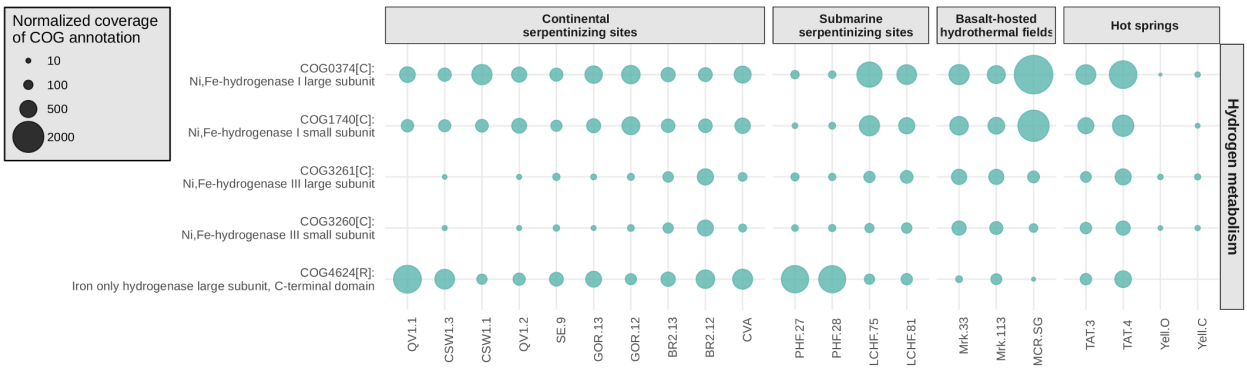


Figure 4.5 – Hydrogen metabolism in diverse hydrothermal environments. Bubble plot of key biomarkers involved in hydrogen metabolism. The size of the circles corresponds to the abundance of normalized COGs.

4.3.4 Comparison of functional profiles

Except for hoxY, the abundance of targeted metabolic genes did not differ between serpentinizing and other environments. Additionally, the hierarchical clustering of all genes annotated with KEGG Orthology (KO) did not group the serpentinizing metagenomes (Figure 4.11). However, this analysis highlighted the functional similarity between LCHF and the basalt-hosted hydrothermal vents. Their metagenomes were enriched in genes encoding transporters (for sugar, phosphate, Mg2+, etc.) and ion channels as well as in genes often associated with coping strategies under fluctuating environmental conditions, including comFC genes related to bacterial competence (Nakasugi et al., 2006) or *dppA* gene linked to chemotaxis (Abouhamad et al., 1991). Likewise, two genes encoding for alginate biosynthesis and transport (*alg8* and *eexD*, respectively) exclusively recovered in metagenomes from deep-sea hydrothermal fields could contribute to the formation of biofilms or desiccation resistant cysts (Hay et al., 2013).

As the global functional profiles of serpentinizing systems cannot be clustered, a supervised approach was then used to specifically select genes with differential abundance between the serpentinizing and the other hydrothermal systems. It is assumed here that the overabundant genes in serpentinizing systems endow a beneficial function on hosted microbial communities. The Random Forest approach finally proved useful in selecting a set of genes whose profile segregated the serpentinizing sites from other hydrothermal environments (Figure 4.6). The gene specifically enriched in serpentinizing ecosystems belonged to various functional categories. These included genes involved in cellular response to exposure to stressful conditions (heat shock response *hcrA* and *rpoE*) as well as genes coding for DNA repair (*recO*) and defense mechanisms (*fitB*, *ndoAI* and *cas7*). Genes encoding membrane transporters of antibacterial compounds (K16907 and *yejA*) and iron complexes (ABC.FEV.A) and a Na^+/H^+ antiporter (*mnhD*, involved in intercellular pH regulation under alkaline conditions (Fukaya et al., 2009)) were also overabundant in serpentinizing metagenomes. As mentioned above, the metagenomes from serpentinizing systems were enriched in *hoxH* and *hoxY*, which encode the large and small subunits of a [NiFe]-hydrogenase of the group 3d. However, the most striking feature was the

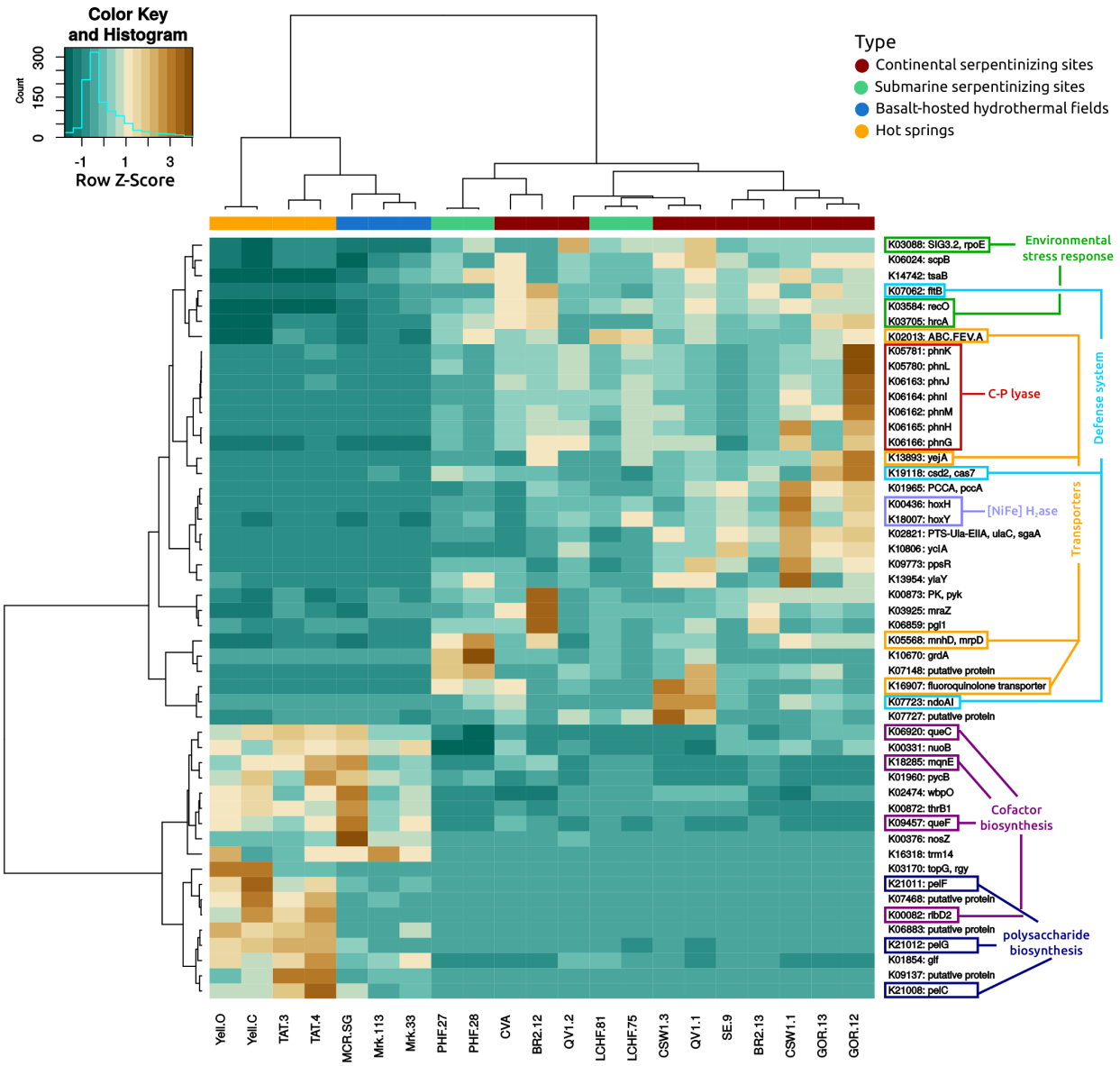


Figure 4.6 – Heatmap showing the normalized abundance of the top 50 KO annotations that best discriminate the 21 functional profiles into two categories of ecosystems. The brown tint indicates a higher relative abundance.

overrepresentation of seven genes from the *phn* operon, encoding enzymes for phosphonate degradation. The same overrepresentation of *phn* genes was detected using a COG-based annotation (Figure 4.6).

4.3.5 The Carbon-Phosphorus lyase operon

The seven identified genes, i.e., *phnGHIJKLM*, belonged to the operon responsible for the C-P lyase activity and, therefore, the degradation of various Pn (White and Metcalf, 2007). These genes constitute the minimal catalytic core unit, essential for C-P bond cleavage (Huang et al., 2005). In the metagenomes of this study, the structural organization and genomic context of the *phn* operon differed from the canonical form of the operon (*phnGHIJKLM*) by rearrangement in gene order and frequent insertion of additional genes encoding acetyltransferases, trans-

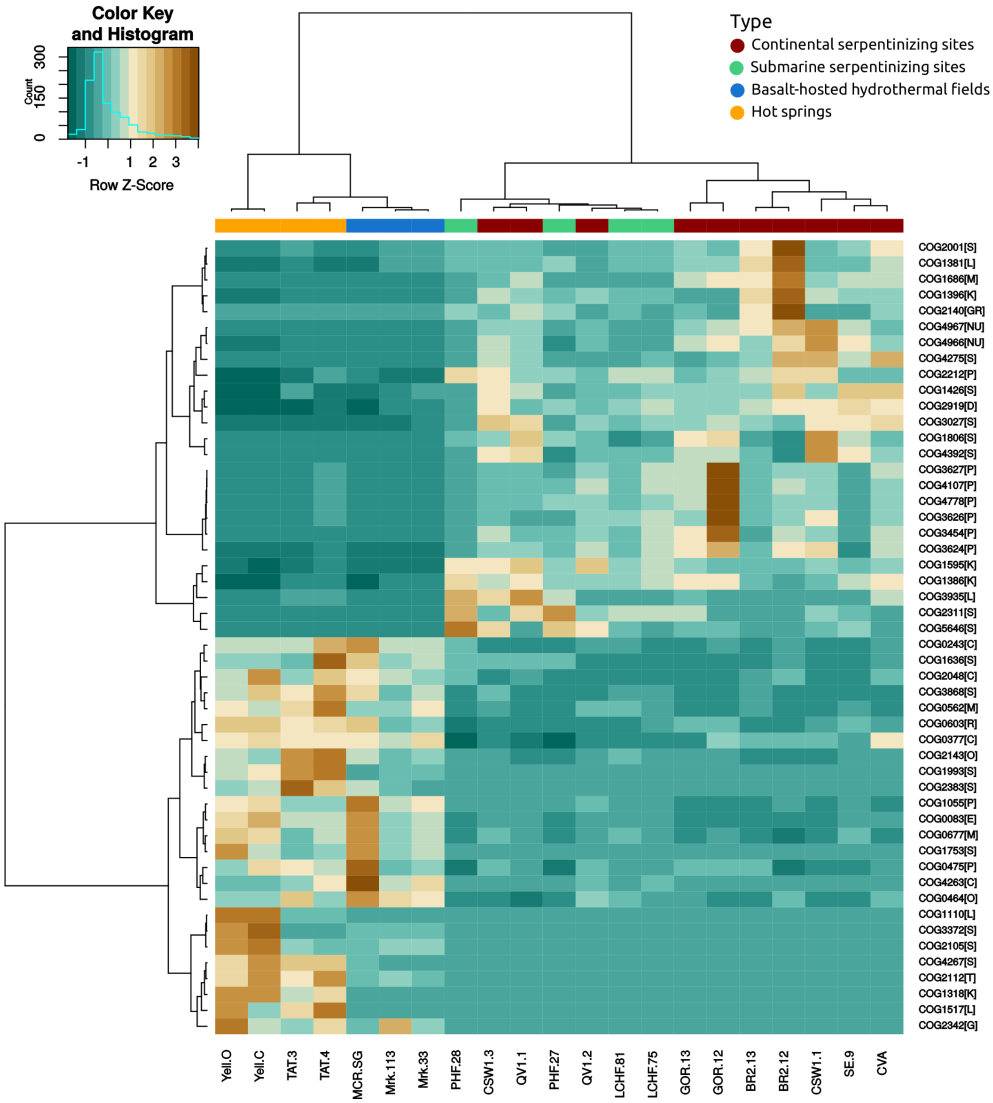


Figure 4.7 – Heatmap showing the normalized abundance of the top 50 COG annotations that best discriminate the 21 functional profiles into two categories of ecosystems. The brown tint indicates a higher relative abundance.

porters or phosphatases (Figure 4.8). Such rearrangements are relatively common as well as the integration of other genes related to phosphorus cycling (Huang et al., 2005). Most of the *phn* gene clusters reconstructed from the metagenomes included genes coding for Pn transporters (*phnCDE*), but none for Pn biosynthesis (i.e., *ppm* and *ppd*). In some metagenomic *phn* clusters, where the *phnF* gene is missing, other genes encoding transcription regulators were identified such as *dasR*, *yvoA*, *yycF*. They may be alternative regulators of the *phn* transcription. Finally, an atypical operon was identified in the metagenomes of PHF that showed the insertion of a NAD-phosphite oxidoreductase gene (*ptxD*, encoding phosphite degradation) between the genes *phnE* and *phnG* (Figure 4.8). This could constitute a novel genomic organization allowing the degradation of Pn or phosphite, since the transporter encoded by the *phnCDE* genes can also be used for the phosphite ion (White and Metcalf, 2004). Compared to C-P lyase, genes for alternative pathways of Pn degradation had comparable abundance in the Santa Elena Ophiolite, but were much rarer in other serpentinizing systems (reported in Table 4.5). Remarkably

the seven genes (*phnGHIJKLM*) of the C-P lyase were almost exclusively present in serpentinite-hosted ecosystems (Figure 4.9). The even abundance of *phn* genes within each metagenome most likely reflected their organization in an operon. In the serpentinizing sites studied herein, from 6 to 44% of microbial genomes could contain the C-P lyase essential genes, a rate comparable to the 30% detected in the metagenomes of the Sargasso Sea, where phosphonate degradation is a prominent process (Table 4.4). The *phn* genes identified in the metagenomes were distributed in several microbial phyla (Figure 4.9). They were predominantly affiliated with *Alphaproteobacteria* or *Firmicutes* at LCHF and PHF, respectively. In the continental serpentinizing ecosystems, most of *phn* genes were affiliated with *Betaproteobacteria* members. Phylogenetic inferences of the *phnJ* gene, which catalyzes the cleavage of the C-P bond, ascertained these assignments (Figure 4.10). However, a high percentage of *phn* genes could not be assigned to known taxa ("not classified" in Figure 4.9), especially in submarines sites.

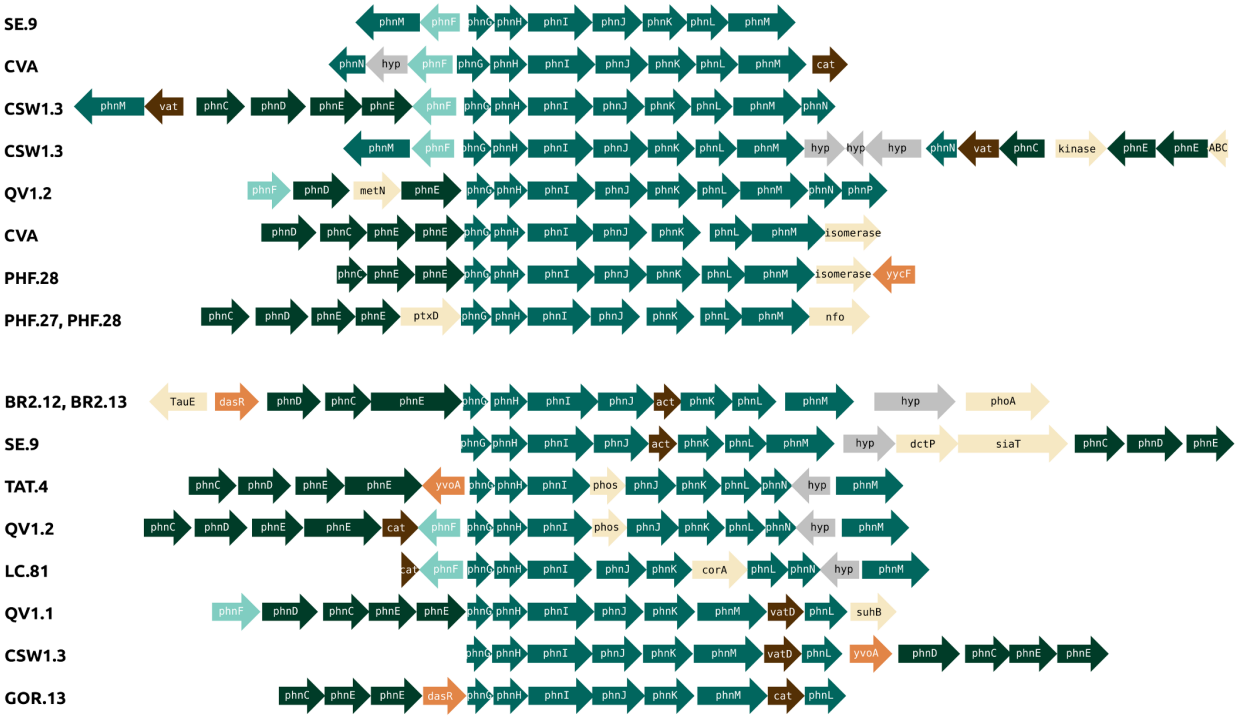


Figure 4.8 – Diversity of *phn* gene clusters structure. Annotated genes, represented by arrows, are drawn to scale. Green arrows correspond to *phn* genes (darkgreen: *phnCDE*, green: *phnGHIJKLMNP* and turquoise: *phnF*), orange arrows indicate transcriptional regulatory proteins, brown arrows correspond to acethyltransferases. The remaining proteins (transporters, phosphate, etc.) are represented by yellow arrows, except for hypothetical proteins which are shown in grey. Annotations are abbreviated as follows: hyp, hypothetical protein; ABC, ABC transporter; phos, phosphatase.

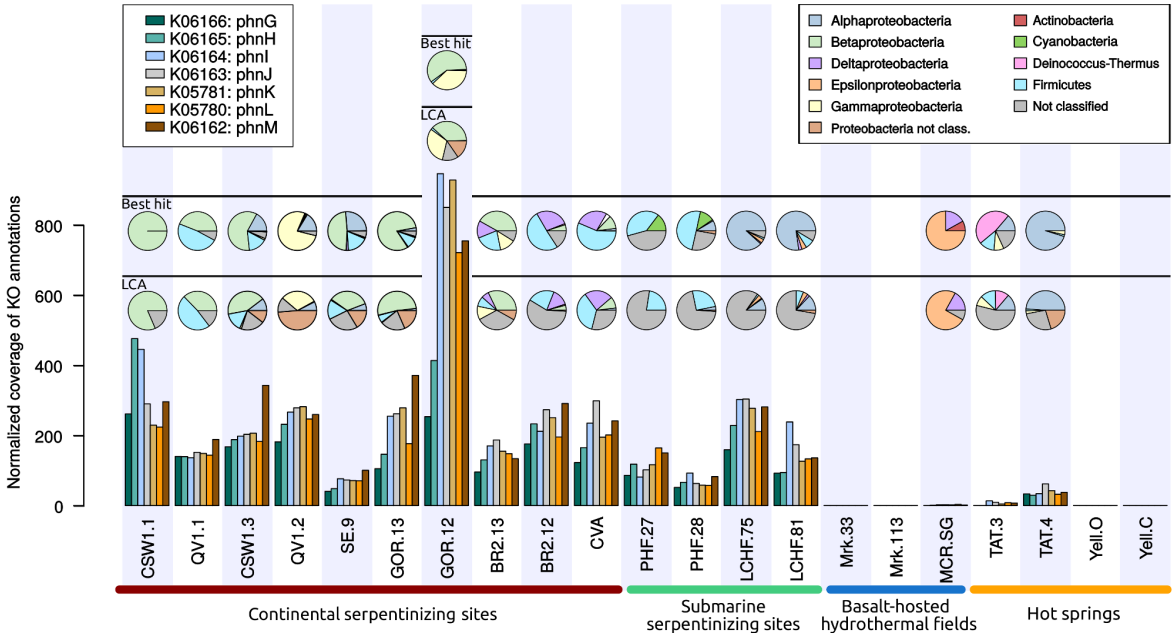


Figure 4.9 – Serpentinizing metagenomes show a higher abundance of genes encoding the catalytic component of C-P lyase than the other metagenomes investigated in the present study. The barplot corresponds to the normalized distribution of each KO annotation associated to *phnGHIJKLM* genes and the pie charts indicate their taxonomic distribution per metagenome among phyla and proteobacterial classes.

Table 4.4 – Frequency of *phn* genes in the 21 studied metagenomes and in a sample from P-depleted waters (Sargasso Sea). Results are expressed as a percentage of microorganisms estimated to contain gene of interest (assuming *recA* is found once per prokaryotic genome).

Type	Sample ID	K06166	K06165	K06164	K06163	K05781	K05780	K06162	Total
Terrestrial serpentinizing sites	QV1.1	18.62	18.57	18.15	20.20	19.78	19.13	25.05	19.93
	CSW1.3	25.22	28.28	29.75	30.58	31.01	27.51	51.57	31.99
	CSW1.1	25.45	46.36	43.39	28.24	22.35	21.81	28.84	30.92
	QV1.2	29.66	37.86	43.51	45.48	46.06	40.32	42.40	40.76
	SE.9	3.67	4.37	6.92	6.61	6.47	6.39	9.09	6.22
	GOR.13	7.37	10.26	17.84	18.34	19.52	12.39	26.00	15.96
	GOR.12	16.18	26.38	60.36	54.22	59.21	46.01	48.14	44.36
	BR2.13	5.76	7.85	10.25	11.26	9.34	8.91	8.05	8.77
	BR2.12	6.73	8.92	8.12	10.46	9.60	7.49	11.14	8.92
	CVA	11.84	15.98	22.72	28.87	18.86	19.46	23.34	20.15
Submarine serpentinizing sites	PHF.27	9.53	13.03	8.98	11.25	12.86	18.12	16.58	12.91
	PHF.28	6.36	8.16	11.39	7.77	7.18	7.09	10.18	8.30
	LC.75	18.11	25.95	34.41	34.56	31.57	23.99	32.01	28.66
	LC.81	11.08	11.28	28.63	20.85	15.20	16.00	16.32	17.05
Basalt-hosted hydrothermal fields	Mrk.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mrk.113	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	MCR.SG	0.00	0.06	0.09	0.10	0.08	0.16	0.07	0.08
Hot springs	TAT.3	0.00	0.00	1.60	1.15	0.55	0.98	0.85	0.73
	TAT.4	4.29	3.83	4.40	7.97	5.52	4.17	4.86	5.01
	Yell.O	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Yell.C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sargasso Sea	Sargasso 7(*)	24.00	32.00	32.00	24.00	19.00	23.00	37.00	27.29

(*) Karl et al. (2008)

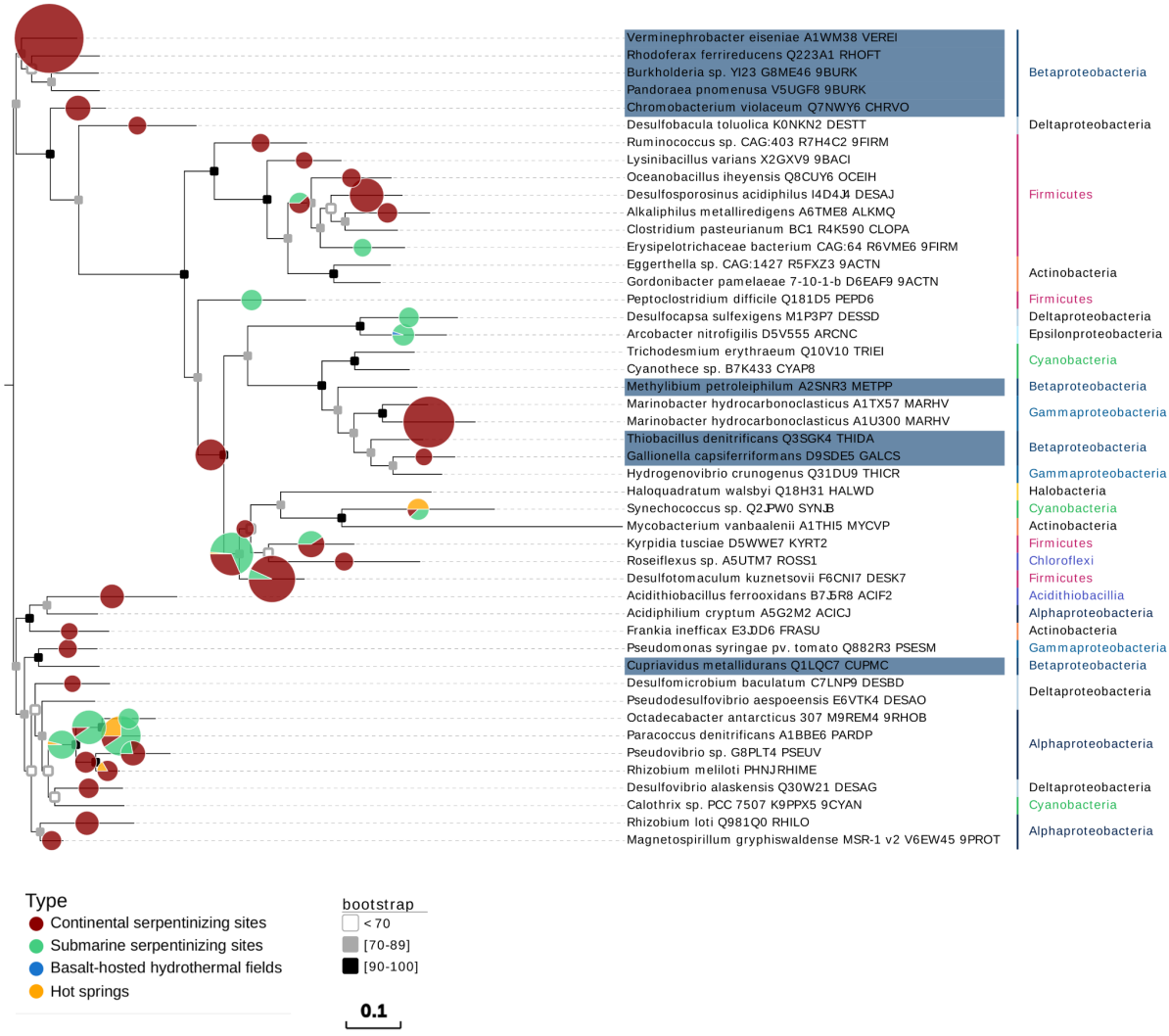


Figure 4.10 – Phylogenetic diversity of putative *phnJ* genes detected in the 21 metagenomes. The taxonomic affiliation of phnJ genes (catalytic function, K06163) was investigated by aligning the genes from metagenomics samples with MUSCLE (Edgar, 2004) and then inserting them with pplacer v1.1.alpha17 (Matsen et al., 2010) onto a reference phylogenetic tree (imported from Pfam, PF06007). The size of the circles is correlated to the coverage of the genes, normalized by the total number of annotated genes per metagenome. *Betaproteobacteria* species are highlighted in blue. The scale bar is equal to 0.1 changes per amino acid residue.

4.4 Discussion

4.4.1 Microbial communities of submarine serpentinizing systems

Our analysis based on taxonomic profile showed that PHF clustered with the continental serpentinizing systems, while LCHF grouped with the basalt-hosted hydrothermal vents (resulting from magmatic contributions). At first sight, this result was surprising because there are numerous phylotypes identified as typical of serpentinizing systems, including the LCMS or members of the genus *Ralstonia* (Frouin et al., 2018, Quéméneur et al., 2014). The clustering of LCHF metagenomes with those of deep-sea hydrothermal fields was mainly due to sequences assigned to genera of the class ϵ-*proteobacteria*, which were highly abundant in LCHF and basalt-hosted hydrothermal sites but nearly absent in PHF and other serpentinizing sites. This is consistent as ϵ-*proteobacteria* members are typically associated with sulfidic environments such as deep-sea hydrothermal vents (Flores et al., 2011, Reveillaud et al., 2016, Urich et al., 2014), where they thrive by oxidation of reduced sulfur compounds at low oxygen tensions. In addition, sulfides are barely detectable in continental serpentinite springs and in PHF, whereas H_2S concentration in LCHF is more in the range of that of basalt-hosted hydrothermal vents, likely owing to its setting proximal to a mid-ocean ridge (Kelley et al., 2001, Konn et al., 2009). Additionally, earlier studies have demonstrated spatiotemporal heterogeneity within the large chimney structure, including differential distributions of archaea and bacteria (Brazelton et al., 2010a, Kelley et al., 2005, Schrenk et al., 2004). It is possible that the chimney sub-samples used in the metagenomic analysis from the LCHF represented a different niche with the chimney habitat that favored sulfur oxidation.

4.4.2 Functional capabilities within serpentinizing ecosystems

In all serpentinizing sites, we observed the overabundance of the Group 3d [NiFe]-hydrogenase (*hoxHY*), which suggested an equal importance of hydrogenotrophic metabolism across terrestrial and submarine systems. This cytoplasmic hydrogenase was presumably essential for autotrophic growth of members of the *Betaproteobacteria* in The Cedars and CROMO serpentinites springs (Suzuki et al., 2014, Twing et al., 2017). Such hydrogenases are known to be oxygen-tolerant in *Ralstonia eutropha* H16 (Burgdorf et al., 2006), which may be of benefit to microbial communities living at the interface between anoxic and oxic zones. No specificity related to nitrogen metabolism was observed for microbial communities in serpentinizing systems compared to other hydrothermal systems. The diversity of geochemical characteristics in the serpentinizing systems made the neutral distribution of the *nifH* gene surprising. Indeed, nitrogen fixation was previously assumed to be more favorable in highly reducing and energetic conditions of submarine ecosystems rather than in continental environments (Brazelton et al., 2011). Further, DNRA can also occur in almost all serpentinizing sites studied, based on the distribution of genetic markers. This pathway is presumed to be more advantageous than denitrification in marine environments, because ammonia can be reused as nitrogen sources with-

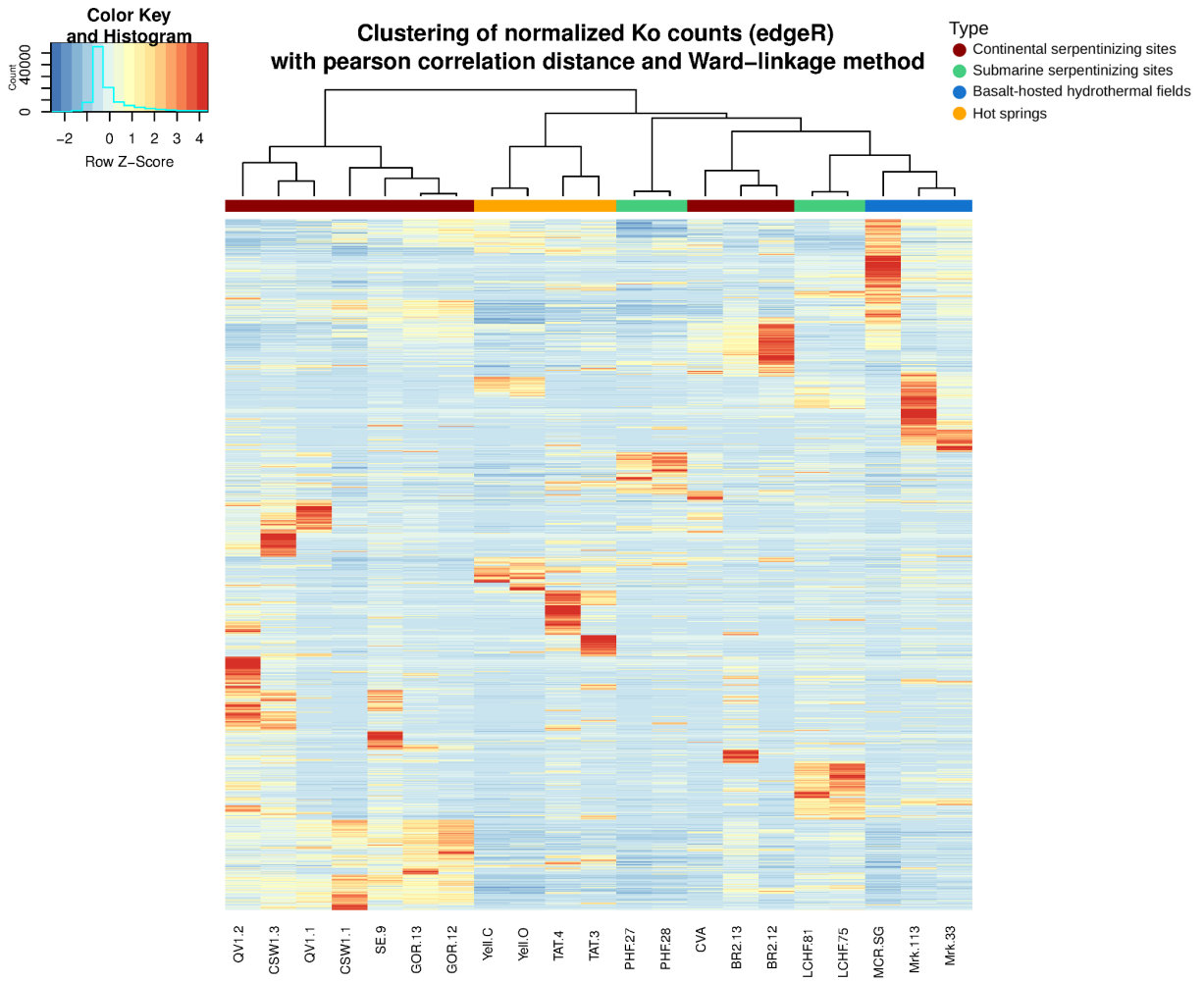


Figure 4.11 – Dissimilarity between functional profiles of the 21 metagenomes. The dendrogram includes the functional annotation KO with at least 20 hits in 2 metagenomes. Colors highlight the kind of ecosystems. To improve the readability of these plots, the KO counts were normalized using a z-score (centered and scaled in row direction) and then filtered with two criteria: i) they must have at least two non-null occurrences and ii) the sum of their occurrence must be higher than an empirical threshold of 10.

out expending energy in nitrogen fixation (Kim et al., 2014). Despite the high concentration of methane in serpentinizing sites, the potential for methanotrophy or methanogenesis differed from one site to another (Figure 4.4). Additionally, biological methane production has been demonstrated in some serpentinizing ecosystems, but is inconsistent between sites (Brazelton et al., 2016, Crespo-Medina et al., 2017, Kohl et al., 2016). The variable origins of methane within these sites probably results from distinct mixtures of thermogenic, abiogenic, and microbial methane as proposed for two serpentinizing systems, CROMO and The Cedars (Wang et al., 2015). At CROMO, notwithstanding methanogenesis pathway were not reported, two OTUs (one *Betaproteobacteria* and one *Clostridia*), were positively correlated with the concentration of methane. The main taxonomic groups linked to C-P lyase and identified in the terrestrial serpentinizing systems nurture this hypothesis that the phosphonate degradation is active and leads to methane production. The extreme conditions of life imposed by reactions of serpentinization do not impact the associated microbial communities to such an extent that an overall

convergence of their functional profiles can be observed (Figure 4.11). However, the similarities between LCHF and other deep-sea hydrothermal systems highlighted by our hierarchical clustering have already been mentioned by Xie and collaborators through enrichments of genes involved in stress responses, homologous recombination and chemotaxis (Xie et al., 2011). The genomes of microbial communities were indeed subject to numerous physical constraints (hydrostatic pressure and temperature) and chemical stresses (heavy metals, radionuclides etc.) that were probably specific to deep-sea hydrothermal vents. Additionally, high abundances of mobile genetic elements such as transposases were also reported as a distinctive characteristic of LCHF and deep-sea hydrothermal vents (Anderson et al., 2014, Brazelton and Baross, 2009). It has been suggested that lateral gene transfer was a frequent event and could be an important source of phylogenetic diversity (Brazelton and Baross, 2009) as well as a way to promote bacterial adaptation to ecological niche. In our study, the high abundance of mobile genetic elements did not appear to be restricted to deep-sea hydrothermal vents, as they were also very common in other systems (data not shown).

4.4.3 Metabolism of reduced phosphorus compounds in serpentinizing systems

The supervised approach highlighted the importance of genes *phnGHIJKLM* in serpentinizing systems. These genes, encoding the functional core of the C-P lyase, should enable microorganisms to utilize Pn as source of P. Their abundance in serpentinizing systems (Table 4.5) supports the hypothesis that phosphonate degradation plays a major role in the microbial communities' growth. We believe, that, in serpentinizing systems, methane or other hydrocarbons could be partially produced during the microbial degradation of Pn to inorganic phosphate (Figure 4.12). The cleavage and incorporation of phosphonates in serpentinizing systems could prove an important means of phosphorus scavenging in this low nutrient environment. The type of Pn substrates remains to be to identified, because the C-P lyase cleaves a broad range of substrates according the species, including alkylphosphonates, 2-aminoethylphosphonate and phenylphosphonate (Schowanek and Verstraete, 1990).

Other pathways are known to cleave C-P bonds; they request very substrate-specific hydrolases such as phosphonoacetaldehyde hydrolase, phosphonoacetate hydrolase, and phosphonopyruvate hydrolase (Sviridov et al., 2012) or an oxidative cleavage with the 2-amino-1-hydroxyethylphosphonate dioxygenase (McSorley et al., 2012). The presence of many alternative pathways in Santa Elena Ophiolite may be surprising as the phosphate concentration of this site was higher than in other serpentinizing environments (Crespo-Medina et al., 2014, 2017). However, this result can be linked to the claim that the catabolism of Pn (via a phosphonoacetate hydrolase) was not restricted to phosphate-limited environment, being inorganic-phosphate insensitive (Chin et al., 2018). Although the C-P lyase pathway is now recognized as widely distributed in contemporary biosphere (Quinn et al., 2007), its identification in ecosystems that may have supported early life suggests a potential link to ancient metabolisms. Phos-

phonates were assumed present on the early Earth (McGrath et al., 2013) because alkylphosphonates were detected in the Murchison meteorite (Cooper et al., 1992). Furthermore, several mechanisms for organophosphate synthesis under prebiotic conditions have been reported (Gull et al., 2017) and may contribute to the tapestry of geochemical processes occurring in modern day serpentinizing environments. At present, geochemical data are missing to identify the origins of Pn in serpentinizing sites. Alternatively, the detection of phosphonate synthesis gene (*ppm*) could be associated with the presence of Pn in dissolved organic matter (Repeta et al., 2016) and be a path to understand Pn distribution. However, the gene markers of this biosynthesis pathway were only retrieved in deep-sea hydrothermal vents, Santa Elena Ophiolite and CROMO (Figure 4.12). The results also point towards the need to resolve sources of phosphorus compounds, including recycled phospholipids in biomass or abiotic synthesis reactions that could contribute to the occurrence of this pathway.

Table 4.5 – Ratio between the abundance of genes involved in alternative pathways of Pn degradation and the C-P lyase pathway. Metagenomic coverage of 4 genes (*phnW*, *phnX*, *phnA* and *phnZ*) were divided by the coverage of *phnJ* gene (C-P lyase) in serpentinizing systems.

		2AEP-pyruvate aminotransferase ***phnW*** **K03430**	**Phosphonatase** ***phnX*** **K05306**	**Phosphonoacetate hydrolase** ***phnA*** **K19670**	**Phosphohydrolase** ***phnZ*** **K21196**
Terrestrial serpentinizing sites	QV1.1	0	0	0	0
	CSW1.3	0,008	0	0	0
	CSW1.1	0,007	0	0	0
	QV1.2	0,544	0,57	0,018	0,177
	SE.9	1,161	1,021	0,059	0,015
	GOR.13	0,019	0,175	0,006	0,01
	GOR.12	0,01	0,059	0	0
	BR2.13	0,214	0,345	0	0
	BR2.12	0,056	0,128	0	0
	CVA	0,106	0,103	0	0
Submarine serpentinizing sites	PHF.27	0	0,269	0	0
	PHF.28	0,463	0,426	0	0
	LC.75	0,576	0,022	0,03	0
	LC.81	0,344	0,027	0	0

4.5 Conclusion

This broad comparative study of serpentinite-hosted ecosystems provides an objective framework to understand their ecology, while pointing out key differences between the sites, that provides wrinkles in our understanding of the microbial diversity of such systems. Taxonomic similarity was therefore demonstrated between PHF and other terrestrial serpentinite-hosted ecosystems. Despite serpentinization influence, our study showed that LCHF microbial communities were closer to those from deep basalt-hosted hydrothermal fields as opposed to continental ophiolites. Moreover, this comparative analysis revealed shared functional capabilities among serpentinite-hosted ecosystems to respond to environmental stress, oxidize abundant

hydrogen resources and catabolize phosphonates. These results are consistent with our generalized picture of serpentinite environments but provide deeper insight into the array of factors that may control microbial activities in these systems. Our results suggested that the Pn metabolism through the C-P lyase pathway is widespread among serpentinizing systems and could play a role that will require further investigations not only in P, but also methane biogeochemical cycles. This metagenomics analysis highlighted opportunities for future studies to quantify the reduced P compounds in the serpentinizing systems in parallel to activity-based studies to evaluate microbial degradation of Pn.

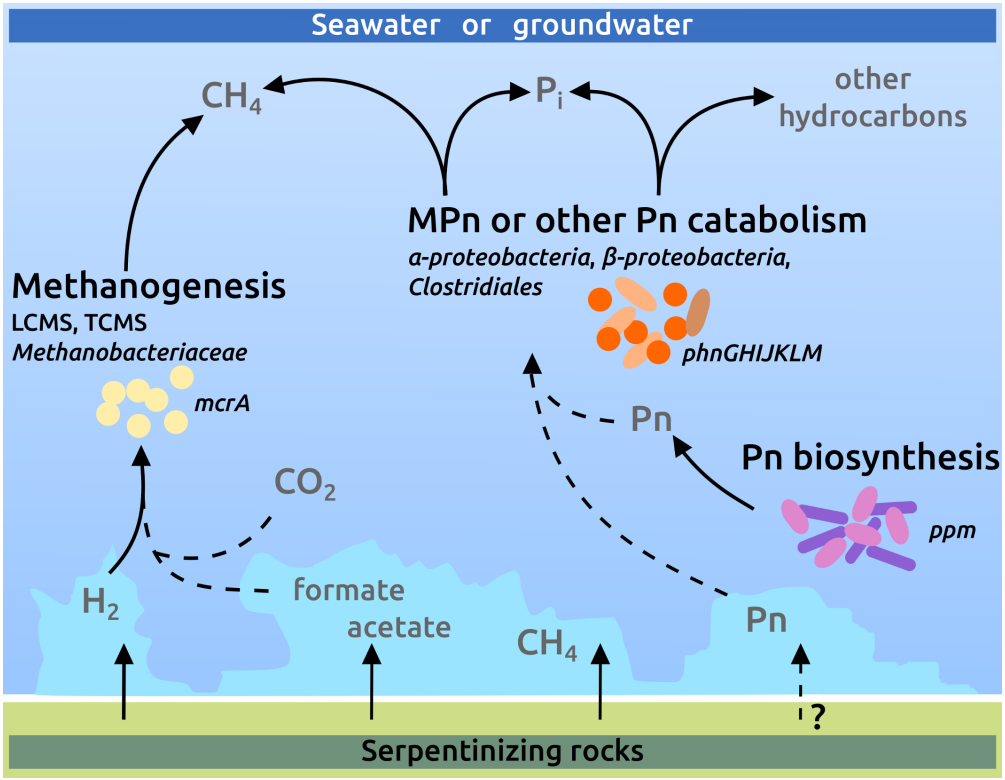


Figure 4.12 – Schematic of the proposed links between carbon and phosphorus cycles in serpentinizing environments. Reactions whose origin of the chemical compounds is unknown are indicated by dashed lines.

Chapter 5

Recovery of metagenome-assembled genomes from PHF [1]

Sommaire

[1] The work described in this chapter is the working material of an article in preparation

Overview (French part)

L'article intitulé "Metabolic diversity of microbial communities inhabiting the serpentinization-influenced Prony Hydrothermal Field" présente la reconstruction de génomes procaryotes détectés dans 5 métagénomes issus de différentes cheminées du site serpentinisé de Prony, en Nouvelle-Calédonie. Trois sites différents ont été échantillonnés: le Bain des Japonais, la Rivière des Kaoris, situés en zone intertidale et le site sous-marin ST07.

Dans un premier temps l'étude de la distribution taxonomique a permis d'estimer la diversité microbienne des 5 échantillons. Cette étude taxonomique a été réalisée à la fois en utilisant des données de metabarcoding et en utilisant des données métagénomiques. Quelques différences notables ont été observées entre ses deux approches. Premièrement, les primers universels ayant servi à l'amplification des séquences des gènes de l'ARNr 16S n'ont amplifiés aucune séquence provenant d'archées. Il s'agit d'un biais important car plus ~ 10% d'archées ont été détectées via la seconde méthode. De plus, l'estimation du nombre d'organismes affiliés aux *Acetothermia* est bien supérieure parmi les séquences d'amplicons du gène de l'ARNr 16S que dans les contigs métagénomiques. La méthode basée sur les données de métagénomique a quand à elle une faible résolution pour l'affiliation taxonomique des espèces.

Des méthodes de reconstruction *in-silico* de génomes ont permis de dresser l'ébauche de 82 génomes procaryotes appelés MAGs (pour metagenomes-assembled genomes). Ces MAGs sont affiliés à 18 phyla différents, incluant notamment des *Firmicutes*, des *Deinococcus-Thermus*, des *Bacteroidetes* et des *Proteobacteria*. Cependant, en raison de la forte diversité des communautés microbiennes retrouvées à Prony, ces génomes reconstruits ne correspondent qu'à une faible proportion de la communauté globale. L'annotation fonctionnelle des MAGs révèle qu'une large partie des MAGs identifiés a un métabolisme aérobie. Certains MAGs peuvent avoir d'autres accepteurs d'électrons que l'oxygène tels que le nitrate ou encore le sulfite. Par ailleurs, le cycle de Calvin-Benson-Bassham est la seule voie de fixation du carbone qui a été retrouvée dans les MAGs. Étonnamment, il a été noté que la majorité des MAGs reconstruits étaient porteurs de gènes de résistance aux métaux (tels que *czcA*, *copA*, *tehB*). L'annotation fonctionnelle de la seule archée identifiée révèle d'importantes lacunes métaboliques. Nos résultats suggèrent que cette archée, affiliée aux *Woesearchaeota* a un mode de vie symbiotique.

Ce chapitre s'intéresse plus particulièrement aux annotations d'un MAG, afflilé au *Burkholderiales* (β-*proteobacteria*). D'après nos inférences phylogéniques, ce MAG semble appartenir au genre candidat *Serpentinomonas*. Finalement, la comparaison génomique de ce MAG et des autres souches du genre *Serpentinimonas* révèle d'importantes similarités. Ce résultat suggère que le phylotype retrouvé à Prony pourrait croître en condition autotrophe. Le phylotype isolé à PHF possède aussi une voie de dégradation du benzène et du phénol qui pourrait lui permettre d'utiliser des composés aromatiques comme source de carbone.

5.1 Introduction

The serpentinization is a geochemical process that results from in-depth water circulation and interactions with ultramafic rocks (peridotite) near slow-spreading oceanic ridges, subduction zones or terrestrial ophiolites. The serpentinization and associated reactions produce abiotic compounds, such as hydrogen, methane and small organic molecules. The serpentinizing environments provide both potential energy sources and organic material to support chemosynthetic microbial communities (Schrenk et al., 2013). Abiotic and biotic interactions in serpentinizing systems are of interest as they could support a hypothetical scenario of early microbial life (Martin et al., 2008, Schulte et al., 2006).

The Prony Hydrothermal Field (PHF) is an active serpentinizing hydrothermal system, located in the Prony Bay on the New Caledonia ophiolitic nappe. The alkaline hydrothermal fluids (pH 9-11) are discharged into lagoon seawater, where they precipitate and form carbonate chimneys (Monnin et al., 2014). The PHF system comprises several hyperalkaline springs emerging at shallow submarine sites (up to 50 m below sea level) or intertidal zones, (covered at high tide). The emitted fluids have a low salinity since they come from meteoric waters percolating through a peridotite nappe. The fluids have a moderate temperature, below 40°C and are enriched in H_2 (12-30% in volume of dry gas), CH_4 (6-14%) and N_2 (56-72%) (Monnin et al., 2014). Contrary to deep-sea hydrothermal systems (e.g., The Lost City Hydrothermal Field, LCHF), which are only colonized by chemosynthetic organisms, shallow hydrothermal systems host complex microbial communities that also includes photosynthetic organisms (Canet et al., 2007). At PHF, microbiological studies identified a highly diversified microbial community, despite the extreme pH (Postec et al., 2015, Quéméneur et al., 2014). Indeed, the microbial diversity found in PHF showed typical populations of serpentinite environments in terrestrial and fully marine environments, such as members of the genera *Dethiobacter* or *Hydrogenophaga*, or the dominant archaeal phylotypes closely related to the Lost City Methanosarcinales (LCMS) and the The Cedars Methanosarcinales (TCMS) (Frouin et al., 2018, Postec et al., 2015). The location of hydrothermal chimneys in intertidal or submarine sites seemed to affect the bacterial diversity, but not the archaeal diversity. The impact of this environmental factor on metabolic potentials remains unexplored.

Anaerobic cultures were performed to gain further insight into the metabolism of microbes living in contact or in the hydrothermal fluids. Five anaerobic strains that belonged to the order *Clostridiales* could thus be isolated from PHF sites (Ben Aissa et al., 2015, 2014, Bes et al., 2015, Mei et al., 2016a, 2014). These fermentative organisms were hypothesized to provide primary production (Schrenk et al., 2013). The characterization of these strains highlighted metabolic capacities of microorganisms inhabiting PHF chimneys, such as their efficiency in producing hydrogen at high pH levels (Mei et al., 2014). However, most of microbes associated with abiotic compounds naturally present in the serpentinizing chimneys of PHF (e.g., H_2, CH_4) could not be cultivated. The best illustration of the limitations of culture-based approaches is the recurrent unsuccessful attempts to isolate LCMS both in the PHF and LCHF (Brazelton et al.,

2011, Quéméneur et al., 2014). There are still many unknowns about the microbial actors of these ecosystems and amongst the most intriguing ones who are the methanogens and what is the metabolic potential of microbes belonging to candidate divisions. Advances in high-throughput sequencing and computational techniques facilitate the cultivation-independent recovery of genomes from metagenomes. Such genome reconstruction has already been addressed in two serpentinizing environments, the alkaline springs from The Cedars and the Voltri Massif, to identify the potential metabolic individual microorganisms (Brazelton et al., 2016, Suzuki et al., 2017). Suzuki and colleagues thus managed to reconstruct dominant microorganisms from the hyperalkaline springs of The Cedars. Among these genomes,some were affiliated with the candidate phylum *Parcubacteria*. These genomes were highly reduced, often with no ATP-synthase and lacked one or more key biosynthetic pathways. The authors suggested that these microbes were incapable of independent growth and could be either symbiont of other microbes or scavengers of dead cells (Suzuki et al., 2017).

The understanding of the functional capabilities of the microorganisms living at the PHF remains very incomplete. Among the microbial diversity of PHF chimneys, some members of *Firmicutes*, *Chloroflexi* and *Betaproteobacteria* were found in other serpentinizing sites. This observation suggests that these phylotypes play a key role in the functioning of serpentinizing ecosystems. To illustrate the life strategies of the microorganisms associated with the serpentinization occurring at PHF, we used a genome-centric metagenomic approach. We present here the taxonomic affiliations and metabolic capabilities of the reconstructed genomes. We focused especially on a genome closely related to alkaliphilic strains belonging to the proposed genus "*Serpentinomonas*" within the class *Betaproteobacteria*. These strains previously isolated from highly alkaline serpentinizing springs at The Cedars. They were reported to be highly adapted to the harsh geological settings observed in serpentinite-hosted terrestrial ecosystems and are capable of autotrophic growth with $CaCO_3$ as source of carbon to overcome limited concentration of dissolved inorganic carbon (Suzuki et al., 2014).

5.2 Material and Methods

5.2.1 Sampling sites and sample collection

Five hydrothermal chimneys of the Prony Bay (New Caledonia) were sampled in 2011 and 2014 for microbiological analyses. Three small chimneys (~10 cm in height and ~3 cm in diameter) were collected the intertidal sites of "Rivière des Kaoris" (RdK) and "Bain des Japonais" (BdJ). Two sections of carbonated chimneys -a core and a top section- were sampled in the shallow submarine site ST07, also called "Aiguille de Prony". Details of sampling locations may be found in previous publication (Mei et al., 2016b). Chimney samples were crushed and stored in sterile FalconTM tubes at -80°C until the DNA was extracted.

5.2.2 Total community DNA extraction, amplicons and sequencing

Whole genomic DNA was extracted using FastDNA SPIN Kit for Soil (MP Biomedicals, Solon, OH, USA). Half of extracted DNA was amplified by PCR reactions using the primers Pro341F and Pro805R, which target the V3-V4 region of the prokaryotic 16S rRNA gene (Takahashi et al., 2014). PCR amplifications were performed by the Molecular Research Laboratory (Texas, USA) as well as the libraries preparation and sequencing on an Illumina Miseq run that generated 300 bp paired-end reads. The remaining half of the extracted DNA was sent to the Get-Plage (GenoToul, Toulouse, France) for whole-genome sequencing. The paired-end libraries (insert size ~380 bp) were prepared using the Illumina TruSeq DNA Nano kit (Illumina, Inc., San Diego, CA, USA) and were sequenced using an Illumina HiSeq3000 for 150bp paired-end reads.

5.2.3 Analyses of 16 rRNA amplicon data

The raw file containing all sequences was demultiplexed into individual files before the barcodes and adaptors were removed using Cutadapt v1.9.1 (Martin, 2011). Amplicon sequences were processed using Divisive Amplicon Denoising Algorithm 2 (DADA2, v1.6) (Callahan et al., 2016). Briefly, quality filtering of amplicons were performed using the function filterAndTrim(), to truncate the forward and reverse reads at position 270 and 220, respectively and discard reads with a maxEE of 5. The samples were pooled to can increase sensitivity of DADA2 to Amplicon Sequence Variants (ASVs) that may be present at very low frequencies in multiple samples. The error errors introduced by PCR amplification and sequencing were identified using the function learnErrors(), based on a parametric model. Chimeric ASVs were removed using the command removeBimeraDenovo(). Taxonomic annotation was assigned to ASVs using the RDP naive Bayesian classifier (Wang et al., 2007) with the SILVA database v123 (Quast et al., 2013).

5.2.4 Metagenomes assembly, mapping, and genome binning

Raw shotgun sequencing reads were dereplicated using cd-hit-dup v4.6.6 (Li and Godzik, 2006) with defaults parameters. To remove low quality bases and adaptors, the reads were trimmed using the software Trimmomatic v0.32 (Bolger et al., 2014) with a minimal quality of 15 over a window of 4 bases and a minimum length of 35 pb. The five metagenomes were individually assembled using the de-novo assembler IDBA-UD and default parameters (Peng et al., 2012). The gene content was identified using Prodigal v47 with the option dedicated to metagenomics sequences (Hyatt et al., 2010). The microbial taxonomic composition of each metagenome was estimated using similarities searches of the gene content against NCBI-nr with DIAMOND v.0.8.34 (Buchfink et al., 2014) and the Lowest Common Ancestor (LCA) algorithm with MEGAN (Huson et al., 2016). Reads from all metagenomes were finally co-assembled with IDBA-UD. The trimmed reads were back mapped on the co-assembled contigs with BWA v.0.7.12-r1039 (Li and Durbin, 2009) using the BWA-MEM algorithm with default parameters. Coassembled contigs >= 2500 bp were binned using MetaBAT v0.26.1 (Kang et al., 2015) under the preset

parameter settings "verysensitive" to recover Metagenome-Assembled Genomes (MAGs). The quality of the genomes (i.e., contamination and completeness) recovered were estimated using CheckM (Parks et al., 2015) with default parameters. Only MAGs with completeness >70% and contamination <5% were considered for further refinement and validation. A total of 82 MAGs were manually refined using the Anvi'o interactive graphical interface (Eren et al., 2015) to visualize contigs and their similarity in terms of sequencing depth and GC content. The quality of the refined MAGs was reassessed using CheckM.

5.2.5 Taxonomic and functional annotation of MAGs

Putative taxonomic annotations were assigned to MAGs using CheckM (Parks et al., 2015). These annotations were further refined with the phylogenomic placement of MAGs using PhyloPhlAn software (Segata et al., 2013). A phylogenetic tree, including the MAGs identified in this study, was built using PhyloPhlAn and a selection of complete genomes and MAGs from GenBank (listed in Table 5.1). Briefly, conserved proteins were identified using USEARCH (Edgar, 2010) and aligned against the built-in database using MUSCLE (Edgar, 2004). Finally, a maximum-likelihood tree was generated using FastTree (Price et al., 2010). Putative genes of each MAG were predicted and functionally annotated using the Prokka pipeline for metagenomics (Seemann, 2014) with the argument -kingdom set according to the predicted taxonomy of the MAGs. The presence of genes encoding metabolic and other functions of interest (listed in the Table 5.2) were sought in the near complete MAGs (39 MAGs were at least 90% complete). Hydrogenase signatures proposed by Vignais and Billoud (2007) were used to precisely identify the group of the hydrogenases detected in MAGs.

5.2.6 Identification of a MAG related to the genus *Serpentinomonas*

The comparison of the MAGs and the 100,000 genomes in GenBank using MinHash sketches (Brown and Irber, 2016) produced very few hits. However, one MAG was closely related to a strain belonging to the candidate genus *Serpentinomonas*, isolated from serpentinizing springs at The Cedars. Several analyses were performed to estimate the similarity between this MAG and the *Serpentinomonas* strains already characterized. First, average nucleotide identity (ANI) was calculated for each combination of pairs of the genomes, using ANI Calculator (Rodriguez-R and Konstantinidis, 2016). A phylogenetic tree comprising the genomes putatively assigned to the genus Serpentinomonas and other *Comamonadaceae* genomes was computed using PhyML and 1,000 bootstrap resamplings (Guindon and Gascuel, 2003). The maximum likelihood tree was inferred from a concatenated alignment of 36 phylogenetically conserved proteins, listed in Table 5.3. The sequences of each protein had previously been individually aligned using MUSCLE (Edgar, 2004). Finally, the *Serpentinomonas* strains were submitted to the tool BlastKOALA (Kanehisa et al., 2016), using default parameters for annotation using the KEGG Ontology (KO) system. KEGGmapper (`http://www.genome.jp/kegg/mapper.html`) was used to identify the main differences between metabolic pathways of these genomes.

Table 5.1 – List of genomes used to refine the phylogeny of the MAGs (Figure 5.2)

phylum	strain	NCBI ID
Armatimonadetes		
	Fimbriimonas ginsengisoli Gsoil 348	NZ_CP007139
	Armatimonadetes bacterium 55-13	MKRX01
	Chthonomonas calidirosea T49	NC_021487
Acidobacteria		
	Acidobacteria bacterium RIFCSPLOWO2_02_FULL_61_28	MEKR01
	Holophaga foetida TMBS4, DSM 6591	AGSB02
	Granulicella rosea DSM 18704	FZOU01
	Pyrinomonas methylaliphatogenes K22	CBXV01
	Acidobacteria bacterium isolate AA35	QHVJ01
	Chloracidobacterium sp. UBA3360	DEQS01
	Occallatibacter savannae	QFFY01
	Candidatus Sulfotelmatobacter kueseliae	OMOD01
Actinobacteria		
	Thermobispora bispora R51, DSM 43833	NC_014165
	Actinobacteria bacterium IMCC25003	NZ_CP015603
	Actinobacteria bacterium HGW-Actinobacteria-10	PHFA01
Balneolaeota		
	Rhodohalobacter barkolensis strain 15182	PISP01.1
	Balneola sp. EhC07	LXYG01
	Gracilimonas tropica DSM 19535	AQXG01
Bacteroidetes		
	Draconibacterium orientale FH5	NZ_CP007451
	Prevotella sp. F0039	NC_022124
	Flavobacteriales bacterium isolate NORP157	NVRW01
	Haliscomenobacter hydrossis O, DSM 1100	NC_015510
	Bacteroidetes bacterium CG2_30_32_10	MNXO01
	Mucilaginibacter mallensis MP1X4	NZ_LT629740
	Cytophaga hutchinsonii ATCC 33406	NC_008255
	Fluviicola taffensis RW262, DSM 16823	NC_015321
	Bacteroidetes bacterium OLB12	LNFR01
	Porphyromonas asaccharolytica VPI 4198, DSM 20707	CP002689
	Belliella baltica DSM 15883	NC_018010
	Chitinophaga pinensis UQM 2034, DSM 2588	NC_013132
	Rhodothermus marinus DSM 4252	NC_013501
	Rhodothermus profundi strain DSM 22212	FRAU01
Chloroflexi		
	Dehalococcoides sp. JdFR-56	MTNR01
	Anaerolineaceae bacterium NAT123	PABI01
	Dehalogenimonas alkenigignens IP3-3	LFDV01
	Chloroflexi bacterium isolate ARS34	NZAI01
Cyanobacteria		
	Gloeobacter violaceus PCC 7421	NC_005125
	Gloeomargarita lithophora Alchichica-D10	NZ_CP017675
	Filamentous cyanobacterium CCP5	PYEQ01
	Leptolyngbya valderiana BDU 20041	LSYZ01
	Synechococcus sp. JA-2-3B'a(2-13)	NC_007776
	Lyngbya confervoides BDU141951	JTHE01
Deinococcus-Thermus		
	Deinococcus radiodurans ATCC 13939	JHYL01
	Thermus thermophilus HB8	NC_006461
	Oceanithermus profundus 506, DSM 14977	NC_014761
	Truepera radiovictrix RQ-24, DSM 17093	NC_014221
	Thermus scotoductus SA-01, ATCC 700910	NC_014974
	Meiothermus ruber DSM 1279	NC_013946
	Deinococcus proteolyticus MRP, DSM 20540	NC_015161
	Marinithermus hydrothermalis T1, DSM 14884	NC_015387
	Thermus aquaticus Y51MC23	NZ_CP010822
	Deinococcus deserti VCD115	CP001114
	Meiothermus silvanus VI-R2, DSM 9946	NC_014212
	Deinococcus maricopensis LB-34, DSM 21211	NC_014958
	Deinococcus geothermalis DSM 11300	NC_008025
Firmicutes		
	Orenia metallireducens strain MSL47	OBDZ01
	Alkaliphilus transvaalensis	JHYF01
	Desulfotomaculum nigrificans CO-1-SRB	NC_015565
	Symbiobacterium thermophilum IAM 14863	NC_006177
	Alkaliphilus peptidifermentans	FMUS01
	Heliobacterium modesticaldum Ice1	NC_010337.2
	Acidaminococcus fermentans DSM 20731	NC_013740
	Dethiobacter alkaliphilus AHT 1	ACJM01
	Bacillus sp. 1NLA3E	NC_021171
	Natranaerobius thermophilus JW/NM-WN-LF	NC_010718
	Peptococcaceae bacterium JdFR-69	MTOE01
	Thermoanaerobacterales bacterium JdFR-68	MTOD01
	Erysipelothrix rhusiopathiae str. Fujisawa	NC_015601
	Anaerovirgula multivorans	FZOJ01
	Thermosinus carboxydivorans Nor1	AAWL01
Gemmatimonadetes		
	Gemmatirosa kalamazoonesis strain KBS708	NZ_CP007128
	Gemmatimonadetes bacterium RBG_16_66_8	MGZE01
	Gemmatimonas phototrophica strain AP64	NZ_CP011454
	Gemmatimonas aurantiaca T-27	NC_012489
Lentisphaerae		
	Lentisphaerae bacterium RIFOXYC12_FULL_60_16	MHBQ01
	Lentisphaera araneosa HTCC2155	ABCK01
	Lentisphaerae bacterium GWF2_45_14	MHBG01
Omnitrophica		
	Omnitrophica WOR_2 bacterium SM23_72	LJUU01
	Candidatus Omnitrophus magneticus SKK-01	JYNY01
	Omnitrophica WOR_2 bacterium SM23_29	LJUB01
Parcubacteria		
	Parcubacteria bacterium GW2011_GWD2_40_9	LCAA01
	Candidatus Parcubacteria bacterium 4484_255	NATD01
	Candidatus Parcubacteria bacterium A4	MTEP01
	Candidatus Nomurabacteria bacterium isolate UBA8515	DPIZ01
Planctomycetes		
	Blastopirellula marina DSM 3645	AANZ01
	Planctomycetaceae bacterium CPC83	NYSY01
	Phycisphaerae bacterium SP96	PAAB01
	Planctomyces sp. SH-PL14	NZ_CP011270
	Pirellula staleyi DSM 6068	NC_013720
	Phycisphaera mikurensis NBRC 102666	NC_017080
Proteobacteria		
α	Parvularcula oceani JLT2013	JPHU01
α	Parvibaculum lavamentivorans DS-1	NC_009719
α	Micavibrio aeruginosavorus ARL-13	NC_016026
α	Methylopila sp. M107	ARWB01000001
α	Albidovulum xiamenense DSM 24422	PVND01
α	Belnapia rosea CPCC 100156	FMZX01
α	Henriciella marina DSM 19595	AQXT01
α	Roseobacter sp. SK209-2-6	AANB01
α	Rhodomicrobium udaipurense JA643	JFZJ01
β	Methylobacillus sp. MM3	LXTQ01
β	Comamonadaceae bacterium B1	NZ_AP014569
γ	Chromatiales bacterium RIFOXYA1_FULL_46_5	MGOR01
γ	Thioalkalivibrio sulfidophilus HL-EbGR7	CP001339
γ	Sulfuricaulis limicola HA5	NZ_AP014879
γ	Arenimonas metalli CF5-1	AVCK01
γ	Acidiferrobacteraceae bacterium isolate ARS46	NYZW01
γ	Acidiferrobacter thiooxydans strain ZJ	MDCF01
γ	Acidiferrobacter sp. SPIII	NZ_CP027663
γ	Methylomicrobium alcaliphilum 20Z	NC_016112
γ	Rhodanobacter thiooxydans LCS2	AJXW01
γ	Methylocaldum szegediense O-12	ATXX01
γ	Methylomarinum vadi IT-4	JPON01
γ	Silanimonas lenta DSM 16282	AUBD01
γ	Legionellales bacterium RIFCSPHIGHO2_12_FULL_35_11	MHBB01
γ	Methylovulum psychrotolerans HV10_M2	NZ_CP022129
δ	Desulfurivibrio alkaliphilus AHT2	NC_014216
δ	Desulfomicrobium baculatum X, DSM 4028	NC_013173
δ	Desulfatibacillum alkenivorans AK-01	NC_011768
δ	Desulfobacca acetoxidans ASRB2, DSM 11109	NC_015388
δ	Myxococcales bacterium isolate fen_1125	PMJO01
δ	Myxococcus fulvus HW-1	NC_015711
o	Bdellovibrio sp. isolate Amazon FNV	PSRN01
o	Pseudobacteriovorax antillogorgiicola RKEM611	FWZT01
o	Halobacteriovorax marinus SJ	NC_016620
o	Oligoflexus tunisiensis Shr3	BDFO01000001
Spirochaetes		
	Marispirochaeta aestuarii strain JC444	MWQY01
	Sphaerochaeta globosa Buddy	NC_015152
	Spirochaetes bacterium GWB1_48_6	MIAP01
	Treponema caldarium DSM 7334	NC_015732
	Alkalispirochaeta americana ASpG1	FTMS01
	Leptospira sp. JW3-C-A1	NPDO01
Verrucomicrobia		
	Methylacidiphilum infernorum V4	NC_010794
	Coraliomargarita akajimensis DSM 45221	NC_014008
	Verrucomicrobium spinosum DSM 4136	ABIZ01
	Terrimicrobium sacchariphilum strain NM-5	BDCO01
Woesearchaeota		
	Woesearchaeota archaeon ARS73	NYYS01
	Candidatus Woesearchaeota archaeon CG06_land_8_20_14	PEWB01
	Candidatus Woesearchaeota archaeon CG1_02_47_18	MNWV01
Archaea		
	Methanosarcina sp. WWM596	NZ_CP009503
	Thermococcus sp. AM4	NC_016051
	Haloarcula japonica DSM 6131	AOLY01

Table 5.2 – List of key genes encoding different metabolisms and searched in the near complete MAGs

Metabolic process or bioenergetic complex	Key gene	Gene product
NITROGEN METABOLISM		
Nitrate reduction	*narG*	Cytoplasmic nitrate reductase,
	narZ	Respiratory nitrate reductase,
	narB	Ferredoxin-nitrate reductase,
	nasA	Assimilatory nitrate reductase,
	napAB	periplasmic nitrate reductase
Nitrite reduction	*nirK**, *nirS*	Nitrite reductase (NO-forming)
	*nirB**	Cytoplasmic nitrite reductase (NH4-forming),
Nitrite reduction	*nrfA*	Periplasmic nitrite reductase (NH4-forming),
	NIT-6	Nitrite reductase (NAD(P)H)
Nitrogen fixation	*nifH*	Nitrogenase
Annamox*	*hzsA*	Hydrazine synthase
SULFUR METABOLISM		
Sulfate reduction*	*aprA*	Adenylyl-sulfate reductase
sulfide oxidation	*fccB*	flavocytochrome sulfide dehydrogenase,
	*sqr**	Sulfide:quinone oxidoreductase
Sulfurtransferase	*dsrEF*	reverse acting dissimilatory sulfite reductase
thiosulfate oxidation	*soxY*	sulfur-oxidizing protein,
	*soxB**	S-sulfosulfanyl-L-cysteine sulfohydrolase
sulfite reduction	*dsrAB*	Dissimilatory sulfite reductase,
	asrAB	Anaerobic sulfite reductase
OXYGEN METABOLISM		
Oxygen respiration	*qoxC*	Cytochrome aa3,
	cydAB	Cytochrome bd quinol oxidase,
	ccoN	cbb3-type oxidase,
	cyoC	Cytochrome bo(3) ubiquinol oxidase
METHANE METABOLISM		
Methanogenesis*	*mcrA*	Methyl-CoM reductase
Methanotrophy	*pmoA*	Particulate methane monooxygenase,
	mmoX	Soluble methane monooxygenase
HYDROGENASES		
Ni-Fe Hydrogenase		Group 1
		Group 2a
		Group 2b
		Group 3a*
		Group 3b
		Group 3c*
		Group 3d
		Group 4*
Fe-Fe Hydrogenase	*hydA*	
AUTOTROPHIC CARBON FIXATION		
Wood-Ljungdahl pathway (methyl branch)	*fdhB*	Formate dehydrogenase
Wood-Ljungdahl pathway (carbonyl branch)	*ascB*	Acetyl-CoA synthase
Calvin-Benson-Bassham cycle	*cbbSL*	Ribulose-1,5-bisphosphate carboxylase
Reverse TCA cycle*	*aclAB*	ATP-citrate lyase
3-Hydroxypropionate bi-cycle*	*mcr*	Malonyl-CoA reductase
Dicarboxylate/4-hydroxybutyrate cycle*	*scr*	Succinyl-CoA reductase
3-Hydroxypropionate/4-hydroxybutyrate cycle*	*mcr*,	Malonyl-CoA reductase,
	scr	Succinyl-CoA reductase
CELL MOBILITY		
Flagellar assembly: Type-III secretion	*fliQ*	Flagellar biosynthesis protein
Chemotaxis; two component system	*cheAY*	Sensor kinase CheA, chemotaxis protein CheY
METAL RESISTANCE		
Cadium	*cadA*	cadmium-transporting ATPase
Cadium/Zinc/Cobalt	*czcA*	cobalt-zinc-cadmium resistance
Copper	*copA*	Copper-exporting P-typeATPase
	pcoCD	Copper resistance protein
Chromate	*chrA*	Chromate transport protein
Silver	*silP*	Silver exporting p-type atpase
Tellurite	*terB*	Tellurite resistance protein,
	tehB	Tellurite methyltransferase
16S rRNA		
SSU rRNA		SSU rRNA

* Processes or genes that were not found in any MAG

Table 5.3 – Details of the 36 phylogenetically conserved proteins used to infer the a tree including the *Serpentinomonas* strains

COG	protein name	gene
COG0504	CTP synthase (UTP - ammonia lyase) 14	pyrG
COG0533	Metal - dependent proteases with possible chaperone activity 2	tsaD
COG0016	Phenylalanyl - tRNA synthetase alpha subunit 31	pheS
COG0072	Phenylalanyl - tRNA synthetase beta subunit 30	pheT
COG0343	Queuine/archaeosine tRNA - ribosyltransferase 36	tgt
COG0164	Ribonuclease HII 38	rnhB
COG0081	Ribosomal protein L1 6	rplA
COG0244	Ribosomal protein L10 25	rplJ
COG0080	Ribosomal protein L11 10	rplK
COG0093	Ribosomal protein L14 1	rplN
COG0200	Ribosomal protein L15 32	rplO
COG0197	Ribosomal protein L16/L10E 26	rplP
COG0256	Ribosomal protein L18 18	rplR
COG0090	Ribosomal protein L2 15	rplB
COG0091	Ribosomal protein L22 7	rplV
COG0198	Ribosomal protein L24 23	rplX
COG0087	Ribosomal protein L3 28	rplC
COG0088	Ribosomal protein L4 12	rplD
COG0094	Ribosomal protein L5 19	rplE
COG0097	Ribosomal protein L6P/L9E 5	rplF
COG0051	Ribosomal protein S10 33	rpsJ
COG0048	Ribosomal protein S12 34	rpsL
COG0184	Ribosomal protein S15P/S13E 16	rpsN
COG0186	Ribosomal protein S17 13	rpsQ
COG0185	Ribosomal protein S19 8	rpsS
COG0052	Ribosomal protein S2 9	rpsB
COG0092	Ribosomal protein S3 3	rpsC
COG0098	Ribosomal protein S5 4	rpsE
COG0049	Ribosomal protein S7 20	rpsG
COG0096	Ribosomal protein S8 27	rpsH
COG0103	Ribosomal protein S9 17	rpsI
COG0541	Signal recognition particle GTPase 11	ffh
COG0532	Translation initiation factor 2 (IF - 2; GTPase) 22	infB
COG0149	Triosephosphate isomerase 35	tpiA

5.3 Results

5.3.1 Microbial biodiversity

The number of amplicon sequences per sample ranged from 108,733 to 150,106. After sequence quality filtering, denoising and chimera checking, 2,319 Amplicon Sequence Variants (ASVs) were obtained. The microbial community in PHF sites included taxa from numerous prokaryotic phyla, 16 of which had a relative abundance greater than 1% (Figure 5.1a). The major phyla identified in all PHF samples were *Proteobacteria* (19-63%), *Acetothermia* (0.04-32%), *Chloroflexi* (1-27%), *Firmicutes* (2-23%), *Cyanobacteria* (1-21%) and *Bacteroidetes* (1-12%). Within the *Proteobacteria*, the *Alphaproteobacteria* class was the most abundant, followed by the *Betaproteobacteria*, and *Gammaproteobacteria* classes. The microbial taxonomic distributions of the intertidal site BdJ (BdJ.11 and BdJ.14) were very similar, unlike those of submarine samples (ST07.11 and ST07.14) that were highly contrasted (Figure 5.1a). As the matter of fact, the *Acetothermia* phylum, overabundant in ST07.11 (central section of the chimney), was nearly absent in the ST07.14 sample (top chimney). Moreover, the ASV number in the ST07.14 sample outnumbered all the others: the 50 most abundant ASVs represented just over one third of the microbial community in ST07.14 compared to 80% in ST07.11 (Figure 5.1b). Overall, the taxonomic survey of the putative genes detected in the metagenomic sequences was consistent with the 16S rRNA amplicon results for the bacterial diversity, except for the *Acetothermia* which was less represented (Figure 5.1c). However, the metagenomic analysis revealed that 2-9% of the gene content was affiliated with Archaea while archaeal ASVs were missing in our 16S rRNA amplicons (Table 5.4). Finally, eukaryotic sequences were identified in metagenomic data (up to 6% in the metagenome ST07.14); they mainly belonged to sponges and mollusks.

5.3.2 Genome recovery from metagenomics dataset in PHF

To gain further insights into the main microbial actors of PHF, we aimed to characterize the metabolic potential of microbes inhabiting submarine sites and intertidal sites using a genome-centric metagenomic approach. Shotgun sequencing of the five libraries produced an average of 76,071,307 paired-end reads per sample. About 74.2% of paired-end reads were kept af-

Table 5.4 – Taxonomic distribution of environmental DNA at PHF. Relative abundance is expressed as percentage of genes (from metagneomic contigs) assigned to different domains of life through the Lowest Common Ancestor algorithm.

Domain \Sample	BdJ.11	BdJ.14	RdK.14	ST07.11	ST07.14
Bacteria	90.2%	89.5%	90.5%	77.8%	86.8%
Archaea	2.1%	2.1%	3.7%	8.5%	0.6%
Eukaryota	4.4%	4.8%	0.5%	5.7%	7.1%
Viruses	0.5%	0.3%	0.1%	0.1%	2.0%
Unclassified	2.7%	3.4%	5.1%	7.9%	3.5%

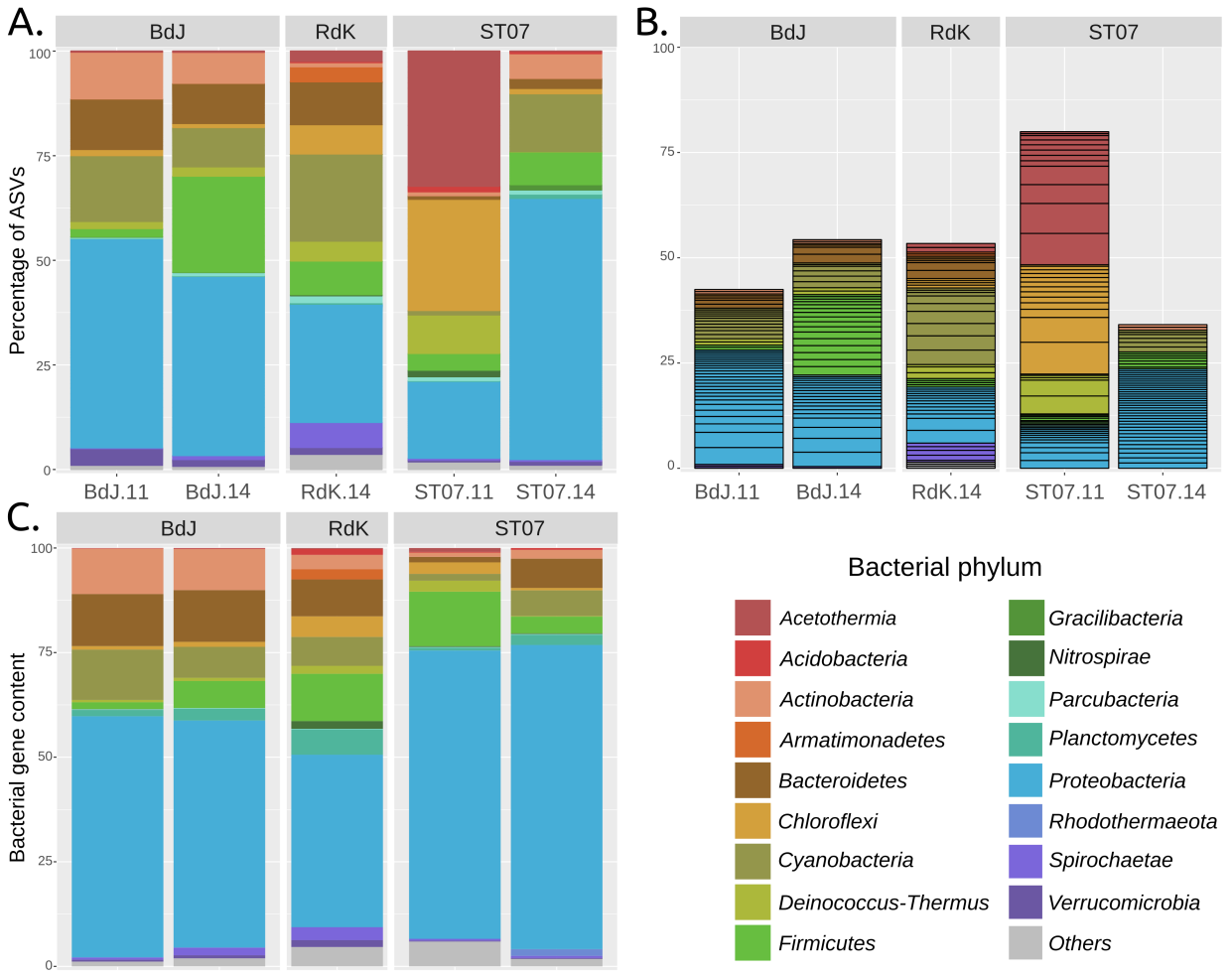


Figure 5.1 – Relative abundance of bacterial phyla in five hydrothermal chimneys of PHF sites. At the top, the barplot show the taxonomic distribution of all amplicons of the 16S rRNA gene (A) and the first 50 most abundant ASVs per sample (B). For comparison, the last barplot represents the taxonomic distribution inferred from the gene content of metagenomic data and limited to bacterial phylotype (C)

ter dereplication and quality trimming (Table 5.5). All five metagenomes were co-assembled into 3,105,427 contigs, and the optimal binning process produced 82 metagenome-assembled genomes (MAGs), encompassing 81 bacterial and 1 archaeal draft genomes. Between 57% and 86% of the trimmed reads were mapped against the co-assembled contigs, whereas only 7-19% were mapped on to the recovered MAGs. Genome statistics including the number of contigs, predicted genes, genome size, GC content, relative frequency, completeness and contamination, are listed in Table 5.6. The archaeal MAG was affiliated with the phylum *Woesearchaeota*, and the 81 bacterial MAGs were distributed across 17 phyla (Figure 5.2). Within the Bacteria domain, members of the phylum *Proteobacteria* are highly represented (26 MAGs, half of which belong to *Gammaproteobacteria*). The so-called abundant MAGs, defined after their percentage of reads mapped per sample and length of the MAGs, belonged to the *Firmicutes*, *Deinococcus-Thermus*, *Bacteroidetes*, *Chloroflexi* and *Proteobacteria* (Table 5.6). Only a few MAGs (6) were closely related to candidate phyla such as *Balneolaeota*, *Omnitrophica* or *Parcubacteria*.

Table 5.5 – Quality control process and assembly statistics for the five metagenomes of PHF

		BdJ-11_ATCACG_L003		D-BdJ-14_ACTGAT_L003		RdK-14_GAGTGG_L003		ST07-11_TGACCA_L003		ST07-14_CAGATC_L003	
		R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
QC process	# of raw reads	75,591,008	75,591,008	86,096,099	86,096,099	75,683,917	75,683,917	74,761,320	74,761,320	68,224,194	68,224,194
	# of reads after cd-hit-dup	63,716,777	63,716,777	69,517,332	69,517,332	62,904,809	62,904,809	55,668,624	55,668,624	56,551,218	56,551,218
	# of paired reads after trimmomatic	59,516,284	59,516,284	65,373,927	65,373,927	58,940,934	58,940,934	45,248,682	45,248,682	53,249,574	53,249,574
	# of single reads after trimmomatic	3,521,990	152,149	3,720,146	176,852	3,290,411	198,528	2,478,482	129,918	2,824,975	120,124
Individual assembly	# of contigs	804,907		807,393		537,987		371,155		754,791	
	# of reads mapped to contigs	87,152,381		92,981,782		105,714,847		78,213,865		64,694,123	
	% of trimmed reads recruited in contigs	73.22%		71.12%		89.68%		86.43%		60.75%	
	# of predicted genes in contigs	1,311,449		1,175,606		1,089,770		507,564		1,297,892	
	# of scaffolds	756,795		762,442		513,100		366,936		705,801	
	# of reads mapped to scaffolds	87,140,966		92,987,529		105,699,851		78,203,010		64,687,600	
	% of trimmed reads recruited in scaffolds	73.21%		71.12%		89.67%		86.41%		60.74%	
Co-assembly	# of contigs	3,105,427									
	# of contigs (>1000bp)	737,141									
	# of contigs (>2500bp)	189,948									
	# of reads mapped to contigs	84,908,439		89,300,188		101,321,788		72,603,481		60,497,555	
	% of trimmed reads recruited in contigs	71,33%		68,30%		85,95%		80,23%		56,81%	
	# of reads mapped to the 82 MAGs	10,626,044		15,940,009		19,275,484		4,885,186		7,559,633	
	% of mapped reads recruited in the 82 MAGs	12,51%		17,85%		19,02%		6,73%		12,50%	

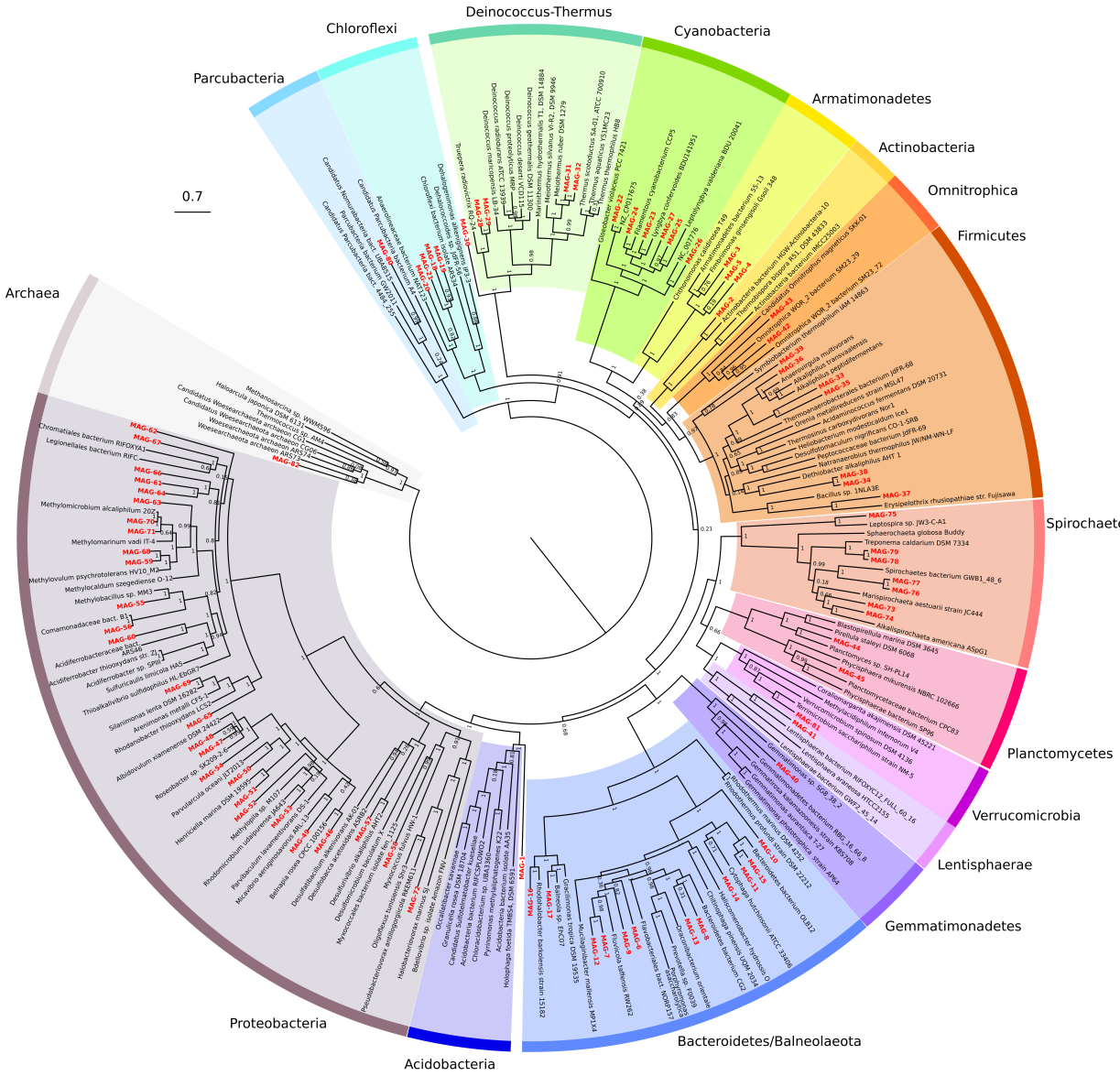


Figure 5.2 – Diversity of metagenome assembled-genomes (MAGs), reconstructed from the 5 PHF samples. Maximum-likelihood phylogenomic tree of all MAGs >70% complete and select genomes, MAGs and single cell amplified genomes from NCBI. The horizontal bar represents the number of substitutions per site. Each red-colored leaf represents one MAG identified in this study. The phyla identified in PHF samples are indicated on the outer ring. The percentage next to the nodes corresponds to the local support values from 1000 replicates.

5.3.3 Functional resolution of MAGs

Of the 82 individual MAGs, 39 reached 90% of genome completeness, and only 5 MAGs (i.e., MAG-1, MAG-20, MAG-40, MAG-52 and MAG-54) meet the stringent rRNA requirements for high-quality draft MAGs defined by Bowers et al. (Bowers et al., 2017). The presence of functional genes in these 39 near-complete MAGs was reported in Figure 5.3 as well as the percentage of mapped reads recruited from the five PHF metagenomes. Surprisingly, the MAGs did not have an abundance profile depending on the nature of the PHF sites (i.e., submarine or intertidal).

Table 5.6 – Overview of all MAGs >70% complete and with <5% contamination.

ID	Completeness	Contamination	CheckM annotation	Phylum or Class extrapolated from PhyloPhlAn	contigs (>2500bp)	Number of predicted genes	Length (bp)	GC content
MAG-1	94.87%	2.56%	k__Bacteria	Acidobacteria	186	2985	3600172	68.5%
MAG-2	77.89%	2.78%	p__Actinobacteria	Actinobacteria	117	1319	1322008	65.7%
MAG-3	75.09%	2.31%	k__Bacteria	Armatimonadetes	220	2192	2236426	56.2%
MAG-4	79.63%	0.93%	k__Bacteria	Armatimonadetes	69	1913	2136053	62.7%
MAG-5	86.88%	4.27%	k__Bacteria	Armatimonadetes	314	2070	2067325	60.2%
MAG-6	84.06%	2.46%	p__Bacteroidetes	Bacteroidetes	251	1958	2264027	41.2%
MAG-7	99.46%	0.72%	k__Bacteria	Bacteroidetes	171	2818	3367353	46.4%
MAG-8	97.94%	0.48%	p__Bacteroidetes	Bacteroidetes	39	2510	3030981	42.4%
MAG-9	97.29%	0.39%	s__algicola	Bacteroidetes	89	2624	2899587	37.0%
MAG-10	95.08%	0.18%	k__Bacteria	Bacteroidetes	111	2346	2961519	56.4%
MAG-11	77.09%	0.98%	k__Bacteria	Bacteroidetes	348	2223	2509711	52.3%
MAG-12	90.86%	2.74%	k__Bacteria	Bacteroidetes	429	2844	3334088	47.9%
MAG-13	86.02%	3.49%	k__Bacteria	Bacteroidetes	332	1940	2337113	44.5%
MAG-14	96.06%	1.82%	o__Cytophagales	Bacteroidetes	432	4007	5083732	39.5%
MAG-15	85.97%	1.06%	o__Cytophagales	Bacteroidetes	259	2210	2487446	52.8%
MAG-16	92.04%	2.19%	k__Bacteria	Balneolaeota	274	2761	3113379	47.3%
MAG-17	98.49%	1.50%	k__Bacteria	Balneolaeota	99	3093	3515685	42.9%
MAG-18	77.27%	0%	k__Bacteria	Chloroflexi	70	2531	3002856	52.8%
MAG-19	87.27%	1.01%	k__Bacteria	Chloroflexi	446	3084	3500207	48.5%
MAG-20	92.73%	0.91%	k__Bacteria	Chloroflexi	100	2495	2726778	50.5%
MAG-21	93.64%	0%	k__Bacteria	Chloroflexi	48	3170	3732735	61.5%
MAG-22	87.89%	1.80%	k__Bacteria	Cyanobacteria	322	2197	2219948	53.0%
MAG-23	91.53%	0.68%	p__Cyanobacteria	Cyanobacteria	246	3942	4298084	55.0%
MAG-24	97.52%	1.18%	p__Cyanobacteria	Cyanobacteria	174	4242	4489261	60.0%
MAG-25	98.07%	0.41%	p__Cyanobacteria	Cyanobacteria	193	3509	4229265	60.2%
MAG-26	95.61%	0%	k__Bacteria	Cyanobacteria	94	2504	2538815	53.3%
MAG-27	99.24%	1.31%	p__Cyanobacteria	Cyanobacteria	152	3591	4013355	44.6%
MAG-28	86.02%	1.69%	k__Bacteria	Deinococcus-Thermus	44	2325	2427901	59.9%
MAG-29	80.51%	0.85%	k__Bacteria	Deinococcus-Thermus	227	1589	1613582	62.7%
MAG-30	81.57%	3.60%	k__Bacteria	Deinococcus-Thermus	181	1810	1928482	70.5%
MAG-31	95.85%	0.85%	k__Bacteria	Deinococcus-Thermus	81	3242	3485611	59.5%
MAG-32	91.58%	0.43%	k__Bacteria	Deinococcus-Thermus	91	2532	2692402	59.2%
MAG-33	83.69%	2.13%	o__Clostridiales	Firmicutes	180	2224	2232016	41.1%
MAG-34	82.42%	1.86%	p__Firmicutes	Firmicutes	155	1627	1813226	48.9%
MAG-35	94.33%	1.11%	o__Clostridiales	Firmicutes	170	2785	2894720	42.7%
MAG-36	78.44%	0.52%	o__Clostridiales	Firmicutes	261	1390	1482154	36.7%
MAG-37	97.17%	2.36%	k__Bacteria	Firmicutes	106	1504	1489890	37.4%
MAG-38	87.08%	1.62%	p__Firmicutes	Firmicutes	127	1653	1870795	50.6%
MAG-39	88.94%	3.96%	p__Firmicutes	Firmicutes	71	1828	1950539	69.7%
MAG-40	97.80%	4.4%	k__Bacteria	Gemmatimonadetes	112	2499	3062271	64.4%
MAG-41	81.08%	1.35%	k__Bacteria	Lentisphaerae	123	1875	2284597	59.8%
MAG-42	90.32%	2.15%	k__Bacteria	Omnitrophica	102	1083	1179698	40.9%
MAG-43	72.83%	0%	k__Bacteria	Omnitrophica	137	1148	1192489	41.5%
MAG-44	97.70%	1.38%	k__Bacteria	Planctomycetes	225	3860	4810023	59.0%
MAG-45	87.03%	0%	k__Bacteria	Planctomycetes	38	2212	2702493	67.1%
MAG-46	89.42%	0.3%	o__Rhodospirillales	Alphaproteobacteria	312	2198	2298908	69.8%
MAG-47	93.85%	0.87%	f__Rhodobacteraceae	Alphaproteobacteria	108	2684	2661550	66.4%
MAG-48	86.62%	1.21%	f__Rhodobacteraceae	Alphaproteobacteria	236	2404	2443917	69.2%
MAG-49	90.87%	1.32%	c__Alphaproteobacteria	Alphaproteobacteria	215	1819	1956525	45.2%
MAG-50	78.94%	1.13%	c__Alphaproteobacteria	Alphaproteobacteria	316	2521	2558024	62.5%
MAG-51	83.24%	1.66%	c__Alphaproteobacteria	Alphaproteobacteria	423	2646	2553804	57.6%
MAG-52	91.68%	0%	c__Alphaproteobacteria	Alphaproteobacteria	12	1392	1482001	47.6%
MAG-53	73.14%	1.96%	o__Rhizobiales	Alphaproteobacteria	180	1920	1987766	63.8%
MAG-54	91.78%	0%	c__Alphaproteobacteria	Alphaproteobacteria	125	2983	3288092	61.3%
MAG-55	71.40%	0.85%	c__Betaproteobacteria	Betaproteobacteria	142	1380	1417946	43.1%
MAG-56	96.17%	0.70%	o__Burkholderiales	Betaproteobacteria	93	2267	2458887	67.3%
MAG-57	84.51%	0%	c__Deltaproteobacteria	Deltaproteobacteria	184	1327	1496224	53.3%
MAG-58	77.46%	1.94%	c__Deltaproteobacteria	Deltaproteobacteria	132	3553	3978769	63.7%
MAG-59	98.10%	0.17%	c__Gammaproteobacteria	Gammaproteobacteria	120	2920	3034721	49.8%
MAG-60	75.00%	0.61%	p__Proteobacteria	Gammaproteobacteria	129	2417	2555657	56.2%
MAG-61	86.69%	1.65%	c__Gammaproteobacteria	Gammaproteobacteria	196	2226	2452549	58.2%
MAG-62	98.08%	0%	c__Gammaproteobacteria	Gammaproteobacteria	11	1711	1824444	41.6%
MAG-63	91.13%	1.94%	c__Gammaproteobacteria	Gammaproteobacteria	554	3750	4466698	46.9%
MAG-64	86.37%	1.06%	c__Gammaproteobacteria	Gammaproteobacteria	231	2371	2431537	50.8%
MAG-65	97.10%	1.45%	c__Gammaproteobacteria	Gammaproteobacteria	16	2263	2557214	51.5%
MAG-66	95.07%	0%	c__Gammaproteobacteria	Gammaproteobacteria	15	2361	2682451	61.8%
MAG-67	87.22%	0.19%	c__Gammaproteobacteria	Gammaproteobacteria	75	2247	2649310	57.1%
MAG-68	72.11%	1.37%	c__Gammaproteobacteria	Gammaproteobacteria	355	1671	1689899	54.9%
MAG-69	78.11%	0.72%	f__Xanthomonadaceae	Gammaproteobacteria	235	1586	1685813	68.1%
MAG-70	91.25%	1.72%	c__Gammaproteobacteria	Gammaproteobacteria	306	3539	4036429	48.1%
MAG-71	96.53%	1.49%	c__Gammaproteobacteria	Gammaproteobacteria	336	2794	3058252	50.0%
MAG-72	87.27%	0%	k__Bacteria	Oligoflexia	29	1669	2110185	41.4%
MAG-73	92.91%	2.30%	k__Bacteria	Spirochaetes	287	1743	1786783	61.1%
MAG-74	87.47%	2.80%	c__Spirochaetia	Spirochaetes	327	3120	3111725	61.2%
MAG-75	94.12%	0%	k__Bacteria	Spirochaetes	267	3097	3395457	39.4%
MAG-76	82.72%	1.23%	c__Spirochaetia	Spirochaetes	276	1837	1979605	53.9%
MAG-77	87.75%	1.09%	c__Spirochaetia	Spirochaetes	314	1764	1849039	56.0%
MAG-78	84.27%	1.14%	k__Bacteria	Spirochaetes	226	1842	1984617	56.4%
MAG-79	91.09%	0%	k__Bacteria	Spirochaetes	140	2198	2458237	56.5%
MAG-80	71.15%	1.12%	k__Bacteria	Unclassified	108	1013	993987	50.1%
MAG-81	98.65%	0.68%	k__Bacteria	Verrucomicrobia	93	2758	2926733	60.9%
MAG-82	75.70%	0.93%	k__Archaea	Woesearchaeota	96	924	925581	36.6%

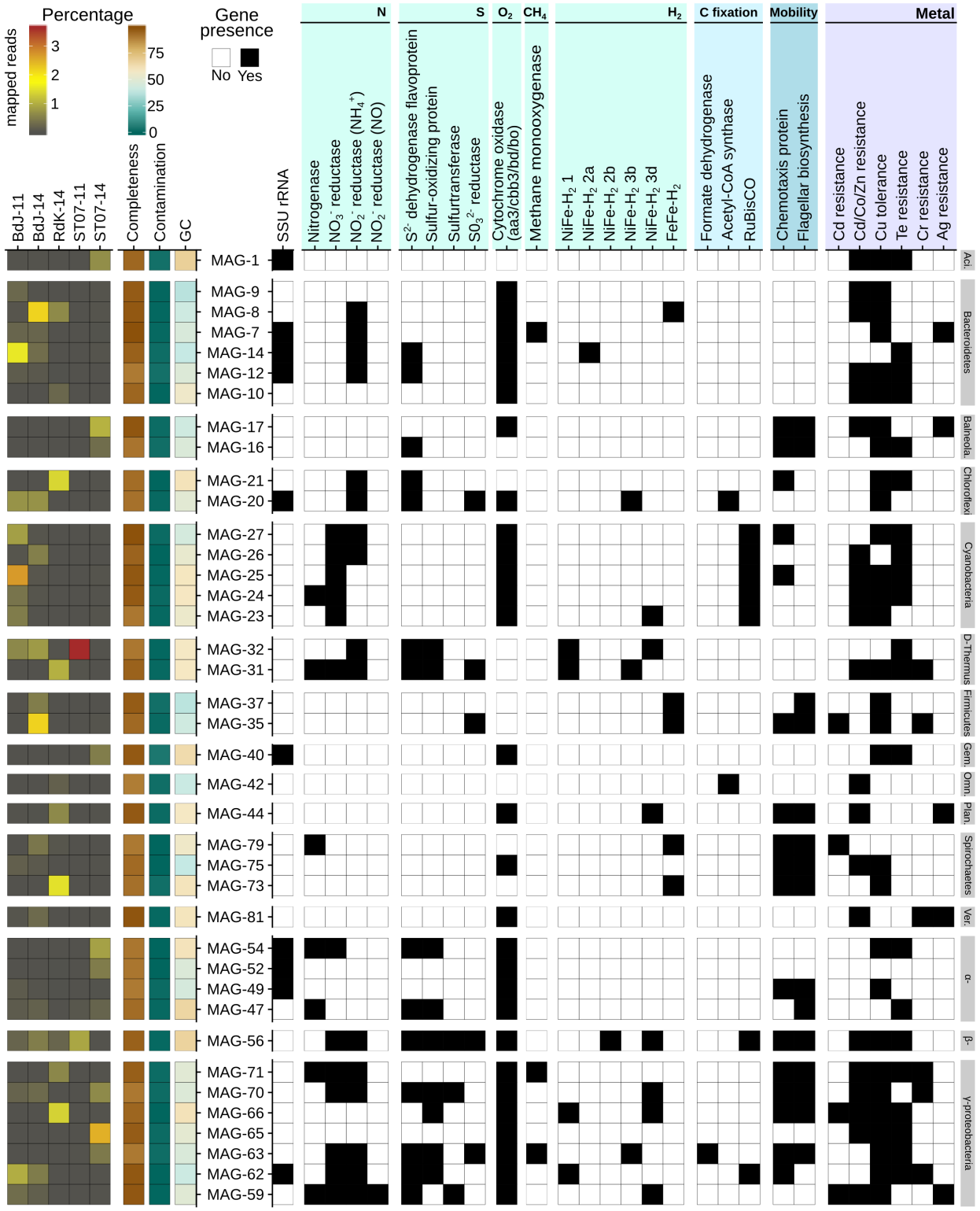


Figure 5.3 – Abundance, genome quality, GC content, subunit 16S rRNA and functional gene of the 39 MAGs . Heatmaps on the left indicate the percentage of mapped reads recruited from the five samples. Black boxes denote the presence of 16S rRNA and genes encoding N, S, O_2, CH_4 and H_2 metabolisms (cyan), carbon fixation (light blue), chemotaxis (blue) and metal resistance (purple). The taxonomic affiliation of each MAG is indicated on the right at the phylum level (or class for the Proteobacteria members). Taxa are abbreviated as follows: Aci., *Acidobacteria*; Balneola., *Balneolaeota*; D-Thermus, *Deinococcus-Thermus*; Gem, *Gemmatimonadetes*; Omn., *Omnitrophica*; Plancto., *Planctomycetes*; Ver., *Verrucomicrobi*, α-, *Alphaproteobaceria*; β-, *Betaproteobacteria*.

Energy metabolisms in MAGs The aerobic respiration potential was detected in 29 MAGs containing low-oxygen sensitivity (aa3 or bo-type) cytochromes or high-oxygen sensitivity (microaerobic conditions; cbb3 or bd-type) cytochromes. The capacity to utilize alternative electron acceptors was assessed in the gene content of the MAGs. The result demonstrated the presence of both nitrate and nitrite reductases (associated genes were listed in the Table 5.2) in MAGs affiliated with *Cyanobacteria*, *Proteobacteria* and *Deinococcus-Thermus*. Additional nitrite reductases were reported in *Bacteroidetes* and *Chloroflexi* MAGs. Regarding the CO_2 respiration, the potential for acetogenesis was suggested by the presence of acetogenesis carbon monoxide dehydrogenase/acetyl-CoA synthase complex (Codh/Acs) in the MAG-42 but remained questionable since the Wood-Ljungdah pathway was not complete. The potential for methanogenesis (assessed with the *mcrA* gene, required for the last step of methanogenesis) was not identified in the recovered MAGs, although tens of sequences of this gene were identified in the unbinned contigs of PHF metagenomes. Most of these *mcrA* genes were affiliated with Methanosarcinales members. This result showed that the reconstructed MAGs and their metabolic potential were not representative of the entire microbial diversity of PHF sites. Methane monooxygenase (*pmoA*/ *mmoX*) was identified in two *Gammaproteobacteria* and one *Bacteroidetes* MAGs, indicated a relatively scarce potential for methanotrophic metabolism despite high CH_4 concertation in fluids. Although no sulfide quinone oxidoreductase (*sqr*) was detected, the potential for sulfide oxidation was suggested in a few MAGs (e.g., *Proteobacteria* and *Deinococcus-Thermus* members) by the presence of the genes encoding the flavocytochrome c/sulfide dehydrogenase. The detection of sulfite reductase genes among the MAGs reconstructed in this study that do not contain sulfate adenylyltransferase and adenosine phosphosulfate reductase suggested the presence of sulfite reducing microorganisms. Finally, the hydrogenases were found in most MAGs reconstructed. Five functional groups of [NiFe]-hydrogenases were identified in the MAGs of PHF, but most of them belonged to the Group 3d. On the other hand, the [FeFe]-hydrogenases were mainly detected in MAGs affiliated with *Firmicutes*, and but also in some *Spirochaetes* and *Bacteroidetes* genomes.

Mode of carbon fixation in MAGs Carbon fixation capabilities was examined in each of the 39 near complete MAGs. All MAGs affiliated with the phylum Cyanobacteria and two MAGs members of the *Proteobacteria* contained the genes encoding for RuBisCO, the canonical enzyme of the Calvin-Benson-Bassham cycle (Figure 5.3). Regarding the Wood-Ljungdah pathway, 3 MAGs contained either the Acetyl-CoA synthase required in the carbonyl branch or the formate dehydrogenase, key enzyme of the methyl branch, but further annotation showed that none included the complete pathway. The key enzymes of the other carbon fixation pathways were not detected in the MAGs, which may suggest a heterotrophic lifestyle for most of microorganisms recovered in PHF.

Microbial mobility and adaptive mechanisms Chemotaxis and mobility appeared to be essential for most of the MAGs identified here. Surprisingly, most MAGs included genes for metal

resistance and tolerance. The gene *czcA*, encoding for Cadium/Cobalt/Zn resistance was included in half of the MAGs. The genes for copper tolerance (*copA* and *pcoCD*) and tellurite resistance (*terB* or *tehB*) were also very abundant in the reconstructed MAGs.

5.3.4 MAGs affiliated with candidate divisions

The two smallest MAGs, reconstructed using MetaBAT, had a total length of less than 100,000bp. The MAG-80 was associated with *Parcubacteria* genomes (Figure 5.2). This draft genome was partial with only 71% of completeness and included 1013 predicted genes (Table 5.6). The gene-content of MAG-80 contained V/A-type ATP synthase. The MAG-82 clustered with four MAGs affiliated with the *Woesearchaeota* phylum in the phylogenetic tree (Figure 5.2). The draft genome of the MAG-82 was relatively small (including 924 genes) even though it was only 75% complete. Based on PROKKA annotation, the MAG-82 had many metabolic deficiencies since the genome lacked a complete Krebs cycle, a complete electron transport chain, or a continuous glycolysis. This result suggested an anaerobic or fermentation-based lifestyle.

5.3.5 Genomic comparison of *Serpentinomonas* phylotypes

After refinement (Figure 5.4), the MAG-56 gathered 0.02-0.92% of reads mapped against the co-assembled contigs and was affiliated with the order *Burkholderiales* (Table 5.6). Based on the functional annotation, the MAG-56 can used oxygen, nitrate or sulfite as electron acceptors (Figure 5.3). The genome of the MAG-56 also comprised two [NiFe]-hydrogenases, that clustered into the functional Group 2b and 3d (Vignais, 2001). The hydrogenases clustered in Group 2b are known to function as H_2 sensors, while being insensitive to oxygen. The bidirectional hydrogenases clustered in Group 3d and located in the cytoplasm, can reduce NAD+ with H_2. The NADH thus produced could be used as energy source. The MAG-56 comprised the genes *cbbS* and *cbbL* that code for the RuBisCO enzyme, suggesting a possible autotrophic growth via the Calvin-Benson-Bassham cycle. Bacterial mobility was assumed due to the presence of genes encoding a flagellar in the MAG-56. This genome is closely related to the three strains isolated from serpentinizing springs at The Cedars, proposed as a new genus *Serpentinomonas* (Suzuki et al., 2014). When comparing the MAG-56 against public genomes of GenBank, best hits were reported for *Serpentinomonas mccroryi* strain B1. Further investigations showed that MAG-65 had 92.6 % average nucleotide identity (ANI) with the strain B1, decreasing down to 85% with A1 and H1 strains. Unfortunately, the MAG-56 did not include the 16S rRNA gene, which prevented comparisons with cultivated microorganisms of the genus *Serpentinomonas*. A tree of concatenated marker genes was used to inferred phylogenetic relationships between the MAG-56, the 3 strains from The Cedars and references genomes of closest genera such as *Hydrogenophaga*, *Malikia* and *Macromonas* (Figure 5.5). The MAG-56 and the three strains affiliated to the genus *Serpentinomonas* formed a cluster separated from closest genera. Based on a genomic comparison, these four strains shared the same major metabolisms (Figure 5.6). Overall, same pathways were identified for the carbon fixation, lipids and carbohydrate metabolisms, except for

the Entner-Doudoroff pathway that was missing in the MAG-56. However, the MAG-56 could most likely degrade aromatic compounds such as benzene or phenol via the phenol hydroxylase and the meta-cleavage pathway of catechol.

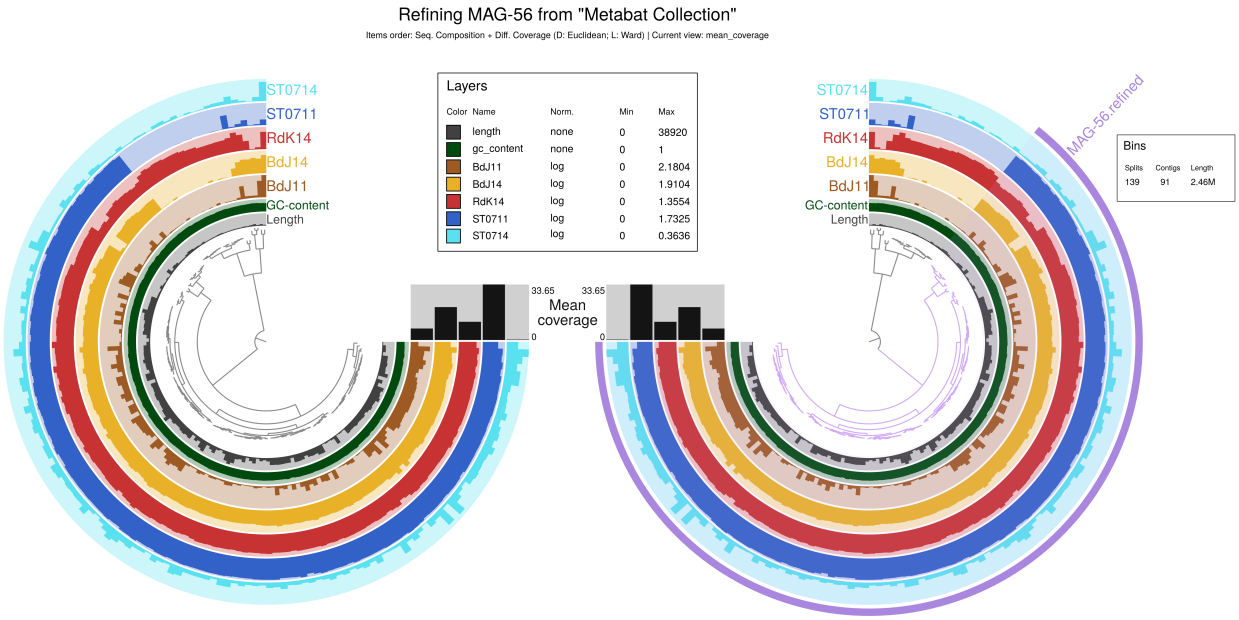


Figure 5.4 – Selection of contigs according to their mean coverage during the refinement of the MAG-56 with Anvi'o.The central tree corresponds to the clustering of the split contigs, using Euclidean distance and Ward's linkage. The most internal circle corresponds to the length of the split contigs

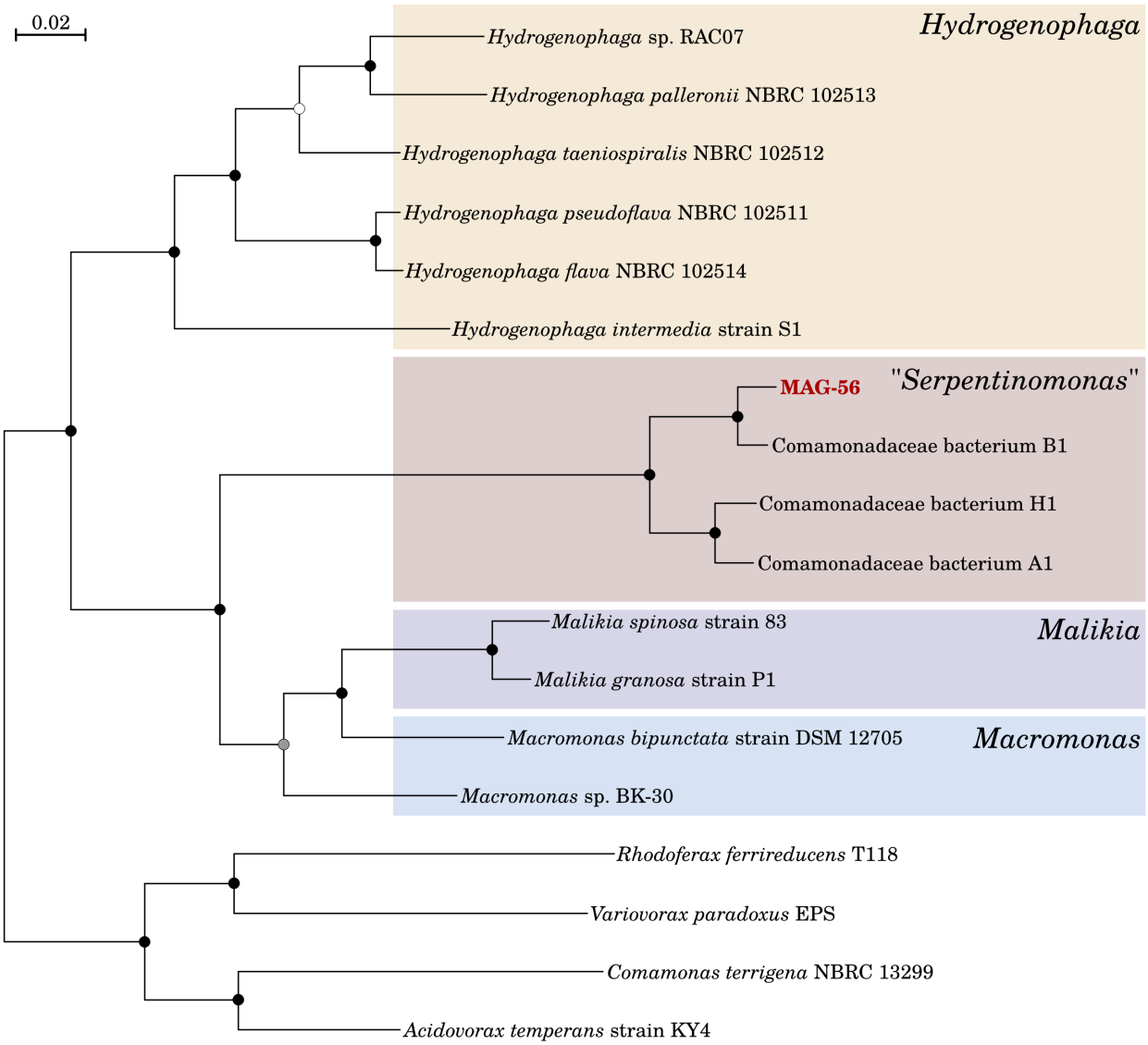


Figure 5.5 – Phylogenetic tree of *Comamonadaceae* genomes containing strains affiliated with the genus *Serpentinomonas*, isolated from serpentinizing springs at The Cedars. The MAG-56, identified in PHF chimneys, is grouped with these three strains. The maximum likelihood tree is based on the concatenated alignments of 36 phylogenetically conserved proteins. Black circles in the tree represent nodes with >90% bootstrap supports, gray circles indicate nodes with >70% bootstrap support and white circles indicate nodes with >50% bootstrap support. Scale bar indicates substitutions per site.

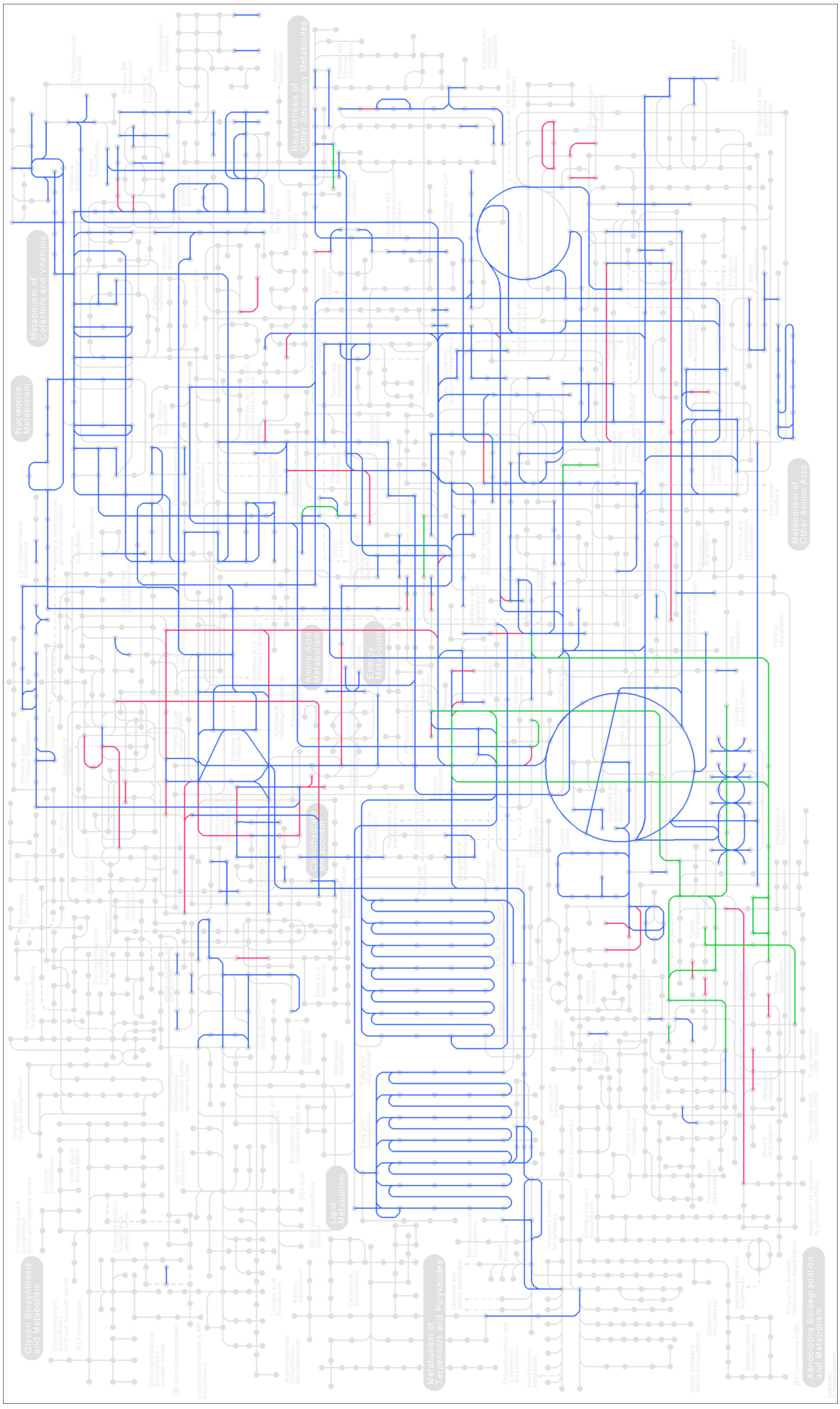


Figure 5.6 – Comparison of metabolic pathways reconstructed from the four complete genomes of *Serpentinomonas*: the MAG-56 and the three strains isolated from the Cedars. Green lines correspond to pathways only detected in the MAG-56, while red lines indicate pathways absent in the MAG-56 but identified in at least one of the strains A1, B1 or H1. Blue lines indicate pathways that are shared by the MAG-56 and at least one of the three other strains.

5.4 Discussion

5.4.1 A highly diversified microbial community

The analysis of the 16S amplicons demonstrated that there were no overabundant microbes in the samples collected at PHF, the most dominant ASV reaching 7% of the microbial community in ST07.11. The reconstruction of genomes from such samples will therefore be challenging because the sequencing depth of each microorganism will be lower. This prediction was confirmed by the low percentage of mapped reads recruited to reconstruct the 82 MAGs. This result demonstrated that a large part of the genomic information was not included in the reconstructed MAGs due to the high diversity of PHF microbial communities. An intriguing result was that we were unable to reconstruct MAGs affiliated with *Acetothermia* despite their abundance estimated by 16S rRNA gene in ST07.11. This result may be due to a primer bias that overestimates the *Acetothermia* organisms in the 16S rRNA gene-based survey or a difficulty in producing high quality contigs in metagenomic analysis.

5.4.2 MAGs associated with serpentinizing ecosystems

Most of the reconstructed MAGs had the genetic potential to use oxygen as electron acceptor. The proportion of anaerobic or fermentative metabolism was rather low, probably because most of the reconstructed MAGs were not necessarily associated with anoxic fluids. In addition, with the exception of the Calvin-Benson-Bassham cycle, no complete carbon fixation pathway could be identified. The Wood-Ljungdahl pathway was not represented, although the acetogenesis was assumed to be potential metabolism of serpentinizing ecosystems (Kohl et al., 2016, Pisapia et al., 2017, Suzuki et al., 2017). However, the diversity of hydrogen metabolism could be highlighted in our study. [FeFe]-hydrogenases were mainly identified in the MAGs belonging to the *Firmicutes*. This observation was consistent with a previous metagenomic study, which reported that the diversity of the gene hydA was dominated by the sequences associated with *Firmicutes* in PHF (Mei et al., 2016a). Moreover, a MAG affiliated with *Bacteroidetes* also had a [FeFe]-hydrogenase and could be considered as a potential actor in the fermentation process as suggested in the serpentinizing site of Voltri Massif (Brazelton et al., 2016). The microorganisms affiliated with the *Parcubacteria* were reported as incapable of independent growth in the terrestrial serpentinizing site of The Cedars (Suzuki et al., 2017), since they did not contain any known ATP synthase genes. Another lifestyle could be considered for the MAG-80 (reconstructed from PHF metagenomes), since this genome contained genes required for independent life such as a V/A-type ATP-synthase. However, our study highlighted numerous metabolic deficiencies in the genome of another candidate phylum, the *Woesearchaeota*. The absence of an electron transport chain, the incomplete TCA cycle and discontinuous glycolysis detected in MAG-82 have already been mentioned by Liu and collaborators in other genomes of *Woesearchaeota* (Liu et al., 2018). In this study, the authors compared 19 publicly available genomes to highlight metabolic pattern shared within Woesearchaeotal lineages. Based on their mod-

eling, the MAG-82 would be specific of anoxic environment because it contained ferredoxin-dependent pathways. By listing all the metabolic deficiencies, the authors suggested symbiotic lifestyles for *Woesearchaeota* phylotypes. This dependency on other organisms is undoubtedly the main challenge in achieving the enrichment or pure cultivation of these lineages.

5.4.3 Genus *Serpentinomonas* as marker of serpentinizing sites?

Serpentinomonas strains were dominant microorganisms in many terrestrial serpentinizing systems, including the Tablelands, Outokumpu borehole, Voltri Massif, Cabeço de Vide Aquifer (Quéméneur et al., 2015, Suzuki et al., 2014, Tiago and Veríssimo, 2013), and were recently detected in lower abundance within the hydrothermal fluids and chimneys of PHF (Mei et al., 2016b). Suzuki and colleagues demonstrated that these species were not limited to serpentinizing environments (Suzuki et al., 2014). However, the systematic presence of *Serpentinomonas* strains indicated that they must be an essential link in the microbial colonization of serpentinizing environments. The isolation of three *Serpentinomonas* strains demonstrated that they were highly adapted to the condition of geochemical settings occurring in serpentinizing sites (high pH, low dissolved inorganic carbon). This study presented a draft of a new species of *Serpentinomonas*. As its sister lineages, the MAG-56 possessed several [NiFe]-hydrogenases, including the Group 3d hydrogenase and genes encoding the RuBisCO enzyme. We can assume that this new species may grow autotrophically using $CaCO_3$, in the same way as the species isolated at The Cedars (Suzuki et al., 2014). There, cyclohexane, a source of aromatic organic carbon, was assumed to contribute to support the life in the springs. Likewise, the genetic potential of MAG-56 to degrade benzene or phenol must be investigated. It would be interesting to quantify aromatic compounds in PHF and if they may be a used as carbon source.

Chapter 6

Discussion and perspectives

“ Comme l’art, la science est une tentative toujours renouvelée d’appréhender, de manière de plus en plus riche, la réalité de l’univers au-delà de ses reflets. ”

Jean-Claude Ameisen

Sommaire

6.1 Methodological issues

Several methodological issues (from sampling to DNA sequence processing) emerged during this doctoral research project and fed into this discussion section. The assessment of our methods highlights the potential biases in our analyses and helps us find alternatives for future challenging studies.

6.1.1 Molecular microbiology

Environmental sampling Although PHF is more easily accessible than deep hydrothermal systems due to its shallow depth, the sampling stage remains challenging. At PHF, two sampling sites are located in intertidal zones, but submarine sites can reach a depth of 50 meters and must be sampled by scuba divers who are rarely biologists. Moreover, the submarine sites of PHF are classified as nature reserves, and authorizations must be granted for sampling. Additionally, the Prony Bay (at the southern end of New Caledonia) is a four-hour drive from the nearest laboratory. During previous field campaigns, the samples were stored in ice during the journey and frozen at -80°C upon receipt at the laboratory. Two types of material can be collected at PHF: alkaline fluids and hydrothermal chimneys. The best way to avoid contamination of hydrothermal fluids by seawater is to collect fluids from intertidal sites, at low tide. Regarding the fluids from submarine sites, the contamination by the surrounding seawater is difficult to avoid and causes biases in physicochemical measurements. For the hydrothermal chimney samples, one strategy to minimize microbial contamination is to retain only the central part of the chimneys for the microbial analysis. However, some chimneys found in two intertidal sites ("Bain des Japonais" and "Rivière des Kaoris") were very tiny and this strategy could not always be applied. Such samples therefore include microbial communities living inside hydrothermal vents and probably also those that colonize the outer walls. Moreover, as the chimney structures are porous, it is difficult to consider setting limits to an environment subjected to gradients.

DNA extraction The crushing of chimneys followed by bead beating before DNA extraction was demonstrated to be effective in previous microbial studies at PHF. However, the deep sequencing of metagenomes described in Chapter 5 required significant amount of DNA for each sample. The technical difficulty of this extraction process was unexpected. Indeed, the quantities previously obtained on chimney samples were largely sufficient to carry out the amplification and sequencing of 16S rRNA genes or even the shallow sequencing of metagenomes. In order to optimize extraction yields, two protocols for DNA extraction were tested: the FastDNA soil kit (MP Biomedicals) and a phenol-chloroform extraction method (see appendices). DNA yields were measured on a Qubit Fluorometer (Invitrogen), since other measurement devices (e.g., BioSpec-nano spectrophotometer by Shimadzu-Biotech) tended to overestimate DNA concentrations. The best yields were obtained with the commercial kit and, especially, with the more recent samples. In contrast, almost all attempts to extract DNA from samples collected

five years ago have failed, despite their storage at -80°C. These limitations highlight the difficulty of conducting ecosystem studies over long time-series using the same methodology.

Primers and PCR amplification PCR amplification and sequencing of 16S rRNA genes are fast and inexpensive methods for characterizing the diversity of a microbial community. Different PCR primers sets were used during this thesis. In Chapter 3, bacterial and archaeal 16S rRNA genes were amplified using distinct primers, while the PCR-amplifications in Chapter 5 relied on universal primers. The advantage of universal primers is the simultaneous amplification of bacterial and archaeal 16S rRNA genes, with reduced time and cost (Takahashi et al., 2014). Prior to the use of these universal probes, we checked their ability to target specific microorganisms expected in PHF sites, such as LCMS, using RDP Probe Match. Moreover, we knew that at PHF, the *Archaea* accounted for ~10% of the total prokaryotic cells according to an estimate obtained from samples of the hydrothermal chimney ST09 (Postec et al., 2015). However, in Chapter 5, the *Archaea* were totally absent from amplified sequences. Our universal primers failed to capture the archaeal diversity identified in Chapter 3. These primers cannot be reasonably described as universal since they failed to amplify some of the expected phylotypes. Moreover, it is impossible to estimate their ability to target the unknown fraction of microbes in these environments. To limit amplification biases, some authors proposed to use primers optimized for each environment (Walters et al., 2011), while others suggested to combine sequencing results from different PCR-amplified regions of 16S rRNA genes (Fuks et al., 2018). The only way to completely overcome amplification bias would be to use whole metagenome sequencing. However, this more expensive methodology that requires higher concentrations of DNA is not always applicable.

6.1.2 Taxonomic and functional annotations

Databases and taxonomic annotations During this doctoral research project, the taxonomic composition of microbial communities was determined using the 16S rRNA genes (metabarcoding in chapters 3 and 5). The resolution power of 16S rRNA gene was typically limited to genus taxonomic level, because of the shortness of the amplified 16S regions. Moreover, the hierarchical taxonomy of 16S rRNA genes is greatly impacted by the selection of the database. Currently, the taxonomic affiliation of sequences is primarily based on four databases: SILVA, RDP, Greengenes, or NCBI. They have different size (3,000 versus 1,500,000), resolution (genus or species) and are not updated concurrently (Balvočiute and Huson, 2017). For example, the last update of Greengenes was made in 2013. The Silva database is regularly updated, but it produces discrepancies between versions. Therefore, caution should be exercised when comparing studies that used different databases or version. Furthermore, significant changes are to be expected in the current databases due to new classifiaction systems. The sequencing of 16S rRNA genes is the main approach for quickly identifying taxa in environmental ecology. However, the reliability and legitimacy of 16S rRNA genes in taxonomic classification are questioned

because of the chimeric sequences inserted in the databases (spurious OTUs). To overcome this problem and better conciliate taxonomy with phylogeny (by removing polyphyletic groupings), Parks and collaborators have suggested a revision of the bacterial taxonomy of the tree of life (Parks et al., 2018). The proposed phylogeny no longer relies on 16S rRNA genes but is inferred from a concatenation of 120 ubiquitous single-copy proteins. This phylogeny includes metagenome-assembled genomes (MAGs) and single-cell genomes (SAGs) as representative of uncultivated microorganisms. The authors proposed a standardized taxonomy that normalized the stratification of taxonomic ranks (such as phylum, class, order) based on a measure of relative evolutionary divergence. To create an accurate genome-based taxonomy, efforts must focus on the reconstruction of high-quality MAGs and SAGs, thus preventing the introduction of chimeras into this new classification. The formal reclassification of prokaryotic organisms, although necessary, will probably be a long transition process. Indeed, a few notable taxonomic changes were suggested for well-known groups. For example, the class β-*proteobacteria* was reclassified as an order within the class γ-*proteobacteria* and the δ- and ϵ-*proteobacteria* were removed from the phylum *Proteobacteria* (Parks et al., 2018). It is noteworthy that the Silva 16S rRNA database has already started the conversion to this new classification in version 132.

Low rate of functional annotation The tools such as PICRUSt (Langille et al., 2013) infer putative biological functions of a microbial community by linking 16S gene marker data with the closest available microbial genomes. These approaches are highly speculative, since there sometimes are discrepancies between the similarities of 16S rRNA and genotypic convergences (Knight et al., 2018). As a consequence, it is more suitable to use metagenomic data to estimate the effective functional capabilities of an entire microbial community or reconstructed genomes. The functional annotation of metagenomic sequences was mostly based on homology approaches (i.e., the comparison of unknown sequences with those of reference databases) and on pattern or sequence signatures. In the two previous chapters (4 and 5), it was noticed that many sequences could not be annotated because they shared no similarity with known sequences. Although this issue has not been investigated in this work, solutions are available to increase the annotated fraction of genes. Prakash and Taylor reviewed some of these techniques, such as tools using the genomic neighborhood of the targeted sequence (Prakash and Taylor, 2012). Alternatively, gene functions can be inferred using tools specifically designed to predict one type of protein, such as lipoproteins, membrane proteins, transporters and CRISPRs.

Set aside the functional annotation In Chapter 4, we focused on the metabolic similarities between different serpentinizing sites. For that purpose, metagenomes were functionally annotated with orthologs databases such as COG and KEGG. However, this process annotated 55% of putative genes on average, which is quite low. The functional profiles of microbial communities were therefore compared after excluding unannotated genes. This is a significant limitation because the annotated fraction reflected only half of the biological diversity. This limitation

suggests that other coping mechanisms could be shared by microbial communities of serpentinizing environments, but we missed them because associated genes did not have known functional annotation. New methods that bypass the annotation step have been proposed to assess the similarity between metagenomes. For example, the method COMMET directly compares metagenomic raw reads, using K-mers (Maillet et al., 2014). This method was successfully used to establish similarities between Tara Oceans stations, where genomic contents were largely unknown (Villar et al., 2015). We could not apply this method in Chapter 4 since it requires high sequencing depth, which was not reached for some metagenomes of our study. This may constitute a novel avenue of research to investigate common features of serpentinizing sites in future studies.

6.1.3 Genome reconstruction

Metagenome-assembled genomes (MAGs) Gene-centric analysis (i.e., the analysis of genes of an entire community) is gradually being replaced by genome-centric analysis when processing metagenomic data. The reconstruction of MAGs links taxa to potential metabolisms. The pre-requisite to perform genome-centric analysis is a high quality assembly, as only long contigs (typically > 2500bp) are binned. This requirement strongly affects the results, as shown in Chapter 5, where only a few genomes could be reconstructed in comparison to the whole diversity. In addition, MAGs reconstruction often includes a manual refinement step (Anderson et al. (2017), Figueroa et al. (2017) and these are only examples). This step is operator-dependent and not reproducible. Given the high number of MAGs identified by metagenomics analyses, this approach cannot be systematized.

Improvement of genome reconstruction Integrative approaches should be considered to enhance the assembly of complete genomes. The meta3C method is an integrated approach, which combines metagenomics and chromosome conformation capture (Marbouty et al., 2014). The chromosome conformation capture (or "3C") reveals the sequences in the neighborhood of physical contact of DNA molecules. Contigs with enriched physical contacts are likely to correspond to DNA molecules from a same cell compartment and thus provides new information to improve the binning stage. A second method to improve the reconstruction of MAGs has been proposed by Frank et al. (2016). The authors used hybrid assemblies of data sets from two platforms: Illumina HiSeq, which produces short reads, and PacBio circular consensus sequencing, which generates longer reads but with a high error rate (Frank et al., 2016).

6.2 Microbial diversity in PHF

Very high taxonomic diversity Compared to previous studies based on clone libraries (Postec et al., 2015, Quéméneur et al., 2014), a greater microbial diversity was reported using high-throughput sequencing (Chapter 3). Six new phyla were thus identified, including *Fibrobac-*

teres, *Lentisphaerae*, *Marinimicrobia*, PAUC34f, *Saccharibacteria*, and *Verrucomicrobia*. Moreover, the pyrosequencing of 16S rRNA genes highlighted numerous rare phylotypes, which constitute the major part of the diversity in PHF. The dominant bacterial phylotypes did not exceed a relative abundance of 10% in the PHF samples (chapters 3 and 5), while other serpentinite-hosted environments, such as CROMO, are dominated by phylotypes (orders *Burkholderiales* or *Clostridiales*) with a relative abundance up to 60% (Twing et al., 2017). One explanation could be that high gradients and variations of hydrothermal regimes favor the renewal of species and inhibit the colonization by highly abundant species. As the ecosystem of PHF is highly diversified, the reconstruction of all MAGs would require a significantly deep sequencing depth. In Chapter 5, a few MAGs were reconstructed but they were not representative of the diversity detected in PHF. A focus on the microbial community of the hydrothermal fluids may help us target phylotypes from subsurface and validate ecological models of PHF.

Early colonizers of PHF chimneys? *Firmicutes* species were suggested to be part of the first inhabitants of the conduits of the PHF young chimneys (Pisapia et al., 2017), raising questions about their metabolic capabilities and their role in the PHF ecosystem. Cultivation, genomics and metagenomics strategies were used to better characterize these phylotypes. Numerous *Clostridiales*, within the phylum *Firmicutes*, were successfully isolated from PHF (Table 2.2). These strains are all fermentative. Fermentation is assumed to be an interesting metabolic strategy in serpentinizing environments due to the low availability of electron acceptors (almost absent in the anoxic fluid). Indeed, if these microorganisms use the organic carbon produced by the reactions of serpentinization, the fermentation could be a form of primary production from carbon and energy of abiotic origin (Schrenk et al., 2013). To gain insight into some members of the *Firmicutes*, a genomic comparison of five species of the genus *Alkaliphilus* isolated from PHF is underway within our team. This study will attempt to highlight metabolic pathways related to the products of serpentinization and adaptation to alkaline pH. Moreover, the study of the proteins involved in these alkalophilic strains isolated from PHF could lead to the discovery of new and potentially interesting enzymes for industrial applications. In Chapter 5, we successfully reconstructed seven MAGs affiliated with the *Firmicutes*, which offers us hints about their metabolisms. These seven MAGs, assumed to be anaerobic, contained at least one[FeFe]-hydrogenase. This result was consistent with previous observations, indicating that *Clostridiales* could be the main producers of hydrogen in serpentinite-hosted ecosystems (Brazelton et al., 2012, Mei et al., 2014). Two of these MAGs even included both formate dehydrogenase and anaerobic sulfite reductase sequences. These genomic contents are new evidence that supports the hypothesis formulated by Lang et al. (2018), where the conversion of formate to CO_2 by sulfate reducers can supply autotrophs, such as the Lost City Methanosarcinales.

6.3 Microbial communities from distinct serpentinizing sites

One of the objectives of this doctoral research project was to establish a core microbial community which, as it would be present in all serpentinizing sites, could help us characterize the microorganisms and pathways that may benefit from subsurface geochemical reactions. A challenge that remains unresolved is to differentiate between organisms originating from the subsurface and the individuals from the surrounding environments (such as seawater, freshwater, soil) that have adapted to the conditions of the hydrothermal systems.

6.3.1 Core microbial community in serpentinizing systems

Taxonomic similarities between microbial communities from distinct serpentinizing systems were identified repeatedly (e.g. Quéméneur et al., 2014, Suzuki et al., 2013). However, these studies presented two significant weaknesses. On one hand, the taxonomic distribution was often compared at the phylum level (Quéméneur et al., 2014, Woycheese et al., 2015) and was not highly informative because various metabolic strategies could be found within a same phylum. In other hand, other comparison was incomplete as it targeted a few clones and their closest relatives, based on the similarity of their 16S rRNA genes (Suzuki et al., 2013). Nevertheless, hypotheses could be raised about the primary actors in serpentinizing environments such as *Firmicutes*, *Betaproteobacteria*. In Chapter 3, we sought to identify which phylotypes from PHF were also detected in other serpentinizing sites. Surprisingly, very few taxa were shared by these ecosystems. At this species taxonomic level, no core microbial community was identified in all serpentinizing systems. The taxonomic signature was subtler: A few uncultivated lineages, notably within the archaeal order *Methanosarcinales* and the bacterial class *Dehalococcoidia* (the candidate division MSBL5) were exclusively found in a few serpentinizing systems while phylotypes belonging to the orders *Clostridiales*, *Thermoanaerobacterales*, or the genus *Hydrogenophaga*, were abundant in several serpentinite-hosted sites. The low number of phylotypes shared between serpentinizing sites can be related to the potentially low proportion of microbes that originate from subsurface environments. Most of the microbial communities may be only ubiquitous microbes that have adapted to the extreme conditions of serpentinizing environments. However, these results should be taken with caution because the comparison was limited to rRNA sequences with representative sequences in the Silva database. Indeed, different regions of the 16S rRNA genes were targeted in the microbial studies of serpentinizing sites (e.g., V2-V3, V3-V4 and V6), preventing exhaustive comparison with *de-novo* analysis. The comparison of microbial communities of the different serpentinizing systems was therefore complicated due to the lack of a gold standard method. As suggested by Knight et al. (2018), this standardization is required to compare or combine separate studies; this could be achieved by providing recommendations and best practices for environmental studies.

6.3.2 Functional convergence under conditions of serpentinization?

Functional profiles A few taxonomic similarities were detected among geographically distant serpentinizing ecosystems, but overall, microbial communities varied considerably from one system to another. These observations suggested that the geochemical conditions and fluid composition variations could strongly affect the microbial composition and structure of these ecosystems. However, the differences in taxonomic pattern identified among the serpentinizing microbial habitats do not necessarily reflect a divergent functional pattern. Indeed, many distant related species possess similar metabolic capabilities and can potentially ensure the same ecological roles among different ecosystems. In Chapter 4, we hypothesized that similar strategies may be employed by the microbial communities from distinct serpentinizing environments, since the reactions of serpentinization produce similar harsh conditions of life for microbial communities. Our approach was to find a set of necessary genes for these microbial communities; we believed they would be genes involved in the metabolism of serpentinization products and in resistance to alkaline pH. The functional profiles of serpentinizing systems appeared to be generally distinct. Surprisingly, the functional profile of LCHF was more similar to the ones of basalt-hosted hydrothermal vents than to other serpentinizing systems. It would have been interesting to see how sites, such as Rainbow, influenced by both serpentinization and magmatic activities, would be clustered in a similar comparison. Unfortunately, there are no metagenomic data sets from such environments to date.

Identification of biomarkers To finally highlight essential functions of serpentinizing systems, we looked for microbial biomarkers among the annotated genes of metagenomes from various hydrothermal environments. In Chapter 4, a supervised approach (Random Forest) was used to identify genes whose abundance best discriminates two groups of the metagenomes: those from serpentinizing systems and those from other hydrothermal systems. We reported enriched functional capabilities among serpentinite-hosted ecosystems compared to other hydrothermal ecosystems. The metabolic capabilities identified were associated with environmental stress response mechanisms, oxidation of hydrogen resources and especially the phosphonate degradation. The random forest outperformed non-parametric tests. Indeed, the tests that compared the medians of gene abundance between the two groups of metagenomes provided results that were more difficult to interpret as biologically informative features. The metabolic capabilities highlighted by the random forest approach are assumed to be linked to the serpentinization. The abundance of hydrogenases was expected in serpentinizing systems, given the high concentration of H_2. It is nevertheless surprising that microbial communities rely on a same metabolic strategy by using a [NiFe]-hydrogenase of Group 3d, as demonstrated in Chapter 4. Likewise, the systematic presence of the phosphonate degradative pathway despite various concentration of phosphates in the serpentinizing sites is puzzling. One hypothesis is that these metabolic capabilities reflect residual metabolic potential of microbial communities from the deep subsurface, subject to extreme conditions and highly similar from one

serpentinizing site to another.

6.4 Perspectives

6.4.1 Gradients of microbial communities in PHF chimneys?

Motivations and sampling At PHF, microbial communities in the fluids, core of chimneys, and young chimneys differ due to numerous gradients of temperature, pH, DIC, and electron acceptors (Mei et al., 2016b, Pisapia et al., 2017, and Chapter 5). Based on these observations, it would be interesting to focus on microbial communities along a longitudinal transect of one chimney.

To conduct such a study, we will have to select a chimney slice with a large circumference (~30cm). This chimney should be subsampled according to the venting structure previously described by Pisapia et al. (2017): (i) the inner conduit, a few centimeters in diameter, where the fluids circulate, (ii) the intermediate part, that is, the inner walls of the chimney and (iii) the external walls of the chimney that are in contact with the surrounding seawater and can be colonized by macrofauna, depending on the age of the chimney (Figure 6.1). The sampling will be completed by collecting the fluids emitted by this chimney.

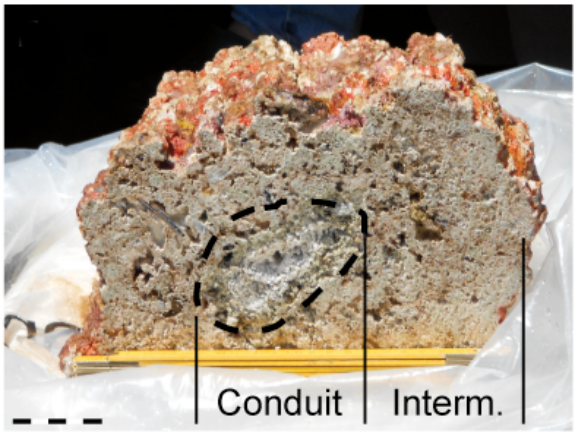


Figure 6.1 – Transversal cut of the hydrothermal chimney ST07, from Pisapia et al. (2017), showing the venting structure. The black scale bar corresponds to 1 cm.

We plan to conduct these analyses in replicates on several chimney slices. This is the only way to ensure that observed variations are statistically significant and unbiased by the sampling and analysis (Prosser, 2010, Thomas et al., 2012). The lack of replicates did not compromise the exploration of new diversity in PHF, but can be problematic when assessing differences between samples in terms of microbial diversity and richness.

Taxonomic distribution A first approach would be to compare the taxonomic distribution among the four compartments (i.e., inner conduit, inner walls, external walls and fluids). The taxonomic profile could be inferred from either 16S rRNA or metagenomic analyses. The analysis of replicates will help us identify the community specificity of each compartment (i.e., those that are not identified in the others). Moreover, we plan to investigate whether some organisms are ubiquitous and can be found in all compartments. Finally, it would be interesting to study the co-occurrence of species, that is, to study species whose presence or absence correlates positively between different samples, and this within each of the four compartments. The co-occurrence patterns may reveal relationships among community members. Indeed, symbiotic

associations are often suggested in serpentinizing sites, but required further investigation. This is the case for members of the candidate phylum OD1, which have very small genomes and lack key biosynthetic pathways, suggesting that they are incapable of independent growth (Suzuki et al., 2017).In addition, the co-occurrence study will provide a chance to assess the presence of syntrophic interactions between *Clostridiales* phylotypes and hydrogenotrophic methanogens suggested in Chapter 3.

Functional profiling Using this subsampling approach, we expect to observe contrasting functional profiles along the mixing gradient of hydrothermal fluids and seawater. The independent study of each compartment will help us complete the ecological model of a PHF chimney, proposed by Postec et al. (2015). In the internal compartments, deprived of oxygen, the main energy metabolisms were assumed to be fermentation, methanogenesis or sulfate reduction. The functional profiling proposed here may facilitate the estimate of their relative abundance in internal compartments. Moreover, other energy metabolisms, such as acetogenesis and dehalorespiration should be considered in anaerobic compartments, based the presence of *Acetothermia* and *Chloroflexi* phylotypes (Pisapia et al., 2017, Suzuki et al., 2017, and Chapter 3). Regarding the oxic compartments, it would be interesting to determine whether the microaerophilic or aerobic microorganisms rely mainly on the abiotic products of the serpentinization, such as H_2, CH_4. To highlight the specificity of the functional profiles of each compartment, we would like to investigate and compare two approaches: (i) a metagenomic analysis to evaluate the potential metabolism in term of genomic content and (ii) a thermodynamic modeling to assess the potential energy available for microbial metabolism in the chimney. The metagenomes of the four compartments will be annotated with functional databases. Functional profiles will be compared to find marker genes either using the same approach as in Chapter 4 (i.e., Random Forest) or using linear discriminant analysis for example LEfSe Segata et al. (2011). These biomarkers could highlight specific metabolisms per compartment. With our second approach, we intend to evaluate the potential energy available for microbial metabolisms at PHF using thermodynamic modeling. Such a strategy was used to estimate feasible metabolisms in serpentinizing springs in the Zambales Ophiolites (Cardace et al., 2015). Values of Gibbs energy (ΔG_r) for a metabolic reaction will be computed using temperatures and chemical compositions of the fluids with the relation:

$$\Delta G_r = \Delta G_r^0 + RT\,ln Q_r \tag{6.1}$$

where ΔG_r^0, corresponds to the standard Gibbs energy values, R represents the gas constant, T denotes temperature, and Q_r is the activity product of the reaction r. In this way, we could estimate the distribution of major metabolisms along the mixing gradient between hydrothermal fluids and seawater, by iterating the energy calculation for different fluid and sea water ratios. The development of thermodynamic modeling is envisaged in collaboration with the laboratory Geosciences Environnement Toulouse. The comparison between thermodynamic mod-

elling and metagenomic profile may exhibit discrepancies regarding the relative abundance of metabolisms, as reported by Reveillaud et al. (2016). Metagenomic analysis alone may lead to slight misinterpretation of microbial activities, since it does not provide the level of gene expression (up- or down-regulated) in natural conditions. Thus, combining the genomic potential (metagenomics) of modeled metabolisms (thermodynamics) would provide better insight into the metabolic potential under natural conditions.

6.4.2 The presence of phosphonates in serpentinizing systems

Chapter 4 highlighted an enrichment of *phnGHIJKLM* in serpentinizing ecosystems compared to other hydrothermal systems. These seven genes encode the catalytic components of the C-P lyase. Recently, aerobic degradation of phosphonates (Pn) via the C-P lyase pathway has attracted strong interest because it may explain the marine methane paradox (i.e., the biotic production of methane in oxic environments). Indeed, this pathway was reported to be associated with methane production in cultures amended with exogenous methylphosphonates (MPn) (Karl et al., 2008) or dissolved organic matter (DOM), including polysaccharide-MPn esters (Repeta et al., 2016). Geochemical measurements, cultures associated with monitoring of the C-P lyase activity, should be carried out to determine whether such mechanisms could contribute to the production of biotic methane or other organic molecules in serpentinizing environments. Such investigation would also provide more information on phosphorus (P) metabolism. The manipulations mentioned above are conditional on the access to new samples. Indeed, the samples collected during previous expeditions are too old and damaged to start new cultures or measure concentration of chemical compounds.

In-situ geochemical measurement As a priority, *in situ* geochemical measurements should be carried out to characterize the main form of P in these systems. Although some studies have investigated concentration of phosphate ions (PO_4^{3-}) in serpentinizing ecosystems (Cardace et al., 2015, Crespo-Medina et al., 2014, 2017, Morrill et al., 2013), none looked at the presence of Pn and their abundance relative to PO_4^{3-}. The dissolved organic phosphorus can be characterized using ^{31}P nuclear magnetic resonance (NMR) spectra (Kolowith et al., 2001, Repeta et al., 2016).

Cultures We will have to confirm the expression of the *phn* genes, because our results (Chapter 4) confirmed only the microbial potential to degrade Pn in serpentinizing environments. In other environments, such as the sea or a freshwater lake, the ability of microorganisms to use C-P lyase was demonstrated using cultures amended with Pn as the sole source of P (Karl et al., 2008, Repeta et al., 2016, Yao et al., 2016). The production of methane and ethylene was monitored by gas chromatography. Repeta and collaborators reported that a mutation of the *phnK* gene in a methane-producing bacterium inhibited methanogenesis (Repeta et al., 2016). Yao and colleagues provided additional evidence linking methane production to C-P lyase activity,

with higher transcript abundance of *phnJ* (marker gene of C-P lyase pathway) in the presence of MPn (Yao et al., 2016). On the basis of these studies, we plan to assess the potential of biotic methane production via the C-P lyase in PHF. For that purpose, enrichment cultures could be grown with a different source of P. In parallel, we want to confirm the activity of C-P lyase in the PHF ecosystem using metatranscriptomic analyses.

Organization of the genes in *phn* operon The analysis of metagenomic data led to characterize the organizations of a few *phn* operons from serpentinizing metagenomes (Chapter 4). However, most of the operon were incomplete because the metagenomes were not properly assembled. To establish a more complete assessment of the genomic diversity as well as the evolutionary history of *phn* operons in serpentinizing environments, we plan to use a gene-capture strategy. This hybridization-based approach reduces the complexity of metagenomic samples before the sequencing step (Denonfoux et al., 2013). The gene-capture approach selects large DNA fragments containing genes of interest and adjacent genomic regions. The genomic sequences reconstructed using a such method will most likely provide a more accurate taxonomic signal.

Isolation of new strains The ability to use phosphonates (via the *phn* genes) or phospites (via *ptxD* gene) as the only source of phosphorus could enable the isolation new microorganisms from PHF or other serpentinizing environments. Cultivation media could therefore be modified by replacing the P source with phosphite, MPn or other Pn. From the MAGs identified in Chapter 5, we know that the C-P lyase is present in some members of the *Proteobacteria* (including the MAG-56, affiliated with the genus *Serpentinomonas*), *Chloroflexi* and *Deinococcus-Thermus*. The knowledge of the metabolic potential of target organisms is an asset in isolating new strains.

6.4.3 Microbial dark matter in PHF

Previous attempts to isolate LCMS or TCMS (affiliated to the order *Methanosarcinales*) from PHF failed. The main obstacle is probably that they need to grow in synergy with other microorganisms. Another explanation would be that culture media do not sufficiently fit their environmental conditions. We thought that culture-independent methods could easily overcome this problem. However, no LCMS or TCMS have been reconstructed from the metagenomes of PHF (Chapter 5). Besides, despite the fact that these phylotypes were identified in many serpentinizing environments, no genome has been reconstructed using binning approaches to date (Brazelton et al., 2016, Suzuki et al., 2017). An alternative way to access their genomic content would be to isolate these *Methanosarcinales* either using laser microdissection (since their morphology is easily recognizable) or using a method that combines fluorescence *in situ* hybridization and flow cytometry (Haroon et al., 2013) to finally carry out a single-cell sequencing and generate Single Amplification Genomes.

Conclusion

During this thesis work, we investigated the taxonomic and functional profiles of the microbial communities that colonized serpentinizing systems. A first study, based on the analysis of 16S rRNA gene, highlighted the very high bacterial diversity of PHF sites. Most of the phylotypes found in these environments tended to be in very low abundance and were part of the rare biosphere. To identify the microbes specifically associated with serpentinizing hydrothermal fluids, all PHF phylotypes where compared with those of other serpentinizing systems. Surprisingly, very few phylotypes were identified in common with other systems. Despite distinct taxonomic profiles, we hypothesized that microbial communities in serpentinizing environments shared functional capacities to cope with similar environmental stresses. This hypothesis was the subject of our second study.

In this comparative metagenomic survey of serpentinizing ecosystems, we aimed to establish a functional core shared by microbes in serpentinizing environments. Overall, the taxonomic and functional annotations associated with microorganisms in serpentinizing systems are divergent. Indeed, the well-known site of Lost City was closer to other deep hydrothermal sites which are not affected by the reactions of serpentinization. Nevertheless, our comparison outlined common metabolic features that seem specific to serpentinizing ecosystems, mainly linked to hydrogen metabolism in agreement with previous studies by our group and others in different sites. Most interestingly, this analysis revealed an unexpected enrichment of a gene cluster encoding a phosphonate degradative pathway. We proposed that phosphonate degradation through the C-P lyase pathway was widespread among serpentinizing systems and could play a significant role in the biogeochemical cycle of phosphorus and production of small hydrocarbon molecules such as methane.

In the second study, the metabolic potentials cannot be precisely related to taxa. For these reasons, in a last study we tried to reconstruct genomes *in silico* from metagenomes obtained with a very deep depth of sequencing. The genome-centric analysis of five metagenomes from PHF generated 82 prokaryotic genomes. Among these genomes, we identified an archaea associated with *Woesearchaeota* phylum that lacks many essential metabolic functions. Another genome was of particular interest because it was closely related to the species *Serpentinomonas mccroryi* strain B1, isolated from The Cedars. Their genomic similarity suggested that the phylotype found at PHF could also grow in autotrophic conditions.

References

Abouhamad, W.N., Manson, M., Gibson, M.M. and Higgins, C.F. (1991). Peptide transport and chemotaxis in Escherichia coli and Salmonella typhimurium: characterization of the dipeptide permease (Dpp) and the dipeptide-binding protein. *Molecular Microbiology*, 5(5):1035–1047. 88

Alfreider, A., Vogt, C. and Babel, W. (2002). Microbial diversity in an *in situ* reactor system treating monochlorobenzene contaminated groundwater as revealed by 16S ribosomal DNA analysis. *Systematic and applied microbiology*, 25(2):232–40. 67

Aller, J.Y. and Kemp, P.F. (2008). Are Archaea inherently less diverse than Bacteria in the same environments? *FEMS Microbiology Ecology*, 65(1):74–87. 61

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F. and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11(11). 45

Amend, J.P., McCollom, T.M., Hentscher, M. and Bach, W. (2011). Catabolic and anabolic energy for chemolithoautotrophs in deep-sea hydrothermal systems hosted in different rock types. *Geochimica et Cosmochimica Acta*, 75(19):5736–5748. 26

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106. 43

Anderson, R.E., Reveillaud, J., Reddington, E., Delmont, T.O., Eren, A.M., McDermott, J.M., Seewald, J.S. and Huber, J.A. (2017). Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. *Nature Communications*, 8(1). 125

Anderson, R.E., Sogin, M.L. and Baross, J.A. (2014). Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics. *PloS one*, 9(10):e109696. 48, 96

Anderson, R.E., Sogin, M.L. and Baross, J.A. (2015). Biogeography and ecology of the rare and abundant microbial lineages in deep-sea hydrothermal vents. *FEMS Microbiology Ecology*, 91(1):1–11. 61

Auguet, J.C., Barberan, A. and Casamayor, E.O. (2010). Global ecological patterns in uncultured Archaea. *The ISME Journal*, 4(2):182–190. 64

Baker, B.J., Saw, J.H., Lind, A.E., Lazar, C.S., Hinrichs, K.U., Teske, A.P. and Ettema, T.J. (2016). Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nature Microbiology*, 1(3):1–7. 18

Balvočiute, M. and Huson, D.H. (2017). SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics*, 18(Suppl 2):1–8. 123

Barnes, I., Lamarche, V.J. and Himmelberg, G. (1967). Geochemical evidence of present-day serpentinization. *Science*, 156:830–832. 10, 11

Baross, J.A. and Hoffman, S.E. (1985). Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Origins of Life and Evolution of the Biosphere*, 15(4):327–345. 15

Bath, A., Christofi, N., Neal, C., Philp, J.C., Cave, M., McKinley, I.G. and Berner, U. (1987). Trace element and microbiological studies of alkaline groundwaters in Oman, Arabian Gulf: A natural analogue for cement pore-waters. *British Geology Survey*, FLPU 87-2. 11, 30, 48

Beaulieu-Jones, B.K. and Greene, C.S. (2017). Reproducibility of computational workflows is automated using continuous analysis. *Nature Biotechnology*, 35(4):342–346. 47

Becraft, E.D., Dodsworth, J.A., Murugapiran, S.K., Ohlsson, I., Briggs, B.R., Kanbar, J., De Vlaminck, I., Quake, S.R., Dong, H., Hedlund, B.P. and Swingley, W.D. (2016). Single-cell-genomics-facilitated read binning of candidate phylum EM19 genomes from geothermal spring metagenomes. *Applied and Environmental Microbiology*, 82(4):992–1003. 50

Ben Aissa, F., Postec, A., Erauso, G., Payri, C., Pelletier, B., Hamdi, M., Fardeau, M.L. and Ollivier, B. (2015). Characterization of Alkaliphilus hydrothermalis sp. nov., a novel alkaliphilic anaerobic bacterium, isolated from a carbonaceous chimney of the Prony Hydrothermal Field, New Caledonia. *Extremophiles*, 19(1):183–188. 30, 48, 49, 55, 101

Ben Aissa, F., Postec, A., Erauso, G., Payri, C., Pelletier, B., Hamdi, M., Ollivier, B. and Fardeau, M.L. (2014). *Vallitalea pronyensis* sp. nov., isolated from a marine alkaline hydrothermal chimney. *International Journal of Systematic and Evolutionary Microbiology*, 64(PART 4):1160–1165. 49, 55, 101

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hügler, M., Alber, B.E. and Fuchs, G. (2010). Autotrophic carbon fixation in archaea. *Nature reviews. Microbiology*, 8(6):447–60. 28, 87

Bes, M., Merrouch, M., Joseph, M., Quéméneur, M., Payri, C., Pelletier, B., Ollivier, B., Fardeau, M.L., Erauso, G. and Postec, A. (2015). Acetoanaerobium pronyense sp. nov., an anaerobic alkaliphilic bacterium isolated from a carbonate chimney of the Prony Hydrothermal Field (New Caledonia). *International Journal of Systematic and Evolutionary Microbiology*, 65(8):2574–2580. 17, 30, 48, 49, 55, 101

Biderre-Petit, C., Dugat-Bony, E., Mege, M., Parisot, N., Adrian, L., Moné, A., Denonfoux, J., Peyretaillade, E., Debroas, D., Boucher, D. and Peyret, P. (2016). Distribution of *Dehalococcoidia* in the anaerobic deep water of a remote meromictic crater lake and detection of *Dehalococcoidia*-derived reductive dehalogenase homologous genes. *PLoS ONE*, 11(1):1–19. 20, 67

Birrien, J.L., Zeng, X., Jebbar, M., Cambon-Bonavita, M.A., Quérellou, J., Oger, P., Bienvenu, N., Xiao, X. and Prieur, D. (2011). *Pyrococcus yayanosii* sp. nov., an obligate piezophilic hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology*, 61(12):2827–2831. 49

Blank, J.G., Green, S.J., Blake, D., Valley, J.W., Kita, N.T., Treiman, A. and Dobson, P.F. (2009). An alkaline spring system within the Del Puerto Ophiolite (California, USA): A Mars analog site. *Planetary and Space Science*, 57(5-6):533–540. 12, 27, 49

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, I., Knight, R., Mills, D.A. and Caporaso, J.G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods*, 10(1):57–59. 58

Bolger, A.M., Lohse, M. and Usadel, B. (2014). Trimmomatic: A flexible read trimming tool for Illumina NGS data. *Bioinformatics*, 30(15):2114–2120. 79, 103

Bowers, R.M. et al (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, 35(8):725–731. 111

Bradley, A.S. (2016). The sluggish speed of making abiotic methane. *Proceedings of the National Academy of Sciences*, 113(49):13944–13946. 6

Bradley, A.S., Fredricks, H., Hinrichs, K.U. and Summons, R.E. (2009). Structural diversity of diether lipids in carbonate chimneys at the Lost City Hydrothermal Field. *Organic Geochemistry*, 40(12):1169–1178. 17, 33, 78

Brazelton, W.J. and Baross, J.A. (2009). Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *The ISME Journal*, 3:1420–1424. 48, 49, 96

Brazelton, W.J. and Baross, J.A. (2010). Metagenomic comparison of two *Thiomicrospira* lineages inhabiting contrasting deep-sea hydrothermal environments. *PLoS ONE*, 5(10). 29, 31, 49, 84

Brazelton, W.J., Ludwig, K.A., Sogin, M.L., Andreishcheva, E.N., Kelley, D.S., Shen, C.C., Edwards, R.L. and Baross, J.A. (2010a). Archaea and bacteria with surprising microdiversity show shifts in dominance over 1,000-year time scales in hydrothermal chimneys. *PNAS*, 107(4):1612–1617. 19, 21, 23, 49, 94

Brazelton, W.J., Mehta, M.P., Kelley, D.S. and Baross, J.A. (2011). Physiological differentiation within a single-species biofilm fueled by serpentinization. *mBio*, 2(4):1–9. 27, 31, 32, 78, 94, 101

Brazelton, W.J., Morrill, P.L., Szponar, N. and Schrenk, M.O. (2013). Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs. *Applied and Environmental Microbiology*, 79:3906–3916. 22, 49, 55

Brazelton, W.J., Nelson, B. and Schrenk, M.O. (2012). Metagenomic evidence for H_2 oxidation and H_2 production by serpentinite-hosted subsurface microbial communities. *Frontiers in Microbiology*, 2. 19, 26, 29, 30, 49, 70, 78, 126

Brazelton, W.J., Schrenk, M.O., Kelley, D.S. and Baross, J.A. (2006). Methane-and sulfur-metabolizing microbial communities dominate the Lost City Hydrothermal Field ecosystem. *Applied and Environmental Microbiology*, 72(9):6257–6270. 18, 20, 49, 58, 69, 71, 72, 77

Brazelton, W.J., Sogin, M.L. and Baross, J.A. (2010b). Multiple scales of diversification within natural populations of archaea in hydrothermal chimney biofilms. *Environmental Microbiology Reports*, 2(2):236–242. 49, 55, 61, 70

Brazelton, W.J., Thornton, C.N., Hyer, A., Twing, K.I., Longino, A.A., Lang, S.Q., Lilley, M.D., Früh-Green, G.L. and Schrenk, M.O. (2016). Metagenomic identification of active methanogens and methanotrophs in serpentinite springs of the Voltri Massif, Italy. *PeerJ*, 5:e2945. 12, 19, 26, 27, 28, 49, 50, 55, 78, 81, 95, 102, 119, 132

Brown, C.T. and Irber, L. (2016). sourmash: a library for MinHash sketching of DNA. *The Journal of Open Source Software*, 1(5):27. 104

Buchfink, B., Xie, C. and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature methods*, 12(1):59–60. 46, 80, 103

Burgdorf, T., Lenz, O., Buhrke, T., Van Der Linden, E., Jones, A.K., Albracht, S.P. and Friedrich, B. (2006). [NiFe]-hydrogenases of Ralstonia eutropha H16: modular enzymes for oxygen-tolerant biological hydrogen oxidation. *Journal of Molecular Microbiology and Biotechnology*, 10(2-4):181–196. 94

Cabria, G.L.B., Argayosa, V.B., Lazaro, J.E.H., Argayosa, A.M. and Arcilla, C.A. (2014). Draft genome sequence of haloalkaliphilic Exiguobacterium sp. strain AB2 from Manleluag Ophiolitic spring, Philippines. *Genome announcements*, 2(4):2–3. 49

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7):581–583. 41, 103

Canet, C., Maria, R. and Ledesma, P. (2007). Mineralizing processes at shallow submarine hydrothermal vents : Examples from Mexico. *Geological Society of America Special papers*, 422:359–376. 101

Cao, H., Wang, Y., Lee, O.O., Zeng, X., Shao, Z. and Qian, P.Y. (2014). Microbial sulfur cycle in two hydrothermal chimneys on the Souhwest Indian Ridge. *mBio*, 5(1):1–11. 50

Cao, J., Maignien, L., Shao, Z., Alain, K. and Jebbar, M. (2016). Genome sequence of the piezophilic, mesophilic sulfate-reducing bacterium *Desulfovibrio indicus* J2T. *Genome Announcements*, 4(2). 49

Caporaso, J.G., Bittinger, K., Bushman, F.D., Desantis, T.Z., Andersen, G.L. and Knight, R. (2010a). PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26(2):266–267. 58

Caporaso, J.G. et al (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nature Publishing Group*, 7(5):335–336. 41, 58

Cardace, D., Meyer-Dombard, D.A.R., Woycheese, K.M. and Arcilla, C.A. (2015). Feasible metabolisms in high pH springs of the Philippines. *Frontiers in Microbiology*, 6. 12, 78, 130, 131

Castelle, C.J., Brown, C.T., Thomas, B.C., Williams, K.H. and Banfield, J.F. (2017). Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate Phyla Radiation. *Scientific Reports*, 7(January):40101. 73

Charlou, J., Donval, J., Fouquet, Y., Jean-Baptiste, P. and Holm, N. (2002). Geochemistry of high H_2 and CH_4 vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36°14'N, MAR). *Chemical Geology*, 191(4):345–359. 9

Charlou, J.L., Donval, J.P., Konn, C., Ondréas, H., Fouquet, Y., Jean-Baptiste, P. and Fourré, E. (2010). *High production and fluxes of H_2 and CH_4 and evidence of biotic hydrocarbon synthesis by serpentinization in ultramafic-hosted hydrothermal systems on the Mid-Atlantic Ridge*. American Geophysical Union (AGU). 9

Chavagnac, V., Monnin, C., Ceuleneer, G., Boulart, C. and Hoareau, G. (2013). Characterization of hyperalkaline fluids produced by low-temperature serpentinization of mantle peridotites in the Oman and Ligurian ophiolites. *Geochemistry, Geophysics, Geosystems*, 14(7):2496–2522. 12, 55

Chin, J.P., Quinn, J.P. and McGrath, J.W. (2018). Phosphate insensitive aminophosphonate mineralisation within oceanic nutrient cycles. *ISME Journal*, pages 1–8. 96

Chun, S.J., Cui, Y., Ko, S.R., Lee, H.G., Oh, H.M. and Ahn, C.Y. (2017). *Silanimonas algicola* sp. nov., isolated from laboratory culture of a bloom-forming cyanobacterium, Microcystis. *International Journal of Systematic and Evolutionary Microbiology*, 67(9):3274–3278. 72

Ciazela, J., Koepke, J., Dick, H.J. and Muszynski, A. (2015). Mantle rock exposures at oceanic core complexes along mid-ocean ridges. *Geologos*, 21(4):207–231. 5

Cooper, G.W., Onwo, W.M. and Cronin, J.R. (1992). Alkyl phosphonic acids and sulfonic acids in the Murchison meteorite. *Geochimica et Cosmochimica Acta*, 56(11):4109–4115. 97

Corliss, J.B., Dymond, J., Gordon, L.I., Edmond, J.M., von Herzen, R.P., Ballard, R.D., Green, K., Williams, D., Bainbridge, A., Crane, K. and van Andel, T.H. (1979). Submarine thermal springs on the Galápagos Rift. *Science*, 203(4385):1073–1083. 8

Cowan, D.A., Ramond, J.B., Makhalanyane, T.P. and De Maayer, P. (2015). Metagenomics of extreme environments. *Current Opinion in Microbiology*, 25(JUNE):97–102. 48

Crépeau, V., Cambon Bonavita, M.A., Lesongeur, F., Randrianalivelo, H., Sarradin, P.M., Sarrazin, J. and Godfroy, A. (2011). Diversity and function in microbial mats from the Lucky Strike hydrothermal vent field. *FEMS Microbiology Ecology*, 76(3):524–540. 84

Crespo-Medina, M., Twing, K.I., Kubo, M.D.Y., Hoehler, T.M., Cardace, D., McCollom, T. and Schrenk, M.O. (2014). Insights into environmental controls on microbial communities in a continental serpentinite aquifer using a microcosm-based approach. *Frontiers in Microbiology*, 5(November):604. 12, 17, 49, 96, 131

Crespo-Medina, M., Twing, K.I., Sánchez-Murillo, R., Brazelton, W.J., McCollom, T.M. and Schrenk, M.O. (2017). Methane dynamics in a tropical serpentinizing environment: the Santa Elena Ophiolite, Costa Rica. *Frontiers in Microbiology*, 8(May). 12, 19, 27, 49, 50, 55, 78, 81, 95, 96, 131

Curtis, A.C., Wheat, C.G., Fryer, P. and Moyer, C.L. (2013). Mariana forearc serpentinite mud volcanoes harbor novel communities of extremophilic Archaea. *Geomicrobiology Journal*, 30(5):430–441. 49

Curtis, T.P. and Sloan, W.T. (2005). Exploring microbial diversity - a vast below. *Science*, 309(5739):1331–1333. 56

Daae, F.L., Økland, I., Dahle, H., Jørgensen, S.L., Thorseth, I.H. and Pedersen, R.B. (2013). Microbial life associated with low-temperature alteration of ultramafic rocks in the Leka ophiolite complex. *Geobiology*, 11(4):318–339. 19, 21, 49, 58, 68, 69, 71, 72

Daffonchio, D. et al (2006). Stratified prokaryote network in the oxic-anoxic transition of a deep-sea halocline. *Nature*, 440(7081):203–7. 66

Delacour, A., Früh-Green, G.L., Bernasconi, S.M., Schaeffer, P. and Kelley, D.S. (2008). Carbon geochemistry of serpentinites in the Lost City Hydrothermal system (30°N, MAR). *Geochimica et Cosmochimica Acta*, 72(15):3681–3702. 30

Denonfoux, J., Parisot, N., Dugat-Bony, E., Biderre-Petit, C., Boucher, D., Morgavi, D.P., Paslier, D.L., Peyretaillade, E. and Peyret, P. (2013). Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Research*, 20(2):185–196. 50, 132

Dinsdale, E.A., Edwards, R.A., Bailey, B.A., Tuba, I., Akhter, S., McNair, K., Schmieder, R., Apkarian, N., Creek, M., Guan, E., Hernandez, M., Isaacs, K., Peterson, C., Regh, T. and Ponomarenko, V. (2013). Multivariate analysis of functional metagenomes. *Frontiers in Genetics*, 4(APR):1–25. 43

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797. 93, 104

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–2461. 58, 104

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16):2194–2200. 40, 58

Ehlmann, B.L., Mustard, J.F. and Murchie, S.L. (2010). Geologic setting of serpentine deposits on Mars. *Geophysical Research Letters*, 37(6). 8

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L. and Delmont, T.O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, 3:e1319. 58, 104

Escobar-Zepeda, A., Vera-Ponce de León, A. and Sanchez-Flores, A. (2015). The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, 6(December):348. 36, 39, 63

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S. and Pascal, G. (2018). FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics*, 34(8):1287–1294. 41

Etiope, G. and Ionescu, A. (2015). Low-temperature catalytic CO_2 hydrogenation with geological quantities of ruthenium: A possible abiotic CH_4 source in chromitite-rich serpentinized rocks. *Geofluids*, 15(3):438–452. 7

Feinberg, L.F., Srikanth, R., Vachet, R.W. and Holden, J.F. (2008). Constraints on anaerobic respiration in the hyperthermophilic archaea Pyrobaculum islandicum and Pyrobaculum aerophilum. *Applied and Environmental Microbiology*, 74(2):396–402. 84

Fernàndez-Guerra, A. and Casamayor, E.O. (2012). Habitat-associated phylogenetic community patterns of microbial ammonia oxidizers. *PLoS ONE*, 7(10):22–26. 64

Figueroa, I.A., Barnum, T.P., Somasekhar, P.Y., Carlström, C.I., Engelbrektson, A.L. and Coates, J.D. (2017). Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO_2 fixation pathway. *Proceedings of the National Academy of Sciences*, 115. 125

Flores, G.E., Campbell, J.H., Kirshtein, J.D., Meneghin, J., Podar, M., Steinberg, J.I., Seewald, J.S., Tivey, M.K., Voytek, M.A., Yang, Z.K. and Reysenbach, A.L. (2011). Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environmental Microbiology*, 13(8):2158–2171. 19, 49, 94

Fortunato, C.S. and Huber, J.A. (2016). Coupled RNA-SIP and metatranscriptomics of active chemolithoautotrophic communities at a deep-sea hydrothermal vent. *The ISME Journal*, 10(8):1–14. 81

Frank, J.A., Pan, Y., Tooming-Klunderud, A., Eijsink, V., McHardy, A., Nederbragt, A.J. and Pope, P. (2016). Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *Scientific reports*, 6:1–26. 125

Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C. and Huttenhower, C. (2015). Sequencing and beyond: Integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology*, 13(6):360–372. 40, 51

Frouin, E., Bes, M., Ollivier, B., Quéméneur, M., Postec, A., Debroas, D., Armougom, F. and Erauso, G. (2018). Diversity of rare and abundant prokaryotic phylotypes in the Prony Hydrothermal Field and comparison with other serpentinite-hosted ecosystems. *Frontiers in Microbiology*, 9(February):102. 49, 77, 94, 101

Früh-Green, G.L., Kelley, D.S., Bernasconi, S.M., Karson, J.A., Ludwig, K.A., Butterfield, D.A., Boschi, C. and Proskurowski, G. (2003). 30,000 years of hydrothermal activity at the Lost City vent field. *Science*, 301(s5632):495–498. 9

Fryer, P. (2012). Serpentinite mud volcanism: Observations, processes, and implications. *Annual Review of Marine Science*, 4(1):345–373. 10

Fukaya, F., Promden, W., Hibino, T., Tanaka, Y., Nakamura, T. and Takabe, T. (2009). An Mrp-like cluster in the halotolerant cyanobacterium *Aphanothece halophytica* functions as a Na+/H+antiporter. *Applied and Environmental Microbiology*, 75(20):6626–6629. 88

Fuks, G., Elgart, M., Amir, A., Zeisel, A., Turnbaugh, P.J., Soen, Y. and Shental, N. (2018). Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. *Microbiome*, 6(1):1–13. 123

Gaill, F. (1993). Aspects of life development at deep-sea hydrothermal vents. *The FASEB Journal*, 7(6):558–565. 9

Gallant, R.M. and Von Damm, K.L. (2006). Geochemical controls on hydrothermal fluids from the kairei and edmond vent fields, 23°-25°S, Central Indian Ridge. *Geochemistry, Geophysics, Geosystems*, 7(6):1–24. 9

Gerasimchuk, A.L., Shatalov, A.A., Novikov, A.L., Butorova, O.P., Pimenov, N.V., Lein, A.Y., Yanenko, A.S. and Karnachuk, O.V. (2010). The search for sulfate-reducing bacteria in mat samples from the Lost City Hydrothermal Field by molecular cloning. *Microbiology*, 79(1):96–105. 22, 31, 49

Goodwin, S., McPherson, J.D. and McCombie, W.R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6):333–351. 37, 39

Gouy, M., Guindon, S. and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, 27(2):221–224. 58

Gubry-Rangina, C., Kratsch, C., Williams, T.A., McHardy, A.C., Embley, T.M., Prosser, J.I. and Macqueen, D.J. (2015). Coupling of diversification and pH adaptation during the evolution of terrestrial *Thaumarchaeota*. *PNAS*, 112(30):9370–9375. 72

Guindon, S. and Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, 52(5):696–704. 104

Gull, M., Cafferty, B., Hud, N. and Pasek, M. (2017). Silicate-promoted phosphorylation of glycerol in non-aqueous solvents: A prebiotically plausible route to organophosphates. *Life*, 7(3):1–10. 97

Haggerty, J.a. and Fisher, J.B. (1992). Short-chain organic acids in interstitial waters from Mariana and Bonin Forearc serpentines: Leg 125. In Fryer, P., Coleman, P., Pearce, J.A. and Stokking, L.B., editors, *Proceedings of the Ocean Drilling Program, Scientific Results*, volume 125, pages 387–395. College Station, TX (Ocean Drilling Program). 9, 10

Haroon, M.F., Skennerton, C.T., Steen, J.A., Lachner, N., Hugenholtz, P. and Tyson, G.W. (2013). *In-solution fluorescence in situ hybridization and Fluorescence-activated cell sorting for single cell and population genome recovery*, volume 531. Elsevier Inc., 1 edition. 132

Hay, I.D., Rehman, Z.U., Moradali, M.F., Wang, Y. and Rehm, B.H.A. (2013). Microbial alginate production, modification and its applications. *Microbial Biotechnology*, 6(6):637–650. 88

He, Y., Xiao, X. and Wang, F. (2013). Metagenome reveals potential microbial degradation of hydrocarbon coupled with sulfate reduction in an oil-immersed chimney from Guaymas Basin. *Frontiers in Microbiology*, 4(JUN):1–13. 50

Hoehler, T.M. (2007). An energy balance concept for habitability. *Astrobiology*, 7(6):824–838. 16

Hoffman, J.I. (2016). Archive computer code with raw data. *Nature*, 534:326. 47

Holm, N.G., Dumont, M., Ivarsson, M. and Konn, C. (2006). Alkaline fluid circulation in ultramafic rocks and formation of nucleotide constituents: A hypothesis. *Geochemical Transactions*, 7:14–16. 5

Horikoshi, K. (1999). Alkaliphiles: some applications of their products for biotechnology. *Microbiology and molecular biology reviews : MMBR*, 63(4):735–750, table of contents. 17

Huang, J., Su, Z. and Xu, Y. (2005). The evolution of microbial phosphonate degradative pathways. *Journal of Molecular Evolution*, 61(5):682–690. 89, 90

Hugenholtz, P. and Tyson, G.W. (2008). Metagenomics. *Nature*, 455(August 2006):481–483. 42, 43

Hügler, M. and Sievert, S.M. (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annual Review of Marine Science*, 3(1):261–289. 29

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Jouan Dufournel, I., Bronner, G., Salter, I., Agogué, H., Mary, I. and Galand, P.E. (2013). Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proceedings of the National Academy of Sciences*, 110(15):6004–9. 58

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J. and Tappu, R. (2016). MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Computational Biology*, 12(6):1–12. 80, 103

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11:119. 79, 103

Itävaara, M., Nyyssönen, M., Kapanen, A., Nousiainen, A., Ahonen, L. and Kukkonen, I. (2011). Characterization of bacterial diversity to a depth of 1500m in the Outokumpu deep borehole, Fennoscandian Shield. *FEMS Microbiology Ecology*, 77(2):295–309. 11, 31, 49

Jonsson, V., Österlund, T., Nerman, O. and Kristiansson, E. (2016). Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. *BMC Genomics*, 17(1):78. 43

Juteau, T. (2003). Identification of a mantle unit in ophiolites: A major step in the evolution of the ophiolite concept. In Dilek, Y. and Newcomb, S., editors, *Ophiolite concept and the evolution of geological thought*, pages 31–53. Geological Society of America, Boulder, Colorado. 10

Kanehisa, M., Sato, Y. and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, 428(4):726–731. 80, 104

Kang, D.D., Froula, J., Egan, R. and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, 3:e1165. 44, 103

Karl, D.M. (2014). Microbially mediated transformations of phosphorus in the sea: New views of an old cycle. *Annual Review of Marine Science*, 6(1):279–337. 78

Karl, D.M., Beversdorf, L., Björkman, K.M., Church, M.J., Martinez, A. and Delong, E.F. (2008). Aerobic production of methane in the sea. *Nature Geoscience*, 1(7):473–478. 82, 92, 131

Karl, D.M. and Tien, G. (1992). Magic - a sensitive and precise method for measuring dissolved phosphorus in aquatic environments. *Limnology and Oceanography*, 37(1):105–116. 17

Kaster, A.K., Mayer-Blackwell, K., Pasarelli, B. and Spormann, A.M. (2014). Single cell genomic study of *Dehalococcoidetes* species from deep-sea sediments of the Peruvian Margin. *The ISME journal*, 8(9):1–12. 20, 67

Katoh, K., Misawa, K., Kuma, K.i. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30(14):3059–3066. 58

Kawasumi, T., Igarashi, Y., Kodama, T. and Minoda, Y. (1984). *Hydrogenobacter thermophilus* gen. nov. , sp. nov. , an Extremely Thermophilic, Aerobic, Hydrogen-Oxidizing Bacterium. *International Journal of Systematic Bacteriology*, 34(1):5–10. 84

Kelemen, P.B. et al (2004). SITE 1271. *Proceedings of the Ocean Drilling Program, Initial Reports*, 209. 6

Kelley, D.S., Karson, J.A., Blackman, D.K., Früh-Green, G.L., Butter, D.A., Lilley, M.D., Olson, E.J., Schrenk, M.O., Roek, K.K. and Lebonk, G.T. (2001). An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30° N. *NATURE*, 412(12):145–149. 8, 9, 19, 77, 94

Kelley, D.S. et al (2005). A serpentinite-hosted ecosystem: the Lost City Hydrothermal Field. *Science*, 307:1428–1434. 9, 27, 94

Kim, S.J., Park, S.J., Cha, I.T., Min, D., Kim, J.S., Chung, W.H., Chae, J.C., Jeon, C.O. and Rhee, S.K. (2014). Metabolic versatility of toluene-degrading, iron-reducing bacteria in tidal flat sediment, characterized by stable isotope probing-based metagenomic analysis. *Environmental Microbiology*, 16(1):189–204. 95

Kim, Y.M., Poline, J.B. and Dumas, G. (2017). Experimenting with reproducibility in bioinformatics. *bioRxiv*, page 143503. 47

Knight, R. et al (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, 16(July):1–13. 41, 42, 43, 124, 127

Kohl, L., Cumming, E., Cox, A., Rietze, A. and Morrissey, L. (2016). Exploring the metabolic potential of microbial communities in ultra-basic, reducing springs at The Cedars, CA, USA: Experimental evidence of microbial methanogenesis and heterotrophic acetogenesis. *Journal of Geophysical Research: Biogeosciences*, 121(4):1203–1220. 95, 119

Kolowith, L.C., Ingall, E.D. and Benner, R. (2001). Composition and cycling of marine organic phosphorus. *Limnology and Oceanography*, 46(2):309–320. 131

Konn, C., Charlou, J.L., Donval, J.P., Holm, N.G., Dehairs, F. and Bouillon, S. (2009). Hydrocarbons and oxidized organic compounds in hydrothermal fluids from Rainbow and Lost City ultramafic-hosted vents. *Chemical Geology*, 258(3-4):299–314. 19, 94

Konopka, A. (2009). What is microbial community ecology. *ISME Journal*, 3(11):1223–1230. 36

Konstantinidis, K.T. and Tiedje, J.M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences*, 102(7):2567–2572. 40

Koonin, E.V. and Martin, W. (2005). On the origin of genomes and cells within inorganic compartments. *Trends in Genetics*, 21(12):647–654. 15

Köster, J. and Rahmann, S. (2012). Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522. 47, 79

Kubo, K., Lloyd, K.G., F Biddle, J., Amann, R., Teske, A. and Knittel, K. (2012). Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *The ISME Journal*, 6(10):1949–1965. 64

Kumagai, H., Nakamura, K., Toki, T., Morishita, T., Okino, K., Ishibashi, J.I., Tsunogai, U., Kawagucci, S., Gamo, T., Shibuya, T., Sawaguchi, T., Neo, N., Joshima, M., Sato, T. and Takai, K. (2008). Geological background of the Kairei and Edmond hydrothermal fields along the Central Indian Ridge: Implications of their vent fluids'distinct chemistry. *Geofluids*, 8(4):239–251. 9

Labrenz, M., Lawson, P.a., Tindall, B.J. and Hirsch, P. (2009). *Roseibaca ekhonensis* gen. nov., sp. nov., an alkalitolerant and aerobic bacteriochlorophyll a-producing alphaproteobacterium from hypersaline Ekho Lake. *International Journal of Systematic and Evolutionary Microbiology*, 59(8):1935–1940. 67

Lang, S.Q., Butterfield, D.A., Schulte, M., Kelley, D.S. and Lilley, M.D. (2010). Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City Hydrothermal Field. *Geochimica et Cosmochimica Acta*, 74(3):941–952. 9, 30, 77

Lang, S.Q., Früh-Green, G.L., Bernasconi, S.M., Brazelton, W.J., Schrenk, M.O. and McGonigle, J.M. (2018). Deeply-sourced formate fuels sulfate reducers but not methanogens at Lost City Hydrothermal Field. *Scientific Reports*, 8(1):1–10. 28, 78, 81, 126

Lang, S.Q., Früh-Green, G.L., Bernasconi, S.M. and Butterfield, D.A. (2013). Sources of organic nitrogen at the serpentinite-hosted Lost City Hydrothermal Field. *Geobiology*, 11(2):154–169. 32

Lang, S.Q., Früh-Green, G.L., Bernasconi, S.M., Lilley, M.D., Proskurowski, G., Méhay, S. and Butterfield, D.A. (2012). Microbial utilization of abiogenic carbon and hydrogen in a serpentinite-hosted system. *Geochimica et Cosmochimica Acta*, 92:82–99. 30

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G. and Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 31(9):814–821. 42, 124

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760. 79, 103

Li, W. and Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659. 103

Liaw, A. and Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2:18–22. 80

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012. 36, 37

Liu, X., Li, M., Castelle, C.J., Probst, A.J., Zhou, Z., Pan, J., Liu, Y., Banfield, J.F. and Gu, J.D. (2018). Insights into the ecology, evolution, and metabolism of the widespread Woesearchaeotal lineages. *Microbiome*, 6(1):1–16. 119

Lloyd, K.G., Lapham, L. and Teske, A. (2006). An anaerobic methane-oxidizing community of ANME-1b archaea in hypersaline gulf of Mexico sediments. *Applied and Environmental Microbiology*, 72(11):7218–7230. 27

Louca, S., Polz, M.F., Mazel, F., Albright, M.B.N., Huber, J.A., O'Connor, M.I., Ackermann, M., Hahn, A.S., Srivastava, D.S., Crowe, S.A., Doebeli, M. and Parfrey, L.W. (2018). Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*. 78

Lozupone, C. and Knight, R. (2005). UniFrac : a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*, 71(12):8228–8235. 58

Lüke, C., Speth, D.R., Kox, M.A., Villanueva, L. and Jetten, M.S. (2016). Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone. *PeerJ*, 4:e1924. 32

Luton, P.E., Wayne, J.M., Sharp, R.J. and Riley, P.W. (2002). The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology*, 148(11):3521–3530. 27

Mahé, F., Rognes, T., Quince, C., de Vargas, C. and Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ*, 2:e593. 41

Mahé, F., Rognes, T., Quince, C., de Vargas, C. and Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3:e1420. 41

Maillet, N., Collet, G., Vannier, T., Lavenier, D. and Peterlongo, P. (2014). Commet: Comparing and combining multiple metagenomic datasets. *Proceedings - 2014 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2014*, pages 94–98. 125

Marbouty, M., Cournac, A., Flot, J.F., Marie-Nelly, H., Mozziconacci, J. and Koszul, R. (2014). Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *eLife*, 3:e03318. 125

Mardanov, A.V., Kadnikov, V.V. and Ravin, N.V. (2018). Chapter 1 - Metagenomics: A Paradigm Shift in Microbiology. In Nagarajan, M., editor, *Metagenomics*, pages 1–13. Academic Press. 42

Mardis, E.R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, 470(7333):198–203. 37

Margulies, M. et al (2005). Genome sequencing in open microfabricated high density picoliter reactors. *Nature biotechnology*, 437(7057):376–380. 38

Marietou, A. and Bartlett, D.H. (2014). Effects of high hydrostatic pressure on coastal bacterial community abundance and diversity. *Applied and Environmental Microbiology*, 80(19):5992–6003. 17

Marques, J.M., Carreira, P.M., Carvalho, M.R., Matias, M.J., Goff, F.E., Basto, M.J., Graça, R.C., Aires-Barros, L. and Rocha, L. (2008). Origins of high pH mineral waters from ultramafic rocks, Central Portugal. *Applied Geochemistry*, 23(12):3278–3289. 12

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10. 103

Martin, W., Baross, J., Kelley, D. and Russell, M.J. (2008). Hydrothermal vents and the origin of life. *Nature Reviews*, 6(11):805–14. 9, 77, 101

Martin, W. and Russell, M.J. (2007). On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1486):1887–1925. 16

Martin, W., Russell, M.J., Horner, D., Blankenship, R., Cavalier-Smith, T. and Nisbet, E. (2003). On the origins of cells: A hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1429):59–85. 15

Martinez, A., Tyson, G.W. and Delong, E.F. (2010). Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environmental Microbiology*, 12(1):222–238. 33

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G. and Neufeld, J.D. (2012). PANDAseq: paired-end assembler for illumina sequences. *BMC bioinformatics*, 13(1):31. 79

Matsen, F.A., Kodner, R.B. and Armbrust, E.V. (2010). Pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1):538. 93

Maurizot, P. and Vendé-Leclerc, M. (2009). *Carte géologique de la Nouvelle-Calédonie au 1/500 000*. Direction de l'Industrie, des Mines et de l'Energie, Service de la Géologie de Nouvelle-Calédonie, Bureau de Recherches Géologiques et Minières, Nouméa. 14

McCollom, T.M. (2016). Abiotic methane formation during experimental serpentinization of olivine. *Proceedings of the National Academy of Sciences*, 113(49):13965–13970. 6, 7

Mccollom, T.M. and Seewald, J.S. (2007). Abiotic synthesis of organic compounds in deep-sea hydrothermal environments. *Nature Geoscience*, pages 382–401. 6, 16

McCollom, T.M. and Seewald, J.S. (2013). Serpentinites, hydrogen, and life. *Elements*, 9(2):129–134. 5, 16, 18, 24

McGrath, J.W., Chin, J.P. and Quinn, J.P. (2013). Organophosphonates revealed: New insights into the microbial metabolism of ancient molecules. *Nature Reviews Microbiology*, 11(6):412–419. 97

McMurdie, P.J. and Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8(4). 58, 80

McSorley, F.R., Wyatt, P.B., Martinez, A., Delong, E.F., Hove-Jensen, B. and Zechel, D.L. (2012). PhnY and PhnZ comprise a new oxidative pathway for enzymatic cleavage of a carbon-phosphorus bond. *Journal of the American Chemical Society*, 134(20):8364–8367. 96

Mei, N., Postec, A., Erauso, G., Joseph, M., Pelletier, B., Payri, C., Ollivier, B. and Quéméneur, M. (2016a). Serpentinicella alkaliphila gen. nov., sp. nov., a novel alkaliphilic anaerobic bacterium isolated from the serpentinite-hosted Prony Hydrothermal Field, New Caledonia. *International Journal of Systematic and Evolutionary Microbiology*, 66(11):4464–4470. 17, 30, 48, 49, 101, 119

Mei, N., Postec, A., Monnin, C., Pelletier, B., Payri, C.E., Ménez, B., Frouin, E., Ollivier, B., Erauso, G. and Quéméneur, M. (2016b). Metagenomic and PCR-based diversity surveys of [FeFe]-hydrogenases combined with isolation of alkaliphilic hydrogen-producing bacteria from the serpentinite-hosted Prony Hydrothermal Field, New Caledonia. *Frontiers in Microbiology*, 7(August). 14, 19, 26, 49, 50, 56, 63, 67, 78, 81, 86, 102, 120, 129

Mei, N., Zergane, N., Postec, A., Erauso, G., Ollier, A., Payri, C., Pelletier, B., Fardeau, M.L., Ollivier, B. and Quéméneur, M. (2014). Fermentative hydrogen production by a new alkaliphilic *Clostridium* sp. (strain PROH2) isolated from a shallow submarine hydrothermal chimney in Prony Bay, New Caledonia. *International Journal of Hydrogen Energy*, 39(34):19465–19473. 17, 49, 55, 101, 126

Mendoza, M.L.Z., Sicheritz-Pontén, T. and Thomas Gilbert, M.P. (2014). Environmental genes and genomes: Understanding the differences and challenges in the approaches and software for their analyses. *Briefings in Bioinformatics*, 16(5):745–758. 39

Metzker, M.L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, 11(1):31–46. 37, 38, 39

Meyer-Dombard, D.R., Woycheese, K.M., Yargıço\u{g}lu, E.N., Cardace, D., Shock, E.L., Güleçal-Pektas, Y. and Temel, M. (2015). High pH microbial ecosystems in a newly discovered, ephemeral, serpentinizing fluid seep at Yanartas (Chimera), Turkey. *Frontiers in Microbiology*, 6(JAN):1–13. 12, 32

Monnin, C., Chavagnac, V., Boulart, C., Ménez, B., Gérard, M., Gérard, E., Pisapia, C., Quéméneur, M., Erauso, G., Postec, A., Guentas-Dombrowski, L., Payri, C. and Pelletier, B. (2014). Fluid chemistry of the low temperature hyperalkaline hydrothermal system of Prony Bay (New Caledonia). *Biogeosciences*, 11(20):5687–5706. 13, 14, 55, 56, 59, 66, 77, 101

Morrill, P.L., Brazelton, W.J., Kohl, L., Rietze, A., Miles, S.M., Kavanagh, H., Schrenk, M.O., Ziegler, S.E. and Lang, S.Q. (2014). Investigations of potential microbial methanogenic and carbon monoxide utilization pathways in ultra-basic reducing springs associated with present-day continental serpentinization: The Tablelands, NL, CAN. *Frontiers in Microbiology*, 5(NOV):1–13. 12, 17

Morrill, P.L., Kuenen, J.G., Johnson, O.J., Suzuki, S., Rietze, A., Sessions, A.L., Fogel, M.L. and Nealson, K.H. (2013). Geochemistry and geobiology of a present-day serpentinization site in California: The Cedars. *Geochimica et Cosmochimica Acta*, 109:222–240. 7, 78, 131

Mottl, M.J., Komor, S.C., Fryer, P. and Moyer, C.L. (2003). Deep-slab fluids fuel extremophilic Archaea on a Mariana forearc serpentinite mud volcano: Ocean drilling program leg 195. *Geochemistry, Geophysics, Geosystems*, 4(11):1–14. 9, 10

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D.J., Salichos, L., Zhang, J., Weinstock, G.M., Isaacs, F., Rozowsky, J. and Gerstein, M. (2016). The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biology*, 17(1):1–9. 45

Müller, A.L., Kjeldsen, K.U., Rattei, T., Pester, M. and Loy, A. (2015). Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *The ISME Journal*, 9(5):1152–1165. 31

Nakasugi, K., Svenson, C.J. and Neilan, B.A. (2006). The competence gene, *comF*, from *Synechocystis* sp. strain PCC 6803 is involved in natural transformation, phototactic motility and piliation. *Microbiology*, 152(12):3623–3631. 88

Narasingarao, P., Podell, S., Ugalde, J.A., Brochier-Armanet, C., Emerson, J.B., Brocks, J.J., Heidelberg, K.B., Banfield, J.F. and Allen, E.E. (2012). De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *The ISME Journal*, 6(1):81–93. 48

Nelson, W.C. and Stegen, J.C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Frontiers in Microbiology*, 6:713. 22, 73

Nercessian, O., Fouquet, Y., Pierre, C., Prieur, D. and Jeanthon, C. (2005). Diversity of Bacteria and Archaea associated with a carbonate-rich metalliferous sediment sample from the Rainbow vent field on the Mid-Atlantic Ridge. *Environmental Microbiology*, 7(5):698–714. 49

Neubeck, A., Duc, N.T., Bastviken, D., Crill, P. and Holm, N.G. (2011). Formation of H_2 and CH_4 by weathering of olivine at temperatures between 30 and 70°C. *Geochemical Transactions*, 12(1):6. 6

Neubeck, A., Sun, L., Müller, B., Ivarsson, M., Hosgörmez, H., Özcan, D., Broman, C. and Schnürer, A. (2017). Microbial community structure of a serpentine-hosted abiotic gas seepage at the Chimaera ophiolite, Turkey. *Applied and environmental microbiology*, 83(12). 20, 49, 55

Ohara, Y., Reagan, M.K., Fujikura, K., Watanabe, H., Michibayashi, K., Ishii, T., Stern, R.J., Pujana, I., Martinez, F., Girard, G., Ribeiro, J., Brounce, M., Komori, N. and Kino, M. (2012). A serpentinite-hosted ecosystem in the Southern Mariana Forearc. *Proceedings of the National Academy of Sciences*, 109(8):2831–5. 10

Okland, I., Huang, S., Dahle, H., Thorseth, I.H. and Pedersen, R.B. (2012). Low temperature alteration of serpentinized ultramafic rock and implications for microbial life. *Chemical Geology*, 318-319:75–87. 12

Okumura, T., Ohara, Y., Stern, R.J., Yamanaka, T., Onishi, Y., Watanabe, H., Chen, C., Bloomer, S.H., Pujana, I., Sakai, S., Ishii, T. and Takai, K. (2016). Brucite chimney formation and carbonate alteration at the Shinkai Seep Field, a serpentinite-hosted vent systemin the southern Mariana forearc. *Geochemistry, Geophysics, Geosystems*, 17:3775–3796. 10

Oulas, A., Polymenakou, P.N., Seshadri, R., Tripp, H.J., Mandalakis, M., Paez-Espino, A.D., Pati, A., Chain, P., Nomikou, P., Carey, S., Kilias, S., Christakis, C., Kotoulas, G., Magoulas, A., Ivanova, N.N. and Kyrpides, N.C. (2016). Metagenomic investigation of the geologically unique Hellenic Volcanic Arc reveals a distinctive ecosystem with unexpected physiology. *Environmental Microbiology*, 18(4):1122–1136. 50

Palandri, J.L. and Reed, M.H. (2004). Geochemical models of metasomatism in ultramafic systems: Serpentinization, rodingitization, and sea floor carbonate chimney precipitation. *Geochimica et Cosmochimica Acta*, 68(5):1115–1133. 7

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A. and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*. 124

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W. (2015). CheckM : assessing the quality of microbial genomes recovered from isolates , single cells , and metagenomes. *Genome Research*, 25(7):1043–1055. 44, 104

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P. and Tyson, G.W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2(November):1. 45

Pelletier, B., Chevillon, C., Menou, J., Butscher, J., Folcher, E., Geoffray, C., Boré, J.M., Panché, J.Y. and Perrier, J. (2006). Plongées, forage et cartographie Baie du Prony et Banc Gail, lagon Sud de Nouvelle-Calédonie, campagne 2005-NC-PL du N.O. ALIS 13-17 Juin 2005 et cartographie baie du Prony et canal Woodin N.O. ALIS. *Rapports de Missions Sciences de la Terre Géologie Géophysique (IRD Noumea)*, 70:1–44. 55, 77

Peng, Y., Leung, H.C.M., Yiu, S.M. and Chin, F.Y.L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11):1420–8. 79, 103

Pérez Bernal, M.F., Brito, E.M.S., Bartoli, M., Aubé, J., Fardeau, M.L., Rodriguez, G.C., Ollivier, B., Guyoneaud, R. and Hirschler-Réa, A. (2017). *Desulfonatronum parangueonense* sp. nov., a sulfate-reducing bacterium isolated from sediment of an alkaline crater lake. *International Journal of Systematic and Evolutionary Microbiology*, 67(12):4999–5005. 31

Perner, M., Kuever, J., Seifert, R., Pape, T., Koschinsky, A., Schmidt, K., Strauss, H. and Imhoff, J.F. (2007). The influence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15°N on the Mid-Atlantic Ridge. *FEMS Microbiology Ecology*, 61(1):97–109. 49

Perner, M., Petersen, J.M., Zielinski, F., Gennerich, H.H. and Seifert, R. (2010). Geochemical constraints on the diversity and activity of H_2-oxidizing microorganisms in diffuse hydrothermal fluids from a basalt- and an ultramafic-hosted vent. *FEMS Microbiology Ecology*, 74(1):55–71. 49

Peters, J.W., Schut, G.J., Boyd, E.S., Mulder, D.W., Shepard, E.M., Broderick, J.B., King, P.W. and Adams, M.W.W. (2015). [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1853(6):1350–1369. 26, 85

Pisapia, C., Gérard, E., Gérard, M., Lecourt, L., Lang, S.Q., Pelletier, B., Payri, C.E., Monnin, C., Guentas, L., Postec, A., Quéméneur, M., Erauso, G. and Ménez, B. (2017). Mineralizing filamentous bacteria from the Prony Bay Hydrothermal Field give new insights into the functioning of serpentinization-based subseafloor ecosystems. *Frontiers in Microbiology*, 8. 13, 19, 20, 23, 49, 55, 56, 63, 66, 67, 69, 71, 72, 78, 119, 126, 129, 130

Pisapia, C., Jamme, F., Duponchel, L. and Ménez, B. (2018). Tracking hidden organic carbon in rocks using chemometrics and hyperspectral imaging. *Scientific Reports*, 8(1):1–14. 78

Pohorelic, B.K.J., Voordouw, J.K., Dolla, A., Harder, J., Voordouw, G. and Lojou, E. (2002). Effects of deletion of genes encoding Fe-only hydrogenase of *Desulfovibrio vulgaris* Hildenborough on hydrogen and lactate metabolism. *Journal of Bacteriology*, 184(3):679–686. 25

Poser, A., Lohmayer, R., Vogt, C., Knoeller, K., Planer-Friedrich, B., Sorokin, D., Richnow, H.H. and Finster, K. (2013). Disproportionation of elemental sulfur by haloalkaliphilic bacteria from soda lakes. *Extremophiles*, 17(6):1003–1012. 31, 70

Postec, A., Quéméneur, M., Bes, M., Mei, N., Benaïssa, F., Payri, C., Pelletier, B., Monnin, C., Guentas-Dombrowsky, L., Ollivier, B., Gérard, E., Pisapia, C., Gérard, M., Ménez, B. and Erauso, G. (2015). Microbial diversity in a submarine carbonate edifice from the serpentinizing hydrothermal system of the Prony Bay (New Caledonia) over a 6-year period. *Frontiers in Microbiology*, 6(AUG):1–19. 13, 18, 20, 22, 23, 31, 49, 55, 56, 57, 63, 64, 65, 69, 71, 72, 77, 101, 123, 125, 130

Prakash, T. and Taylor, T.D. (2012). Functional assignment of metagenomic data: Challenges and applications. *Briefings in Bioinformatics*, 13(6):711–727. 124

Preiss, L., Hicks, D.B., Suzuki, S., Meier, T. and Krulwich, T.A. (2015). Alkaliphilic bacteria with impact on industrial applications, concepts of early life forms, and bioenergetics of ATP synthesis. *Frontiers in bioengineering and biotechnology*, 3(June):75. 17

Price, M.N., Dehal, P.S. and Arkin, A.P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, 5(3). 104

Proskurowski, G., Lilley, M.D., Seewald, J.S., Früh-Green, G.L., Olson, E.J., Lupton, J.E., Sylva, S.P. and Kelley, D.S. (2008). Abiogenic hydrocarbon production at Lost City Hydrothermal Field. *Science*, 319(5863):604–607. 6

Prosser, J.I. (2010). Replicate or lie. *Environmental Microbiology*, 12(7):1806–1810. 129

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1):590–596. 58, 103

Quéméneur, M., Bes, M., Postec, A., Mei, N., Hamelin, J., Monnin, C., Chavagnac, V., Payri, C., Pelletier, B., Guentas-Dombrowsky, L., Gérard, M., Pisapia, C., Gérard, E., Ménez, B., Ollivier, B. and Erauso, G. (2014). Spatial distribution of microbial communities in the shallow submarine alkaline hydrothermal field of the Prony Bay, New Caledonia. *Environmental Microbiology Reports*, 6:665–674. 13, 19, 20, 21, 22, 23, 27, 49, 55, 56, 63, 65, 68, 69, 70, 71, 72, 77, 86, 94, 101, 102, 125, 127

Quéméneur, M., Palvadeau, A., Postec, A., Monnin, C., Chavagnac, V., Ollivier, B. and Erauso, G. (2015). Endolithic microbial communities in carbonate precipitates from serpentinite-hosted hyperalkaline springs of the Voltri Massif (Ligurian Alps, Northern Italy). *Environmental Science and Pollution Research*, 22(18):13613–13624. 20, 21, 23, 49, 55, 58, 68, 69, 70, 71, 72, 120

Quesnel, B., Gautier, P., Cathelineau, M., Boulvais, P., Couteau, C. and Drouillet, M. (2016). The internal deformation of the Peridotite Nappe of New Caledonia: A structural study of serpentine-bearing faults and shear zones in the Koniambo Massif. *Journal of Structural Geology*, 85:51–67. 77

Quince, C., Lanzen, A., Davenport, R.J. and Turnbaugh, P.J. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, 12(36). 41

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842. 80

Quinn, J.P., Kulakova, A.N., Cooley, N.A. and McGrath, J.W. (2007). New ways to break an old bond: The bacterial carbon-phosphorus hydrolases and their role in biogeochemical phosphorus cycling. *Environmental Microbiology*, 9(10):2392–2400. 96

Quéméneur, M., Erauso, G., Frouin, E., Zeghal, E., Vandecasteele, C., Ollivier, B., Tamburini, C., Garel, M., Ménez, B. and Postec, A. (submitted). Hydrostatic pressure helps to cultivate original and anaerobic bacteria from the Atlantis Massif subseafloor. 17

Rempfert, K.R., Miller, H.M., Bompard, N., Nothaft, D., Matte, J.M., Kelemen, P., Fierer, N. and Templeton, A.S. (2017). Geological and geochemical controls on subsurface microbial life in the Samail Ophiolite, Oman. *Frontiers in Microbiology*, 8(56). 12, 23, 49, 55, 70

Repeta, D.J., Ferrón, S., Sosa, O.A., Johnson, C.G., Repeta, L.D., Acker, M., DeLong, E.F. and Karl, D.M. (2016). Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nature Geoscience*, 9(12):884–887. 79, 97, 131

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J.L., Sylva, S., Seewald, J., German, C.R. and Huber, J.A. (2016). Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environmental Microbiology*, 18(6):1970–1987. 24, 25, 30, 81, 94, 131

Rinke, C. et al (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459):431–437. 22

Rizoulis, A., Milodowski, A.E., Morris, K. and Lloyd, J.R. (2014). Bacterial diversity in the hyperalkaline Allas Springs (Cyprus), a natural analogue for cementitious radioactive waste repository. *Geomicrobiology Journal*, pages 73–84. 12, 23, 49

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140. 43, 80

Robinson, M.D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 11(3):R25. 80

Rodriguez-R, L.M. and Konstantinidis, K.T. (2016). The enveomics collection : a toolbox for specialized analyses of microbial genomes and metagenomes. *Peer J Preprints*, 4:e1900v1. 104

Rosen, M.J., Callahan, B.J., Fisher, D.S. and Holmes, S.P. (2012). Denoising PCR-amplified metagenome data. *BMC Bioinformatics*, 13(1). 41

Roussel, E.G., Konn, C., Charlou, J.L., Donval, J.P., Fouquet, Y., Querellou, J., Prieur, D. and Cambon Bonavita, M.A. (2011). Comparison of microbial communities associated with three Atlantic ultramafic hydrothermal systems. *FEMS Microbiology Ecology*, 77(3):647–665. 27, 49

Russell, M.J., Hall, A.J. and Martin, W. (2010). Serpentinization as a source of energy at the origin of life. *Geobiology*, 8(5):355–371. 55, 72

Sánchez-Murillo, R., Gazel, E., Schwarzenbach, E.M., Crespo-Medina, M., Schrenk, M.O., Boll, J. and Gill, B.C. (2014). Geochemical evidence for active tropical serpentinization in the Santa Elena Ophiolite, Costa Rica: An analog of a humid early Earth? *Geochemistry, Geophysics, Geosystems*, 15(5):1783–1800. 11, 19, 21, 49, 55

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467. 36

Sangwan, N., Xia, F. and Gilbert, J.A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4(1):8. 44, 45

Saxena, R., Dhakan, D.B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A. and Sharma, V.K. (2017). Metagenomic analysis of hot springs in central India reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Frontiers in Microbiology*, 7(January). 81, 83

Schmidt, T.M., Delong, E.F. and Pace, N.R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology*, 173(14):4371–4378. 39

Schowanek, D. and Verstraete, W. (1990). Phosphonate utilization by bacterial cultures and enrichments from environmental samples. *Applied and Environmental Microbiology*, 56(4):895–903. 96

Schrenk, M.O., Brazelton, W.J. and Lang, S.Q. (2013). Serpentinization, carbon, and deep life. *Reviews in Mineralogy & Geochemistry*, 75:575–606. 1, 7, 8, 9, 11, 12, 16, 17, 24, 26, 30, 55, 68, 73, 77, 78, 101, 126

Schrenk, M.O., Kelley, D.S., Bolton, S.A. and Baross, J.A. (2004). Low archaeal diversity linked to subseafloor geochemical processes at the Lost City Hydrothermal Field, Mid-Atlantic Ridge. *Environmental Microbiology*, 6(10):1086–1095. 21, 49, 58, 63, 69, 70, 71, 72, 77, 94

Schulte, M., Blake, D., Hoehler, T. and McCollom, T.M. (2006). Serpentinization and its implications for life on the early Earth and Mars. *ASTROBIOLOGY*, 6(2):364–376. 101

Schwarzenbach, E.M., Gill, B.C., Gazel, E. and Madrigal, P. (2016). Sulfur and carbon geochemistry of the Santa Elena peridotites: Comparing oceanic and continental processes during peridotite alteration. *Lithos*, 252-253:92–108. 11

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069. 82, 104

Segata, N., Börnigen, D., Morgan, X.C. and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications*, 4. 104

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. and Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6):R60. 130

Shannon, C.E. and Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press. 58

Simpson, E.H. (1949). Measurement of Diversity. *Nature*, 163(4148):688–688. 58

Sleep, N.H., Bird, D.K. and Pope, E.C. (2011). Serpentinite and the dawn of life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1580):2857–2869. 16

Sobieraj, M., and Boone, D.R. (2006). Syntrophomonadaceae. In Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H. and Stackebrandt, E., editors, *The Prokaryotes: Bacteria: Firmicutes, Cyanobacteria, Vol. 4*, pages 1041–1049. Springer, New York, NY. 67

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the USA*, 103(32):12115–12120. 23

Sorokin, D.Y., Abbas, B., Geleijnse, M., Kolganova, T.V. and Loosdrecht, M.C.M.V. (2016). Syntrophic associations from hypersaline soda lakes converting organic acids and alcohols to methane at extremely haloalkaline conditions. *Environmental Microbiology*, 18(9):3189–3202. 68

Sorokin, D.Y., Berben, T., Melton, E.D., Overmars, L., Vavourakis, C.D. and Muyzer, G. (2014). Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles*, 18(5):791–809. 16

Sorokin, D.Y., Tourova, T.P., Mußmann, M. and Muyzer, G. (2008). *Dethiobacter alkaliphilus* gen. nov. sp. nov., and *Desulfurivibrio alkaliphilus* gen. nov. sp. nov.: Two novel representatives of reductive sulfur cycle from soda lakes. *Extremophiles*, 12(3):431–439. 70

Sosa, O.A. and Delong, E.F. (2017). Isolation and characterization of bacteria that degrade phosphonates in marine dissolved organic matter. *Frontiers in Microbiology*, 8(September). 79

Sousa, F.L., Thiergart, T., Landan, G., Nelson-Sathi, S., Pereira, I.a.C., Allen, J.F., Lane, N. and Martin, W.F. (2013). Early bioenergetic evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1622):20130088. 15, 16

Steinegger, M. and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028. 46

Stephens, Z.D., Lee, S.Y., Faghri, F., Campbell, R.H., Zhai, C., Efron, M.J., Iyer, R., Schatz, M.C., Sinha, S. and Robinson, G.E. (2015). Big data: Astronomical or genomical? *PLoS Biology*, 13(7):1–11. 45, 46

Suzuki, S., Gijs Kuenen, J., Schipper, K., Van Der Velde, S., Ishii, S., Wu, A., Sorokin, D.Y., Tenney, A., Meng, X., Morrill, P.L., Kamagata, Y., Muyzer, G. and Nealson, K.H. (2014). Physiological and genomic features of highly alkaliphilic hydrogen-utilizing *Betaproteobacteria* from a continental serpentinizing site. *Nature Communications*, 5. 17, 26, 29, 49, 70, 78, 94, 102, 115, 120

Suzuki, S., Ishii, S., Hoshino, T., Rietze, A., Tenney, A., Morrill, P.L., Inagaki, F., Kuenen, J.G. and Nealson, K.H. (2017). Unusual metabolic diversity of hyperalkaliphilic microbial communities associated with subterranean serpentinization at The Cedars. *The ISME Journal*, pages 1–15. 22, 49, 50, 55, 73, 102, 119, 130, 132

Suzuki, S., Ishii, S., Wu, A., Cheung, A., Tenney, A., Wanger, G., Gijs Kuenen, J., Nealson, K.H. and Canfield, D.E. (2013). Microbial diversity in The Cedars, an ultrabasic, ultrareducing, and low salinity serpentinizing ecosystem. *PNAS*, 110:15336–15341. 19, 21, 22, 23, 49, 50, 51, 58, 63, 66, 68, 69, 70, 71, 72, 73, 77, 127

Sviridov, A.V., Shushkova, T.V., Zelenkova, N.F., Vinokurova, N.G., Morgunov, I.G., Ermakova, I.T. and Leontievsky, A.A. (2012). Distribution of glyphosate and methylphosphonate catabolism systems in soil bacteria *Ochrobactrum anthropi* and *Achromobacter* sp. *Applied Microbiology and Biotechnology*, 93(2):787–796. 96

Szponar, N., Brazelton, W.J., Schrenk, M.O., Bower, D.M., Steele, A. and Morrill, P.L. (2013). Geochemistry of a continental site of serpentinization, the Tablelands Ophiolite, Gros Morne National Park: A Mars analogue. *Icarus*, 224(2):286–296. 12

Taberlet, P., Coissac, E., Hajibabaei, M. and Rieseberg, L.H. (2012). Environmental DNA. *Molecular Ecology*, 21(8):1789–1793. 39

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T. and Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PLoS ONE*, 9(8). 103, 123

Takai, K., Gamo, T., Tsunogai, U., Nakayama, N., Hirayama, H., Nealson, K.H. and Horikoshi, K. (2004). Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field. *Extremophiles*, 8(4):269–282. 9, 49

Takai, K., Moyer, C.L., Miyazaki, M., Nogi, Y., Hirayama, H., Nealson, K.H. and Horikoshi, K. (2005). *Marinobacter alkaliphilus* sp. nov., a novel alkaliphilic bacterium isolated from subseafloor alkaline serpentine mud from Ocean Drilling Program Site 1200 at South Chamorro Seamount, Mariana Forearc. *Extremophiles*, 9(1):17–27. 49

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Smirnov, S., Nikolskaya, A.N., Rao, B.S., Mekhedov, S.L., Sverlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J. and Natale, D.A. (2003). The COG database: An updated vesion includes eukaryotes. *BMC Bioinformatics*, 4:1–14. 80

Thomas, T., Gilbert, J. and Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1):3. 42, 129

Tiago, I., Chung, A.P. and Veríssimo, A. (2004). Bacterial diversity in a nonsaline alkaline environment: Heterotrophic aerobic populations. *Applied and Environmental Microbiology*, 70(12):7378–7387. 16, 30, 48, 49

Tiago, I., Mendes, V., Pires, C., Morais, P.V. and Veríssimo, A. (2005a). Phenylobacterium falsum sp. nov., an Alphaproteobacterium isolated from a nonsaline alkaline groundwater, and emended description of the genus *Phenylobacterium*. *Systematic and Applied Microbiology*, 28(4):295–302. 49

Tiago, I., Mendes, V., Pires, C., Morais, P.V. and Veríssimo, A. (2006a). *Chimaereicella alkaliphila* gen. nov., sp. nov., a Gram-negative alkaliphilic bacterium isolated from a nonsaline alkaline groundwater. *Systematic and Applied Microbiology*, 29(2):100–108. 30, 49

Tiago, I., Morais, P.V., da Costa, M.S. and Veríssimo, A. (2006b). *Microcella alkaliphila* sp. nov., a novel member of the family *Microbacteriaceae* isolated from a nonsaline alkaline groundwater, and emended description of the genus *Microcella*. *International Journal of Systematic and Evolutionary Microbiology*, 56(10):2313–2316. 49

Tiago, I., Pires, C., Mendes, V., Morais, P.V., Da Costa, M. and Veríssimo, A. (2005b). *Microcella putealis* gen. nov., sp. nov., a Gram-positive alkaliphilic bacterium isolated from a nonsaline alkaline groundwater. *Systematic and Applied Microbiology*, 28(6):479–487. 30, 49

Tiago, I., Pires, C., Mendes, V., Morais, P.V., da Costa, M.S. and Veríssimo, A. (2006c). *Bacillus foraminis* sp. nov., isolated from a non-saline alkaline groundwater. *International Journal of Systematic and Evolutionary Microbiology*, 56(11):2571–2574. 49

Tiago, I. and Veríssimo, A. (2013). Microbial and functional diversity of a subterrestrial high pH groundwater associated to serpentinization. *Environmental Microbiology*, 15(6):1687–1706. 21, 27, 49, 55, 58, 68, 69, 71, 72, 120

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-liggett, C., Knight, R. and Gordon, J.I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature*, 449(7164):804–810. 40

Twing, K.I., Brazelton, W.J., Kubo, M.D.Y., Hyer, A.J., Cardace, D., Hoehler, T., McCollom, T.M. and Schrenk, M.O. (2017). Serpentinization-influenced groundwater harbors extremely low diversity microbial communities adapted to high pH. *Frontiers in Microbiology*, 8(March):308. 21, 23, 26, 31, 49, 50, 51, 55, 59, 69, 70, 78, 81, 94, 126

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. and Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978):37–43. 39

Urich, T., Lanzén, A., Stokke, R., Pedersen, R.B., Bayer, C., Thorseth, I.H., Schleper, C., Steen, I.H. and Øvreas, L. (2014). Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics. *Environmental Microbiology*, 16(9):2699–2710. 94

van der Walt, A.J., van Goethem, M.W., Ramond, J.B., Makhalanyane, T.P., Reva, O. and Cowan, D.A. (2017). Assembling metagenomes, one community at a time. *BMC Genomics*, 18(1):1–13. 42, 46

Van Mooy, B.A.S., Rocap, G., Fredricks, H.F., Evans, C.T. and Devol, A.H. (2006). Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proceedings of the National Academy of Sciences*, 103(23):8607–8612. 33

Vance, S., Harnmeijer, J., Kimura, J., Hussmann, H., Demartin, B. and Brown, J.M. (2007). Hydrothermal systems in small ocean planets. *ASTROBIOLOGY*, 7(6):987–1005. 8

Vartoukian, S.R. (2016). Cultivation strategies for growth of uncultivated bacteria. *Journal of Oral Biosciences*, 58(4):143–149. 50

Venter, J.C. et al (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304(5667):66–74. 39

Vieites, J.M., Guazzaroni, M.E., Beloqui, A., Golyshin, P.N. and Ferrer, M. (2009). Metagenomics approaches in systems microbiology. *FEMS Microbiology Reviews*, 33(1):236–255. 40

Vignais, P.M. (2001). Classification and phylogeny of hydrogenases. *FEMS Microbiology Reviews*, 25. 26, 85, 115

Vignais, P.M. and Billoud, B. (2007). Occurrence, classification, and biological function of hydrogenases: An overview. *Chemical Reviews*, 107(10):4206–4272. 25, 104

Vignais, P.M. and Colbeau, A. (2004). Molecular biology of microbial hydrogenases. *Applied and Environmental Microbiology*, 6:159–188. 25

Villar, E. et al (2015). Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science*, 348(6237):1261447. 125

Walters, W.A., Caporaso, J.G., Lauber, C.L., Berg-Lyons, D., Fierer, N. and Knight, R. (2011). PrimerProspector: De novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics*, 27(8):1159–1161. 123

Wang, D.T. et al (2015). Nonequilibrium clumped isotope signals in microbial methane. *Science*, 348(6233):428–431. 95

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16):5261–5267. 58, 103

Wang, S., Huang, H., Kahnt, J. and Thauer, R.K. (2013). A reversible electron-bifurcating ferredoxin- and NAD-Dependent [FeFe]-Hydrogenase (HydABC) in *Moorella thermoacetica*. *Journal of Bacteriology*, 195(6):1267–1275. 25

Wasmund, K., Schreiber, L., Lloyd, K.G., Petersen, D.G., Schramm, A., Stepanauskas, R., Jørgensen, B.B. and Adrian, L. (2013). Genome sequencing of a single cell of the widely distributed marine subsurface *Dehalococcoidia*, phylum *Chloroflexi*. *The ISME Journal*, 8:383–397. 67

White, A.K. and Metcalf, W.W. (2004). Two C-P lyase operons in *Pseudomonas stutzeri* and their roles in the oxidation of phosphonates, phosphite, and hypophosphite. *Journal of Bacteriology*, 186(14):4730–4739. 90

White, A.K. and Metcalf, W.W. (2007). Microbial metabolism of reduced phosphorus compounds. *Annual Review of Microbiology*, 61(1):379–400. 33, 78, 79, 89

Willems, A. (2014). The Family *Comamonadaceae*. In Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F., editors, *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*, pages 777–851. Springer Berlin Heidelberg, Berlin, Heidelberg. 67

Woese, C.R. and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11):5088–5090. 36

Woycheese, K.M., Meyer-Dombard, D.R., Cardace, D., Argayosa, A.M. and Arcilla, C.A. (2015). Out of the dark: transitional subsurface-to-surface microbial diversity in a terrestrial serpentinizing seep (Manleluag, Pangasinan, the Philippines). *Frontiers in Microbiology*, 6:44. 11, 20, 49, 55, 127

Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S.M., Meng, J., Huang, G., Li, Y., Yan, Q., Wu, S., Wang, X., Chen, S., He, G., Xiao, X. and Xu, A. (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *The ISME Journal*, 5:414–426. 32, 50, 96

Yao, M., Henny, C. and Maresca, J.A. (2016). Freshwater bacteria release methane as a by-product of phosphorus acquisition. *Applied and Environmental Microbiology*, 82(23):6994–7003. 131, 132

Appendix A

Miscellaneous appendices

A.1 Protocols

Extraction de l'ADN d'échantillons environnementaux

A) Phénol-Chloroforme-isoamyl

Matériel nécessaire

- Hotte de chimie
- Centrifugeuse de paillasse (Eppendorf)
- Microtubes Eppendorf de 2 mL. (microtubes 2 mL « Phase Lock Gel™ » en option)
- Tubes Falcon de 15 mL
- Microtubes Eppendof de 1,5 mL
- Micropipettes 20-100 µL, 200-1000 µL
- Cônes jaunes et cônes bleus
- TNE : 100 mM Tris.HCl pH 8.0, 100 mM NaCl, 50 mM EDTA pH 8.0
- TE: 10 mM Tris.HCl, pH 8.0, 1 mM EDTA pH 8.0
- Pour la lyse: 10% SDS, 10% Sarkosyl
- Lysozyme 10 mg/mL dans 10 mM Tris.HCl pH 8.0
- RNase A (Sigma R6513) 5 mg/mL dans du TE à 50% glycérol autoclavé, stockée à -20°C
- Protéinase K : 20 mg/mL dissous dans 10 mM Tris.HCl pH 8.0, 0.5 mM $CaCl_2$
- Mélange PCI phénol/chloroforme/alcool isoamylique (25 :24 :1) BIOSOLVE
- Chloroforme
- Isopropanol
- Ethanol à 75 %

Toutes les manipulations avec du phénol ou du chloroforme doivent être réalisées sous la hotte ventilée et en portant des gants; les déchets correspondants doivent être stockés dans les containers prévus.

1. Transférer 450 µL de suspension de sédiment ou 450 mg de sol dans un tube E (Lysing Matrix).
    i. *Dans le cas d'une culture d'enrichissement, centrifuger entre 5 et 10 mL de culture transférés en tube stérile à usage unique type Falcon 15 ml (8000 rpm , 4°C , 10 min)*
2. Ajouter 800 µL de TNE au tube E
    i. *ou reprendre le culot cellulaire dans 800 µL de TNE et transférer le tout dans un tube lysing matrix E (MP Biomedicals).*
3. Ajouter 100 µL de lysozyme et revisser le bouchon mais sans trop serrer.
4. Mélanger en secouant le tube plusieurs fois pour homogénéiser.
5. Incuber 15 à 20 min à température ambiante
6. Ajouter 100 µL de SDS 10% et 100 µL de Sarkosyl 10%

7. Si nécessaire, enlever un peu du mélange pour que **le niveau du remplissage du tube ne dépasse pas le haut du rainurage du tube**.
8. Positionner les tubes E dans l'appareil FastPrep et visser les supports du couvercle.
9. Mettre en marche pour 40 sec à la vitesse 6.
10. Centrifuger les tubes E à 14 000 X g (11000 rpm centri Eppendorf) pendant 20 min.
11. Transférer le surnageant (~900 µL) dans un tube de 2 mL stérile.
12. Ajouter 50µL de protéinase K et incuber 1h à 55°C.
13. Transférer le lysat dans un microtube de 2 mL « Phase Lock Gel™ »
14. Ajouter 900 µL de PCI et secouant vigoureusement le tube une dizaine de fois (doit former une émulsion visible).
15. Centrifuger les tubes E à 14 000 X g (11000 rpm centri Eppendorf) 4°C pendant 20 min.
16. Récupérer la phase aqueuse (supérieure, incolore) et transférer dans un tube 2 mL
17. Ajouter 900 µL de Chloroforme et mélanger secouant le tube une dizaine de fois.
18. Centrifuger les tubes à 14 000 X g (11000 rpm centri Eppendorf) 4°C pendant 5 min.
19. Récupérer la phase aqueuse (supérieure) et transférer dans un tube 2 mL.
20. Ajouter 700 µL d'isopropanol (précipitation de l'ADN) et 4 µL de DNA carrier
21. Homogénéiser en retournant le tube plusieurs fois
22. Centrifuger les tubes à 14 000 X g (11000 rpm centri Eppendorf) 4°C pendant 20 min
23. Eliminer délicatement l'alcool avec la pipette sous pompe à vide (me demander !)
24. Rincer le culot (probablement invisible) avec 400 µL d'éthanol à 75%.
25. Eliminer délicatement l'alcool avec la pipette sous pompe à vide (me demander !) attention étape délicate : on risque d'aspirer le culot, me demander la démo.
26. Mettre les culots à sécher au speedvac.
27. Resuspendre le culot dans 30 µL de TE (ou d'H_2O SIGMA)

B) Kit Fast DNA SPIN Kit for Soil

Matériel nécessaire

- Homogénéisateur FastPrep
- Centrifugeuse de paillasse (Eppendorf)

- Kit Fast DNA SPIN Kit for Soil (MP Biomedicals)
- Tubes Falcon de 15 mL
- Microtubes Eppendof de 1,5 mL
- Micropipettes 20-100 µL, 200-1000 µL
- Cônes jaunes et cônes bleus

Etapes du protocole :

1. Ajouter 450 µL de suspension de sédiment ou 450 mg de sol dans un tube E (Lysing Matrix).
2. Ajouter 978 µL de Sodium Phosphate buffer au tube E.
3. Ajouter 122 µL de MT Buffer et revisser le bouchon mais sans trop serrer.
4. Mélanger en inversant le tube plusieurs fois pour homogénéiser.
5. Si nécessaire, enlever un peu du mélange pour que le niveau du remplissage du tube ne dépasse pas le haut du rainurage du tube.
6. Positionmer les tubes E dans l'appareil FastPrep et visser les supports du couvercle.
7. Mettre en marche pour 40 sec à la vitesse 6.
8. Centrifuger les tubes E à 14 000 X g (11000 rpm centri Eppndorf) pendant 20 min.
9. Transférer le surnageant dans un tube de 2 mL stérile. Ajouter 250 µL de solution PPS (protein precipitation solution) et mélanger à la main en secouant le tube une dizaine de fois ;
10. Centrifuger les tubes E à 14 000 X g (11000 rpm centri Eppndorf) pendant 20 min.
11. Transférer le surnageant dans un tube stérile Falcon de 15 mL stérile.
12. Resuspendre complètement la Binding Matrix puis en ajouter 1 ml au surnageant, dans le tube de 15 mL.
13. Bien homogénéiser le tout en inversant régulièrement mais doucement, les tubes pour assurer une bonne fixation de l'ADN sur la résine (Binding Matrix) pendant environ 4 min.
14. Placer les tubes verticalement dans un portoir et laisser la suspension sédimenter pendant 5 min ; un culot de résine se forme peu à peu.
15. Enlever environ 800 µL du surnageant et le jeter.
16. Re-mélanger la supension et en transférer 600 µL de suspension à un SPIN Filter et centrifuger à 11 000 rpm pendant 1 min. Vider l'effluat du tube collecteur (catch tube).
17. Rajouter 600 µL de la suspension au même SPIN filter et recommencer l'étape 16 , jusqu'à avoir fait passé toute la suspension au travers du même filtre.
18. Ajouter 500 µL de la soltion de lavage SEWS-M et délicatement remettre la résine en solution en pipettant up and down mais en prenant soin de ne pas toucher la surface du filtre, pour ne pas le percer.
19. Centrifuger les SPIN filters à 11 000 rpm pendant 1 min. Vider l'éluat de lavage du tube collecteur.

20. Sans rajouter de liquide, centrifuger à nouveau à 11 000 rpm pendant 2 min pour « sécher » la résine (matrix) du reste de solution de lavage. Eliminer l'éluat et placer le SPIN filter dans un nouveau tube collecteur (catch tube).
21. Laisser sécher la résine à l'air libre pendant 5 min.
22. Ajouter 60 µL de DES (DNase/Pyrogen-Free Water) à la résine et très délicatement avec le bout d'une pointe jaune avec la micropipette, remettre la résine en suspension dans l'eau. Attention à ne surtout pas percer le filtre !!! Cette étape permet d'éluer l'ADN fixé sur la résine.
23. Placer les tubes dans un bain à sec à 55°C (ce qui améliore le rendement d'élution de l'ADN) pendant 5 min.
24. Centrifuger à 11 000 X rpm pdt 1 min pour récupérer l'ADN élué dans le nouveau tube collecteur.
25. Récupérer l'éluat d'ADN et le transférer dans un microtube de 1,5 mL stérile. Votre ADN est prêt pour une PCR.

A.2 List of publications

Publications

1. **E. Frouin**, M. Bes, B. Ollivier, M. Quéméneur, A. Postec, D. Debroas, F. Armougom, G. Erauso (2018). Diversity of rare and abundant prokaryotic phylotypes in the Prony Hydrothermal Field and comparison with other serpentinite-hosted ecosystems. *Frontiers in Microbiology*. 9:102. doi: 10.3389/fmicb.2018.00102

2. **E. Frouin**, F. Armougom, M. Schrenk, G. Erauso. Comparative metagenomics highlight a pathway involved in the catabolism of phosphonates in multiple serpentinizing ecosystems. Submitted to *The ISME Journal*.

3. G. Pillot, **E. Frouin**, E. Pasero, A. Godfroy, Y. Combet-Blanc, S. Davidson, P.P. Liebgott (2018). Specific enrichment of hyperthermophilic electroactive Archaea from deep-sea hydrothermal vent on electrically conductive support. *Bioresource Technology*. 7:1301. doi:10.1016/j.biortech.2018.03.053

4. N. Mei, A. Postec, C. Monnin, B. Pelletier, C. Payri, B. Ménez, **E. Frouin**, B. Ollivier, G. Erauso, M. Quéméneur (2016). Metagenomic and PCR-Based diversity surveys of [FeFe]-hydrogenases combined with isolation of alkaliphilic hydrogen-producing bacteria from the serpentinite-hosted Prony Hydrothermal Field, New Caledonia. *Frontiers in Microbiology*. 7:1301. doi: 10.3389/fmicb.2016.01301

5. R. Amiraux, J.-F. Rontani, F. Armougom, **E. Frouin**, M. Babin, L. Artigue, P. Bonin. Deciphering biodiversity and stress state of bacteria associated with ice algae and sinking particles in Canadian Artic: impact to carbon flux. Resubmission in progress.

6. M. Quéméneur, G. Erauso, **E. Frouin**, E. Zeghal, C. Vandecasteele, B. Ollivier, C. Tamburini, M. Garel, B. Ménez, A. Postec. Hydrostatic pressure helps to cultivate original and anaerobic bacteria from the Atlantis Massif subseafloor. Submitted to *The ISME Journal*.

Oral communications

1. **E. Frouin**, F. Armougom, G. Erauso. *Comparative metagenomics highlighted a pathway involved in the catabolism of phosphonates in multiple serpentinizing ecosystems.* International Society for Microbial Ecology 17, Leipzig. August 2018

2. **E. Frouin**, F. Armougom, G. Erauso. *Comparative metagenomics highlighted a core of metabolic capabilities in multiple serpentinizing ecosystems.* Journées Ouvertes en Biologie, Informatique et Mathématiques, Marseille. July 2018

3. **E. Frouin**, F. Armougom, G. Erauso. *Comparative metagenomics revealed a core of metabolic capabilities in multiple serpentinizing ecosystems.* CRG Scientific Sessions, Barecelone. June 2018

4. **E. Frouin**, F. Armougom, G. Erauso. *Métagénomique comparée des communautés microbiennes d'un système hydrothermal serpentinisé en baie de Prony (NC) avec celles d'autres systèmes hydrothermaux.* Colloque de l'Association Francophone d'Ecologie Microbienne, Camaret-sur-Mer. October 2017

Poster communications

1. **E. Frouin**, F. Armougom, G. Erauso. *Comparative metagenomics of microbial communities inhabiting Prony hydrothermal field and related ecosystems.* Congrès de l'Ecole Doctorale Sciences de l'Environnement, Aix-en-Provence, May 2017.

2. **E. Frouin**, F. Armougom, G. Erauso. *Functional metagenomics of microbial communities inhabiting deep and shallow serpentinite-hosted ecosystems.* Journées Ouvertes en Biologie, Informatique et Mathématiques, Lyon. June 2016

frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 February 2018
doi: 10.3389/fmicb.2018.00102

Diversity of Rare and Abundant Prokaryotic Phylotypes in the Prony Hydrothermal Field and Comparison with Other Serpentinite-Hosted Ecosystems

*Eléonore Frouin[1], Méline Bes[1], Bernard Ollivier[1], Marianne Quéméneur[1], Anne Postec[1], Didier Debroas[2], Fabrice Armougom[1] and Gaël Erauso[1]**

[1] Aix-Marseille Univ, Université de Toulon, CNRS, IRD, MIO UM 110, Marseille, France, [2] CNRS UMR 6023, Laboratoire "Microorganismes – Génome et Environnement", Université Clermont Auvergne, Clermont-Ferrand, France

OPEN ACCESS

Edited by:
Anna-Louise Reysenbach, Portland State University, United States

Reviewed by:
Donnabella Castillo Lacap-Bugler, Auckland University of Technology, New Zealand
William Brazelton, East Carolina University, United States

***Correspondence:**
Gaël Erauso
gael.erauso@mio.osupytheas.fr

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

Received: *12 October 2017*
Accepted: *17 January 2018*
Published: *06 February 2018*

Citation:
Frouin E, Bes M, Ollivier B, Quéméneur M, Postec A, Debroas D, Armougom F and Erauso G (2018) Diversity of Rare and Abundant Prokaryotic Phylotypes in the Prony Hydrothermal Field and Comparison with Other Serpentinite-Hosted Ecosystems. Front. Microbiol. 9:102. doi: 10.3389/fmicb.2018.00102

The Bay of Prony, South of New Caledonia, represents a unique serpentinite-hosted hydrothermal field due to its coastal situation. It harbors both submarine and intertidal active sites, discharging hydrogen- and methane-rich alkaline fluids of low salinity and mild temperature through porous carbonate edifices. In this study, we have extensively investigated the bacterial and archaeal communities inhabiting the hydrothermal chimneys from one intertidal and three submarine sites by 16S rRNA gene amplicon sequencing. We show that the bacterial community of the intertidal site is clearly distinct from that of the submarine sites with species distribution patterns driven by only a few abundant populations, affiliated to the *Chloroflexi* and *Proteobacteria* phyla. In contrast, the distribution of archaeal taxa seems less site-dependent, as exemplified by the co-occurrence, in both submarine and intertidal sites, of two dominant phylotypes of *Methanosarcinales* previously thought to be restricted to serpentinizing systems, either marine (Lost City Hydrothermal Field) or terrestrial (The Cedars ultrabasic springs). Over 70% of the phylotypes were rare and included, among others, all those affiliated to candidate divisions. We finally compared the distribution of bacterial and archaeal phylotypes of Prony Hydrothermal Field with those of five previously studied serpentinizing systems of geographically distant sites. Although sensu stricto no core microbial community was identified, a few uncultivated lineages, notably within the archaeal order *Methanosarcinales* and the bacterial class *Dehalococcoidia* (the candidate division MSBL5) were exclusively found in a few serpentinizing systems while other operational taxonomic units belonging to the orders *Clostridiales, Thermoanaerobacterales*, or the genus *Hydrogenophaga*, were abundantly distributed in several sites. These lineages may represent taxonomic signatures of serpentinizing ecosystems. These findings extend our current knowledge of the microbial diversity inhabiting serpentinizing systems and their biogeography.

Keywords: microbial communities, prony, shallow hydrothermal field, alkaliphiles, methanosarcinales, serpentinization

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 August 2016
doi: 10.3389/fmicb.2016.01301

Metagenomic and PCR-Based Diversity Surveys of [FeFe]-Hydrogenases Combined with Isolation of Alkaliphilic Hydrogen-Producing Bacteria from the Serpentinite-Hosted Prony Hydrothermal Field, New Caledonia

*Nan Mei[1], Anne Postec[1], Christophe Monnin[2], Bernard Pelletier[3], Claude E. Payri[3], Bénédicte Ménez[4], Eléonore Frouin[1], Bernard Ollivier[1], Gaël Erauso[1] and Marianne Quéméneur[1]**

[1] Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France, [2] GET UMR5563 (Centre National de la Recherche Scientifique/UPS/IRD/CNES), Géosciences Environnement Toulouse, Toulouse, France, [3] Institut pour la Recherche et le Développement (IRD) Centre de Nouméa, MIO UM 110, Nouméa, Nouvelle-Calédonie, [4] Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Univ Paris Diderot, Centre National de la Recherche Scientifique, Paris, France

OPEN ACCESS

Edited by:
Andreas Teske, University of North Carolina at Chapel Hill, USA

Reviewed by:
John R. Spear, Colorado School of Mines, USA
Igor Tiago, University of Coimbra, Portugal

***Correspondence:**
Marianne Quéméneur
marianne.quemeneur@ird.fr

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

Received: *18 February 2016*
Accepted: *08 August 2016*
Published: *30 August 2016*

Citation:
Mei N, Postec A, Monnin C, Pelletier B, Payri CE, Ménez B, Frouin E, Ollivier B, Erauso G and Quéméneur M (2016) Metagenomic and PCR-Based Diversity Surveys of [FeFe]-Hydrogenases Combined with Isolation of Alkaliphilic Hydrogen-Producing Bacteria from the Serpentinite-Hosted Prony Hydrothermal Field, New Caledonia. Front. Microbiol. 7:1301. doi: 10.3389/fmicb.2016.01301

High amounts of hydrogen are emitted in the serpentinite-hosted hydrothermal field of the Prony Bay (PHF, New Caledonia), where high-pH (∼11), low-temperature (<40°C), and low-salinity fluids are discharged in both intertidal and shallow submarine environments. In this study, we investigated the diversity and distribution of potentially hydrogen-producing bacteria in Prony hyperalkaline springs by using metagenomic analyses and different PCR-amplified DNA sequencing methods. The retrieved sequences of *hydA* genes, encoding the catalytic subunit of [FeFe]-hydrogenases and, used as a molecular marker of hydrogen-producing bacteria, were mainly related to those of *Firmicutes* and clustered into two distinct groups depending on sampling locations. Intertidal samples were dominated by new *hydA* sequences related to uncultured *Firmicutes* retrieved from paddy soils, while submarine samples were dominated by diverse *hydA* sequences affiliated with anaerobic and/or thermophilic submarine *Firmicutes* pertaining to the orders *Thermoanaerobacterales* or *Clostridiales*. The novelty and diversity of these [FeFe]-hydrogenases may reflect the unique environmental conditions prevailing in the PHF (i.e., high-pH, low-salt, mesothermic fluids). In addition, novel alkaliphilic hydrogen-producing *Firmicutes* (*Clostridiales* and *Bacillales*) were successfully isolated from both intertidal and submarine PHF chimney samples. Both molecular and cultivation-based data demonstrated the ability of *Firmicutes* originating from serpentinite-hosted environments to produce hydrogen by fermentation, potentially contributing to the molecular hydrogen balance *in situ*.

Keywords: hydrogen, microbial diversity, hydrogen producers, serpentinization, *hydA* genes, [FeFe]-hydrogenase, metagenomics

Bioresource Technology 259 (2018) 304–311

Contents lists available at ScienceDirect

Bioresource Technology

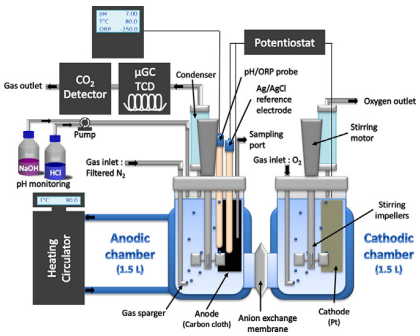
journal homepage: www.elsevier.com/locate/biortech

Specific enrichment of hyperthermophilic electroactive *Archaea* from deep-sea hydrothermal vent on electrically conductive support

Guillaume Pillot[a], Eléonore Frouin[a], Emilie Pasero[a], Anne Godfroy[b], Yannick Combet-Blanc[a], Sylvain Davidson[a], Pierre-Pol Liebgott[a,*]

[a] Aix Marseille Université, IRD, Université de Toulon, CNRS, MIO UM 110, Marseille, France
[b] IFREMER, CNRS, Université de Bretagne Occidentale, Laboratoire de Microbiologie des Environnements Extrêmes – UMR6197, Ifremer, Centre de Brest CS10070, Plouzané, France

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Hyperthermophilic exoelectrogens
Deep-sea hydrothermal vent
Microbial electrolysis cell
External electron transfer

ABSTRACT

While more and more investigations are done to study hyperthermophilic exoelectrogenic communities from environments, none have been performed yet on deep-sea hydrothermal vent. Samples of black smoker chimney from Rainbow site on the Atlantic mid-oceanic ridge have been harvested for enriching exoelectrogens in microbial electrolysis cells under hyperthermophilic (80 °C) condition. Two enrichments were performed in a BioElectrochemical System specially designed: one from direct inoculation of crushed chimney and the other one from inoculation of a pre-cultivation on iron (III) oxide. In both experiments, a current production was observed from 2.4 A/m^2 to 5.8 A/m^2 with a set anode potential of −0.110 V vs Ag/AgCl. Taxonomic affiliation of the exoelectrogen communities obtained on the electrode exhibited a specific enrichment of *Archaea* belonging to *Thermococcales* and *Archeoglobales* orders, even when both inocula were dominated by *Bacteria*.

1. Introduction

Since the discovery of the first deep-sea hydrothermal vent in 1977, many studies have expanded our understanding of extremophilic life forms in those environments. Considered as primordial environments similar to the early Earth conditions (Martin et al., 2008; Takai et al., 2006), the deep-sea hydrothermal ecosystems are regarded as a good model to study the origins and evolution of life on Earth (Takai et al., 2006). In addition, in recent years many authors have focused on the great potential of extremophilic marine microbes for their use as

* Corresponding author at: Mediterranean Institute of Oceanography, Campus de Luminy, Bâtiment OCEANOMED, 13288 Marseille Cedex 09, France.
E-mail address: pierre-pol.liebgott@mio.osupytheas.fr (P.-P. Liebgott).

https://doi.org/10.1016/j.biortech.2018.03.053
Received 8 January 2018; Received in revised form 9 March 2018; Accepted 10 March 2018
Available online 14 March 2018
0960-8524/ © 2018 Elsevier Ltd. All rights reserved.

Résumé. Les systèmes hydrothermaux serpentinisés proviennent de l'hydratation de roches ultramafiques à proximité de dorsales océaniques, de zones de subduction ou dans des ophiolites. Les fluides hydrothermaux issus des réactions de serpentinisation sont anoxiques et riches en hydrogène, méthane et petites molécules organiques. Ces composants minéraux et organiques alimentent les micro-organismes qui colonisent les systèmes serpentinisés, et ce en dépit d'un pH élevé et de faibles concentrations en accepteurs d'électrons et en carbone dissous. Dans ce travail, les communautés microbiennes ont été étudiées suivant deux axes : un premier axe focalisé sur Prony, un écosystème serpentinisé côtier de Nouvelle Calédonie et un second, comparant différents écosystèmes serpentinisés, pour faire émerger des similarités taxonomiques et fonctionnelles. À Prony, nos analyses en metabarcoding ont mis en évidence l'importance de la biosphère rare, au vu du grand nombre de taxons présents en très faible abondance. De plus, nos résultats suggèrent que la composition des communautés bactériennes à Prony est influencée par la nature des sites (sous-marins ou intertidaux), ce qui n'est pas le cas pour les archées. Enfin, l'analyse de cinq métagénomes a permis de reconstruire 82 génomes procaryotes provenant des sites hydrothermaux de Prony. Un de ces génomes est phylogénétiquement proche des espèces du genre *Serpentinomonas*, bactéries chimiolithotrophes isolées du site serpentinisé The Cedars, en Californie qui détiennent le record d'alcalophilie. Ces espèces et d'autres phylotypes, tels que les taxons affiliés aux Lost City Methanosarcinales ou au genre *Dethiobacter*, ont été trouvés dans plusieurs sites serpentinisés et pourraient contribuer à la définition d'une "signature biologique" des phénomènes de serpentinisation. La comparaison de métagénomes issus de systèmes hydrothermaux a révélé des divergences importantes entre les sites serpentinisés. Les profils fonctionnels des communautés microbiennes de Lost City (un site serpentinisé sous-marin profond) ont davantage de similarités avec ceux des sites hydrothermaux non influencés par la serpentinisation qu'avec ceux des sites serpentinisés côtiers et continentaux. Ces résultats s'expliquent par la diversité des habitats microbiens et l'influence des paramètres environnementaux, tels que le contexte géologique ou la géochimie des fluides hydrothermaux. En ciblant spécifiquement les métabolismes enrichis dans les milieux serpentinisés, nous avons pu mettre en évidence l'importance du métabolisme de l'hydrogène et des mécanismes de réponse aux stress environnementaux. L'analyse des métagénomes serpentinisés a indiqué l'importance d'une voie de dégradation des phosphonates, reposant sur l'activité d'une C-P lyase. Cette voie métabolique, qui a un rôle clé dans l'assimilation du phosphore et la libération de molécules organiques, vient enrichir les modèles écologiques des systèmes serpentinisés.

Mots clés : Serpentinisation, Écologie microbienne, Comparaison métagénomique, Dégradation des phosphonates, Reconstruction de génomes, Extrêmophiles.

Abstract. Serpentinizing hydrothermal systems result from the hydration of ultramafic rocks, near mid-ocean ridges, subduction zones or in ophiolites. Hydrothermal fluids associated with serpentinization are anoxic and enriched in hydrogen, methane and small organic molecules. These mineral and organic compounds support microbes that colonize serpentinizing systems, despite high pH and low concentrations of electron acceptors and dissolved carbon. In this work, two axes were explored to study the microbial communities. On the one hand, we focused on the Prony Hydrothermal Field, a coastal serpentinizing ecosystem in New Caledonia, and on the other hand we compared different serpentinizing ecosystems to reveal taxonomic and functional similarities. At Prony, our metabarcoding analyses highlighted the importance of the rare biosphere, given the large number of taxa present in low abundance. In addition, our results suggested that the composition of bacterial communities depended on the nature of the sites -submarine or intertidal-, while the archaea did not appear to be impacted. Finally, 82 prokaryotic genomes were successfully reconstructed using five metagenomes from the hydrothermal sites of Prony. One of these genomes was phylogenetically close to the species of the genus *Serpentinomonas*, chemolithotrophic bacteria isolated at the serpentinizing site The Cedars in California that are capable of growth up to pH 12.5. These species, and other phylotypes, such as taxa affiliated with Lost City Methanosarcinales or genus *Dethiobacter*, were identified in several serpentinizing sites and could contribute to the definition of a "biological signature" associated with serpentinization. Comparison of metagenomes from hydrothermal systems revealed significant differences between serpentinizing sites. The functional profiles of the microbial communities from Lost City (a deep submarine serpentinizing site) are more similar to those of hydrothermal sites not influenced by serpentinization than those of coastal and continental serpentinizing sites. These results are explained by the variability of microbial habitats and the influence of environmental parameters such as the geological context or the geochemistry of hydrothermal fluids. By specifically targeting enriched metabolisms in serpentinizing environments, we highlighted key functions associated with hydrogen metabolism and environmental stress response mechanisms. Analysis of serpentinizing metagenomes revealed the importance of a phosphonate degradative pathway, based on C-P lyase activity. This metabolic pathway, which plays a key role in the uptake of phosphorus and the release of organic molecules, was integrated into the ecological models of serpentinizing systems.

Keywords: Serpentinization, Microbial ecology, Metagenomics comparison, Phosphonate degradation, Genome reconstruction, Extremophiles.