國立臺灣大學生命科學院動物學研究所

碩士論文

Institute of Zoology, College of Life Science
National Taiwan University

Master Thesis

北臺灣酸性熱泉中微生物群集的總體基因體分析

Metagenomic Analysis of Microbial Communities from an Acidic Hot Spring in Northern Taiwan

林奎含

Kuei-Han Lin

指導教授：于宏燦 博士

Advisor: Hon-Tsen Yu, Ph.D.

中華民國 102 年 1 月

Jan, 2013

國立臺灣大學碩士學位論文
口試委員會審定書

北臺灣酸性熱泉中微生物群集的總體基因體分析
Metagenomic Analysis of Microbial Communities from an Acidic Hot Spring in Northern Taiwan

本論文係林奎含君（學號 R97B4 1017）在國立臺灣大學動物學研究所完成之碩士學位論文，於民國 102 年 1 月 2 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

于宏燦 （簽名）

（指導教授）

于宏燦

黃曉薇

動物學研究所所長 潘建源 （簽名）

誌謝

於此為期三年半的時間，本篇論文得以完成必須歸功於許多人。從地球科學跨越到溫泉微生物總體基因體，源於書報討論時林立虹老師所開啟的門，爾後又因緣際會地加入動物所于宏燦老師的實驗室。于老師是整個過程中最重要的指引者以及支持者，他以身作則、不遺餘力地導正與教誨，以及在其他各方面的協助與體諒，學生深切感念。感謝地質科學系楊燦堯老師實驗室陳乃禎、溫心怡與陳艾荻在溫泉探勘上的協助。溫泉水的大量採集是本次實驗材料的來源，感謝韓偉力、劉昂宇及何致煒，協助搬運百公升的泉水，以及提醒我開車不要睡著。實驗操作上，與這段日子以來實驗室生活的大小事，特別要對黃曉薇學姊致上敬意，感謝學姊教導我所有重要的實做能力，細心解答疑惑，平日也總是默默關心我。感謝呂曉沛學姊，時常與我分享她的經驗與見解，提供新的想法和啟示，同時也是不可缺的精神夥伴以及學習典範。國家衛生研究院的廖本揚老師、張庭諺先生，本實驗室的王育彬學長，他們的協助是此研究得以完成的另一關鍵。生物資訊於我是全然陌生的領域，廖老師與張庭諺先生不僅耐心引導我思考分析資料的方法，提供軟硬體各方面的支援，也費了甚多心神與我討論。倘若沒有他們，研究成果實難達到預期的質量。另外再次感謝于宏燦老師，在整體研究架構的設想以及彈性上調度有方，使得多數問題總是能獲得解決。感謝所有口試委員對於論文撰寫的指正與建議，尤其感謝黃曉薇學姊與廖本揚老師。感謝地質所林立虹老師與海研所王珮玲老師，總是樂意在我需要幫助的時候傾聽我的問題以及分享他們的心得。感謝環工所童心欣老師實驗室協助溫泉水中有機碳濃度的測量。感謝外籍生接待服務義工社、國際事務處和所有曾經合作過的朋友與長輩們，與生科院的國際交流補助，他們為我開了一扇有著不同風景的窗。感謝好鄰居劉昂宇和老同學何致煒。感謝韓偉力，總是為我補充常常不夠用的希望。感謝家龜和家犬的陪伴。

最後，感謝我的家人：本篇論文獻給你們。

摘要

極端環境微生物生態系的探查，數十年來是為地球歷史、生物演化、天體生物學領域發展的基礎研究之一。溫泉是臺灣常見的極端環境生態系統，早期在北部大屯山區對溫泉微生物生態系的調查偏重使用聚合酶連鎖反應(polymerase chain reaction, PCR) 偵測原核生物 16 S rRNA 基因，再加以轉殖與定序，對於微生物族群結構與功能的探究深度及廣度較為不足。近五年來，次世代定序的技術逐漸成熟與普及，使序列資料的獲得更有效率，也能得知更接近真實的自然界樣貌。在這篇研究中，選定於大屯山區四磺坪地區溫泉進行採樣，利用總體基因體學結合 Illumina™ 定序平臺分析大屯火山群的四磺坪酸性熱泉中的微生物群集，希望了解其主要的群集結構與功能。

經過調查，四磺坪區域的溫泉，泉溫平均為 73°C，pH 平均為 2.8，估計溫泉水中浮游原核生物密度約為 7.3×10^5/mL，低於一般陸地淡水常溫水體與鹽湖大約十倍。浮游微生物總體基因隨後建構為一包含九千餘株樣本的 fosmid library，自 library 中取一千餘株樣本與另一批直接自四磺坪泉水萃取的總體基因同時進行次世代定序。經由生物資訊分析定許結果，發現細菌為此溫泉生態系的強勢種，估計與古細菌的比例約為 7:3 到 8:2 之間。細菌主要以化學自營性細菌 phylum Aquificae 成員為主，其次為 Proteobacteria 和 Firmicutes；古細菌則以 phylum Crenarchaeota 佔主要位置。在本次調查中，部分分析顯示 Aquificae 之下的 *Hydrogenobaculum* 屬佔有相當高的相對豐度。在比對模式種基因體序列後，發現無法完整重建本區 *Hydrogenobaculum* 類似物種的基因體，可能暗示著此二者間存有某種程度上的序列差異性。藉由 Protein COG 的分佈模式分析，並與其他相似案例比較，推測此熱泉微生物群集的整體基因體組成偏重於代謝相關功能。

關鍵字：溫泉，微生物，多樣性，總體基因體，次世代定序，生物資訊

Abstract

Researches of ecosystems in extreme environments have been critical to the accumulation of fundamental knowledge in fields of earth history, evolution biology and astrobiology. Novel enzymes, molecules possessed by microorganisms in extreme environments also provide sources for potent molecular biological and biotechnical applications. Extreme environments such as hot springs are commonly seen in Taiwan. Previous studies relevant to hot spring microbial ecosystems in the Tatun Volcanic Group area (TVG, northern Taiwan) relied mainly on SSU rRNA gene sequence analyses, which were unilateral and could be possibly biased. To obtain a more realistic view of prokaryotic diversity by acquiring sufficient sequence information, next-generation sequencing (NGS) was applied to this study with the aid of bioinformatics. Through the investigations on metagenomes from microbial community in a hot spring in SHP area, this study was aimed to obtain an overview on prokaryotic diversity, community structure and function in the hydrothermal spring system.

According to the results, the average pH of the hot spring in SHP area was 2.8, and average temperature was 73°C. Cell density of planktonic prokaryotes was estimated to be 7.3×10^5/mL, which was about 10 times lower than that of terrestrial fresh water and salt lakes. Metagenomic DNAs extracted from SHP hot spring were used for construction of a fosmid library containing 9481 clones. DNAs extracted from hot spring water and DNAs of 1485 clones from fosmid library were sequenced using Illumina™ platform. Analyses of community composition revealed dominance of bacteria over archaea by 7:3 to 8:2. Members of bacterial phylum Aquificae were

abundant, followed by Proteobacteria and Firmicutes; members of phylum Crenarchaeota were most abundant among archaea. Sequences from predominant members of genus *Hydrogenobaculum*, affiliated with Aquificae, were mapped to the reference genome and certain level of sequence difference was found in the comparison. By analysis of the protein COG distribution pattern and comparing results from the other research, the hot spring prokaryotic community in SHP displayed a higher proportion of metabolism-related functions encoded in their genomes.

Key words: hot spring, microorganism, diversity, metagenomics, next-generation sequencing, bioinformatics

Table of Contents

List of Figures and Tables

1. Introduction

1.1 Extreme Environments and Extremophiles

1.1.1 Definitions

From an ecological view, the whole Earth system encompasses numerous niches with unique environmental settings and each of them has its own biota that consists of functional related consortia of organism. For not only survival but also prosperous growth and reproduction, moderate environments of comparatively less evolutionary stress are preferred by vast majority of life. The criteria for moderate conditions are somehow anthropocentric, referred to those adequate physical and chemical property that are close to the range favored by human beings: pH value around neutral, temperatures between 20 to 40°C, air pressure about 1 atm, with sufficient water activity (better higher than 0.7) and nutrients accessibility, suitable concentration of salts (between normal fresh water and sea water), without excessive exposure to radiation (compared with the average amount that is received on the surface of the Earth), and lower levels of heavy metal or toxic compounds (such as organic solvents). In contrast, those harsh environments which are featured by unusual physicochemical conditions, such as high acidity and alkalinity, extreme temperatures, high pressure or vacuum state, low water activity, low amount of nutrition, high salinity, intensive radiation, and places with high concentration of heavy metal or toxic substance, are defined as extreme environments (Rothschild and Mancinelli, 2001; Wilson and Brimble, 2009).

Survival and reproduction are the most important goals to achieve for all kinds of life. In order to adapt to the dynamic material surroundings, organisms are usually capable of tolerating a certain degree of physical and chemical fluctuations in environments. However, a state of tolerating is different from optimal growth; thus, organisms which can survive in extreme environments are sometimes just extremotolerant instead of extremophilic—the latter describes the characteristics that an organism has its optimal growth under extreme conditions, and the organism *per se* is called an extremophile (Macelroy, 1974; Kristjansson and Hreggvidsson, 1995; Wilson and Brimble, 2009). Accordingly, extremophiles thriving in environments with multiple harsh properties are grouped as "polyextremophiles".

Increasing interest and efforts in researches relevant to extreme environments and organisms started around the 1950's. Biotopes such as deep sea, hypersaline environments, hot springs, deserts, ice, permafrost and atmosphere are the most focused regions for studying and all of them have various combinations of geochemical backgrounds. For instance, general features of deep sea environment are low temperature around 1 to 2°C, anoxic and lack of photosynthesis. However, at some specific locations, when hydrothermal vents or ancient evaporite beddings exist, environmental conditions would then be a mixture of its original geochemical context plus individual variables: hydrothermal vents heat up surrounding sea water and provide an additional chemical source, and dissolution of evaporite beddings would greatly increase

the local salinity. The same is true for other locales: environmental conditions are results of interplays of regional backgrounds and local variables.

In the conventional classification of extremophiles, the main categories are divided based on physical and chemical conditions optimal for growth. Those commonly mentioned are thermophiles (high-temperature-loving), psychrophiles (low-temperature-loving), acidophiles (low-pH-loving), alkaliphiles (high-pH-loving), halophiles (high-salinity-loving), xerophiles (low-water-activity-loving) and barophiles (pressure-loving, also called piezophiles). In some cases, those being able to resist particular stringent environmental factors are included into extremophiles, such as radioresistant and endolithic (living in cracks or pores of rocks or minerals) organisms, although this is somehow incongruent with the definition of extremophiles.

To view extremophiles as a whole, interest of research are of three categories: (1) evolution and earth history, (2) astrobiology and extraterrestrial trace of life and (3) biotechnological applications (Wilson and Brimble, 2009). From the formation of the planet Earth to the existence of life, the proto-environments are regarded to be analogous to some current extreme biotopes, which are generally anaerobic with extreme high temperature, low organic matters, and sparse nutrients. Through the expansion of knowledge on extreme biogeochemical systems, mechanisms of evolution, and biodiversity, the paleoclimatic change along the geological time could be better deduced. In

a broader scale, the current search for extraterrestrial life within the solar system is supported by knowledge from studying environment analogues on the Earth. In practical aspects, extremozymes (Hough and Danson, 1999) produced by extremophiles with distinct functions have tremendous potential for the biotechnology industry, bioremediation, chemistry and pharmacy—the well-known Taq polymerase isolated from *Thermus aquaticus* (Brock and Freeze, 1969) used in polymerase chain reaction has served as a remarkable example.

1.1.2 The Underexplored Microbial Extreme Ecosystems

Microbial ecosystems, compared with other ecosystems of macro-organisms, are still poorly understood. Microscopic size made them inconspicuous for human eyes, therefore receiving less than sufficient attention. Furthermore, their almost-universal existence, difficulties in cultivation and technical constraints (although great progression has been made molecularly and bioinformatically, abundance and diversity of microorganism are not completely assessable within a short period of time) have made the discovery and in-depth understanding of microorganisms in a slower process than their significance deserved. The same situation occurs in the exploration of microbial extreme ecosystems.

Extreme biotopes, located only in limited geographic regions with unusual physical and chemical characteristics, usually harbor a significantly lower

amount of cells compared to normal niches and thus make it more laborious to collect sufficient amount of samples (Ferrer et al., 2009) for adequate analyses. In the study of extreme ecosystems, prokaryotic organisms have long been the major target of interest because of their higher abundance and diversity over other multicellular or eukaryotic life in extreme environments. Prokaryotic microorganisms, which consist of members from bacteria and archaea domains, are thought to be the most wide-spread form of life on Earth and the number of prokaryotes was about 10^{30} cells in total (Turnbaugh and Gordon, 2008). As inhabitants and decomposers in ecosystems, their remarkable physiological functions are closely connected to the biogeochemical cycles of the Earth.

As progress of microbiology started from traditional culture-based methods into a culture-independent era and by the aid of extensive global explorations, rapid analysis and large amount of sequence data also revealed the underestimation of prokaryotic diversity in extreme environments (Hugenholtz et al., 1998; Reysenbach et al., 2000; Takai et al., 2001b; Huber et al., 2002b; Satyanarayana et al., 2005; Sogin et al., 2006; Yim et al., 2006; Huber et al., 2007; Wilson et al., 2008; Kato et al., 2011). Cultivation had identified 10 or 12 divisions in Bacteria domain and 2 or 3 divisions in Archaea domain (Woese, 1987). However, more than 40 divisions in bacteria (Pace, 1997) and more than 12 divisions in archaea (DeLong and Pace, 2001) are now discovered through culture-independent approaches and the total number of

genospecies approximates 10^6 to 10^8 (Sleator et al., 2008) with 1316 complete prokaryotic genomes available in online databases (updated in October 2012). According to statistics, in recent prokaryotic sequencing projects only c.a. 4% of them belong to extremophiles (Ferrer et al., 2007), which strongly indicates that there is still room and needs in the field of extreme ecosystem explorations.

1.2 General Features of Hot Spring Environments

1.2.1 Geological, Physical and Chemical Conditions

A spring is a concentrated discharge of groundwater that appears at the surface as a current of flowing water (Todd, 1980). Conventionally, hot springs are terrestrial. When geothermal water issues under the sea, although it is also a kind of hot spring under water, a better term is hydrothermal or geothermal vent. Therefore, in this research we followed the general principle and declined the usage of "hot spring" in continental regime. To date, there is no single definition for hot springs. Usually the term is applied when the spring water meets one of the three descriptions (Todd, 1980; LaMoreaux and Tanner, 2001): (1) spring water with temperatures higher than that of its local underground water; (2) spring water that is 5°C higher than the average temperature of its local annual air average temperature or (3) springs with warm water above body temperature.

Hot springs are classified according to different purposes: temperatures (Table 1), geological backgrounds (such as volcanic, sedimentary or metamorphic), physical features (such as fluid states, activity patterns and geomorphic appearance)(Table 2) and chemical properties (pH and dissolved ions in spring water)(Table 3). Geothermally and volcanically active zones are locations suitable for hot spring development. Precipitations of meteoric water and underground water are usually the sources of hot spring water, but hydrological properties of each hot spring varies because of the influences by various geochemical environments.

1.2.2 Lives and Biological Activities

Hot springs in some regions are part of local cultures (hydrotherapy, bathing etiquettes, etc.) and recreations, and its geothermal nature is also close related to the mining of resources in peripheral areas (elemental sulfur, silica, bentonite, pyrite and gold). Generally speaking, the variety and intensiveness of biological activity depend almost always on environmental conditions a hot spring possesses. It is possible for large animals, plants and invertebrates such as insects to thrive nearby the hot spring. However, taking high temperature or acidic spring water as an example, there are only a few kinds of life that might survive—most of them are microorganisms. For photosynthetic microorganisms, the maximum temperature for survival is 73 to 75°C (Brock, 1967; Cox et al., 2011) and better with low sulfide concentration (Cox et al., 2011). While the environment becomes harsher, certain branches of

prokaryotes could be found under certain conditions (Satyanarayana et al., 2005).

1.3 Microbial Activities in Acidic Hot Springs in Tatun Volcanic Area

1.3.1 Regional Background

Sitting on the lively collision zone between the Eurasian Plate and the Philippine Sea Plate, Taiwan is an island featured by frequent geothermal activities. The mountain belt has been well-accepted to be a result from collision between the Luzon arc and the Asian continent (Chai, 1972). Our study focused on prokaryotic microorganisms in hot springs of Tatun Volcanic Group (TVG), the largest volcanic group in Taiwan (Wang and Chen, 1990) (400 km^2, including more than 20 volcanoes), which is dormant in modern times. TVG has an average elevation of 800 to 1000 m and consists of two principal volcanic ridges, E–W and SW-NE; each is about 15 km long.

Previous dating data (Wang and Chen, 1990; Song et al., 2000) suggested the volcanic activity started from 2.8 to 2.5 Ma with relatively mild eruptions due to the regional compressive stress field. The massive explosive event occurred between 0.8 to 0.2 (some suggested after 0.1) Ma, while collision between plates in northern Taiwan had been weakened or even stopped and thus created a divergent stress environment which facilitated the extrusion and release of magma through surface cracks generated by normal faults. The latest evidence

specifically indicated the last eruption record was around 6000 years ago (Belousov et al., 2010).

Although generally being considered dormant or even extinct, TVG still displays several traces that strongly suggest the existence of a magma chamber and its active geothermal activity: frequent shallow micro-earthquakes, harmonic codas, seismic tremors (Lin et al., 2005; Konstantinou et al., 2007), high heat flows, measurements of volcanic gases (Yang et al., 1999) and abnormally high helium ratios (Song et al., 2000). Many distinguishing volcanic characters such as solfataras, fumaroles, sulfur crystals (yellowish needle or dendritic shaped) and irritating sulfuric gases are common scenes in this area.

Hydrothermal systems in TVG area, primarily with water of meteoric origin, are heated by the geothermal environment and react with peripheral rocks and gases. Therefore, the water property is mostly determined by the regional geological background. TVG is an andesitic volcanic unit with volcanic gases comprising mainly of H_2O, CO_2, SO_2, HCl (Lee et al., 2005) and its hot springs are of principally three types defined by properties of host rocks, pH values, and the major elements: (1) SO_4^{2-} acidic water (Type I); (2) HCO_3^- a nearly neutral water (Type II); and (3) Cl^- -rich acidic water (Type III) (Liu et al., 2011). The three types of hot springs are scattered in TVG. Type I is usually with pH and temperature around 1.5 to 3.2 and 42 to 93°C, TDS (total

dissolved solids) 245 to 12900 mg/L; type II is generally with near neutral pH, 40 to 62°C, and TDS concentration about 957 to 1149 mg/L; type III has its pH, temperature and TDS in between 1.2 to 1.9, 71 to 91°C, 17400 to 19800 mg/L, respectively.

Although there are some neutral carbonate hot springs in the TVG area, sulfuric acidic hot springs are predominant in this area. The sampling site of this research belongs to Type I, featured by sulfur crystalized on the rim of wide-spread fumaroles. This type of hot springs is geochemically characterized by a high concentration of SO_4^{2-}, but with low HCO_3^- and Cl^- anions and low Na^+, K^+, Mg^{2+} and Ca^{2+} cations.

1.3.2 Previous Surveys on Microbial Diversity in TVG Area

Researches related to thermophilic microorganisms in Taiwan have been started since 1960s. Most of them were focused on physiological and biochemical aspects through cultivation or selective functions of specific genes; only a few had put the interest on a broad spectrum of surveys on microbial diversity and community function in TVG hydrothermal systems (陳懋彥, 2002; 何珮楨, 2004; Ng et al., 2005; 李重義, 2006; 鄭婷文 and 林立虹, 2007; 李信緯, 2009; 李重義 and 張怡塘, 2009; Cheng et al., 2013; Lu et al., unpublished data).

In environments such as deep sea hypersaline basins (Ferrer et al., 2012),

hydrothermal vents (Takai et al., 2001b; Nakagawa et al., 2005b), mud volcanoes (Cheng et al., 2012) and continental geothermal springs (Macur et al., 2004; Cheng et al., 2013), it is well-documented that microbial community structures are dynamic in accordance with physical and chemical profiles of the environment. In a terrestrial hydrothermal spring or hot pond, sediment and water layer are viewed as two distinguished habitats for microbes due to many differences of environmental variables such as photic/aphotic, aerobic/anaerobic, etc. Therefore, it would be more realistic to separate the sediment and the water layer when the microbial diversity in a hot pond is to be investigated.

Earlier works on microbial diversity of acidic thermal springs in TVG area were done by Li's group (李重義, 2006; 李重義 and 張怡塘, 2009) using 1:1 shallow layer sediments (10 cm at depth) mixed with spring water as material. They pointed out that each sample site has a unique microbial community composition with specific patterns of archaea and bacteria constituents.

Applying the 16 S rDNA sequence amplification to water-sediment samples from three sites DRG (地熱谷), LHG (硫磺谷) and GZP (庚子坪), 李 (2006) constructed libraries and set up an indicative reference for comparing the relative amount of microbes. The author observed that in low pH environments bacteria were more abundant than archaea, whereas an

increasing amount of archaea was found in mildly low pH environments. However, there were no strict criteria for defining the environment conditions and for quantifying the amount of microorganisms. Conclusively, two orders of Crenarcheota, Thermoprotei and Sulfolobales were discovered in four selected springs in the report: *Caldisphaera* sp., *Metallosphaera* sp. and *Sulfolobus* sp.. About 78% were uncultured archaeon clones. In bacteria domain, microbes such as *Aquabacterium* sp., *Delftia* sp., *Desulfurella* sp., *Methylophilus* sp., *Pseudomonas* sp., *Stenotrophomonas* sp. and *Thiobacillus* sp. were detected, which belong to classes Beta-proteobacteria (Burkholderiales, Methylophilales, Hydrogenophilales), Delta-proteobacteria (Desulfurellales) and Gamma-proteobacteria (Pseudomonadales, Xanthomonadales). Approximately 45% of bacteria clones were uncultured.

In the research conducted by Cheng et al. (2013), samples from the two locations HS (磺山) and LHG (硫磺谷) in TVG acidic hot springs demonstrated that diversity and segregated niches (water body and sediment layer) of hot spring microbes were likely correlated with the oxygen level in the system. In the water column, microbial community structures were homogeneous in different depth (since the hot ponds were not deep) but remarkably different from that in the sediment layer: archaea displayed different community structures between water and sediment layers, while those for bacteria did not differ as much. According to the PCR amplification on 16 S rDNA sequences, this research also indicated that at least during this

sampling period some sample sites contained no bacteria. Based on the conclusion derived from the clone library screening and sequencing, they concluded that community structures in the two sample sites were relatively simple, only 12 phylotypes discovered. Crenarchaeal orders Sulfolobales (*Sulfolobus* sp. and *Acidianus* sp.), Thermoproteales (*Vulcanisaeta* sp.), Caldisphaerales (*Caldisphaera* sp.) and euryarchaeal order Thermoplasmatales (*Thermoplasma* sp.) were identified to be the closest cultivated representatives, and bacterial order Aquificales (*Hydrogenobaculum* sp.) was predominant. About 40% and 12% in the archaeal and bacterial clone libraries, respectively, were unclassified phylotypes.

In another study, Lu et al. (unpublished data) analyzed thermal spring planktonic microbial community structures at five locations (DRG 地熱谷, LHG 硫磺谷, LFG 龍鳳谷, HS 磺山, FY 富源 and HG 磺港) along the west-east ridge of TVG area. At the east-most parts of the sampling region no bacterial 16 S rRNA gene sequences were amplified by PCR. Although diversity and community structures varied from one another, Caldisphaerales, Desulfurococcales, Sulfolobales and Thermoproteales were common archaeal orders seen in those sites, and Aquificales, Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Epsilonproteobacteria and Clostridia were phylogenetic groups harbored the sampled bacterial members. Uncultured archaeal were 33% of total archaeal phylotypes, and 21% was that of bacterial ones.

The research completed by 李重義 and 張怡塘 (2009) collected water-sediment samples from five sites (LHG 硫磺谷, ZSL 中山樓, MC 馬槽, DYK 大油坑 and GZP 庚子坪) for microbial diversity analysis. Significant differences of bacterial and archaeal community compositions among sampling sites were again confirmed, which revealed that dominant microbes differed among locations. Most of the archaea and bacteria were uncultivable species. Of archaea, 4 orders were discovered under phylum Crenarchaeota: Caldisphaerales (*Caldisphaera* sp., *lagunensis* sp.), Desulfurococcales (*Acidilobus* sp.), Sulfolobales (*Acidianus* sp., *Metallosphaera* sp., *Sulfolobus* sp.) and Thermoproteales; only order Thermoplasmatales was of Euryarcharchaeota. In bacteria, 9 phyla were found in this research: Acidobacteria (*Acidobacterium* sp.), Actinobacteria, Aquificae (*Hydrogenobaculum* sp.), Chlorobi (Phylum), Chloroflexi (*Anaerolinea thermophila)*, Firmicutes (*Alicyclobacillus* sp., *Thermoanaerobacter* sp.), Proteobacteria (*Sphingomonas* sp., *Ralstonia* sp., *Thiomonas* sp., *Thiobacillus* sp., *Acidithiobacillus* sp., *Rheinheimera* sp., *Desulfurella* sp. *Desulfurella kamchatkensis* and *Desulfovibrio* sp.), Spirochaetes and Thermotogae (*Thermotoga* sp.).

1.3.3 Previous Research from Hot Springs Worldwide

Thermophilic communities are assumed to be geographically highly isolated from one another due to thermal barriers between habitats, but certain genera appear to be quite common to many continents. So far the most studied

microbial ecosystem of terrestrial hot spring environments are in the Yellowstone National Park (YNP) of the United States (Barns et al., 1994; Pace, 1997; Meyer-Dombard et al., 2005; Hall et al., 2008) , Kamchatka of Russia (Bonch-Osmolovskaya et al., 1999; Reigstad et al., 2010), Japan (Yamamoto et al., 1998; Nakagawa and Fukui, 2002), Tibet (Yim et al., 2006; Lau et al., 2009; Huang et al., 2011), Iceland (Marteinsson et al., 2001; Reigstad et al., 2010), New Zealand (Hetzer et al., 2007; Childs et al., 2008), Bulgaria (Tomova et al., 2010; Ivanova et al., 2011), Indonesia (Baker et al., 2001), Thailand (Kanokratana et al., 2004) and South Africa (Tekere et al., 2011). Due to the different regional geological history, each hot spring has its unique physiochemical features which are relevant to the components, structures and functions of its microbiomes.

Although detailed tectonic histories vary, hydrothermal systems hotspots share one important characteristic: the existence of geothermal energy sources. Yellowstone National Park (YNP), featured by Yellowstone Caldera, is the largest volcanic system in North America. Several volcanic eruptions driven by North American Plate drifting over a stationary mantle hotspot have created diversified geothermal environments with active hydrothermal in YNP. Iceland lies on the mid-Atlantic ridge (also: mid-ocean ridge) rift zone, the boundary between the North American and Eurasian tectonic plates. Upwelling of high-temperature mantle material generates the geothermal energy and results in active volcanic events. Most of the geothermal springs are with pH around 8

to 9, mild-acidic to neutral are of the minority. Na^+ and Cl^- ions are usually abundant (Arnórsson et al., 1983). The extremely isolated Tibetan geothermal region has recently become a focus of microbial diversity exploration. Among the youngest mountain ranges on the planet, the Himalaya was formed as a result of a continental collision or orogeny along the convergent boundary between the Indo-Australian Plate and the Eurasian Plate since Upper Cretaceous (70 Ma).

Almost all eukaryotic microorganisms cannot thrive in environments with temperature higher than 68°C. In terrestrial hydrothermal systems, regardless of pH, Crenarchaeota of archaea domain are more prevalent than Euryarchaeota and other archaeal phylum (Thaumarchaeota, Korarchaeota and Nanoarchaeota) in hydrothermal systems. In bacteria domain, Cyanobacteria are widely distributed and commonly seen to be associated with Chloroflexi (green non-sulfur bacteria) in environments with temperatures under 73 to 75°C (Ferris and Ward, 1997) where photosynthesis are still possible. At temperatures above 75°C to boiling degree, chemolithoautotrophic communities appear as gray or pink filaments/streamers in waters, often tolerating high dissolved sulfide levels (Reysenbach et al., 1994). Other bacteria such as Proteobacteria (α, β, γ, δ), Aquificae, Firmicutes, Acidobacteria, Actinobacteria, Thermodesulfobacteria, Deinococcus-Thermus, Bacteroidetes, Nitrospirae, Spirochaetes and Verrucomicrobia were also discovered in hydrothermal systems (although not all of these groups of

microbes occurred in every hot spring).

For acidic hot springs, well-studied research area were more restricted to YNP (Burton and Norris, 2000; Jackson et al., 2001; Inskeep et al., 2010), and other studies are in Andes (Bohorquez et al., 2012), Japan (Kato et al., 2011), New Zealand (Ellis et al., 2005) and Tibet. Bacteria phyla such as Proteobacteria, Firmicutes, Planctomycetes, Spirochaetes and Aquificae were found; archaeal phylum Crenarcheota is still in general more prevalent than Euryarchaeota and others. In some cases, especially when the environments were more acidic (pH value lower than 5) and had higher temperatures (above 65°C), archaea populations are sometimes become dominant over those of bacteria (Inskeep et al., 2010). Within Crenarchaea, Sulfolobales were found to be dominant in many surveys (Ellis et al., 2005; Inskeep et al., 2010). In addition to widely-distributed Proteobacteria, deep-branched Aquificae is another phylum which was revealed to be dominant among bacteria.

Microbial composition in communities is thought to be correlated with environmental factors; however so far there are no universal quantitative standards to clarify the definite relationship between them. According to Mathur et al. (2007), phylogenetic diversity of bacterial communities seems to be more related to temperature.

1.4 Study of Microbial Community

1.4.1 Metagenomics and DNA Sequencing

The largest proportion of individual organisms on earth is represented by prokaryotes, which were estimated to comprise 10^6 to 10^8 genospecies (Sleator et al., 2008). At present, only a few thousand species of microorganisms have been formally described, while more than 99% are still under discovery and most of them cannot be readily cultured (Amann et al., 1995; Rappe and Giovannoni, 2003; DeLong, 2005). Basic microbiology stemmed in the accumulation of knowledge from species which had been isolated and cultured. Life history, morphological, physiological and metabolic characteristics were described in detail.

Based on sequence similarity of small subunit ribosomal RNA gene, Woese and Fox constructed the first tree of life that portrayed the concept of three domains of life in 1977. A decade later, the utilization of polymerase chain reaction (PCR) started (Mullis and Faloona, 1987). These key revolutionary methods combined with gene cloning technique (Weisburg et al., 1991; Reysenbach et al., 1992a), molecular identification and phylogenetic analysis of prokaryotes thus became another core procedure in microbiology studies that offered a culture-independent option for systematic assessments of microbes.

With the aid of development in modern molecular biotechnology and

methodology, metagenomics (Handelsman et al., 1998), a study of a whole microbial community through genetic material recovered directly from field samples, further expanded the realm of microbiology to the ecological scale. It allowed scientists to look into the microbial diversity and interactions with a broad view bypassing cultivations of single isolates, and has gradually become a regular strategy for probing into the structure and function of microscopic-biospheres. From genomics to metagenomics, an enlarged frame of research brought increasing amount of specimens and genetic information as well. DNA sequencing is one of those critical innovations that help the effective transformation from hardly visible microorganism specimens to readable genetic codes.

DNA-sequencing techniques, first appeared in 1968 and became prevalent after 1977 (Maxam and Gilbert, 1977; Sanger et al., 1977), has played a key role concurrently with the trend of culture-independent microbiology research. In its early stage, the focus of metagenomics (Schmidt et al., 1991) was the analyses of single representative genes (e.g. the small subunit ribosomal RNA gene) in a community of interest. Routine procedures included target gene amplification, cloning, library construction, and sequencing. To explore the ecological diversity (Biddle et al., 2008) or metabolic profile (Tringe et al., 2005), sequences generated were often compared with references from well-studied cultivable relative species, in order to depict possible physiological or metabolic traits as well as evolutionary distances. In addition

to genetic information, metagenomic libraries constructed in the process are themselves valuable reservoirs of novel enzymes and molecules ready to be discovered by means of function- or sequence-based screening (Handelsman, 2004; Simon and Daniel, 2011).

Automated Sanger DNA-sequencing was the chief approach used in metagenomic research in the early stage. Due to technical constraints, efforts were mainly put on certain gene markers or genes encoding molecules with essential functions. However, discoveries driven by only a few marker genes provided restrictive or sometimes even biased information. Taxonomically, information tends to suffer from loss or distortion during PCR and cloning steps; functionally, microbial ecology analysis represented by only a few genes is far from convincing. Therefore, completeness of every genome in the surveyed community was then pursued in attempt to create a more objective and systematic vision in metagenomic studies. One of the strategies to achieve this is environmental shotgun sequencing (ESS), which sequences massive small pieces from a fragmented genomes, and assemble the sequences by the overlapping regions through the assist of algorithms and computational power (Venter et al., 2004; Edwards et al., 2006). Evidence had supported that, compared to traditional ribosomal RNA–based investigation, direct sequencing of metagenomic DNA (ESS) provides the most accurate approach for assessing taxonomic composition (von Mering et al., 2007). By now ESS has been one of the most conventional ways for environmental microbiology

studies and the general concept of shotgun sequencing has also been the core strategy involved in genome sequencing projects.

1.4.2 Next Generation Sequencing

Regardless of what purposes to achieve, the genetic sequence has so far been a standard information format to access the world of microbiology and most of the fields relevant to biology. Conventional Sanger DNA-sequencing technology (Sanger et al., 1977), the so-called first-generation sequencing, is capable of recovering up to 1 kb of sequence data from an individual sample at a time. The most advanced Sanger sequencers can elevate the efficiency to processing 96 samples at a time with up to 1 kb sequence for each sample. Since one reaction allows only one specimen to be analyzed with Sanger sequencing, it is quite time-consuming, labor-intensive and expensive for metagenomic studies with huge amount of clones to be sequenced.

In 2005, the “next-generation” sequencing (NGS) technology (also: second generation sequencing, massive parallel sequencing or high-throughput sequencing) was first commercially available with the 454 sequencing system (Margulies et al., 2005). Being developed in 1996 (Ronaghi et al., 1996; Kawashima et al., 1998) and are now with several derived core technologies (Mardis, 2008; Shokralla et al., 2012), NGS possesses some revolutionary features that are distinguishable from capillary-based Sanger sequencing. First is the throughput of data, which is at least an order of magnitude higher than

Sanger method, measured by base pairs per day of run time. Second, the cost of NGS by base pair is at least an order of magnitude lower. Third, the reads from NGS are generally shorter than those generated by capillary sequencing. Last, an environmental sample can be recovered and analyzed directly through NGS and bypass the cloning procedure as well as the PCR amplification. Suppose that one human genome is to be sequenced by the two approaches. By utilizing ABI 3700 (Mitchelson, 2001) of Sanger capillary sequencing technology (could read up to 1 million base per day), eight years of non-stop processing is needed to complete the task. In contrast, only 2 hours are required if an Illumina HiSeq 2000 (Ajay et al., 2011) of NGS technology is applied (which is capable of producing 50 Gbp per day).

Bypassing the cloning step in NGS sequencing makes it possible to obtain result in an objective way: DNA fragments are usually not cloned into host with equal possibility, some may lost during processing, some may subject to selective bias such as GC content (Temperton et al., 2009). However, for NGS samples which have no library archives, genes are discovered only as a series of sequences, their actual physiological functions or relative gene locations are so far impossible to be re-accessed from picking up clones that harbor the desired inserts. NGS has advantages over capillary-based method mainly in its affordable price for high-throughput data production, but reduction in average read length has always been a drawback even advances in bioinformatics are gradually narrowing the gap between immense data and proper solutions for

analysis and interpretation.

Since 2005, researches incorporating NGS technique have increased dramatically. Unlike first generation sequencing, NGS constitutes various strategies that rely on complex interplay of enzymology, chemistry, high-resolution optics, hardware and software engineering. By now, 454-prosequencing GS FLX + system (Roche) and HiSeq system (Illumina) are the main players in high throughput sequencing despite the fact that there are some other competitors and new comers.

1.4.3. From Sequencing, Genetic Codes to Information

From capillary-based sequencing to the era of NGS massive parallel sequencing, output data has dramatically expanded in size but meanwhile shortened in the read length. Careful considerations must be taken when evaluating if a sample is suitable to be analyzed through the shotgun sequencing approach, since the output sequences from all members within the community will be fragmented and mixed. To choose a community with less structural complexity (such as one dominant species and low diversity) can effectively reduce the difficulty during assembly, especially when shotgun sequencing is to be carried out by an NGS sequencer. Supporting evidences came from several metagenomic studies on extreme or oligotrophic environments (although were sequenced through capillary-based method) such as acidophilic biofilm (Tyson et al., 2004), acid mine drainage (Edwards et al.,

2000) and the Sargasso Sea (Venter et al., 2004), which were successfully profiling the community structures or even recovered whole genomes of some dominant species (Baker and Banfield, 2003).

Sufficient quantity and high quality of extracted DNA is critical to sequencing. In general, for pyrosequencing or cloning, some micrograms of genomic DNA are required. Although library construction is not necessary for other NGS systems such as Illumina, at least 20 μg would be suggested. For low-biomass samples derived from environments with harsh conditions, usually it would be the major challenge to obtain enough metagenomic material for starting up (Ferrer et al., 2009).

Once the sequencing stage has been finished, reads are base-called from signals generated by sequencer, screened to remove vectors (if vectors were used) and barcodes and then undergo a quality check to trim off bases or reads with poor-quality. The next step is to assemble reads according to overlapping regions into longer fragments termed contigs. Assembly is a process which consumes enormous computing power and also a step which might introduce artifacts to the final output (Kunin et al., 2008). The finished length of contig can vary widely from around 100 bp to more than 100 kb, depending on microbial community composition and sample quality.

After contigs are generated, the diversity, gene functions and metabolic

pathways of the sample are three basic directions for further analyses. Databases such as SILVA (Pruesse et al., 2007), Greengenes (DeSantis et al., 2006), or Ribosomal Database Project II (RDP II) (Cole et al., 2003) offer useful resources for rRNA gene-based classification of microorganisms so that sequences containing taxonomic information or sequences derived directly from 16 S rRNA gene amplicon can be analyzed readily.

For gene function annotation and metabolic pathway investigation, the Gene Ontology (GO) database (Ashburner et al., 2000), Clusters of Orthologous Groups (COG) database (Tatusov et al., 2001), Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2008), Pfam (Finn et al., 2010), NCBI (Sayers et al., 2009), SEED (Overbeek et al., 2005) are commonly used databases for metagenomic data sets generated by NGS. To compare data on hand with other known species, reference genomes deposited in database like IMG (Joint Genome Institute) can be downloaded to a local server and mapped with reads or contigs.

After the sequence information was annotated, taxonomic relative abundance, main metabolic functions and unique pathways are usually integrated with environmental factors and cross-compared with other studies. Public domain depositions of all available data are also essential for environmental metagenomic research; the accumulation of results from surveys facilitates later cross-comparison and also contributes to the discovery of the whole

picture of microbial biosphere on Earth.

1.5 Motivations and Aims of This Study

Hot springs, one of extreme environments that had been proven to be a reservoir of valuable knowledge to the past (earth history), present (microbial ecology) and future (bioprospecting and astrobiology), are worth being further explored and understood. In the past few decades, studies of this field in Taiwan have accumulated a certain amount of knowledge that provided a rough image of local hot spring ecosystems. However, more thorough and systematic researches on prokaryotic diversity are needed. In this study, the high-throughput sequencing technique was applied to obtain sufficient sequence data that facilitated a more comprehensive analysis on prokaryotic community structure in an acidic hot spring ecosystem.

1.6 Framework of Research

Samples of hot spring water were collected, concentrated and from which metagenomic DNAs were extracted for fosmid library construction and direct shotgun sequencing. Results of sequencing were analyzed for assessments of prokaryotic community composition, genome mapping and community gene function distribution. Prokaryotic community structure analysis and gene functions were the main topics of this study (blue and green boxes of Figure 1).

Prior to the sample collection, a general survey on the bacteria, archaea and nanoarchaea distribution of six candidate hydrothermal springs in TVG area was done (white box of Figure 1 and Appendix 1).

References

Ajay, S. S., Parker, S. C. J., Abaan, H. O., Fajardo, K. V. F. and Margulies, E. H. (2011). "Accurate and comprehensive sequencing of personal genomes." Genome Research **21**(9): 1498-1505.

Amann, R. I., Ludwig, W. and Schleifer, K. H. (1995). "Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells without Cultivation." Microbiological Reviews **59**(1): 143-169.

Arnórsson, S., Gunnlaugsson, E. and Svavarsson, H. (1983). "The chemistry of geothermal waters in Iceland. II. Mineral equilibria and independent variables controlling water compositions." Geochimica Et Cosmochimica Acta **47**(3): 547-566.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. and Consortium, G. O. (2000). "Gene Ontology: tool for the unification of biology." Nature Genetics **25**(1): 25-29.

Baker, B. J. and Banfield, J. F. (2003). "Microbial communities in acid mine drainage." Fems Microbiology Ecology **44**(2): 139-152.

Baker, G. C., Gaffar, S., Cowan, D. A. and Suharto, A. R. (2001). "Bacterial community analysis of Indonesian hot springs." Fems Microbiology Letters **200**(1): 103-109.

Barns, S. M., Fundyga, R. E., Jeffries, M. W. and Pace, N. R. (1994). "Remarkable Archaeal Diversity Detected in a Yellowstone-National-Park Hot-Spring Environment." Proceedings of the National Academy of Sciences of the

United States of America **91**(5): 1609-1613.

Belousov, A., Belousova, M., Chen, C. H. and Zellmer, G. F. (2010). "Deposits, character and timing of recent eruptions and gravitational collapses in Tatun Volcanic Group, Northern Taiwan: Hazard-related issues." Journal of Volcanology and Geothermal Research **191**(3-4): 205-221.

Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E. and House, C. H. (2008). "Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment." Proceedings of the National Academy of Sciences of the United States of America **105**(30): 10583-10588.

Bohorquez, L. C., Delgado-Serrano, L., Lopez, G., Osorio-Forero, C., Klepac-Ceraj, V., Kolter, R., Junca, H., Baena, S. and Zambrano, M. M. (2012). "In-depth Characterization via Complementing Culture-Independent Approaches of the Microbial Community in an Acidic Hot Spring of the Colombian Andes." Microbial Ecology **63**(1): 103-115.

Bonch-Osmolovskaya, E. A., Miroshnichenko, M. L., Slobodkin, A. I., Sokolova, T. G., Karpov, G. A., Kostrikina, N. A., Zavarzina, D. G., Prokof'eva, M. I., Rusanov, I. I. and Pimenov, N. V. (1999). "Biodiversity of Anaerobic lithotrophic prokaryotes in terrestrial hot springs of Kamchatka." Microbiology **68**(3): 343-351.

Brock, T. D. (1967). "Life at High Temperatures - Evolutionary Ecological and Biochemical Significance of Organisms Living in Hot Springs Is Discussed." Science **158**(3804): 1012-&.

Brock, T. D. and Freeze, H. (1969). "Thermus Aquaticus Gen N and Sp N a Nonsporulating Extreme Thermophile." Journal of Bacteriology **98**(1): 289-&.

Burton, N. P. and Norris, P. R. (2000). "Microbiology of acidic, geothermal springs of

Montserrat: environmental rDNA analysis." Extremophiles **4**(5): 315-320.

Chai, B. H. T. (1972). "Structure and Tectonic Evolution of Taiwan." American Journal of Science **272**(5): 389-&.

Cheng, T.-W., Chang, Y.-H., Tang, S.-L., Tseng, C.-H., Chiang, P.-W., Chang, K.-T., Sun, C.-H., Chen, Y.-G., Kuo, H.-C., Wang, C.-H., Chu, P.-H., Song, S.-R., Wang, P.-L. and Lin, L.-H. (2012). "Metabolic stratification driven by surface and subsurface interactions in a terrestrial mud volcano." Isme Journal.

Cheng, T. W., Wang, P. L., Chen, K. C., Wu, J. J., Song, S. R. and Lin, L. H. (2013). "Segregated planktonic and bottom-dwelling archaeal communities in high-temperature acidic, sulfuric ponds of the Tatung Volcano Group, northern Taiwan." Terrestrial, Atmospheric and Oceanic Sciences **(in press)**.

Childs, A. M., Mountain, B. W., O'toole, R. and Stott, M. B. (2008). "Relating Microbial Community and Physicochemical Parameters of a Hot Spring: Champagne Pool, Wai-o-tapu, New Zealand." Geomicrobiology Journal **25**(7-8): 441-453.

Cole, J. R., Chai, B., Marsh, T. L., Farris, R. J., Wang, Q., Kulam, S. A., Chandra, S., Mcgarrell, D. M., Schmidt, T. M., Garrity, G. M. and Tiedje, J. M. (2003). "The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy." Nucleic Acids Research **31**(1): 442-443.

Cox, A., Shock, E. L. and Havig, J. R. (2011). "The transition to microbial photosynthesis in hot spring ecosystems." Chemical Geology **280**(3-4): 344-351.

Delong, E. E. (2005). "Microbial community genomics in the ocean." Nature Reviews Microbiology **3**(6): 459-469.

Delong, E. E. and Pace, N. R. (2001). "Environmental diversity of Bacteria and Archaea." Systematic Biology **50**(4): 470-478.

Desantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G. L. (2006). "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB." Applied and Environmental Microbiology **72**(7): 5069-5072.

Edwards, K. J., Bond, P. L., Gihring, T. M. and Banfield, J. F. (2000). "An archaeal iron-oxidizing extreme acidophile important in acid mine drainage." Science **287**(5459): 1796-1799.

Edwards, R. A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D. M., Saar, M. O., Alexander, S., Alexander, E. C. and Rohwer, F. (2006). "Using pyrosequencing to shed light on deep mine microbial ecology." Bmc Genomics **7**.

Ellis, D. G., Bizzoco, R. L. W., Maezato, Y., Baggett, J. N. and Kelley, S. T. (2005). "Microscopic examination of acidic hot springs of Waiotapu, North Island, New Zealand." New Zealand Journal of Marine and Freshwater Research **39**(5): 1001-1011.

Ferrer, M., Beloqui, A., Timmis, K. N. and Golyshin, P. N. (2009). "Metagenomics for Mining New Genetic Resources of Microbial Communities." Journal of Molecular Microbiology and Biotechnology **16**(1-2): 109-123.

Ferrer, M., Golyshina, O., Beloqui, A. and Golyshin, P. N. (2007). "Mining enzymes from extreme environments." Current Opinion in Microbiology **10**(3): 207-214.

Ferrer, M., Werner, J., Chernikova, T. N., Bargiela, R., Fernandez, L., La Cono, V., Waldmann, J., Teeling, H., Golyshina, O. V., Glockner, F. O., Yakimov, M. M.,

Golyshin, P. N. and Consortium, M. S. (2012). "Unveiling microbial life in the new deep-sea hypersaline Lake Thetis. Part II: a metagenomic study." Environmental Microbiology **14**(1): 268-281.

Ferris, M. J. and Ward, D. M. (1997). "Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis." Applied and Environmental Microbiology **63**(4): 1375-1381.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L. L., Eddy, S. R. and Bateman, A. (2010). "The Pfam protein families database." Nucleic Acids Research **38**: D211-D222.

Hall, J. R., Mitchell, K. R., Jackson-Weaver, O., Kooser, A. S., Cron, B. R., Crossey, L. J. and Takacs-Vesbach, C. D. (2008). "Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes." Applied and Environmental Microbiology **74**(15): 4910-4922.

Handelsman, J. (2004). "Metagenomics: Application of genomics to uncultured microorganisms." Microbiology and Molecular Biology Reviews **68**(4): 669-+.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. and Goodman, R. M. (1998). "Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products." Chemistry & Biology **5**(10): R245-R249.

Hetzer, A., Morgan, H. W., Mcdonald, I. R. and Daughney, C. J. (2007). "Microbial life in champagne pool, a geothermal spring in Waiotapu, New Zealand." Extremophiles **11**(4): 605-614.

Hough, D. W. and Danson, M. J. (1999). "Extremozymes." Current Opinion in Chemical Biology **3**(1): 39-46.

Huang, Q. Y., Dong, C. Z., Dong, R. M., Jiang, H. C., Wang, S., Wang, G. H., Fang, B., Ding, X. X., Niu, L., Li, X., Zhang, C. L. and Dong, H. L. (2011). "Archaeal and bacterial diversity in hot springs on the Tibetan Plateau, China." Extremophiles **15**(5): 549-563.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C. and Stetter, K. O. (2002). "A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont." Nature **417**(6884): 63-67.

Huber, J. A., Mark Welch, D., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A. and Sogin, M. L. (2007). "Microbial population structures in the deep marine biosphere." Science **318**(5847): 97-100.

Hugenholtz, P., Pitulle, C., Hershberger, K. L. and Pace, N. R. (1998). "Novel division level bacterial diversity in a Yellowstone hot spring." Journal of Bacteriology **180**(2): 366-376.

Inskeep, W. P., Rusch, D. B., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Richardson, T. H., Macur, R. E., Hamamura, N., Jennings, R. D., Fouke, B. W., Reysenbach, A. L., Roberto, F., Young, M., Schwartz, A., Boyd, E. S., Badger, J. H., Mathur, E. J., Ortmann, A. C., Bateson, M., Geesey, G. and Frazier, M. (2010). "Metagenomes from High-Temperature Chemotrophic Systems Reveal Geochemical Controls on Microbial Community Structure and Function." Plos One **5**(3).

Ivanova, I., Atanassov, I., Lyutskanova, D., Stoilova-Disheva, M., Dimitrova, D., Tomova, I., Derekova, A., Radeva, G., Buchvarova, V. and Kambourova, M. (2011). "High Archaea diversity in Varvara hot spring, Bulgaria." Journal of

Basic Microbiology **51**(2): 163-172.

Jackson, C. R., Langner, H. W., Donahoe-Christiansen, J., Inskeep, W. P. and Mcdermott, T. R. (2001). "Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring." Environmental Microbiology **3**(8): 532-542.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T. and Yamanishi, Y. (2008). "KEGG for linking genomes to life and the environment." Nucleic Acids Research **36**: D480-D484.

Kanokratana, P., Chanapan, S., Pootanakit, K. and Eurwilaichitr, L. (2004). "Diversity and abundance of Bacteria and Archaea in the Bor Khlueng hot spring in Thailand." Journal of Basic Microbiology **44**(6): 430-444.

Kato, S., Itoh, T. and Yamagishi, A. (2011). "Archaeal diversity in a terrestrial acidic spring field revealed by a novel PCR primer targeting archaeal 16S rRNA genes." Fems Microbiology Letters **319**(1): 34-43.

Kawashima, E., Farinelli, L. and Mayer, P. (1998). Method of nucleic acid amplification.

Konstantinou, K. I., Lin, C. H. and Liang, W. T. (2007). "Seismicity characteristics of a potentially active Quaternary volcano: The Tatun Volcano Group, northern Taiwan." Journal of Volcanology and Geothermal Research **160**(3-4): 300-318.

Kristjansson, J. K. and Hreggvidsson, G. O. (1995). "Ecology and Habitats of Extremophiles." World Journal of Microbiology & Biotechnology **11**(1): 17-25.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K. and Hugenholtz, P. (2008). "A Bioinformatician's Guide to Metagenomics." Microbiology and Molecular

Biology Reviews **72**(4): 557-578.

Lamoreaux, P. E. and Tanner, J. T. (2001). Springs and Bottled Waters of the World: Ancient History, Source, Occurrence, Quality and Use Springer.

Lau, M. C. Y., Aitchison, J. C. and Pointing, S. B. (2009). "Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet." Extremophiles **13**(1): 139-149.

Lee, H. F., Yang, T. F., Lan, T. F., Song, S. R. and Tsao, S. (2005). "Fumarolic gas composition of the Tatun Volcano Group, northern Taiwan." Terrestrial Atmospheric and Oceanic Sciences **16**(4): 843-864.

Lin, C. H., Konstantinou, K. I., Liang, W. T., Pu, H. C., Lin, Y. M., You, S. H. and Huang, Y. P. (2005). "Preliminary analysis of volcanoseismic signals recorded at the Tatun Volcano Group, northern Taiwan." Geophysical Research Letters **32**(10).

Liu, C. M., Song, S. R., Chen, Y. L. and Tsao, S. (2011). "Characteristics and Origins of Hot Springs in the Tatun Volcano Group in Northern Taiwan." Terrestrial Atmospheric and Oceanic Sciences **22**(5): 475-489.

Lu, H. P., Lin, Y. K., Lin, L. H., Wang, P. L. and Yu, H. T. (unpublished data). The microbial community structure in acidic hot springs.

Macelroy, R. D. (1974). "Some Comments on Evolution of Extremophiles." Biosystems **6**(1): 74-75.

Macur, R. E., Langner, H. W., Kocar, B. D. and Inskeep, W. P. (2004). "Linking geochemical processes with microbial community analysis: successional dynamics in an arsenic-rich, acid-sulphate-chloride geothermal spring." Geobiology **2**(3): 163-177.

Mardis, E. R. (2008). "Next-generation DNA sequencing methods." Annual Review

of Genomics and Human Genetics **9**: 387-402.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z. T., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., Mcdade, K. E., Mckenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P. G., Begley, R. F. and Rothberg, J. M. (2005). "Genome sequencing in microfabricated high-density picolitre reactors." Nature **437**(7057): 376-380.

Marteinsson, V. T., Hauksdottir, S., Hobel, C. F. V., Kristmannsdottir, H., Hreggvidsson, G. O. and Kristjansson, J. K. (2001). "Phylogenetic diversity analysis of subterranean hot springs in Iceland." Applied and Environmental Microbiology **67**(9): 4242-4248.

Mathur, J., Bizzoco, R. W., Ellis, D. G., Lipson, D. A., Poole, A. W., Levine, R. and Kelley, S. T. (2007). "Effects of abiotic factors on the phylogenetic diversity of bacterial communities in acidic thermal springs." Applied and Environmental Microbiology **73**(8): 2612-2623.

Maxam, A. M. and Gilbert, W. (1977). "A new method for sequencing DNA." Proceedings of the National Academy of Sciences **74**(2): 560-564.

Metzker, M. L. (2010). "Applications of Next-Generation Sequencing Sequencing Technologies - the Next Generation." Nature Reviews Genetics **11**(1): 31-46.

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

Mitchelson, K. (2001). Overview: The Application of Capillary Electrophoresis for DNA Polymorphism Analysis. Capillary Electrophoresis of Nucleic Acids. K. Mitchelson and J. Cheng, Humana Press. **162:** 3-26.

Mullis, K. B. and Faloona, F. A. (1987). "Specific Synthesis of DNA Invitro Via a Polymerase-Catalyzed Chain-Reaction." Methods in Enzymology **155**: 335-350.

Nakagawa, S., Takai, K., Inagaki, F., Chiba, H., Ishibashi, J., Kataoka, S., Hirayama, H., Nunoura, T., Horikoshi, K. and Sako, Y. (2005). "Variability in microbial community and venting chemistry in a sediment-hosted backarc hydrothermal system: Impacts of subseafloor phase-separation." Fems Microbiology Ecology **54**(1): 141-155.

Nakagawa, T. and Fukui, M. (2002). "Phylogenetic characterization of microbial mats and streamers from a Japanese alkaline hot spring with a thermal gradient." Journal of General and Applied Microbiology **48**(4): 211-222.

Ng, C. C., Chang, C. C. and Shyu, Y. T. (2005). "Archaeal community revealed by 16s rRNA and fluorescence in situ hybridization in a sulphuric hydrothermal hot spring, northern Taiwan." World Journal of Microbiology & Biotechnology **21**(6-7): 933-939.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., De Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B.,

Mchardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O. and Vonstein, V. (2005). "The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes." Nucleic Acids Research **33**(17): 5691-5702.

Pace, N. R. (1997). "A molecular view of microbial diversity and the biosphere." Science **276**(5313): 734-740.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W. G., Peplies, J. and Glockner, F. O. (2007). "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB." Nucleic Acids Research **35**(21): 7188-7196.

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

Reigstad, L. J., Jorgensen, S. L. and Schleper, C. (2010). "Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka." Isme Journal **4**(3): 346-356.

Reysenbach, A. L., Ehringer, H. and Hershberger, K. (2000). "Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist." Extremophiles **4**(1): 61-67.

Reysenbach, A. L., Giver, L. J., Wickham, G. S. and Pace, N. R. (1992). "Differential Amplification of Ribosomal-Rna Genes by Polymerase Chain-Reaction." Applied and Environmental Microbiology **58**(10): 3417-3418.

Reysenbach, A. L., Wickham, G. S. and Pace, N. R. (1994). "Phylogenetic Analysis of the Hyperthermophilic Pink Filament Community in Octopus Spring,

Yellowstone-National-Park." Applied and Environmental Microbiology **60**(6): 2113-2119.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. and Nyrén, P. (1996). "Real-Time DNA Sequencing Using Detection of Pyrophosphate Release." Analytical Biochemistry **242**(1): 84-89.

Rothschild, L. J. and Mancinelli, R. L. (2001). "Life in extreme environments." Nature **409**(6823): 1092-1101.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). "DNA Sequencing with Chain-Terminating Inhibitors." Proceedings of the National Academy of Sciences of the United States of America **74**(12): 5463-5467.

Satyanarayana, T., Raghukumar, C. and Shivaji, S. (2005). "Extremophilic microbes: Diversity and perspectives." Current Science **89**(1): 78-90.

Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L., Yaschenko, E. and Ye, J. (2009). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Research **37**: D5-D15.

Schmidt, T. M., Delong, E. F. and Pace, N. R. (1991). "Analysis of a Marine Picoplankton Community by 16s Ribosomal-Rna Gene Cloning and Sequencing." Journal of Bacteriology **173**(14): 4371-4378.

Shokralla, S., Spall, J. L., Gibson, J. F. and Hajibabaei, M. (2012). "Next-generation sequencing technologies for environmental DNA research." Molecular

Ecology **21**(8): 1794-1805.

Simon, C. and Daniel, R. (2011). "Metagenomic Analyses: Past and Future Trends." Applied and Environmental Microbiology **77**(4): 1153-1161.

Sleator, R. D., Shortall, C. and Hill, C. (2008). "Metagenomics." Letters in Applied Microbiology **47**(5): 361-366.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., 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 United States of America **103**(32): 12115-12120.

Song, S. R., Yang, T. F., Yeh, Y. H., Tsao, S. J. and Lo, H. J. (2000). "The Tatun Volcano Group is active or extinct?" Journal of Geological Society of China **43**: 521-534.

Stein, L. D. (2011). An Introduction to the Informatics of "Next-Generation" Sequencing. Current Protocols in Bioinformatics, John Wiley & Sons, Inc.

Takai, K., Komatsu, T., Inagaki, F. and Horikoshi, K. (2001). "Distribution of archaea in a black smoker chimney structure." Applied and Environmental Microbiology **67**(8): 3618-3629.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. and Koonin, E. V. (2001). "The COG database: new developments in phylogenetic classification of proteins from complete genomes." Nucleic Acids Research **29**(1): 22-28.

Tekere, M., Lotter, A., Olivier, J., Jonker, N. and Venter, S. (2011). "Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa." African Journal of Biotechnology **10**(78): 18005-18012.

Temperton, B., Field, D., Oliver, A., Tiwari, B., Muhling, M., Joint, I. and Gilbert, J.

A. (2009). "Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing." Isme Journal **3**(7): 792-796.

Todd, D. K. (1980). Groundwater Hydrology. New York, Wiley.

Tomova, I., Stoilova-Disheva, M., Lyutskanova, D., Pascual, J., Petrov, P. and Kambourova, M. (2010). "Phylogenetic analysis of the bacterial community in a geothermal spring, Rupi Basin, Bulgaria." World Journal of Microbiology & Biotechnology **26**(11): 2019-2028.

Tringe, S. G., Von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholtz, P. and Rubin, E. M. (2005). "Comparative metagenomics of microbial communities." Science **308**(5721): 554-557.

Turnbaugh, P. J. and Gordon, J. I. (2008). "An invitation to the marriage of metagenomics and metabolomics." Cell **134**(5): 708-713.

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.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D. Y., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H. and Smith, H. O. (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." Science **304**(5667): 66-74.

Von Mering, C., Hugenholtz, P., Raes, J., Tringe, S. G., Doerks, T., Jensen, L. J., Ward, N. and Bork, P. (2007). "Quantitative phylogenetic assessment of microbial

communities in diverse environments." Science **315**(5815): 1126-1130.

Wang, W. S. and Chen, C. H. (1990). "The volcanology and fission track age dating of pyroclastic deposits in Tatun Volcano Group, northern Taiwan." Acta Geologica Taiwanica **28**: 1-30.

Weisburg, W. G., Barns, S. M., Pelletier, D. A. and Lane, D. J. (1991). "16s Ribosomal DNA Amplification for Phylogenetic Study." Journal of Bacteriology **173**(2): 697-703.

Wilson, M. S., Siering, P. L., White, C. L., Hauser, M. E. and Bartles, A. N. (2008). "Novel archaea and bacteria dominate stable microbial communities in North America's largest hot spring." Microbial Ecology **56**(2): 292-305.

Wilson, Z. E. and Brimble, M. A. (2009). "Molecules derived from the extremes of life." Natural Product Reports **26**(1): 44-71.

Woese, C. R. (1987). "Bacterial Evolution." Microbiological Reviews **51**(2): 221-271.

Yamamoto, H., Hiraishi, A., Kato, K., Chiura, H. X., Maki, Y. and Shimizu, A. (1998). "Phylogenetic evidence for the existence of novel thermophilic bacteria in hot spring sulfur-turf microbial mats in Japan." Applied and Environmental Microbiology **64**(5): 1680-1687.

Yang, T. F., Sano, Y. and Song, S. R. (1999). "$^{3}He/^{4}He$ ratio of fumaroles and bubbling gases of hot spring in Tatun Volcano Group, North Taiwan." Nuovo Cimento **22**(0304): 281-286.

Yim, L. C., Jing, H. M., Aitchison, J. C. and Pointing, S. B. (2006). "Highly diverse community structure in a remote central Tibetan geothermal spring does not display monotonic variation to thermal stress." Fems Microbiology Ecology **57**(1): 80-91.

何珮楨 (2004). 台灣地區不同溫泉中微生物組成的研究. 碩士, 東吳大學.

李信緯 (2009). 陽明山國家公園溫泉微生物多樣性分析. 碩士, 東吳大學.

李重義 (2006). 陽明山國家公園溫泉微生物多樣性與利用之研究. 陽明山國家公園管理處. 臺北市.

李重義 and 張怡塘 (2009). 陽明山國家公園酸性溫泉物種調查及生存機制之研究. 陽明山國家公園管理處. 臺北市.

陳懋彥 (2002). 台灣地熱區嗜熱性細菌之研究. 博士, 國立臺灣大學.

鄭婷文 and 林立虹 (2007). 酸性熱泉中底生性與懸浮性微生物族群結構之比較. 2007 台灣地球科學聯合學術研討會.

2. Materials and Methods

2.1 Sample Collection

The spring water for fosmid library construction was collected from 四磺坪 (abbreviated as SHP) on 3rd, 10th and 21st in December 2010, 8th February and 17th in March 2011. Volumes of collected water samples were 18, 36, 36, 36 and 36 liters, respectively. Average temperature was 64.2°C and average pH was 3. The hot spring water was collected directly by using sterile polypropylene (PP) containers or being pumped through sterile tubing by a peristaltic pump. Water samples in sterile PP containers (Vitlab, Germany) were then transported for 1 hour under ambient temperature (25 to 30°C, depending on season) and were stored at 4°C afterwards waiting for filtration and concentration.

For direct shotgun sequencing 40 liters of spring water were collected from the same location (SHP) on April 5th, 2012. The same conditions and procedures for collecting, transportation and storage were applied as mentioned above.

2.2 Sample Concentration

The tangential flow system (Giovannoni et al., 1990) was used in the concentration process to remove excessive spring water. The system consisted of a peristaltic pump (Masterflex® E/S™ Portable Sampler, Cole-Parmer, USA), tubing (C-Flex® 06424-24, Cole-Parmer, USA) and a hollow fiber cartridge (Hollow Fiber Cartridge CFP-2-E-3MA, GE Healthcare, USA) with pore size 0.2 μm in diameter (Figure 2). The sterile tubing was fixed to the

peristaltic pump. One end was put into the container containing the water sample and the other end was connected to a pressure gauge. The other side of the gauge was connected to the cartridge with a short sterile silicon tube. The cartridge had 3 connectors: The feed stream end, the retentate end and the permeate end. The water sample flowed from the PP bottle into the cartridge through the feed stream end and microorganisms as well as particles larger than the pore size of the filter flowed back to the bottle as retentate and stayed in the system. The excessive water and particles smaller than the pore size of the filter were discarded as filtrate.

The concentration procedure was carried out under room temperature and included three steps: Starting, concentration and finishing.

Starting

The new cartridge was primed with a rinsing process according to the user manual. Later in every single operation after the concentration system was set up (Figure 2), the whole system was flushed with deionized water at full speed of peristaltic pump until the filtrate and retentate are of pH value 7.

Concentration

The deionized water was replaced with the sample water which was to be filtered. The peristaltic pump was turned on and pressure applied to retentate was maintained at 5 psi throughout the process. (If clotting occurred during the concentration process, the system was temporarily stopped. After applying the finishing procedure described below, the starting procedure was performed

again before resuming the concentration process.)

Finishing

All permeate ends were closed and the system was flushed with 500 mL of filtrate to obtain the remaining cells (this retentate was collected with concentrated spring water). The bottle with the concentrated sample was removed. Then the system was cleaned and sterilized by flushing 2 liters of 0.5 N NaOH. For the first 1.5 liter, all permeate ends were closed. Later, one permeate end was opened and the flushing continued. Another 2 liters of 0.1 N NaOH was applied using the same procedure, but with one permeate end open. Since the cartridge needed to be filled with 0.1 N NaOH for preservation, the pump was stopped and all permeate ends were closed before the NaOH was used up. All components of the system were disassembled carefully and caps to both ends of the cartridge were applied. To prevent leaking of NaOH, a parafilm was utilized. The cartridge was stored at room temperature away from sunlight.

2.3 DNA Extraction

The resulting concentrated water sample was usually about 500 mL after each filtration operation (initiated using approximately 40 liters of spring water collected in field), which was not suitable for conventional DNA extraction procedures to be applied. Thus, further deduction of extra spring water was necessary and two different methods were tested for optimizing the efficiency of this secondary concentration prior to DNA extraction.

A. Vacuum Filtration

Vacuum filtration was completed at room temperature within one day. A cellulose acetate filter with pore size 0.2 μm (Sartorius GmbH, Germany) was utilized during the whole process. Depending mostly on the amount of suspension particles in concentrated spring water, up to six filters were required to complete the filtration process. Used filter membranes were temporarily stored inside sterile tubes at 8°C. Membranes were then fragmented under aseptic conditions in laminar flow cabinet by sterile stainless steel scissor with each piece being smaller than 0.5 cm^2. Fragmented membranes were treated as sediments and extracted by UltraClean® Mega Soil DNA Isolation Kit (MO BIO Laboratories, Inc. (USA)) according to the provided user manual. The amount of extracted DNA was estimated through concentration measurements by fluorescence meter (Qubit® 2.0 Fluorometer, Invitrogen (USA)). Afterwards the DNA solution was frozen at -20°C.

B. High-Speed Centrifugation

The concentrated spring water (c.a. 500 mL) was divided into 4 bottles for the centrifugation in a high-speed refrigerated centrifuge (Himac CR-21, Hitachi (Japan)). Samples were centrifuged at 6000 rpm for 30 minutes, with acceleration speed and deceleration speed adjusted to 9 and 4, respectively. An adequate amount of supernatant was kept for re-suspension of pallets into a muddy solution (a mixture of suspension particles, fine sediments and microorganisms). The muddy solution was

then pooled together into a bottle (weight was measured before use), and the same centrifugation process was repeated again for another 30 minutes. Supernatant were discarded and sediments were weighted. According to the suggestions in the user manual of UltraClean® Mega Soil DNA Isolation Kit, an appropriate number of kits were used to extract the DNA with standard procedures. The yield of recovered environmental DNA was then estimated by a concentration measurement using the fluorescence meter.

2.4 Fosmid Library Construction and Random Shotgun Sequencing

2.4.1 Fosmid Library Construction

The sufficient amount of DNA derived from hot spring water at the SHP area was collected and accumulated for fosmid library construction. Procedures were based on manufacturer's instructions (CopyControl™ HTP Fosmid Library Production Kits, Epicentre® Biotechnologies). Some steps were slightly modified. According to the manual, the whole process to construct a fosmid library could be briefly summarized into the following steps: (1) Shearing the insert DNA, (2) end-repair of the insert DNA, (3) size selection of the end-repaired DNA, (4) recovery of the size-fractionated DNA, (5) ligation reaction, (6) packaging of the CopyControl fosmid clone, (7) titering the packaged CopyControl fosmid clones and (8) plating and selecting the CopyControl fosmid library.

The first step was skipped since previous experience suggested the extracted

DNAs were already fragmented at some point. In the testing period, the size selection was bypassed to avoid the further loss of DNA samples. However, due to the negative effect on the final result, the size selection was then incorporated back into the standard operation. The seventh step was not performed since the density of plated colonies was already ideal.

Randomly selected colonies were chosen and checked for insert size through pulse-field gel electrophoresis (PFGE). Each selected colony was prepared as 5-mL overnight culture and was centrifuged (Himac CR-21, Hitachi (Japan)) at 6000 rpm, 4 °C for 15 min. Supernatant was discarded and 250 μL Buffer P1 (from QIAGEN Plasmid Mini Kit, QIAGEN) were added for cell resuspension by vortex. 250 μL of Buffer P2 (from QIAGEN Plasmid Mini Kit, QIAGEN) were then added. Tubes were inverted 6 times and 250 μL ice-cold Buffer P3 were added. After 6 times of tube inverting, reactions were put on ice for 30 min. Tubes were then centrifuged at 13krpm in 4°C for 30 min. Supernatant was transferred to new tubes and isopropanol (with the amount which is equal to 60% of supernatant in volume) was added into reactions and well-mixed. All reactions were centrifuged at 13krpm in 4°C for 30 min. Supernatant was discarded and 500 μL of 70% EtOH were added to wash the pellets. Tubes were centrifuged at 13krpm in 4°C for 10 min. Supernatant was discarded and pellets were air dried. DNAs were resolved by adding 20 μL of sterile water into each tube.

The resolved DNAs were digested with restriction enzyme *Not*I. The reaction volume was 12 μL, with 10 μL of DNA (could be diluted), 1.2 μL of 10X No.3

Buffer (NEB), 0.12 μL of 100X BSA, 0.57 μL of dH_2O and 0.2 μL of *Not*I. After 16 to 18 hours of incubation in 37°C, PFGE was performed to confirm the insert size of fosmid clones. The PFGE system consisted of a gel chamber, a Standard Power Pack (both are of Rotaphor® System, Biometra) and a circulator tank (Refrigerated Circulator RCB411, TKS) for cooling. A gel running program was set with interval of 2 sec; the angle was 130° to 110°, linear; 130V to 90V in voltage, log; 10°C in temperature; 16 hours of running time and "rotate with power" function was turned on. 1% agarose gel was made and *Not*I digested fosmid DNA samples were tested in the system with 1/3X Loening Buffer diluted from stock solution (10X) as running buffer.

After the confirmation of insert size, a fosmid library MG-HSTL was constructed (#1 to #9558, 9481 viable clones in total) and deposited in the Food Industry Research and Development Institute (FIRDI, Taiwan), and to be publicly available since Aug 1, 2013.

2.4.2 Random Shotgun Sequencing

Random Shotgun Sequencing was performed with HiSeq™ 2000 (Illumina) in this study. Two samples were sent: one was 1485 clones (of total 9481 clones) from fosmid library, and the other was metagenomic DNA directly extracted from hot spring water. Fosmid DNAs of randomly picked clones were cultured, induced and extracted with the alkaline lysis method in National Yang-Ming University VYM Genome Research Center. The concentration of fosmid DNAs was measured using a fluorescence meter and then mixed into the bulk sample for sequencing with an equal amount of DNA derived from each clone.

The second sample (abbreviated into DSS, from Direct Shotgun Sequencing) was prepared as follows: 40 liters of hot spring water was later collected using the same method mentioned previously on April 5, 2012. DNAs were extracted, purified and checked for concentration and amount prior to sequencing.

Samples were processed in Yourgene Bioscience Co., Ltd. (Taiwan) using the standard procedure: (1) Sample quality check, (2) DNA library construction (fragmentation, end-repair, A tailing, adapter ligation, PCR), (3) library quality check (size and concentration check), (4) sequencing and (5) bioinformatic processing (assembly, customized solution).

2.5 Information Processing and Data Analysis

2.5.1 Sequence Assembly

Raw reads generated were trimmed by criteria of 35 bp at minimum length and error probability lower than 0.05. Removal of vector sequence was applied to sample from fosmid library. Raw reads of fosmid library were assembled into contigs with CLC (CLC bio, Denmark) and MetaVelvet (Namiki et al., 2012); assembly of DSS contigs was accomplished only by MetaVelvet.

2.5.2 Analysis on Community Structure

Community structure was analyzed through strategies that focused on 4 different objects:

A. Coverage of Contigs Containing 16 S rDNA sequence

16 S sequences in fosmid clones and DSS were predicted by online blastn_rRNA and hmm_rRNA programs in WebMGA server (Wu et al., 2011), using 5s Ribosomal Database and European rRNA database that were latest updated on Oct 9, 2009. Primary results were examined and 16 S rRNA sequences were then uploaded and classified with RDP Classifier (Cole et al., 2003; Lan et al., 2012). For accuracy, results were compared with another independent operation, which aligned both sets of contigs (blastn criteria: e-value cutoff was 10^{-5}, aligned length of query should be ≥ 200 bp) against SILVA rRNA database (Pruesse et al., 2007) version 106. Top 5 hits of each query sequence were picked and processed by LCA algorithm for determination of the least common ancestor (Harel and Tarjan, 1984), through which a taxonomic name was generated and assigned to each of the contigs from fosmid clones and DSS sample.

In order to estimate the relative abundance of each taxonomic group, contigs with reliable taxonomic identities (e-value < 10^{-10}, derived from RDP classifier) were used as references for raw reads blastn-mapping. Raw reads were filtered by identity higher than 95%, query sequence length longer than 95% of itself and e-value less than 10^{-10} to reference contigs. In each blast step of a raw read, final hit with highest identity was picked (multiple hits were picked if identity were equal). Each contig hit by a single raw read then acquired a score normalized by number of hit (N) and contig length (L):

$$S_{16s-ctg} = \frac{1}{N * L}$$

Where:

$S_{16\ S\text{-}ctg}$ is score assigned to a specific contig after a single blast; N is the number of top hit(s) obtained from a blast process of a single raw read against 16 S-containing contigs; L is the length of the contig hit by the raw read.

Supposed a raw read had 3 top hits on contig A, B and C. Contig A would get a score in this blast process:

$$\frac{1}{3 * length\ of\ contig\ A}$$

After blast of all raw reads was complete, final score of each genus was merged form scores of contigs with same genus names.

B. Raw Reads Bearing 16 S rDNA Sequence

Two ends of DSS raw reads were separately mapped to SILVA SSU reference database (version 111) with criteria set as following: (1) Sequence identity higher than 80%, (2) aligned query sequence length longer than 90% of itself and (3) e-value less than 10^{-10}. The top hit with highest identity of a single read was counted (multiple hits with the same identity were all picked). Each count assigned to a reference 16 S rDNA sequence hit by a raw read was weighted as:

$$S_{SILVA} = \frac{1}{N * L}$$

Where:

S_{SILVA} is the score assigned to a specific reference 16 S rDNA from SILVA database after a single blast of a raw read; N is the number of top hit(s) obtained from a blast process of a single raw read against SILVA database; L is the length of 16 S rDNA sequence hit by the raw read.

Supposed a raw read had 4 top hits on 16 S rDNA sequences A, B, C and D. Sequence B in this single blast process would get a score:

$$\frac{1}{4 * length\ of\ sequence\ B}$$

After blast of all raw reads was complete, final score of each genus was merged from scores of sequences with identical names in SILVA database.

C. Reference Genomes Mapping

Raw reads from DSS were mapped (using nucleotide blast) against IMG genome database (The Integrated Microbial Genomes system) of the U.S. Department of Energy Joint Genome Institute (DOE JGI). Mapping conditions were set as following: longer than 90% length of query sequence could be mapped to reference and the identity should be higher than 80%; e-value should be less than 10^{-10} and both ends of a read must hit to the same reference genome. Each reference hit by a raw read obtains

a score calculated as described:

$$\frac{L_r}{L_{max}}$$

Where:

L_r stands for length of raw read (in this case, it is a constant 98) and L_{max} is the longest genome size among the least common ancestor (LCA) of the hit reference.

For instance, suppose there were 50 strains of *E.coli* were hit during the whole mapping operation, and the LCA would be genus *Escherichia*. We picked the longest genome among genus *Escherichia* and set its genome size as L_{max} for all *E. coli* strain which was mapped by raw reads. Bias possibly caused by unequal number of taxa information deposit in JGI was not eliminated.

D. Coverages of Contigs

Firstly, each DSS contigs were mapped using nucleotide blast (cutoff: 10^{-5}) against manually filtered NR database of NCBI (complete database was downloaded on June 22, 2012) that contained only prokaryotes. LCA was performed on top 5 hits of each contig (sorted by e-value) to assign appropriate taxa names. Secondly, coverage information of each contig derived from the assembly by MetaVelvet were extracted and applied to each contig and relative abundance of taxa was calculated.

2.5.3 Analysis on *Hydrogenobaculum*-like Sequences

A. Mapping and *de novo* assembly of *Hydrogenobaculum*-like Sequences

DSS raw reads was mapped to reference *Hydrogenobaculum* sp. Y04AAS1 genome (Reysenbach et al., 2009) by Bowtie2 (Langmead and Salzberg, 2012) with default settings. De novo assembly was then done by SOAPdenovo (Li et al., 2010) with default settings.

Mapped DSS raw reads to reference genome and contigs assembled by SOAPdenovo were visualized by Circos (Krzywinski et al., 2009). To compare the completeness of sequence recovery, contigs from fosmid library and DSS contigs were also mapped to *Hydrogenobaculum* sp. Y04AAS1 reference genome using nucleotide blast. Criteria were: e-value lower than 10^{-10} and alignable query length longer than 90% of contig itself. Supposed that there was a contig mapped to multiple regions of reference genome, it was then selected by highest blastn score to decide which region was mapped by the contig. If multiple mapped regions shared the same score, the same contig was to assign to those regions.

B. Relationship between GC content and coverage

DSS raw reads mapped to *Hydrogenobaculum* sp. Y04AAS1 reference genome were analyzed for the relationship between sequence coverage and GC-content. Estimation strategy was to set a 1000 bp window started from the beginning of reference genome. Each time a 50 bp slide was made and average GC-content and coverage in the 1000 bp window were calculated and plotted.

2.5.4 Overview on Gene Functions of the Hot Spring Community

On WebMGA server, ORF prediction by metagene (Noguchi et al., 2006) was applied to DSS contigs and results were subsequently analyzed and annotated by COG classification (e-value: 10^{-5}). Each category of COG count was normalized with contig coverage information from MetaVelvet assembly. COG distribution pattern of DSS sample were compared to that of *Hydrogenobaculum* sp. Y04AAS1 reference genome in JGI database.

3. Results

3.1 Samples for Fosmid Library Construction and Direct Shotgun Sequencing

Because of sample accessibility and environmental conditions suitable for thermophiles, a hot spring in 四磺坪 (abbreviated as SHP) area was chosen as the metagenomic DNA source for fosmid library construction and also served as sample which was directly shotgun-sequenced.

3.3.1 Physicochemical Properties of the Hot Spring in SHP

Since the hydrothermal spring in SHP area was the object in this study, measurements of environmental parameters was done and compared with previous research in the same area (Table 4, Appendix 2), but exact sampling locations in reference documents were uncertain. In general, pH value remained stably around 3 but temperatures fluctuated from 65 to 78°C depending on locations of sampling. Recent measurements revealed high concentration of sulfate (approximately 350 mg/L). Concentrations of Cl^-, HCO_3^-, Ca^{2+}, Mg^{2+}, K^+ and Na^+ were low , consistent with descriptions of Type I hot spring water suggested by Liu et al. (2011). Unlike some other hot springs in this area (Cheng et al., 2013), concentration of total As in SHP was low.

3.3.2 Fosmid Library

Fosmid library was established with accumulated metagenomic DNAs extracted from SHP area spanned from May 2010 to March 2011, containing viable 9481 clones of 9558 clones in total. The complete fosmid library was

deposited at Bioresource Collection and Research Center of FIRDI (The Food Industry Research and Development Institute, Taiwan) with BCRC number g1067 and will become available to the public after Aug 2013 (Appendix 3).

3.2 Random Shotgun Sequencing

Metagenomic DNAs for direct shotgun sequencing (DSS) was extracted from concentrated hot spring water in SHP (original volume was 40 liters). A total 118.5 μg of DNAs were extracted but only 28 μg left after final concentration.

Pair ends random shotgun sequencing on 1485 fosmid clones and DSS sample have generated 20G (each end) and 50G (each end) of data for fosmid clones and DSS, respectively (Table 5a and 6a). *De novo* assembly of both data sets were done by MetaVelvet and generated 52,221 and 126,849 contigs from fosmid clones and DSS data (Table 5b and 6b). Distribution pattern of contig length of both samples is plotted in Figure 3. Fosmid contig length (559 bp) on average was slightly longer than that of the DSS sample (478 bp); however, percentage of contigs longer than 300bp was quite close for both: 30.29% of total fosmid clones and 29.36% of DSS.

3.3 Analyses on Community Structure

Community structure was analyzed through 4 approaches that estimated the relative abundance by: (1) coverage of contigs containing 16 S rDNA sequences; (2) raw reads containing 16 S rDNA sequences; (3) mapping raw

reads to reference genomes and (4) coverage of all annotated contigs (Figure 4). In each approach, percentages of reads being used for relative abundance estimate were listed in Table 7.

16 S (SSU) rRNA gene sequence similarity has long been a standard method for phylogenetic analysis and taxa identification. Two strategies, coverage of contigs containing 16 S rDNA sequences (abbreviated as DSS > 16 SCTG) and raw reads containing 16 S rDNA sequences (abbreviated as DSS > SILVA), yielded similar community structure which suggested a single dominant role (over 80%) played by deep-branched bacterial phylum Aquificae, followed by 10% of archaeal phylum Crenarcheota and another 2% of Nanoarchaeota. The proportion represented by Proteobacteria ranged between 1 to 4%.

Free from the constraint that only 16 S rRNA information was available (since information from other genomic sequences were discarded), IMG database containing 4631 prokaryotic genomes (4455 of bacteria, 176 of archaea; both included finished and draft genomes) was set as reference database for nucleotide blast for fosmid and DSS raw reads. In the results from fosmid clones, since *E. coli* made up over 73% of the community (Appendix 4) and was suspected to be contaminated by fosmid host genomic DNA during the extraction process, only analytical results of DSS was used in further analyses and discussions. In the result from mapping DSS raw reads onto IMG reference genomes (DSS > IMG), bacteria was dominant over archaea with 9:1 ratio, similar to results generated by DSS > 16sCTG and DSS > SILVA (Figure 5b). Bacterial phylum Aquificae (genus *Hydrogenobaculum*) was

dominant, comprising 67% in abundance of the community. Phyla Proteobacteria, Firmicutes and Tenericutes were represented by 12%, 7% and 1% in the abundance, respectively. Archaea were represented by 10% of the community with 4 phyla being detected: Crenarchaeota (8%), Euryarchaeota (1%), Thaumarchaeota (0.2%) and Nanoarchaeota (0.07%)(Table 8).

In order to extract most of the information from DSS contigs, resource from NR database of NCBI was utilized for recognizing taxa identities of contig. By applying contig coverage to each taxa, this process (abbreviated as DSS > tCTG, since the operation is equal to map raw reads onto total annotated contigs) revealed a less abundant but still dominant role of phylum Aquificae (43%) within the community, followed by 24% of unassigned bacteria, 20% of Crenarcheota, 6% of unassigned sequences and 2% of each unassigned archaea, Firmicutes and Proteobacteria.

To sum up, 4 approaches all suggested prokaryotic community in SHP hot spring was bacteria-dominant with bacteria / archaea ratio around 9 to 1 (revealed by DSS > 16 SCTG, DSS > SILVA and DSS > IMG) and 3 to 1 (revealed by DSS > tCTG; 71% of bacteria, 23% of archaea and 6% unassigned). Although different approaches might implicate different relative abundance, bacterial phylum Aquificae (specifically, genus *Hydrogenobaculum*) was still dominant by representing at least 43% of the community; aside from unassigned sequences, phylum Crenarchaeota was the most abundant phylum among archaea and the second abundant member in this community by at least 8%, followed by Proteobacteria (at least 2%) and

Firmicutes (at least 2%).

Species with lower abundance were also detected in the community structure assessment in addition to predominant members (Table 8). All phyla of archaea were found in this study, including Crenarcheota, Euryarchaeota, Nanoarchaeota, Thaumarchaeota and Korarchaeota. A few phototrophic bacteria such as Cyanobacteria, Chloroflexi and Chlorobi were also present. Sequences affiliated with archaeal genus *Ignicoccus*, one of them was reported to be host of *Nanoarchaea equitans*, were also detected by DSS > SILVA, DSS > IMG and DSS > tCTG, although the abundance was very low, no higher than 0.14% (0.00005 to 0.003% in DSS > SILVA; 0.14% in DSS > IMG; 0.004% in DSS > tCTG, data not shown).

3.4 Analysis on *Hydrogenobaculum*-like Sequences

In reference genome mapping, 7,164,045 DSS raw reads (1.1% of total reads) were mapped to *Hydrogenobaculum* sp. Y04AAS1 (genome size 1,559,514 bp) with overall coverage 463.97. Status of coverage varied according to positions (Figure 5), and the two copies of rRNA operons possessed the highest coverage (around 15,000 to 24,000) among all.

Contigs of 3 sources were mapped to reference genome to evaluate the completeness of *Hydrogonobaculum*-like sequences recovered after random shotgun sequencing. Contigs assembled by MetaVelvet were more equally mapped to reference than the other two contig sets. Similar with coverage

status of DSS raw read, higher overlap of contigs can also be found at regions of rRNA operon. Contigs from fosmid clones tended to aggregate in several positions; GC-content of these regions averaged 35%, which is congruent with the GC-content of *Hydrogenobaculum* sp. Y04AAS1. In contrast, the GC-content in rRNA operons was higher (c.a. 55%) than the genome average.

Analysis on GC content of DSS raw reads that mapped to reference *Hydrogenobaculum* sp. Y04AAS1 genome was done and plotted with the mapping coverage through the reference genome to examine if the coverage of mapping was affected by GC content of reads (Figure 6). Vast majority of data point fell in the region with coverage lower than 4000, GC content ranged from 25 to 45% and an average around 35%. However, some data points also aggregated at upper right region which had relatively high GC content (an average close to 55%) and high mapping coverage (around 20000). In addition, although most of the data points gathered at the lower left quadrant, the rest of points displayed a trend that suggested a positive relation between GC content and mapping coverage.

3.5 Gene Function Distribution of the Hot Spring Community

Gene function of the community was assessed by protein COG analysis (Clusters of Orthologous Groups of proteins, COGs)(Figure 7). Two sets of contigs, DSS contigs (DSS) and DSS subset *Hydrogenobaculum*-contigs (DSS Hb), and COG information of reference *Hydrogenobaculum* sp. Y04AAS1 were compared with each other. In search of the *Hydrogenibaculum*-like

sequence missed by sequencing process, COGs of Y04AAS1 and DSS Hb were compared. To probe into the gene function of the whole community, COG of DSS was set as an object to compare with the other two.

25 COG classes were further grouped into 4 main categories (Figure 7a): (1) information storage and processing, containing classes J, A, K, L and B; (2) cellular processes and signaling, containing classes D, Y, V, T, M, N, Z, W, U and O; (3) metabolism, containing classes C, G, E, F, H, I and P; (4) poorly characterized, containing classes R and S. In the first group, reference genome Y04AAS1 was with overall 10% less of abundance compared with others, in classes J (translation, ribosomal structure and biogenesis), K (transcription), and L (replication, recombination and repair). In the second group, DSS showed about 10% less in abundance with others, mainly in classes D (cell cycle control, cell division, chromosome partitioning), T (signal transduction mechanisms), M (cell wall/membrane/envelope biogenesis), N (cell motility) and U (intracellular trafficking, secretion, and vesicular transport). However, in class V (defense mechanism) gene function distribution of DSS was slightly more abundant than the rest. In group 3, three samples were of similar overall abundance, but minor differences occurred from class to class. In the fourth group, DSS Hb had almost 10% less in abundance than others, in both R (general functional prediction only) and S (function unknown) classes.

As a whole, abundance of gene functions related with metabolism was consistent among the three data sets, occupied 35 to 40% of total function. Functions for information storage and processing was especially low in

Y04AAS1 data set; functions for cellular process and signaling was lower in DSS dataset than that in others; the proportion of genes with unknown function was lowest in DSS Hb dataset.

4. Discussions

4.1 Sequencing: Fosmid Library Clones and Direct Shotgun Sample

Direct shotgun sample (DSS) was 3.4 times larger than fosmid library clones (Fos) in total size of reads after removal of vectors, and assembled contigs from DSS were 2.4 times more than that of Fos sample. The proportions of total bases being used for contig assembly were 0.19% and 0.11% in Fos and DSS, respectively. If the proportion of contig length was compared (Figure 3, Table 5 and 6), contigs longer than 100 kb occupied 0.01% of total contigs in both samples. Contigs with length in between 10 kb and 100 kb were 0.7% and 0.4% of Fos and DSS total contig sets, respectively. Contigs with length ranged from 1 kb to 10 kb were 5.3% of Fos total contig and 6.1% in DSS total contig. Apparently there was no significant difference in the assembly of two sets of contig under same settings of assembly parameters, which might indirectly indicated the genetic complexity of both samples were quite similar, although DNAs from fosmid clones were contaminated by host *E. coli* genomic DNAs during the extraction.

Two kinds of software for reads assembly (CLC, MetaVelvet) were tested on Fos sample, and general comparison by proportion of certain length of contigs was obviously incongruent (see also Table 5b and c): contigs longer than 100 kb occupied 0.01% and 0.1% of total contigs in samples assembled by MetaVelvet and CLC, respectively. Contigs with length in between 10 kb and 100 kb were 0.7% and 4.9% of MetaVelvet and CLC, respectively. Contigs with length ranged from 1 kb to 10 kb were 5.3% of MetaVelvet assembled total contig and 27% in that from CLC. Therefore, the overview suggested that

the choice on methods and tools in assembly step played a critical role which might affect data interpretation to some extent.

4.2 Analyses on Community Structure

4.2.1 Different Analytic Approaches

Due to the fact that high proportion of *E. coli* was found in Fos sample after mapping raw reads against reference genomes in IMG database (Appendix 4), complete analysis on community structure was applied only to DSS sample.

Each method of the four had its advantages and limitations in data interpretation. Analysis on SSU rDNA sequences was one of the most conventional systems for taxonomic identification. Although some have pointed out sequence diversity of different copies of rRNA operons in individual prokaryote sometimes are higher than those between well-defined species (Pei et al., 2010), 16 S rRNA analysis offers a reliable way to characterize species diversity without the complication by horizontal gene transfer (HGT) or other highly variable sequences. The other advantage of taxa assigning through 16 S rDNA sequences is the efficiency aided by abundant information accumulated in well-established online database. The extraction of contigs containing 16 S rDNA sequences attempted to fulfill two purposes: to know which prokaryotes were in the community through 16 S rDNA sequences identification and to estimate the abundance through contig length. Theoretically, the longer the contig is, the more abundant the species represented by the contig is (Inskeep et al., 2010). However, the analytic

results of abundance determined by length of 16 S rDNA containing contigs were not like what has been expected. Contigs with same taxa id were all of different length, which was problematic to apply an objective weighting principle. Furthermore, after comparing this result with relative abundances that derived from mapping raw reads against SILVA (DSS > SILVA) and IMG database, it was clear that many taxa information was overlooked in this analysis. To reinforce the original idea, relative abundance was estimated again by mapping raw reads onto contigs containing 16 S rDNA sequences (DSS > 16 SCTG) instead of judging by contig length.

The disadvantage of using 16 S-based concepts for relative abundance measurement, as described above, might be insufficient sampling of sequences. Prokaryotic rRNA genes are averagely 1.5 kb, which only counts for 0.03% of average prokaryotic genome (5 Mb). For environmental shotgun sequencing, the chance to obtained sufficient 16 S rDNA sequences is relatively difficult by using PCR amplification, even though evidence has shown that results from PCR might be biased by preferential amplification (Reysenbach et al., 1992b).

To compensate the weakness of previous two 16 S-centered analyses, mapping raw reads against reference genomes (DSS > IMG) and against all contigs with taxa names (DSS > tCTG) were done. The pros of DSS > IMG and DSS > tCTG was to use most of all sequence information, not only restricted to SSU rRNA genes. The main cons of DSS > IMG were two. First, sequences belonged to species that lacked information in IMG database was thrown away, especially those of archaea (such as *Stygiolobus* sp., which could be found in

SILVA database). Second, there was no better weighting method to solve the unbalance of extra scores assigned to genus that consisted of more species. This distortion could result in a community structure represented by underestimated archaeal members, since the ratio of bacterial and archaeal genome number in IMG was almost 10:1. DSS > tCTG analysis was the least restricted since NCBI's non-redundant (NR) database was used, which contained the broadest range of taxonomic information (although quality was not completely guaranteed) among databases utilized in previous analyses. Blast against NR database also revealed the proportion of unassigned prokaryotes, which made DSS > tCTG analysis a comprehensive estimation of prokaryotic community structure instead of community structure of known taxa (represented by DSS > 16 SCTG, DSS > SILVA and DSS > IMG). In the fourth approach, although it revealed the proportions of unassigned bacteria and archaea, there were several factors that might be influential to the results: no delicate strategy other than LCA to make sure that every assigned taxa name of contigs were accurate, sequence quality in NR database was relatively unstable compared to that of IMG's, and no additional process to estimate and reduce the bias cause by horizontal gene transfer.

The fourth approach served more like an overview (but not so precise) on the community structure since it reflected the abundance contributed by uncategorized prokaryotic members. The other three, on the other hand, provided better abundance estimations on prokaryotes that have already been documented.

In the 16 S-centered analyses (DSS > 16sCTG and DSS > SILVA), proportions of utilized reads were 0.06 to 0.1%, which were close to the average percentage occupied by SSU rRNA genes in a genome (suppose a prokaryotic genome size is 5 Mb and SSU rRNA gene is 1 to 1.5 Kb). In the DSS > IMG analysis, due to stringent criteria and limited number of reference genomes, only 1.41% of total reads were used for abundance assessment. In contrast, 66% of reads was utilized in DSS > tCTG, which seemed to be more convincing for obtaining a general view on community structure by its better representativeness.

4.2.2 Prokaryotes Composition of Hot Spring in SHP

From the four approaches, it was clear that bacteria were dominant over archaea, although each approach generated different ratios. Since the copy number of prokaryotic rRNA operon varied, this likely source of bias could be introduced to DSS > 16sCTG but not to DSS > SILVA (see materials and method). Thus, the bacteria / archaea (of known species) ratio derived from latter should be more accurate than the former. Even though DSS > IMG was known to be possibly affected by overestimation of bacterial abundance by the weighting method, and DSS > tCTG was vulnerable to low quality reference sequences and HGT effect, the abundance of bacteria estimated by this study was around 70 or 80%.

Most documented members of phylum Aquificae were chemolithoautotrophic hydrogen oxidizers and many lived in terrestrial hydrothermal systems. Genus

Hydrogenobaculum, comprised at least 16% relative abundance of whole community (based on the DSS > tCTG result), was once reported to be capable of arsenite oxidation (Donahoe-Christiansen et al., 2004). Due to the lack of further investigation, it remained uncertain whether there is a relationship between the limited abundance of genus *Hydrogenobaculum* and the low amount of total As in SHP, or members of *Hydrogenobaculum* did not possess the function of arsenite oxidation in this community.

Many bacteria found in the hot spring water were also commonly seen in soil, such as members of Firmicutes and Proteobacteria. Since the sample was from a narrow flow of heated water next to soil and weathered rock debris, some prokaryotes of soil origin might be introduced to water. Phototrophic bacteria such as members of phyla Cyanobacteria, Chloroflexi and Chlorobi were also detected with low abundance from 0.001 to 0.145% (Table 8). According to the work done by Cox et al. (2011), it would be difficult for these bacteria to survive in environment with low pH and high temperature like SHP. Thus, the existence of sequences from phototrophic bacteria in this study could be originated from small amount of living cells (could be allopatric or sympatric), and some inactive or dead cells.

In archaea, members of genus *Sulfolobus,* which have been isolated almost exclusively from continental solfataric fields, were the most abundant, contributed 5% to the total abundance of the community. Almost all *Sulfolobus* sp. are extremely acidophilic thermophiles, and are able to oxidize elemental sulfur to sulfuric acid, which seemed to be close related to the environmental

conditions that featured by high sulfuric acid concentration and rich sulfur compounds.

Sequences belonged to archaeal phyla Nanoarchaeota, Thaumarchaeota and Korarchaeota were discovered as well. Earlier research has proposed the symbiosis of *Nanoarchaeum equitans* and its partner host *Ignicoccus hospitalis* (Huber et al., 2002b) but the latter was not found by previous work (Lu et al., unpublished data) through PCR and cloning. Three approaches of four in this study have discovered the genus *Ignicoccus* sequences and estimated the relative abundances. If the abundance calculated by DSS > tCTG was taken, the abundance ratio of phylum Nanoarchaeota to genus Ignicoccus was 7.25 to 1. Based on the co-culture experiment of *N. equitans* and *I. hospitalis*, a single *I. hospitalis* could carry as much as 10 *N. equitans* (Jahn et al., 2008). In this context, the occurrence of symbiotic relationship between the two was still possible since the abundance of nanoarchaea-like prokaryotes did not exceed the host capacity of *Ignicoccus*, even though some have suggested that discovered nanoarchaea-like prokaryotes might have free-living (non-symbiotic) life style (Casanueva et al., 2008).

Members of phylum Thaumarchaeota was originally found to be mesophilic ammonia oxidizers (Brochier-Armanet et al., 2008) but later were also discovered in acidic high temperature spring (Kozubal et al., 2012). Korarchaeota is still a poorly understood phylum; many have reported the coexistence of Korarchaeotes and members of bacterial phylum Aquificae (Reysenbach et al., 2000; Reigstad et al., 2010). Observations in previous

work mainly suggested this co-occurrence preferred a higher pH environment (close to neutral), but our finding in this study revealed another possibility under acidic condition.

4.3 *Hydrogenobaculum*-like Sequences Recovered in this Study

4.3.1 Aquificae, a Dominant Bacterial Phylum in Hydrothermal Springs

Numerous inspections on hot spring microbial systems have been performed in geothermally active areas such as Yellowstone National Park, Grand Canyon Basin and Lassen Volcanic National Park of the United States (Brock et al., 1972; Barns et al., 1994; Reysenbach et al., 1994; Barns et al., 1996; Ferris et al., 1996; Hugenholtz et al., 1998; Reysenbach et al., 2000; Jackson et al., 2001; Donahoe-Christiansen et al., 2004; Meyer-Dombard et al., 2005; Spear et al., 2005; Siering et al., 2006; Mathur et al., 2007; Hall et al., 2008; Costa et al., 2009; Inskeep et al., 2010; Vick et al., 2010; Klatt et al., 2011; Das, 2012; Kozubal et al., 2012), Iceland (Hjorleifsdottir et al., 2001; Kvist et al., 2007; Perevalova et al., 2008; Reigstad et al., 2010; Tobler and Benning, 2011), Kamchatka of Russia (Perevalova et al., 2008; Kublanov et al., 2009; Reigstad et al., 2010; Kochetkova et al., 2011; Mardanov et al., 2011; Zhao et al., 2011), Tibet of China (Yim et al., 2006; Lau et al., 2009; Huang et al., 2011; Song et al., 2012), the Andes in Colombia (Bohorquez et al., 2012), Caribbean island of Montserrat (Burton and Norris, 2000), Bulgaria (Tomova et al., 2010; Ivanova et al., 2011), Japan (Yamamoto et al., 1998; Kato et al., 2011), Taiwan (Ng et al., 2005), Australia (Kimura et al., 2005), New Zealand (Ellis et al., 2005; Hetzer et al., 2007; Childs et al., 2008), Thailand

(Kanokratana et al., 2004), Indonesia (Baker et al., 2001), South Africa (Tekere et al., 2011) and Tunisia (Sayeh et al., 2010).

Although physicochemical properties differ in each hot springs even within the same geographic region, certain microbial clades seem to be more widespread and cosmopolitan than others. In terrestrial hydrothermal systems, despite the fact that some rare cases have reported the lack of detected bacteria or archaea (Siering et al., 2006; Tobler and Benning, 2011), bacteria and archaea usually coexist. As crenarchaea are often the main proportion of archaeal community in the majority of hot spring environments that have been surveyed, evolutionary deep-branched Aquificae (Burggraf et al., 1992) is one of the chief members of Domain Bacteria found in volcanically or geothermally heated microbial ecosystems (Huber and Eder, 2007; Hall et al., 2008).

Aquificae consisting of a single class and the single order Aquificales with two affiliated families “Aquificaceae” and “Hydrogenothermaceae”. Cultivated members of Aquificales lineage (Kryukov et al., 1983; Kawasumi et al., 1984; Kristjansson et al., 1985; Bonjour and Aragno, 1986; Nishihara et al., 1990; Shima and Suzuki, 1993; Skirnisdottir et al., 2001; Takai et al., 2001a; Eder and Huber, 2002) were discovered in continental and/or marine hydrothermal environments, belong to extremely thermophilic or hyperthermophilic bacteria with optimal growth in the neutral pH range (with some exceptions, such as *Hydrogenobaculum acidophilum* and *Hydrogenobaculum* sp. NOR3L3B (Eder and Huber, 2002)). Most Aquificales are chemolithoautotrophic, although few

heterotrophs were reported (Huber et al., 1998; Nakagawa et al., 2005a) such as chemoorganoheterotrophic Thermocrinis ruber and Sulfurihydrogenibium sp.. Some of them are also strict aerobic, growing preferentially under microaerophilic culture conditions and are able to grow with hydrogen as sole electron donor and oxygen as electron acceptor, performing the "Knallgas" reaction, the reduction of O_2 with H_2:

$$2H_2 + O_2 \rightarrow 2H_2O$$

However, evidence based on culture studies and field samplings suggested Aquificales can utilize a diversity of metabolic reactions in addition to or instead of hydrogen oxidation: oxidation of elemental sulfur, thiosulfate, or ferrous iron and reduction of nitrate, ferric iron, arsenate, selenate, selenite, or elemental sulfur were also observed (Huber et al., 1998; L'Haridon et al., 1998; Takai et al., 2001a; Blank et al., 2002; Eder and Huber, 2002; Gotz et al., 2002; Huber et al., 2002a; Alain et al., 2003; Nakagawa et al., 2003; Takai et al., 2003a; Takai et al., 2003b; Aguiar et al., 2004; Nakagawa et al., 2004; Vetriani et al., 2004; Nakagawa et al., 2005a; L'Haridon et al., 2006). In high-sulfide hot springs, Aquificales is one of the dominant phyla that manage to oxidize sulfur to sulfuric acid or reduce it to hydrogen sulfide proven by culture studies (Huber et al., 1998; Skirnisdottir et al., 2000).

Affiliated with Aquificales, members of (or closely related to) genus Hydrogenobaculum has been documented in some terrestrial hot spring environments in Japan and YNP of the US (Shima and Suzuki, 1993;

Donahoe-Christiansen et al., 2004; Spear et al., 2005; Mathur et al., 2007; D'Imperio et al., 2008; Hall et al., 2008; Wilson et al., 2008; Inskeep et al., 2010; Kozubal et al., 2012; Song et al., 2012) as well as in sulfuric hydrothermal systems of northern Taiwan volcanic area (李重義 and 張怡塘, 2009; Cheng et al., 2013).

Hydrogenobaculum spp. in some cases were found to dominant among bacteria in hot springs (Spear et al., 2005; Mathur et al., 2007; Inskeep et al., 2010; Kozubal et al., 2012; Song et al., 2012; Cheng et al., 2013). Metabolically, it is a chemolithoautotrophic hydrogen oxidizer as most Aquificales members are, while some studies have pointed out its extra or alternative arsenite-oxidation potential (Donahoe-Christiansen et al., 2004).

4.3.2 Analysis on Recovered *Hydrogenobaculum*-like Sequences

Based on the results of community structure analyses, Aquificae (solely consisted of genus *Hydrogenobaculum* in this study) was dominant in this hydrothermal spring. Due to the lack of full-length-SSU rRNA gene sequences of *Hydrogenobaculum*-like (Hb-like) prokaryotes, the phylogenetic relationship in between reference strain Y04AAS1 and that found in SHP area was unable to be known by sequence alignment. To understand the completeness of Hb-like genome recovered from shotgun sequencing, Y04AAS1 genome was mapped by DSS read coverage, status of DSS and Fos contig, and contigs assembled by SOAPdenovo (Figure 5). Coverage of read was mostly consistent with DSS contig mapping density while Fos contigs had its distinct pattern. Fos contigs in this study was known to have higher GC

content than that of DSS contigs. A previous research also indicated that sequences with higher GC content are more likely to be packaged into fosmid clones (Temperton et al., 2009). Hb-like Fos contigs were examined to check if they displayed higher GC content trait, but only regions of rRNA operons were of 55% in GC content and the average was only 35%, equal to that of *Hydrogenobaculum* sp.. According to the result derived from mapping Fos raw reads to IMG database (Appendix 4), if *E. coli* was excluded, top five abundant genera were of average 57% in GC content, closer to 65% (average GC content of Fos contigs) compared to 35% of *Hydrogenobaculum* sp.. Possible explanation could be, the higher GC content found in Fos contigs might come from other prokaryotes with higher GC content. To be more specific, the insert of fosmid clone is around 40±5 Kb. If high GC sequence is preferred, it should be easier to pick up genome fragments from high GC organisms than to pick up high GC fragments in a low GC content organism since it is harder to find a high GC fragment with a length about 40 Kb in a low GC genome. Thus, if the postulation is true, it might be a reasonable inference that the fact *Hydrogenobaculum* sp. was not dominant in Fos sample could be somehow affected by the process of fosmid library construction.

The difference between SOAPdenovo contig set and Fos / DSS contig sets was, reads comprised the former was filtered first to be Hb-like and then assembled into contigs while the latter two were Hb-like subsets of total contigs assembled by total reads from Fos and DSS samples. SOAPdenovo contigs were significant less than that of DSS mapped to reference genome. This might be caused by over-stringent conditions filtered out possible raw reads

that originally belonged to Hb-like genome for SOAPdenovo assembly, or loosely incorporate raw reads that did not belong to Hb-like genome into MetaVelvet assembly, or a mixed effect of both.

4.3.3 *Hydrogenobaculum*-associated Reads and Genome Mapping

From the reads coverage mapped to strain Y04AAS1 genome, except some rare regions with conspicuous higher value (which might be regions with highly conserved sequences among different species or multiple copies of certain genes), the over-all pattern seemed to be close to randomly sequenced. The relationship between GC content of Hb-like raw reads and average read mapping coverage displayed a positive correlation trend which was also reported as an Illumina platform-specific phenomenon (Dohm et al., 2008; Minoche et al., 2011; Benjamini and Speed, 2012; Luo et al., 2012). However, the distribution of data points aggregated densely at coverage lower than 5000, with average GC content about 35% which is consistent with that of *Hydrogenobaculum* sp. (Figure 6). Unlike the proposed unimodal pattern, our plot did not have the other portion, where GC content higher than 60%, since *Hydrogenobaculum* sp. is a bacteria with low GC genome.

Assumed that the pattern of Hb-like read mapping coverage to reference genome could be applied to every genome occurred in this community, reads mapping to another reference genome with higher GC content should generate a typical unimodal pattern that reflects the GC bias in Illumina platform, which is considered to be bias generated in the PCR step before sequencing.

4.4 Characteristics of Gene Function Distribution

Gene function distribution in the community was presented by COG chart of DSS contigs and Hydrogenobaculum-like contigs (DSS Hb). In general, COG pattern of DSS and DSS Hb showed similarity in metabolism and information storage and processing but quite different in cellular processes and signaling and genes with unknown functions (Figure 7). Phylum Aquificae was discovered to be dominant in this community (67% in DSS > IMG and 43% in DSS > tCTG), which indicated that there were still many other prokaryotes thrived in this community might possess different gene function and contributed the functional differences.

Through comparing COG pattern of DSS Hb and reference genome Y04AAS1, it is possible to reasonably conjecture that the low read mapping coverage regions might be genes relevant with functional group L (replication, recombination and repair), T (signal transduction mechanisms), G (carbohydrate transport and metabolism), E (amino acid transport and metabolism) and a large part belongs to unknown function.

Comparing top five abundant COG classes of DSS total contig with the that of the other hot spring with lower temperature but same acidity in Columbian Andes (Jiménez et al., 2012), class L (replication, recombination and repair) was of higher proportion in the hot spring of Andes than here in SHP, followed by classes R (general function prediction only), S (function unknown), E (amino acid transport and metabolism), J (translation, ribosomal structure and biogenesis). In SHP, however, class J (translation, ribosomal structure and

biogenesis) was the most abundant function found in this study, followed by class C (energy production and conversion), class R (general function prediction only), class E (amino acid transport and metabolism) and class L (replication, recombination and repair), which suggested a plausible enrichment of gene function related to metabolism in this community.

5. Summary

This study was aimed to obtain an overview on prokaryotic diversity, community structure and function of hydrothermal spring system in TVG of northern Taiwan thorough comprehensive approaches from traditional PCR to next generation sequencing technique and bioinformatic analysis.

Structural and functional analyses of prokaryotic community in selected hot spring in SHP revealed bacteria-dominated characteristics. Members of bacterial phyla Aquificae, Proteobacteria and Firmicutes, archaeal phylum Crenarchaeota were main players in this hot spring system, comprising more than 60% of the community. Different methodologies for estimating relative abundance of taxa provided different results, but dominant taxa was less affected than taxa with lower abundance.

In DSS sample, prokaryotes of genus *Hydrogenobaculum* were the most abundant bacteria found in the chosen hot spring in SHP. However, reads derived from sequencing were still not sufficient enough (around 30% of gaps after reference genome mapping) to recover a complete genome. GC content bias of Illumina NGS platform was detected in analysis of Hb-like sequences.

Gene function distribution deduced by protein COG categorizing suggested a high proportion of metabolism-related function of this hot spring community. Further investigations on critical metabolic pathways and gene expressions are vital for a more thorough and concrete insight to the terrestrial hydrothermal spring systems.

References

Aguiar, P., Beveridge, T. J. and Reysenbach, A. L. (2004). "Sulfurihydrogenibium azorense, sp nov., a thermophilic hydrogen-oxidizing microaerophile from terrestrial hot springs in the Azores." International Journal of Systematic and Evolutionary Microbiology **54**: 33-39.

Ajay, S. S., Parker, S. C. J., Abaan, H. O., Fajardo, K. V. F. and Margulies, E. H. (2011). "Accurate and comprehensive sequencing of personal genomes." Genome Research **21**(9): 1498-1505.

Alain, K., Rolland, S., Crassous, P., Lesongeur, F., Zbinden, M., Le Gall, C., Godfroy, A., Page, A., Juniper, S. K., Cambon-Bonavita, M. A., Duchiron, F. and Querellou, J. (2003). "Desulfurobacterium crinifex sp.nov., a novel thermophilic, pinkish-streamer forming, chemolithoautotrophic bacterium isolated from a Juan de Fuca Ridge hydrothermal vent and amendment of the genus Desulfurobacterium." Extremophiles **7**(5): 361-370.

Amann, R. I., Ludwig, W. and Schleifer, K. H. (1995). "Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells without Cultivation." Microbiological Reviews **59**(1): 143-169.

Arnórsson, S., Gunnlaugsson, E. and Svavarsson, H. (1983). "The chemistry of geothermal waters in Iceland. II. Mineral equilibria and independent variables controlling water compositions." Geochimica Et Cosmochimica Acta **47**(3): 547-566.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. and Consortium, G. O. (2000).

"Gene Ontology: tool for the unification of biology." Nature Genetics **25**(1): 25-29.

Baker, B. J. and Banfield, J. F. (2003). "Microbial communities in acid mine drainage." Fems Microbiology Ecology **44**(2): 139-152.

Baker, G. C., Gaffar, S., Cowan, D. A. and Suharto, A. R. (2001). "Bacterial community analysis of Indonesian hot springs." Fems Microbiology Letters **200**(1): 103-109.

Barns, S. M., Delwiche, C. F., Palmer, J. D. and Pace, N. R. (1996). "Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences." Proceedings of the National Academy of Sciences of the United States of America **93**(17): 9188-9193.

Barns, S. M., Fundyga, R. E., Jeffries, M. W. and Pace, N. R. (1994). "Remarkable Archaeal Diversity Detected in a Yellowstone-National-Park Hot-Spring Environment." Proceedings of the National Academy of Sciences of the United States of America **91**(5): 1609-1613.

Belousov, A., Belousova, M., Chen, C. H. and Zellmer, G. F. (2010). "Deposits, character and timing of recent eruptions and gravitational collapses in Tatun Volcanic Group, Northern Taiwan: Hazard-related issues." Journal of Volcanology and Geothermal Research **191**(3-4): 205-221.

Benjamini, Y. and Speed, T. P. (2012). "Summarizing and correcting the GC content bias in high-throughput sequencing." Nucleic Acids Research **40**(10).

Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E. and House, C. H. (2008). "Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment." Proceedings of the National Academy of Sciences of the United States of America **105**(30): 10583-10588.

Blank, C. E., Cady, S. L. and Pace, N. R. (2002). "Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park." Applied and Environmental Microbiology **68**(10): 5123-5135.

Bohorquez, L. C., Delgado-Serrano, L., Lopez, G., Osorio-Forero, C., Klepac-Ceraj, V., Kolter, R., Junca, H., Baena, S. and Zambrano, M. M. (2012). "In-depth Characterization via Complementing Culture-Independent Approaches of the Microbial Community in an Acidic Hot Spring of the Colombian Andes." Microbial Ecology **63**(1): 103-115.

Bonch-Osmolovskaya, E. A., Miroshnichenko, M. L., Slobodkin, A. I., Sokolova, T. G., Karpov, G. A., Kostrikina, N. A., Zavarzina, D. G., Prokof'eva, M. I., Rusanov, I. I. and Pimenov, N. V. (1999). "Biodiversity of Anaerobic lithotrophic prokaryotes in terrestrial hot springs of Kamchatka." Microbiology **68**(3): 343-351.

Bonjour, F. and Aragno, M. (1986). "Growth of thermophilic, obligatorily chemolithoautotrophic hydrogen-oxidizing bacteria related to Hydrogenobacter with thiosulfate and elemental sulfur as electron and energy source." Fems Microbiology Letters **35**(1): 11-15.

Brochier-Armanet, C., Boussau, B., Gribaldo, S. and Forterre, P. (2008). "Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota." Nature Reviews Microbiology **6**(3): 245-252.

Brock, T. D. (1967). "Life at High Temperatures - Evolutionary Ecological and Biochemical Significance of Organisms Living in Hot Springs Is Discussed." Science **158**(3804): 1012-&.

Brock, T. D., Brock, K. M., Belly, R. T. and Weiss, R. L. (1972). "Sulfolobus : A new genus of sulfur-oxidizing bacteria living at low pH and high temperature."

Archives of Microbiology **84**(1): 54-68.

Brock, T. D. and Freeze, H. (1969). "Thermus Aquaticus Gen N and Sp N a Nonsporulating Extreme Thermophile." Journal of Bacteriology **98**(1): 289-&.

Burggraf, S., Olsen, G. J., Stetter, K. O. and Woese, C. R. (1992). "A phylogenetic analysis of Aquifex pyrophilus." Syst. Appl. Microbiol. **15**: 352-356.

Burton, N. P. and Norris, P. R. (2000). "Microbiology of acidic, geothermal springs of Montserrat: environmental rDNA analysis." Extremophiles **4**(5): 315-320.

Casanueva, A., Galada, N., Baker, G. C., Grant, W. D., Heaphy, S., Jones, B., Ma, Y. H., Ventosa, A., Blamey, J. and Cowan, D. A. (2008). "Nanoarchaeal 16S rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments." Extremophiles **12**(5): 651-656.

Chai, B. H. T. (1972). "Structure and Tectonic Evolution of Taiwan." American Journal of Science **272**(5): 389-&.

Chen, W. F. and Sung, M. H. (2009). "The Redox Potential of Hot Springs in Taiwan." Terrestrial Atmospheric and Oceanic Sciences **20**(3): 465-479.

Cheng, T.-W., Chang, Y.-H., Tang, S.-L., Tseng, C.-H., Chiang, P.-W., Chang, K.-T., Sun, C.-H., Chen, Y.-G., Kuo, H.-C., Wang, C.-H., Chu, P.-H., Song, S.-R., Wang, P.-L. and Lin, L.-H. (2012). "Metabolic stratification driven by surface and subsurface interactions in a terrestrial mud volcano." Isme Journal.

Cheng, T. W., Wang, P. L., Chen, K. C., Wu, J. J., Song, S. R. and Lin, L. H. (2013). "Segregated planktonic and bottom-dwelling archaeal communities in high-temperature acidic, sulfuric ponds of the Tatung Volcano Group, northern Taiwan." Terrestrial, Atmospheric and Oceanic Sciences **(in press)**.

Childs, A. M., Mountain, B. W., O'toole, R. and Stott, M. B. (2008). "Relating Microbial Community and Physicochemical Parameters of a Hot Spring:

Champagne Pool, Wai-o-tapu, New Zealand." Geomicrobiology Journal **25**(7-8): 441-453.

Cole, J. R., Chai, B., Marsh, T. L., Farris, R. J., Wang, Q., Kulam, S. A., Chandra, S., Mcgarrell, D. M., Schmidt, T. M., Garrity, G. M. and Tiedje, J. M. (2003). "The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy." Nucleic Acids Research **31**(1): 442-443.

Costa, K. C., Navarro, J. B., Shock, E. L., Zhang, C. L. L., Soukup, D. and Hedlund, B. P. (2009). "Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin." Extremophiles **13**(3): 447-459.

Cox, A., Shock, E. L. and Havig, J. R. (2011). "The transition to microbial photosynthesis in hot spring ecosystems." Chemical Geology **280**(3-4): 344-351.

D'imperio, S., Lehr, C. R., Oduro, H., Druschel, G., Kuhl, M. and Mcdermott, T. R. (2008). "Relative importance of H(2) and H(2)S as energy sources for primary production in geothermal springs." Applied and Environmental Microbiology **74**(18): 5802-5808.

Das, D. (2012). Effects of abiotic factors on the phylogenetic diversity of archaeal communities in acidic thermal springs of Yellowstone National Park, San Diego State University.

Delong, E. E. (2005). "Microbial community genomics in the ocean." Nature Reviews Microbiology **3**(6): 459-469.

Delong, E. E. and Pace, N. R. (2001). "Environmental diversity of Bacteria and Archaea." Systematic Biology **50**(4): 470-478.

Desantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber,

T., Dalevi, D., Hu, P. and Andersen, G. L. (2006). "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB." Applied and Environmental Microbiology **72**(7): 5069-5072.

Dohm, J. C., Lottaz, C., Borodina, T. and Himmelbauer, H. (2008). "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing." Nucleic Acids Research **36**(16).

Donahoe-Christiansen, J., D'imperio, S., Jackson, C. R., Inskeep, W. P. and Mcdermott, T. R. (2004). "Arsenite-oxidizing Hydrogenobaculum strain isolated from an acid-sulfate-chloride geothermal spring in Yellowstone National Park." Applied and Environmental Microbiology **70**(3): 1865-1868.

Eder, W. and Huber, R. (2002). "New isolates and physiological properties of the Aquificales and description of Thermocrinis albus sp nov." Extremophiles **6**(4): 309-318.

Edwards, K. J., Bond, P. L., Gihring, T. M. and Banfield, J. F. (2000). "An archaeal iron-oxidizing extreme acidophile important in acid mine drainage." Science **287**(5459): 1796-1799.

Edwards, R. A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D. M., Saar, M. O., Alexander, S., Alexander, E. C. and Rohwer, F. (2006). "Using pyrosequencing to shed light on deep mine microbial ecology." Bmc Genomics **7**.

Ellis, D. G., Bizzoco, R. L. W., Maezato, Y., Baggett, J. N. and Kelley, S. T. (2005). "Microscopic examination of acidic hot springs of Waiotapu, North Island, New Zealand." New Zealand Journal of Marine and Freshwater Research **39**(5): 1001-1011.

Ferrer, M., Beloqui, A., Timmis, K. N. and Golyshin, P. N. (2009). "Metagenomics for

Mining New Genetic Resources of Microbial Communities." Journal of Molecular Microbiology and Biotechnology **16**(1-2): 109-123.

Ferrer, M., Golyshina, O., Beloqui, A. and Golyshin, P. N. (2007). "Mining enzymes from extreme environments." Current Opinion in Microbiology **10**(3): 207-214.

Ferrer, M., Werner, J., Chernikova, T. N., Bargiela, R., Fernandez, L., La Cono, V., Waldmann, J., Teeling, H., Golyshina, O. V., Glockner, F. O., Yakimov, M. M., Golyshin, P. N. and Consortium, M. S. (2012). "Unveiling microbial life in the new deep-sea hypersaline Lake Thetis. Part II: a metagenomic study." Environmental Microbiology **14**(1): 268-281.

Ferris, M. J., Muyzer, G. and Ward, D. M. (1996). "Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community." Applied and Environmental Microbiology **62**(2): 340-346.

Ferris, M. J. and Ward, D. M. (1997). "Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis." Applied and Environmental Microbiology **63**(4): 1375-1381.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L. L., Eddy, S. R. and Bateman, A. (2010). "The Pfam protein families database." Nucleic Acids Research **38**: D211-D222.

Giovannoni, S. J., Delong, E. F., Schmidt, T. M. and Pace, N. R. (1990). "Tangential Flow Filtration and Preliminary Phylogenetic Analysis of Marine Picoplankton." Applied and Environmental Microbiology **56**(8): 2572-2575.

Gotz, D., Banta, A., Beveridge, T. J., Rushdi, A. I., Simoneit, B. R. T. and Reysenbach, A. (2002). "Persephonella marina gen. nov., sp nov and Persephonella guaymasensis sp nov., two novel, thermophilic, hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents." International Journal of Systematic and Evolutionary Microbiology **52**: 1349-1359.

Hall, J. R., Mitchell, K. R., Jackson-Weaver, O., Kooser, A. S., Cron, B. R., Crossey, L. J. and Takacs-Vesbach, C. D. (2008). "Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes." Applied and Environmental Microbiology **74**(15): 4910-4922.

Handelsman, J. (2004). "Metagenomics: Application of genomics to uncultured microorganisms." Microbiology and Molecular Biology Reviews **68**(4): 669-+.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. and Goodman, R. M. (1998). "Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products." Chemistry & Biology **5**(10): R245-R249.

Harel, D. and Tarjan, R. E. (1984). "Fast Algorithms for Finding Nearest Common Ancestors." Siam Journal on Computing **13**(2): 338-355.

Hetzer, A., Morgan, H. W., Mcdonald, I. R. and Daughney, C. J. (2007). "Microbial life in champagne pool, a geothermal spring in Waiotapu, New Zealand." Extremophiles **11**(4): 605-614.

Hjorleifsdottir, S., Skirnisdottir, S., Hreggvidsson, G. O., Holst, O. and Kristjansson, J. K. (2001). "Species composition of cultivated and noncultivated bacteria from short filaments in an Icelandic hot spring at 88 degrees C." Microbial Ecology **42**(2): 117-125.

Hough, D. W. and Danson, M. J. (1999). "Extremozymes." Current Opinion in Chemical Biology **3**(1): 39-46.

Huang, Q. Y., Dong, C. Z., Dong, R. M., Jiang, H. C., Wang, S., Wang, G. H., Fang, B., Ding, X. X., Niu, L., Li, X., Zhang, C. L. and Dong, H. L. (2011). "Archaeal and bacterial diversity in hot springs on the Tibetan Plateau, China." Extremophiles **15**(5): 549-563.

Huber, H., Diller, S., Horn, C. and Rachel, R. (2002a). "Thermovibrio ruber gen. nov., sp. nov., an extremely thermophilic, chemolithoautotrophic, nitrate-reducing bacterium that forms a deep branch within the phylum Aquificae." International Journal of Systematic and Evolutionary Microbiology **52**(5): 1859-1865.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C. and Stetter, K. O. (2002b). "A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont." Nature **417**(6884): 63-67.

Huber, J. A., Mark Welch, D., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A. and Sogin, M. L. (2007). "Microbial population structures in the deep marine biosphere." Science **318**(5847): 97-100.

Huber, R. and Eder, W. (2007). Aquificales. The Prokaryotes. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer and E. Stackebrandt, Springer New York**:** 925-938.

Huber, R., Eder, W., Heldwein, S., Wanner, G., Huber, H., Rachel, R. and Stetter, K. O. (1998). "Thermocrinis ruber gen. nov., sp. nov., a pink-filament-forming hyperthermophilic bacterium isolated from Yellowstone National Park." Applied and Environmental Microbiology **64**(10): 3576-3583.

Hugenholtz, P., Pitulle, C., Hershberger, K. L. and Pace, N. R. (1998). "Novel division

level bacterial diversity in a Yellowstone hot spring." Journal of Bacteriology **180**(2): 366-376.

Inskeep, W. P., Rusch, D. B., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Richardson, T. H., Macur, R. E., Hamamura, N., Jennings, R. D., Fouke, B. W., Reysenbach, A. L., Roberto, F., Young, M., Schwartz, A., Boyd, E. S., Badger, J. H., Mathur, E. J., Ortmann, A. C., Bateson, M., Geesey, G. and Frazier, M. (2010). "Metagenomes from High-Temperature Chemotrophic Systems Reveal Geochemical Controls on Microbial Community Structure and Function." Plos One **5**(3).

Ivanova, I., Atanassov, I., Lyutskanova, D., Stoilova-Disheva, M., Dimitrova, D., Tomova, I., Derekova, A., Radeva, G., Buchvarova, V. and Kambourova, M. (2011). "High Archaea diversity in Varvara hot spring, Bulgaria." Journal of Basic Microbiology **51**(2): 163-172.

Jackson, C. R., Langner, H. W., Donahoe-Christiansen, J., Inskeep, W. P. and Mcdermott, T. R. (2001). "Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring." Environmental Microbiology **3**(8): 532-542.

Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K. O., Rachel, R. and Huber, H. (2008). "Nanoarchaeum equitans and Ignicoccus hospitalis: New insights into a unique, intimate association of two archaea." Journal of Bacteriology **190**(5): 1743-1750.

Jiménez, D. J., Andreote, F. D., Chaves, D., Montaña, J. S., Osorio-Forero, C., Junca, H., Zambrano, M. M. and Baena, S. (2012). "Structural and Functional Insights from the Metagenome of an Acidic Hot Spring Microbial Planktonic Community in the Colombian Andes." Plos One **7**(12): e52069.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T. and Yamanishi, Y. (2008). "KEGG for linking genomes to life and the environment." Nucleic Acids Research **36**: D480-D484.

Kanokratana, P., Chanapan, S., Pootanakit, K. and Eurwilaichitr, L. (2004). "Diversity and abundance of Bacteria and Archaea in the Bor Khlueng hot spring in Thailand." Journal of Basic Microbiology **44**(6): 430-444.

Kato, S., Itoh, T. and Yamagishi, A. (2011). "Archaeal diversity in a terrestrial acidic spring field revealed by a novel PCR primer targeting archaeal 16S rRNA genes." Fems Microbiology Letters **319**(1): 34-43.

Kawashima, E., Farinelli, L. and Mayer, P. (1998). Method of nucleic acid amplification.

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.

Kimura, H., Sugihara, M., Yamamoto, H., Patel, B. K. C., Kato, K. and Hanada, S. (2005). "Microbial community in a geothermal aquifer associated with the subsurface of the Great Artesian Basin, Australia." Extremophiles **9**(5): 407-414.

Klatt, C. G., Wood, J. M., Rusch, D. B., Bateson, M. M., Hamamura, N., Heidelberg, J. F., Grossman, A. R., Bhaya, D., Cohan, F. M., Kuhl, M., Bryant, D. A. and Ward, D. M. (2011). "Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential." Isme Journal **5**(8): 1262-1278.

Kochetkova, T. V., Rusanov, I. I., Pimenov, N. V., Kolganova, T. V., Lebedinsky, A. V., Bonch-Osmolovskaya, E. A. and Sokolova, T. G. (2011). "Anaerobic transformation of carbon monoxide by microbial communities of Kamchatka hot springs." Extremophiles **15**(3): 319-325.

Konstantinou, K. I., Lin, C. H. and Liang, W. T. (2007). "Seismicity characteristics of a potentially active Quaternary volcano: The Tatun Volcano Group, northern Taiwan." Journal of Volcanology and Geothermal Research **160**(3-4): 300-318.

Kozubal, M. A., Macur, R. E., Jay, Z. J., Beam, J. P., Malfatti, S. A., Tringe, S. G., Kocar, B. D., Borch, T. and Inskeep, W. P. (2012). "Microbial iron cycling in acidic geothermal springs of Yellowstone National Park: Integrating molecular surveys, geochemical processes and isolation of novel Fe-active microorganisms." Frontiers in Microbiology **3**.

Kristjansson, J. K. and Hreggvidsson, G. O. (1995). "Ecology and Habitats of Extremophiles." World Journal of Microbiology & Biotechnology **11**(1): 17-25.

Kristjansson, J. K., Ingason, A. and Alfredsson, G. A. (1985). "Isolation of Thermophilic Obligately Autotrophic Hydrogen-Oxidizing Bacteria, Similar to Hydrogenobacter-Thermophilus, from Icelandic Hot Springs." Archives of Microbiology **140**(4): 321-325.

Kryukov, V. R., Saveleva, N. D. and Pusheva, M. A. (1983). "Calderobacterium-Hydrogenophilum Gen Et Sp-Nov, an Extremely Thermophilic Hydrogen Bacterium and Its Hydrogenase Activity." Microbiology **52**(5): 611-618.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. and Marra, M. A. (2009). "Circos: An information aesthetic for comparative

genomics." Genome Research **19**(9): 1639-1645.

Kublanov, I. V., Perevalova, A. A., Slobodkina, G. B., Lebedinsky, A. V., Bidzhieva, S. K., Kolganova, T. V., Kaliberda, E. N., Rumsh, L. D., Haertle, T. and Bonch-Osmolovskaya, E. A. (2009). "Biodiversity of Thermophilic Prokaryotes with Hydrolytic Activities in Hot Springs of Uzon Caldera, Kamchatka (Russia)." Applied and Environmental Microbiology **75**(1): 286-291.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K. and Hugenholtz, P. (2008). "A Bioinformatician's Guide to Metagenomics." Microbiology and Molecular Biology Reviews **72**(4): 557-578.

Kvist, T., Ahring, B. K. and Westermann, P. (2007). "Archaeal diversity in Icelandic hot springs." Fems Microbiology Ecology **59**(1): 71-80.

L'haridon, S., Cilia, V., Messner, P., Raguenes, G., Gambacorta, A., Sleytr, U. B., Prieur, D. and Jeanthon, C. (1998). "Desulfurobacterium thermolithotrophum gen. nov., sp. nov., a novel autotrophic, sulphur-reducing bacterium isolated from a deep-sea hydrothermal vent." International Journal of Systematic Bacteriology **48**: 701-711.

L'haridon, S., Reysenbach, A. L., Tindall, B. J., Schonheit, P., Banta, A., Johnsen, U., Schumann, P., Gambacorta, A., Stackebrandt, E. and Jeanthon, C. (2006). "Desulfurobacterium atlanticum sp nov., Desulfurobacterium pacificum sp nov and Thermovibrio guaymasensis sp nov., three thermophilic members of the Desulfurobacteriaceae fam. nov., a deep branching lineage within the Bacteria." International Journal of Systematic and Evolutionary Microbiology **56**: 2843-2852.

Lamoreaux, P. E. and Tanner, J. T. (2001). Springs and Bottled Waters of the World:

Ancient History, Source, Occurrence, Quality and Use Springer.

Lan, Y. M., Wang, Q., Cole, J. R. and Rosen, G. L. (2012). "Using the RDP Classifier to Predict Taxonomic Novelty and Reduce the Search Space for Finding Novel Organisms." Plos One **7**(3).

Langmead, B. and Salzberg, S. L. (2012). "Fast gapped-read alignment with Bowtie 2." Nature Methods **9**(4): 357-U354.

Lau, M. C. Y., Aitchison, J. C. and Pointing, S. B. (2009). "Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet." Extremophiles **13**(1): 139-149.

Lee, H. F., Yang, T. F., Lan, T. F., Song, S. R. and Tsao, S. (2005). "Fumarolic gas composition of the Tatun Volcano Group, northern Taiwan." Terrestrial Atmospheric and Oceanic Sciences **16**(4): 843-864.

Li, R. Q., Zhu, H. M., Ruan, J., Qian, W. B., Fang, X. D., Shi, Z. B., Li, Y. R., Li, S. T., Shan, G., Kristiansen, K., Li, S. G., Yang, H. M., Wang, J. and Wang, J. (2010). "De novo assembly of human genomes with massively parallel short read sequencing." Genome Research **20**(2): 265-272.

Lin, C. H., Konstantinou, K. I., Liang, W. T., Pu, H. C., Lin, Y. M., You, S. H. and Huang, Y. P. (2005). "Preliminary analysis of volcanoseismic signals recorded at the Tatun Volcano Group, northern Taiwan." Geophysical Research Letters **32**(10).

Liu, C. M., Song, S. R., Chen, Y. L. and Tsao, S. (2011). "Characteristics and Origins of Hot Springs in the Tatun Volcano Group in Northern Taiwan." Terrestrial Atmospheric and Oceanic Sciences **22**(5): 475-489.

Lu, H. P., Lin, Y. K., Lin, L. H., Wang, P. L. and Yu, H. T. (unpublished data). The microbial community structure in acidic hot springs.

Luo, C. W., Tsementzi, D., Kyrpides, N., Read, T. and Konstantinidis, K. T. (2012). "Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample." Plos One **7**(2).

Macelroy, R. D. (1974). "Some Comments on Evolution of Extremophiles." Biosystems **6**(1): 74-75.

Macur, R. E., Langner, H. W., Kocar, B. D. and Inskeep, W. P. (2004). "Linking geochemical processes with microbial community analysis: successional dynamics in an arsenic-rich, acid-sulphate-chloride geothermal spring." Geobiology **2**(3): 163-177.

Mardanov, A. V., Gumerov, V. M., Beletsky, A. V., Perevalova, A. A., Karpov, G. A., Bonch-Osmolovskaya, E. A. and Ravin, N. V. (2011). "Uncultured archaea dominate in the thermal groundwater of Uzon Caldera, Kamchatka." Extremophiles **15**(3): 365-372.

Mardis, E. R. (2008). "Next-generation DNA sequencing methods." Annual Review of Genomics and Human Genetics **9**: 387-402.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z. T., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., Mcdade, K. E., Mckenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P. G., Begley, R. F. and Rothberg, J. M. (2005). "Genome

sequencing in microfabricated high-density picolitre reactors." Nature **437**(7057): 376-380.

Marteinsson, V. T., Hauksdottir, S., Hobel, C. F. V., Kristmannsdottir, H., Hreggvidsson, G. O. and Kristjansson, J. K. (2001). "Phylogenetic diversity analysis of subterranean hot springs in Iceland." Applied and Environmental Microbiology **67**(9): 4242-4248.

Mathur, J., Bizzoco, R. W., Ellis, D. G., Lipson, D. A., Poole, A. W., Levine, R. and Kelley, S. T. (2007). "Effects of abiotic factors on the phylogenetic diversity of bacterial communities in acidic thermal springs." Applied and Environmental Microbiology **73**(8): 2612-2623.

Maxam, A. M. and Gilbert, W. (1977). "A new method for sequencing DNA." Proceedings of the National Academy of Sciences **74**(2): 560-564.

Metzker, M. L. (2010). "Applications of Next-Generation Sequencing Sequencing Technologies - the Next Generation." Nature Reviews Genetics **11**(1): 31-46.

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

Minoche, A. E., Dohm, J. C. and Himmelbauer, H. (2011). "Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems." Genome Biology **12**(11).

Mitchelson, K. (2001). Overview: The Application of Capillary Electrophoresis for DNA Polymorphism Analysis. Capillary Electrophoresis of Nucleic Acids. K. Mitchelson and J. Cheng, Humana Press. **162:** 3-26.

Mullis, K. B. and Faloona, F. A. (1987). "Specific Synthesis of DNA Invitro Via a Polymerase-Catalyzed Chain-Reaction." Methods in Enzymology **155**:

335-350.

Nakagawa, S., Nakamura, S., Inagaki, F., Takai, K., Shirai, N. and Sako, Y. (2004). "Hydrogenivirga caldilitoris gen. nov., sp nov., a novel extremely thermophilic, hydrogen- and sulfur-oxidizing bacterium from a coastal hydrothermal field." International Journal of Systematic and Evolutionary Microbiology **54**: 2079-2084.

Nakagawa, S., Shtaih, Z., Banta, A., Beveridge, T. J., Sako, Y. and Reysenbach, A.-L. (2005a). "Sulfurihydrogenibium yellowstonense sp. nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus Sulfurihydrogenibium, Sulfurihydrogenibium subterraneum and Sulfurihydrogenibium azorense." International Journal of Systematic and Evolutionary Microbiology **55**(6): 2263-2268.

Nakagawa, S., Takai, K., Horikoshi, K. and Sako, Y. (2003). "Persephonella hydrogeniphila sp nov., a novel thermophilic, hydrogen-oxidizing bacterium from a deep-sea hydrothermal vent chimney." International Journal of Systematic and Evolutionary Microbiology **53**: 863-869.

Nakagawa, S., Takai, K., Inagaki, F., Chiba, H., Ishibashi, J., Kataoka, S., Hirayama, H., Nunoura, T., Horikoshi, K. and Sako, Y. (2005b). "Variability in microbial community and venting chemistry in a sediment-hosted backarc hydrothermal system: Impacts of subseafloor phase-separation." Fems Microbiology Ecology **54**(1): 141-155.

Nakagawa, T. and Fukui, M. (2002). "Phylogenetic characterization of microbial mats and streamers from a Japanese alkaline hot spring with a thermal gradient." Journal of General and Applied Microbiology **48**(4): 211-222.

Namiki, T., Hachiya, T., Tanaka, H. and Sakakibara, Y. (2012). "MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads." Nucleic Acids Research.

Ng, C. C., Chang, C. C. and Shyu, Y. T. (2005). "Archaeal community revealed by 16s rRNA and fluorescence in situ hybridization in a sulphuric hydrothermal hot spring, northern Taiwan." World Journal of Microbiology & Biotechnology **21**(6-7): 933-939.

Nishihara, H., Igarashi, Y. and Kodama, T. (1990). "A New Isolate of Hydrogenobacter, an Obligately Chemolithoautotrophic, Thermophilic, Halophilic and Aerobic Hydrogen-Oxidizing Bacterium from Seaside Saline Hot-Spring." Archives of Microbiology **153**(3): 294-298.

Noguchi, H., Park, J. and Takagi, T. (2006). "MetaGene: prokaryotic gene finding from environmental genome shotgun sequences." Nucleic Acids Research **34**(19): 5623-5630.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., De Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., Mchardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O. and Vonstein, V. (2005). "The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes." Nucleic Acids Research **33**(17): 5691-5702.

Pace, N. R. (1997). "A molecular view of microbial diversity and the biosphere." Science **276**(5313): 734-740.

Pei, A. Y., Oberdorf, W. E., Nossa, C. W., Agarwal, A., Chokshi, P., Gerz, E. A., Jin, Z. D., Lee, P., Yang, L. Y., Poles, M., Brown, S. M., Sotero, S., Desantis, T., Brodie, E., Nelson, K. and Pei, Z. H. (2010). "Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes." Applied and Environmental Microbiology **76**(12): 3886-3897.

Perevalova, A. A., Kolganova, T. V., Birkeland, N.-K., Schleper, C., Bonch-Osmolovskaya, E. A. and Lebedinsky, A. V. (2008). "Distribution of Crenarchaeota Representatives in Terrestrial Hot Springs of Russia and Iceland." Applied and Environmental Microbiology **74**(24): 7620-7628.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W. G., Peplies, J. and Glockner, F. O. (2007). "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB." Nucleic Acids Research **35**(21): 7188-7196.

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

Reigstad, L. J., Jorgensen, S. L. and Schleper, C. (2010). "Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka." Isme Journal **4**(3): 346-356.

Reysenbach, A. L., Ehringer, H. and Hershberger, K. (2000). "Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist." Extremophiles **4**(1): 61-67.

Reysenbach, A. L., Giver, L. J., Wickham, G. S. and Pace, N. R. (1992a). "Differential Amplification of Ribosomal-Rna Genes by Polymerase Chain-Reaction." Applied and Environmental Microbiology **58**(10): 3417-3418.

Reysenbach, A. L., Giver, L. J., Wickham, G. S. and Pace, N. R. (1992b). "Differential amplification of rRNA genes by polymerase chain reaction." Applied and Environmental Microbiology **58**(10): 3417-3418.

Reysenbach, A. L., Hamamura, N., Podar, M., Griffiths, E., Ferreira, S., Hochstein, R., Heidelberg, J., Johnson, J., Mead, D., Pohorille, A., Sarmiento, M., Schweighofer, K., Seshadri, R. and Voytek, M. A. (2009). "Complete and Draft Genome Sequences of Six Members of the Aquificales." Journal of Bacteriology **191**(6): 1992-1993.

Reysenbach, A. L., Wickham, G. S. and Pace, N. R. (1994). "Phylogenetic Analysis of the Hyperthermophilic Pink Filament Community in Octopus Spring, Yellowstone-National-Park." Applied and Environmental Microbiology **60**(6): 2113-2119.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. and Nyrén, P. (1996). "Real-Time DNA Sequencing Using Detection of Pyrophosphate Release." Analytical Biochemistry **242**(1): 84-89.

Rothschild, L. J. and Mancinelli, R. L. (2001). "Life in extreme environments." Nature **409**(6823): 1092-1101.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). "DNA Sequencing with Chain-Terminating Inhibitors." Proceedings of the National Academy of Sciences of the United States of America **74**(12): 5463-5467.

Satyanarayana, T., Raghukumar, C. and Shivaji, S. (2005). "Extremophilic microbes: Diversity and perspectives." Current Science **89**(1): 78-90.

Sayeh, R., Birrien, J. L., Alain, K., Barbier, G., Hamdi, M. and Prieur, D. (2010). "Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches." Extremophiles **14**(6): 501-514.

Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L., Yaschenko, E. and Ye, J. (2009). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Research **37**: D5-D15.

Schmidt, T. M., Delong, E. F. and Pace, N. R. (1991). "Analysis of a Marine Picoplankton Community by 16s Ribosomal-Rna Gene Cloning and Sequencing." Journal of Bacteriology **173**(14): 4371-4378.

Shima, S. and Suzuki, K. I. (1993). "Hydrogenobacter-Acidophilus Sp-Nov, a Thermoacidophilic, Aerobic, Hydrogen-Oxidizing Bacterium Requiring Elemental Sulfur for Growth." International Journal of Systematic Bacteriology **43**(4): 703-708.

Shokralla, S., Spall, J. L., Gibson, J. F. and Hajibabaei, M. (2012). "Next-generation sequencing technologies for environmental DNA research." Molecular Ecology **21**(8): 1794-1805.

Siering, P. L., Clarke, J. M. and Wilson, M. S. (2006). "Geochemical and biological diversity of acidic, hot springs in Lassen Volcanic National Park." Geomicrobiology Journal **23**(2): 129-141.

Simon, C. and Daniel, R. (2011). "Metagenomic Analyses: Past and Future Trends." Applied and Environmental Microbiology **77**(4): 1153-1161.

Skirnisdottir, S., Hreggvidsson, G. O., Hjorleifsdottir, S., Marteinsson, V. T., Petursdottir, S. K., Holst, O. and Kristjansson, J. K. (2000). "Influence of

sulfide and temperature on species composition and community structure of hot spring microbial mats." Applied and Environmental Microbiology **66**(7): 2835-2841.

Skirnisdottir, S., Hreggvidsson, G. O., Holst, O. and Kristjansson, J. K. (2001). "A new ecological adaptation to high sulfide by a Hydrogenobacter sp growing on sulfur compounds but not on hydrogen." Microbiological Research **156**(1): 41-47.

Sleator, R. D., Shortall, C. and Hill, C. (2008). "Metagenomics." Letters in Applied Microbiology **47**(5): 361-366.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., 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 United States of America **103**(32): 12115-12120.

Song, S. R., Yang, T. F., Yeh, Y. H., Tsao, S. J. and Lo, H. J. (2000). "The Tatun Volcano Group is active or extinct?" Journal of Geological Society of China **43**: 521-534.

Song, Z.-Q., Wang, F.-P., Zhi, X.-Y., Chen, J.-Q., Zhou, E.-M., Liang, F., Xiao, X., Tang, S.-K., Jiang, H.-C., Zhang, C. L., Dong, H. and Li, W.-J. (2012). "Bacterial and archaeal diversities in Yunnan and Tibetan hot springs, China." Environmental Microbiology: n/a-n/a.

Spear, J. R., Walker, J. J., Mccollom, T. M. and Pace, N. R. (2005). "Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem." Proceedings of the National Academy of Sciences of the United States of America **102**(7): 2555-2560.

Stein, L. D. (2011). An Introduction to the Informatics of “Next-Generation”

Sequencing. Current Protocols in Bioinformatics, John Wiley & Sons, Inc.

Takai, K., Kobayashi, H., Nealson, K. H. and Horikoshi, K. (2003a). "Sulfurihydrogenibium subterraneum gen. nov., sp nov., from a subsurface hot aquifer." International Journal of Systematic and Evolutionary Microbiology **53**: 823-827.

Takai, K., Komatsu, T. and Horikoshi, K. (2001a). "Hydrogenobacter subterraneus sp nov., an extremely thermophilic, heterotrophic bacterium unable to grow on hydrogen gas, from deep subsurface geothermal water." International Journal of Systematic and Evolutionary Microbiology **51**: 1425-1435.

Takai, K., Komatsu, T., Inagaki, F. and Horikoshi, K. (2001b). "Distribution of archaea in a black smoker chimney structure." Applied and Environmental Microbiology **67**(8): 3618-3629.

Takai, K., Nakagawa, S., Sako, Y. and Horikoshi, K. (2003b). "Balnearium lithotrophicum gen. nov., sp nov., a novel thermophilic, strictly anaerobic, hydrogen-oxidizing chemolithoautotroph isolated from a black smoker chimney in the Suiyo Seamount hydrothermal system." International Journal of Systematic and Evolutionary Microbiology **53**: 1947-1954.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. and Koonin, E. V. (2001). "The COG database: new developments in phylogenetic classification of proteins from complete genomes." Nucleic Acids Research **29**(1): 22-28.

Tekere, M., Lotter, A., Olivier, J., Jonker, N. and Venter, S. (2011). "Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa." African Journal of Biotechnology **10**(78): 18005-18012.

Temperton, B., Field, D., Oliver, A., Tiwari, B., Muhling, M., Joint, I. and Gilbert, J.

A. (2009). "Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing." Isme Journal **3**(7): 792-796.

Tobler, D. J. and Benning, L. G. (2011). "Bacterial diversity in five Icelandic geothermal waters: temperature and sinter growth rate effects." Extremophiles **15**(4): 473-485.

Todd, D. K. (1980). Groundwater Hydrology. New York, Wiley.

Tomova, I., Stoilova-Disheva, M., Lyutskanova, D., Pascual, J., Petrov, P. and Kambourova, M. (2010). "Phylogenetic analysis of the bacterial community in a geothermal spring, Rupi Basin, Bulgaria." World Journal of Microbiology & Biotechnology **26**(11): 2019-2028.

Tringe, S. G., Von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholtz, P. and Rubin, E. M. (2005). "Comparative metagenomics of microbial communities." Science **308**(5721): 554-557.

Turnbaugh, P. J. and Gordon, J. I. (2008). "An invitation to the marriage of metagenomics and metabolomics." Cell **134**(5): 708-713.

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.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D. Y., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H. and Smith, H. O. (2004). "Environmental genome shotgun sequencing of the Sargasso Sea."

Science **304**(5667): 66-74.

Vetriani, C., Speck, M. D., Ellor, S. V., Lutz, R. A. and Starovoytov, V. (2004). "Thermovibrio ammonificans sp nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents." International Journal of Systematic and Evolutionary Microbiology **54**: 175-181.

Vick, T. J., Dodsworth, J. A., Costa, K. C., Shock, E. L. and Hedlund, B. P. (2010). "Microbiology and geochemistry of Little Hot Creek, a hot spring environment in the Long Valley Caldera." Geobiology **8**(2): 140-154.

Von Mering, C., Hugenholtz, P., Raes, J., Tringe, S. G., Doerks, T., Jensen, L. J., Ward, N. and Bork, P. (2007). "Quantitative phylogenetic assessment of microbial communities in diverse environments." Science **315**(5815): 1126-1130.

Wang, W. S. and Chen, C. H. (1990). "The volcanology and fission track age dating of pyroclastic deposits in Tatun Volcano Group, northern Taiwan." Acta Geologica Taiwanica **28**: 1-30.

Weisburg, W. G., Barns, S. M., Pelletier, D. A. and Lane, D. J. (1991). "16s Ribosomal DNA Amplification for Phylogenetic Study." Journal of Bacteriology **173**(2): 697-703.

Wilson, M. S., Siering, P. L., White, C. L., Hauser, M. E. and Bartles, A. N. (2008). "Novel archaea and bacteria dominate stable microbial communities in North America's largest hot spring." Microbial Ecology **56**(2): 292-305.

Wilson, Z. E. and Brimble, M. A. (2009). "Molecules derived from the extremes of life." Natural Product Reports **26**(1): 44-71.

Woese, C. R. (1987). "Bacterial Evolution." Microbiological Reviews **51**(2): 221-271.

Wu, S. T., Zhu, Z. W., Fu, L. M., Niu, B. F. and Li, W. Z. (2011). "WebMGA: a

customizable web server for fast metagenomic sequence analysis." Bmc Genomics **12**.

Yamamoto, H., Hiraishi, A., Kato, K., Chiura, H. X., Maki, Y. and Shimizu, A. (1998). "Phylogenetic evidence for the existence of novel thermophilic bacteria in hot spring sulfur-turf microbial mats in Japan." Applied and Environmental Microbiology **64**(5): 1680-1687.

Yang, T. F., Sano, Y. and Song, S. R. (1999). "$^{3}He/^{4}He$ ratio of fumaroles and bubbling gases of hot spring in Tatun Volcano Group, North Taiwan." Nuovo Cimento **22**(0304): 281-286.

Yim, L. C., Jing, H. M., Aitchison, J. C. and Pointing, S. B. (2006). "Highly diverse community structure in a remote central Tibetan geothermal spring does not display monotonic variation to thermal stress." Fems Microbiology Ecology **57**(1): 80-91.

Zhao, W. D., Song, Z. Q., Jiang, H. C., Li, W. J., Mou, X. Z., Romanek, C. S., Wiegel, J., Dong, H. L. and Zhang, C. L. (2011). "Ammonia-oxidizing Archaea in Kamchatka Hot Springs." Geomicrobiology Journal **28**(2): 149-159.

何珮楨 (2004). 台灣地區不同溫泉中微生物組成的研究. 碩士, 東吳大學.

李信緯 (2009). 陽明山國家公園溫泉微生物多樣性分析. 碩士, 東吳大學.

李重義 (2006). 陽明山國家公園溫泉微生物多樣性與利用之研究. 陽明山國家公園管理處. 臺北市.

李重義 and 張怡塘 (2009). 陽明山國家公園酸性溫泉物種調查及生存機制之研究. 陽明山國家公園管理處. 臺北市.

陳懋彥 (2002). 台灣地熱區嗜熱性細菌之研究. 博士, 國立臺灣大學.

鄭婷文 and 林立虹 (2007). 酸性熱泉中底生性與懸浮性微生物族群結構之比較. 2007 台灣地球科學聯合學術研討會.

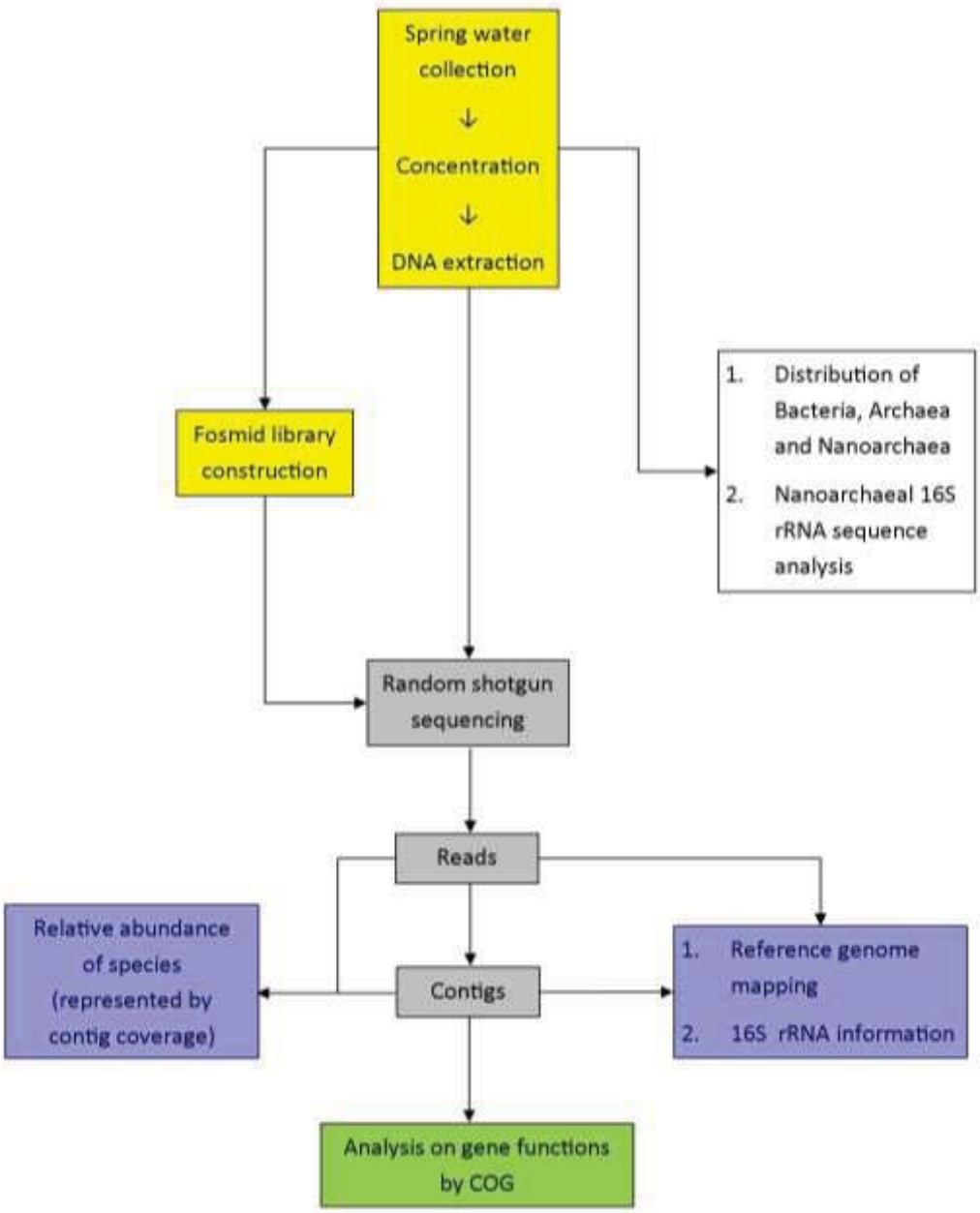


Figure 1. Workflow of this study. Yellow: sample preparation stage; white: general survey of microbial distribution; gray: sequencing and data processing stage; blue: community structure analyses; green: overview on gene functions of the community.

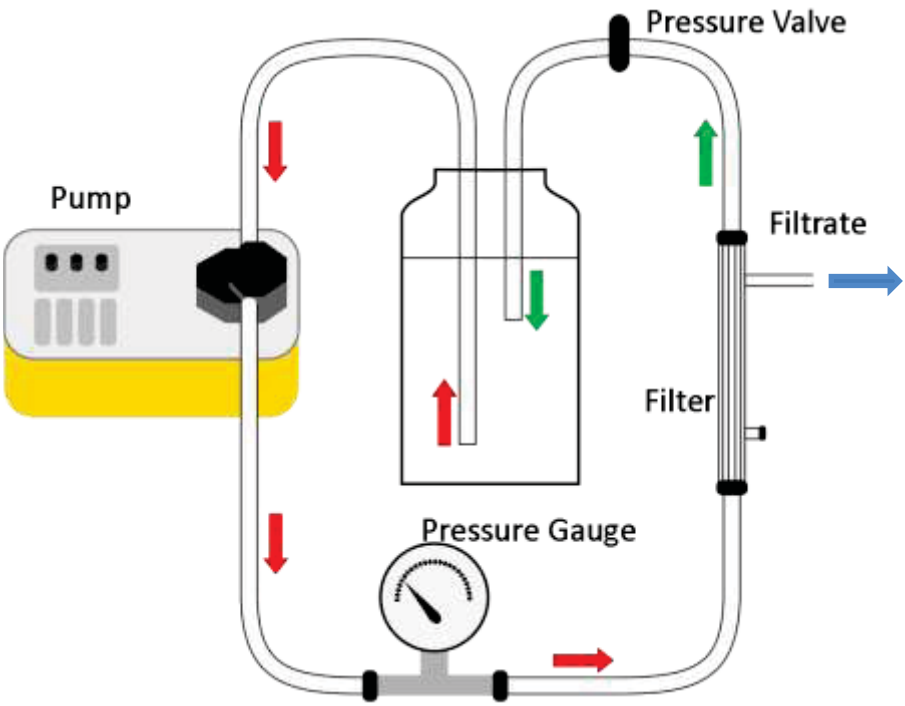
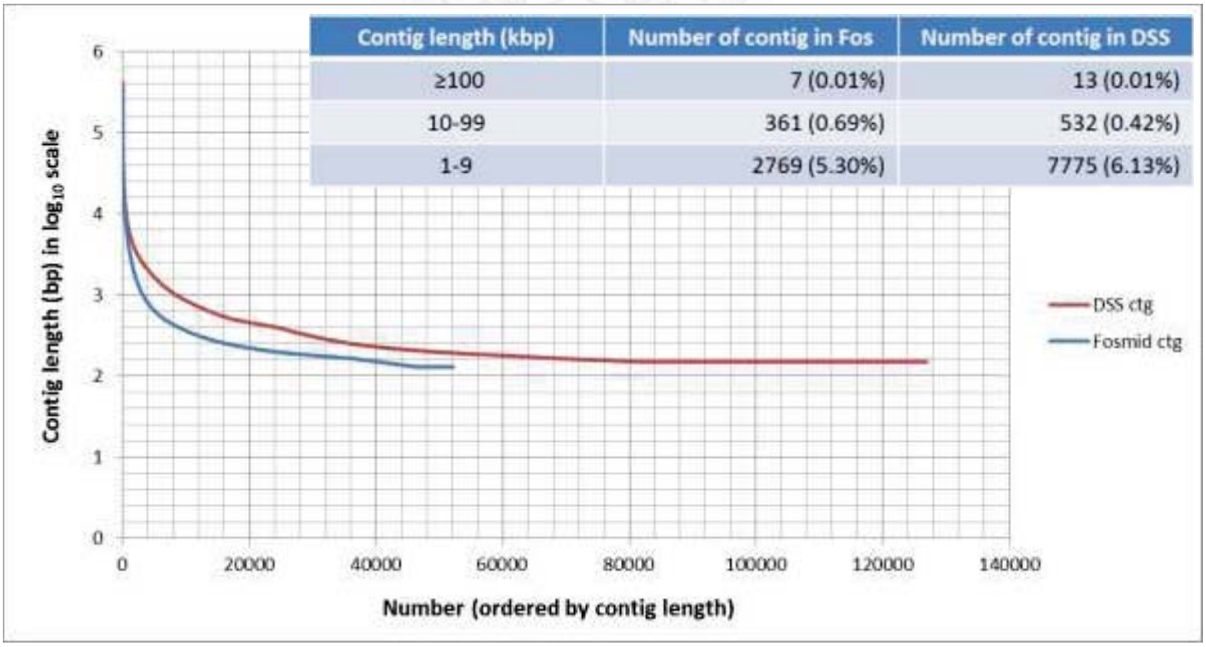


Figure 2. Tangential flow system for spring water concentration. Red arrow: water to be concentrated; green arrow: retentate obtained after concentration; blue arrow: removal of excessive water from the system through filtration.



Contig length (kbp)	Number of contig in Fos	Number of contig in DSS
≥100	7 (0.01%)	13 (0.01%)
10-99	361 (0.69%)	532 (0.42%)
1-9	2769 (5.30%)	7775 (6.13%)

Figure 3. Contig length distribution of fosmid and DSS sample

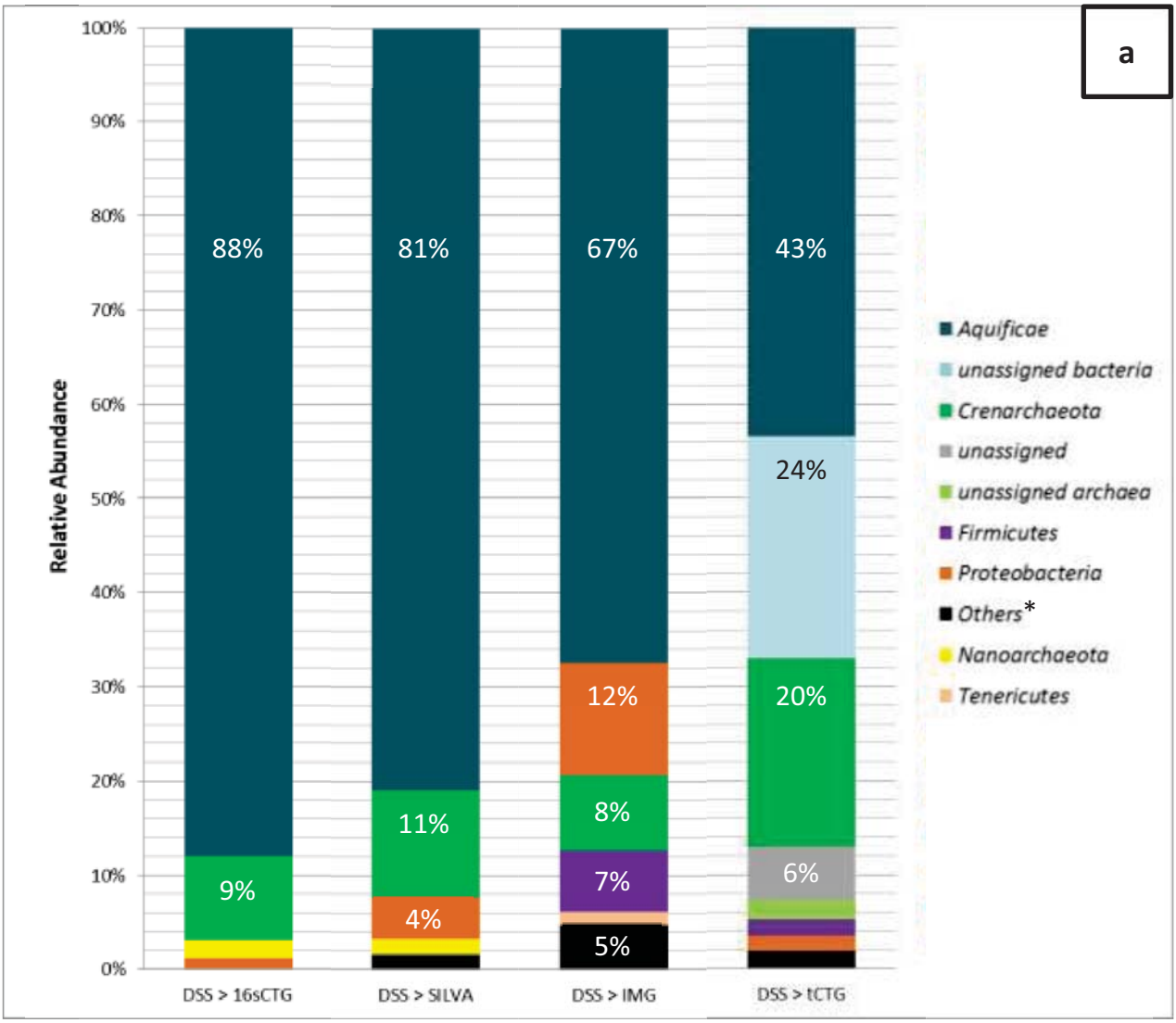


Figure 4. Relative abundance distribution of DSS sample. (a) Abundance assessed by 4 approaches.*Phyla with relative abundance lower than 1% were grouped into "Others". Please refer to Table 8 for details.

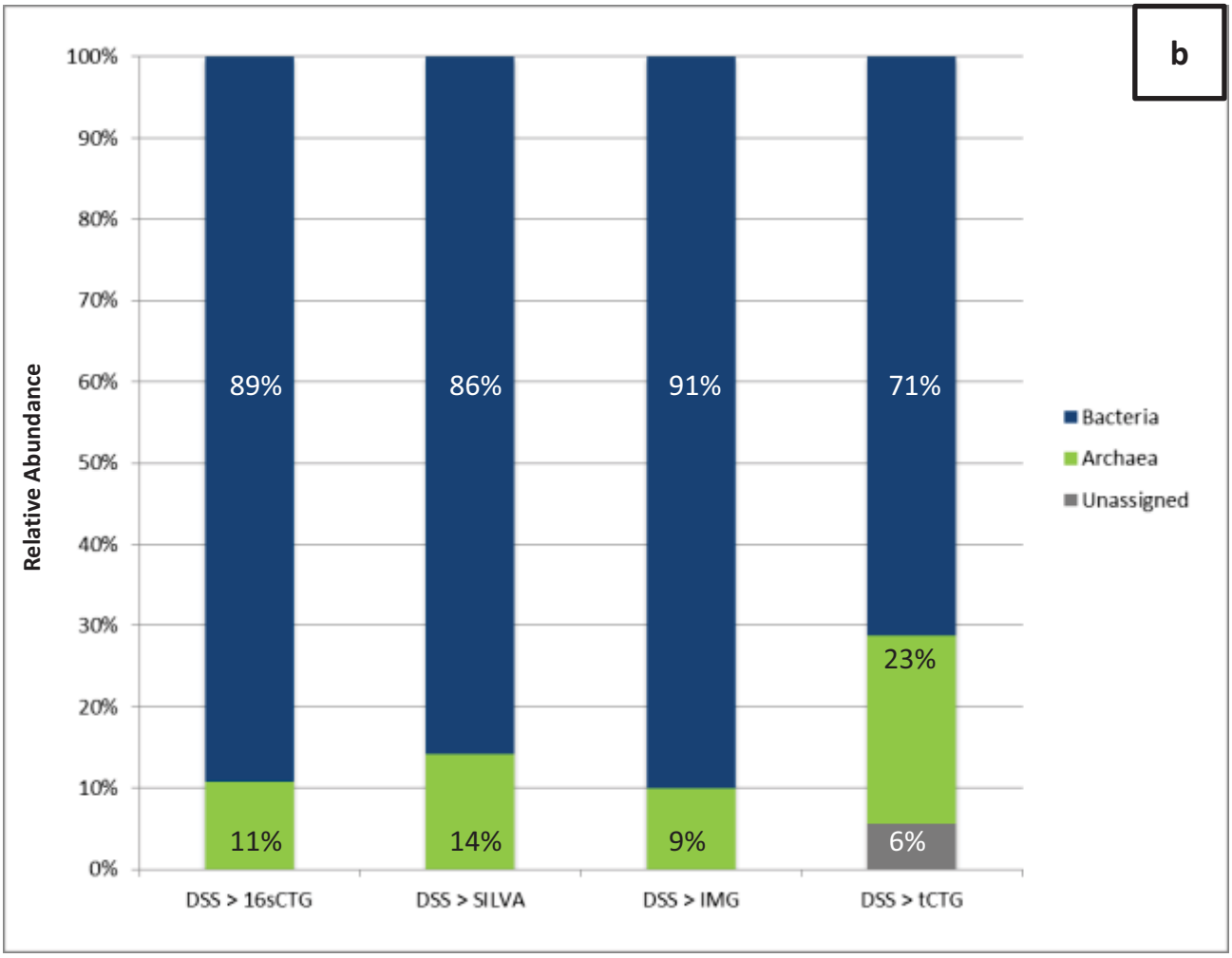


Figure 4. Relative abundance distribution of DSS sample. (b) Proportions of bacteria, archaea and unassigned prokaryotes

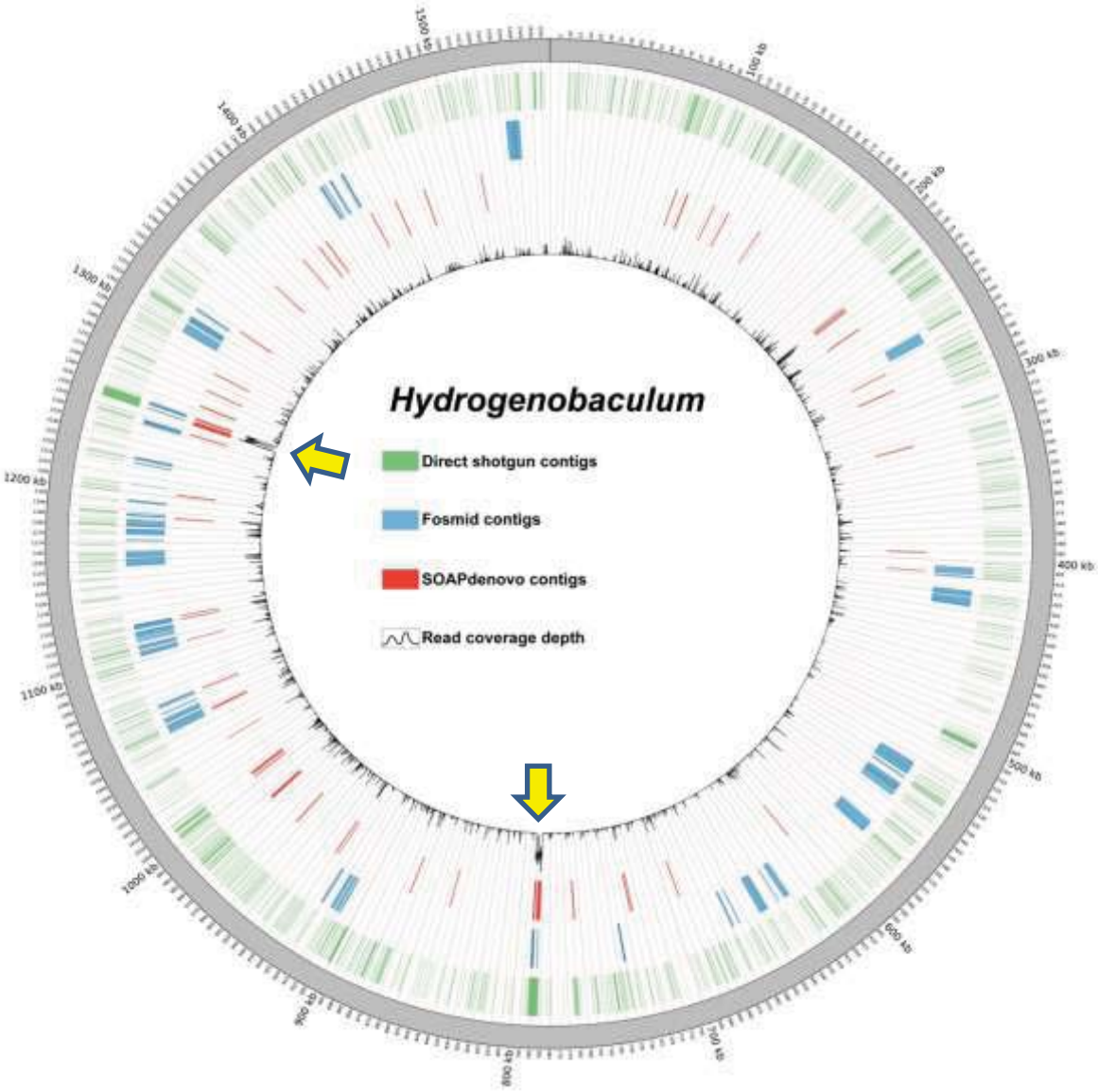


Figure 5. Mapping status of reference Hydrogenobaculum sp. Y04AAS1. Reference genome was mapped by DSS raw reads (black), de novo assembled contigs by SOAPdenovo (red), fosmid contigs (blue) and DSS contigs (green). Yellow arrows marked the locations of 2 rRNA operons.

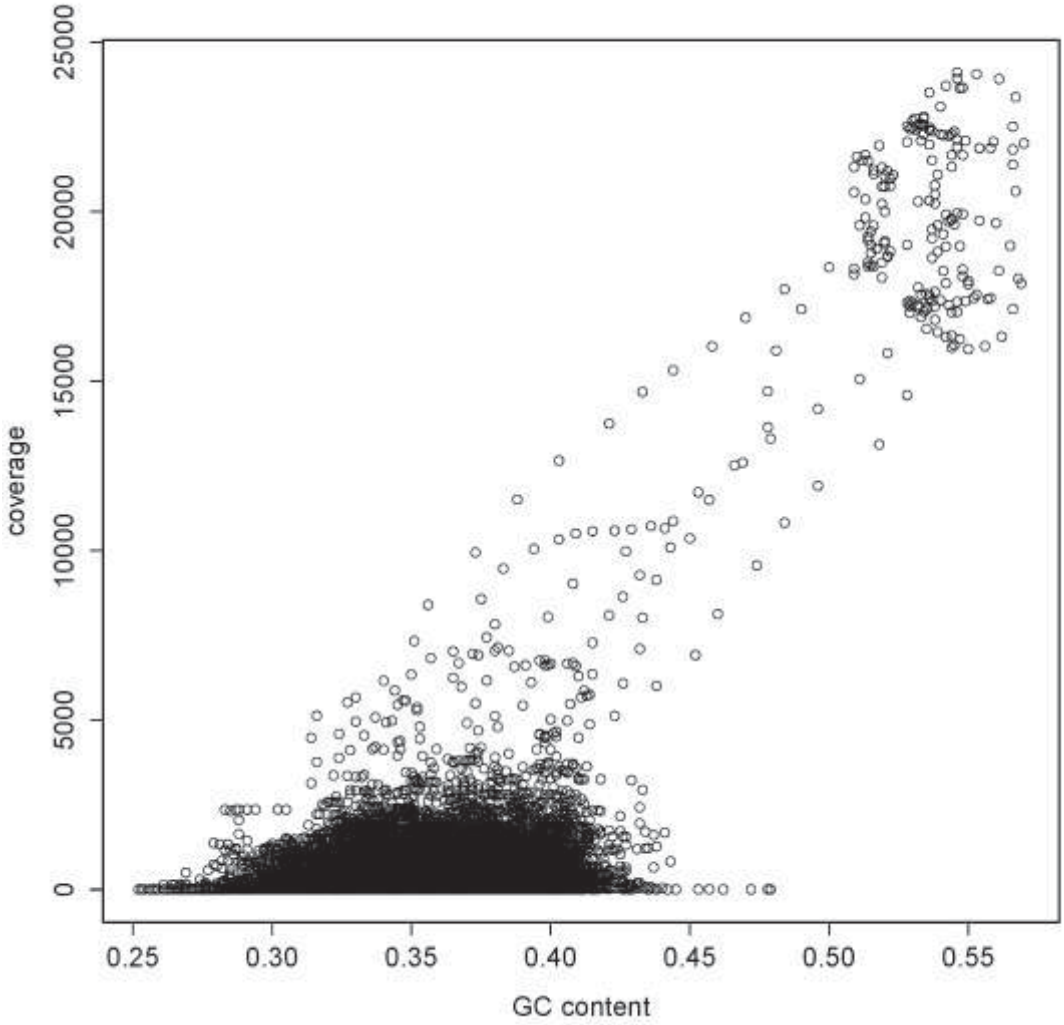


Figure 6. Relationship between sequence coverage and GC content.

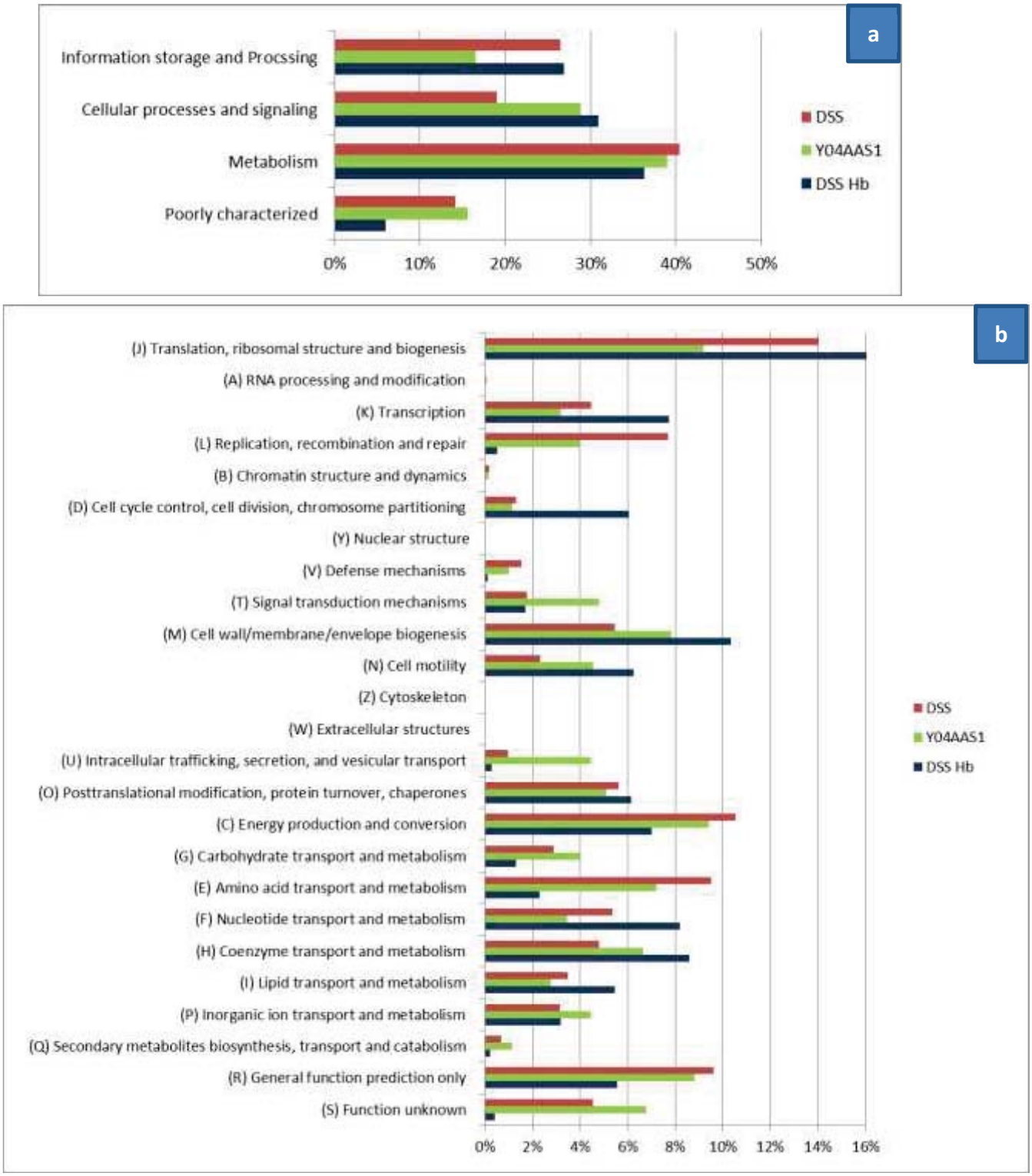


Figure 7. COG distribution of gene functions. (a) COGs grouped by general category. (b) Distribution pattern of each COG class. DSS: COG distribution represented by total DSS contigs. Y04AAS1: COG distribution of reference genome Hydrogenobaculum sp. Y04AAS1. DSS Hb: COG distribution represented by Hydrogenobaculum contigs extracted from DSS contigs

Table 1. Hot spring classification based on temperature

Types	Temperature Range
Low-Temperature Hot Spring	Lower than 49°C
Medium-Temperature Hot Spring	50-74°C
High-Temperature Hot Spring	75-96°C
Boiling Spring	Higher than 97°C

Table 2. Hot spring classification based on physical properties

Types	Physical Properties
Ordinary Hot Spring	With temperatures below boiling point, and in most cases spring water is constantly welling. One of the most widely distributed types of terrestrial hot springs.
Geyser	Characterized by intermittent discharge of water. Pause time and water temperatures vary, some are above boiling point.
Boiling Spring	With temperatures from 93 to 100°C (depends on altitude).
Fumaroles or Solfataras	Underground water was heated and vaporized (the source temperature is usually higher than 100°C), and then migrated upwards to surface. If the emitting gas is mainly water vapor, then the degassing structure is called fumarole; if H_2S is dominant, then it is called solfatara.
Hot Mud Spring	The discharged hot water is with large amount of mud. The major constituents in mud are clay minerals, which are derived from the long term interaction between hot water and surrounding rocks.

Table 3. Hot spring classification based on chemical properties

Category: subcategory	**Major Ions**	**pH Range**	**Chinese Nomenclature**
1. Chloride Spring	**Cl^-**		**氯化物泉**
Carbonate-chloride spring	Na^+, Cl^- > HCO_3^-	8-9	碳酸氫鈉氯化物泉
Acidic sulfate-chloride spring	H^+ > Na^+, Cl^- > SO_4^{2-}	1-4	酸性硫酸鹽氯化物泉
Neutral sulfate-chloride spring	Na^+ > Ca^{2+}, Cl^- > $SO4^{2-}$	8-9	中性硫酸鹽氯化物泉
2. Hydrocarbonate Spring	**HCO3-**		**碳酸氫鹽泉**
Sodium hydrocarbonate spring	H^+ > Na^+, HCO_3^-	7-10	碳酸氫鈉泉
Sodium/Calcium-hydrocarbonate spring	Na^+ > Ca^{2+}, HCO_3^-	7-8	碳酸氫鈣鈉泉
Sulfate-sodium hydrocarbonate spring	Na^+, HCO_3^- > SO_4^{2-}	7-8	硫酸鹽碳酸氫鈉泉
Chloride-hydrocarbonate spring	Na^+, HCO_3^- > Cl^-	6-7	氯化物碳酸氫鈉泉
3. Sulfate Spring	**SO_4^{2-}**		**硫酸鹽泉**
Acidic sulfate spring	H^+ > Na^+, Ca^{2+}, $SO4^{2-}$	1-3	酸性硫酸鹽泉
Neutral sulfate spring	Ca^{2+} > Na^+, $SO4^{2-}$ > HCO_3^-	6-7	中性硫酸鹽泉
4. Others			
	$NaHCO_3$ (Baking Soda)		重曹泉
	$Ca(HCO_3)_2$ or $Mg(HCO_3)_2$		重碳酸土類泉
	NaCl (Salt)		食鹽泉
	$CaCl_2$ or $MgCl_2$		氯化土鹽類
	Na_2SO_4 (Glauber's Salt)		芒硝泉
	$CaSO_4$ (Gypsum)		石膏泉
	$MgSO_4$ (Epsom Salts)		正苦味泉

Table 4. Average geochemical and physical parameters in SHP (四磺坪)

Parameters	This study	Liu et al. (2011)	Chen and Sung (2009)
pH	2.8*	3.2	2.74
Temperature (°C)	73.0*	77.6	68.5
DO	-	-	2.74
TDS	707	245	-
EC (μ S/cm^2)	1709*	-	275
ORP (mv)	-62	-	-145
DOC	1.0	-	-
HCO_3^{2-}	< 0.03	-	< 0.1
Cl^- and other halides	9.3	20.3 (Cl^-)	0.21 (F^-) / 8.9 (Cl^-) / <0.1 (Br^-)
SO_4^{2-}	378	346	97.8
S^0	0.50	-	-
H_2S	52.7	-	> 5
$S_2O_3^{2-}$	0.12	-	-
Total Fe (Fe^{2+} and Fe^{3+})	111	-	26.7
Ca^{2+}	1.32	5.3	9.8
Na^+	12.0	5.5	6.6
Mg^{2+}	1.21	2.0	2.8
K^+	2.17	1.5	< 0.1
Al^{3+}	19.0	-	-
Total As	0.0012	-	0.03

Unit mg/L was applied to parameters except pH.

DO: dissolved oxygen; TDS: total dissolved solid; EC: electrical conductivity; ORP: oxidation/reduction potential; DOC: dissolved organic carbon.

*: Average value derived from multiple samples at the same location

Table 5. Sequencing and assembly of fosmid clones (Fos)

(a) Sequencing summary of 1485 fosmid clones

Total Reads	224,299,204
Total Reads after QT	219,083,057
Reads after vector removal	159,891,223
Percentage of vector in total reads after QT	27%
Read Length (bp)	101
Read Length after QT (bp)	94.4
Read Length after vector removal (bp)	93.6

QT: quality trim (criteria: min length = 35 bp, error probability <0.05)

(b) De novo assembly of Fos by MetaVelvet

Number of contig	52,221
Average contig length (bp)	559.45
Max / min contig length (bp)	325,311 / 129
Contig ≥ 300 bp, number (percentage)	12,648 (24.2%)
Contig ≥ 1 kb, number (percentage)	3,137 (6.0%)
Contig ≥ 40 kb, number (percentage)	45 (0.09%)
N50	2,693
Number of "N"	48,975
N-ratio	0.002

(c) De novo assembly of Fos by CLC

Number of contig	10,121
Average contig length (bp)	2454.69
Max / min contig length (bp)	326,294 / 183
Contig ≥ 300 bp, number (percentage)	7,388 (73%)
Contig ≥ 1 kb, number (percentage)	3,241 (32%)
Contig ≥ 40 kb, number (percentage)	56 (0.5%)
N50	14,643
Number of "N"	539,244
N-ratio	0.02

Table 6. Direct shotgun sequencing and assembly of sample (DSS)

(a) Summary of sample derived from direct shotgun sequencing (DSS)

Total Reads	557,415,266
Total Reads after QT	548,895,370
Read Length (bp)	101
Read Length after QT (bp)	98

QT: quality trim (criteria: min length = 35 bp, error probability <0.05)

(b) De novo assembly of DSS by MetaVelvet

Number of contig	126,849
Average contig length (bp)	478.4
Max / min contig length (bp)	313,819 / 149
Contig ≥ 300 bp, number (percentage)	43,015 (23.8%)
Contig ≥ 1 kb, number (percentage)	10,014 (5.5%)
Contig ≥ 40 kb, number (percentage)	112 (0.06%)
N50	1,191
Number of "N"	1,811,850
N-ratio	0.02

Table 7. Percentage of DSS reads used in each analysis on community structure

Analyses	**Read 1** (number / percentage)	**Read 2** (number / percentage)	**Average Percentage**
DSS > 16sCTG	565,058 / 0.10 %	525,852 / 0.09 %	0.10 %
DSS > SILVA	352,312 / 0.06 %	342,532 / 0.06 %	0.06 %
DSS > IMG	15,722,577 / 1.41 %		1.41 %
DSS > tCTG	367,480,112 / 66.0 %		66 %

Total read number of DSS shotgun sequencing was 557,415,266

Table 8. Assessments on relative abundance of DSS sample through 4 approaches

DSG > 16sCTG		DSS > SILVA		DSS > IMG		DSS > tCTG	
Aquificae	88%	*Aquificae*	80.9833%	*Aquificae*	67.438%	*Aquificae*	43.389%
Crenarchaeota	8.9%	*Crenarchaeota*	11.3544%	*Proteobacteria*	11.915%	*unassigned bacteria*	23.595%
Nanoarchaeota	1.9%	*Proteobacteria*	4.4329%	*Crenarchaeota*	8.008%	*Crenarchaeota*	20.059%
Proteobacteria	1.1%	*Nanoarchaeota*	1.6300%	*Firmicutes*	6.516%	*unassigned*	5.670%
Actinobacteria	0.1%	*Thaumarchaeota*	0.7035%	*Tenericutes*	1.346%	*unassigned archaea*	2.017%
		Euryarchaeota	0.5481%	*Euryarchaeota*	0.744%	*Firmicutes*	1.750%
		Actinobacteria	0.1758%	*Actinobacteria*	0.655%	*Proteobacteria*	1.652%
		Thermotogae	0.0560%	*Bacteroidetes*	0.567%	*Euryarchaeota*	0.965%
		Firmicutes	0.0443%	*Thermotogae*	0.363%	*Deferribacteres*	0.209%
		Cyanobacteria	0.0272%	*Deinococcus-Thermus*	0.344%	*Bacteroidetes*	0.200%
		Tenericutes	0.0133%	*Chlamydiae*	0.289%	*Cyanobacteria*	0.145%
		Chloroflexi	0.0064%	*Fusobacteria*	0.273%	*Nitrospirae*	0.062%
		Gemmatimonadetes	0.0063%	*Spirochaetes*	0.271%	*Thermotogae*	0.060%
		Nitrospirae	0.0056%	*Thaumarchaeota*	0.209%	*Actinobacteria*	0.035%
		Korarchaeota	0.0036%	*Synergistetes*	0.200%	*Thermodesulfobacteria*	0.032%
		Deferribacteres	0.0034%	*Cyanobacteria*	0.175%	*Nanoarchaeota*	0.029%
		Spirochaetes	0.0013%	*Deferribacteres*	0.142%	*Planctomycetes*	0.028%
		Armatimonadetes	0.0012%	*Dictyoglomi*	0.111%	*Thaumarchaeota*	0.023%
		Deinococcus-Thermus	0.0011%	*Chloroflexi*	0.109%	*Chloroflexi*	0.017%
		Planctomycetes	0.0006%	*Nanoarchaeota*	0.075%	*Chlamydiae*	0.012%
		Synergistetes	0.0006%	*Acidobacteria*	0.068%	*Spirochaetes*	0.011%
		Fusobacteria	0.0005%	*Nitrospirae*	0.050%	*Dictyoglomi*	0.011%
		Caldiserica	0.0004%	*Verrucomicrobia*	0.047%	*Tenericutes*	0.005%
		Elusimicrobia	0.0003%	*Planctomycetes*	0.031%	*Korarchaeota*	0.004%
				Elusimicrobia	0.027%	*Synergistetes*	0.004%
				Chlorobi	0.009%	*Ignavibacteria*	0.003%
				Lentisphaerae	0.008%	*Verrucomicrobia*	0.002%
				Fibrobacteres	0.006%	*Deinococcus-Thermus*	0.002%
				Chrysiogenetes	0.003%	*Fusobacteria*	0.001%
				TM7	0.001%	*Chlorobi*	0.001%
				Gemmatimonadetes	0.001%	*Acidobacteria*	0.001%
						Elusimicrobia	0.0004%

6. Appendix

Appendix 1. General Survey of Prokaryotes in TVG Area

1. Introduction

Previous studies on hot spring microbial diversity in TVG area have revealed the unexpected distribution pattern of bacteria and archaea (Cheng et al., 2013; Lu et al., unpublished data): archaea or bacteria were absent at some locations, defying the concept of ubiquitous distribution of microbes. Also, members of an archaeal phylum Nanoarchaeota (Huber et al., 2002) were discovered in the two studies; however, the suspected host *Ignicoccus hospitalis* was not found in those studies. To further confirm the distribution of bacteria and archaea and to assess the phylogenetic affinity of nanoarchaeum-like prokaryotes in TVG area, supplemental analyses were done during the period of main research.

2. Material and Methods

2.1 Sample Collection, Concentration and DNA Extraction

Six sites in TVG area were chosen for sampling (Supplemental Figure 1 and 2, Supplemental Table 1). Same procedures were used as described in Chapter 2.

2.2 Polymerase Chain Reactions on 16 S rRNA Genes

The purpose of amplification of 16 S rDNA genes was to detect the existence of bacteria, archaea and nanoarchaea in TVG area. Primer sets used in this study for detection were listed in Supplemental Table 2. Each 10 μL PCR reaction contained 1 μL DNA, 1 mM of each dNTP, 1.5 mM $MgCl_2$ (Promega), 1× PCR buffer (Promega), 1 μM of each primer, and 0.5 U DNA Polymerase (GoTaq® Flexi DNA polymerase, Promega). All primers were first tested in order to select the most suitable sets. Amplification included initiation, chain-reaction (denature, annealing and extension) and termination phases. Conditions of several steps were identical for all reactions. Initiation was 94°C for 5 min, followed by chain reaction: denature step was 94°C for 36 30-sec cycles. Termination was 72°C for 7 min and then preserved in 4°C. Annealing temperature and extension time differed with each primer sets: annealing temperature for B27F/U1406R, 8mcF/934mcR, 8mcF/1513mcR and

9bF/515mcR was 53 to 55°C; for B27F/U1492R, A109F/A958R, 23F/U1492R, A21F/A958R and 519mcF/1114mcR was 50 to 53°C. Duration for extension depended on expected product length of each primer sets and efficiency of polymerase: 90 sec for B27F/U1406R, B27F/U1492R, 23F/U1492R and 8mcF/1513mcR; 60 sec for A109F/A958R, A21F/A958R, 8mcF/934mcR, 9bF/515mcR and 519mcF/1114mcR.

PCR products were checked with 1% agarose gel, and then purified with QIAquick® PCR purification Kit (QIAGEN, USA) according to user manual provided by manufacturer.

2.3 Sequence Analysis on Nanoarchaeal 16 S rRNA Gene

Since nanoarchaea were detected previously in this area but their 16 S rDNA sequences revealed certain level of distance to reference species *Nanoarchaeum equitans* clone CU-1 and OP-9 (Lu et al., unpublished data), analysis on nanoarchaeal 16 S rDNA sequences were performed in this study to confirm the earlier result. Cloning process contained ligation of PCR fragments into vectors and transformation. Purified nanoarchaeal 16 S rDNA partial genes were ligated with T&A® Cloning Vector Kit (Yeastern Biotech Co. Ltd.) followed the instructions in user manual.

50 µL of competent cells (ECOS^{TM} 101 competent cells, Yeastern Biotech Co. Ltd.) in each eppendorf tube were put on ice. 10 µL of ligated DNA were added to each tube of competent cells when 1/3 of competent cells thawed. Ligated DNA and competent cells were mixed gently by tapping the tube. Competent cells were heat-shocked for in 42°C water bath for 45 sec, and then put on ice for 15 min. 50 µL of X-Gal (20 µg /µL) and 50 µL IPTG (0.1M) were added to each tube. Each reaction was plated on an Ampicillin-containing (100 µg / mL) LB plate, and incubated in 37°C for 14 to 16 hours.

Positive colonies were picked to a new Amp-plate and insert size were checked by PCR with primer sets 8mcF/934mcR and same conditions (except cycle number was reduced to 30). Colonies bore successfully transformed

nanoarchaeal 16 S rDNA genes were sent to sequencing (Mission Biotech Co., Ltd., Taipei, Taiwan).

Sequenced nanoarchaea-like sequences amplified by PCR were first searched using blastn in the nr database for closest known species. Several reference 16 S rRNA genes were then selected for alignment: *Nanoarchaeum equitans* strain Kin4-M (AJ318041.1), uncultured nanoarchaeotes clone CU-1 (AJ458437.1) and clone OP-9 (AJ458436.1), *Escherichia coli* (AY776275.1), *Stygiolobus azoricus* strain DSM 6296 (NR_043434.1), *Vulcanisaeta distributa* DSM 14429 strain IC-017 (NR_040876.1), *Sulfolobus tokodaii* strain TW (EU545801.1), *Acidilobus saccharovorans* strain 345-15 (AY350586.2) and *Thermoplasma acidophilum* strain 122-1B2 (NR_028235.1). Alignment was done by MEGA5 (Tamura et al., 2011) with ClustalW method. Default parameters settings were used: gap opening penalty and gap extension penalty were 15 and 6.66 respectively for pairwise and multiple alignments; DNA weight matrix was IUB and transition weight was 0.5 with 30% as delay divergent cutoff. Neighbor-joining tree was constructed using bootstrap method according to the alignment result. Number of bootstrap replications was 500; substitution type was nucleotide and model was p-distance. Transitions and transversions were both included in substitution. Uniform rates and same (homogenous) were selected in the option rate among sites and pattern among lineages, respectively. Gaps or missing data were completely deleted; all position including noncoding sites were selected in the data subset. Parameters set for final calculation on pairwise distance were the same as that in neighbor-joining tree construction (Supplemental Table 3).

3. Results

3.1. Environmental Metadata and DNA Extraction

3.1.1. Sample Sites Descriptions and Properties of Hot Spring Waters

In order to know the distribution pattern of bacteria, archaea and nanoarchaea in TVG region, hot spring waters from 6 locations LFG, ZSL, XYK, MC, HS and SHP (Supplemental Figure 1) were collected for metagenomic DNAs

extraction (Supplemental Table 1).

LFG is a highly developed hot spring recreation center. Water was sampled in a low temperature pool (Supplemental Figure 2a) without measuring pH value and temperature (tested only by hand, around 40°C). The color of water was greyish white and non-transparent. ZSL sample was collected in a hot spring pool surrounded by an artificial crossable fence at government-possessed Zhong Shan Building, which is a restricted area only for permitted visitors and was relatively less disturbed by human activities. At the edge of pool there was certain amount of deposit sediments in greyish white color and the water was muddy grey with visible suspension particles (Supplemental Figure 2b). Water temperatures measured in ZSL were slightly higher than that in LFG, around 50°C. XYK was located in a protected national park region and entry of tourists is strictly forbidden. Water sample was obtained at a collapsing rocky slope under a huge fumarole. Without any pool or pond structure formed on the steep slope, spouting water with temperature close to boiling point was collected (Supplemental Figure 2c). Compared to other sites in this study, water from XYK was with the highest transparency. Samples of MC were from two artificial pools (Supplemental Figure 2d) at a hill slope of the valley under Ma-Cao Bridge (馬槽橋). HS (Supplemental Figure 2e) is located in a sulfur mining field and was characterized by conspicuous fumaroles, yellow crystals of elemental sulfur and strong smell of H_2S gas. Samples were taken from artificial hot spring pools selected for higher temperatures. On the other side of mountain to HS, SHP (Supplemental Figure 2f) is an area also partly belongs to a mining field. In general, geomorphic features of HS and SHP are quite similar.

3.1.2. DNA Extraction

Waters obtained from XYK and ZSL were the two with distinguishable appearances: the former was relatively transparent compared to others, and the latter was highly particle-rich, which was also found to be problematic and filter membrane was frequently blocked during the concentration process. The amounts of extracted DNA from ZSL (sampling under heavy rain) and SHP

samples were quantified. Metagenomic DNAs extracted after sample concentration according to protocol provided by manufacturer yielded averagely 3.73 μg and 11.8 μg of DNAs per liter of hot spring water from SHP and ZSL, respectively. Due to requirements of further experimental steps, these extracted DNAs were concentrated with isopropanol and ethanol-based protocol according to user manual of DNA extraction kit. From the individual results derived in this study, if no bias from DNA quantitation process was introduced, great loss of DNAs during second concentration done by both methods was usually about 30 to 50%. Occasionally, the loss could be up to 90%.

3.2. Distribution of Hot Spring Prokaryotes in TVG Area

3.2.1. Polymerase Chain Reactions on 16 S rRNA Genes

Previous surveys performed in this area have reported the absence of bacteria or archaea or nanoarchaea in certain hydrothermal ponds (Cheng et al., 2013; Lu et al., unpublished data). On the other hand, the absence of another archaea *Ignicoccus hospitalis* in the libraries constructed in these researches also triggered the curiosity since the discovery of nanoarchaeum also unveiled the parasitic or symbiotic relationship with its host *I. hospitalis* (Huber et al., 2002; Huber et al., 2003; Waters et al., 2003; Paper et al., 2007; Jahn et al., 2008; Podar et al., 2008).

The detections of bacteria, archaea and nanoarchaea using specific primer sets (Supplemental Table 2) were first tested for optimizing PCR results (in terms of successfully generating clear and single band) and three sets of primer were then set as default in each detection: B27F / U1406R for bacteria; A109F / A958R for archaea and 8mcF / 934 mcR for nanoarchaea. Repeated tests revealed that archaeal 16 S rDNA sequences and nanoarchaea-like 16 S rDNA sequences existed in all sites surveyed in this study (LFG, ZSL, XYK, MC, SHP and HS). However, signals of bacterial SSU rDNA sequences were only detected in LFG, MC, SHP and the sample with lower temperature (37°C) in HS (Supplemental Figure 3).

3.2.2. Sequence Analysis on Nanoarchaeal 16 S rRNA Gene

To assess the phylogenetic position of nanoarchaea-like archaeotes discovered in ZSL, MC, SHP and HS in relation to known references (such as *Nanoarchaeum equitans* strain Kin4-M, uncultured nanoarchaeote clones CU-1 and OP-9), partial 16 S rDNA sequences (961 bp) were aligned and compared through 3 different approaches for tree construction: maximum likelihood, neighbor-joining and minimum evolution. All three displayed highly similar pattern (Supplemental Figure 4 and 5, Supplemental Table 3) and neighbor-joining tree was chosen to serve as representative figure (Supplemental Figure 5). In general, the tree was well represented by three major archaeal clades (phyla Nanoarcheota, Crenarcheota and Euryarcheota) and a bacteria representative *Escherichia coli*. 16 S rDNA signals amplified by nanoarchaeal primer set 8mcF / 934mcR in this area were conspicuously grouped together, except one sample from MC which was classified into Euryarchaeota clade with *Thermoplasma acidophilum* reference sequence. The closest known relative to the group belongs to nanoarchaea; however, the similarity between strain CU-1 and sequences found in this cluster is ranged from 91.7% to 94.7% (Supplemental Table 4), which clearly indicated the two are not of the same species. Within the sequence group discovered in TVG area, only sample derived from ZSL seemed to be more regional specific in comparison with samples from SHP and HS. In addition, the pairwise distances within nanoarchaeal-like sequences found in this area were also higher than 3% (Supplemental Table 4), which indicated the existence of a diverse nanoarchaea-like prokaryotes cluster.

4. Discussions

4.1. Distribution of Hot Spring Prokaryotes in TVG Area

Archaeal and nanoarchaea-like 16 S rDNA sequences were found to be present in every surveyed location in this study; however, bacterial 16 S rDNA was absent in ZSL, XYK and HS (the hot pond with higher temperature). Previous researches have revealed a distinguishable separation which was reflected by

the restricted distribution of bacterial 16 S rDNA sequence only at the eastern part of TVG area (Cheng et al., 2013; Lu et al., unpublished data). Although bacterial signals were not detected in samples of HS with temperatures over 80°C collected by previous studies and this study, the existence of geographic barrier could not be confirmed since bacteria was found in hot springs in SHP and MC, which are located in between XYK and HS. The absence of bacterial signals in XYK and HS in this study also provided a different fact from that of the research done by 李重義 and 張怡塘 (2009), in which bacteria were found to be more tolerant to high temperatures (could be as high as 95°C) than archaea. Based on what was observed in this study, high temperature was not the factor that limits the distribution of bacteria, if the absence of bacterial SSU rDNA sequences in ZSL was not a result of unknown error.

Separation of archaea was also reported previously (Lu et al., unpublished data), that archaeal 16 S rDNA sequences were absent at the most eastern edge of sampling area and nanoarchaea-like sequences were absent at both east and west distal regions LHG (硫磺谷), HG (磺港) and FY (富源). Even though more hot springs in central TVG were sampled in this study than previous one, but the three locations mentioned above were skipped, only LFG and HS were sampled by both studies. From current results, it can be only concluded that there was no separation among archaea and nanoarchaea-like prokaryotes in the region being surveyed in this study.

Conclusively, distribution of bacteria, archaea and nanoarchaea-like 16 S rDNA sequences was found to be wide-spread in this study. The pattern suggested by earlier research that bacteria were more likely to occur at the eastern part of TVG was not supported by our results. Varied conclusions obtained from different surveys also indicated the dynamic prokaryotic community structures in TVG hydrothermal environments could be possibly higher than expected.

4.1.1. Biomass Estimation of Hot Spring Prokaryotic Community

The average DNA amounts recovered from hot spring water were 3.73 and 11.8

ng/ml in SHP and ZSL samples, respectively. Since the sampling conditions in SHP (high temperature, low pH) were set to remove the possibility of the existence of eukaryotic organisms as much as possible, we could assumed that the vast majority of extracted DNAs were of prokaryotic origin (this was also proven later by the result of sequencing). For 978 Mbp of double-stranded DNA is around 1 pg, supposed that the average prokaryotic genome size is 5 Mb, then the deduced cell density of the hot spring in SHP was equivalent to 7.3×10^5 cells/ml, which was similar to that of other terrestrial hydrothermal systems (Burton and Norris, 2000) but higher than that of deep marine environment (Huber et al., 2007). Compared with the unit 10^6 cells/ml for estimating cell density in fresh water and saline lakes (Whitman et al., 1998), apparently geothermal spring system possesses less prokaryotic cell density.

4.1.2. Nanoarchaea-like 16 S rDNA Sequences in TVG Area

Nanoarchaea-like 16 S rDNA sequences detected in TVG area did not belong to the same species with known reference *Nanoarchaeum equitans* Kin4-M. The closest relatives were uncultured nanoarchaeote clones CU-1 (from the Uzon Caldera, Kamchatka, Russia) and OP-9 (from Obsidian Pool, Yellowstone National Park, USA) (Hohn et al., 2002). The sample from MC was obviously grouped with Euryarchaeota clade, which was thought to be caused by unknown artifact. Though almost all nanoarchaea-like SSU sequences found in this area were grouped together, in some cases the pairwise differences were still above species level (Supplemental Table 4).

5. Summary

In the six sites during the time interval of sampling, there was no significant effect of geographic isolation found among the distribution pattern of Domain Bacteria, Archaea and phylum Nanoarchaeota by PCR confirmation, but archaea and nanoarchaea-like sequences were cosmopolitan in six sampling sites and bacterial ones were slightly less prevalent in hot springs of this region.

References

Burggraf, S., Stetter, K. O., Rouviere, P. and Woese, C. R. (1991). "Methanopyrus kandleri: An Archaeal Methanogen Unrelated to all Other Known Methanogens." Systematic and Applied Microbiology **14**(4): 346-351.

Burton, N. P. and Norris, P. R. (2000). "Microbiology of acidic, geothermal springs of Montserrat: environmental rDNA analysis." Extremophiles **4**(5): 315-320.

Cheng, T. W., Wang, P. L., Chen, K. C., Wu, J. J., Song, S. R. and Lin, L. H. (2013). "Segregated planktonic and bottom-dwelling archaeal communities in high-temperature acidic, sulfuric ponds of the Tatung Volcano Group, northern Taiwan." Terrestrial, Atmospheric and Oceanic Sciences **(in press)**.

Delong, E. F. (1992). "Archaea in Coastal Marine Environments." Proceedings of the National Academy of Sciences of the United States of America **89**(12): 5685-5689.

Eder, W., Ludwig, W. and Huber, R. (1999). "Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of Kebrit Deep, Red Sea." Archives of Microbiology **172**(4): 213-218.

Hansen, M. C., Tolker-Nielsen, T., Givskov, M. and Molin, S. (1998). "Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region." Fems Microbiology Ecology **26**(2): 141-149.

Hohn, M. J., Hedlund, B. P. and Huber, H. (2002). "Detection of 16S rDNA sequences representing the novel phylum "Nanoarchaeota": Indication for a wide distribution in high temperature biotopes." Systematic and Applied Microbiology **25**(4): 551-554.

Hongoh, Y., Ohkuma, M. and Kudo, T. (2003). "Molecular analysis of bacterial microbiota in the gut of the termite Reticulitermes speratus (Isoptera; Rhinotermitidae)." Fems Microbiology Ecology **44**(2): 231-242.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C. and Stetter, K. O. (2002). "A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont." Nature **417**(6884): 63-67.

Huber, H., Hohn, M. J., Stetter, K. O. and Rachel, R. (2003). "The phylum Nanoarchaeota: Present knowledge and future perspectives of a unique form of life." Research in Microbiology **154**(3): 165-171.

Huber, J. A., Mark Welch, D., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D.

A. and Sogin, M. L. (2007). "Microbial population structures in the deep marine biosphere." Science **318**(5847): 97-100.

Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K. O., Rachel, R. and Huber, H. (2008). "Nanoarchaeum equitans and Ignicoccus hospitalis: New insights into a unique, intimate association of two archaea." Journal of Bacteriology **190**(5): 1743-1750.

Lane, D. J. (1991). 16S/23S rRNA sequencing. Chichester, United Kingdom, John Wiley & Sons.

Loy, A., Maixner, F., Wagner, M. and Horn, M. (2007). "probeBase - an online resource for rRNA-targeted oligonucleotide probes: new features 2007." Nucleic Acids Research **35**: D800-D804.

Lu, H. P., Lin, Y. K., Lin, L. H., Wang, P. L. and Yu, H. T. (unpublished data). The microbial community structure in acidic hot springs.

Paper, W., Jahn, U., Hohn, M. J., Kronner, M., Nather, D. J., Burghardt, T., Rachel, R., Stetter, K. O. and Huber, H. (2007). "Ignicoccus hospitalis sp nov, the host of 'Nanoarchaeum equitans'." International Journal of Systematic and Evolutionary Microbiology **57**: 803-808.

Podar, M., Anderson, I., Makarova, K. S., Elkins, J. G., Ivanova, N., Wall, M. A., Lykidis, A., Mavromatis, K., Sun, H., Hudson, M. E., Chen, W., Deciu, C., Hutchison, D., Eads, J. R., Anderson, A., Fernandes, F., Szeto, E., Lapidus, A., Kyrpides, N. C., Saier, M. H., Richardson, P. M., Rachel, R., Huber, H., Eisen, J. A., Koonin, E. V., Keller, M. and Stetter, K. O. (2008). "A genomic analysis of the archaeal system Ignicoccus hospitalis-Nanoarchaeum equitans." Genome Biology **9**(11).

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). "MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods." Molecular Biology and Evolution **28**(10): 2731-2739.

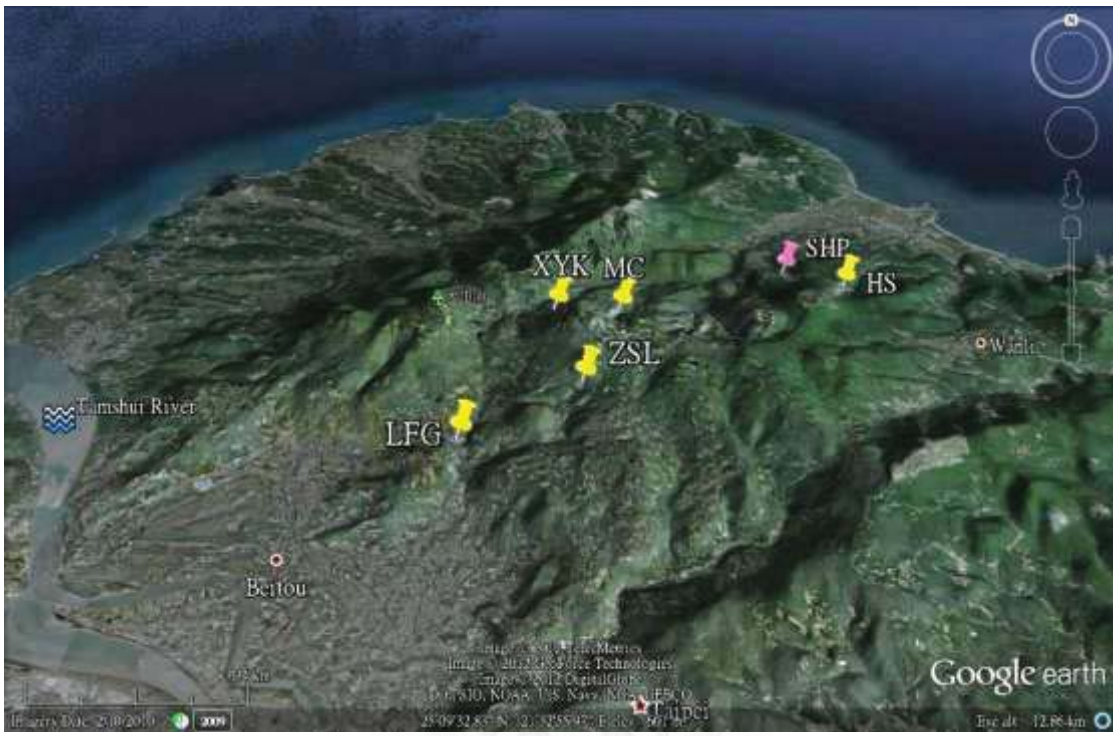
Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X. Y., Mathur, E., Ni, J. W., Podar, M., Richardson, T., Sutton, G. G., Simon, M., Soll, D., Stetter, K. O., Short, J. M. and Noordewier, M. (2003). "The genome of Nanoarchaeum equitans: Insights into early archaeal evolution and derived parasitism."

Proceedings of the National Academy of Sciences of the United States of America **100**(22): 12984-12988.

Whitehead, T. R. and Cotta, M. A. (1999). "Phylogenetic diversity of methanogenic archaea in swine waste storage pits." Fems Microbiology Letters **179**(2): 223-226.

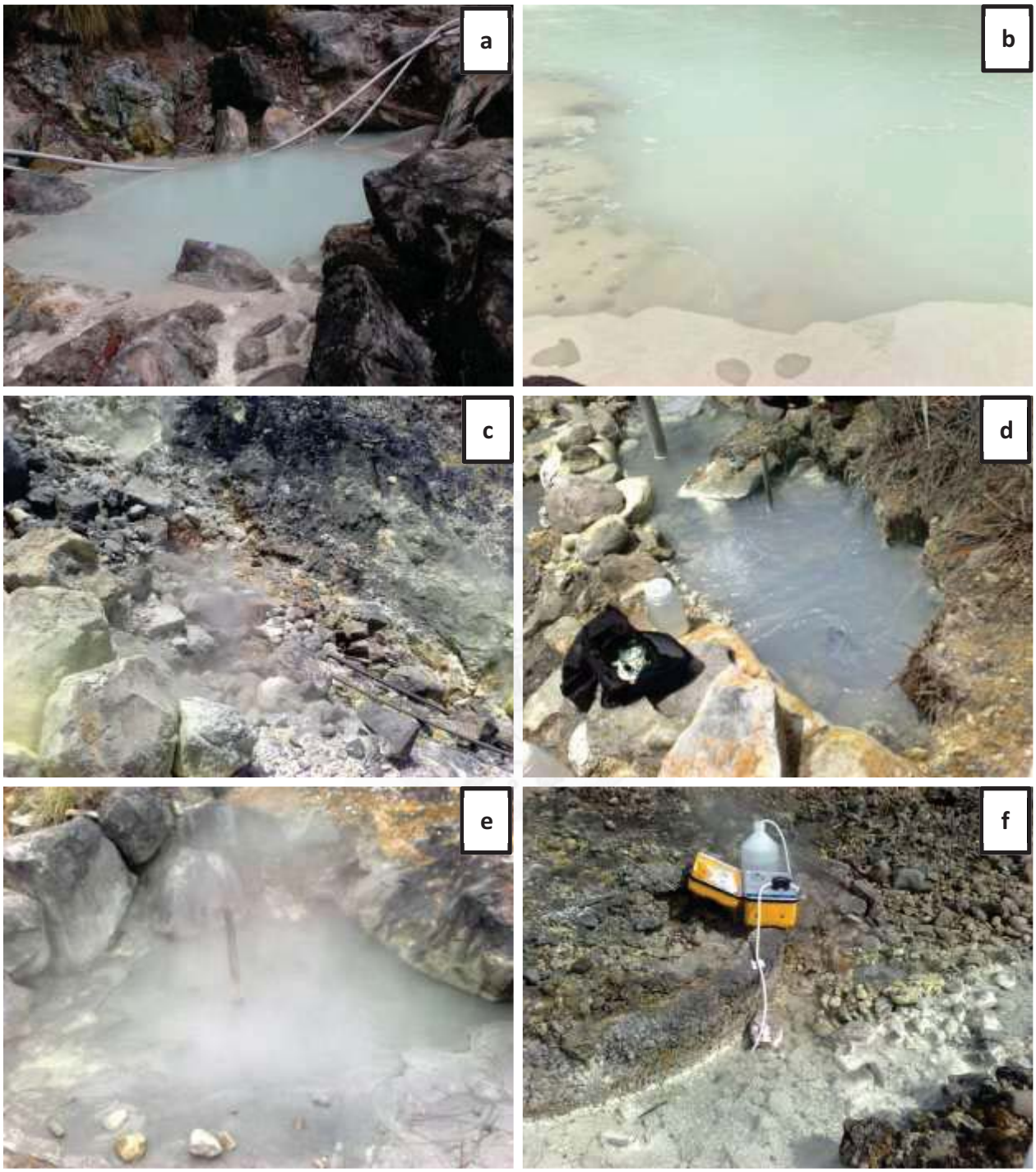
Whitman, W. B., Coleman, D. C. and Wiebe, W. J. (1998). "Prokaryotes: The unseen majority." Proceedings of the National Academy of Sciences of the United States of America **95**(12): 6578-6583.

李重義 and 張怡塘 (2009). 陽明山國家公園酸性溫泉物種調查及生存機制之研究. 陽明山國家公園管理處. 臺北市.

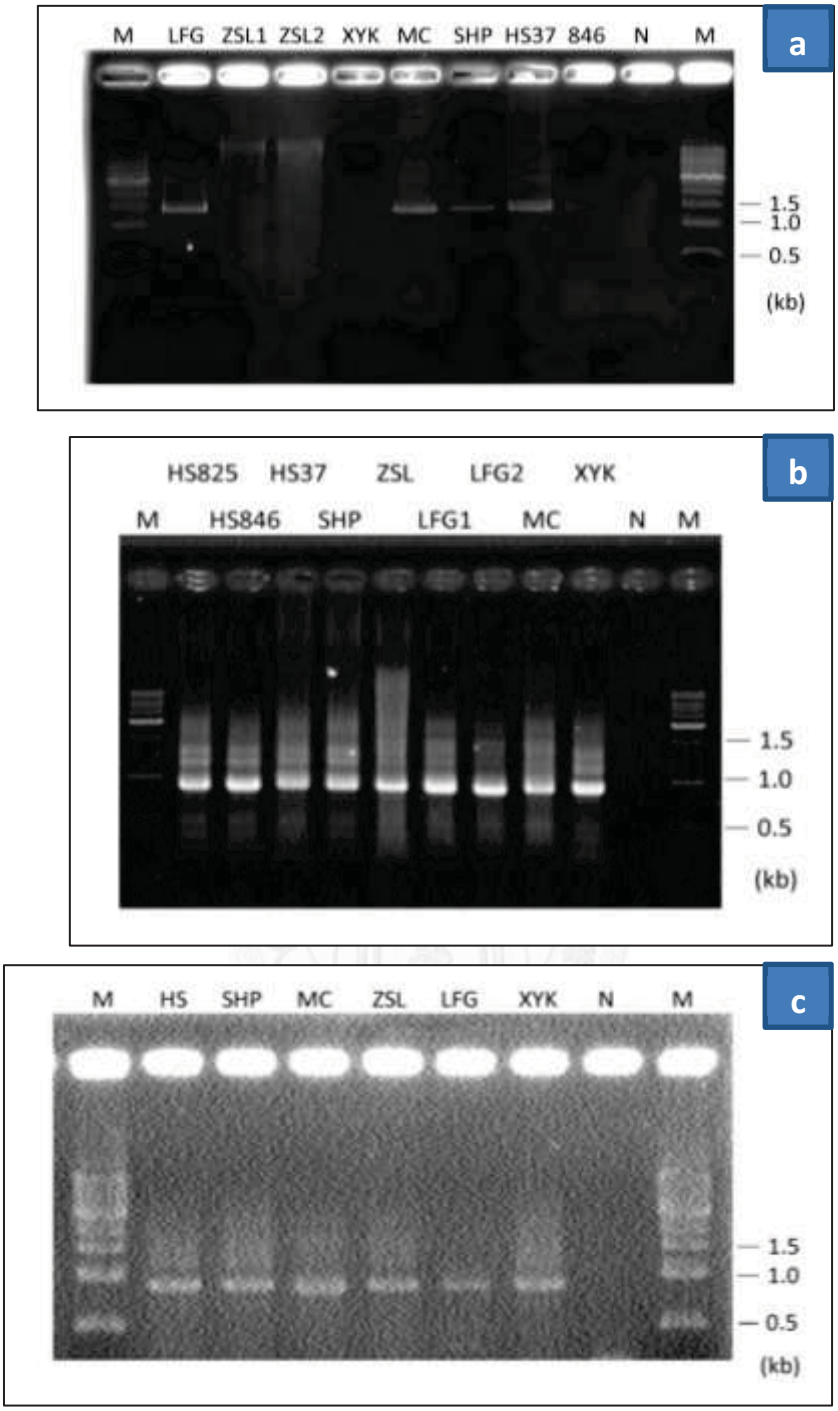


Supplemental Figure 1. Map of sample sites.

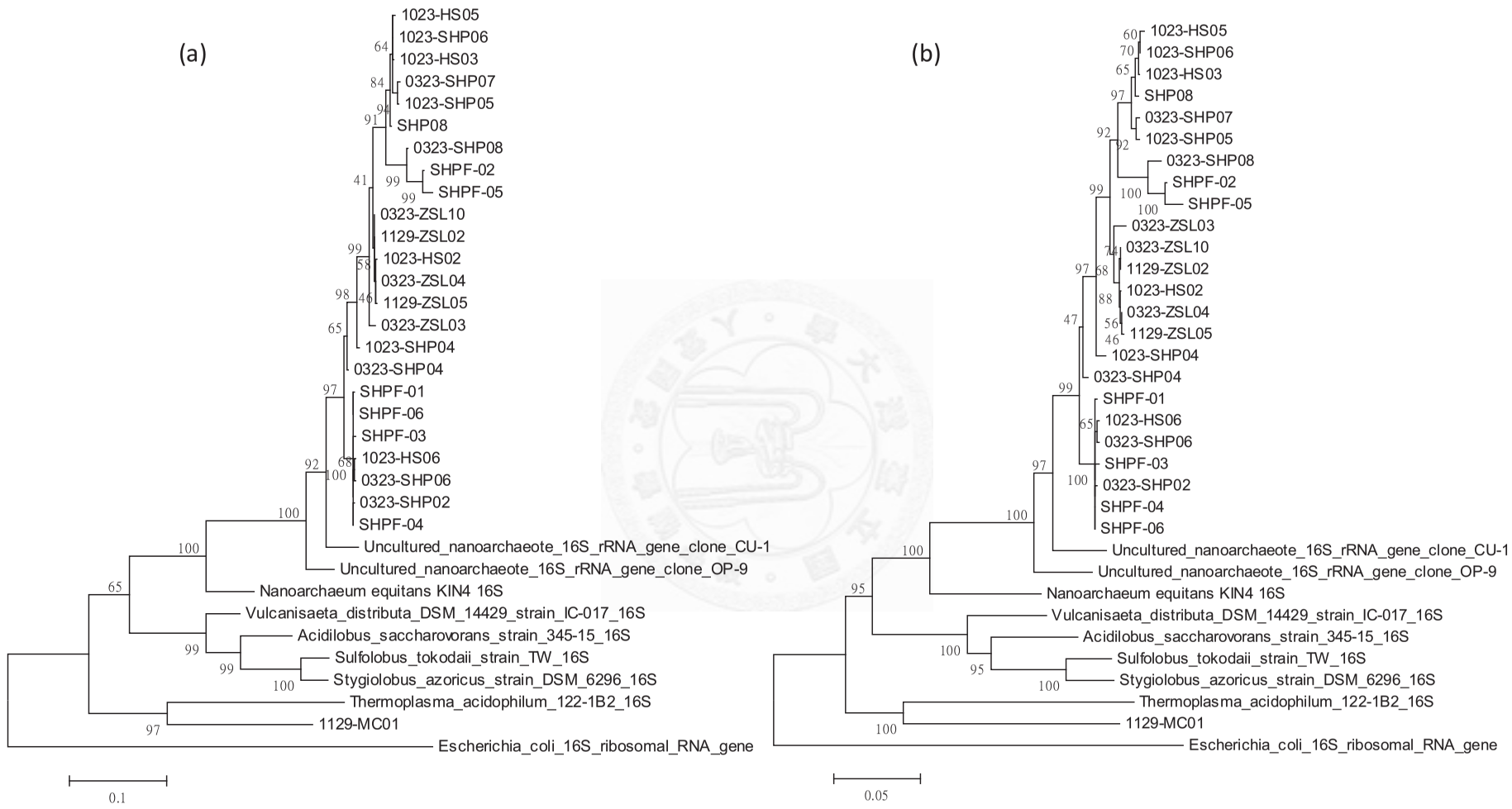
Pink pin represents the source of material for fosmid library construction and shotgun sequencing.



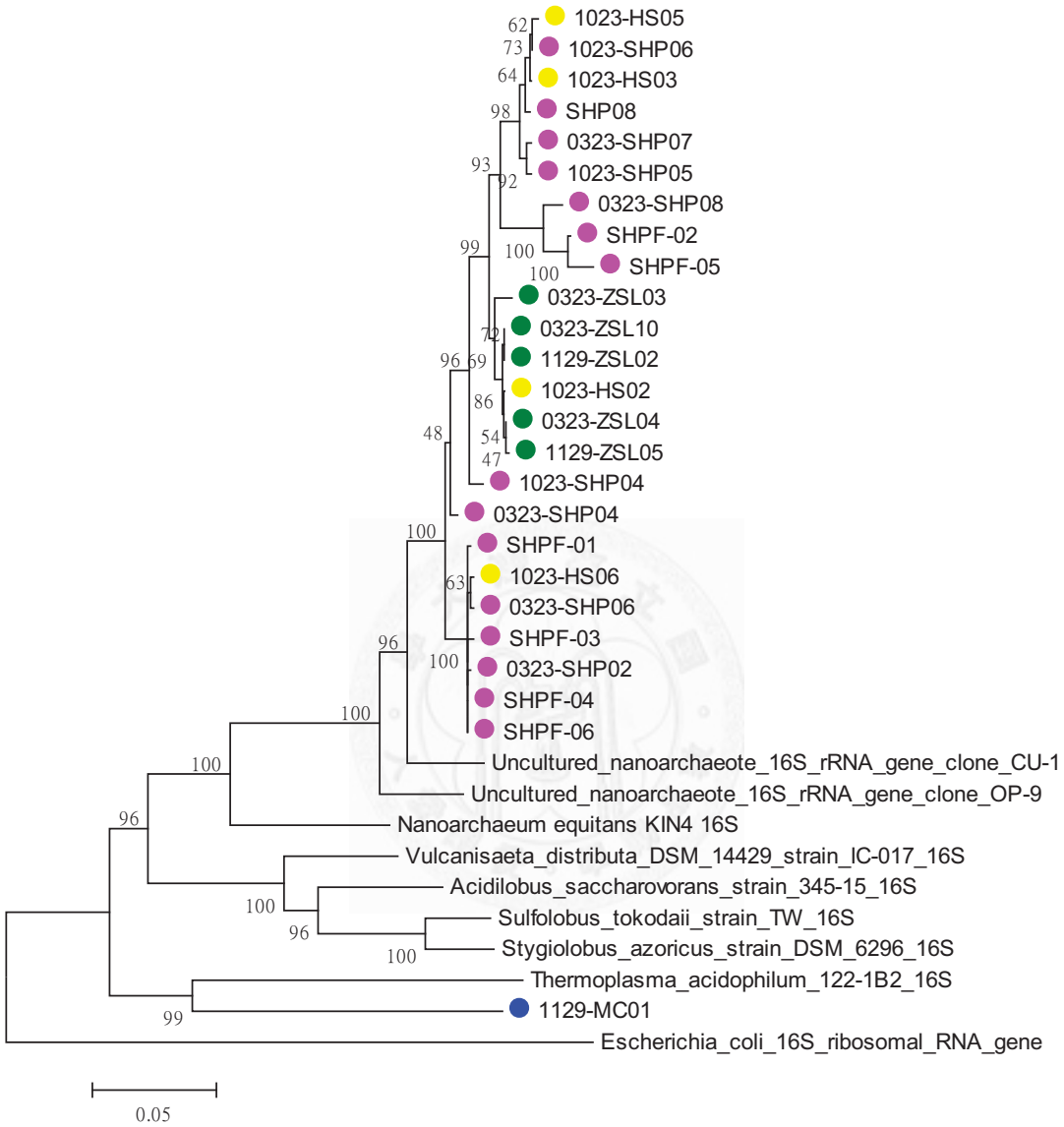
Supplemental Figure 2. Field views of sample sites in this study. (a: LFG, b: ZSL, c: XYK, d: MC, e: HS, f: SHP)



Supplemental Figure 3. PCR detection of bacterial, archaeal and nanoarchaeal SSU rDNA sequences on metagenomic DNAs from 6 sample sites. (a) Bacteria-specific primer: B27F / U1492R; (b) archaea-specific primers: A109F / A958R; (c) nanoarchaea-specific primers: 8mcF / 934mcR. M: marker; N: negative control; ZSL1 and 2 were sampled on different date; LFG1 and LFG2 were sampled in the same day but in different containers; HS37, HS846 (846), HS825 all came from HS and the number represent temperatures of water samples: 37 is 37°C, 846 is 84.6°C and 825 is 82.5°C.



Supplemental Figure 4. (a) ML tree of nanoarchaeal SSU sequences, (b) Minimum evolution tree of nanoarchaeal SSU sequences



Supplemental Figure 5. 16s rDNA phylogenetic tree of Nanoarchaea-like sequences in northern Taiwan volcanic hot spring area. ● : Samples from SHP; ● : samples from HS; ● : samples from MC and ● : samples from ZSL.

Supplemental Table 1. General information and physicochemical characteristics of hot spring water samples collected

Sample Site (Abbreviation)	Location	Sample Date (yyyy/mm/dd)	Temperature (°C)	pH	Sample volume (L)
龍鳳谷 (LFG)	25° 8'37.27"N, 121°31'41.95"E	2010/05/12	NA	NA	4
中山樓 (ZSL)	25° 9'20.94"N, 121°33'10.76"E	2010/07/07	52.5	NA	4
		2011/01/28	37.0	5.0	36 (sampled under heavy rain)
小油坑 (XYK)	25°10'30.59"N, 121°32'51.35"E	2010/07/07	90.9	NA	4
馬 槽 (MC)	25°10'38.82"N, 121°33'42.60"E	2010/07/07	44.2	NA	4
		2010/09/10	56.7	NA	6
磺 山 (HS)	25°11'18.35"N, 121°36'53.91"E	2010/07/02	82.5	NA	4
		2010/07/02	84.6	NA	4
四磺坪 (SHP)	25°11'43.60"N, 121°36'8.82"E	2010/05/26	70.0 / 37.0	NA	4
		2010/09/10	77.6	NA	8
		2010/12/03	69.0	2.8	18
		2010/12/10	78.8	2.8	36
		2010/12/21	71.7	2.9	36
		2011/02/08	64.0 / 37.0	3.5	36
		2011/03/17	82.0 / 46.0	3.2	40
		2012/04/25	69.0	2.5	40
		2012/10/04	74.6	2.4	*

*: No water was collected for DNA extraction, only parameters were measured.

Supplemental Table 2. Primer sets used in this study

Bacteria-Specific Primers			
Forward **Reverse**	Sequence	Product length (bp)	Reference
B27F **U1406R**	AGA GTT TGA TCM TGG CTC AG GAC GGG CGG TGT GTR CA	1465	Hongoh et al. (2003) Hansen et al. (1998)
B27F **U1492R**	AGA GTT TGA TCM TGG CTC AG GGY TAC CTT GTT ACG ACT T	1379	Hongoh et al. (2003) Lane (1991)
Archaea-Specific Primers			
Forward **Reverse**	Sequence	Product length (bp)	Reference
A109F **A958R**	ACK GCT CAG TAA CAC GT YCC GGC GTT GAM TCC AAT T	849	Whitehead and Cotta (1999) Delong (1992)
23F **U1492R**	TGC AGA YCT GGT YGA TYC TGC C GGC TAC CTT GTT ACG ACT T	1469	Burggraf et al. (1991) Lane (1991)
A21F **A958R**	TTC CGG TTG ATC CTG CCG GA YCC GGC GTT GAM TCC AAT T	937	Delong (1992)
Nanoarchaea-Specific Primers			
Forward **Reverse**	Sequence	Product length (bp)	Reference
8mcF **934mcR**	TCC CGT TGA TCC TGC G GTG CTC CCC CGC CTA TTC CT	926	Huber et al. (2002)
8mcF **1513mcR**	TCC CGT TGA TCC TGC G ACG GCT ACC TTG TGT CGA CTT	1507	Huber et al. (2002)
9bF **515mcR**	CCC GTT GAT CCT GCC GGA G CCC CTC TTG CCC ACC GCT	506	Eder et al. (1999) Huber et al. (2002)
519mcF **1114mcR**	CAG CCG CCG CGG GAA CAC GGG TCT CGC CTG TTT CC	595	Huber et al. (2002)

Supplemental Table 3. Parameters of nanoarchaeal-16s rRNA sequences phylogenetic tree reconstruction using MEGA5

Parameter	Maximum likelyhood	Neighbor joining	Minimum evolution
Test of phylogeny	Bootstrap method	Bootstrap method	Bootstrap method
No. of Bootstrap replication	500	500	500
Substitutions type	nucleotide	nucleotide	nucleotide
Substitution model	Tamura-Nei model	p-distance	p-distance
Rates among sites	uniform rates	uniform rates	uniform rates
Substitution to include	-	d: transitions + transversions	d: transitions + transversions
Gaps/missing data treatment	complete deletion	complete deletion	complete deletion
Select codon position	1 + 2 + 3 + noncoding sites	1 + 2 + 3 + noncoding sites	1 + 2 + 3 + noncoding sites
ML heuristic method	nearest-neighbor-interchange	-	-
Initial tree for ML	make initial tree automatically	-	-
ME heuristic method	-	-	close-neighbor-interchange
Initial tree for ME	-	-	neighbor-joining
ME search level	-	-	1

0323-ZSL03		0.004	0.004	0.004	0.004	0.004	0.005	0.005	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.007	0.005	0.007	0.006	0.006	0.006	0.007	0.006	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
0323-ZSL04	0.010		0.001	0.001	0.001	0.001	0.005	0.005	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.006	0.005	0.007	0.006	0.006	0.006	0.007	0.006	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
0323-ZSL10	0.011	0.001		0.000	0.002	0.002	0.005	0.005	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.006	0.005	0.007	0.006	0.006	0.006	0.007	0.006	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
1129-ZSL02	0.011	0.001	0.000		0.002	0.002	0.005	0.005	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.006	0.005	0.007	0.006	0.006	0.006	0.007	0.006	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
1129-ZSL05	0.011	0.001	0.003	0.003		0.002	0.005	0.005	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.006	0.005	0.007	0.006	0.007	0.006	0.007	0.006	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
1023-HS02	0.011	0.001	0.003	0.003	0.003		0.005	0.005	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.006	0.005	0.006	0.006	0.006	0.006	0.007	0.006	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
1023-HS03	0.025	0.021	0.020	0.020	0.023	0.023		0.002	0.007	0.007	0.007	0.007	0.003	0.006	0.003	0.001	0.006	0.002	0.008	0.007	0.007	0.007	0.008	0.007	0.013	0.010	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
1023-HS05	0.026	0.025	0.024	0.024	0.026	0.026	0.004		0.007	0.007	0.007	0.007	0.003	0.006	0.003	0.002	0.006	0.003	0.007	0.007	0.007	0.007	0.008	0.007	0.013	0.009	0.010	0.015	0.015	0.015	0.016	0.016	0.017	0.015
1023-HS06	0.039	0.036	0.035	0.035	0.038	0.035	0.048	0.051		0.002	0.004	0.002	0.007	0.005	0.007	0.007	0.008	0.007	0.002	0.008	0.003	0.002	0.009	0.002	0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
0323-SHP02	0.038	0.035	0.034	0.034	0.036	0.034	0.046	0.050	0.004		0.004	0.002	0.007	0.005	0.007	0.007	0.008	0.007	0.002	0.008	0.002	0.001	0.009	0.001	0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
0323-SHP04	0.030	0.030	0.029	0.029	0.031	0.029	0.040	0.041	0.016	0.015		0.005	0.006	0.004	0.007	0.007	0.008	0.007	0.004	0.008	0.004	0.004	0.008	0.004	0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
0323-SHP06	0.039	0.036	0.035	0.035	0.038	0.035	0.048	0.051	0.003	0.004	0.016		0.007	0.006	0.007	0.007	0.008	0.007	0.002	0.008	0.003	0.002	0.009	0.002	0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
0323-SHP07	0.026	0.025	0.024	0.024	0.026	0.026	0.009	0.010	0.046	0.045	0.034	0.046		0.005	0.002	0.003	0.006	0.004	0.007	0.007	0.007	0.007	0.008	0.007	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
1023-SHP04	0.021	0.021	0.020	0.020	0.023	0.020	0.030	0.031	0.028	0.026	0.014	0.028	0.024		0.005	0.006	0.007	0.006	0.006	0.007	0.005	0.005	0.008	0.005	0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
1023-SHP05	0.025	0.024	0.023	0.023	0.025	0.025	0.008	0.009	0.048	0.046	0.035	0.048	0.004	0.025		0.003	0.006	0.003	0.008	0.007	0.007	0.007	0.008	0.007	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
1023-SHP06	0.024	0.023	0.021	0.021	0.024	0.024	0.001	0.003	0.049	0.048	0.039	0.049	0.008	0.029	0.006		0.006	0.002	0.008	0.007	0.007	0.007	0.008	0.007	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
0323-SHP08	0.043	0.038	0.036	0.036	0.039	0.039	0.030	0.031	0.061	0.058	0.050	0.061	0.036	0.048	0.035	0.029		0.005	0.009	0.005	0.008	0.008	0.005	0.008	0.013	0.010	0.011	0.015	0.015	0.015	0.016	0.016	0.017	0.015
SHP08	0.023	0.021	0.020	0.020	0.023	0.023	0.005	0.006	0.045	0.046	0.038	0.045	0.011	0.033	0.010	0.004	0.028		0.008	0.007	0.007	0.007	0.008	0.007	0.013	0.009	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
SHPF-01	0.038	0.035	0.034	0.034	0.036	0.034	0.046	0.050	0.004	0.003	0.015	0.004	0.045	0.026	0.046	0.048	0.060	0.046		0.008	0.002	0.001	0.009	0.001	0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
SHPF-02	0.041	0.036	0.035	0.035	0.038	0.038	0.044	0.048	0.060	0.056	0.051	0.060	0.050	0.049	0.049	0.045	0.019	0.044	0.059		0.008	0.008	0.003	0.008	0.014	0.010	0.010	0.015	0.016	0.015	0.016	0.016	0.017	0.015
SHPF-03	0.039	0.036	0.035	0.035	0.038	0.035	0.048	0.051	0.005	0.004	0.016	0.005	0.046	0.028	0.048	0.049	0.059	0.048	0.004	0.058		0.002	0.009	0.002	0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
SHPF-04	0.036	0.034	0.033	0.033	0.035	0.033	0.045	0.049	0.003	0.001	0.014	0.003	0.044	0.025	0.045	0.046	0.059	0.045	0.001	0.058	0.003		0.009	0.000	0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
SHPF-05	0.053	0.048	0.046	0.046	0.049	0.049	0.053	0.056	0.071	0.068	0.060	0.071	0.056	0.058	0.055	0.054	0.028	0.053	0.070	0.011	0.069	0.069		0.009	0.014	0.010	0.010	0.015	0.015	0.015	0.016	0.015	0.017	0.015
SHPF-06	0.036	0.034	0.033	0.033	0.035	0.033	0.045	0.049	0.003	0.001	0.014	0.003	0.044	0.025	0.045	0.046	0.059	0.045	0.001	0.058	0.003	0.000	0.069		0.013	0.008	0.009	0.015	0.016	0.015	0.016	0.016	0.017	0.015
Nanoarchaeum_equitans_KIN4_16S	0.178	0.174	0.173	0.173	0.174	0.173	0.179	0.180	0.165	0.165	0.158	0.164	0.173	0.163	0.177	0.178	0.188	0.177	0.163	0.197	0.167	0.164	0.199	0.164		0.013	0.013	0.013	0.016	0.016	0.014	0.014	0.017	0.014
Uncultured_nanoarchaeote_16S_rRNA_gene_clone_CU-1	0.070	0.070	0.069	0.069	0.069	0.069	0.081	0.083	0.055	0.054	0.054	0.055	0.075	0.060	0.076	0.080	0.093	0.079	0.054	0.091	0.055	0.053	0.095	0.053	0.163		0.009	0.015	0.016	0.016	0.015	0.016	0.017	0.015
Uncultured_nanoarchaeote_16S_rRNA_gene_clone_OP-9	0.079	0.079	0.080	0.080	0.080	0.080	0.091	0.095	0.074	0.073	0.066	0.074	0.088	0.071	0.089	0.093	0.107	0.093	0.073	0.100	0.074	0.071	0.104	0.071	0.155	0.075		0.014	0.016	0.016	0.015	0.015	0.017	0.015
Vulcanisaeta_distributa_DSM_14429_strain_IC-017_16S	0.246	0.239	0.241	0.241	0.239	0.241	0.253	0.254	0.234	0.234	0.229	0.233	0.247	0.238	0.251	0.254	0.259	0.252	0.232	0.259	0.236	0.233	0.263	0.233	0.173	0.236	0.221		0.015	0.016	0.011	0.011	0.016	0.010
Thermoplasma_acidophilum_122-1B2_16S	0.331	0.326	0.326	0.326	0.326	0.325	0.325	0.328	0.322	0.320	0.318	0.322	0.321	0.321	0.325	0.326	0.320	0.325	0.320	0.325	0.322	0.321	0.326	0.321	0.288	0.326	0.316	0.291		0.015	0.016	0.015	0.017	0.016
1129-MC01	0.316	0.313	0.312	0.312	0.315	0.312	0.321	0.321	0.308	0.310	0.307	0.308	0.316	0.312	0.318	0.320	0.321	0.317	0.307	0.320	0.310	0.308	0.326	0.308	0.281	0.323	0.321	0.279	0.258		0.016	0.016	0.017	0.016
Sulfolobus_tokodaii_strain_TW_16S	0.284	0.281	0.281	0.281	0.282	0.282	0.288	0.291	0.271	0.271	0.272	0.269	0.282	0.281	0.286	0.289	0.298	0.287	0.268	0.297	0.271	0.269	0.301	0.269	0.234	0.278	0.267	0.135	0.305	0.297		0.008	0.017	0.011
Stygiolobus_azoricus_strain_DSM_6296_16S	0.287	0.284	0.284	0.284	0.286	0.286	0.291	0.293	0.273	0.273	0.277	0.272	0.284	0.283	0.288	0.292	0.301	0.291	0.271	0.296	0.273	0.272	0.301	0.272	0.227	0.286	0.272	0.135	0.303	0.287	0.054		0.017	0.011
Escherichia_coli_16S_ribosomal_RNA_gene	0.439	0.436	0.436	0.436	0.436	0.437	0.437	0.439	0.425	0.427	0.431	0.426	0.434	0.432	0.435	0.439	0.441	0.439	0.425	0.445	0.427	0.426	0.449	0.426	0.421	0.434	0.420	0.402	0.441	0.440	0.429	0.415		0.017
Acidilobus_saccharovorans_strain_345-15_16S	0.261	0.258	0.259	0.259	0.257	0.259	0.272	0.273	0.251	0.251	0.251	0.249	0.266	0.256	0.269	0.273	0.287	0.272	0.248	0.282	0.252	0.249	0.286	0.249	0.193	0.248	0.253	0.104	0.297	0.299	0.122	0.119	0.411	

Supplemental Table 4 . Distance matrix of nanoarchaea-like 16s rRNA gene partial sequence neighbor-joining tree

Appendix 2. Measurements of environmental parameters in SHP (四磺坪)

版次	5.2
修訂日期	101.05.07
文件編號	RS-WIE-23-01-03

謙德檢驗股份有限公司溫泉水檢測報告

行政院環保署認可字號:環署環檢字第006號　　電話:(02)2643-0126
交通部觀光局認可字號:觀技字第0944001704號　　傳真:(02)2643-0099
地址:221 新北市汐止區大同路一段263號9樓

委託單位：國立台灣大學生命科學館動物學研究所于宏燦教授實驗室　　樣品基質：溫泉水
業　　別：＊＊＊　　採樣日期：101 年 10 月 04 日 14 時 25 分
採樣地點：新北市金山區重和里名流路1-7號(天籟會館)　　收樣日期：101 年 10 月 04 日
採樣單位：謙德檢驗股份有限公司　　報告日期：101 年 10 月 16 日
報告編號：EF101GF083　　聯 絡 人：李佩珊　　分機 908

項次	檢驗項目	單位	原樣名稱	檢驗方法	溫泉基準
			溫泉水		
			檢測值		
1	氫離子濃度指數(pH)	—	2.6	NIEA W424.52A	—
2	泉溫(Temperature)	℃	72.1	溫泉泉溫檢測方法-溫度計法	>30
3	總溶解固體量(TDS)	mg/L	707	溫泉水中總溶解固體檢測方法-原水過濾法	>500
4	碳酸氫根離子(HCO_3^-)	mg/L	ND(MDL=0.03)	溫泉水中碳酸氫根離子檢測方法-滴定法	>250
5	硫酸根離子(SO_4^{2-})	mg/L	378	溫泉水中硫酸根離子檢測方法-濁度法	>250
6	氯離子含其他鹵族離子(Cl^-)	mg/L	9.3	溫泉水中氯離子檢測方法-硝酸銀滴定法	>250
7	硫化物(S^{2-})	mg/L	0.50	溫泉水中硫化物及硫化氫檢測方法-分光光度計/甲烯藍法	>0.1
	硫化氫 (H2S)	mg/L	52.7		
8	過硫酸根離子($S_2O_8^{2-}$)	mg/L	0.12	溫泉水中過硫酸根離子檢測方法-定色時間定量法	>0.1

備 註 ：1.本報告共 2 頁，分離使用無效。
2.低於方法偵測極之測定以"ND"表示，並註明其方法偵極限值(MDL)。
3.本報告僅對該樣品負責，並不得隨意複製及作為宣傳廣告之用。

聲明書
(一)茲保證本報告內容完全依照行政院環境保護署及有關機關之標準方法及品保品管等相關規定，秉持公正，誠實進行採樣、檢測，絕無虛偽不實，如有違反，就政府機關所受損失願負連帶賠償責任之外，並接受主管機關依法令所為之行政處分及刑事處罰。
(二)吾人瞭解如自身受政府機關委任從事公務，亦屬於刑法上之公務員，並瞭解刑法上圖利罪，公務員登載不實偽造公文書及貪污治罪條例之相關規定，如有違反，亦為刑法及貪污治罪條例之適用對象，願受嚴厲之法律制裁。

公司名稱：謙德檢驗股份有限公司

負責人：周錦煌　　檢驗室主任：王宜梅

第一頁/共二頁

版次	5.2
修訂日期	101.05.07
文件編號	RS-WIE-23-01-03

謙德檢驗股份有限公司溫泉水檢測報告

行政院環保署認可字號:環署環檢字第006號　　電話:(02)2643-0126
交通部觀光局認可字號:觀技字第0944001704號　　傳真:(02)2643-0099
地址:221 新北市汐止區大同路一段263號9樓

委託單位：國立台灣大學生命科學館動物學研究所 于宏燦教授實驗室　　樣品基質：溫泉水
案　　別：***　　採樣日期：101 年 10 月 04 日 14 時 25 分
採樣地點：新北市金山區重和里石流路1-7號(天籟會館)　　收樣日期：101 年 10 月 04 日
採樣單位：謙德檢驗股份有限公司　　報告日期：101 年 10 月 16 日
報告編號：EF101GF083　　聯 絡 人：李佩珊　　分機 908

項次	檢驗項目	單位	原樣名稱 溫泉水 檢測值	檢驗方法	溫泉基準
9	導電度(EC)	μmho/cm	1260	NIEA W203.51B	—
10	總鐵 ($Fe^{+2}+Fe^{+3}$)	mg/L	111	溫泉水中總鐵離子檢測方法-火焰原子吸收光譜法	>10
11	鈣(Ca)	mg/L	1.32	APHA 3111B	
12	鈉(Na)	mg/L	12.0	NIEA W306.52A	
13	鎂(Mg)	mg/L	1.21	APHA 3111B	
14	鉀(K)	mg/L	2.17	NIEA W306.52A	
15	鋁(Al)	mg/L	19.0	NIEA W311.52C	
16	砷(As)	mg/L	0.0012	NIEA W434.53B	
	以下空白				

備　註：1.本報告共 2 頁，分離使用無效。
2.低於方法偵測極之測定以"ND"表示，並註明其方法偵極限值(MDL)。
3.本報告僅對該樣品負責，並不得隨意複製及作為宣傳廣告之用。

聲明書
(一)茲保證本報告內容完全依照行政院環境保護署及有關機關之標準方法及品保品管等相關規定，秉持公正，誠實進行採樣、檢測。絕無虛偽不實，如有違反，就政府機關所受損失願負連帶賠償責任之外，並接受主管機關依法令所為之行政處分及刑事處罰。
(二)吾人瞭解如自身受政府機關委任從事公務，亦屬於刑法上之公務員，並瞭解刑法上圖利罪、公務員登載不實偽造公文書及貪污治罪條例之相關規定，如有違反，亦為刑法及貪污治罪條例之適用對象，願受最嚴厲之法律制裁。

公司名稱：謙德檢驗股份有限公司

負責人：周錦煌　　檢驗室主任：王宜梅

第二頁/共二頁

Appendix 3. Document of public deposition of fosmid library

財團法人 食品工業發展研究所
生物資源保存及研究中心
新竹市食品路 331 號 http://www.bcrc.firdi.org.tw
Tel:03-5223191-6 ext.513 Fax:03-5224172

ISO 9001 2000 VERSION

公開寄存菌種證明

台端於 100 年 8 月 4 日交付本所公開寄存之下列菌種，業經通過本所初步複核與污染檢測，這些菌種目前已進入本所菌種保存與管理體系，並分別獲得如下表所列之菌種編號。

■ 個人寄存 □ 機構寄存

菌名與來源編號		BCRC 編號
Tien-Lai Hot Spring microorganisms Fosmid Library	MG-HSTL(#1~#9558)	g1067

本所竭誠感謝台端的協助與支持，此致

寄存人：于宏燦

身分證或營利事業統一編號：[redacted]

地址：國立台灣大學動物所，生科館847室 國立臺灣大學 動物學研究所

Tel：02-33662456 Fax：02-23638179

財團法人 食品工業發展研究所
生物資源保存及研究中心
中華民國 100 年 12 月 2 日

說明：1. 本所生資中心接受並給予 BCRC 編號之公開寄存菌種，可依下列二項，選擇其中之一方式獲得優惠。
(a)可於寄存菌種被接受後取用十管該菌種。
(b)以一種對一種方式交換生資中心(除細胞株外)之其他菌種。
2. 寄存菌種亦將公開刊登於生資中心之簡訊、目錄及網路資訊，以供各界參考應用。

受理編號：A20110802

第 1 頁/共 1 頁 T-PS-043(2/2)-F

Appendix 4. Mapping Fos raw reads to reference genomes in IMG database
(only GC contents of the top 20's were included in this table)

Genus Name	Kingdom	Weighted Scores	Relative Abundance	GC-content
Escherichia	Bacteria	6288.59	73.68%	51%
Thiomonas	Bacteria	359.36	4.21%	64%
Acidithiobacillus	Bacteria	125.40	1.47%	59%
Neisseria	Bacteria	110.94	1.30%	48%
Yersinia	Bacteria	94.86	1.11%	47%
Acidovorax	Bacteria	49.17	0.58%	69%
Brucella	Bacteria	46.56	0.55%	57%
Xanthomonas	Bacteria	44.27	0.52%	64%
Bordetella	Bacteria	43.98	0.52%	62%
Actinobacillus	Bacteria	39.66	0.46%	41%
Buchnera	Bacteria	33.97	0.40%	25%
Haemophilus	Bacteria	32.55	0.38%	38%
Ruminococcus	Bacteria	31.57	0.37%	44%
Citrobacter	Bacteria	31.24	0.37%	54%
Bacillus	Bacteria	30.70	0.36%	44%
Cupriavidus	Bacteria	29.71	0.35%	64%
Klebsiella	Bacteria	29.20	0.34%	57%
Enterobacter	Bacteria	27.85	0.33%	56%
Hydrogenobaculum	Bacteria	21.86	0.26%	35%
Polaromonas	Bacteria	21.79	0.26%	62%
Variovorax	Bacteria	21.56	0.25%	
Comamonas	Bacteria	20.94	0.25%	
Methylibium	Bacteria	19.47	0.23%	
Leptothrix	Bacteria	18.17	0.21%	
Alicycliphilus	Bacteria	17.03	0.20%	
Francisella	Bacteria	16.90	0.20%	
Shewanella	Bacteria	16.84	0.20%	
Staphylococcus	Bacteria	15.30	0.18%	
Lactobacillus	Bacteria	13.98	0.16%	
Achromobacter	Bacteria	12.28	0.14%	
Edwardsiella	Bacteria	12.01	0.14%	
Legionella	Bacteria	11.49	0.13%	
Thioalkalivibrio	Bacteria	11.46	0.13%	
Thauera	Bacteria	11.38	0.13%	
Thiobacillus	Bacteria	11.32	0.13%	

Delftia	Bacteria	11.06	0.13%
Coxiella	Bacteria	10.97	0.13%
Stenotrophomonas	Bacteria	10.51	0.12%
Pectobacterium	Bacteria	10.51	0.12%
Lautropia	Bacteria	10.41	0.12%
Cronobacter	Bacteria	10.30	0.12%
Providencia	Bacteria	10.19	0.12%
Azoarcus	Bacteria	10.04	0.12%
Verminephrobacter	Bacteria	9.89	0.12%
Rhodopseudomonas	Bacteria	9.83	0.12%
Thermoplasma	Archaea	9.64	0.11%
Dickeya	Bacteria	9.41	0.11%
Lutiella	Bacteria	9.38	0.11%
Candidatus	Archaea	9.16	0.11%
Laribacter	Bacteria	9.14	0.11%
Methylovorus	Bacteria	9.02	0.11%
Herbaspirillum	Bacteria	8.85	0.10%
Chromobacterium	Bacteria	8.49	0.10%
Oxalobacter	Bacteria	8.39	0.10%
Bifidobacterium	Bacteria	8.37	0.10%
Helicobacter	Bacteria	8.31	0.10%
Sideroxydans	Bacteria	8.13	0.10%
Rhizobium	Bacteria	7.62	0.09%
Rhodobacter	Bacteria	7.56	0.09%
Kingella	Bacteria	7.50	0.09%
Pseudoxanthomonas	Bacteria	7.19	0.08%
Sulfolobus	Archaea	7.02	0.08%
Polynucleobacter	Bacteria	6.98	0.08%
Rhodoferax	Bacteria	6.62	0.08%
Aromatoleum	Bacteria	6.58	0.08%
Methylobacterium	Bacteria	6.56	0.08%
Serratia	Bacteria	6.29	0.07%
Caldivirga	Archaea	6.27	0.07%
Allochromatium	Bacteria	6.23	0.07%
Xylella	Bacteria	6.17	0.07%
Marinobacter	Bacteria	6.11	0.07%
Geobacter	Bacteria	6.09	0.07%
Grimontia	Bacteria	5.98	0.07%
Acetobacter	Bacteria	5.96	0.07%

Nitrosomonas	Bacteria	5.54	0.06%
beta	Bacteria	5.52	0.06%
Photobacterium	Bacteria	5.42	0.06%
Methylococcus	Bacteria	5.37	0.06%
Aggregatibacter	Bacteria	5.13	0.06%
Marinomonas	Bacteria	5.09	0.06%
Methylophilales	Bacteria	5.01	0.06%
Desulfovibrio	Bacteria	5.00	0.06%
Gamma	Bacteria	4.99	0.06%
Methylotenera	Bacteria	4.94	0.06%
Nitrosococcus	Bacteria	4.78	0.06%
Eikenella	Bacteria	4.75	0.06%
Streptomyces	Bacteria	4.71	0.06%
Methylobacillus	Bacteria	4.66	0.05%
Dechloromonas	Bacteria	4.51	0.05%
Mycoplasma	Bacteria	4.43	0.05%
Caulobacter	Bacteria	4.42	0.05%
Alcanivorax	Bacteria	4.41	0.05%
Taylorella	Bacteria	4.40	0.05%
Alkalilimnicola	Bacteria	4.32	0.05%
Rickettsia	Bacteria	4.25	0.05%
Herminiimonas	Bacteria	4.15	0.05%
Proteus	Bacteria	4.14	0.05%
Enterobacteriaceae	Bacteria	4.10	0.05%
Marine	Bacteria	4.04	0.05%
Azotobacter	Bacteria	3.77	0.04%
Mannheimia	Bacteria	3.73	0.04%
Gallionella	Bacteria	3.68	0.04%
Halorhodospira	Bacteria	3.65	0.04%
Sutterella	Bacteria	3.63	0.04%
Janthinobacterium	Bacteria	3.61	0.04%
Anaeromyxobacter	Bacteria	3.58	0.04%
Methylophaga	Bacteria	3.57	0.04%
Cardiobacterium	Bacteria	3.40	0.04%
Gluconacetobacter	Bacteria	3.32	0.04%
Deinococcus	Bacteria	3.31	0.04%
Raphidiopsis	Bacteria	3.29	0.04%
Mesorhizobium	Bacteria	3.22	0.04%
Xenorhabdus	Bacteria	3.19	0.04%

Limnobacter	Bacteria	3.19	0.04%
Glaciecola	Bacteria	3.16	0.04%
Oceanicola	Bacteria	3.10	0.04%
Metallosphaera	Archaea	3.07	0.04%
Halothiobacillus	Bacteria	3.05	0.04%
Nitrobacter	Bacteria	3.03	0.04%
Idiomarina	Bacteria	3.03	0.04%
Chromohalobacter	Bacteria	2.94	0.03%
Nitrosospira	Bacteria	2.92	0.03%
Borrelia	Bacteria	2.91	0.03%
Halomonas	Bacteria	2.86	0.03%
Dichelobacter	Bacteria	2.80	0.03%
Micrococcus	Bacteria	2.76	0.03%
Erythrobacter	Bacteria	2.63	0.03%
Reinekea	Bacteria	2.63	0.03%
Bradyrhizobium	Bacteria	2.62	0.03%
Vibrionales	Bacteria	2.61	0.03%
Rhodobacterales	Bacteria	2.60	0.03%
Thermoanaerobacter	Bacteria	2.58	0.03%
Brevundimonas	Bacteria	2.57	0.03%
Pasteurella	Bacteria	2.56	0.03%
Wolbachia	Bacteria	2.55	0.03%
Carboxydibrachium	Bacteria	2.54	0.03%
Ochrobactrum	Bacteria	2.54	0.03%
Azospirillum	Bacteria	2.53	0.03%
Wigglesworthia	Bacteria	2.51	0.03%
Bartonella	Bacteria	2.49	0.03%
Agrobacterium	Bacteria	2.47	0.03%
Simonsiella	Bacteria	2.45	0.03%
Burkholderiales	Bacteria	2.37	0.03%
Endoriftia	Bacteria	2.37	0.03%
Bermanella	Bacteria	2.32	0.03%
Rahnella	Bacteria	2.32	0.03%
Caminibacter	Bacteria	2.28	0.03%
Acidiphilium	Bacteria	2.27	0.03%
Histophilus	Bacteria	2.26	0.03%
Magnetospirillum	Bacteria	2.24	0.03%
Actinomyces	Bacteria	2.24	0.03%
Corynebacterium	Bacteria	2.23	0.03%

Sphingomonas	Bacteria	2.16	0.03%
Pseudovibrio	Bacteria	2.07	0.02%
Psychrobacter	Bacteria	2.06	0.02%
Sphingobium	Bacteria	2.05	0.02%
Aurantimonas	Bacteria	2.02	0.02%
Rhodocista	Bacteria	2.01	0.02%
Moritella	Bacteria	2.00	0.02%
Nitrococcus	Bacteria	1.98	0.02%
Baumannia	Bacteria	1.93	0.02%
Photorhabdus	Bacteria	1.84	0.02%
Geobacillus	Bacteria	1.83	0.02%
Ensifer	Bacteria	1.79	0.02%
Kangiella	Bacteria	1.79	0.02%
Selenomonas	Bacteria	1.77	0.02%
Starkeya	Bacteria	1.74	0.02%
Roseovarius	Bacteria	1.74	0.02%
Moraxella	Bacteria	1.74	0.02%
Paenibacillus	Bacteria	1.72	0.02%
gamma	Bacteria	1.69	0.02%
Psychromonas	Bacteria	1.67	0.02%
Listeria	Bacteria	1.67	0.02%
Ruegeria	Bacteria	1.67	0.02%
Phaeobacter	Bacteria	1.66	0.02%
Xanthobacter	Bacteria	1.64	0.02%
Ferrimonas	Bacteria	1.63	0.02%
Paracoccus	Bacteria	1.63	0.02%
Novosphingobium	Bacteria	1.60	0.02%
Thermaerobacter	Bacteria	1.59	0.02%
Rickettsiella	Bacteria	1.56	0.02%
Caldicellulosiruptor	Bacteria	1.55	0.02%
Congregibacter	Bacteria	1.54	0.02%
Pelagibaca	Bacteria	1.52	0.02%
Pelobacter	Bacteria	1.51	0.02%
Tolumonas	Bacteria	1.48	0.02%
Azorhizobium	Bacteria	1.46	0.02%
Phenylobacterium	Bacteria	1.46	0.02%
Rhodospirillum	Bacteria	1.44	0.02%
Citreicella	Bacteria	1.44	0.02%
Afipia	Bacteria	1.43	0.02%

Silicibacter	Bacteria	1.43	0.02%
Citromicrobium	Bacteria	1.43	0.02%
Roseibium	Bacteria	1.41	0.02%
Mariprofundus	Bacteria	1.37	0.02%
Zymomonas	Bacteria	1.37	0.02%
Labrenzia	Bacteria	1.35	0.02%
Alteromonas	Bacteria	1.33	0.02%
Methylobacter	Bacteria	1.31	0.02%
Sphingopyxis	Bacteria	1.31	0.02%
Faecalibacterium	Bacteria	1.29	0.02%
Basfia	Bacteria	1.28	0.01%
Collinsella	Bacteria	1.26	0.01%
Cellvibrio	Bacteria	1.24	0.01%
Asticcacaulis	Bacteria	1.21	0.01%
Hoeflea	Bacteria	1.21	0.01%
Methylocystis	Bacteria	1.20	0.01%
Oceanithermus	Bacteria	1.16	0.01%
Atopobium	Bacteria	1.16	0.01%
Frankia	Bacteria	1.16	0.01%
Aliivibrio	Bacteria	1.16	0.01%
Oceanicaulis	Bacteria	1.14	0.01%
Gemella	Bacteria	1.13	0.01%
Leuconostoc	Bacteria	1.10	0.01%
Eubacterium	Bacteria	1.10	0.01%
Nisaea	Bacteria	1.08	0.01%
Rhodomicrobium	Bacteria	1.08	0.01%
Granulibacter	Bacteria	1.07	0.01%
Rhodococcus	Bacteria	1.06	0.01%
Thiomicrospira	Bacteria	1.06	0.01%
Dermacoccus	Bacteria	1.04	0.01%
Sulfurihydrogenibium	Bacteria	1.03	0.01%
Fulvimarina	Bacteria	1.03	0.01%
Desulfarculus	Bacteria	1.00	0.01%
Sagittula	Bacteria	1.00	0.01%
Enhydrobacter	Bacteria	1.00	0.01%
Anaplasma	Bacteria	1.00	0.01%
Sulfitobacter	Bacteria	0.99	0.01%
Chlorobium	Bacteria	0.98	0.01%
Eggerthella	Bacteria	0.96	0.01%

Dinoroseobacter	Bacteria	0.96	0.01%
Ureaplasma	Bacteria	0.93	0.01%
Arthrobacter	Bacteria	0.93	0.01%
Leptospira	Bacteria	0.92	0.01%
Methylosinus	Bacteria	0.91	0.01%
Neorickettsia	Bacteria	0.90	0.01%
Micromonospora	Bacteria	0.89	0.01%
Acholeplasma	Bacteria	0.89	0.01%
Chelativorans	Bacteria	0.88	0.01%
Teredinibacter	Bacteria	0.87	0.01%
Mobiluncus	Bacteria	0.87	0.01%
Maritimibacter	Bacteria	0.85	0.01%
Saccharophagus	Bacteria	0.85	0.01%
Maricaulis	Bacteria	0.85	0.01%
Stigmatella	Bacteria	0.84	0.01%
Slackia	Bacteria	0.83	0.01%
Hyphomicrobium	Bacteria	0.82	0.01%
Parvibaculum	Bacteria	0.80	0.01%
Hahella	Bacteria	0.80	0.01%
Capnocytophaga	Bacteria	0.78	0.01%
Thermus	Bacteria	0.77	0.01%
Saccharopolyspora	Bacteria	0.76	0.01%
Neptuniibacter	Bacteria	0.76	0.01%
Gluconobacter	Bacteria	0.75	0.01%
Aeromicrobium	Bacteria	0.73	0.01%
Succinatimonas	Bacteria	0.72	0.01%
Nitratiruptor	Bacteria	0.72	0.01%
Desulfurispirillum	Bacteria	0.72	0.01%
Thermotoga	Bacteria	0.72	0.01%
Pediococcus	Bacteria	0.72	0.01%
Meiothermus	Bacteria	0.71	0.01%
Desulfurivibrio	Bacteria	0.71	0.01%
Beggiatoa	Bacteria	0.71	0.01%
Desulfobulbus	Bacteria	0.70	0.01%
Methylocella	Bacteria	0.70	0.01%
Thermomonospora	Bacteria	0.69	0.01%
Cyanobium	Bacteria	0.69	0.01%
Hydrogenobacter	Bacteria	0.69	0.01%
Nodularia	Bacteria	0.69	0.01%

Pyramidobacter	Bacteria	0.67	0.01%
Nitratifractor	Bacteria	0.67	0.01%
Desulfurococcus	Archaea	0.66	0.01%
Arcobacter	Bacteria	0.65	0.01%
Alicyclobacillus	Bacteria	0.64	0.01%
Chlamydia	Bacteria	0.63	0.01%
Kytococcus	Bacteria	0.63	0.01%
Cyanothece	Bacteria	0.61	0.01%
Carnobacterium	Bacteria	0.60	0.01%
Loktanella	Bacteria	0.60	0.01%
Ketogulonicigenium	Bacteria	0.60	0.01%
Nocardia	Bacteria	0.59	0.01%
Dehalococcoides	Bacteria	0.59	0.01%
Acidimicrobium	Bacteria	0.58	0.01%
delta	Bacteria	0.57	0.01%
Campylobacterales	Bacteria	0.55	0.01%
Olsenella	Bacteria	0.55	0.01%
Sulfurimonas	Bacteria	0.54	0.01%
Granulicatella	Bacteria	0.54	0.01%
Thermosinus	Bacteria	0.54	0.01%
Orientia	Bacteria	0.53	0.01%
Gordonibacter	Bacteria	0.53	0.01%
Nautilia	Bacteria	0.52	0.01%
Ehrlichia	Bacteria	0.51	0.01%
Exiguobacterium	Bacteria	0.51	0.01%
Treponema	Bacteria	0.50	0.01%
Rhodothermus	Bacteria	0.50	0.01%
Cellulomonas	Bacteria	0.50	0.01%
Syntrophus	Bacteria	0.49	0.01%
Brachybacterium	Bacteria	0.48	0.01%
Myxococcus	Bacteria	0.48	0.01%
Fusobacterium	Bacteria	0.46	0.01%
Segniliparus	Bacteria	0.46	0.01%
Desulfuromonas	Bacteria	0.46	0.01%
Rothia	Bacteria	0.46	0.01%
Parvularcula	Bacteria	0.45	0.01%
Coprococcus	Bacteria	0.45	0.01%
Ahrensia	Bacteria	0.45	0.01%
Desulfomicrobium	Bacteria	0.45	0.01%

Anaerotruncus	Bacteria	0.45	0.01%
Catenibacterium	Bacteria	0.45	0.01%
Symbiobacterium	Bacteria	0.44	0.01%
Oceanibulbus	Bacteria	0.44	0.01%
Staphylothermus	Archaea	0.44	0.01%
Octadecabacter	Bacteria	0.44	0.01%
Nautella	Bacteria	0.44	0.01%
Acidobacterium	Bacteria	0.44	0.01%
Tropheryma	Bacteria	0.43	0.01%
Gloeobacter	Bacteria	0.43	0.01%
Granulicella	Bacteria	0.43	0.01%
Desulfotalea	Bacteria	0.43	0.00%
Leifsonia	Bacteria	0.42	0.00%
Magnetococcus	Bacteria	0.42	0.00%
Haliangium	Bacteria	0.42	0.00%
Bacteroides	Bacteria	0.42	0.00%
Veillonella	Bacteria	0.41	0.00%
Psychroflexus	Bacteria	0.41	0.00%
Xylanimonas	Bacteria	0.40	0.00%
Thermobifida	Bacteria	0.40	0.00%
Jannaschia	Bacteria	0.40	0.00%
Anaerococcus	Bacteria	0.40	0.00%
Kocuria	Bacteria	0.39	0.00%
Gardnerella	Bacteria	0.39	0.00%
Dietzia	Bacteria	0.39	0.00%
Hyphomonas	Bacteria	0.38	0.00%
Prevotella	Bacteria	0.37	0.00%
Anaerostipes	Bacteria	0.37	0.00%
Peptoniphilus	Bacteria	0.37	0.00%
Rickettsiales	Bacteria	0.36	0.00%
Beijerinckia	Bacteria	0.36	0.00%
Thermosphaera	Archaea	0.36	0.00%
Plesiocystis	Bacteria	0.35	0.00%
Ethanoligenens	Bacteria	0.35	0.00%
Oenococcus	Bacteria	0.35	0.00%
Thermobispora	Bacteria	0.35	0.00%
Thermanaerovibrio	Bacteria	0.35	0.00%
Actinosynnema	Bacteria	0.34	0.00%
Ignicoccus	Archaea	0.34	0.00%

Arthrospira	Bacteria	0.34	0.00%
Chlorobaculum	Bacteria	0.34	0.00%
Truepera	Bacteria	0.34	0.00%
Gemmata	Bacteria	0.34	0.00%
Nakamurella	Bacteria	0.34	0.00%
Nocardiopsis	Bacteria	0.33	0.00%
Desulfonatronospira	Bacteria	0.33	0.00%
Calditerrivibrio	Bacteria	0.33	0.00%
Geodermatophilus	Bacteria	0.33	0.00%
Thermocrinis	Bacteria	0.33	0.00%
Pyrobaculum	Archaea	0.32	0.00%
Conexibacter	Bacteria	0.32	0.00%
Cylindrospermopsis	Bacteria	0.32	0.00%
Alkaliphilus	Bacteria	0.31	0.00%
Amycolatopsis	Bacteria	0.31	0.00%
Kribbella	Bacteria	0.30	0.00%
Kineococcus	Bacteria	0.30	0.00%
Pseudoflavonifractor	Bacteria	0.30	0.00%
Halothermothrix	Bacteria	0.30	0.00%
Sulfurovum	Bacteria	0.30	0.00%
Pelodictyon	Bacteria	0.30	0.00%
Rubrobacter	Bacteria	0.30	0.00%
Pseudonocardia	Bacteria	0.30	0.00%
Colwellia	Bacteria	0.29	0.00%
Catenulispora	Bacteria	0.29	0.00%
Desulfohalobium	Bacteria	0.29	0.00%
Heliobacterium	Bacteria	0.29	0.00%
Acidilobus	Archaea	0.29	0.00%
Saccharomonospora	Bacteria	0.29	0.00%
Janibacter	Bacteria	0.29	0.00%
Spirochaeta	Bacteria	0.28	0.00%
Nocardioides	Bacteria	0.28	0.00%
Mitsuokella	Bacteria	0.28	0.00%
Sulfuricurvum	Bacteria	0.28	0.00%
Methanobrevibacter	Archaea	0.28	0.00%
Hyperthermus	Archaea	0.28	0.00%
Thermosynechococcus	Bacteria	0.28	0.00%
Thermodesulfovibrio	Bacteria	0.27	0.00%
Agreia	Bacteria	0.27	0.00%

Erysipelothrix	Bacteria	0.27	0.00%
Bilophila	Bacteria	0.27	0.00%
Intrasporangium	Bacteria	0.27	0.00%
Shuttleworthia	Bacteria	0.26	0.00%
Deferribacter	Bacteria	0.26	0.00%
Dialister	Bacteria	0.26	0.00%
Ruminococcaceae	Bacteria	0.26	0.00%
Victivallis	Bacteria	0.26	0.00%
Halanaerobium	Bacteria	0.26	0.00%
Jonquetella	Bacteria	0.26	0.00%
Syntrophobacter	Bacteria	0.26	0.00%
Desulfurobacterium	Bacteria	0.26	0.00%
Oribacterium	Bacteria	0.26	0.00%
Desulfococcus	Bacteria	0.26	0.00%
Streptosporangium	Bacteria	0.25	0.00%
Vulcanisaeta	Archaea	0.25	0.00%
Desulfatibacillum	Bacteria	0.25	0.00%
Sulfurospirillum	Bacteria	0.25	0.00%
Sphaerobacter	Bacteria	0.25	0.00%
Salinispora	Bacteria	0.25	0.00%
Stackebrandtia	Bacteria	0.25	0.00%
Weissella	Bacteria	0.24	0.00%
Ammonifex	Bacteria	0.24	0.00%
Dorea	Bacteria	0.24	0.00%
Korebacter	Bacteria	0.24	0.00%
Thermovibrio	Bacteria	0.23	0.00%
Sanguibacter	Bacteria	0.23	0.00%
Microcoleus	Bacteria	0.23	0.00%
Planococcus	Bacteria	0.23	0.00%
Ignisphaera	Archaea	0.22	0.00%
Candidate	Bacteria	0.22	0.00%
Wolinella	Bacteria	0.22	0.00%
Opitutaceae	Bacteria	0.22	0.00%
Epulopiscium	Bacteria	0.22	0.00%
Aminobacterium	Bacteria	0.22	0.00%
Megasphaera	Bacteria	0.22	0.00%
Dethiobacter	Bacteria	0.22	0.00%
Moorella	Bacteria	0.22	0.00%
Sorangium	Bacteria	0.22	0.00%

Cyanobacterium	Bacteria	0.21	0.00%
Lawsonia	Bacteria	0.21	0.00%
Subdoligranulum	Bacteria	0.21	0.00%
Brevibacterium	Bacteria	0.21	0.00%
Brachyspira	Bacteria	0.20	0.00%
Crocosphaera	Bacteria	0.20	0.00%
Terriglobus	Bacteria	0.20	0.00%
Bulleidia	Bacteria	0.19	0.00%
Alistipes	Bacteria	0.19	0.00%
Acidothermus	Bacteria	0.19	0.00%
Natranaerobius	Bacteria	0.19	0.00%
Gemmatimonas	Bacteria	0.19	0.00%
Hydrogenivirga	Bacteria	0.19	0.00%
Clostridiales	Bacteria	0.19	0.00%
Salinibacter	Bacteria	0.18	0.00%
Thermosediminibacter	Bacteria	0.18	0.00%
Thalassiobium	Bacteria	0.18	0.00%
Catonella	Bacteria	0.18	0.00%
Oscillochloris	Bacteria	0.18	0.00%
Nostoc	Bacteria	0.17	0.00%
Opitutus	Bacteria	0.17	0.00%
Halobacterium	Archaea	0.17	0.00%
Roseburia	Bacteria	0.17	0.00%
Mesoplasma	Bacteria	0.17	0.00%
Pelotomaculum	Bacteria	0.17	0.00%
Methanocaldococcus	Archaea	0.17	0.00%
Aquifex	Bacteria	0.17	0.00%
Persephonella	Bacteria	0.16	0.00%
Bacteriovorax	Bacteria	0.16	0.00%
Parascardovia	Bacteria	0.16	0.00%
Coprobacillus	Bacteria	0.16	0.00%
Anaerobaculum	Bacteria	0.16	0.00%
Macrococcus	Bacteria	0.16	0.00%
Syntrophomonas	Bacteria	0.15	0.00%
Thermococcus	Archaea	0.15	0.00%
Desulfobacterium	Bacteria	0.15	0.00%
Eremococcus	Bacteria	0.15	0.00%
Lyngbya	Bacteria	0.15	0.00%
Anoxybacillus	Bacteria	0.15	0.00%

Lysinibacillus	Bacteria	0.15	0.00%
Solibacter	Bacteria	0.15	0.00%
Cryptobacterium	Bacteria	0.14	0.00%
Peptostreptococcus	Bacteria	0.14	0.00%
Lachnospiraceae	Bacteria	0.14	0.00%
Beutenbergia	Bacteria	0.14	0.00%
Prosthecochloris	Bacteria	0.14	0.00%
Aerococcus	Bacteria	0.14	0.00%
Ilyobacter	Bacteria	0.14	0.00%
Holdemania	Bacteria	0.14	0.00%
Chthoniobacter	Bacteria	0.14	0.00%
Flavobacteria	Bacteria	0.14	0.00%
Marinitoga	Bacteria	0.14	0.00%
Planctomyces	Bacteria	0.14	0.00%
Waddlia	Bacteria	0.14	0.00%
Thermincola	Bacteria	0.13	0.00%
Dehalogenimonas	Bacteria	0.13	0.00%
Pseudoramibacter	Bacteria	0.13	0.00%
Dictyoglomus	Bacteria	0.13	0.00%
Thermofilum	Archaea	0.13	0.00%
Desulfotomaculum	Bacteria	0.13	0.00%
Solobacterium	Bacteria	0.13	0.00%
Aeropyrum	Archaea	0.13	0.00%
Turicibacter	Bacteria	0.13	0.00%
Anaerofustis	Bacteria	0.13	0.00%
Pirellula	Bacteria	0.12	0.00%
Parvimonas	Bacteria	0.12	0.00%
Butyrivibrio	Bacteria	0.12	0.00%
Roseiflexus	Bacteria	0.12	0.00%
Blautia	Bacteria	0.11	0.00%
Acidaminococcus	Bacteria	0.11	0.00%
Fervidobacterium	Bacteria	0.11	0.00%
Renibacterium	Bacteria	0.11	0.00%
Thermosipho	Bacteria	0.11	0.00%
Thermoproteus	Archaea	0.11	0.00%
Syntrophobotulus	Bacteria	0.11	0.00%
Parabacteroides	Bacteria	0.11	0.00%
Chloroflexus	Bacteria	0.10	0.00%
Sphingobacterium	Bacteria	0.10	0.00%

Arcanobacterium	Bacteria	0.10	0.00%
Finegoldia	Bacteria	0.10	0.00%
Phascolarctobacterium	Bacteria	0.10	0.00%
Halalkalicoccus	Archaea	0.10	0.00%
Syntrophothermus	Bacteria	0.10	0.00%
Oscillatoria	Bacteria	0.10	0.00%
Oceanobacillus	Bacteria	0.10	0.00%
Pyrococcus	Archaea	0.10	0.00%
Desulfitobacterium	Bacteria	0.09	0.00%
Carboxydothermus	Bacteria	0.09	0.00%
Denitrovibrio	Bacteria	0.09	0.00%
Anaerolinea	Bacteria	0.09	0.00%
Microcystis	Bacteria	0.09	0.00%
Elusimicrobium	Bacteria	0.08	0.00%
Porphyromonas	Bacteria	0.08	0.00%
Blastopirellula	Bacteria	0.08	0.00%
Acetivibrio	Bacteria	0.08	0.00%
Synergistetes	Bacteria	0.08	0.00%
Synechocystis	Bacteria	0.08	0.00%
Fibrobacter	Bacteria	0.08	0.00%
Megamonas	Bacteria	0.08	0.00%
Robiginitalea	Bacteria	0.08	0.00%
Lentisphaera	Bacteria	0.08	0.00%
Methanococcus	Archaea	0.07	0.00%
Brevibacillus	Bacteria	0.07	0.00%
Anabaena	Bacteria	0.07	0.00%
Akkermansia	Bacteria	0.07	0.00%
Abiotrophia	Bacteria	0.07	0.00%
Jonesia	Bacteria	0.07	0.00%
Erysipelotrichaceae	Bacteria	0.07	0.00%
Leptotrichia	Bacteria	0.07	0.00%
Nanoarchaeum	Archaea	0.06	0.00%
Petrotoga	Bacteria	0.06	0.00%
Acaryochloris	Bacteria	0.06	0.00%
Archaeoglobus	Archaea	0.06	0.00%
Verrucomicrobiales	Bacteria	0.06	0.00%
Verrucomicrobium	Bacteria	0.06	0.00%
Rhodopirellula	Bacteria	0.06	0.00%
Hirschia	Bacteria	0.06	0.00%

Trichodesmium	Bacteria	0.06	0.00%
Bryantella	Bacteria	0.06	0.00%
Methylacidiphilum	Bacteria	0.06	0.00%
Coraliomargarita	Bacteria	0.06	0.00%
Thermomicrobium	Bacteria	0.06	0.00%
Thermoanaerobacterium	Bacteria	0.06	0.00%
Cellulophaga	Bacteria	0.06	0.00%
Methanothermus	Archaea	0.05	0.00%
Isosphaera	Bacteria	0.05	0.00%
Polaribacter	Bacteria	0.05	0.00%
Chloroherpeton	Bacteria	0.05	0.00%
Methanopyrus	Archaea	0.05	0.00%
Riemerella	Bacteria	0.05	0.00%
Kordia	Bacteria	0.05	0.00%
Cytophaga	Bacteria	0.05	0.00%
Halorubrum	Archaea	0.04	0.00%
Haloferax	Archaea	0.04	0.00%
Dyadobacter	Bacteria	0.04	0.00%
Natronomonas	Archaea	0.04	0.00%
Blattabacterium	Bacteria	0.04	0.00%
Leeuwenhoekiella	Bacteria	0.04	0.00%
Weeksella	Bacteria	0.04	0.00%
Kosmotoga	Bacteria	0.04	0.00%
Halomicrobium	Archaea	0.04	0.00%
Acetohalobium	Bacteria	0.04	0.00%
Maribacter	Bacteria	0.03	0.00%
Coprothermobacter	Bacteria	0.03	0.00%
Ktedonobacter	Bacteria	0.03	0.00%
Methanocella	Archaea	0.03	0.00%
Methanohalophilus	Archaea	0.03	0.00%
Flavobacteriaceae	Bacteria	0.03	0.00%
Ulvibacter	Bacteria	0.03	0.00%
Pedobacter	Bacteria	0.03	0.00%
Haladaptatus	Archaea	0.03	0.00%
Flavobacteriales	Bacteria	0.03	0.00%
Haloterrigena	Archaea	0.03	0.00%
Chlamydophila	Bacteria	0.03	0.00%
Halogeometricum	Archaea	0.03	0.00%
Methanothermobacter	Archaea	0.03	0.00%

Croceibacter	Bacteria	0.03	0.00%
Pedosphaera	Bacteria	0.03	0.00%
Zunongwangia	Bacteria	0.03	0.00%
Parachlamydia	Bacteria	0.02	0.00%
Ferroglobus	Archaea	0.02	0.00%
Halorhabdus	Archaea	0.02	0.00%
Aciduliprofundum	Archaea	0.02	0.00%
Flavobacterium	Bacteria	0.02	0.00%
Gramella	Bacteria	0.02	0.00%
Methanosarcina	Archaea	0.02	0.00%
Picrophilus	Archaea	0.02	0.00%
Scardovia	Bacteria	0.02	0.00%
Methanoculleus	Archaea	0.02	0.00%
Nitrosopumilus	Archaea	0.02	0.00%
Microscilla	Bacteria	0.02	0.00%
Dokdonia	Bacteria	0.02	0.00%
Thermotogales	Bacteria	0.02	0.00%
Chryseobacterium	Bacteria	0.02	0.00%
Methanohalobium	Archaea	0.02	0.00%
Natrialba	Archaea	0.01	0.00%
Cenarchaeum	Archaea	0.01	0.00%
Methanosphaerula	Archaea	0.01	0.00%
Methanococcoides	Archaea	0.01	0.00%
Herpetosiphon	Bacteria	0.01	0.00%
Methanosphaera	Archaea	0.01	0.00%
Sebaldella	Bacteria	0.01	0.00%
Spirosoma	Bacteria	0.01	0.00%
Methanosaeta	Archaea	0.01	0.00%
Thermobaculum	Bacteria	0.01	0.00%
Marivirga	Bacteria	0.01	0.00%
Algoriphagus	Bacteria	0.01	0.00%
Streptobacillus	Bacteria	0.01	0.00%
Bacteroidetes	Bacteria	0.01	0.00%
Methanoplanus	Archaea	0.01	0.00%
Ferroplasma	Archaea	0.01	0.00%
Odoribacter	Bacteria	0.01	0.00%
Mucilaginibacter	Bacteria	0.01	0.00%
Methanocorpusculum	Archaea	0.01	0.00%
Paludibacter	Bacteria	0.00	0.00%

Methanospirillum	Archaea	0.00	0.00%
Leadbetterella	Bacteria	0.00	0.00%
Chitinophaga	Bacteria	0.00	0.00%
Haloquadratum	Archaea	0.00	0.00%

Appendix 5. Mapping DSS raw reads to reference genomes in IMG database

Genus	Domain	Weighted Scores	Phylum	Class	Order	Family
Hydrogenobaculum	Bacteria	2303.55	Aquificae	Aquificae	Aquificales	Aquificaceae
Sulfurihydrogenibium	Bacteria	117.60	Aquificae	Aquificae	Aquificales	Hydrogenothermaceae
Sulfolobus	Archaea	117.00	Crenarchaeota	Thermoprotei	Sulfolobales	Sulfolobaceae
Hydrogenobacter	Bacteria	77.86	Aquificae	Aquificae	Aquificales	Aquificaceae
Metallosphaera	Archaea	76.76	Crenarchaeota	Thermoprotei	Sulfolobales	Sulfolobaceae
Acidithiobacillus	Bacteria	56.84	Proteobacteria	Gammaproteobacteria	Acidithiobacillales	Acidithiobacillaceae
Bacillus	Bacteria	44.88	Firmicutes	Bacilli	Bacillales	Alicyclobacillaceae
Thermocrinis	Bacteria	39.95	Aquificae	Aquificae	Aquificales	Aquificaceae
Ureaplasma	Bacteria	34.97	Tenericutes	Mollicutes	Mycoplasmatales	Mycoplasmataceae
Caldivirga	Archaea	31.06	Crenarchaeota	Thermoprotei	Thermoproteales	Thermoproteaceae
Staphylococcus	Bacteria	26.36	Firmicutes	Bacilli	Bacillales	Staphylococcaceae
Escherichia	Bacteria	25.66	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Hydrogenivirga	Bacteria	24.42	Aquificae	Aquificae	Aquificales	Aquificaceae
Neisseria	Bacteria	23.44	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
Lactobacillus	Bacteria	23.28	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae
Vulcanisaeta	Archaea	21.96	Crenarchaeota	Thermoprotei	Thermoproteales	Thermoproteaceae
Aquifex	Bacteria	21.75	Aquificae	Aquificae	Aquificales	Aquificaceae
Thiomonas	Bacteria	21.70	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified
Brucella	Bacteria	21.69	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae
Persephonella	Bacteria	16.43	Aquificae	Aquificae	Aquificales	Hydrogenothermaceae
Mycoplasma	Bacteria	14.59	Tenericutes	Mollicutes	Mycoplasmatales	Mycoplasmataceae

Caldicellulosiruptor	Bacteria	12.86	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacterales Family III. Incertae Sedis
Francisella	Bacteria	12.11	Proteobacteria	Gammaproteobacteria	Thiotrichales	Francisellaceae
Pyrobaculum	Archaea	11.24	Crenarchaeota	Thermoprotei	Thermoproteales	Thermoproteaceae
Rickettsia	Bacteria	10.87	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae
Desulfurococcus	Archaea	10.74	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae
Desulfurobacterium	Bacteria	9.34	Aquificae	Aquificae	Aquificales	Desulfurobacteriaceae
Thermoanaerobacter	Bacteria	9.14	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacteraceae
Thermosinus	Bacteria	8.97	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae
Staphylothermus	Archaea	8.50	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae
Coxiella	Bacteria	8.23	Proteobacteria	Gammaproteobacteria	Legionellales	Coxiellaceae
Thermovibrio	Bacteria	8.21	Aquificae	Aquificae	Aquificales	Desulfurobacteriaceae
Fusobacterium	Bacteria	8.14	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae
Thermotoga	Bacteria	7.86	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae
Faecalibacterium	Bacteria	7.30	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
Pseudovibrio	Bacteria	7.26	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Selenomonas	Bacteria	7.15	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae
Peptoniphilus	Bacteria	7.11	Firmicutes	Clostridia	Clostridiales	Clostridiales Family XI. Incertae Sedis
Ruminococcus	Bacteria	7.10	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
Candidatus	Archaea	6.95	Thaumarchaeota	Nitrosopumilales	Nitrosopumilales	Nitrosopumilaceae
Coprococcus	Bacteria	6.74	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Thermosphaera	Archaea	6.65	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae
Chlamydia	Bacteria	6.51	Chlamydiae	Chlamydiia	Chlamydiales	Chlamydiaceae
Bordetella	Bacteria	6.22	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae
Xanthomonas	Bacteria	6.18	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae

Eubacterium	Bacteria	5.78	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae
Thermus	Bacteria	5.68	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae
Methanocaldococcus	Archaea	5.49	Euryarchaeota	Methanococci	Methanococcales	Methanocaldococcaceae
Ignicoccus	Archaea	5.36	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae
Geobacter	Bacteria	5.30	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	Geobacteraceae
Hyperthermus	Archaea	5.25	Crenarchaeota	Thermoprotei	Desulfurococcales	Pyrodictiaceae
Finegoldia	Bacteria	5.25	Firmicutes	Clostridia	Clostridiales	Clostridiales Family XI. Incertae Sedis
Shewanella	Bacteria	5.20	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae
Dorea	Bacteria	4.88	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Anaerococcus	Bacteria	4.64	Firmicutes	Clostridia	Clostridiales	Clostridiales Family XI. Incertae Sedis
Wolbachia	Bacteria	4.63	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bartonellaceae
Roseibium	Bacteria	4.56	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Thermoplasma	Archaea	4.56	Euryarchaeota	Thermoplasmata	Thermoplasmatales	Thermoplasmataceae
Acidilobus	Archaea	4.51	Crenarchaeota	Thermoprotei	Acidilobales	Acidilobaceae
Rhodopseudomonas	Bacteria	4.46	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae
Dictyoglomus	Bacteria	4.30	Dictyoglomi	Dictyoglomia	Dictyoglomales	Dictyoglomaceae
Legionella	Bacteria	4.28	Proteobacteria	Gammaproteobacteria	Legionellales	Legionellaceae
Yersinia	Bacteria	4.26	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Chlamydophila	Bacteria	4.21	Chlamydiae	Chlamydiia	Chlamydiales	Chlamydiaceae
Acidovorax	Bacteria	4.18	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
Rhizobium	Bacteria	4.09	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
Paenibacillus	Bacteria	4.03	Firmicutes	Bacilli	Bacillales	Paenibacillaceae
Thermococcus	Archaea	3.93	Euryarchaeota	Thermococci	Thermococcales	Thermococcaceae
Rhodobacter	Bacteria	3.92	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae

Deinococcus	Bacteria	3.83	Deinococcus-Thermus	Deinococci	Deinococcales	Deinococcaceae
Phaeobacter	Bacteria	3.79	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Acetobacter	Bacteria	3.76	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
Ignisphaera	Archaea	3.73	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae
Actinobacillus	Bacteria	3.67	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
Anaerostipes	Bacteria	3.47	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Labrenzia	Bacteria	3.37	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Acidimicrobium	Bacteria	3.36	Actinobacteria	Actinobacteria	Acidimicrobiales	Acidimicrobiaceae
Zymomonas	Bacteria	3.26	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
Desulfovibrio	Bacteria	3.24	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae
Epulopiscium	Bacteria	3.23	Firmicutes	Clostridia	Clostridiales	unclassified
Calditerrivibrio	Bacteria	3.17	Deferribacteres	Deferribacteres	Deferribacterales	Deferribacteraceae
Pedobacter	Bacteria	3.16	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae
Haemophilus	Bacteria	3.08	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
Cupriavidus	Bacteria	3.06	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
Aeropyrum	Archaea	2.98	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae
Thermoproteus	Archaea	2.98	Crenarchaeota	Thermoprotei	Thermoproteales	Thermoproteaceae
Rhodobacterales	Bacteria	2.94	Proteobacteria	Alphaproteobacteria	Rhodobacterales	-
Nanoarchaeum	Archaea	2.91	Nanoarchaeota	Unclassified	unclassified	unclassified
Veillonella	Bacteria	2.88	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae
Anaerotruncus	Bacteria	2.79	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
Coprothermobacter	Bacteria	2.78	Firmicutes	Clostridia	Thermoanaerobacterales	Thermodesulfobiaceae
Anaerofustis	Bacteria	2.66	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae
Photobacterium	Bacteria	2.62	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae

Nitrosococcus	Bacteria	2.62	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae
Borrelia	Bacteria	2.60	Spirochaetes	Spirochaetia	Spirochaetales	Spirochaetaceae
Brachyspira	Bacteria	2.60	Spirochaetes	Spirochaetia	Spirochaetales	Brachyspiraceae
Bartonella	Bacteria	2.56	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bartonellaceae
Roseovarius	Bacteria	2.54	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Jonquetella	Bacteria	2.53	Synergistetes	Synergistia	Synergistales	Synergistaceae
Nitrobacter	Bacteria	2.53	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae
Treponema	Bacteria	2.49	Spirochaetes	Spirochaetia	Spirochaetales	Spirochaetaceae
Buchnera	Bacteria	2.49	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Thioalkalivibrio	Bacteria	2.44	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae
Halanaerobium	Bacteria	2.39	Firmicutes	Clostridia	Halanaerobiales	Halanaerobiaceae
Deferribacter	Bacteria	2.33	Deferribacteres	Deferribacteres	Deferribacterales	Deferribacteraceae
Thermosipho	Bacteria	2.32	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae
Orientia	Bacteria	2.31	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae
Xylella	Bacteria	2.31	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
Thermofilum	Archaea	2.31	Crenarchaeota	Thermoprotei	Thermoproteales	Thermofilaceae
Oceanicola	Bacteria	2.29	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Methylobacterium	Bacteria	2.24	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae
Polaromonas	Bacteria	2.21	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
Anaeromyxobacter	Bacteria	2.20	Proteobacteria	Deltaproteobacteria	Myxococcales	Myxococcaceae
Leptospira	Bacteria	2.17	Spirochaetes	Spirochaetia	Spirochaetales	Leptospiraceae
Porphyromonas	Bacteria	2.15	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae
Marinomonas	Bacteria	2.10	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae
Variovorax	Bacteria	2.08	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae

Idiomarina	Bacteria	2.05	Proteobacteria	Gammaproteobacteria	Alteromonadales	Idiomarinaceae
Sphingomonas	Bacteria	2.04	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
Meiothermus	Bacteria	2.01	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae
Pyrococcus	Archaea	2.01	Euryarchaeota	Thermococci	Thermococcales	Thermococcaceae
Erythrobacter	Bacteria	2.00	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Erythrobacteraceae
Coprobacillus	Bacteria	1.98	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae
Dehalococcoides	Bacteria	1.95	Chloroflexi	Dehalococcoidetes	Dehalococcoidales	Dehalococcoidaceae
Sphingobium	Bacteria	1.95	Firmicutes	Bacilli	Bacillales	Bacillaceae
Thermodesulfovibrio	Bacteria	1.95	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae
Polynucleobacter	Bacteria	1.91	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
Silicibacter	Bacteria	1.89	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Comamonas	Bacteria	1.87	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
Methylibium	Bacteria	1.83	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified
Grimontia	Bacteria	1.82	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae
Aurantimonas	Bacteria	1.81	Proteobacteria	Alphaproteobacteria	Rhizobiales	Aurantimonadaceae
Achromobacter	Bacteria	1.77	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae
Vibrionales	Bacteria	1.74	Proteobacteria	Gammaproteobacteria	Vibrionales	-
Mesoplasma	Bacteria	1.73	Tenericutes	Mollicutes	Entomoplasmatales	Entomoplasmataceae
Aminobacterium	Bacteria	1.72	Synergistetes	Synergistia	Synergistales	Synergistaceae
Polaribacter	Bacteria	1.71	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Halorhodospira	Bacteria	1.68	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae
Marinitoga	Bacteria	1.66	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae
Gemella	Bacteria	1.65	Firmicutes	Bacilli	Bacillales	Bacillales Family XI. Incertae Sedis
Thermanaerovibrio	Bacteria	1.65	Synergistetes	Synergistia	Synergistales	Synergistaceae

Algoriphagus	Bacteria	1.64	Bacteroidetes	Cytophagia	Cytophagales	Cyclobacteriaceae
Helicobacter	Bacteria	1.63	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae
Methylophaga	Bacteria	1.62	Proteobacteria	Gammaproteobacteria	Thiotrichales	Piscirickettsiaceae
Leptothrix	Bacteria	1.61	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified
Caulobacter	Bacteria	1.60	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
Corynebacterium	Bacteria	1.59	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae
Sulfitobacter	Bacteria	1.59	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Ehrlichia	Bacteria	1.56	Proteobacteria	Alphaproteobacteria	Rickettsiales	Anaplasmataceae
Fulvimarina	Bacteria	1.55	Proteobacteria	Alphaproteobacteria	Rhizobiales	Aurantimonadaceae
Roseburia	Bacteria	1.53	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Carboxydibrachium	Bacteria	1.53	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacteraceae
Gluconacetobacter	Bacteria	1.53	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
Mesorhizobium	Bacteria	1.53	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae
Azoarcus	Bacteria	1.52	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae
Oceanithermus	Bacteria	1.51	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae
Oenococcus	Bacteria	1.50	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae
Kingella	Bacteria	1.49	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
Leptotrichia	Bacteria	1.49	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae
Methylophilales	Bacteria	1.49	Proteobacteria	Betaproteobacteria	Methylophilales	-
Agrobacterium	Bacteria	1.48	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
Dichelobacter	Bacteria	1.47	Proteobacteria	Gammaproteobacteria	Cardiobacteriales	Cardiobacteriaceae
Bacteroides	Bacteria	1.45	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae
Geobacillus	Bacteria	1.45	Firmicutes	Bacilli	Bacillales	Bacillaceae
Thauera	Bacteria	1.43	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae

Prevotella	Bacteria	1.43	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae
Glaciecola	Bacteria	1.43	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae
Alicycliphilus	Bacteria	1.43	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
Nitrosomonas	Bacteria	1.42	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae
Butyrivibrio	Bacteria	1.41	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Thermosynechococcus	Bacteria	1.41	Cyanobacteria	unclassified	Chroococcales	unclassified
Fervidobacterium	Bacteria	1.37	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae
Lutiella	Bacteria	1.37	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
Actinomyces	Bacteria	1.37	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae
Methanopyrus	Archaea	1.36	Euryarchaeota	Methanopyri	Methanopyrales	Methanopyraceae
Sphingobacterium	Bacteria	1.35	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae
Bradyrhizobium	Bacteria	1.34	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae
Dokdonia	Bacteria	1.34	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Leuconostoc	Bacteria	1.32	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae
Bifidobacterium	Bacteria	1.31	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae
Lautropia	Bacteria	1.31	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
Stenotrophomonas	Bacteria	1.30	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
Azospirillum	Bacteria	1.30	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
Laribacter	Bacteria	1.30	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
Rhodospirillum	Bacteria	1.29	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
Thiobacillus	Bacteria	1.27	Proteobacteria	Betaproteobacteria	Hydrogenophilales	Hydrogenophilaceae
Carboxydothermus	Bacteria	1.26	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacteraceae
Gallionella	Bacteria	1.25	Proteobacteria	Betaproteobacteria	Gallionellales	Gallionellaceae
Mobiluncus	Bacteria	1.24	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae

Thermaerobacter	Bacteria	1.23	Firmicutes	Clostridia	Clostridiales	Clostridiales Family XVII. Incertae Sedis
Acidiphilium	Bacteria	1.23	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
Megasphaera	Bacteria	1.22	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae
Ammonifex	Bacteria	1.21	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacteraceae
Ruegeria	Bacteria	1.21	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Sideroxydans	Bacteria	1.20	Proteobacteria	Betaproteobacteria	Gallionellales	Gallionellaceae
Collinsella	Bacteria	1.20	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae
Methanothermus	Archaea	1.20	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanothermaceae
Lachnospiraceae	Bacteria	1.14	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Arthrospira	Bacteria	1.14	Cyanobacteria	unclassified	Oscillatoriales	unclassified
Nitratiruptor	Bacteria	1.12	Proteobacteria	Epsilonproteobacteria	unclassified	unclassified
Archaeoglobus	Archaea	1.12	Euryarchaeota	Archaeoglobi	Archaeoglobales	Archaeoglobaceae
Gluconobacter	Bacteria	1.11	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
Chromobacterium	Bacteria	1.10	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
Citreicella	Bacteria	1.10	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Exiguobacterium	Bacteria	1.10	Firmicutes	Bacilli	Bacillales	Bacillales Family XII. Incertae Sedis
Hirschia	Bacteria	1.07	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Hyphomonadaceae
Micrococcus	Bacteria	1.07	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae
Pelagibaca	Bacteria	1.06	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Citromicrobium	Bacteria	1.06	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
Elusimicrobium	Bacteria	1.05	Elusimicrobia	Elusimicrobia	Elusimicrobiales	Elusimicrobiaceae
Delftia	Bacteria	1.05	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
Ochrobactrum	Bacteria	1.05	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae
Ethanoligenens	Bacteria	1.05	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae

Cardiobacterium	Bacteria	1.02	Proteobacteria	Gammaproteobacteria	Cardiobacteriales	Cardiobacteriaceae
Anaerobaculum	Bacteria	1.00	Synergistetes	Synergistia	Synergistales	Synergistaceae
Sagittula	Bacteria	1.00	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Flavobacteria	Bacteria	0.99	Bacteroidetes	Flavobacteria	-	-
gamma	Bacteria	0.99	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae
Thermoanaerobacterium	Bacteria	0.98	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacterales Family III. Incertae Sedis
Methylovorus	Bacteria	0.98	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae
Ensifer	Bacteria	0.98	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
Alteromonas	Bacteria	0.97	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae
Eikenella	Bacteria	0.97	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
Acholeplasma	Bacteria	0.97	Tenericutes	Mollicutes	Acholeplasmatales	Acholeplasmataceae
Aromatoleum	Bacteria	0.97	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae
Halothermothrix	Bacteria	0.96	Firmicutes	Clostridia	Halanaerobiales	Halanaerobiaceae
Afipia	Bacteria	0.96	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae
Blautia	Bacteria	0.96	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Pseudoxanthomonas	Bacteria	0.95	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
Streptomyces	Bacteria	0.95	Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae
Paracoccus	Bacteria	0.95	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Listeria	Bacteria	0.94	Firmicutes	Bacilli	Bacillales	Listeriaceae
Loktanella	Bacteria	0.94	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Oxalobacter	Bacteria	0.94	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
Marine	Bacteria	0.93	Actinobacteria	Actinobacteria	Actinomycetales	Tsukamurellaceae
Ilyobacter	Bacteria	0.93	Fusobacteria	Fusobacteria (class)	Fusobacteriales	Fusobacteriaceae

Eremococcus	Bacteria	0.92	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae
Nitrosospira	Bacteria	0.92	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae
Herbaspirillum	Bacteria	0.92	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
Aciduliprofundum	Archaea	0.92	Euryarchaeota	unclassified	unclassified	unclassified
Thermosediminibacter	Bacteria	0.92	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacterales Family III. Incertae Sedis
Granulicatella	Bacteria	0.92	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae
Microcoleus	Bacteria	0.92	Cyanobacteria	unclassified	Oscillatoriales	Oscillatoriaceae
Acetohalobium	Bacteria	0.92	Firmicutes	Clostridia	Halanaerobiales	Halobacteroidaceae
Nodularia	Bacteria	0.90	Cyanobacteria	Nostocales	Nostocales	Nostocaceae
Dechloromonas	Bacteria	0.89	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae
Methylococcus	Bacteria	0.88	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylococcaceae
Xanthobacter	Bacteria	0.88	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae
Anaplasma	Bacteria	0.87	Proteobacteria	Alphaproteobacteria	Rickettsiales	Anaplasmataceae
Rhodoferax	Bacteria	0.87	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
Verminephrobacter	Bacteria	0.86	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
Alkalilimnicola	Bacteria	0.85	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae
Starkeya	Bacteria	0.85	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae
Nitrococcus	Bacteria	0.84	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae
Pseudoramibacter	Bacteria	0.84	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae
Clostridiales	Bacteria	0.84	Firmicutes	Clostridia	Clostridiales	-
Novosphingobium	Bacteria	0.84	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
Pyramidobacter	Bacteria	0.84	Synergistetes	Synergistia	Synergistales	Synergistaceae
Alkaliphilus	Bacteria	0.83	Firmicutes	Clostridia	Clostridiales	Clostridiaceae
Thermobaculum	Bacteria	0.82	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae

Chromohalobacter	Bacteria	0.81	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae
Erysipelothrix	Bacteria	0.81	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae
Methylobacillus	Bacteria	0.81	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae
Granulicella	Bacteria	0.80	Acidobacteria	Acidobacteria (class)	Acidobacteriales	Acidobacteriaceae
Catenibacterium	Bacteria	0.80	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae
Akkermansia	Bacteria	0.80	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae
Chelativorans	Bacteria	0.79	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae
Methanohalobium	Archaea	0.79	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae
Hyphomicrobium	Bacteria	0.79	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae
Herminiimonas	Bacteria	0.79	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
Frankia	Bacteria	0.79	Actinobacteria	Actinobacteria	Actinomycetales	Frankiaceae
Aggregatibacter	Bacteria	0.78	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
Rhodomicrobium	Bacteria	0.77	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae
Leeuwenhoekiella	Bacteria	0.77	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Providencia	Bacteria	0.77	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Heliobacterium	Bacteria	0.77	Firmicutes	Clostridia	Clostridiales	Heliobacteriaceae
Parascardovia	Bacteria	0.76	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae
Sphingopyxis	Bacteria	0.75	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
Hoeflea	Bacteria	0.75	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae
Granulibacter	Bacteria	0.75	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae
Reinekea	Bacteria	0.75	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae
Methanosarcina	Archaea	0.74	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae
Rhodocista	Bacteria	0.73	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
Azorhizobium	Bacteria	0.72	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae

Shuttleworthia	Bacteria	0.72	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Marinobacter	Bacteria	0.71	Firmicutes	Bacilli	Bacillales	Bacillaceae
Pelobacter	Bacteria	0.70	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	Pelobacteraceae
Microscilla	Bacteria	0.70	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae
Taylorella	Bacteria	0.70	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae
Moritella	Bacteria	0.69	Proteobacteria	Gammaproteobacteria	Alteromonadales	Moritellaceae
Nitrosopumilus	Archaea	0.69	Thaumarchaeota	Nitrosopumilales	Nitrosopumilales	Nitrosopumilaceae
Allochromatium	Bacteria	0.68	Proteobacteria	Gammaproteobacteria	Chromatiales	Chromatiaceae
Mitsuokella	Bacteria	0.67	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae
Moorella	Bacteria	0.66	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacteraceae
Moraxella	Bacteria	0.66	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae
Slackia	Bacteria	0.66	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae
Magnetospirillum	Bacteria	0.66	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
Spirochaeta	Bacteria	0.65	Spirochaetes	Spirochaetia	Spirochaetales	Spirochaetaceae
Janthinobacterium	Bacteria	0.65	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae
Bermanella	Bacteria	0.64	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae
Limnobacter	Bacteria	0.64	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
Lysinibacillus	Bacteria	0.64	Firmicutes	Bacilli	Bacillales	Bacillaceae
Kangiella	Bacteria	0.64	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Alcanivoracaceae
Methanohalophilus	Archaea	0.63	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae
Tropheryma	Bacteria	0.63	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcineae
Kordia	Bacteria	0.63	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Micromonospora	Bacteria	0.62	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae
Acidobacterium	Bacteria	0.62	Acidobacteria	Actinobacteria	Acidobacteriales	Acidobacteriaceae

Methanoculleus	Archaea	0.62	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Methanomicrobiaceae
Thermobispora	Bacteria	0.62	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae
Desulfonatronospira	Bacteria	0.61	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfohalobiaceae
Methanococcoides	Archaea	0.61	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae
Cellulophaga	Bacteria	0.61	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Chloroflexus	Bacteria	0.61	Chloroflexi	Chloroflexi	Chloroflexales	Chloroflexaceae
Megamonas	Bacteria	0.61	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae
Psychrobacter	Bacteria	0.61	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae
Acidaminococcus	Bacteria	0.60	Firmicutes	Negativicutes	Selenomonadales	Acidaminococcaceae
Ahrensia	Bacteria	0.60	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Nautella	Bacteria	0.60	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Phenylobacterium	Bacteria	0.59	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
Nisaea	Bacteria	0.59	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae
Endoriftia	Bacteria	0.58	Proteobacteria	Gammaproteobacteria	unclassified	unclassified
Octadecabacter	Bacteria	0.58	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Methanothermobacter	Archaea	0.57	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae
Thalassiobium	Bacteria	0.57	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Psychroflexus	Bacteria	0.57	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Picrophilus	Archaea	0.56	Euryarchaeota	Thermoplasmata	Thermoplasmatales	Picrophilaceae
Pelotomaculum	Bacteria	0.56	Firmicutes	Clostridia	Clostridiales	Peptococcaceae
Edwardsiella	Bacteria	0.56	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Asticcacaulis	Bacteria	0.56	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
Dinoroseobacter	Bacteria	0.56	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Simonsiella	Bacteria	0.56	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae

Atopobium	Bacteria	0.56	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae
Methylotenera	Bacteria	0.56	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae
Halomonas	Bacteria	0.55	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae
Ferroglobus	Archaea	0.55	Euryarchaeota	Archaeoglobi	Archaeoglobales	Archaeoglobaceae
Desulfotalea	Bacteria	0.55	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfobulbaceae
Terriglobus	Bacteria	0.55	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae
Alcanivorax	Bacteria	0.54	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Alcanivoracaceae
Symbiobacterium	Bacteria	0.54	Firmicutes	Clostridia	Clostridiales	Clostridiales Family XVIII. Incertae Sedis
Kytococcus	Bacteria	0.53	Actinobacteria	Actinobacteria	Actinomycetales	Dermacoccaceae
Hyphomonas	Bacteria	0.53	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Hyphomonadaceae
Aerococcus	Bacteria	0.53	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae
Subdoligranulum	Bacteria	0.52	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
Maricaulis	Bacteria	0.52	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Hyphomonadaceae
Dialister	Bacteria	0.51	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae
Oceanicaulis	Bacteria	0.51	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Hyphomonadaceae
Arcobacter	Bacteria	0.51	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae
Korebacter	Bacteria	0.51	Acidobacteria	Unclassified	unclassified	unclassified
Cryptobacterium	Bacteria	0.51	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae
Brevundimonas	Bacteria	0.51	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
Sphaerobacter	Bacteria	0.50	Chloroflexi	Thermomicrobia	Sphaerobacterales	Sphaerobacteraceae
Magnetococcus	Bacteria	0.50	Proteobacteria	Alphaproteobacteria	Magnetococcales	Magnetococcaceae
Macrococcus	Bacteria	0.50	Firmicutes	Bacilli	Bacillales	Staphylococcaceae
Alicyclobacillus	Bacteria	0.49	Firmicutes	Bacilli	Bacillales	Alicyclobacillaceae
Parvularcula	Bacteria	0.49	Proteobacteria	Alphaproteobacteria	Parvularculales	Parvularculaceae

Parvibaculum	Bacteria	0.49	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae
Lawsonia	Bacteria	0.49	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae
Salinibacter	Bacteria	0.48	Bacteroidetes	Bacteroidetes Order II. Incertae sedis	Bacteroidetes Order II. Incertae sedis	Rhodothermaceae
Caminibacter	Bacteria	0.48	Proteobacteria	Epsilonproteobacteria	Nautiliales	Nautiliaceae
Enterobacteriaceae	Bacteria	0.48	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Ketogulonicigenium	Bacteria	0.48	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Kosmotoga	Bacteria	0.47	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae
Cenarchaeum	Archaea	0.47	Thaumarchaeota	Cenarchaeales	Cenarchaeales	Cenarchaeaceae
Arcanobacterium	Bacteria	0.47	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae
Methanococcus	Archaea	0.46	Euryarchaeota	Methanococci	Methanococcales	Methanococcaceae
Salinispora	Bacteria	0.46	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae
Ulvibacter	Bacteria	0.46	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Rickettsiella	Bacteria	0.46	Proteobacteria	Gammaproteobacteria	Legionellales	Coxiellaceae
Thermomonospora	Bacteria	0.46	Actinobacteria	Actinobacteria	Actinomycetales	Thermomonosporaceae
Klebsiella	Bacteria	0.46	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Natranaerobius	Bacteria	0.46	Firmicutes	Clostridia	Natranaerobiales	Natranaerobiaceae
Desulfotomaculum	Bacteria	0.45	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Dickeya	Bacteria	0.45	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Methylosinus	Bacteria	0.44	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylocystaceae
Jannaschia	Bacteria	0.44	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Enterobacter	Bacteria	0.44	Proteobacteria	Alphaproteobacteria	Rickettsiales	Anaplasmataceae
Halothiobacillus	Bacteria	0.44	Proteobacteria	Gammaproteobacteria	Chromatiales	Halothiobacillaceae
Blastopirellula	Bacteria	0.44	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae
Azotobacter	Bacteria	0.44	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae

Methylocella	Bacteria	0.44	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae
Cyanobium	Bacteria	0.43	Cyanobacteria	unclassified	Chroococcales	Synechococcaceae
Methylobacter	Bacteria	0.42	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylococcaceae
Methanoplanus	Archaea	0.42	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Methanomicrobiaceae
Acidothermus	Bacteria	0.41	Actinobacteria	Actinobacteria	Actinomycetales	Acidothermaceae
Haliangium	Bacteria	0.40	Proteobacteria	Deltaproteobacteria	Myxococcales	Kofleriaceae
Methylocystis	Bacteria	0.40	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylocystaceae
Desulfobulbus	Bacteria	0.40	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfobulbaceae
Sutterella	Bacteria	0.40	Proteobacteria	Betaproteobacteria	Burkholderiales	Sutterellaceae
Neorickettsia	Bacteria	0.40	Proteobacteria	Alphaproteobacteria	Rickettsiales	Anaplasmataceae
Methanosphaerula	Archaea	0.39	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Methanoregulaceae
Ferroplasma	Archaea	0.38	Euryarchaeota	Thermoplasmata	Thermoplasmatales	Ferroplasmaceae
Oceanibulbus	Bacteria	0.38	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Desulfuromonas	Bacteria	0.38	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	Desulfuromonadaceae
Cyanothece	Bacteria	0.38	Cyanobacteria	unclassified	Chroococcales	Cyanobacteriaceae
Petrotoga	Bacteria	0.38	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae
Oceanobacillus	Bacteria	0.38	Firmicutes	Bacilli	Bacillales	Bacillaceae
Maribacter	Bacteria	0.38	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Saccharophagus	Bacteria	0.37	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae
Brevibacillus	Bacteria	0.36	Firmicutes	Bacilli	Bacillales	Paenibacillaceae
Streptosporangium	Bacteria	0.36	Actinobacteria	Actinobacteria	Actinomycetales	Streptosporangiaceae
Tolumonas	Bacteria	0.35	Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae
Dehalogenimonas	Bacteria	0.35	Chloroflexi	Dehalococcoidetes	unclassified	unclassified
Croceibacter	Bacteria	0.35	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae

Teredinibacter	Bacteria	0.35	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria incertae sedis	unclassified
Raphidiopsis	Bacteria	0.35	Cyanobacteria	Nostocales	Nostocales	Nostocaceae
Truepera	Bacteria	0.34	Deinococcus-Thermus	Deinococci	Deinococcales	Trueperaceae
Beijerinckia	Bacteria	0.34	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae
Maritimibacter	Bacteria	0.34	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Ferrimonas	Bacteria	0.34	Proteobacteria	Gammaproteobacteria	Alteromonadales	Ferrimonadaceae
Olsenella	Bacteria	0.34	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae
Plesiocystis	Bacteria	0.33	Proteobacteria	Deltaproteobacteria	Myxococcales	Nannocystaceae
Chthoniobacter	Bacteria	0.33	Verrucomicrobia	Spartobacteria	unclassified	unclassified
Citrobacter	Bacteria	0.33	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Desulfohalobium	Bacteria	0.31	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfohalobiaceae
Planctomyces	Bacteria	0.31	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae
Nostoc	Bacteria	0.31	Cyanobacteria	Nostocales	Nostocales	Nostocaceae
Waddlia	Bacteria	0.30	Chlamydiae	Chlamydiia	Chlamydiales	Waddliaceae
Pectobacterium	Bacteria	0.30	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Psychromonas	Bacteria	0.30	Proteobacteria	Gammaproteobacteria	Alteromonadales	Psychromonadaceae
Thermincola	Bacteria	0.29	Firmicutes	Clostridia	Clostridiales	Peptococcaceae
Wigglesworthia	Bacteria	0.29	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Mariprofundus	Bacteria	0.29	Proteobacteria	Zetaproteobacteria	Mariprofundales	Mariprofundaceae
Cronobacter	Bacteria	0.29	Proteobacteria	Alphaproteobacteria	Rickettsiales	Anaplasmataceae
Mannheimia	Bacteria	0.28	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
Anoxybacillus	Bacteria	0.28	Firmicutes	Bacilli	Bacillales	Bacillaceae
Capnocytophaga	Bacteria	0.28	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Ktedonobacter	Bacteria	0.28	Chloroflexi	Ktedonobacteria	Ktedonobacterales	Ktedonobacteraceae

Desulfurivibrio	Bacteria	0.27	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfobulbaceae
Stigmatella	Bacteria	0.27	Proteobacteria	Deltaproteobacteria	Myxococcales	Cystobacteraceae
Scardovia	Bacteria	0.27	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae
Arthrobacter	Bacteria	0.26	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae
Pasteurella	Bacteria	0.26	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
Methylacidiphilum	Bacteria	0.26	Verrucomicrobia	Methylacidiphilales	Methylacidiphilales	Methylacidiphilaceae
Thiomicrospira	Bacteria	0.26	Proteobacteria	Gammaproteobacteria	Thiotrichales	Piscirickettsiaceae
Rubrobacter	Bacteria	0.25	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae
Histophilus	Bacteria	0.25	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
Syntrophobacter	Bacteria	0.25	Proteobacteria	Deltaproteobacteria	Syntrophobacterales	Syntrophobacteraceae
Mucilaginibacter	Bacteria	0.25	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae
Methanocella	Archaea	0.25	Euryarchaeota	Methanomicrobia	Methanocellales	Methanocellaceae
Baumannia	Bacteria	0.24	Proteobacteria	Gammaproteobacteria	unclassified	unclassified
Serratia	Bacteria	0.23	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Methanobrevibacter	Archaea	0.23	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae
Halobacterium	Archaea	0.23	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Fibrobacter	Bacteria	0.23	Fibrobacteres	Fibrobacteres	Fibrobacterales	Fibrobacteraceae
Burkholderiales	Bacteria	0.22	Proteobacteria	Betaproteobacteria	Burkholderiales	-
Holdemania	Bacteria	0.22	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae
Oribacterium	Bacteria	0.22	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Desulfomicrobium	Bacteria	0.22	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfomicrobiaceae
Gemmata	Bacteria	0.22	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae
Congregibacter	Bacteria	0.22	Proteobacteria	Gammaproteobacteria	unclassified	unclassified
Parachlamydia	Bacteria	0.21	Chlamydiae	Chlamydiia	Chlamydiales	Parachlamydiaceae

Rothia	Bacteria	0.21	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae
Zunongwangia	Bacteria	0.21	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Rickettsiales	Bacteria	0.21	Schizomycetes	Microtatobiotes	Rickettsiales	-
Methanosphaera	Archaea	0.20	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae
Trichodesmium	Bacteria	0.20	Cyanobacteria	Cyanophyceae	Oscillatoriales	Phormidiaceae
Brachybacterium	Bacteria	0.20	Actinobacteria	Actinobacteria	Actinomycetales	Dermabacteraceae
Verrucomicrobium	Bacteria	0.20	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae
Bilophila	Bacteria	0.20	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae
Rhodothermus	Bacteria	0.20	Bacteroidetes	Bacteroidetes Order II. Incertae sedis	Bacteroidetes Order II. Incertae sedis	Rhodothermaceae
Lyngbya	Bacteria	0.19	Cyanobacteria	Cyanophyceae	Oscillatoriales	Oscillatoriaceae
Victivallis	Bacteria	0.18	Lentisphaerae	Victivallaceae	Victivallaceae	Victivallaceae
Eggerthella	Bacteria	0.18	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae
Rhodopirellula	Bacteria	0.18	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae
Solibacter	Bacteria	0.18	Acidobacteria	Solibacteres	Solibacterales	Solibacteraceae
Chlorobium	Bacteria	0.18	Chlorobi	Chlorobia	Chlorobiales	Chlorobiaceae
Rhodococcus	Bacteria	0.18	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae
Phascolarctobacterium	Bacteria	0.18	Firmicutes	Negativicutes	Selenomonadales	Acidaminococcaceae
Herpetosiphon	Bacteria	0.18	Chloroflexi	Chloroflexi	Herpetosiphonales	Herpetosiphonaceae
Crocosphaera	Bacteria	0.17	Cyanobacteria	Cyanophyceae	Chroococcales	Cyanobacteriaceae
Xylanimonas	Bacteria	0.17	Actinobacteria	Actinobacteria	Actinomycetales	Promicromonosporaceae
Cellulomonas	Bacteria	0.17	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae
Stackebrandtia	Bacteria	0.16	Actinobacteria	Actinobacteria	Actinomycetales	Glycomycetaceae
Cytophaga	Bacteria	0.16	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae
Saccharopolyspora	Bacteria	0.16	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae

Colwellia	Bacteria	0.16	Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae
Methanosaeta	Archaea	0.16	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosaetaceae
Cellvibrio	Bacteria	0.16	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
Syntrophus	Bacteria	0.15	Proteobacteria	Deltaproteobacteria	Syntrophobacterales	Syntrophaceae
Basfia	Bacteria	0.15	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
Oscillochloris	Bacteria	0.15	Chloroflexi	Chloroflexi	Chloroflexales	Oscillochloridaceae
Planococcus	Bacteria	0.15	Firmicutes	Bacilli	Bacillales	Planococcaceae
Turicibacter	Bacteria	0.15	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae
Succinatimonas	Bacteria	0.15	Proteobacteria	Gammaproteobacteria	Aeromonadales	Succinivibrionaceae
Thermomicrobium	Bacteria	0.14	Chloroflexi	Thermomicrobia	Thermomicrobiales	Thermomicrobiaceae
Bryantella	Bacteria	0.14	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Desulfitobacterium	Bacteria	0.14	Firmicutes	Clostridia	Clostridiales	Peptococcaceae
Lentisphaera	Bacteria	0.14	Lentisphaerae	Lentisphaerales	Lentisphaerales	Lentisphaeraceae
Sanguibacter	Bacteria	0.14	Actinobacteria	Actinobacteria	Actinomycetales	Sanguibacteraceae
Aeromicrobium	Bacteria	0.14	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae
Desulfarculus	Bacteria	0.14	Proteobacteria	Deltaproteobacteria	Desulfarculales	Desulfarculaceae
Proteus	Bacteria	0.12	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Robiginitalea	Bacteria	0.12	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Myxococcus	Bacteria	0.12	Proteobacteria	Deltaproteobacteria	Myxococcales	Myxococcaceae
Desulfurispirillum	Bacteria	0.12	Chrysiogenetes	Chrysiogenetes	Chrysiogenales	Chrysiogenaceae
Xenorhabdus	Bacteria	0.12	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Gramella	Bacteria	0.12	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Campylobacterales	Bacteria	0.12	Proteobacteria	Epsilonproteobacteria	Campylobacterales	-
Beutenbergia	Bacteria	0.11	Actinobacteria	Actinobacteria	Actinomycetales	Beutenbergiaceae

Enhydrobacter	Bacteria	0.11	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae
Sulfurimonas	Bacteria	0.11	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae
Pediococcus	Bacteria	0.11	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae
Conexibacter	Bacteria	0.11	Actinobacteria	Actinobacteria	Solirubrobacterales	Conexibacteraceae
Segniliparus	Bacteria	0.11	Actinobacteria	Actinobacteria	Actinomycetales	Segniliparaceae
Flavobacteriales	Bacteria	0.11	Bacteroidetes	Flavobacteria	Flavobacteriales	-
Neptuniibacter	Bacteria	0.10	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae
Nautilia	Bacteria	0.10	Proteobacteria	Epsilonproteobacteria	Nautiliales	Nautiliaceae
Gordonibacter	Bacteria	0.10	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae
Nitratifractor	Bacteria	0.10	Proteobacteria	Epsilonproteobacteria	Campylobacterales	unclassified
Hahella	Bacteria	0.10	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Hahellaceae
Gardnerella	Bacteria	0.10	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae
Microcystis	Bacteria	0.10	Cyanobacteria	not assigned	Chroococcales	Microcystaceae
Coraliomargarita	Bacteria	0.10	Verrucomicrobia	Opitutae	Puniceicoccales	Puniceicoccaceae
Carnobacterium	Bacteria	0.09	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae
Sorangium	Bacteria	0.09	Proteobacteria	Deltaproteobacteria	Myxococcales	Polyangiaceae
Dermacoccus	Bacteria	0.08	Actinobacteria	Actinobacteria	Actinomycetales	Dermacoccaceae
Cylindrospermopsis	Bacteria	0.08	Cyanobacteria	Nostocales	Nostocales	Nostocaceae
Photorhabdus	Bacteria	0.08	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Halorubrum	Archaea	0.07	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Nakamurella	Bacteria	0.07	Actinobacteria	Actinobacteria	Actinomycetales	Nakamurellaceae
Kocuria	Bacteria	0.07	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae
Alistipes	Bacteria	0.07	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae
Chitinophaga	Bacteria	0.07	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Crenotrichaceae

Aliivibrio	Bacteria	0.07	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae
Nocardia	Bacteria	0.07	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae
Beggiatoa	Bacteria	0.07	Proteobacteria	Gammaproteobacteria	Thiotrichales	Thiotrichaceae
Peptostreptococcus	Bacteria	0.07	Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae
Methanocorpusculum	Archaea	0.07	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Methanocorpusculaceae
Geodermatophilus	Bacteria	0.07	Actinobacteria	Actinobacteria	Actinomycetales	Geodermatophilaceae
Haloferax	Archaea	0.07	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Pelodictyon	Bacteria	0.06	Chlorobi	Chlorobia	Chlorobiales	Chlorobiaceae
Weissella	Bacteria	0.06	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae
Gloeobacter	Bacteria	0.06	Cyanobacteria	Gloeobacteria	Gloeobacterales	Gloeobacteraceae
Dethiobacter	Bacteria	0.06	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae
Intrasporangium	Bacteria	0.06	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae
Leifsonia	Bacteria	0.06	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae
Paludibacter	Bacteria	0.06	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae
Cyanobacterium	Bacteria	0.06	Cyanobacteria	unclassified	Chroococcales	unclassified
Nocardioides	Bacteria	0.06	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae
Actinosynnema	Bacteria	0.06	Actinobacteria	Actinobacteria	Actinomycetales	Actinosynnemataceae
Rahnella	Bacteria	0.06	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
Halogeometricum	Archaea	0.06	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Chlorobaculum	Bacteria	0.06	Chlorobi	Chlorobia	Chlorobiales	Chlorobiaceae
Thermobifida	Bacteria	0.06	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiopsaceae
Kineococcus	Bacteria	0.06	Actinobacteria	Actinobacteria	Actinomycetales	Kineosporiaceae
Sulfurovum	Bacteria	0.05	Proteobacteria	Epsilonproteobacteria	unclassified	unclassified
Dietzia	Bacteria	0.05	Actinobacteria	Actinobacteria	Actinomycetales	Dietziaceae

Halomicrobium	Archaea	0.05	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Odoribacter	Bacteria	0.05	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae
Sulfuricurvum	Bacteria	0.05	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae
Wolinella	Bacteria	0.05	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae
Desulfococcus	Bacteria	0.05	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfobacteraceae
Janibacter	Bacteria	0.05	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae
Marivirga	Bacteria	0.05	Bacteroidetes	Cytophagia	Cytophagales	Flammeovirgaceae
Candidate	Bacteria	0.05	-	-	-	-
Desulfatibacillum	Bacteria	0.05	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfobacteraceae
Syntrophothermus	Bacteria	0.05	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae
Opitutus	Bacteria	0.05	Verrucomicrobia	Opitutae	Opitutaceae	Opitutaceae
Nocardiopsis	Bacteria	0.05	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiopsaceae
Ruminococcaceae	Bacteria	0.05	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
Bulleidia	Bacteria	0.05	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae
Pseudonocardia	Bacteria	0.04	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae
Syntrophomonas	Bacteria	0.04	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae
Sulfurospirillum	Bacteria	0.04	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae
Bacteriovorax	Bacteria	0.04	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae
Kribbella	Bacteria	0.04	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae
Gemmatimonas	Bacteria	0.04	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae
Agreia	Bacteria	0.04	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae
Catenulispora	Bacteria	0.04	Actinobacteria	Actinobacteria	Actinomycetales	Catenulisporaceae
Pseudoflavonifractor	Bacteria	0.04	Firmicutes	Clostridia	Clostridiales	unclassified
Pedosphaera	Bacteria	0.04	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobia subdivision 3

Amycolatopsis	Bacteria	0.04	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae
Methanospirillum	Archaea	0.04	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Methanospirillaceae
Catonella	Bacteria	0.04	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
Haloterrigena	Archaea	0.04	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Brevibacterium	Bacteria	0.04	Actinobacteria	Actinobacteria	Actinomycetales	Brevibacteriaceae
Isosphaera	Bacteria	0.04	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae
Roseiflexus	Bacteria	0.04	Chloroflexi	Chloroflexi	Chloroflexales	Chloroflexaceae
Saccharomonospora	Bacteria	0.04	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae
Natrialba	Archaea	0.03	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Desulfobacterium	Bacteria	0.03	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfobacteraceae
Solobacterium	Bacteria	0.03	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae
Denitrovibrio	Bacteria	0.03	Deferribacteres	Deferribacteres	Deferribacterales	Deferribacteraceae
Syntrophobotulus	Bacteria	0.03	Firmicutes	Clostridia	Clostridiales	Peptococcaceae
Jonesia	Bacteria	0.03	Actinobacteria	Actinobacteria	Actinomycetales	Jonesiaceae
Prosthecochloris	Bacteria	0.03	Chlorobi	Chlorobia	Chlorobiales	Chlorobiaceae
Thermotogales	Bacteria	0.03	Thermotogae	Thermotogae	Thermotogales	-
Synechocystis	Bacteria	0.03	Cyanobacteria	unclassified	Chroococcales	Cyanobacteriaceae
Haladaptatus	Archaea	0.03	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Natronomonas	Archaea	0.03	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Renibacterium	Bacteria	0.02	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae
Streptobacillus	Bacteria	0.02	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae
Halalkalicoccus	Archaea	0.02	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Opitutaceae	Bacteria	0.02	Verrucomicrobia	Opitutae	Opitutaceae	Opitutaceae
Parabacteroides	Bacteria	0.02	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae

Abiotrophia	Bacteria	0.02	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae
Anaerolinea	Bacteria	0.02	Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae
Synergistetes	Bacteria	0.02	Synergistetes	-	-	-
Parvimonas	Bacteria	0.02	Firmicutes	Clostridia	Clostridiales	Clostridiales Family XI. Incertae Sedis
Oscillatoria	Bacteria	0.02	Cyanobacteria	unclassified	Oscillatoriales	Oscillatoriaceae
Acetivibrio	Bacteria	0.02	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
Acaryochloris	Bacteria	0.02	Cyanobacteria	unclassified	unclassified	unclassified
Anabaena	Bacteria	0.02	Cyanobacteria	Nostocales	Nostocales	Nostocaceae
Chloroherpeton	Bacteria	0.02	Chlorobi	Chlorobia	Chlorobiales	Chlorobiaceae
Pirellula	Bacteria	0.01	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae
Blattabacterium	Bacteria	0.01	Bacteroidetes	Flavobacteriia	Flavobacteriales	Blattabacteriaceae
Halorhabdus	Archaea	0.01	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Erysipelotrichaceae	Bacteria	0.01	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae
Riemerella	Bacteria	0.01	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Sebaldella	Bacteria	0.01	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae
Haloquadratum	Archaea	0.01	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
Flavobacterium	Bacteria	0.01	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Flavobacteriaceae	Bacteria	0.01	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Dyadobacter	Bacteria	0.01	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae
Weeksella	Bacteria	0.01	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Leadbetterella	Bacteria	0.01	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae
Verrucomicrobiales	Bacteria	0.01	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	-
Spirosoma	Bacteria	0.01	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae