BMC Genomics

Metabolic characteristics of dominant microbes and key rare species from an acidic hot spring in Taiwan revealed by metagenomics

Kuei-Han Lin[1†], Ben-Yang Liao[2†], Hao-Wei Chang[1,3†], Shiao-Wei Huang[1], Ting-Yan Chang[2], Cheng-Yu Yang[4], Yu-Bin Wang[1,5], Yu-Teh Kirk Lin[1,6], Yu-Wei Wu[7,8], Sen-Lin Tang[4*] and Hon-Tsen Yu[1,9*]

Abstract

Background: Microbial diversity and community structures in acidic hot springs have been characterized by 16S rRNA gene-based diversity surveys. However, our understanding regarding the interactions among microbes, or between microbes and environmental factors, remains limited.

Results: In the present study, a metagenomic approach, followed by bioinformatics analyses, were used to predict interactions within the microbial ecosystem in Shi-Huang-Ping (SHP), an acidic hot spring in northern Taiwan. Characterizing environmental parameters and potential metabolic pathways highlighted the importance of carbon assimilatory pathways. Four distinct carbon assimilatory pathways were identified in five dominant genera of bacteria. Of those dominant carbon fixers, *Hydrogenobaculum* bacteria outcompeted other carbon assimilators and dominated the SHP, presumably due to their ability to metabolize hydrogen and to withstand an anaerobic environment with fluctuating temperatures. Furthermore, most dominant microbes were capable of metabolizing inorganic sulfur-related compounds (abundant in SHP). However, *Acidithiobacillus ferrooxidans* was the only species among key rare microbes with the capability to fix nitrogen, suggesting a key role in nitrogen cycling. In addition to potential metabolic interactions, based on the 16S rRNAs gene sequence of *Nanoarchaeum*-related and its potential host *Ignicoccus*-related archaea, as well as sequences of viruses and CRISPR arrays, we inferred that there were complex microbe-microbe interactions.

Conclusions: Our study provided evidence that there were numerous microbe-microbe and microbe-environment interactions within the microbial community in an acidic hot spring. We proposed that *Hydrogenobaculum* bacteria were the dominant microbial genus, as they were able to metabolize hydrogen, assimilate carbon and live in an anaerobic environment with fluctuating temperatures.

Background

Microbial diversity surveys based on 16S rRNA gene are the most common culture-independent method to characterize composition of a microbial community and to compare microbial diversity among habitats [1]. This method has been widely used to characterize microbial community structure in a variety of acidic hot springs in diverse locations, including the Azores Islands in the North Atlantic Ocean [2], Colombian Andes [3], Iceland [4], Lassen Volcanic National Park [5, 6], Montserrat in the Caribbean sea [7], New Mexico [8], Philippines [9], St. Luica in the Lesser Antilles [10], Tengchong [11, 12], Tibet [13], West Java-Indonesia [14], and Yellowstone National Park [15, 16]. Overall, there was evidence that microbial communities in hot springs were closely linked to local environmental conditions. For example, *Sulfolobus* and *Metallosphaera*, two microbial taxa that

* Correspondence: sltang@gate.sinica.edu.tw; ayu@ntu.edu.tw
†Equal contributors
[4]Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan, Republic of China
[1]Department of Life Science, National Taiwan University, Taipei 10617, Taiwan, Republic of China
Full list of author information is available at the end of the article

Lin *et al. BMC Genomics* (2015) 16:1029

Page 2 of 16

metabolize sulfur-containing compounds, were dominant organisms in sulfur-rich hot springs [14].

Although characterizing composition of microbial communities could provide insights regarding potential metabolic interactions among microbes or between microbes and environmental factors, additional methods are required for more comprehensive understanding. For example, a metagenomic approach could be used to characterize potential metabolic activities of microbial communities. In that regard, Jiménez and coworkers successfully identified several key genes involved in metabolism of nitrogen (e.g., *narGHI*, *nirS*, *norBCDQ* and *nosZ*) and sulfur (e.g., *cysDN*, *cysNC* and *aprA*) in a microbial community in El Coquito spring, National Nature Park Los Nevados, Colombian Andes, Columbia [17]. Furthermore, in a comparative metagenomic study, Inskeep et al. characterized diverse metabolic strategies related to geochemical characteristics of two acidic hot springs (Crater Hills and Norris Geysers Basin) in Yellowstone National Park [18]. In contrast to the El Coquito spring, microbes in Crater Hills and Norris Geysers Basin adopted non-photosynthetic carbon assimilation pathways (reductive citrate cycle), apparently because water temperatures of these springs (>65 °C) approached upper temperature limits for photosynthesis [19].

Taiwan, located on the "Ring of Fire" in the West Pacific (with copious geothermal activity), is well suited for studying interactions between thermophiles and environmental factors. Tatun Volcanic Group (TVG) area (ca. 400 km^2) in northern Taiwan comprises the largest volcanic group (>20 volcanoes) on the island. Volcanic activity started 2.8–2.5 Ma ago, with the last massive explosive event between 0.8 and 0.2 Ma ago [20]. The TVG has a ubiquitous smell of sulfuric gases, solfataras rimmed with sulfur crystals, and some sulfur mines [21]. Hot springs in this area are primarily of meteoric origin, surfacing after being heated by geothermal energy [22, 23]. Typical hot springs in the TVG area, such as Shi-Huang-Ping (SHP; also denoted Szehuangtzeping in some reports), are highly acidic (pH ~ 2.5) due to dissolved inorganic sulfur-containing compounds, and water temperature ranges from ~ 50 to 85 °C [21].

Well-documented geological features make SHP ideal for conducting a metagenomic study to characterize metabolic potential of the microbial community in acidic hyperthermal environments and relationships between the microbial community and local geochemistry. In reports that used 16S rRNA gene-based diversity surveys, *Hydrogenobaculum* was the dominant microbe at two acidic hot springs in the TVG area [24]. Furthermore, *Hydrogenobaculum*-dominant features were reported in two acidic hot springs in Yellowstone National Park (Dragon Spring and One Hundred Spring; [25, 26]). However, microbe-microbe and microbe-environment interactions within the SHP ecosystem have not been characterized.

In this study, a metagenomic approach was used to elucidate putative interactions within an acidic hot spring ecosystem. Metabolic interactions were predicted by searching metagenomic data against the KEGG database and information derived from the literature. Analyzing microbial community structure revealed potential interactions between *Ignicoccus* and an archaeal parasite *Nanoarchaea*. Furthermore, based on CRISPR array analysis, there were also potential microbe-virus interactions.

Results and discussion

Hydrological parameters of SHP

Limnological parameters of the SHP are shown (Table 1). Temperature and pH of the sample were 69 °C and 2.5, respectively. Concentrations of several ions (Cl$^-$, HCO^{3-}, Ca^{2+}, Mg^{2+}, K$^+$ and Na$^+$) were low, as was that of dissolved organic carbon (~1 mg/L). In addition, concentrations of sulfate (378 mg/L), hydrogen sulfide (52.7 mg/L) thiosulfate (0.12 mg/L) and elemental sulfur (0.50 mg/L) in SHP water were also determined.

Table 1 Geochemical and physical parameters of SHP hot spring water

Parameter[a]	Current study	Song et al., 2005[b] [21]
pH	2.5	2.77 – 3.25
Temperature (°C)	69.0	49.8 – 85.1
TDS	707	177 – 1674
EC (μS/cm^2)	1760	266 – 1039
ORP (mv)	−62	–
DOC	1.0	–
HCO$_3^{2-}$	<0.03	L
NO^{3-}	–	L – 0.9
PO$_4^{3-}$	–	L
Cl$^-$ and other halides	9.3	L (F$^-$) / 4.93 – 10.3 (Cl$^-$) / L – 3.78 (Br$^-$)
SO$_4^{2-}$	378	100 – 450
S^0	0.50	–
H$_2$S	52.7	–
S$_2$O$_3^{2-}$	0.12	–
Fe^{2+} and Fe^{3+}	111	1.8 – 54.9 (Fe^{2+})
Ca^{2+}	1.32	5.8 – 977
Na$^+$	12.0	L – 13.4
Mg^{2+}	1.21	1.21 – 9.70
K$^+$	2.17	1.40 – 6.00
Al^{3+}	19.0	0.70 – 17.3
Total As	0.0012	–

"L" represented concentrations below detection limits
TDS total dissolved solids, *EC* electrical conductivity, *ORP* oxidation/reduction potential, *DOC* dissolved organic carbon
[a]Units were mg/L for all end points except pH
[b]Source: Geological survey and potential application of hot springs and geothermal energy of Yangmingshan (in Chinese); monthly report of environmental parameters in the SHP acidic hot springs in 2005

Lin *et al. BMC Genomics* (2015) 16:1029

Page 3 of 16

The abundant genera in SHP

The top 20 abundant genera (eight bacterial and 12 archaeal genera; Fig. 1) were selected (from 16S rRNA gene-based diversity surveys) for phylogenetic analyses. Although there were more archaeal than bacterial genera in the top 20, bacteria clearly dominated the microbial community in SHP, based on relative abundance of 16S rRNA. In that regard, *Hydrogenobaculum* bacteria accounted for 86.30 % of RA_{16S} (relative abundance in 16S rRNA gene-based diversity survey), whereas the second most abundant genus, *Nanoarchaeum* (an archaeal genus) only accounted for approximately 0.99 % (Additional file 1: Table S1).

Relative abundances of the microbial community were also analyzed based on direct shotgun sequence (DSS) contigs. Major genera identified using this method were designated genomic information-rich genera, because contig information was used in analyses of metabolic ability. Nine genomic information-rich genera were identified, namely, *Hydrogenobaculum, Vulcanisaeta, Thermoproteus, Caldisphaera, Sulfolobus, Caldivirga, Acidithiobacillus, Thiomonas,* and *Metallosphaera* (Table 2 and Additional file 1: Figure S1). It was noteworthy that the order of the ranking between the two lists, the information-rich genera and the 20 abundant genera from 16S rRNA gene-based method, were similar, with *Hydrogenobaculum* at the top of both lists and all genomic information-rich genera in the top 20 of the 16S rRNA gene-based diversity survey.

Inconsistencies between compositional lists of 16S rRNA gene-based diversity and metagenomic information

Ranking and composition of dominant microbes differed between the genomic information-rich genera list and the dominant microbe list (derived from 16S rRNA gene-based diversity surveys) identified in the present study. There were several potential reasons, including variations among microbes in genome sizes and copy numbers of 16S rRNA gene, and the threshold used. For example, although *Nanoarchaeum* was one of the most abundant genera in the top 20 16S rRNA gene-based list, it was absent from the list of the genomic information-rich genera (Table 2). This was attributed to its small genome (~490 kb; [27]), which would reduce the probability of being detected during sequencing.

Advantageous characteristics of *Hydrogenobaculum* in SHP

Hydrogenobaculum was the predominant genus in SHP, where the temperature and pH were 69 °C and 2.5, respectively. Similarly, bacteria of the same genus also predominated in other acidic hot springs with variable (albeit harsh) environmental conditions, including Dragon Spring
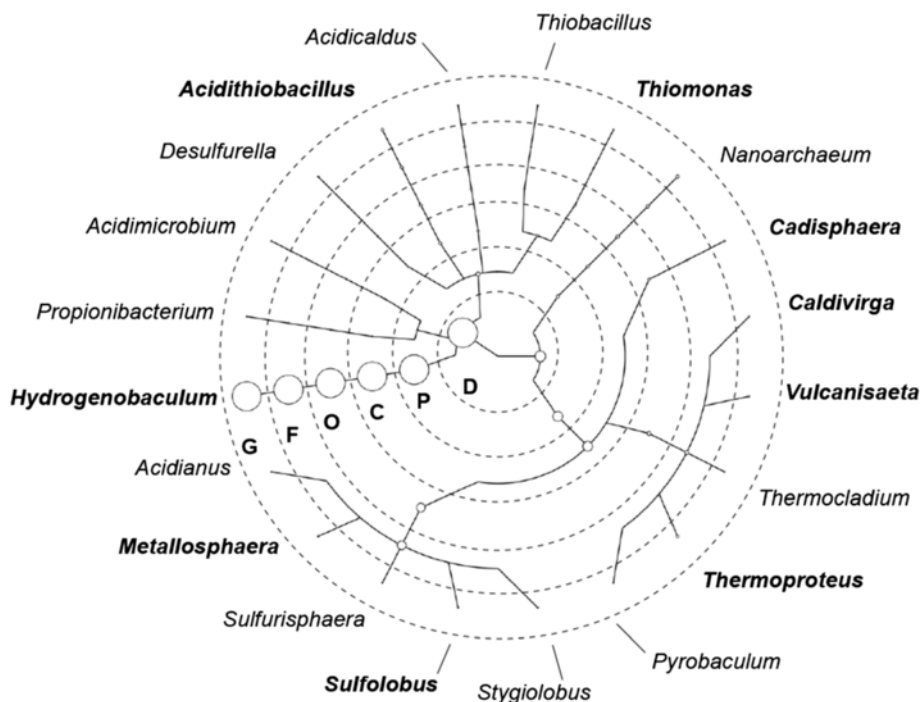
Fig. 1 Phylogenetic tree generated from top 20 most abundant genera identified with 16S rRNA gene-based diversity survey. Diameters of circles are proportional to abundances; the smallest circles represent RA_{16S} ~ 0.02 %, whereas the largest circle (*Hydrogenobaculum*) represent RA_{16S} = 86.3 %. Letters within each circle represent taxonomy levels: domain (D), phylum (P), class (C), order (O), family (F) and genus (G). Genera names are outside the circle. Genera names in bold were common genera (top 20 most abundant and information-rich genera)

Lin *et al. BMC Genomics* (2015) 16:1029

Page 4 of 16

Table 2 Information-rich genera and the top 20 most abundant genera derived from SNP hot springs water

Rank in abundance	Information-rich genera[a]		Top 20 most abundant genera identified by 16S rDNA analysis	
1	**Hydrogenobaculum**	5.73 %	**Hydrogenobaculum**	86.31 %
2	**Vulcanisaeta**	3.64 %	*Nanoarchaeum*	0.99 %
3	**Thermoproteus**	2.82 %	**Acidithiobacillus**	0.85 %
4	**Caldisphaera**	2.66 %	**Thermoproteus**	0.67 %
5	**Sulfolobus**	2.43 %	**Caldisphaera**	0.47 %
6	**Caldivirga**	2.39 %	**Thiomonas**	0.26 %
7	**Acidithiobacillus**	2.17 %	*Acidicaldus*	0.23 %
8	**Thiomonas**	1.66 %	*Sulfurisphaera*	0.21 %
9	*Metallosphaera*	1.50 %	*Acidianus*	0.17 %
10			**Caldivirga**	0.16 %
11			**Metallosphaera**	0.16 %
12			**Vulcanisaeta**	0.16 %
13			*Thiobacillus*	0.14 %
14			**Sulfolobus**	0.12 %
15			*Stygiolobus*	0.11 %
16			*Thermocladium*	0.10 %
17			*Pyrobaculum*	0.05 %
18			*Desulfurella*	0.04 %
19			*Acidimicrobium*	0.02 %
20			*Propionibacterium*	0.02 %

[a]Genera contained the relative abundance of contigs exceeding 1 %
The name in bold represents the genus share between two lists

(70 ~ 72 °C; pH 3.1; [25, 26]), One Hundred Spring (73 °C; pH 3.5; 25, 26) and Norris Geyser (65 °C; pH 3.0; [26]).

The abilities of *Hydrogenobaculum* bacteria to assimilate carbon and metabolize hydrogen were suggested as crucial characteristics for living in an acidic hot spring [28, 29]. Indeed, carbon assimilation ability would be important for bacteria residing in SHP, due to the low dissolved organic carbon (DOC) concentration (1 mg/L) in spring water. However, *Hydrogenobaculum* was not the only microbial genus in SHP that assimilated inorganic carbon. Based on our metagenomic analysis, genes for carbon assimilation pathways were present in five of the nine genomic information-rich genera, including *Acidithiobacillus*, *Hydrogenobaculum*, *Metallosphaera*, *Sulfolobus*, and *Thiomonas* (Fig. 2). Also, physiological studies indicated that *Thermoproteus tenax* [30], *Sulfolobus tokodaii* [31], *Acidithiobacillus* [32], and *Metallosphaera* [33] were also capable of utilizing hydrogen as an energy source.

Although carbon assimilation metabolism and hydrogen metabolism (Additional file 1: Table S2) were regarded as important, they were not the only advantageous characteristics enabling the genus *Hydrogenobaculum* to dominate in SHP. Given substantial environmental variations among various hot springs, *Hydrogenobaculum* bacteria seemed to adapt to a broader temperature range compared to other detected genera; this could be another characteristic

contributing to their dominance in variable acidic hot springs, with temperatures ranging from 50 to 82 °C [6, 15, 18, 25, 26, 34]. In SHP, water temperature ranged from 50 to 85 °C in a-year-long survey [21], similar to the temperature range in other *Hydrogenobaculum*-dominated hot springs. On the contrary, two other relatively less well represented genera identified in SHP, e.g. *Acidithiobacillus* and *Thiomonas*, were reported to only grow under mild thermophilic conditions. For example, temperature ranges of *A. caldus*, *A. ferrooxidans*, *Thiomonas arsenitoxydans*, and *Thiomonas intermedia*, were 32 ~ 52 °C [35], 10 ~ 37 °C [36], 30 °C [37], and 30 ~ 35 °C [37], respectively (Additional file 1: Table S3). Whether those bacterial strains have evolved additional heat tolerance mechanisms is apparently unknown.

Low oxygen concentrations in SHP water could also have affected microbial dominance, as they might not have been favorable for aerobic carbon assimilators, e.g. *Sulfolobus* and *Metallosphaera* [31, 38–40]. However, the oxygen requirement of *Hydrogenobaculum* Y04AAS1-related strain has apparently not been reported. Regardless, Y04AAS1-related strain seemed well adapted to anaerobic or microaerobic conditions, due to the presence of oxygen-sensitive pyruvate synthase and phosphoenolpyruvate carboxylase, which catalyze carboxylation steps in the reductive citrate cycle [41]. In addition, based on previous metagenomic

Lin *et al. BMC Genomics* (2015) 16:1029

Page 5 of 16

Module	Function	Ez/ gene	hya	hys	vdi	vmo	tuz	ttn	clg	sto	sso	sai	cma	acu	afe	tin	thi	msc
Reductive citric acid cycle (Arnon-Buchanan cycle)	Acetyl-CoA + CO_2 → Pyruvate	1.2.7.1																
		1.2.7.-																
	Pyruate → Oxoloactate	6.4.1.1																
	Pyruvate → Phosphoenolpyruvate	2.7.9.1																
		2.7.9.2																
	Phosphoenolpyruvate + CO_2 → Oxoloactate	4.1.1.31																
	Oxoloactate → Malate	1.1.1.37																
	Malate ←→ Fumarate	4.2.1.2																
	Furmarate ←→ Succinate	1.3.99.1																
	Succinate → Succinyl-CoA	6.2.1.5																
	Succinyl-CoA + CO_2 → 2-Oxoglutarate	*1.2.7.3																
	2-Oxoglutarate + CO_2 → Isocitrate	1.1.1.42																
	Isocitrate → cis-Aconitate	4.2.1.3																
	cis-Aconitate → Citrate	4.2.1.3																
	Citrate → Oxoloactate + Acetyl-CoA	2.3.3.8																
	Citrate → Citryl-CoA	*CCS																
	Citryl-CoA → Acetyl-CoA	*CCL																
Hydroxypropionate-hydroxybutyrate cycle	Acetyl-CoA + HCO_3^- → Malonyl-CoA	6.4.1.2																
	Malonyl-CoA → Malonate Semialdehyde	*1.2.1.75																
	Malonate Semialdehyde → 3-Hydroxypropanoate	*1.1.1.298																
	3-Hydropropanoate → 3-Hydroxypropionyl-CoA	*6.2.1.36																
	3-Hydroxypropionyl-CoA → Acryloyl-CoA	*4.2.1.116																
	Acryloyl-CoA → Propionyl-CoA	*1.3.1.84																
	Propionyl-CoA + HCO_3^- → (S)Methylmalonyl-CoA	*6.4.1.3																
	(S)Methylmalonyl-CoA → (R)Methylmalonyl-CoA	5.1.99.1																
	(R)Methylmalonyl-CoA → Succinyl-CoA	*5.4.99.2																
	Succinyl-CoA → Succinate Semialdehyde	1.2.1.76																
	Succinate Semialdehyde → 4-Hydroxybutyric acid	1.1.1.-																
	4-Hydroxybutyric acid → 4-Hydroxybutyl-CoA	6.2.1.-																
	4-Hydroxybutyl-CoA → Vinylacetyl-CoA	*4.2.1.120																
	Vinylacetyl-CoA ←→ Crotonoyl-CoA	*5.3.3.3																
	Crotonoyl-CoA → (S)-3-Hydroxybutanoyl-CoA	4.2.1.17																
	(S)-3-Hydroxybutanoyl-CoA → Acetoacetyl-CoA	1.1.1.35																
	Acetoacyl-CoA → Acetyl-CoA	2.3.1.9																
Dicarboxylate Hydroxybutyrate Cycle	Acetyl-CoA + HCO_3^- → Malonyl-CoA	6.4.1.2																
	Malonyl-CoA → Malonate Semialdehyde	1.2.1.75																
	Malonate Semialdehyde → 3-Hydroxypropanoate	1.1.1.298																
	3-Hydropropanoate → 3-Hydroxypropionyl-CoA	6.2.1.36																
	3-Hydroxypropionyl-CoA → Acryloyl-CoA	4.2.1.116																
	Acryloyl-CoA → Propionyl-CoA	1.3.1.84																
	Propionyl-CoA + HCO_3^- → (S)Methylmalonyl-CoA	6.4.1.3																
	(S)Methylmalonyl-CoA → (R)Methylmalonyl-CoA	5.1.99.1																
	(R)Methylmalonyl-CoA → Succinyl-CoA	5.4.99.1																
	Succinyl-CoA → Succinate Semialdehyde	1.2.1.76																
	Succinate Semialdehyde → 4-Hydroxybutyric acid	1.1.1.-																
	4-Hydroxybutyric acid → 4-Hydroxybutyl-CoA	6.2.1.-																
	4-Hydroxybutyl-CoA → Vinylacetyl-CoA	*4.2.1.120																
	Vinylacetyl-CoA ←→ Crotonoyl-CoA	*5.3.3.3																
	Crotonoyl-CoA → (S)-3-Hydroxybutanoyl-CoA	4.2.1.17																
	(S)-3-Hydroxybutanoyl-CoA → Acetoacetyl-CoA	1.1.1.35																
	Acetoacyl-CoA → Acetyl-CoA	2.3.1.9																
	Acetyl-CoA + CO_2 → Pyruvate	1.2.7.1																
		1.2.7.-																
	Pyruvate → Phosphoenolpyruvate	2.7.9.2																
	Phosphoenolpyruvate + CO_2 → Oxoloactate	4.1.1.31																
	Oxoloactate → Malate	1.1.1.37																
	Malate ←→ Fumarate	4.2.1.2																
	Furmarate ←→ Succinate	1.3.99.1																
	Succinate → Succinyl-CoA	6.2.1.5																
Carbon fixation pathway in prokeryotes (Photosynthesis)	CO_2 + H_2O +Ribulose-1,5P2 → Glycerate-3P	*4.1.1.39																
	Glycerate-3P → 1,3-Bisphosphoglycerate	2.7.2.3																
	1,3-Bisphosphoglycerate → Glyceraldehyde-3P	1.2.1.13																
		1.2.1.59																
		1.2.1.12																
	Glyceraldehyde-3P → D-Fructose 1,6P2	4.1.2.13																
	D-Fructose 1,6s → D-Fructose 6P	3.1.3.11																
	D-Fructose 6P + Xylulose-5P → Erythrose-4P	2.2.1.1																
		4.1.2.22																
	Erythrose-4P + Glycerone-P → Sedoheptulose 1,7-bisphosphate	4.1.2.13																
	Sedoheptulose 1,7-bisphosphate → Sedoheptulose-7P	3.1.3.37																
	Sedoheptulose ←→ Sedoheptulose-7P	2.7.1.14																
	Sedoheptulose-7P + Glyceraldehyde-3P→Ribose-5P + Xylulose-5P	2.2.1.1																
	Ribose-5P → Ribulose-5P	5.3.1.6																
	Ribulose-5P → Ribulose-1,5Ps	*2.7.1.19																
	Glyceraldehyde-3P → Glycerone-P	5.3.1.1																
	Xylulose-5P → Ribuose-5P	5.1.3.1																
	Xylulose → Glyceraldehyde-3P	4.1.2.9																

Module	Function	Ez/ gene	hya	hys	vdi	vmo	tuz	ttn	clg	sto	sso	sai	cma	acu	afe	tin	thi	msc
3-Hydroxyropionate Bicycle	Acetyl-CoA + HCO_3^- → Malonyl-CoA	6.4.1.2																
	Malonyl-CoA → Malonate Semialdehyde	*1.2.1.75																
	Malonate Semialdehyde → 3-Hydroxypropanoate	1.1.1.298																
	3-Hydropropanoate → 3-Hydroxypropionyl-CoA	*6.2.1.36																
	3-Hydroxypropionyl-CoA → Acryloyl-CoA	4.2.1.116																
	Acryloyl-CoA → Propionyl-CoA	1.3.1.84																
	Propionyl-CoA + HCO_3^- → (S)Methylmalonyl-CoA	6.4.1.3																
	(S)Methylmalonyl-CoA → (R)Methylmalonyl-CoA	5.1.99.1																
	(R)Methylmalonyl-CoA → Succinyl-CoA	5.4.99.1																
	Succinate ←→ Succinyl-CoA	2.8.3.22																
	Succinate ←→ Furmarate	1.3.99.1																
	Fumarate ←→ Malate	4.2.1.2																
	Malate → L-Malyl-CoA	2.8.3.22																
	L-Malyl-CoA → Acetyl-CoA + Glyoxylate	*4.1.3.24																
Wood-Ljungdhl Pathway (Reductive Acetyl-CoA pathway)	CO_2 → CO	1.2.99.2																
		1.2.7.4																
	CO_2 → Formate	1.2.1.43																
	Formate + Tetrathiofolate → 10-Formyl-THF	6.3.4.3																
	10-Formaly-THF → 5,10-Methenyl-THF	3.5.4.9																
	5,10-Methenyl-THF → 5,10-Methene-THF	1.5.1.5																
	5,10-Methene-THF → 5-Methenylhydrotetrafolate	1.5.1.20																
	5-Methenylhydrotetrefolate → Tetrahydrofloate	2.1.1.258																
	5-Methenyltetrefolate + CO → Tetrahydrofolate + Acetyl-CoA	*2.3.1.69																

Fig. 2 (See legend on next page.)

Lin *et al. BMC Genomics* (2015) 16:1029

Page 6 of 16

(See figure on previous page.)
Fig. 2 Carbon-metabolizing enzymes identified from dominant microbes using KEGG mapping. Asterisk: key enzymes in metabolic pathways. Abbreviations: hya, *Hydrogenobaculum* sp. Y04AAS1; hys, *Hydrogenobaculum* sp. SN; vdi, *Vulcanisaeta distributa*; vmo, *Vulcanisaeta moutnovskia*; tuz, *Thermoproteus uzoniensis*; ttn, *Thermoproteus tenax*; clg, *Caldisphaera lagunensis*; sto, *Sulfolobus tokodaii*; sso, *Sulfolobus solfataricus*; sai, *Sulfolobus acidocaldarius*; cma, *Caldivirga maquilingensis*; acu, *Acidithiobacillus caldus*; afe, *Acidithiobacillus ferrooxidans*; tin, *Thiomonas arsenitoxydan*; thi, *Thiomonas intermedia*; mse, *Metallosphaera sedula*. Cell in white, not listed in the KEGG reference pathway; grey, listed in the KEGG reference pathway; black, listed in the KEGG reference pathway and identified in our metagenomic dataset. Carbon-metabolizing enzymes identified from dominant microbes using KEGG mapping

studies, *Hydrogenobaculum* bacteria dominated in two acidic hot springs with radically different dissolved oxygen concentrations (>3 and 22 μM in Dragon Spring and One Hundred Spring, respectively; [25, 26]), suggesting substantial physiological flexibility of *Hydrogenobaculum* bacteria to variations in oxygen concentration. In short, dominance of *Hydrogenobaculum* bacteria in SHP was attributed to their inherent adaptability to withstand fluctuations in both temperature and oxygen concentration, as well as their metabolic capacity to assimilate carbon or use hydrogen as an energy source.

Genomic map of *Hydrogenobaculum* bacteria
Hydrogenobaculum was the predominant genus in SHP. Mapping DSS reads covered >90 % of the length of the *Hydrogenobaculum* sp. Y04SSA1 reference genome (Fig. 3, Additional file 1: Table S4), consistent with analysis of genomic information-rich genera, which designated *Hydrogenobaculum* Y04SSA1-related strain as the dominant microbe (Additional file 1: Table S3). In addition, 16S rRNA genes and key carbon metabolic gene *ccl*, which encodes citryl-CoA lyase, in the genome of Y04SSA1-related strain were also on the genome map (Fig. 3).

Comparison among acidic hot spring metagenomes
Cheng et al. reported that *Hydrogenobaculum* was a major genus in an acidic hot spring (Huang-Shan: 82.9 °C, pH 2.2) in TVG, based on amplification and analysis of full-length 16S rRNA genes [24]. *Hydrogenobaculum* was also designated the major genus in acidic hot springs in Yellowstone National Park [25, 26]. Comparative metagenomics characterize interactions between microbes and their environment. Currently, there are only two published acidic hot spring metagenome datasets [17, 18], one from Yellowstone National Park and the other from El Coquito spring, National Natural Park Los Nevados. Functional profiles (based on KEGG or COGs) of the SHP metagenome were compared to metagenomes of Yellowstone National Park and National Natural Park Los Nevados (for the latter, see Fig. 4 and Additional file 1: Figure S2). The SHP metagenome was closer to the metagenome from Yellowstone National Park than to National Natural Park Los Nevados. The four major pathways of the COGs category that differed between SHP/Yellowstone National Park and National Natural Park Los Nevados were: (a)

amino acid transport and metabolism; (b) nucleotide transport and metabolism; (c) replication, recombination and repair; and (d) general function prediction (Additional file 1: Fig. S2).

Environmental conditions shape microbial community structure, which would in turn affect functional profiles. At SHP and YNP, conditions were: temperatures >50 °C, pH approximately 2–3, concentrations of sulfur-related compounds were high, and major microbial genera were *Hydrogenobaculum*, *Sulfolobus* and *Metallosphaera*. That these two hot springs were on distant continents and derived by distinct geological events, we concluded that microbial communities in acidic hot springs have undergone persistent and common selection, characterized by phenotypic conservation (Additional file 1: Fig. S2).

Diverse microenvironments in SHP implied by microbial composition
To further elucidate interactions between microbes and environmental factors in SHP, we critically reviewed previous reports of dominant microbes in SHP. Analyzing microbial community structures and metagenomes contribute to understanding geochemical conditions in acidic hot springs [42]. Dominant microbes in SHP microbial community had diverse oxygen preferences, including aerobic microbes (e.g., *Sulfolobus* and *Metallosphaera*), facultative aerobic microbes (e.g., *Acidithiobacillus*), microaerobic microbes (e.g., *Vulcanisaeta* and *Caldvirga*), and anaerobic microbes (e.g., *Thiomonas* and *Caldisphaera*). We inferred that the water environment of the hot spring had at least three distinct microhabitats, namely aerobic, microaerobic and anaerobic (Fig. 5 and Additional file 1: Table S3). Furthermore, the lowest reported oxygen condition in SHP (2.74 mg/L; [43]) indirectly supported the presence of habitats with varying oxygen concentrations.

Although metagenomic information in the present study clearly supported the presence of microaerobic or anaerobic microenvironments in SHP water, potential sources of error could not be excluded. For example, some microaerobic or anaerobic microbes from the sediment or the soil nearby the pond might have contaminated our sample. However, that pond water was clear and calm during sampling, and sampling was carefully conducted, the probability that contamination occurred was extremely low.
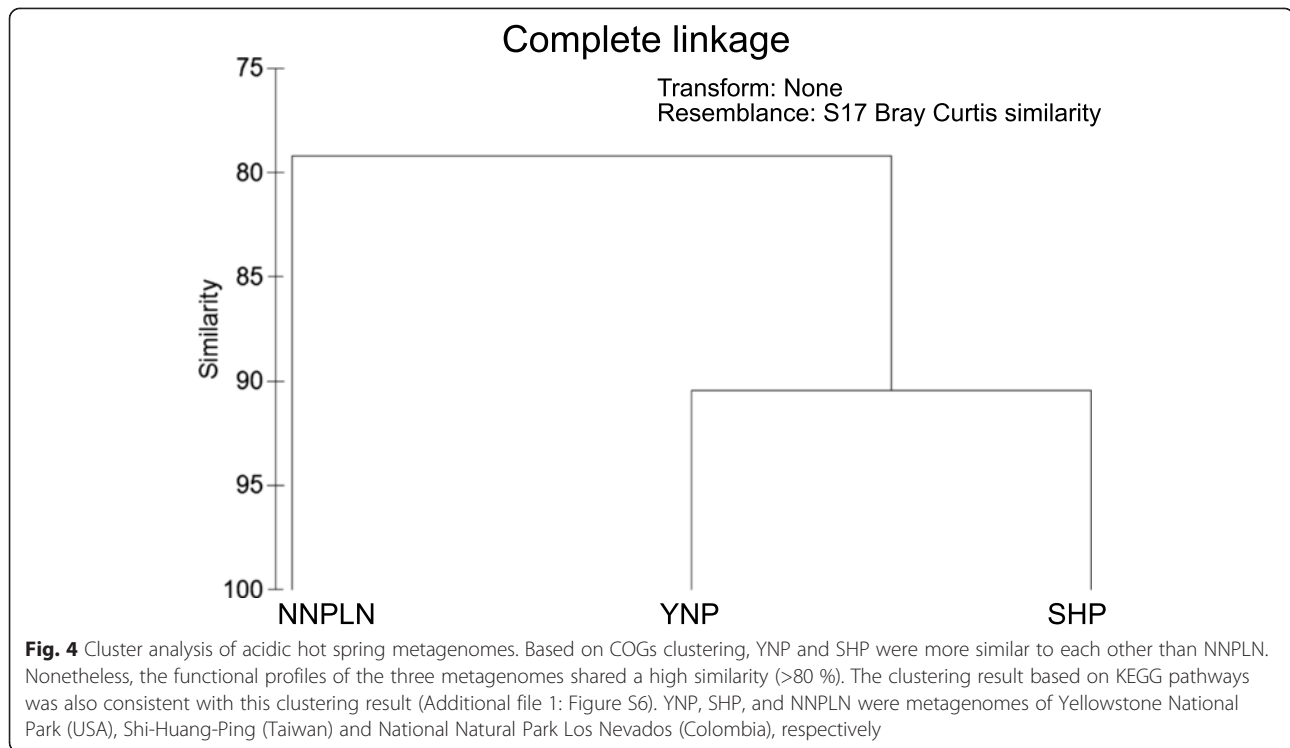
Lin *et al. BMC Genomics* (2015) 16:1029

Page 7 of 16

Fig. 3 Mapping both fosmid and DSS contigs and raw reads to *Hydrogenobaculum* sp. Y04AAS1 reference genome. From outer to inter circles are: DSS raw reads (*dark gray*), DSS contigs (*orange*), fosmid raw reads (*light gray*), fosmid contigs (*green*) and specific highlighted genes, respectively

Carbon cycle in the SHP

The carbon cycle in the acidic hot spring is highly dependent upon chemotrophic processing, as the combination of high temperature and low pH hamper photosynthesis [19, 44]. The upper limit for photosynthesis is ~ 56 °C in an acidic (pH <4.0) environment [44]. Thus, microbes detected in the springs of Cater Hills and Horris Greyser Basin (USA) presumably used non-photosynthetic chemotrophic pathways for carbon assimilation [16].

Four chemotrophic carbon assimilation pathways were identified in our metagenomic data. *Hydrogenobaculum* had a reductive citrate cycle, *Sulfolobus* and *Metallosphaera* used the hydroxypropionate-hydroxybutyrate cycle [45, 46], and *T. uzoniensis* and *T. tenax* had genes for both a reductive citrate cycle and a dicarboxylate-hydroxybutyrate cycle [47, 48]. We identified two chemosynthesis-based carbon assimilation pathways, including a reductive citrate cycle in genus *Hydrogenobaculum*, and a hydroxypropionate-hydroxybutyrate cycle in genus *Sulfolobus* and genus *Metallosphaera* (Fig. 2 and Additional file 1: Table S5).

Two key genes in the reductive citrate cycle, *korA* (EC 1.2.7.3) and *korB* (EC 1.2.7.3), encoding the α and β subunits

Lin *et al. BMC Genomics* (2015) 16:1029

Page 8 of 16

Complete linkage

Transform: None
Resemblance: S17 Bray Curtis similarity

Fig. 4 Cluster analysis of acidic hot spring metagenomes. Based on COGs clustering, YNP and SHP were more similar to each other than NNPLN. Nonetheless, the functional profiles of the three metagenomes shared a high similarity (>80 %). The clustering result based on KEGG pathways was also consistent with this clustering result (Additional file 1: Figure S6). YNP, SHP, and NNPLN were metagenomes of Yellowstone National Park (USA), Shi-Huang-Ping (Taiwan) and National Natural Park Los Nevados (Colombia), respectively

of critical enzyme 2-oxoglutarate ferredoxin oxidoreductase, were identified in the metagenome data and assigned to *Hydrogenobaculum* and *T. uzoniensis*. Furthermore, based on metagenomic data, *Hydrogenobaculum* bacteria had a gene encoding citryl-CoA lyase (*ccl*), an enzyme capable of catalyzing a biochemical reaction similar to another essential enzyme, ATP-citrate lyase, in the reductive citrate cycle [49].

In addition to the reductive citrate cycle, both *T. uzoniensis* and *T. tenax* had the key enzyme 4-hydroxybutyryl-CoA dehydratase (4.2.1.120 and 5.3.3.3) for dicarboxylate-hydroxybutyrate, and the key enzyme 2-oxoglutarate synthase (KorA and KorB, EC 1.7.3.2) of the reductive citrate cycle.

Sulfolobus and *Metallosphaera* bacterial rely on the hydroxypropionate-hydroxybutyrate cycle to convert carbon dioxide into organic carbons. Genes encoding key enzymes in this pathway, including acetyl-CoA/propionyl-CoA carboxylase (EC 6.4.1.2), malonyl-CoA reductase (EC 1.2.1.75 and 1.1.1.298), methylmalonyl-CoA mutase (EC 6.4.99.2) and 4-hydroxybutyryl-CoA dehydratase (EC 4.2.1.120) [41] were identified in contigs assigned to *Sulfolobus tokodaii* and *Metallosphaera* genus.

Acidithiobacillus and *Thiomonas* bacteria use the Calvin cycle to assimilate inorganic carbon [50–52]. Notably, *Acidithiobacillus* bacteria use electrons generated from sulfur metabolism for the Calvin cycle [50, 51], thereby circumventing temperature limitations for photosynthesis [19]. However, it remains unclear whether *Thiomonas*

bacteria could invoke a mechanism similar to *Acidithiobacillus* bacteria, enabling it to fix carbon [21] when the water temperature increased. However, the presence of *cbbSL* genes that encode the key enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (EC 4.1.1.39) in the Calvin cycle identified in the current study and a previous report (Fig. 2; [52]), provided additional evidence that *Thiomonas* bacteria can assimilate carbon.

Nitrogen cycle in SHP

The only dominant SHP microbe capable of fixing nitrogen (Fig. 6, Additional file 1: Table S5; [53]) was *A. ferrooxidans*; therefore, we inferred it played a key role in the SHP nitrogen cycle. It is noteworthy that the SHP spring is a nitrogen-limited environment (nitrate concentrations ranged from 0.9 ppm to below detection limits; Table 1; [21]). In addition, microbes living in SHP might have to obtain organic nitrogen from an alternative source (e.g. metabolizing existing nitrogen-containing compounds in the water). For example, several bacteria (*Hydrogenobaculum*, *A. ferrooxidans*, and *Thiomonas*) had nitronate monooxygenase (EC 1.13.12.16), the enzyme for transforming nitroalkane compounds (R-NO$_2$) to nitrite (Fig. 6).

Nitrite could be converted to ammonia (nitrogen reduction) and used to synthesize amino acids, or be converted into nitrogen (through denitrification) to generate energy. Two groups of bacteria, genus *Thiomonas* and *A. ferrooxidans* encoded several genes (*narG*, *narH*, *narI*,

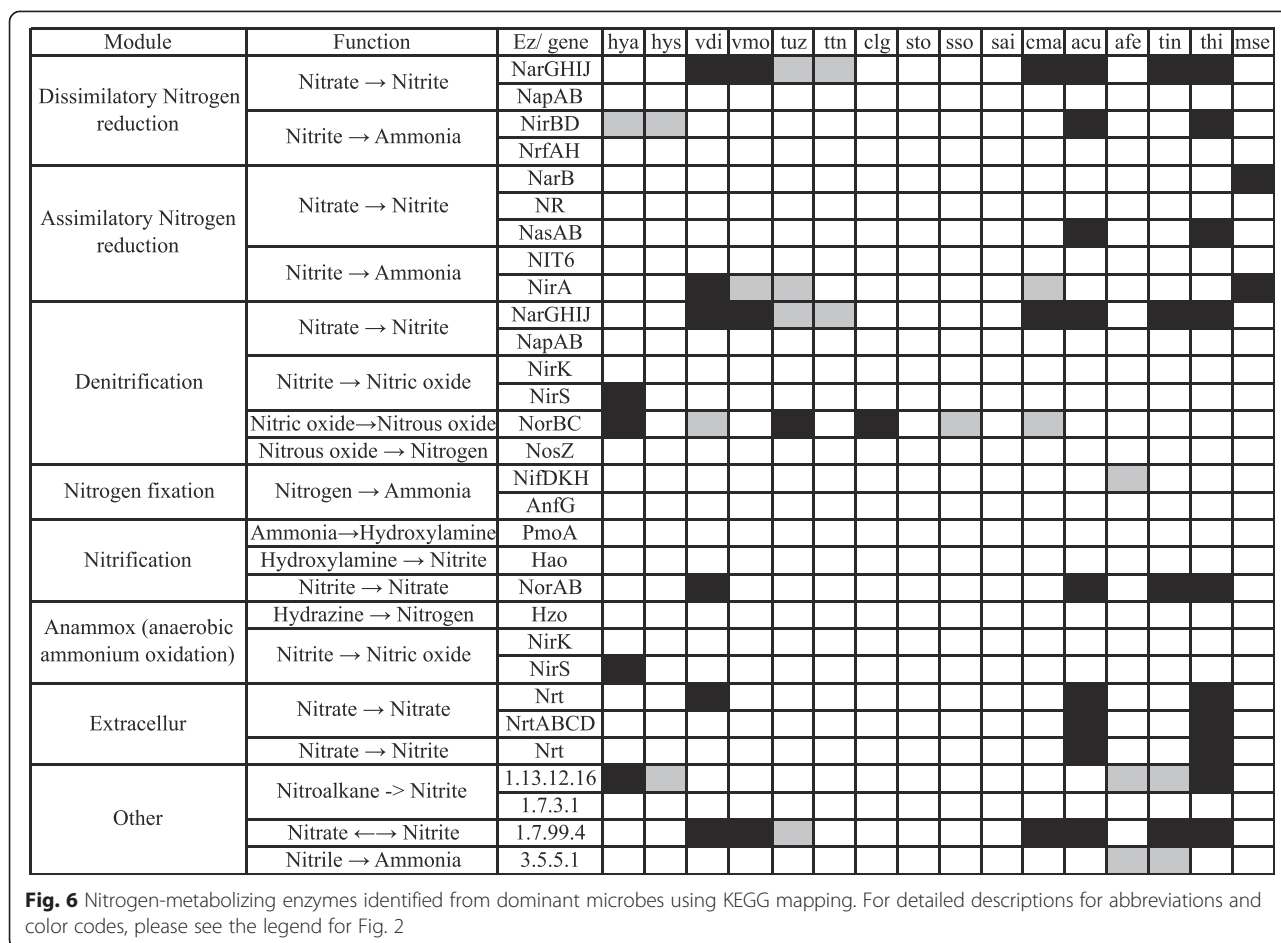Lin *et al. BMC Genomics* (2015) 16:1029

Page 9 of 16
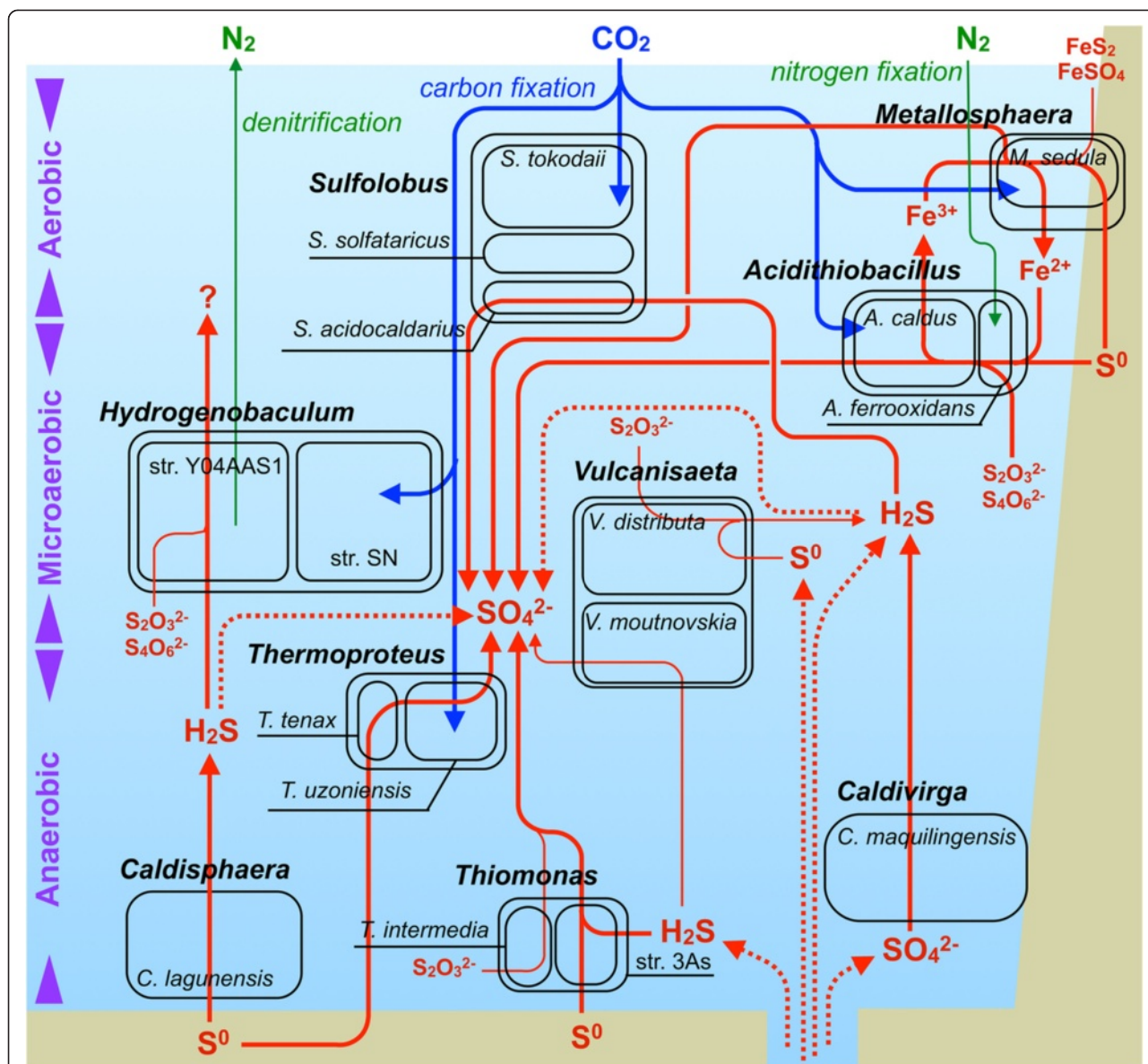
Fig. 5 Hypothetical metabolic interactions between microbes and environments in the SHP acidic hot spring. These are potential relationships between dominant microbes and biogeochemical pathways of carbon, nitrogen and sulfur in the SHP acidic hot spring. Dominant microbial genera, species or strains are in rectangles. Dotted lines show compounds potentially derived from other sources (e.g., sediments, hot spring from underground, etc.). Thick lines are metabolic potentials detected in genera, whereas thin lines highlight alternative metabolic pathways in dominant microbial genera

narJ, nirB and *nirD*) involved in dissimilatory nitrate reduction pathways. Nevertheless, according to the KEGG reference pathway, none of the dominant microbes had enzymes for a complete denitrification pathway (Fig. 6). Even though Reysenbach et al. analyzed the *Hydrogenobaculum* bacterial genome and suggested that str. Y04AAS1 genome harbored all genes required for this pathway, they did not detect reduced nitrate under experimental conditions [28]. However, that SHP has a low concentration of organic nitrogen compounds, microbes might prefer to use nitrate to synthesize building blocks

in lieu of generating energy. Clearly, further investigations are needed to elucidate the nitrogen nutrient cycle in SHP.

Sulfur metabolism

Dominant microbes were dexterous in sulfur metabolism (Fig. 7). *Vulcanisaeta* archaea, *Thermoproteus tenax* and *Caldivirga maquilingensis* had the capacity to transform trithionate into sulfite with sulfite reductase (EC 1.8.99.3). Since archaea of genus *Vulcanisaeta*, *T. tenax* and *C. maquilingensis* had all enzymes required for dissimilatory

Lin *et al. BMC Genomics* (2015) 16:1029

Page 10 of 16

Module	Function	Ez/ gene	hya	hys	vdi	vmo	tuz	ttn	clg	sto	sso	sai	cma	acu	afe	tin	thi	mse
Dissimilatory Nitrogen reduction	Nitrate → Nitrite	NarGHIJ			■	■	▨	▨					■	■		■	■	
		NapAB																
	Nitrite → Ammonia	NirBD	▨	▨										■		■		
		NrfAH																
Assimilatory Nitrogen reduction	Nitrate → Nitrite	NarB																■
		NR																
		NasAB												■		■		
	Nitrite → Ammonia	NIT6																
		NirA			■	■		▨					■	▨				■
Denitrification	Nitrate → Nitrite	NarGHIJ			■	■		▨					■	■		■	■	
		NapAB																
	Nitrite → Nitric oxide	NirK																
		NirS	■															
	Nitric oxide→Nitrous oxide	NorBC	■		▨		■			■		▨	▨					
	Nitrous oxide → Nitrogen	NosZ																
Nitrogen fixation	Nitrogen → Ammonia	NifDKH												▨				
		AnfG																
Nitrification	Ammonia→Hydroxylamine	PmoA																
	Hydroxylamine → Nitrite	Hao																
	Nitrite → Nitrate	NorAB			■								■	■		■		
Anammox (anaerobic ammonium oxidation)	Hydrazine → Nitrogen	Hzo																
	Nitrite → Nitric oxide	NirK																
		NirS	■															
Extracellur	Nitrate → Nitrate	Nrt			■									■			■	
		NrtABCD												■				
	Nitrate → Nitrite	Nrt																
Other	Nitroalkane -> Nitrite	1.13.12.16	■	▨											▨	▨		
		1.7.3.1																
	Nitrate ←→ Nitrite	1.7.99.4			■		▨						■			■		
	Nitrile → Ammonia	3.5.5.1												▨	▨			

Fig. 6 Nitrogen-metabolizing enzymes identified from dominant microbes using KEGG mapping. For detailed descriptions for abbreviations and color codes, please see the legend for Fig. 2

sulfate reduction, they were capable of utilizing sulfate or sulfite for energy metabolism. Furthermore, sulfite converted from trithionate could be used for dissimilatory sulfate reduction. *Thiomonas* bacteria encoded genes for complete SOX complex, enabling them to convert thiosulfate into sulfate. Thiosulfate could also be converted into sulfite via thiosulfate/3-mercaptopyruvate sulfurtransferase (EC 2.8.1.1), present in most dominant microbes in SHP. Although several dominant microbes could convert thiosulfate into tetrathiosulfate, *Hydrogenobaculum* bacteria were the only dominant microbes capable of converting either tetrathionate or trithionate into thiosulfate.

Searching against the KEGG database provided a basic understanding of the SHP sulfur cycle in the SHP. However, an extended literature search revealed additional or uncommon sulfur-related metabolic pathways, absent from the KEGG reference pathways, but identified in dominant microbes. For example, in addition to genus *Hydrogenobaculum*, based on genomic and transcriptomic analyses, we inferred that *S. tokodaii* [54], genus *Acidithiobacillus* [50, 55, 56] and genus *Metallosphaera* [54] could also convert tetrathionate to thiosulfate (Additional file 1: Figure S3, Table S5). For bioleaching microbes like genus

Acidithiobacillus and genus *Metallosphaera*, thiosulfate served as an oxidizer for Fe(II), which could be used to generate protons as a driving force for respiration [54, 57]. Polysulfide mechanism is another Fe(II) oxidizing pathway [55, 57]. Interestingly, based on the KEGG reference pathway and a literature search, almost all dominant microbes (including non-bioleaching microbes) in SHP were capable of transforming hydrogen sulfide to polysulfide (Additional file 1: Figure S3, Table S5). Regardless, *T. tenax* and *A. caldus* were the only two dominant microbes with enzymes to recycle polysulfide [48, 50] and thereby replenish the hydrogen sulfide pool, which would be beneficial for *A. caldus* during bioleaching.

Hydrogen metabolism

Genus *Hydrogenobaculum* could use hydrogen as its major energy source [29]. To explore hydrogen metabolism-related genes in our metagenomics data, we searched our DDS dataset against the NCBI database, and summarized the results (Additional file 1: Table S2). The gene encoding Ni/Fe hydrogenase, which catalyzes the reaction: $H_2 \leftrightarrow 2H^+ + 2e^-$, was identified. In addition, genes encoding Hyp, a group of

Module		Function	Ez/ gene	hya	hys	vdi	vmo	tuz	ttn	clg	sto	sso	sai	cma	acu	afe	tin	thi	mse
Assimilatory sulfate reduction		Sulfate → APS	PAPSS																
			Sat (2.7.7.4)																
			CysND																
		APS → PAPS	PAPSS																
			CysC																
		PAPS → Sulfite	CysH																
		Sulfite → sulfide	CysJI																
			Sir																
Dissimilatory sulfate reduction		Sulfate ←→ APS	Sat (2.7.7.4)																
		APS ←→ Sulfite	AprAB																
		Sulfite ←→ Sulfide	DsrAB (18.99.1)																
SOX		Thiosulfate → SoxY Z-S-S-SO₃	SoxA																
			SoxX																
		SoxY Z-S-S-SO₃ → Sulfate + SoxY Z-S-SH	SoxB																
		SoxY Z-S-SH → SoxY Z-S-SO₃	SoxC																
			SoxD																
		SoxY Z-S-SO₃ → SoxY Z-SH	SoxB																
		Important components	SoxY																
			SoxZ																
Other sulfur related metabolism pathways		Tetrathionate → Thiosulfate	Ttr																
		Trithionate → thiosulfate	3.12.1.1																
		Trithionate → sulfite	1.8.99.3																
		Thiosulfate → Tetrathionate	1.8.5.2																
		Thiosulfate → Sulfite	2.8.1.1																
		Thiosulfate → Sulfite + Sulfide	2.8.1.5																
		Thiosulfate → Sulfide	Phs																
		Sulfide → (Sulfide)ₙ₋₁	1.8.5.-																
		(Sulfide)ₙ → (Sulfide)ₙ₋₁ + Sulfide	1.12.98.4																
		Sulfide → Sulfur	1.13.11.18																
			1.8.2.3																
		Sulfur → Sulfide	1.13.11.55																
			Sre																

Fig. 7 Sulfur-metabolizing enzymes identified from dominant microbes using KEGG mapping. For detailed descriptions for abbreviations and color codes, please see the legend for Fig. 2

proteins required during maturation of Ni/Fe hydrogenase [58], were also present in our DSS dataset.

Microbial interactions in acidic hot springs

In addition to several potential metabolic interactions, 16S rRNA gene-based diversity and CRISPR arrays also revealed microbe-microbe interactions. In that regard, the presence of genus *Nanoarchaea* and numerous viral sequences/CRISPR arrays were consistent with robust microbial interactions in the SHP.

Genus *Nanoarchaea* (represented by *Nanoarchaea*-like 16S rRNA gene sequences) was a dominant genus in SHP (Fig. 1 and Table 2). There were apparently no previous reports of genus *Nanoarchaea* in an acidic thermal environment with a low NaCl concentration. The sole species of genus *Nanoarchaea* (*Nanoarchaeum equitans*) previously reported had a much-reduced genome and could only be grown in the presence of *Ignicoccus sp.*, an archaeal genus [59]. Furthermore, that an *Ignicoccus*-like 16S rRNA gene sequence was also detected in the present survey (highlighted in green in Additional file 1: Table S1), suggested a potential host-parasite interaction between *Nanoarchaea* and *Ignicoccus*.

It is well known that CRISPR is an antiviral defense system common in microbial genomes [60, 61]. Furthermore, repeat sequences and spacers in CRISPR assays can be used to assign taxa, as they are strain-specific [62]. In the SHP metagenome, 1711 CRISPR-like arrays (comprising 15130 spacers) were identified, of which 123 were assigned to specific microbes (based on their unique repeat sequences; Additional file 1: Table S6). In addition, there were several kinds of viral DNA sequences in the SHP metagenome (Additional file 1: Table S7), providing evidence of viral infection.

Spacer sequences of the CRISPR array could be used to characterize microbial evolution. Six of the CRISPR-like arrays identified from DSS dataset were assigned to *Metallosphaera sedula* based on their repeat sequence. The *M. sedula* reference genome contained four CRISPR arrays, each with a unique repeat sequence. Six CRISPR-like arrays were compared to known CRISPR arrays in the *M. sedula* reference genome [63]; two of the CRISPR-like arrays had identical repeat sequences with that of the longest CRISPR array (161 spacers) from the *M. sedula* reference genome. Furthermore, there were 65 identical spacers identified by comparing spacer

Lin *et al. BMC Genomics* (2015) 16:1029

Page 12 of 16

sequences of those two arrays to the reference array (Fig. 8). More importantly, identical spacers were arranged in the same order as the reference. Since spacers are added to a CRISPR array in a chronological order [64], with 65 identical spacers in the reference genome on the 3′-end, we inferred that the two *M. sedula* populations, the reference strain isolated in Italy, and another identified by analyzing metagenomic data from SHP in this study, were both derived from the same ancestral population (with a common infection history). The *M. sedula* type strain was isolated from a hot water pond at Pisciarelli Solfatara, Italty. Multiple water samples were collected for microbial isolation, water pH was ~2 and temperature ranged from 25 to 52 °C [40] (cooler than SHP).

Conclusions

Using a metagenomic approach to acquire copious sequence data from members of SHP planktonic microbial

Fig. 8 Alignment of CRISPR arrays with *M. sedula* reference arrays. Numbers of spacer are denoted. According to the CRISPR database, *M. sedula* has four CRISPR arrays in its genome. The repeat sequence of Array 647 and 646 (detected in the SHP metagenome by BLAST) matched repeat sequences of the longest CRISPR array (CRISPR ID: NC_009440_4, refer as "reference array" hereafter) in the *M. sedula* DSM 5348 reference genome. Since the order of spacers could be associated with time of virus infection, older spacers were located near the "ancient" end of the array. With a comparative analysis of the spacer sequences, all 27 spacers in Array 647 perfectly matched the array in the reference genome (with regards to sequence similarity and orientation). The 27 spacers were located at the end of the reference array. Array 646 contained 42 spacers; the last 38 spacers perfectly matched the 83th to 121st spacers in the reference array. Sequences of the first four spacers in Array 646 were different from all of the spacers in reference array; therefore, there was divergence of virus infection history between the *M. sedula*-related species in the SHP and *M. sedula* DSM5348. Moreover, since no spacers similar to the 1st to 89th and 122nd to 124th spacers of the reference array were detected, those spacers were designated "unknown"

community enabled us to not only identify community composition, but also to postulate potential interactions within the ecosystem. Specifically, we used metagenomic data to predict potential metabolite exchange, microbe-phage interaction (CRISPR analyses) and archaeal parasite-host interactions (*Nanoarchaea* and *Ignicoccus*) within SHP. Potential metabolite exchanges among microbes is shown (Fig. 8), based on existing physiological and biochemical studies of dominant microbes. Predicting potential metabolic pathways for carbon, sulfur, carbon and hydrogen with the NCBI and/or KEGG databases enabled us to elucidate metabolic ability of each dominant microbe. However, metabolic analyses cannot fully explain the *Hydrogenobaculum*-dominant feature. Previous studies attributed a *Hydrogenobaculum*-dominant feature based on carbon assimilation pathways and hydrogen utilization features of this genus. However, *Hydrogenobaculum* is not the only microbial genus capable of utilizing hydrogen and assimilating inorganic carbon. Thus, we proposed that *Hydrogenobaculum* bacteria dominate SHP due to additional abilities, e.g. temperature tolerance and ability to survive in an anaerobic environment. Together with our analytical results and literature mining, this study provided a comprehensive understanding of interactions within the microbial ecosystem in an acidic thermal environment.

Methods

Sample collection and preparation

For direct shotgun sequencing (DSS), water samples were collected from an SHP hot spring (25°11′43. 60″N, 121°36′8. 82″E; water depth was about 15 ~ 20 cm at sampling site) into sterile polypropylene (PP) containers on April 25th, 2012. Water temperature and pH were measured *in situ*; all environmental parameters listed in Table 1 were measured according to National Institute of Environmental Analysis (NIEA, Taiwan) or American Public Health Associate (APHA) protocols. A tangential flow system [65] equipped with hollow-fiber cartridge (pore diameter 0.2 μm; Hollow Fiber Cartridge CFP-2-E-3MA, GE Healthcare, Little Chalfont, United Kingdom) was used to reduce the volume of spring water to approximately 500 mL. Thereafter, the water sample was subjected to high-speed centrifugation ($7000 \times g$ for 30 min; Himac CR-21, Hitachi, Japan). Sampled water was stored at 4 °C before further processing. Ample supernatant was retained for re-suspension of pellets into a muddy solution that yielded a mixture of suspension particles, fine sediments and microorganisms. Thereafter, DNA was extracted with an UltraClean® Mega Soil DNA Isolation Kit, (MO BIO Laboratories, Inc., Carlsbad, CA, USA), according to the manufacturer's protocol.

For fosmid library construction, water samples were collected on December 3rd, 10th and 21st 2010, and on

Lin *et al. BMC Genomics* (2015) 16:1029

Page 13 of 16
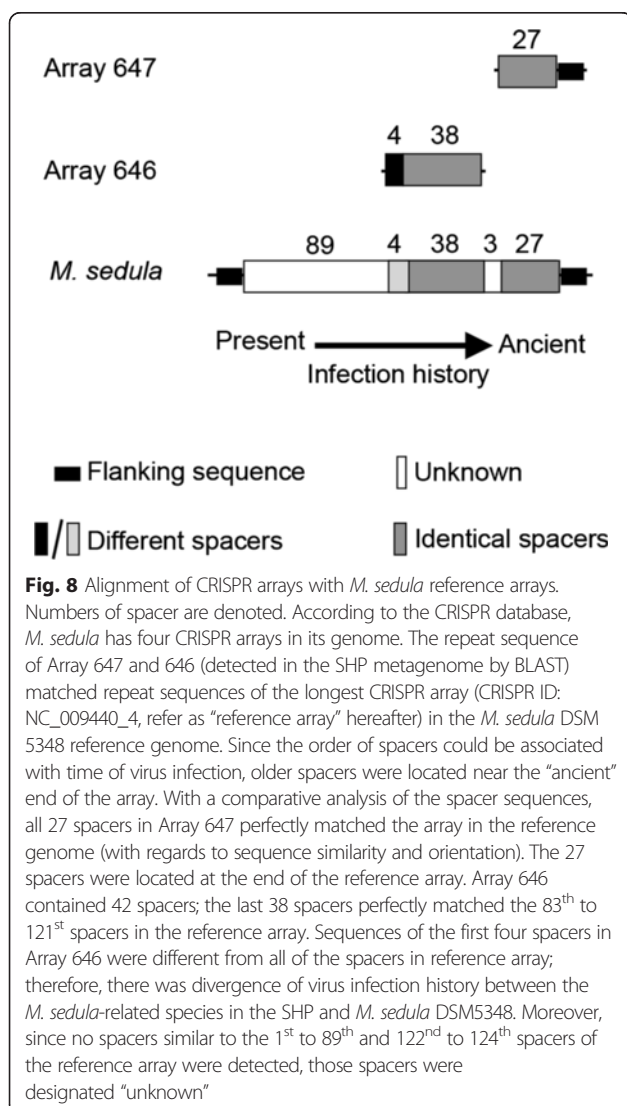
February 8th and March 17th, 2011 at the same location, and processed as described above.

Fosmid library construction

Fosmid libraries were constructed according to manufacturer's instructions (Copy Control™ HTP Fosmid Library Production Kits, Epicentre® Biotechnologies, Madison, WI, USA). Pulse-field gel electrophoresis (PFGE) was done to verify insert sizes of fosmid DNA from randomly selected colonies. Overnight cultures of selected colonies were harvested by centrifugation ($7000 \times g$ at 4 °C for 30 min; Himac CR-21, Hitachi, Japan). Fosmid DNA was extracted according to manufacturer's instructions (QIAGEN Plasmid Mini Kit, QIAGEN, Venlo, The Netherlands). Insert DNA was removed with a restriction enzyme (*Not*I) at 37 °C for 16–18 h. For PFGE, digested DNA was analyzed using 1 % agarose gel in $1/3 \times$ Loening Buffer. The PFGE system consisted of a Standard Power Pack (Rotaphor® System, Biometra, Goettingen, Germany) and a circulator tank (Refrigerated Circulator RCB411, TKS, Kaiserslautern, Germany) for cooling. After confirmation of insert size, a fosmid library MG-HSTL (9481 clones) was constructed and deposited in the Food Industry Research and Development Institute (FIRDI, Taiwan; publicly available as of Aug 1st, 2013. Website: http://www.firdi.org.tw/En_Firdi_Index.ASPX).

Fosmid DNA was extracted from randomly picked 1485 clones according to the alkaline lysis method, provided by VYM Genome Research Center, National Yang-Ming University, Taiwan. Concentration of fosmid DNA was measured using a Qubit Fluorometer (Life Technologies, Carlsbad, CA, USA) and then mixed (equal amounts of DNA from each clone) into a bulk sample for sequencing.

Random shotgun sequencing and contig assembly

Metagenomic DNA was directly extracted from the water sample and sequenced on a HiSeq™ 2000 (Illumina, San Diego, CA, USA) at Yourgene Bioscience Co., Ltd. (Taipei, Taiwan). Metagenomic DNA was sequenced separately from the Fosmid library. Raw sequencing reads were trimmed (35 bp minimum length and error probability <0.05). All DSS and fosmid contigs were assembled by MetaVelvet [66]. Raw sequence data to be assembled were ∼ 56.3 and 22.4 Gb for DSS and the fosmid library, respectively. Statistics regarding sequencing and contig assembling for DSS and fosmid are shown (Additional file 1: Tables S8 and S9, respectively).

Sequence information from DSS was used to determine composition and metabolic potential of microbial community, and reconstruct the *Hydrogenobaculum* bacterial genome. In addition, sequence information obtained from the fosmid library was used to facilitate reconstruction of the *Hydrogenobaculum* bacterial genome.

Analysis of microbial community structure

Microbial community structure in SHP was characterized by 16S rRNA gene-based diversity surveys. Qualified reads were blasted against the SILVA SSU reference database (Version 115, download date: Sep 7th, 2013) with the following criteria: a) sequence identity >95 %; b) alignment coverage >90 % of the length of the query sequence; c) E-value $<10^{-15}$; and d) highest bit-score. The taxonomic affiliation of top hit in the blast search was assigned for a single read. The relative abundance (RA_{16S}) of a specific genus was calculated by the total read number in a genus, divided by the total 16S rRNA gene sequence encoded reads (Additional file 1: Table S1). The top 20 abundant genera were selected for phylogenetic analyses (Fig. 1), using web-based software (GraPhlAn; http://huttenhower. org/galaxy, [67]).

Relative abundance of genomic information-rich genera

Each contig was assigned to a specific organism by blasting it (E-value $\leq 10^{-5}$) against the NCBI database. Taxa of the contigs were determined by annotation of the best hit. Relative abundance (RA_{contig}) of each genus was calculated as total number of qualified reads in contigs belonging to the same genus, divided by total number of qualified reads. Genera with a RA_{contig} >1 % (Additional file 1: Figure S1), were designated as genomic information-rich genera. Furthermore, to present diverse metabolic capabilities within the same genera, metagenomic information or relevant literature for major strains or species listed under genomic information-rich genera with RA_{contig} >0.2 % were also retrieved (Additional file 1: Table S3). Two lists of dominant microbial genera were generated: a) top 20, the most abundant genera (based on 16S rRNA gene-based diversity surveys); and b) genomic information-rich genera (based on taxonomic affiliations of contigs) in SHP, according to relative abundance analyses.

Mapping contigs onto *Hydrogenobaculum* bacterial genome

The majority of the DNA recovered was from the genus *Hydrogenobaculum* (see Results). Therefore, the genome of *Hydrogenobaculum sp.* Y04AAS1 [28], downloaded from Joint Genome Institute (JGI), was used as a reference genome for contig mapping. Qualified sequencing reads of DSS and fosmid were mapped to the reference genome by CLC Genomics Workbench (similarity 0.7, mapping length 0.9, website: http://www.clcbio.com), whereas DSS and fosmid contigs were mapped by MUMmer 3.0, using default settings (website: http://mummer.sourceforge.net, [68]). Mapping results were visualized with Circos [69].

Comparative metagenomics analysis

To compare putative functional profiles between the metagenome of this study and those of the other two acidic hot spring metagenomes from the Americas,

Lin *et al. BMC Genomics* (2015) 16:1029

Page 14 of 16

two metagenomic datasets, namely Yellowstone National Park (#41119) and National Natural Park Los Nevados (#4449206.3), were downloaded from the NCBI and MG-RAST servers, respectively. Thereafter, all open reading frames retrieved from the three metagenomes were compared against COGs and KEGG databases using the WebMGA service (website: http://weizhong-lab.ucsd.edu/metagenomic-analysis/) and the BBH-method service in KAAS (website: http://www.genome.jp/kegg/kaas/), respectively. Normalization to samples was done by the matched ORFs in each category, divided by the total matched ORFs for each sample. Relative abundance was presented with a line plot (non-prokaryotic functions or pathways were excluded). Furthermore, Primer 6 (website: http://www.primer-e.com/primer.htm) was used to compare putative functional profiles of three metagenomes, using a Bray-Curtis model with complete-linkage cluster [70].

Reconstruction of potential metabolic networks of dominant genera

The KEGG reference pathway mapping and blastp (E-value $<10^{-5}$ and bit-score >100) were used to identify proteins associated with dominant microbes from DSS contigs. With KEGG Mapper (http://www.genome.jp/kegg/tool/map_pathway2.html), proteins involved in carbon (Fig. 2), nitrogen (Fig. 6), or sulfur (Fig. 7) metabolic pathways were identified by mapping gi codes obtained in blastp to KEGG reference pathways (Release 72.0, October 1st 2014) for each dominant microbe. However, the KEGG database per se was regarded as insufficient as a reference for all acquired metagenomic information from an extreme microbial ecosystem [18, 25, 26]. Therefore, in addition to KEGG reference pathway mapping, information extracted from an extensive literature search was also used to predict metabolic networks and potential relationships among dominant microbes. Thereafter, relevant physiological, biochemical and genomic reports (Additional file 1: Table S3) were reviewed to provide additional information regarding carbon, sulfur and nitrogen sources (Additional file 1: Figure S3).

Identification of CRISPR-like arrays

The PILER-CR software (website: http://drive5.com/piler/) was used to identify CRISPR-like arrays in our DSS dataset [71]. Repeats and spacers of identified CRISPR-like arrays were searched against CRISPRdb (website: http://crispr.u-psud.fr/crispr/; [63]) with blastN-short, E-value $<10^{-5}$, to assign arrays to specific microbial species and to identify correlated viral sequences.

Data deposition

The microbial metagenome elucidated in this study was deposited in the NCBI Sequence Read Archive (Accession Number SRP041649).

Availability of supporting data

The microbial metagenome elucidated in this study was deposited in the NCBI Sequence Read Archive (Accession Number SRP041649).

Additional file

> **Additional file 1: Supplimentary material on metagemic analysis.** (DOCX 312 kb)

Author details
[1]Department of Life Science, National Taiwan University, Taipei 10617, Taiwan, Republic of China. [2]Division of Biostatistics & Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan Town, Miaoli County 35053, Taiwan, Republic of China. [3]Molecular Microbiology and Microbial Pathogenesis Program, Division of Biology and Biomedical Science, Washington University in St. Louis, St. Louis, MO 63130, USA. [4]Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan, Republic of China. [5]Institute of Information Science, Academia Sinica, Taipei 11529, Taiwan, Republic of China. [6]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei 10617, Taiwan, Republic of China. [7]Joint BioEnergy Institute, Emeryville, CA 94608, USA. [8]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [9]Degree Program of Genome and Systems Biology, National Taiwan University and Academia Sinica, Taipei 10617, Taiwan, Republic of China.

References
1. Woese CR. Bacterial evolution. Microbiol Rev. 1987;51:221–71.
2. Sahm K, John P, Nacke H, Wemheuer B, Grote R, Daniel R, et al. High abundance of heterotrophic prokaryotes in hydrothermal springs of the

Lin *et al. BMC Genomics* (2015) 16:1029

Page 15 of 16

Azores as revealed by a network of 16S rRNA gene-based methods. Extremophiles. 2013;17:649–62.

3. Bohorquez LC, Delgado-Serrano L, López G, Osorio-Forero C, Klepac-Ceraj V, Kolter R, et al. In-depth characterization via complementing culture-independent approaches of the microbial community in an acidic hot spring of the colombian Andes. Microbiol Ecol. 2011;63:103–15.

4. Kvist T, Ahring BK, Westermann P. Archaeal diversity in Icelandic hot springs. FEMS Microbiol Ecol. 2007;59:71–80.

5. Siering PL, Clarke JM, Wilson MS. Geochemical and biological diversity of acidic, hot springs in Lassen Volcanic National Park. Geomicrobiology J. 2006;23:129–41.

6. Wilson MS, Siering PL, White CL, Hauser ME, Bartles AN. Novel archaea and bacteria dominate stable microbial communities in North America's Largest Hot Spring. Microbiol Ecol. 2007;56:292–305.

7. Burton NP, Norris PR. Microbiology of acidic, geothermal springs of Montserrat: environmental rDNA analysis. Extremophiles. 2000;4:315–20.

8. Rzonca B, Schulze-Makuch D. Correlation between microbiological and chemical parameters of some hydrothermal springs in New Mexico, USA. J Hydro. 2003;280:272–84.

9. Huang Q, Jiang H, Briggs BR, Wang S, Hou W, Li G, et al. Archaeal and bacterial diversity in acidic to circumneutral hot springs in the Philippines. FEMS Microbiol Ecol. 2013;85:452–64.

10. Stout LM, Blake RE, Greenwood JP, Martini AM, Rose EC. Microbial diversity of boron-rich volcanic hot springs of St. Lucia, Lesser Antilles. FEMS Microbiol Ecol. 2009;70:402–12.

11. Pagaling E, Grant WD, Cowan DA, Jones BE, Ma Y, Ventosa A, et al. Bacterial and archaeal diversity in two hot spring microbial mats from the geothermal region of Tengchong, China. Extremophiles. 2012;16:607–18.

12. Hou W, Wang S, Dong H, Jiang H, Briggs BR, Peacock JP, et al. A Comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. PLoS ONE. 2013;8:e53350.

13. Song Z-Q, Wang F-P, Zhi X-Y, Chen J-Q, Zhou E-M, Liang F, et al. Bacterial and archaeal diversities in Yunnan and Tibetan hot springs, China. Environ Microbiol. 2012;15:1160–75.

14. Aditiawati P, Yohandini H, Madayanti F, Akhmaloka. Microbial diversity of acidic hot spring (Kawah Hujan B) in geothermal field of Kamojang area, West Java-Indonesia. Open Microbiol J. 2009;3:58–66.

15. Jackson CR, Langner HW, Donahoe-Christiansen J, Inskeep WP, McDermott TR. Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. Environ Microbiol. 2001;3:532–42.

16. Macur RE, Jay ZJ, Taylor WP, Kozubal MA, Kocar BD, Inskeep WP. Microbial community structure and sulfur biogeochemistry in mildly-acidic sulfidic geothermal springs in Yellowstone National Park. Geobiology. 2012;11:86–99.

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

18. Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, et al. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS ONE. 2010;5:e9773.

19. Rothschild LJ, Mancinelli RL. Life in extreme environments. Nature. 2001;409:1092–101.

20. Konstantinou KI, Lin C-H, Liang W-T. Seismicity characteristics of a potentially active Quaternary volcano: the Tatun Volcano Group, northern Taiwan. J Volcanol and Geotherm Res. 2007;160:300–18.

21. Song SR. Geological survey on the potential application of hot springs and geothermal resurce in Yangmingshan (Chinese). Taiwan: Management Office of Yangmingshan National Park, Construction and Panning Agency, Ministry of the Interior; 2005.

22. Liu C-M, Song S-R, Chen Y-L, Tsao S. Characteristics and origins of hot springs in the Tatun Volcano Group in northern Taiwan. Terr Atmos Ocean Sci. 2011;22:475–89.

23. Fournier RO. Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. Annu Rev Earth Planet Sci. 1989;17:13–53.

24. Cheng T-W, Wang P-L, Song S-R, Lin L-H. Segregated planktonic and bottom-dwelling archaeal communities in high-temperature acidic/sulfuric ponds of the Tatun Volcano Group, Northern Taiwan. Terr Atmos Ocean Sci. 2013;24:345–56.

25. Takacs-Vesbach C, Inskeep WP. Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three Aquificales lineages. Front Microbiol. 2013;84:1–25.

26. Inskeep WP, Jay ZJ, Tringe SG, Herrgård MJ, Rusch DB, Members YMPSCaWG. The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front Microbiol. 2013;4:1–15.

27. Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, et al. The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci USA. 2003;100:12984–8.

28. Reysenbach AL, Hamamura N, Podar M, Griffiths E, Ferreira S, Hochstein R, et al. Complete and draft genome sequences of six members of the aquificales. J Bacteriol. 2009;191:1992–3.

29. Spear JR, Walker JJ, McCollom TM, Pace NR. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. Proc Natl Acad Sci USA. 2005;102:2555–60.

30. Bonch-Osmolovskaya EA, Miroshnichenko ML, Kostrikina NA, Chernych NA, Zavarzin GA. *Thermoproteus uzoniensis* sp. nov., a new extremely thermophilic archaebacterium from Kamchatka continental hot springs. Arch Microbiol. 1990;154:556–9.

31. Suzuki T, Iwasaki T, Uzawa T, Hara K, Nemoto N, Kon T, et al. *Sulfolobus tokodaii* sp. nov. (f. *Sulfolobus* sp. strain 7), a new member of the genus *Sulfolobus* isolated from Beppu Hot Springs, Japan. Extremophiles. 2002;6:39–44.

32. Hedrich S, Johnson DB. Aerobic and anaerobic oxidation of hydrogen by acidophilic bacteria. FEMS Microbiol Lett. 2013;349:40–5.

33. Auernik KS, Kelly RM. Impact of molecular hydrogen on chalcopyrite bioleaching by the extremely thermoacidophilic archaeon *Metallosphaera sedula*. Appl Environ Microbiol. 2010;76:2668–72.

34. Ferrera I, Longhorn S, Banta AB, Liu Y, Preston D, Reysenbach AL. Diversity of 16S rRNA gene, ITS region and aclB gene of the Aquificales. Extremophiles. 2006;11:57–64.

35. Hallberg KB, Lindström EB. Characterization of *Thiobacillus caldus* sp. nov., a moderately thermophilic acidophile. Micriobiology. 1994;140:3451–6.

36. Kelly DP, Wood AP. Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov. and *Thermithiobacillus* gen. nov. Int J Syst Evol Microbiol. 2000;50:511–6.

37. Slyemi D, Moinier D, Brochier-Armanet C, Bonnefoy V, Johnson DB. Characteristics of a phylogenetically ambiguous, arsenic-oxidizing *Thiomonas* sp., *Thiomonas arsenitoxydans* strain 3As^{T} sp. nov. Arch Microbiol. 2011;193:439–49.

38. Chen L, Brügger K, Skovgaard M, Redder P, She Q, Torarinsson E, et al. The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. J Bacteriol. 2005;187:4992–9.

39. Jaubert C, Danioux C, Oberto J, Cortez D, Bize A, Krupovic M, et al. Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. Open Biol. 2013;3:130010.

40. Huber G, Spinnler C, Gambacorta A, Stetter KO. *Metallosphaera sedula* gen. and sp. nov. represents a new genus of aerobic, metal-mobilizing, thermoacidophilic archaebacteria. Syst Appl Microbiol. 1989;12:38–47.

41. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, et al. Autotrophic carbon fixation in archaea. Nat Rev Microbiol. 2010;8:447–60.

42. Mathur J, Bizzoco RW, Ellis DG, Lipson DA, Poole AW, Levine R, et al. Effects of abiotic factors on the phylogenetic diversity of bacterial communities in acidic thermal springs. Appl Environ Microbiol. 2007;73:2612–23.

43. Chen W-F, Menghau S. The redox potential of hot springs in Taiwan. Terr Atmos Ocean Sci. 2009;20:465–79.

44. Boyd ES, Leavitt WD, Geesey GG. CO_2 uptake and fixation by a thermoacidophilic microbial community attached to precipitated sulfur in a geothermal spring. Appl Environ Microbiol. 2009;75:4289–96.

45. Alber B, Olinger M, Rieder A, Kockelkorn D, Jobst B, Hugler M, et al. Malonyl-coenzyme A reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal *Metallosphaera* and *Sulfolobus* spp. J Bacteriol. 2006;188:8551–9.

46. Teufel R, Kung JW, Kockelkorn D, Alber BE, Fuchs G. 3-Hydroxypropionyl-coenzyme A dehydratase and acryloyl-coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in the Sulfolobales. J Bacteriol. 2009;191:4572–81.

47. Mardanov AV, Gumerov VM, Beletsky AV, Prokofeva MI, Bonch-Osmolovskaya EA, Ravin NV, et al. Complete genome sequence of the thermoacidophilic crenarchaeon *Thermoproteus uzoniensis* 768-20. J Bacteriol. 2011;193:3156–7.

Lin *et al. BMC Genomics* (2015) 16:1029

Page 16 of 16

48. Siebers B, Zaparty M, Raddatz G, Tjaden B, Albers S-V, Bell SD, et al. The complete genome sequence of *Thermoproteus tenax*: a physiologically versatile member of the Crenarchaeota. PLoS ONE. 2011;6:e24222.

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

50. Chen L, Ren Y, Lin J, Liu X, Pang X, Lin J. *Acidithiobacillus caldus* sulfur oxidation model based on transcriptome analysis between the wild type and sulfur oxygenase reductase defective mutant. PLoS ONE. 2012;7:e39470.

51. You X-Y, Guo X, Zheng H-J, Zhang M-J, Liu L-J, Zhu Y-Q, et al. Unraveling the *Acidithiobacillus caldus* complete genome and its central metabolisms for carbon assimilation. J Genet Genomics. 2011;38:243–52.

52. Duquesne K, Lieutaud A, Ratouchniak J, Muller D, Lett M-C, Bonnefoy V. Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. Environ Microbiol. 2008;10:228–37.

53. Pretorius IM, Rawlings DE, Woods DR. Identification and cloning of *Thiobacillus ferrooxidans* structural *nif* genes in *Escherichia coli*. Gene. 1986; 45:59–65.

54. Auernik KS, Kelly RM. Identification of components of electron transport chains in the extremely thermoacidophilic crenarchaeon *Metallosphaera sedula* through iron and sulfur compound oxidation transcriptomes. Appl Environ Microbiol. 2008;74:7723–32.

55. Bugaytsova Z, Lindström EB. Localization, purification and properties of a tetrathionate hydrolase from *Acidithiobacillus caldus*. Eur J Biochem. 2004;271:272–80.

56. Kanao T, Kamimura K, Sugio T. Identification of a gene encoding a tetrathionate hydrolase in *Acidithiobacillus ferrooxidans*. J Biotechnol. 2007;132:16–22.

57. Rohwerder T, Gehrke T, Kinzler K, Sand W. Bioleaching review part A. Appl Microbiol Biotechnol. 2003;63:239–48.

58. Watanabe S, Sasaki D, Tominaga T, Miki K. Structural basis of [NiFe] hydrogenase maturation by Hyp proteins. Biol Chem. 2012;393:1089–100.

59. Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature. 2002;417:63–7.

60. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol. 2005;60:174–82.

61. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 2010;11:181–90.

62. Nathan L, Bachmann NKP, Ben Zakour NL, Szubert JM, Savill J, Beatson SA. Genome analysis and CRISPR typing of Salmonella enterica serovar Virchow. BMC Genomics. 2014;15:389.

63. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics. 2007;8:172.

64. Pourcel C. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology. 2005;151:653–63.

65. Giovannoni SJ, DeLong EF, Schmidt TM, Pace NR. Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. Appl Environ Microbiol. 1990;56:2572–5.

66. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res. 2012;40:e155.

67. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ. 2015;18:e1029.

68. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.

69. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Research. 2009;19:1639–45.

70. Clarke K. Non-parametric multivariate analyses of changes in community structure. Aust J Ecol. 1993;18:117–43.

71. Edgar RC. PILER-CR: fast and accurate identification of CRISPR repeats. BMC Bioinformatics. 2007;8:18.