Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica

Craig W. Herbold[1,2], Charles K. Lee[1,2], Ian R. McDonald[1,2], S. Craig Cary[1,2,3]

Corresponding Author: Craig Cary
Department of Biological Sciences
The University of Waikato
Private Bag 3105
Hamilton, New Zealand
Telephone: +64 7 838 4543
Fax: +64 7 838 4324
Email: caryc@waikato.ac.nz

Author affiliations:
1) Department of Biological Sciences, University of Waikato, Hamilton, New Zealand
2) International Centre for Terrestrial Antarctic Research, University of Waikato, Hamilton, New Zealand
3) College of Marine and Earth Studies, University of Delaware, Lewes, DE, USA.

ABSTRACT

New evidence in aerobiology challenges the assumption that geographical isolation is an effective barrier to microbial transport. However, given the uncertainty with which aerobiological organisms are recruited into existing communities, the ultimate impact of microbial dispersal is difficult to assess. To evaluate the ecological significance of global-scale microbial dispersal, molecular genetic approaches were used to examine microbial communities inhabiting fumarolic soils on Mt. Erebus, the southernmost geothermal site on Earth. There, hot, fumarolic soils provide an effective environmental filter to test the viability of organisms that have been distributed via aeolian transport over geological time. We find that cosmopolitan thermophiles dominate the surface, whereas endemic Archaea and members of poorly understood Bacterial candidate divisions dominate the immediate subsurface. These results imply that aeolian processes readily disperse viable organisms globally, where they are incorporated into pre-existing complex communities of endemic and cosmopolitan taxa.

INTRODUCTION

Aeolian transport of microbes is the primary assumption inherent to Louren Baas Becking's hypothesis that "everything is everywhere, but the environment selects"[1]. Aerosolized microbes may originate from agricultural fields [2], sewage treatment centers [3], geothermal springs [4], any surface exposed to sufficient wind force [5], and large-scale volcanic eruptions[6]. Organisms that have been independently aerosolized or attached to dust particles can be transported thousands of kilometers, which presumably allows microbes to easily move between continents and hemispheres [7-11].

The study of aerobiology is making great strides in showing that viable cells are attached to particles that are moved atmospherically [11,12]. However, these findings support only one of the two requirements for aeolian dispersal. While survival of cells on aeolian particles is necessary, successful dispersal of microorganisms also requires successful colonization of new habitats[13]. The presence of viable cells on actively transported particulates is therefore not sufficient to show that aeolian transport is capable of microbial dispersal.

An ideal site for examining global-scale microbial dispersal would be one that is geographically isolated from similar sites and has habitat that is highly selective for viable organisms[13]. Mt. Erebus, located on Ross Island in Victoria Land, Antarctica, fits these criteria extremely well [14]. It is the southernmost active volcano on Earth and possesses unique high-elevation geothermal features, such as ice caves and ice-free

63 fumarolic ground that are located near its 3794m summit [15]. The geothermal features
64 on Mt. Erebus are separated from similar features on distant volcanoes by hundreds of
65 kilometers of snow and ice, ensuring that locally-sourced microbes are psychrophilic.
66 The warm environments that are created by geothermal activity actively select against
67 local psychrophiles, and encourage the growth of viable thermophiles and mesophiles
68 that are endemic or have been sourced from distant features.

69 The Tramway Ridge Antarctic Specially Protected Area (ASPA) (Figure 1a-d), located
70 approximately 1.5 km NW of the main crater of Erebus at an elevation between 3350
71 and 3400 m, is an extensive warm fumarolic area protected by international treaty as a
72 site of particular biological interest (ASPA 130 Management Plan). It is located at the
73 terminus of a lava flow that is approximately 10,000 years old and is composed primarily
74 of phonolite, a fine-grained volcanic rock of alkali feldspars and nepheline [16]. A loose
75 layer, as deep as 10 cm, of highly altered mineral soils lies atop the phonolitic base[15].
76 Within this ice-free area, small fumaroles emit steam and CO_2, maintaining year-round
77 average surface temperatures of 60-65°C even during the coldest part of winter when
78 air temperatures drop below -55°C [17]. Subsurface concentrations of CO_2 and gas efflux
79 rates vary considerably, even between fumaroles in close proximity[18]. Fumaroles are
80 characterized by a neutral to mildly alkaline pH and are surrounded by steep lateral
81 decreasing pH and temperature gradients that support unique assemblages of mosses
82 and cyanobacterial mats[7,17].

A preliminary molecular community analysis of shallow (~2-4 cm depth) lateral transects showed that bacterial community structure varies significantly across much of the cool (<60°C), low-pH (<7) expanse of Tramway Ridge; however, hot, neutral-pH steaming fumaroles share bacterial communities that are similar to one another [17]. The study showed that the fumaroles themselves are likely dominated by endemic taxa, including deep-branching Planctomycetes, Acidobacteria and Chloroflexi, as well as many novel, potentially division-level lineages with no known relatives [17]. The few Archaea observed at Tramway Ridge were all classified as Crenarchaeota and were most similar to environmental clones from subsurface environments in South Africa [19] and central Europe [20].

The aim of the current study is to examine the relevance of aeolian transport in the assembly of the microbial community inhabiting the fumarolic soils of Tramway Ridge. We characterize the composition of microbial communities with respect to depth within fumaroles at Tramway Ridge, determine which taxa are dominant, and identify physicochemical factors that might structure communities within fumaroles. Vertical profiles were collected at two 65°C fumaroles: one fumarole that vigorously emitted CO_2 and steam (active site, Figure 1e) and another that was visibly less active (passive site, Figure 1f). Amplicon pyrosequencing is used to assess community composition and structure within the soil profile at each fumarole, while shotgun DNA sequencing enables the reconstruction of full-length 16S rRNA genes that are used for high-resolution phylogenetic analysis. This survey reveals the co-occurrence of two distinct

104 microbial communities, a surface community that is dominated by globally distributed
105 populations and a subsurface community unique to Mt. Erebus, demonstrating the
106 importance of large-scale aeolian transport in the assembly of complex thermophilic
107 communities.

108 **RESULTS**

109 **Physicochemistry**

110 Biologically relevant physicochemical parameters were measured for two soil profiles
111 from Tramway Ridge (Table 1) and limited variation between the active and passive
112 fumaroles was observed: 65°C at all sampling depths; 2.85%-4.4% gravimetric water
113 content; pH of 8.35-8.63; 64.7-81.6 μS conductivity; 0.20-0.44% total nitrogen; 0.12-
114 0.62% total carbon. Subsurface oxygen concentrations revealed the subsurface
115 atmosphere to be suboxic (~28% saturation) at each fumarole. While subtle, differences
116 in physicochemistry were greater between samples taken from the active site (site A)
117 than those from the passive site (site B) (Supplementary Fig. S1)

118 **Community structure of the Tramway Ridge ASPA fumaroles**

119 Microbial community structure was assessed by pyrosequencing multiplexed amplicons
120 (see methods). In total, 111 OTUs were observed across all samples while the number of
121 OTUs observed in any given sample ranged from 22 to 94(Table 2). Because different
122 numbers of sequences were generated across all samples, the read depth was
123 normalized through bootstrap resampling to facilitate meaningful α-diversity and β-

diversity calculations. (see methods). Pairwise dissimilarity indices were greater (T-test p=0.051 ± 0.004) within the vertical profile of the passive fumarole (average Morisita-horn dissimilarity = 0.662 95%CI: {0.197 0.957}) than the more active fumarole (average Morisita-horn dissimilarity = 0.069 95%CI: {0.030 0.103}) over all resamplings. The microbial community at both fumaroles had a marginally higher richness at the surface than in deeper substrata (Table 2) and an inverse relationship was observed between the number of OTUs observed in a sample and sampling depth (Spearman rho=-0.837, p=0.019). This relationship is consistent (bootstrap support) but weaker in resampled communities for both number of OTUs (Spearman rho=-0.717, p=0.054, 95%CI: {-0.717 -0.788}, >99.9% bootstrap support for cor<0) and Chao1 diversity index (Spearman rho=-0.717, p=0.054, 95%CI: {-0.717 -0.788}, 99.8% bootstrap support for cor<0). Diversity metrics that include evenness were found to be less correlated with depth but showed consistency across resampled communities: ShannonH (Spearman rho=-0.239, p=0.32, 95%CI: {-0.239 -0.120}, >99.9% bootstrap support for cor<0); SimpsonsD (Spearman rho= 0.120, p=0.41, 95%CI: { 0.120 0.120}, 99.8% bootstrap support for cor>0). BEST analysis [21] failed to identify any physicochemical parameters that significantly describe variation in community structure using either Jaccard incidence distances (global p>0.10) or Morisita-Horn distances (global p>0.6).

Phylogenetic and thermal classification of fumarolic taxa

The Tramway Ridge fumarolic community was found to be phylogenetically diverse, spanning 16 recognized Prokaryotic divisions, 8 candidate divisions and at least 5

unclassified lineages that conceivably represent novel divisions (Figure 2, Suplementary Table S1). A few phyla such as Chloroflexi and Proteobacteria were well represented by large numbers of OTUs (10 and 16 respectively); however, they constitute vastly different portions of the community. Chloroflexi were present in all samples comprising up to 9% of all reads of a single amplicon library, while Proteobacteria was observed in only 2 samples and represent at most 1.1% of the reads in either. A single Archaeal OTU dominated the amplicon libraries, accounted for 40-60% of reads from each substratum at the active fumarole and showed much greater variation in samples from the passive fumarole (1.5-50.4%). Other prevalent taxa included Planctomycetes, Meiothermus, Cyanobacteria, and members of the OP1_GAL35 and OctSpA1-106 lineages. Of the 111 OTUs that were observed at Tramway Ridge, 68 (61%) were found to be cosmopolitan and comprised between 32% and 92% of the summed relative abundances of any given sample. Cosmopolitan OTUs were further classified into one of five environmental categories based on the types of environment in which their non-Antarctic members have been observed (Table 3, Supplementary Table S1). Almost 31% (34/111) of all OTUs were classified as non-thermal and made up 5-30% of the total number of OTUs observed in resampled communities (Table 3). The surface (0-2 cm) sample from the passive fumarole in particular showed the greatest diversity of non-thermal taxa, with over 25% of all OTUs observed related to non-thermal lineages, collectively comprising 8.5% of the total reads in that sample. Over both profiles, the proportion of OTUs that were classified as non-thermal decreased marginally with depth (Spearman rho=-0.717, p=0.054, 95%CI: {-0.837 -0.478}, 99.9% bootstrap support for cor<0) and the proportion

of OTUs classified as thermal increased marginally with depth (Spearman rho=0.717, p=0.054, 95%CI: {0.598 0.837}, 99.9% bootstrap support for cor>0).

Networks reveal structured groups of organisms

Microbial association network analysis was used to cluster OTUs into intra-correlated groups (ICGs). Network techniques are increasingly used in microbial ecology studies to infer "interactions" between individuals in a community and indicate that taxa vary in a coordinated fashion across different samples[22,23]. ICGs within our association network were defined as a set of OTUs that are interconnected with significant positive correlation values ($r>0.8$ and $p<0.001$). While a minority of the total diversity was captured in the six robust ICGs (26 OTUs), together these taxa account for 88-95% of all reads from any given sample. (Figure 2). Positive correlation values below the significance threshold also reveal potential relationships among OTUs belonging to different ICGs (Figure 3a). These potential connections were strongest between ICGs 1 and 2 (average $r=0.600 \pm 0.162$), between ICGs 3 and 4 (average $r=0.577 \pm 0.108$) and between ICGs 5 and 6 (average $r=0.468 \pm 0.180$). Negative correlations, which may either reflect an antagonistic or an independent relationship between sets of OTUs were also observed between the OTUs that comprise individual ICGs (Figure 3b). Negative correlations were strongest between OTUs assigned to ICG 2 and those assigned to OTUs in ICG 6 ($r= -0.806 \pm 0.098$).

Full length rRNA gene analysis

Twelve full-length 16S rRNA genes were reconstructed from shotgun sequencing of reads from bulk environmental DNA (Supplementary Table S2) and used for phylogenetic analysis (Supplementary Fig. S2). These full-length sequences corresponded to the most abundant (>10% relative abundance in any library) amplicon-derived OTUs from each sample and collectively represented the majority of total amplicons from the passive site (54-82%) and the active site (78-88%). These full-length sequences were affiliated with cyanobacterium *Mastigocladus laminosus*, three distinct lineages of *Meiothermus*, two Acidobacterial lineages, a thermally restricted lineage of Armatimonadetes, three thermally restricted Candidate Division lineages (OctSpA106[24], GAL15 [25,26], and OP1_GAL35[25,27,28], as well as a novel Archaeal sequence.

Two of the full-length sequences were examined in detail because they appear to represent lineages restricted to the Tramway Ridge ASPA. The first, belonging to a Planctomycete, closely matched the sequence from an earlier clone library from the Tramway Ridge ASPA [17], but matched no other sequence in the NCBI or Greengenes databases with more than 89% nucleotide identity. This taxon groups within a clade of environmental clones (99.97% posterior support) as a sister lineage to the Gemmataceae (>99.99% posterior support) within the Gemmatales (97.8% posterior support) (Figure 4a). A full-length sequence was also reconstructed for the novel, dominant Archaeon observed in the amplicon dataset. This particular Archaeon was most closely related (94% identity) to a single clone (GI# 14028778) from a subterranean

hot spring in Iceland[29] and shared no more than 87.1% identity to any other sequence in the NCBI database. Phylogenetic analysis of this full-length 16S rRNA gene grouped it with Crenarchaeota and Thaumarchaeota, specifically suggesting that it is a deep-branching relative of the Thaumarchaeota (Figure 4b). This novel Archaeon evaded amplification in a previous study [17], most likely due to mismatches in both forward and reverse primers. Archaeal sequences found in the previous study were present in the current study as rare taxa and phylogenetically affiliated with group 1.1b of the Thaumarchaeota (Supplementary Table S1 and Supplementary Fig. S2).

DISCUSSION

Amplicon libraries of the 16S rRNA gene were used to examine the distribution of the geothermal microbial community in vertical profiles of fumaroles found on Mt. Erebus. Two profiles were collected for analysis, one from an "active" fumarole, vigorously emitting steam, and another from a "passive" fumarole. Measured physicochemistry varied more at the active fumarole than at the passive fumarole (Supplementary Fig. S1), yet the community structure showed greater stratification at the passive fumarole, as is evident from greater community dissimilarities observed in the depth profile. Subsurface oxygen measurements were similar between the two fumaroles and no additional physicochemical parameters (pH, conductivity, Carbon, Nitrogen, Moisture etc.) were identified to play a significant role in determining community structure. This suggests that currently unidentified soil physicochemistry variables, subsurface gas composition or other unknown characteristics may play a more significant role in

228 determining the community structure within fumaroles than the variables measured in
229 the current study.

230 The microbial ecosystem of the Tramway Ridge ASPA fumaroles was dominated by an
231 enigmatic relative of Thaumarchaeota, Aigarchaeota and Crenarchaeota (Figure 4B)
232 that showed very little sequence identity at the 16S rRNA gene level to other members
233 of these groups (<89% identity). Similarly, a single lineage of Planctomycetes, nearly
234 identical to a clone sequence previously reported for the Tramway Ridge ASPA [17] was
235 identified as a deep-branching member of an uncultivated clade of mostly mesophilic
236 members of the Gemmatales. In contrast, nearly all other dominant organisms were
237 indistinguishable from cosmopolitan thermophiles that have been detected in neutral to
238 alkaline terrestrial hot springs and associated sediments. These lineages include
239 reasonably well-understood microbial mat taxa (Cyanobacteria and Meiothermus) as
240 well as poorly understood divisions (Acidobacteria, OP1_GAL35, GAL15, OctSpA1-106,
241 Armatimonadetes).

242 Mat-associated taxa, such as Cyanobacteria, were detected in the lowest substratum in
243 both profiles. Given the proximity of Cyanobacterial mats to fumaroles, this was not
244 surprising, yet it presented a challenge for identifying which taxa best represented
245 thermophiles that were specifically associated with the fumarolic mineral soils. A
246 microbial association network (Figure 3) based on correlation of relative abundance
247 profiles was used to separate individual OTUs into structured subcommunities of

248 dominant OTUs or "intra-correlated groups" (ICGs). Although only 26 out of 111 OTUs
249 robustly clustered into 6 ICGs throughout multiple *in silico* resamplings, these OTUs
250 nevertheless comprised 88-95% of amplicon reads from each sample (Figure 2) and
251 enabled the separation of surface-associated taxa from the subsurface microbial
252 community.

253 Surface mat-associated subcommunities (ICGs 1 and 2) were dominated by
254 *Meiothermus sp.* and *Mastigocladus laminosus*, the latter of which is known to
255 dominate cyanobacterial mats at the Tramway Ridge ASPA [7]. The upper temperature
256 limit for cultures of Tramway Ridge ASPA-specific strains of *Mastigocladus laminosus* is
257 approximately 60°C and at the Tramway Ridge ASPA, mats do not extend across the hot
258 steaming fumaroles [7,30,31]. The presence of *Mastigocladus laminosus* in ICG 1 suggests
259 that this subcommunity was restricted to cooler temperatures than 65°C, the measured
260 temperature of mineral soils at the time of sampling and the annual mean temperature
261 previously reported for individual fumaroles at the Tramway Ridge ASPA [17].
262 Subcommunities defined by ICGs 1 and 2 include numerous additional lineages that are
263 known to be associated with phototrophic microbial mats: *Meiothermus sp.*,
264 Bacteroidetes, Chlorobi and Chloroflexi. It is uncertain at this time whether the
265 correlation of these taxonomic signatures with Cyanobacterial signatures indicates that
266 these taxa were physically associated with fumarole-adjacent mats, or if it indicates that
267 these taxa were inhabitants of the fumarole that were directly dependent on mat-
268 derived materials. These subcommunities were proportionately greater in the surface

269 (0-2 cm) of the highly stratified passive fumarole, and also included non-thermal OTUs
270 (Supplementary Table S1).

271 Subsurface subcommunities (ICGs 5 and 6) were negatively correlated (Figure 3b) with
272 surface mat-associated subcommunities (ICGs 1 and 2) and therefore OTUs in these
273 subcommunities are not likely to have been derived from surface mat-derived materials.
274 We found that a novel relative of the Crenarchaeota, Thaumarchaeota and Aigarchaeota
275 (Figure 2 and Figure 4) dominated the subsurface, as evidenced from amplicon library
276 abundances and 16S reads from shotgun sequencing of environmental DNA. The exact
277 relationship among the Aigarchaeota, Hot Thaumarchaeota-related Clade (HTC) and
278 Thaumarchaeota is currently being debated in the literature[32-34] and therefore
279 phylogenetic analysis of a full-length reconstructed gene could not fully resolve whether
280 this lineage should be included as a member of the Thaumarchaeota. Several
281 demarcations are presented (Figure 4) that indicate robust monophyletic groupings with
282 >99.0% posterior support and might provide reasonable phylogenetic limits to the
283 Thaumarchaeota. In the present survey, we chose to refer to this lineage as
284 Thaumarchaeota-like to capture this uncertainty and to differentiate between this
285 relative of Thaumarchaeota and the recently described HTC lineages[34], which forms its
286 own, separate, well-supported clade.

287 In addition the Thaumarchaeota-like taxon, subsurface subcommunities contained
288 lineages of candidate divisions currently known only through 16S clone sequences.

289 OctSpA1-106 was originally detected at Octopus Spring in Yellowstone National Park
290 (USA)[24] and has since been detected at El Tatio Geyser Field, Chile[35], Uzon Caldera
291 (Russia)[36] and terrestrial hot-spring areas in North America[28,37]. It has been detected
292 only in neutral to alkaline, geothermally heated environments. The full-length 16S
293 sequence from Tramway Ridge is 96% identical with the original type clone sequence
294 (GI# 3800711) and 97% identical with clone GAL39 (GI# 84322458), both from
295 Yellowstone National Park. The third lineage within ICG 6 is a member of Candidate
296 Division OP1. However, the sequences in our dataset are highly similar (98.55%
297 nucleotide identity over the full-length 16S rRNA gene) to members of the GAL35 Class
298 of Candidate Division OP1, and in our phylogenetic reconstructions (Supplementary Fig.
299 S2), this class forms a distinct Division-level lineage, as has been observed by other
300 researchers[28].

301 OTUs were assigned to ICGs 3 and 4 were difficult to identify as either surface or
302 subsurface-associated. The abundance of OTUs in ICGs 3 and 4 decrease with depth
303 over the profile at the active site (site A), but are most abundant in the deepest
304 substratum at the passive site (siteB). OTUs assigned to these ICGs include novel
305 lineages of Actinobacteria and Chloroflexi, as well as thermally restricted lineages of
306 Armatimonadetes, Acidobacteria, Candidate Division BH1 and OP1 (GAL35). ICG 4 also
307 included a novel lineage of Planctomycetes that was detected in an earlier study[17] and
308 shown to be present in relatively high abundance in the current study. This full-length
309 sequence could confidently be assigned to the Gemmatales (97.8% posterior support)

(Figure 4a) and potentially redefines the upper temperature limit for this Order at 60-65°C. While the "Anammox" lineages of Planctomycetes have been detected in high-temperature (>75°C) environments[38,39], the previous upper temperature limit for characterized members of the Gemmatales was approximately 55°C (*Isosphaera pallida*[40]).

We found that amplicons for the majority of OTUs (68/111) were highly similar (>97% sequence identity over 400 nt of the 16S rRNA gene) to clones reported from non-Antarctic environments (Supplementary Table S1 and Supplementary Fig. S4). These OTUs collectively accounted for 32-92% of the amplicons in any given sample library. We propose that the taxa represented by these OTUs have likely been introduced to the Tramway Ridge ASPA through recent (in geologic terms) aeolian transport. Aeolian dispersal has also been invoked to explain the incidence and movement of mat-associated Cyanobacteria, moss and algae among geothermal sites in Antarctica[7,30,41,42].

Atmospheric patterns suggest that South America is a likely source for Antarctic microbes. Over the glacial periods of the last 160,000 years, the majority of dust particles that were trapped in ice in Antarctica have been derived primarily from Patagonia[43,44]. Consistent with the aeolian input of mesophilic organisms into Antarctica, we found many non-thermal cosmopolitan OTUs at Tramway Ridge that were predominately found in the surface strata (0-2 cm) of both profiles. It is possible that these are signatures of organisms that have colonized the cooler areas adjacent to

hot fumaroles and are distributed to fumarolic locations by the harsh winds at Tramway Ridge. Alternatively, some of these organisms may actually be living in a cooler, fumarolic surface crust that is too thin to detect within the resolution of our temperature probes (~1 cm). Regardless, the detection of diverse mesophilic cosmopolitan signatures at Tramay Ridge hints that the dispersal of organisms to the site may not be limited to thermophiles.

Abundant, cosmopolitan thermophiles in our dataset were classified as *Mastigocladus laminosus*, *Meiothermus sp*., Acidobacteria, OctSpA1-106, GAL35 and OP1_GAL35 and are indistinguishable from those found in geothermal environments elsewhere. This suggests that each of these lineages possess adaptations for long-range dispersal. Interestingly, these OTUs match sequences from various areas known to be capable of "super-volcano" scale eruptions, including the Yellowstone Caldera (USA)[24,27], Altiplano-Puna Volcanic Complex (Chile, Bolivia)[35], and the Taupo Volcanic zone (New Zealand)[45]. Specific organisms, such as *Mastigocladus laminosus,* which form conspicuous microbial mats at the Tramway Ridge ASPA [7,17] have previously been hypothesized to have originated from Yellowstone Caldera and it has been suggested that they have subsequently become globally distributed through aeolian dispersal [46]. Our dataset alone cannot be used to identify sources of the microbiota at Tramway Ridge, but volcanic events are readily detected in Antarctic ice cores and several of these can be traced to off-continent events[47]. While it remains difficult to understand how organisms might survive a typical pyroclastic eruption, a directed-blast explosion with enough

force to eject material before superheating could potentially launch organisms and debris high into the atmosphere. Granted, the types of eruptions that might be capable of such an accomplishment would be large and rare, but it has been shown that even moderately large eruptions can disperse larger, eukaryotic organisms over vast distances[6].

Our results show that much of the diversity at Tramway Ridge is cosmopolitan and therefore that viable cells, from diverse divisions of the tree of life, are successful at global dispersal. While this ability is well known for particular Bacterial species [11,14,48-50] others, especially thermophiles, are known to exhibit evidence of dispersal limitation [46,51-54]. Although these categorizations are largely dependent on temporal scaling [55] and all organisms probably exhibit some level of dispersal limitation[56], here we have shown that global-scale dispersal nevertheless plays a significant role in the assembly of microbial communities. On the other hand, the presence of highly abundant, endemic populations suggests that specific organisms are uniquely adapted to this particular combination of environmental factors. These endemic taxa are defined explicitly by their ability to outcompete exogenous microbes, and their inability to disperse to other habitats. In our work, this principle is shown for thermophilic communities, however there is no reason to suspect that this principle cannot be extended to mesophilic communities. We conclude that the process of microbial dispersal over global-scale distances must be considered as an important component of the general concept of microbial community assembly.

METHODS

Sample collection

Sediment samples were collected within the Tramway Ridge Antarctic Specially Protected Area (ASPA 130) in February 2009 from two sites (site A: 77° 31.103' S, 167° 6.682' S and site B: 77° 31.106' S, 167° 6.668' E). All suggested sterilization protocols for entering into this protected site were adhered to, following the ASPA 130 Management Plan (http://www.scar.org/publications/bulletins/151/aspa130.html). Sites were chosen based on measuring a surface temperature of 65°C with a stainless steel Checktemp1 temperature probe (Hanna Instruments, Rhode Island, USA), sterilized with 70% ethanol immediately prior to each use. Surface "crust" was set aside prior to collecting samples. Samples were collected by aseptically removing the top 2 cm of sediment in an approximately 25cm^2 area. Sediment was placed into a fresh 50 mL Falcon tube. Sampling continued with the collection of a second (2-4 cm depth) and third (4-8 cm depth) layer following the same procedures. Temperature measurements were repeated for each layer sampled. All samples were immediately frozen, transported back to the University of Waikato frozen and maintained at -80°C in the laboratory until analyzed.

Physicochemistry

Samples were thawed and aliquots removed for pH/conductivity, total moisture content, and carbon/nitrogen analyses. Preparation and analyses followed previously

published procedures[57,58], and results are summarized in Table 1. Moisture content, pH, total carbon and total nitrogen were combined with previously collected physicochemistry data[17] and examined in a principal components analysis using the "rda" function from the vegan package in R (http://www.R-project.org). Subsurface fumarolic Oxygen concentrations were measured in a subsequent trip to the Tramway Ridge ASPA (November, 2011) using a Fibox 3 LCD trace minisensor oxygen meter (PreSens Precision sensing, Germany) having been calibrated for temperature and altitude at our base camp.

DNA extraction, library preparation and sequencing

DNA was extracted from samples using a modified CTAB (cetyltrimethylammonium bromide) bead-beating protocol[59] and quantified using the Quant-IT dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA). For shot-gun sequencing, a portion of extracted genomic DNA was sequenced using standard protocols for the 454-Ti platform (Roche 454 Life Sciences, Branford, CT, USA) at the UCLA GenoSeq CORE. PCR amplicons containing V5–V7 hypervariable regions of the 16S rRNA gene were generated from the same genomic DNA samples using primers Tx9 (5'-GGATTA GAWACCCBGGTAGTC-3') and 1391R (5'-GACGGGCRGTGWGTRCA-3') [60]. PCR was performed in triplicate on each sample and pooled to reduce stochastic variation[61]. Three samples (site A 0-2 cm, site B 0-2 cm, site B 2-4 cm) were sequenced using the 454-GS-FLX platform by Taxon Biosciences (Tiburon, CA, USA) and three samples (site A 2-4 cm, site A 4-8 cm and site B 4-8 cm) were sequenced using the 454 Junior platform at the Waikato DNA Sequencing

Facility (Hamilton, New Zealand). For the three samples sequenced using the 454-GS-FLX platform, each 30 µl reaction contained 2-10 ng of DNA extract, Pfx polymerase and platinum polymerase (0.5 U each; Invitrogen), 1× Pfx PCR buffer with Pfx enhancer, 0.2 mM dNTPs, 1 mM $MgCl_2$, 0.02 mg/ml BSA, 0.8 µM of forward and reverse primer, and PCR-grade water. Thermal cycling conditions were 94°C for 2min; 24 cycles of 94°C for 15 s, 55°C for 30 s and 68°C for 1 min; and 68°C for 3 min. Amplicons were size-selected and purified using polyacrylamide gel electrophoresis before being prepared for pyrosequencing by Taxon Biosciences. For the three samples sequenced using the 454-junior platform, each 30 µl reaction contained 2-10 ng of DNA extract, PrimeStar polymerase (0.625 U; Takara), 1× PCR buffer, 0.2 mM dNTPs, 0.4 µM of forward and reverse primer, and PCR-grade water. Thermal cycling conditions were 94°C for 3min; 24 cycles of 94°C for 20s, 52°C for 20s and 72°C for 45s; and 72°C for 3 min. Triplicate PCR reactions were pooled and gel-purified using the UltraCleanTM 15 DNA Purification Kit (MO BIO Laboratories Inc.), cleaned using the Agencourt AMPure XP Bead Cleanup kit (Beckman Coulter Inc.) and quantified (Quant-iTTM dsDNA HS Assay Kit, Invitrogen Ltd.). Cleaned amplicons were used as template (25 ng) in a second PCR reaction using fusion primers (Forward: 5'-{454 Adapter A}-TCAG-MID-Tx9-3'; Reverse: 5'-{454 Adapter B}-TCAG-1391R-3'). PCR conditions were exactly the same as the first round except only 10 cycles of PCR were performed. Triplicate PCR reactions were pooled and gel-purified using the UltraCleanTM 15 DNA Purification Kit, cleaned using the Agencourt AMPure XP Bead Cleanup kit and quantified (Quant-iTTM dsDNA HS Assay Kit) before being prepared for pyrosequencing using the 454 Junior platform by the University of Waikato

sequencing facility. Original pyrosequencing flowgram files have been deposited at the European Nucleotide Archive under study ERP002340 [currently private but sff files are available upon request for review].

Sequence processing

Shot-gun sequencing reads were queried against a modified 16S rRNA gene database [62] using BLAST[63]. 1033 reads returned a bit score greater than 50 (see Supplementary Fig. S3 for histogram of all bit scores) and were assembled in Newbler (Roche 454 Life Sciences) using an overlap of 200 nt and percent identity cutoff of 99% to produce near-full-length 16S rRNA genes. Flowgrams for the shot-gun sequencing reads that were used to reconstruct 16S rRNA genes are available from the European Nucleotide Archive under study ERP002340 [currently private, but available upon request for review]. Near-full length 16S rRNA gene sequences were aligned to 180 additional reference sequences using SINA[64] and the resulting alignment was used in a Bayesian phylogenetic analysis in Bali-Phy[65] with the following parameters: alignment="traditional", iterations=11000, burnin=6000. 20 independently seeded chains were initiated and independent posterior tree populations were pooled into a single posterior tree population from which consensus trees were built. 16S amplicon pyrosequencing results were processed using AmpliconNoise v1.22 to remove noise and filtered for chimeric reads using Perseus[66]. Reads were required to perfectly match MID sequences for processing. Sequence predictions from fewer than 3 individual raw reads or with at least one primer mismatch were discarded. Pairwise alignments/distances between sequence

predictions were calculated using ESPRIT[67]and reads were clustered into Operational Taxonomic Units (OTUs) using a distance cutoff of 0.03 and the average neighbor clustering algorithm (Table 2) in Mothur[68]. OTUs that correspond to full-length sequences (>99% identity) were classified using our calculated phylogeny with a strict phylogenetic nesting criterion while remaining OTUs were classified taxonomically using classify.seqs in Mothur with 80% confidence as a threshold, using the Greengenes alignment and taxonomic outline released in November, 2012 [69].

Environmental classification of OTUs

Each unique amplicon sequence was queried against the NCBI nr nucleotide database using BLAST[63]. Reads were classified as "thermal" in the case that all database entries that matched the representative unique sequence with >97% identity had been previously observed in thermal environments, as "non-thermal" if all matches were observed in non-thermal environments and as "polythermal" if the matches were observed in a mixture of thermal and non-thermal environments. Reads for which the representative unique sequences only matched with <97% identity to any entries in the database were classified as "sub-novel" or "novel" if the nearest match was between 95.5% and 97%, or <95.5%, respectively (see Supplementary Fig. S4 for histogram justifying these boundaries).

Determination of Intra-Correlated Groups

OTUs were partitioned into "Intra-Correlated Groups" (ICGs) based on resampled correlation between relative abundance across all samples. First, "adjusted" relative abundances were calculated by adding a single Bayesian pseudo-count to the number of counts observed for each OTU. This allowed every OTU to have a non-zero probability of occurrence in any sample. These revised counts were used as a probability vector to resample OTUs to the same depth (1431 counts) within each sample using the sample() function in R[70]. Resampling was carried out 2000 times. For each resampling, Pearson correlations were calculated in a pairwise fashion between all OTU pairs using log-transformed relative abundances. OTU networks were constructed based on Pearson correlation coefficients exceeding 0.8 in >99.9% of resampled communities.

REFERENCES

1. De Wit, R. & Bouvier, T. *'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ. Microbiol.* **8,** 755–758 (2006).
2. Lighthart, B. & Shaffer, B. Airborne Bacteria in the Atmospheric Surface Layer: Temporal Distribution above a Grass Seed Field. *Appl. Environ. Microbiol* **61,** 1492–1496
3. Folmsbee, M. & Strevett, K. A. Bioaerosol Concentration at an Outdoor Composting Center. *Journal of the Air & Waste Management …* **49,** 554–561 (1999).
4. Bonheyo, G. T. & Frias-Lopez, J. A test for Airborne Dispersal of Thermophilic Bacteria from Hot Springs. *Geothermal Biology and …* (2005).
5. Jones, A. M. & Harrison, R. M. The effects of meteorological factors on atmospheric bioaerosol concentrations--a review. *Sci. Total Environ.* **326,** 151–180 (2004).
6. Van Eaton, A. R., Harper, M. A. & Wilson, C. J. N. High-flying diatoms: Widespread dispersal of microorganisms in an explosive volcanic eruption. *Geology* (2013). doi:10.1130/G34829.1
7. Broady, P. A. Taxonomic and ecological investigations of algae on steam-warmed soil on Mt Erebus, Ross Island, Antarctica. *Phycologia* **23,** 257–271 (1984).
8. Prospero, J. M., Blades, E., Mathison, G. & Naidu, R. Interhemispheric transport of viable fungi and bacteria from Africa to the Caribbean with soil dust. *Aerobiologia* **21,** 1–19 (2005).
9. Burrows, S. M. & Elbert, W. Bacteria in the global atmosphere- Part 1: Review and synthesis of literature data for different ecosystems. *Atmospheric …* (2009).
10. Pöschl, U. & Lawrence, M. G. Bacteria in the global atmosphere- Part 2: Modelling of emissions and transport between different ecosystems. *Atmospheric …* (2009).
11. Smith, D. J. *et al.* Free Tropospheric Transport of Microorganisms from Asia to North America. *Microb. Ecol.* **64,** 973–985 (2012).
12. Yamaguchi, N., Ichijo, T., Sakotani, A., Baba, T. & Nasu, M. Global dispersion of bacterial cells on Asian dust. *Sci Rep* **2,** 525 (2012).
13. Martiny, J. B. H. *et al.* Microbial biogeography: putting microorganisms on the map. *Nat Rev Micro* **4,** 102–112 (2006).
14. Wynn-Williams, D. D. Aerobiology and colonization in Antarctica — the BIOTAS Programme. *Grana* **30,** 380–393 (1991).
15. Ugolini, F. C. & Starkey, R. L. Soils and Micro-organisms from Mount Erebus, Antarctica. *Nature* **211,** 440–441 (1966).
16. Harpel, C. J., Kyle, P. R., Esser, R. P., McIntosh, W. C. & Caldwell, D. A. 40Ar/39Ar dating of the eruptive history of Mount Erebus, Antarctica: summit flows, tephra, and caldera collapse. *Bull Volcanol* **66,** 687–702 (2004).
17. Soo, R. M., Wood, S. A., Grzymski, J. J., McDonald, I. R. & Cary, S. C. Microbial

biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica. *Environmental Microbiology* **11,** 715–728 (2009).
18. Wardell, L. J., Kyle, P. R. & Campbell, A. R. Carbon dioxide emissions from fumarolic ice towers, Mount Erebus volcano, Antarctica. *Geological Society, London, Special Publications* **213,** 231–246 (2003).
19. Gihring, T. M., Moser, D. P., Lin, L. H. & Davidson, M. Taylor & Francis Online :: The Distribution of Microbial Taxa in the Subsurface Water of the Kalahari Shield, South Africa - Geomicrobiology Journal - Volume 23, Issue 6. *Geomicrobiology …* (2006).
20. Weidler, G. W., Dornmayr-Pfaffenhuemer, M., Gerbl, F. W., Heinen, W. & Stan-Lotter, H. Communities of Archaea and Bacteria in a Subsurface Radioactive Thermal Spring in the Austrian Central Alps, and Evidence of Ammonia-Oxidizing Crenarchaeota. *Applied and Environmental Microbiology* **73,** 259–270 (2007).
21. Clarke, K. R. & Somerfield, P. J. Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *Journal of Experimental Marine …* 56–69 (2008).
22. Gilbert, J. A. *et al.* Defining seasonal marine microbial community dynamics. *ISME J* **6,** 298–308 (2012).
23. Faust, K. *et al.* Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* **8,** e1002606 (2012).
24. Blank, C. E., Cady, S. L. & Pace, N. R. Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. *Applied and Environmental Microbiology* **68,** 5123–5135 (2002).
25. Ackerman, G. G. Biogeochemical gradients and energetics in geothermal systems of Yellowstone National Park. *Department of Land Resources and Envi- ronmental Science. Bozeman, MT, USA: Montana State University* (2006).
26. Lin, X., Kennedy, D., Fredrickson, J., Bjornstad, B. & Konopka, A. Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. *Environmental Microbiology* **14,** 414–425 (2011).
27. Hugenholtz, P., Pitulle, C., Hershberger, K. L. & Pace, N. R. Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol* **180,** 366–376 (1998).
28. Cole, J. K. *et al.* Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J* (2012). doi:10.1038/ismej.2012.157
29. Marteinsson, V. T. *et al.* Phylogenetic diversity analysis of subterranean hot springs in Iceland. *Applied and Environmental Microbiology* **67,** 4242–4248 (2001).
30. Broady, P., Given, D., Greenfield, L. & Thompson, K. The biota and environment of fumaroles on Mt Melbourne, northern Victoria Land. *Polar Biology* **7,** 97–113 (1987).
31. Melick, D., Broady, P. & Rowan, K. Morphological and physiological characteristics of a non-heterocystous strain of the cyanobacterium Mastigocladus laminosus Cohn from fumarolic soil on Mt Erebus, Antarctica. *Polar Biology* **11,** (1991).

32. Brochier-Armanet, C., Gribaldo, S. & Forterre, P. Spotlight on the Thaumarchaeota. *ISME J* **6,** 227–230 (2012).
33. Gribaldo, S. & Brochier-Armanet, C. Time for order in microbial systematics. *Trends Microbiol.* **20,** 209–210 (2012).
34. Eme, L. *et al.* Metagenomics of Kamchatkan hot spring filaments reveal two new major (hyper)thermophilic lineages related to Thaumarchaeota. *Research in Microbiology* **164,** 425–438 (2013).
35. Engel, A. S., Johnson, L. R. & Porter, M. L. Arsenite oxidase gene diversity among Chloroflexi and Proteobacteria from El Tatio Geyser Field, Chile. *FEMS Microbiology Ecology* **83,** 745–756 (2013).
36. Wemheuer, B., Taube, R., Akyol, P., Wemheuer, F. & Daniel, R. Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka Peninsula. *Archaea* **2013,** 136714 (2013).
37. Costa, K. C. *et al.* Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. *Extremophiles* **13,** 447–459 (2009).
38. Byrne, N. *et al.* Presence and activity of anaerobic ammonium-oxidizing bacteria at deep-sea hydrothermal vents. *ISME J* **3,** 117–123 (2009).
39. Li, H., Chen, S., Mu, B.-Z. & Gu, J.-D. Molecular detection of anaerobic ammonium-oxidizing (anammox) bacteria in high-temperature petroleum reservoirs. *Microb. Ecol.* **60,** 771–783 (2010).
40. Giovannoni, S. J., Godchaux, W., Schabtach, E. & Castenholz, R. W. Cell wall and lipid composition of Isosphaera pallida, a budding eubacterium from hot springs. *J. Bacteriol* **169,** 2702–2707 (1987).
41. Broady, P. A. The morphology, distribution and ecology of Pseudococcomyxa simplex (Mainx) Fott (Chlorophyta, Chlorellaceae), a widespread terrestrial Antarctic alga. *Polar Biology* **7,** 25–30 (1987).
42. Bargagli, R. & Broady, P. A. Preliminary investigation of the thermal biosystem of Mount Rittmann fumaroles (northern Victoria Land, Antarctica). *Antarctic Science- …* **8,** 121–126 (1996).
43. Basile, I. *et al.* Patagonian origin of glacial dust deposited in East Antarctica (Vostok and Dome C) during glacial stages 2, 4 and 6. *Earth and Planetary Science Letters* **146,** 573–589 (1997).
44. Iriondo, M. Patagonian dust in Antarctica. *Quaternary International* **68-71,** 83–86 (2000).
45. Stott, M. B. *et al.* Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environmental Microbiology* **10,** 2030–2041 (2008).
46. Miller, S. R., Castenholz, R. W. & Pedersen, D. Phylogeography of the thermophilic cyanobacterium Mastigocladus laminosus. *Applied and Environmental Microbiology* **73,** 4751–4759 (2007).
47. Bay, R. C., Bramall, N. E. & Price, P. B. Globally synchronous ice core volcanic tracers and abrupt cooling during the last glacial period. *Journal of …* (2006).
48. Roberts, M. S. & Cohan, F. M. Recombination and migration rates in natural

populations of Bacillus subtilis and Bacillus mojavensis. *Evolution* 1081–1094 (1995).
49. Jungblut, A. D., Lovejoy, C. & Vincent, W. F. Global distribution of cyanobacterial ecotypes in the cold biosphere. *ISME J* **4,** 191–202 (2010).
50. DeLeon-Rodriguez, N. *et al.* Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proceedings of the National Academy of Sciences* **110,** 2575–2580 (2013).
51. Papke, R. T., Ramsing, N. B., Bateson, M. M. & Ward, D. M. Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology* **5,** 650–659 (2003).
52. Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301,** 976–978 (2003).
53. Dvořák, P., Hašler, P. & Poulíčková, A. PLoS ONE: Phylogeography of the Microcoleus vaginatus (Cyanobacteria) from Three Continents – A Spatial and Temporal Characterization. *PLoS ONE* (2012).
54. Bahl, J. *et al.* Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nat Comms* **2,** 163 (2011).
55. Green, J. & Bohannan, B. J. M. Spatial scaling of microbial biodiversity. *Trends Ecol. Evol. (Amst.)* **21,** 501–507 (2006).
56. Papke, R. T. & Ward, D. M. The importance of physical isolation to microbial diversification. *FEMS Microbiology Ecology* **48,** 293–303 (2004).
57. Lee, C. K. *et al.* Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. *PLoS ONE* **7,** e44224 (2012).
58. Tiao, G., Lee, C. K., McDonald, I. R., Cowan, D. A. & Cary, S. C. Rapid microbial response to the presence of an ancient relic in the Antarctic Dry Valleys. *Nat Comms* **3,** 660 (2012).
59. Barrett, J. E., Virginia, R. A., Wall, D. H. & Cary, S. C. Cambridge Journals Online - Abstract - Co-variation in soil biodiversity and biogeochemistry in northern and southern Victoria Land, Antarctica. *Antarctic …* (2006).
60. Ashby, M. N., Rine, J., Mongodin, E. F., Nelson, K. E. & Dimster-Denk, D. Serial analysis of rRNA genes and the unexpected dominance of rare members of microbial communities. *Applied and Environmental Microbiology* **73,** 4532–4542 (2007).
61. Suzuki, M. T. & Giovannoni, S. J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology* **62,** 625–630 (1996).
62. Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W. & Banfield, J. F. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* **12,** R44 (2011).
63. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25,** 3389–3402 (1997).
64. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple

sequence alignment of ribosomal RNA genes. *Bioinformatics* **28,** 1823–1829 (2012).
65. Suchard, M. A. & Redelings, B. D. BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22,** 2047–2048 (2006).
66. Quince, C., Lanzén, A., Davenport, R. J. & Turnbaugh, P. J. Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics* **12,** 38 (2011).
67. Sun, Y. *et al.* ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Res* **37,** e76–e76 (2009).
68. Schloss, P. D. & Handelsman, J. *Genome Biol* **6,** 229 (2005).
69. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6,** 610–618 (2012).
70. R, C. T. R: A language and environment for statistical computing. *R-project.org* (2012).

END NOTES

ACKNOWLEDGEMENTS

Financial support was provided by grant UOW0802 from the New Zealand Marsden Fund to SCC and IRM and a CRE award from the National Geographic Society to SCC. Antarctic logistic support for Event K-023 was provided by Antarctica New Zealand. Additional salary support was provided by the New Zealand Marsden fund to CKL (UOW1003) and CWH (UOW0802). We thank Rochelle Soo and Joseph J Grzymski for assistance in collection of samples, Grace Tiao for assistance in physicochemical analyses and Sarah Kelly for assistance in amplicon library preparation. The authors have no conflict of interest to declare.

AUTHOR CONTRIBUTIONS

SCC and IRM conceived of the study, secured funding and collected samples. CKL and CWH adapted laboratory and bioinformatic protocols to ensure quality data production. CWH led analysis of data and authorship of the manuscript. All authors contributed to the writing of this manuscript.

COMPETING FINANCIAL INTERESTS

The authors have no competing financial interests to declare

SEQUENCE DATA ACCESSION

All raw data (*.sff files) for this study are available to the Sequence Read Archive as study number ERP002340 (http://www.ebi.ac.uk/ena/data/view/ERP002340). Assembled near full-length ribosomal RNA sequences are available at Genbank (Accession # KF923316-KF923327)

FIGURE LEGENDS

Figure 1

Location of sampling sites in the present study. a) Location of Ross Island in context of the continent of, Antarctica. b) Location of the summit of Mt. Erebus in context of Ross Island. c) View of the summit Caldera of Mt. Erebus showing location of the Tramway Ridge Antarctica Specially Protected Area (ASPA) 130. d) Location of sampling sites A (active fumarole - orange) and B (passive fumarole - green) within ASPA 130. Boundaries of the ASPA are denoted with red lines. Entry into the northern half of the ASPA is strictly prohibited. e) Fumarole sampled as site A (active fumarole). f) Fumarole sampled as site B (passive fumarole). Imagery provided by Polar Geospatial Center; © 2012 DigitalGlobe, Inc. (fig 1a, 1c, 1d), NASA Earth Observatory (fig 1b) and S. Craig Cary (fig 1e, 1f).

Figure 2

Abundance profiles for twenty-six OTUs that were clustered into Intra-Correlated Groups (ICGs). ICGs are referred to in the text according to numeric designations in the far left column. Bar charts were show the relative abundance of OTUs belonging to each ICG within each sample. Height of bars represents log-transformed relative abundance of each OTU in each sample. The y-axis ranges from 5×10^{-5} to 1 and is identical across samples and ICGs. Non-transformed relative abundances for each OTU may be found in supplemental Table 1.

718 **Figure 3**

719 Complete correlation network for twenty-six OTUs that clustered into Intra-Correlated
720 Groups (ICGs). Nodes represent individual OTUs and are colored by taxonomy. OTUs in
721 each ICG are bound within a yellow area. a) Blue lines represent positive correlation
722 between OTUs. b) red lines indicate negative correlation between OTUs. The shade of
723 lines are scaled according to strength of correlation as indicated by the color key. c)
724 color key for taxonomic classifications.

725 **Figure 4**

726 Bayesian consensus phylogenetic trees of full-length SSU-rRNA genes reconstructed
727 from shot-gun DNA sequencing for the dominant Planctomycetes (A) and Archaeal (B)
728 taxa at the Tramway Ridge ASPA. Bipartitions were drawn at the 80% consensus level
729 from a posterior distribution generated from 20 independent Markov chains.
730 Bipartitions are designated as follows for posterior support: 90% < * < 99% < ** < 99.9%
731 < ***

732 **TABLES**

733 **Table 1**

734 Physicochemical data for the Tramway Ridge ASPA samples.

Table 1 | Physicochemical Parameters for Tramway Ridge Samples						
Site	Active Site			Passive Site		
Depth (cm)	0-2	2-4	4-8	0-2	2-4	4-8
Temperature	65	65	65	65	65	65
Extracted DNA (ng/g soil)	14.9 (±5.3)	8.0 (±2.8)	0.04 (±0.02)	18.6 (±6.6)	0.9 (±0.3)	1.6 (±0.6)
Moisture Content	3.25%	2.85%	4.40%	3.63%	4.17%	4.00%
pH	8.35 (±0.17)	8.52 (±0.18)	8.45 (±0.18)	8.42 (±0.18)	8.63 (±0.24)	8.61 (±0.27)
Conductivity (µS)	81.6 (±5.8)	69.1 (±1.9)	75.2 (±0.9)	76.6 (±4.4)	70.1 (±5.3)	64.7 (±2.2)
Total Nitrogen Mass	0.20%	0.26%	0.33%	0.40%	0.44%	0.40%
Total Carbon Mass	0.62%	0.21%	0.12%	0.20%	0.15%	0.13%
Organic Carbon Mass	0.57%	0.26%	0.15%	0.22%	0.16%	0.15%
Oxygen saturation	28.3% (1.03 mg/L) @ 65.5 C in Nov, 2011			27.98% (1.20 mg/L) @ 62.7 C in Nov, 2011		

735

736

Table 2

Sequencing statistics and diversity metrics for total and resampled communities.

Table 2 | Sequencing and Diversity Statistics for each sample. 95% confidence estimates are reported within curly brackets.

		Active Site			Passive Site		
Sequencing Stats	Depth range (cm)	0-2	2-4	4-8	0-2	2-4	4-8
	raw	75682	8219	5535	67320	86022	7944
	Passed QC	55459	4556	2862	49496	58378	3740
	Unique (N=370)	103	60	41	120	64	75
Average Neighbor Clustering, d=0.03	No. of OTUs (N=111)	75	32	22	94	46	40
	Simpsons D	0.207	0.351	0.321	0.212	0.286	0.105
	Shannon H'	2.45	1.69	1.69	2.44	1.93	2.78
Bootstrapped Sub-Sampling (N_{reads}=1431, 2000 iterations)	No. of OTUs	54.5 {50 60}	30.6 {28 32}	21.8 {21 22}	65.6 {60-72}	35.3 {31 39}	39 {37 40}
	Simpsons D	0.207 {0.193 0.222}	0.352 {0.333 0.371}	0.321 {0.305 0.339}	0.212 {0.199 0.225}	0.286 {0.270 0.304}	0.105 {0.097 0.113}
	Shannon H'	2.44 {2.37 2.5}	1.68 {1.62 1.74}	1.68 {1.63 1.74}	2.43 {2.37 2.49}	1.93 {1.87 1.98}	2.78 {2.73 2.82}

741

742 **Table 3**

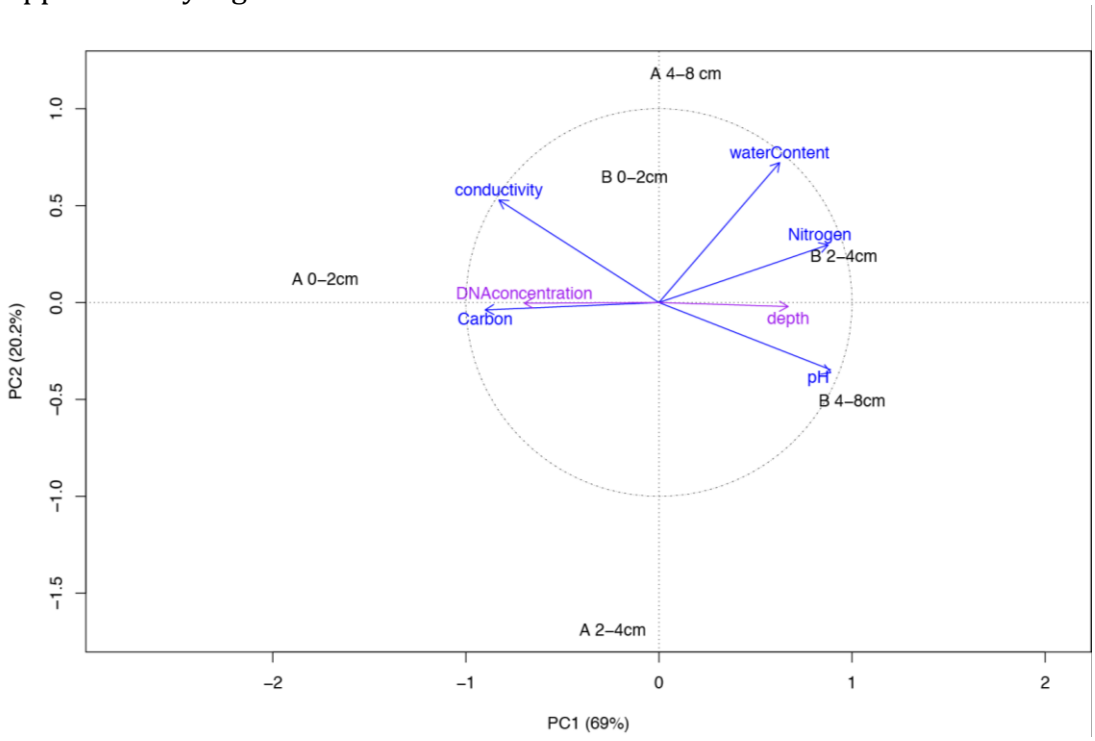
743 Incidence of environmental classifications for each total and resampled dataset.

Table 3 | OTU Frequency and Abundance by Classification

		Active Site			Passive Site		
	Depth (cm)	0-2	2-4	4-8	0-2	2-4	4-8
Average Neighbor Clustering, d=0.03	OTUs (N=111)	75	32	22	94	46	40
	Nonthermal OTUs (N=34)	16	4	2	33	3	4
	Polythermal OTUs (N=10)	9	4	2	10	6	5
	Thermal OTUs (N=24)	20	14	9	19	17	15
	Subnovel OTUs (N=6)	3	1	0	3	1	2
	Novel OTUs (N=35)	25	8	9	27	17	12
	Supernovel OTUs (N=2)	2	1	0	2	2	2
Bootstrapped Sub-Sampling (N_{reads}=1431, 1000 iterations)	OTUs	54.5 {50 60}	30.6 {28 32}	21.8 {21 22}	65.6 {60-62}	35.3 {31 39}	39 {37 40}
	Nonthermal OTUs	9.7 {7 12}	3.8 {3 4}	1.9 {1 2}	18.9 {15 23}	2.2 {1 3}	4 {3 4}
	Polythermal OTUs	7.5 {5 9}	3.4 {2 4}	2 {2 2}	8.1 {7 10}	4.4 {2 6}	4.7 {3 5}
	Thermal OTUs	18 {16 20}	13.7 {13 14}	9 {9 9}	18.1 {17 19}	13.3 {11 15}	14.6 {13 15}
	Subnovel OTUs	2 {2 2}	0.9 {0 1}	0 {0 0}	2 {2 2}	2 {2 2}	2 {2 2}
	Novel OTUs	15.7 {12 19}	7.9 {7 8}	8.9 {8 9}	17.4 {14 21}	12.4 {10 15}	11.9 {11 12}
	Supernovel OTUs	1.6 {0 3}	0.9 {0 1}	0 {0 0}	1.1 {0 3}	1 {1 1}	1.9 {1 2}

744

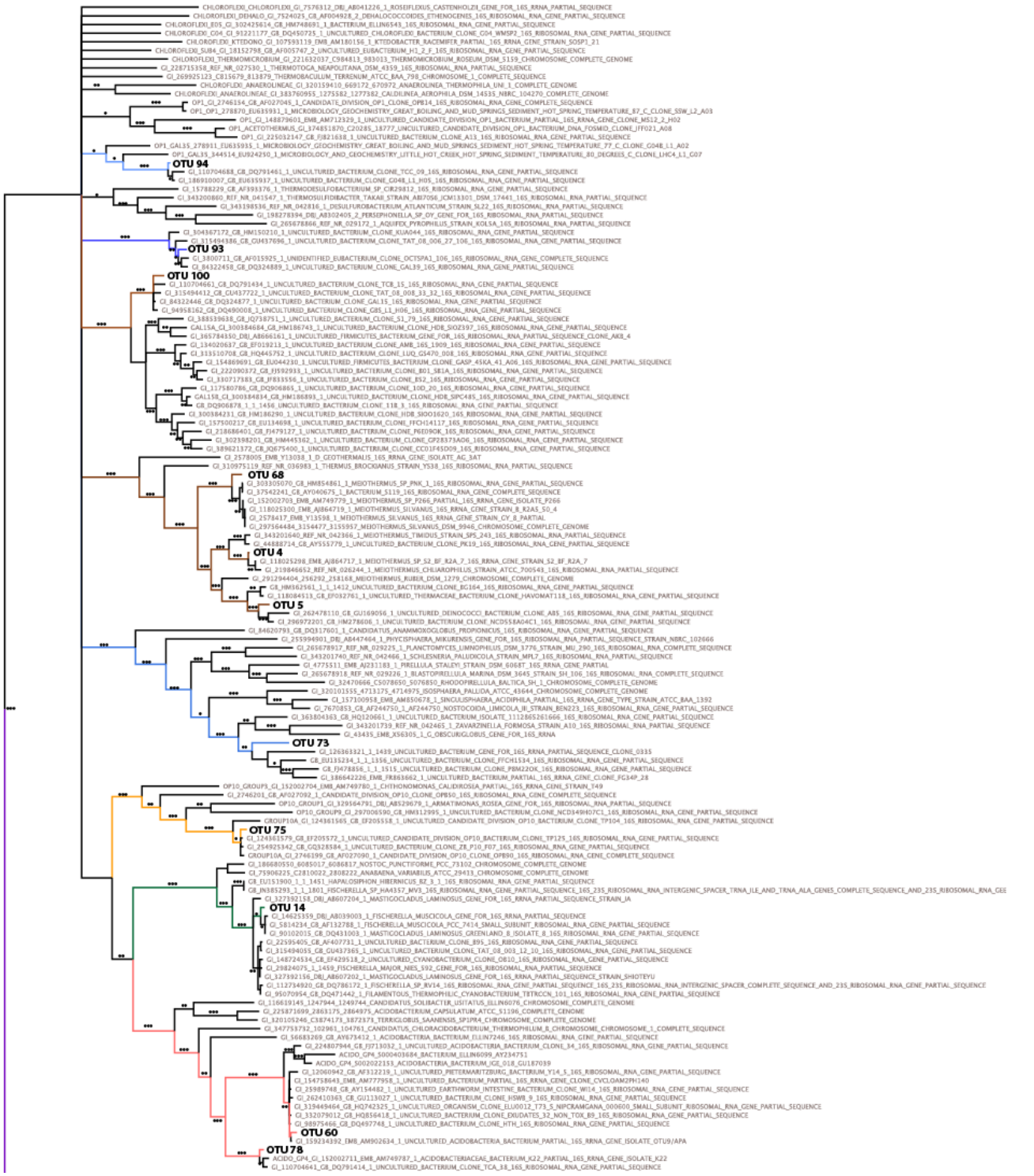
Supplementary Figure S1



Supplementary Figure S1

Principal Component Analysis of physicochemical data of samples from the current study. The percentage of variance captured by each principal component is denoted along each axis. Samples are plotted along Principal Component axes using sample name (site and depth). The correlations of factors along Principal Component axes are denoted with arrows. Blue arrows indicate factors that were used in the PCA. Purple arrows denote additional factors (extractable DNA and depth) that were mapped onto axes post-analysis.

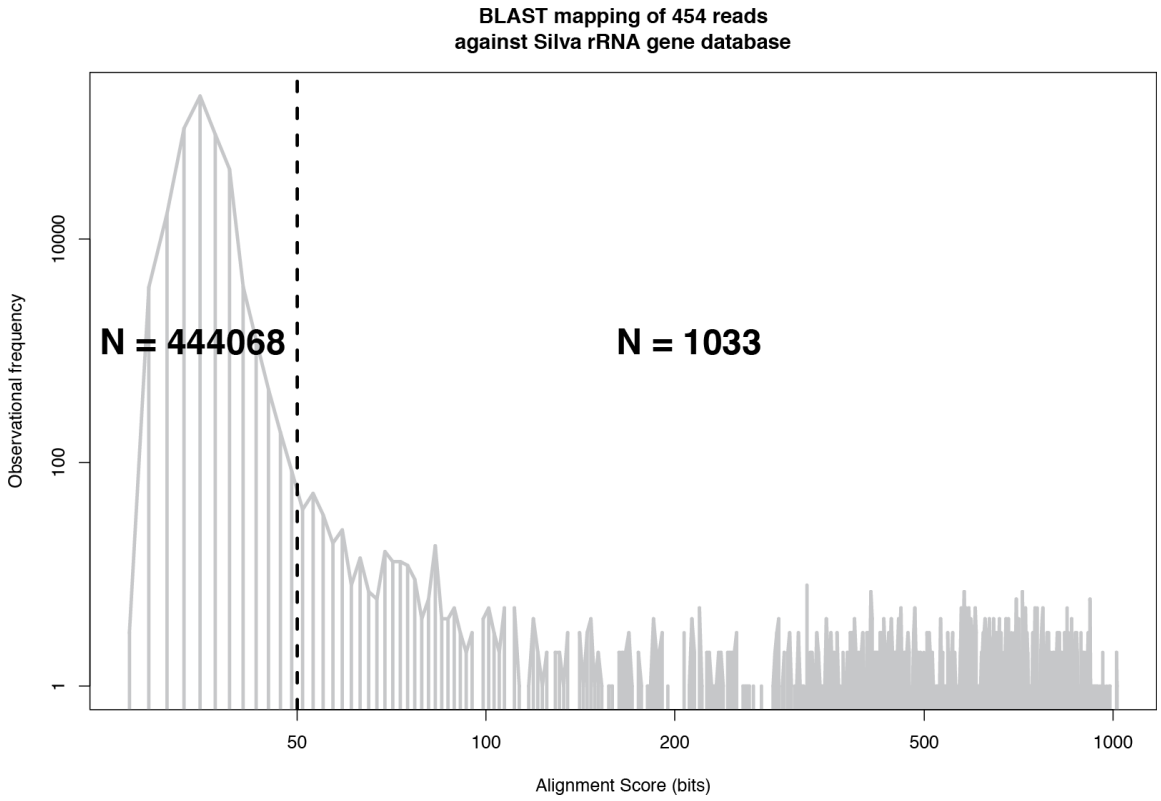
Supplementary Figure S2



Supplementary Figure S2

Bayesian consensus phylogenetic tree placing full-length sequences reconstructed from shot-gun sequencing in context with relatives. Bipartitions were drawn at the 80% consensus level from the posterior distribution generated from 20 independent Markov chains. ***: >99.9% consensus support. **: >99.0% consensus support. *: >90% consensus support. Branches leading to taxa for which full-length sequences were reconstructed are colored to match figure 1 in the main text.

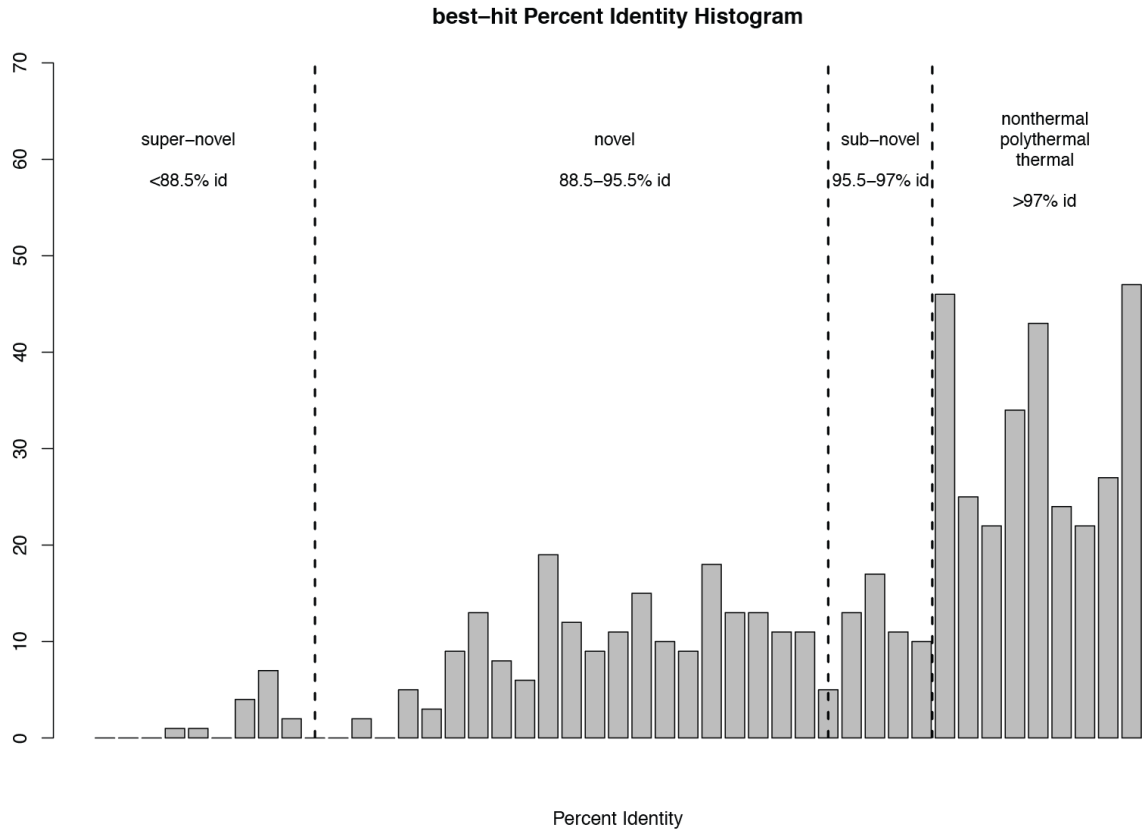
Supplementary Figure S3



Supplementary Figure S3

Histogram used to justify BLAST bit score cutoff. Each metagenomic read was used as a query in a BLAST search against the Silva SSU rRNA database. 1033 reads that hit the database with a bit score larger than 50 were assembled into 12 full-length sequences.

Supplementary Figure S4



Supplementary Figure S4

Histogram used for environmental classification of amplicons. Each unique read prediction was used as a BLAST query against the NCBI nr database. Environmental metadata were recovered for all hits that matched the query read with >97% identity and used to identify the read as either thermal or non-thermal depending on the type of environment the hits had been observed in. Reads were further identified as polythermal if the closest matches had been observed in both thermal and non-thermal datasets. Reads with no database matches showing >97% identity were classified as novel, sub-novel and super-novel based on level of percent identity with their best hits. These boundaries are depicted as dashed lines in the figure.

Supplementary Table S1

Classification and relative abundance of each OTU observed at Tramway Ridge.										
OTU	PhyloNeighborhood	ICG	Env. class	A (0-2cm)	A (2-4cm)	A (4-8cm)	B (0-2cm)	B (2-4cm)	B (4-8cm)	Full-length?
1	Thermoprotei	0	novel	0.04%	0.00%	0.00%	0.00%	0.05%	0.00%	N
2	Chloroflexi	0	novel	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	N
3	Thaumarchaeota	0	thermal	0.82%	0.15%	0.00%	0.09%	0.00%	0.00%	N
4	Meiothermus	1	thermal	0.38%	0.00%	0.91%	3.96%	0.02%	0.72%	Y
5	Meiothermus	1	thermal	1.43%	0.13%	0.77%	2.46%	0.00%	0.00%	Y
6	unclassified	0	novel	0.08%	0.00%	0.00%	0.31%	0.00%	0.00%	N
7	Armatimonadetes	0	novel	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	N
8	Armatimonadetes	0	thermal	1.14%	0.09%	0.00%	0.50%	0.47%	0.05%	N
9	Acidobacteria	0	novel	0.15%	0.00%	0.00%	0.01%	0.00%	0.00%	N
10	unclassified	0	novel	0.02%	0.00%	0.00%	0.01%	0.02%	0.00%	N
11	Firmicutes	0	novel	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	N
12	Firmicutes	0	novel	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
13	Firmicutes	0	novel	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	N
14	Cyanobacteria	1	thermal	4.18%	1.01%	13.28%	13.49%	0.18%	2.17%	Y
15	Chloroflexi	2	nonthermal	0.52%	0.35%	0.00%	2.13%	0.07%	0.56%	N
16	Actinobacteria	0	nonthermal	0.10%	0.00%	0.00%	0.03%	0.00%	0.00%	N
17	Actinobacteria	0	nonthermal	0.01%	0.00%	0.00%	0.01%	0.00%	0.00%	N
18	Actinobacteria	3	novel	0.99%	0.26%	0.00%	0.71%	0.56%	0.83%	N
19	Actinobacteria	0	polythermal	0.03%	0.00%	0.00%	0.01%	0.13%	0.40%	N
20	Actinobacteria	0	polythermal	0.21%	0.00%	0.00%	0.43%	0.05%	0.08%	N
21	Actinobacteria	0	thermal	0.01%	0.00%	0.00%	0.04%	0.01%	0.00%	N
22	Actinobacteria	0	nonthermal	0.98%	0.90%	0.10%	0.78%	0.17%	1.18%	N
23	Actinobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	N
24	Gemmatimonadetes	0	nonthermal	0.09%	0.00%	0.00%	0.26%	0.00%	0.11%	N
25	Gemmatimonadetes	0	sub-novel	0.07%	0.00%	0.00%	0.03%	0.21%	0.11%	N
26	Actinobacteria	0	sub-novel	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	N
27	Chloroflexi	0	nonthermal	0.02%	0.00%	0.00%	0.01%	0.00%	0.00%	N
28	Chloroflexi	2	thermal	0.12%	0.00%	0.00%	0.63%	0.01%	0.00%	N
29	AD3	0	sub-novel	0.00%	0.09%	0.00%	0.00%	0.00%	0.00%	N
30	unclassified	0	novel	0.01%	0.00%	0.00%	0.06%	0.00%	0.00%	N
31	Chloroflexi	0	novel	0.03%	0.00%	0.10%	0.02%	0.04%	0.16%	N
32	unclassified	0	novel	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	N
33	unclassified	0	novel	0.27%	0.29%	0.21%	0.81%	1.69%	4.06%	N
34	unclassified	0	novel	0.00%	0.00%	0.00%	0.00%	0.02%	0.08%	N
35	unclassified	0	thermal	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	N
36	unclassified	5	novel	0.23%	0.09%	0.84%	0.08%	1.19%	1.93%	N
37	Chloroflexi	0	thermal	0.11%	0.00%	0.00%	0.24%	0.06%	0.21%	N
38	Chloroflexi	5	thermal	0.07%	0.07%	1.01%	0.14%	1.41%	3.13%	N
39	Chloroflexi	3	novel	0.49%	0.15%	0.00%	0.51%	0.36%	5.08%	N
40	unclassified	0	super-novel	0.23%	0.00%	0.00%	0.24%	0.21%	0.64%	N
41	unclassified	0	super-novel	0.29%	0.07%	0.00%	0.16%	0.16%	0.64%	N
42	Nitrospirae	0	thermal	0.47%	0.15%	0.00%	0.10%	0.00%	0.00%	N
43	Nitrospirae	0	thermal	0.04%	0.00%	0.00%	0.03%	0.00%	0.00%	N
44	unclassified	5	novel	0.03%	0.00%	0.35%	0.03%	0.60%	0.43%	N
45	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	N
46	unclassified	0	novel	0.01%	0.00%	0.00%	0.00%	0.01%	0.00%	N
47	unclassified	0	novel	0.01%	0.00%	0.00%	0.00%	0.02%	0.13%	N
48	Chloroflexi	0	novel	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
49	OP1_GAL35	5	thermal	1.47%	1.89%	2.31%	0.35%	4.62%	6.68%	N
50	unclassified	4	polythermal	2.01%	1.34%	0.00%	0.13%	0.14%	2.73%	N
51	OP1_GAL35	4	novel	1.62%	1.01%	0.10%	0.25%	0.23%	1.95%	N
52	unclassified	0	novel	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	N
53	Planctomycetes	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
54	Proteobacteria	0	nonthermal	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	N
55	Chlorobi	2	novel	1.89%	0.00%	0.00%	1.06%	0.01%	0.00%	N
56	Bacteroidetes	0	sub-novel	0.01%	0.00%	0.00%	0.01%	0.00%	0.00%	N
57	Bacteroidetes	0	nonthermal	0.02%	0.00%	0.00%	0.59%	0.00%	0.00%	N
58	Bacteroidetes	1	nonthermal	0.18%	0.07%	1.15%	3.22%	0.01%	0.00%	N
59	SBR1093	2	polythermal	0.52%	0.22%	0.00%	1.74%	0.02%	0.08%	N
60	Acidobacteria	2	polythermal	1.07%	0.09%	0.00%	5.91%	0.04%	0.00%	Y
61	Cyanobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N

61	Cyanobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
62	Cyanobacteria	0	polythermal	0.04%	0.00%	4.44%	0.58%	0.00%	0.00%	N
63	unclassified	0	polythermal	0.04%	0.00%	0.00%	0.05%	0.03%	0.08%	N
64	Firmicutes	0	thermal	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	N
65	Proteobacteria	0	nonthermal	0.01%	0.00%	0.00%	0.23%	0.00%	0.00%	N
66	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	N
67	Proteobacteria	0	nonthermal	0.03%	0.00%	0.00%	0.02%	0.00%	0.00%	N
68	Meiothermus	2	thermal	11.42%	1.40%	1.47%	42.69%	0.47%	3.80%	Y
69	Proteobacteria	0	nonthermal	0.01%	0.00%	0.00%	0.09%	0.00%	0.00%	N
70	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
71	unclassified	0	polythermal	0.07%	0.02%	0.14%	0.45%	0.00%	0.00%	N
72	Acidobacteria	0	novel	0.01%	0.00%	0.17%	0.00%	0.00%	0.00%	N
73	Planctomycetes	4	novel	3.95%	3.47%	0.42%	0.86%	2.32%	1.60%	Y
74	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
75	Armatimonadetes	3	thermal	3.75%	0.90%	0.00%	4.31%	1.78%	2.65%	N
76	Acidobacteria	0	novel	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
77	Deinococcus	0	novel	0.00%	0.00%	0.00%	0.17%	0.00%	0.00%	N
78	unclassified	3	thermal	2.04%	0.42%	0.00%	2.83%	0.55%	7.57%	Y
79	Thaumarchaeota	0	nonthermal	0.05%	0.00%	0.00%	0.04%	0.00%	0.00%	N
80	Meiothermus	0	novel	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	N
81	Proteobacteria	0	polythermal	0.04%	0.00%	0.00%	0.20%	0.00%	0.00%	N
82	Actinobacteria	0	polythermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
83	Thermus	3	thermal	1.31%	0.57%	0.00%	0.85%	2.50%	8.98%	N
84	BH1	4	novel	4.62%	5.82%	0.98%	0.90%	2.85%	4.95%	N
85	Verrucomicrobia	0	nonthermal	0.29%	0.09%	0.00%	0.07%	0.00%	0.00%	N
86	unclassified	0	nonthermal	0.05%	0.00%	0.00%	0.23%	0.00%	0.24%	N
87	Planctomycetes	0	novel	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	N
88	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
89	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	N
90	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.26%	0.00%	0.00%	N
91	Armatimonadetes	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
92	Armatimonadetes	0	novel	0.00%	0.00%	0.00%	0.23%	0.00%	0.00%	N
93	OctSpA1-106	6	thermal	2.05%	5.64%	4.23%	0.25%	8.12%	1.50%	Y
94	OP1_GAL35	6	thermal	3.43%	15.65%	13.70%	0.69%	13.24%	5.64%	Y
95	Chloroflexi	0	novel	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	N
96	Thaumarchaeota-Like	6	novel	42.69%	56.37%	52.83%	1.53%	50.42%	26.44%	Y
97	Planctomycetes	0	sub-novel	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
98	Firmicutes	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
99	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	N
100	GAL15	0	thermal	1.52%	1.25%	0.49%	0.32%	4.90%	2.17%	Y
101	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
102	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.14%	0.00%	0.00%	N
103	Proteobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
104	SC4	0	nonthermal	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	N
105	TM6	0	nonthermal	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	N
106	Cyanobacteria	0	nonthermal	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	N
107	unclassified	0	novel	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	N
108	unclassified	0	thermal	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	N
109	unclassified	0	sub-novel	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	N
110	OP1_GAL35	0	thermal	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	N
111	Cyanobacteria	0	thermal	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	N

Supplementary Table S1

Classification and relative abundance of each OTU observed at the Tramway Ridge ASPA.

Supplementary Table S2

Classification and relative abundances of Full Length Sequences Reconstructed from Metagenomic data (Roche 454-Ti)												
	OTU ID	Metagenomic bases incorporated	Accession #	Site 1 (0-2 cm)	Site 1 (2-4 cm)	Site 1 (4-8 cm)	Site 3 (0-2 cm)	Site 3 (2-4 cm)	Site 3 (4-8 cm)	Best BLAST hit (GI num)	Best hit annotation(nr database)	Best hit % identity (nr database)
Meiothermus sp.	4	8337	KF923326	0.38%	0.00%	0.91%	3.96%	0.02%	0.72%	118025298	Meiothermus sp. S2-bf-R2A-7	98.55
Meiothermus sp.	5	7619	KF923323	1.43%	0.13%	0.77%	2.46%	0.00%	0.00%	302028075	Uncultured bacterium clone BG165	95.41
Mastigocladus laminosus	14	14613	KF923320	4.18%	1.01%	13.28%	13.49%	0.18%	2.17%	112734921	Fischerella sp. RV14	98.3
Acidobacteria	60	12138	KF923324	1.10%	0.10%	0.00%	5.90%	0.00%	0.00%	154758643	Uncultured bacterium partial 16S rRNA gene, clone CVCloAm2Ph140	98.79
Meiothermus silvanus	68	54463	KF923322	11.40%	1.40%	1.50%	42.70%	0.50%	3.80%	296848933	Meiothermus silvanus DSM 9946, complete genome	97.43
Planctomycetes	73	6128	KF923321	3.90%	3.50%	0.40%	0.90%	2.30%	1.60%	169412070	Uncultured bacterium clone ERB-D10 16S ribosomal RNA gene, partial sequence	99.72
Armatimonadetes group10A	75	6370	KF923317	3.70%	0.90%	0.00%	4.30%	1.80%	2.60%	124361579	Uncultured candidate division OP10 bacterium clone TP125 16S ribosomal RNA gene, partial sequence	98.49
Acidobacteria	78	5670	KF923318	2.00%	0.40%	0.00%	2.80%	0.50%	7.60%	152002711	Acidobacteriaceae bacterium K22 partial 16S rRNA gene, isolate K22	98.33
Candidate Division OctSpA106	93	5202	KF923319	2.10%	5.60%	4.20%	0.20%	8.10%	1.50%	3800711	Unidentified eubacterium clone OctSpA1-106 16S ribosomal RNA gene, complete sequence	96.07
Candidate Division OP1_GAL35	94	3328	KF923325	3.40%	15.70%	13.70%	0.70%	13.20%	5.60%	186910007	Uncultured bacterium clone G04b_L1_H05 16S ribosomal RNA gene, partial sequence	98.55
Candidate Division GAL15	100	2589	KF923327	1.50%	1.30%	0.50%	0.30%	4.90%	2.20%	84322446	Uncultured bacterium clone GAL15 16S ribosomal RNA gene, partial sequence	98.34
Novel Archaeon	96	11165	KF923316	42.70%	56.40%	52.80%	1.50%	50.40%	26.40%	14028778	SUBT-13	94.34
Cumulative abundance				77.79%	86.44%	88.05%	79.21%	81.91%	54.19%			
Maximum abundance of amplicon OTU not reflected in full-length SSU gene set				4.62%	5.82%	4.44%	4.31%	4.62%	8.98%			

Supplementary Table S2

Classification and relative abundances of near-full-length sequences reconstructed from metagenomic data (Roche 454-Ti)