

ORIGINAL ARTICLE

Environmental microarray analyses of Antarctic soil microbial communities

Etienne Yergeau^{1,6}, Sung A Schoondermark-Stolk^{1,7}, Eoin L Brodie², Sébastien Déjean³, Todd Z DeSantis², Olivier Gonçalves⁴, Yvette M Piceno², Gary L Andersen² and George A Kowalchuk^{1,5}

¹Netherlands Institute of Ecology (NIOO-KNAW), Centre for Terrestrial Ecology, Heteren, The Netherlands;

²Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA;

³Institut de mathématiques, Université Paul Sabatier, Toulouse, France; ⁴Laboratoire 'Microorganismes:

Génome et Environnement', UMR CNRS 6023, Université Blaise Pascal, Clermont-Ferrand II, France and

⁵Institute of Ecological Science, Free University of Amsterdam, Amsterdam, The Netherlands

Antarctic ecosystems are fascinating in their limited trophic complexity, with decomposition and nutrient cycling functions being dominated by microbial activities. Not only are Antarctic habitats exposed to extreme environmental conditions, the Antarctic Peninsula is also experiencing unequalled effects of global warming. Owing to their uniqueness and the potential impact of global warming on these pristine systems, there is considerable interest in determining the structure and function of microbial communities in the Antarctic. We therefore utilized a recently designed 16S rRNA gene microarray, the PhyloChip, which targets 8741 bacterial and archaeal taxa, to interrogate microbial communities inhabiting densely vegetated and bare fell-field soils along a latitudinal gradient ranging from 51 °S (Falkland Islands) to 72 °S (Coal Nunatak). Results indicated a clear decrease in diversity with increasing latitude, with the two southernmost sites harboring the most distinct Bacterial and Archaeal communities. The microarray approach proved more sensitive in detecting the breadth of microbial diversity than polymerase chain reaction-based bacterial 16S rRNA gene libraries of modest size (~190 clones per library). Furthermore, the relative signal intensities summed for phyla and families on the PhyloChip were significantly correlated with the relative occurrence of these taxa in clone libraries. PhyloChip data were also compared with functional gene microarray data obtained earlier, highlighting numerous significant relationships and providing evidence for a strong link between community composition and functional gene distribution in Antarctic soils. Integration of these PhyloChip data with other complementary methods provides an unprecedented understanding of the microbial diversity and community structure of terrestrial Antarctic habitats.

The ISME Journal (2009) 3, 340–351; doi:10.1038/ismej.2008.111; published online 20 November 2008

Subject Category: integrated genomics and post-genomics approaches in microbial ecology

Keywords: Antarctic soil ecosystems; GeoChip microarray; microbial community structure; microbial diversity; PhyloChip microarray

Introduction

Antarctic environments are extraordinary in the harshness of their climates, far more severe than northern climates at similar latitudes (Convey, 2001). Antarctic food webs are consequently rela-

tively simple, with a general absence of insect and mammalian herbivores (Davis, 1981; Heal and Block, 1987). Cold temperatures and low moisture availability are probably the main limiting factors responsible for the depauperate status of Antarctic habitats (Kennedy, 1996). The relatively simplified food-web structure of Antarctic terrestrial habitats provides a reasonably tractable system to disentangle the drivers of soil microbial activities and the consequences of system perturbation. Recent studies on the soils of this area have aimed to establish baseline knowledge of microbial community structure and function across a range of environments, and to assess the impacts of global warming (Lawley *et al.*, 2004; Brinkmann *et al.*, 2007; Bokhorst *et al.*, 2007a, 2008; Yergeau *et al.*, 2007a, b, c; Yergeau and Kowalchuk, 2008). Environmental conditions, such

Correspondence: GA Kowalchuk, Netherlands Institute of Ecology (NIOO-KNAW), Centre for Terrestrial Ecology, PO BOX 40, Heteren 6666 ZG, The Netherlands.

E-mail: g.kowalchuk@nioo.knaw.nl

⁶Present address: Biotechnology Research Institute, National Research Council of Canada, Montréal, Canada.

⁷Present address: Diakonessenhuis, Medical Microbiology and Immunology, Unit Molecular Diagnostics, Utrecht, The Netherlands.

Received 4 August 2008; revised 8 October 2008; accepted 9 October 2008; published online 20 November 2008

as temperature and freeze–thaw cycles, appear to have profound effects on soil microbial communities (Bokhorst *et al.*, 2007a; Yergeau and Kowalchuk, 2008). Bacterial diversity, community structure, abundance and functional gene density were all reported to be affected to different degrees by environmental conditions, most of the time in interaction with the type of aboveground cover (Yergeau *et al.*, 2007a, b, c).

The application of microarrays to study complex microbial communities is a relatively new practice, but the rapid increase in genetic databases (Cole *et al.*, 2005; DeSantis *et al.*, 2006) has facilitated the development of comprehensive platforms encompassing the known range of bacterial and archaeal diversity based on 16S rRNA gene sequences (for example, the PhyloChip, Brodie *et al.*, 2006; DeSantis *et al.*, 2007). The PhyloChip platform allows for the simultaneous detection of 8741 bacterial and archaeal taxa and has been shown to reveal a broader range of diversity than modestly sized 16S rRNA gene libraries for soil, water and aerosol samples (Brodie *et al.*, 2006, 2007; DeSantis *et al.*, 2007). However, it is not yet clear how such PhyloChip results might be compared with more traditional molecular methods like PCR-DGGE, T-RFLP, cloning–sequencing and quantitative PCR, or how such data can be integrated into studies of microbial community ecology. Furthermore, microarray platforms are highly dependent on the amount of information already known and cannot detect taxa that have not been described earlier in databases. Thus, it is imperative that such methods be tested across novel environments, such as the Antarctic soils examined in this study.

Most earlier reports about microbial communities in Antarctic soils have relied on relatively labor-intensive methods with low levels of taxonomic resolution. With the increasing interest in linking microbial identity and function, PhyloChip analyses also offer the opportunity to link microbial community composition with analyses of enzyme activity, density of functional gene families and the distribution of nutrient-cycle-related functional gene sequences. Thus, the aims of this study were: (1) to determine the suitability of 16S rRNA gene microarrays to monitor Antarctic soil bacteria and archaea, (2) to describe Antarctic soil-borne bacterial and archaeal communities using microarrays, thereby providing a more complete description of bacterial and archaeal diversity than possible earlier, (3) to relate bacterial and archaeal community composition and diversity to important environmental parameters and (4) to assess the feasibility of linking functional gene and 16S rRNA gene microarray data. To achieve these ends, the recently expanded PhyloChip of DeSantis *et al.* (2007) was used on PCR-amplified DNA directly extracted from soils sampled at five different sites ranging from the Falkland Islands (51°S) to Coal Nunatak (72°S), with a comparison of extensively vegetated patches

vs bare, fell-field environments. The resulting patterns of bacterial and archaeal community composition and diversity were compared with similar data recovered from clone libraries and real-time PCR assays and integrated into studies of function, including functional gene microarray analyses.

Materials and methods

Sampling sites

During the austral summer of 2003–2004, 2 × 2 m plots were established at the following sites (see Supplementary Figure S1 for a map): the Falklands Islands (cool temperate zone; 51°76'S 59°03'W), Signy Islands (South Orkney Islands, maritime Antarctic; 60°43'S, 45°38'W) and Anchorage Islands (near Rothera Research Station, western Antarctic Peninsula; 67°34'S, 68°08'W). At each location, two habitat types were selected for soil sampling: (1) 'vegetated', where dense vegetation cover was present with retention of underlying soil, and (2) 'fell-field', with rocky or gravel terrain and scarce vegetation or cryptogam coverage. Data with respect to vegetation cover within these environments were reported earlier (Bokhorst *et al.*, 2007b). Twelve plots were delineated per location, with half of the plots positioned over each soil type. The Falkland Islands fell-field habitat was not sufficiently extensive to allow for such a design, and nine of the twelve plots were therefore placed in the vegetated environment. Two additional sites were chosen for sampling, but without delineation of permanent plots. Six frost-sorted soil polygons at two different sites were sampled near Fossil Bluff (71°19'S, 68°18'W) and five adjacent polygons were sampled at Coal Nunatak (72°03'S, 68°31'W).

Soil samples

For molecular analyses, five 1-cm-diameter (from 2–3 to 15 cm deep, depending on the depth of soil per habitat) cores were sampled from each plot or polygon. They were frozen to –20 °C as soon as possible (within 24 h) and maintained at that temperature until further analysis. Material for soil analyses was collected from a 10-cm-diameter core taken directly adjacent to the established plots in order to minimize destructive sampling in the long-term plots. Sampling was conducted during 26–28 October 2004 for the Falkland Islands, during 2–3 January 2005 for Signy Islands, during 18–19 January 2005 for Anchorage Islands and during 22–23 February 2005 for Coal Nunatak and Fossil Bluff.

Nucleic acid extractions

DNA was extracted from 500 mg soil samples after bead-beating for 30 s at 50 ms⁻¹ in a hexadecyltrimethylammonium bromide (CTAB) buffer using a phenol–chloroform purification protocol as

detailed in Yergeau *et al.* (2007a). DNA extractions were performed separately for each of the five sub-samples taken per experimental plot. After PCR-DGGE analysis that confirmed low intra-plot variability (Yergeau *et al.*, 2007a), equal volumes of these five extractions were pooled to create the mixed environmental DNA used for further analysis.

Real-time PCR and PCR for microarray analyses

Real-time PCR quantifications for *Acidobacteria*, *Actinobacteria*, *Firmicutes*, *Alphaproteobacteria*, *Betaproteobacteria* and bacteria were performed using primers and cycling conditions described in Fierer *et al.* (2005). Real-time PCR quantifications were carried out on soil DNA using Absolute QPCR SYBR green mixes (AbGene, Epsom, UK) on a Rotor-Gene 3000 (Corbett Research, Sydney, Australia) as described earlier (Yergeau *et al.*, 2007a). Known template standards were made from plasmids containing previously characterized full-length 16S rRNA gene inserts. Samples and standards were assessed in at least two different runs to confirm the reproducibility of the quantification. Results of real-time PCR quantifications for the different phyla/classes are presented as a percentage of the total number of bacteria, by dividing the phyla/class 16S rRNA gene abundance by total bacterial 16S rRNA gene abundance.

PCR amplification for microarray hybridization was carried out using a bacterial-specific 16S rRNA gene primer set (27f.1 and 1492R, DeSantis *et al.*, 2007) and an archaeal-specific 16S rRNA gene primer set (4fa (5'-TCCGGTTGATCCTGCCRG-3') and 1492R). Both for bacteria and archaea, four independent PCRs were performed in a C1000 thermocycler (BioRad, CA, USA) with annealing temperatures of 48, 51.9, 54.4 and 58 °C and pooled. For reagent composition of PCR mixture and detailed PCR conditions, see DeSantis *et al.* (2007). PCR products originating from each location-vegetation type combination were further pooled together to provide one sample per location-vegetation type combination to give a total of 12 representative samples. These pooled samples were then concentrated to a volume of <40 µl with a Microcon YM100 spin filter (Millipore, Billerica, MA, U.S.S.S.A.). Five hundred nanograms of bacterial PCR product and 100 ng of archaeal PCR product were subsequently used per microarray hybridization.

PhyloChip processing, scanning and probe set scoring

Most samples were assessed on two independent chips (technical duplicates), with the exception of Fossil Bluff and Coal Nunatak samples, for which only one hybridization worked satisfactorily. Duplicates were processed separately, and the resulting data were combined. The pooled PCR products of

each sample were spiked with known concentrations of amplicons derived from yeast and bacterial metabolic genes. This mix was fragmented to 50–200 bp using DNase I (0.02 U µg⁻¹ DNA, Invitrogen, Carlsbad, CA, USA) and One-Phor-All buffer (GE Healthcare, Piscataway, NJ, USA) following the manufacturer's protocols. The mixture was then incubated at 25 °C for 20 min and 98 °C for 10 min before biotin labeling with a GeneChip DNA labeling reagent kit (Affymetrix, Santa Clara, CA, USA) following the manufacturer's instructions. Next, the labeled DNA was denatured at 99 °C for 5 min and hybridized to custom-made Affymetrix GeneChips (16S rRNA genes PhyloChips) at 48 °C and 60 rpm for 16 h. PhyloChip washing and staining were performed according to the standard Affymetrix protocols described by Masuda and Church (2002).

Each PhyloChip was scanned and recorded as a pixel image, and initial data acquisition and intensity determination were performed using standard Affymetrix software (GeneChip microarray analysis suite, version 5.1). Background subtraction, data normalization and probe pair scoring were done essentially as reported earlier (Brodie *et al.*, 2006; DeSantis *et al.*, 2007). The positive fraction (PosFrac) was calculated for each probe set as the number of positive probe pairs divided by the total number of probe pairs in a probe set. Taxa were deemed present when the PosFrac value exceeded 0.92. We used the resulting binary data (taxon presence-absence) directly after this step for some analyses. Technical replicate data were merged by considering a taxon as present if one or both technical replicates identified it as present, and averaging the intensity values of the present taxa. For relative abundance analyses, relative taxon signals were calculated by dividing the average signal of the probes targeting a given taxon by the total average signal for all the taxa identified as present. The relative abundance of taxa whose PosFrac did not exceed 0.92 was set to zero (since these taxa are scored as absent). Relative abundance values represent the fraction of the summed intensity that is because of a single taxon. These data were used directly for single taxon-level analyses and summed up to the phylum or family level for other analyses.

Other data

Complete soil analyses and detailed soil biological characterization of the sites are available in Yergeau *et al.* (2007a). For the analyses presented in this study, we used the following soil parameters: NH₄⁺, NO₃⁻ and total N concentrations, pH, C:N ratio, soil water and organic matter content. All other molecular data were collected from the same DNA extracts used for the PhyloChip analyses. Clone library data consisted of 192 16S rRNA gene sequences per sample, and a detailed description of this dataset is presented in Yergeau *et al.* (2007c).

Functional gene microarray analyses were carried out using the GeoChip (He *et al.*, 2007) and are presented in Yergeau *et al.* (2007b).

Statistical analyses

Mantel tests were based on Mantel's r (r_m) with 999 permutations and were performed in P Legendre's statistical software (Casgrain and Legendre, 2001). The choice of similarity indices for the different datasets followed the rationale outlined in Legendre and Legendre (1998): Steinhaus similarity (one-complement of Bray–Curtis distance) for taxon relative abundance, Jaccard similarity for the presence or absence of taxon and Gower similarity for soil data. Principal coordinates analyses (PCoA) were carried out with taxon relative abundance data in P Legendre's statistical software, whereas the phylum and functional gene information was entered in ordination graphs as supplementary variables, that is, variables that did not interfere in the calculations. The effects of the location and plant cover on the community structure as determined by phylogenetic microarray analysis were tested by distance-based redundancy analysis (db-RDA) (Legendre and Anderson, 1999) with 999 permutations in Canoco 4.5 for Windows (ter Braak and Šmilauer, 2002). Canonical correspondence analyses were also carried out in Canoco as follows: relative taxon abundance was used as 'species' data, whereas soil and environmental data were included in the analysis as 'environmental' variables. All correlation analyses (Pearson r or Spearman r_s) were carried out in Statistica 7.0 (StatSoft Inc., Tulsa, OK, USA). Correlations were considered significant at a $P < 0.05$ baseline and to be nearly significant at $0.05 < P < 0.10$. GeoChip and PhyloChip datasets were related to each other using regularized canonical correlation analyses (RCCorA) in the R package (González *et al.*, 2008). The strongest associations in the resulting graphs were identified by calculating Bray–Curtis distance and Pearson's linear correlation between functional genes and taxa.

Results

Community structure and phyla–sites association

Of the 8741 taxa represented on the PhyloChip, 616 were detected in at least one sample across our different study sites. Between 106 and 427 taxa were detected per sample, with 87 taxa being common to all samples. Principal coordinate analysis (PCoA) of the community composition at the taxon level (using relative intensity data) showed a clear separation of sites, mainly between Fossil Bluff and Coal Nunatak and the other sites (Figure 1). These last two sites were separated from all the other sites on the first ordination axis, which explained a large part of the total variation (65.5%). The second ordination axis explained much less variation (12.8%), separating the two

Falkland Islands environments from the Signy and Anchorage environments. Vegetated environments from Signy and Anchorage Islands are grouped together in the ordination. Superimposition of the summed phylum-level data (Supplementary Table S1) over the ordination of the sampling sites allowed for visualization of the association of these phyla with particular sites. Some phyla showed relatively higher presence at some sites (Supplementary Table S1 and Figure 1). For instance, *Alphaproteobacteria*, *Bacteroidetes* and *Firmicutes* were present more at Fossil Bluff and Coal Nunatak. *Cyanobacteria* were also, in general, present more in these southernmost sites. *Actinobacteria* were present more in the Falkland Islands than in any other site, whereas the remaining taxa were less abundant in Fossil Bluff and Coal Nunatak samples.

Community structure in relation to soil factors, location and presence of vegetation

Distance-based redundancy analysis (db-RDA) of the relative abundance data at the taxon level showed a significant effect of location on community structure ($P = 0.0090$). No such effect was observed with respect to vegetation cover ($P = 0.2940$). Similar analyses using presence–absence (binary) data showed the same relationships and similar significance levels. Mantel tests were also performed to examine whether soil factors or geographical distance between sites was significantly correlated with similarity in community composition. Geographical distance had a nearly significant effect on community composition ($r_m = -0.204$, $P = 0.0940$). Similarity in soil physico-chemical characteristics was not related to similarity in bacterial and archaeal community composition ($r_m = 0.00429$, $P = 0.4240$). Canonical correspondence analyses were used to highlight the effect of individual soil or environmental factors on the community structure at the taxon level. Using the relative intensity data, latitude ($P = 0.0280$) and pH ($P = 0.0080$) were the only two factors chosen by forward selection. Altogether, they formed a model that significantly explained the taxon–environmental variables relationships ($P = 0.0010$). Correlation analyses were used to characterize further the association of phyla/classes relative abundance with specific soil factors. *Chloroflexi* and *Betaproteobacteria* were negatively and significantly correlated to soil pH, whereas *Firmicutes* and *Verrucomicrobia* showed significant positive correlations with soil pH. *Chloroflexi* and *Planctomycetes* were positively and significantly correlated with soil water and organic matter content. *Betaproteobacteria* and *Crenarchaeota* showed nearly significant ($0.05 < P < 0.10$) positive correlations with soil NH_4^+ concentration. Furthermore, *Betaproteobacteria* was significantly and positively correlated with soil NO_3^- and total N. *Planctomycetes* and *Chloroflexi* also showed significant positive correlations with soil total N. *Actinobacteria*, *Euryarchaeota*, *Epsilonproteobacteria*

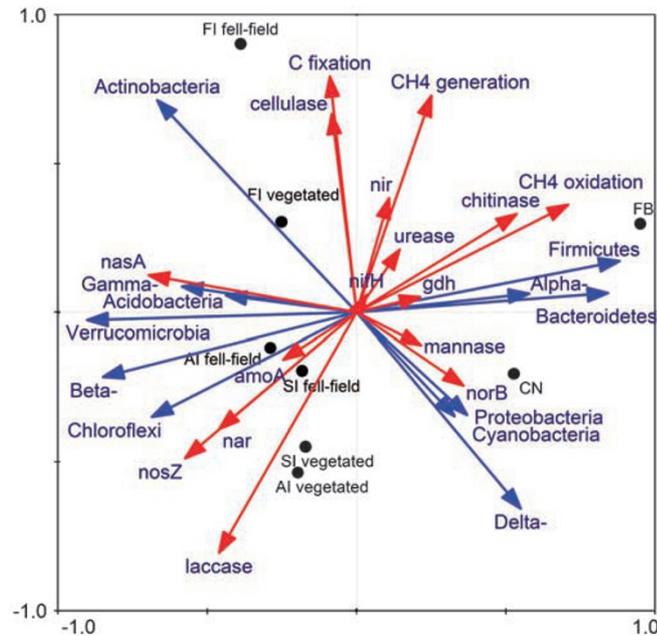


Figure 1 Principal coordinates analysis (PCoA) ordination based on Steinhaus similarity of the relative abundances of taxa detected in Falkland (FI), Signy (SI) and Anchorage (AI) Islands and Fossil Bluff (FB) and Coal Nunatak (CN) soil samples. Phylum/class data (blue arrows) and functional gene categories (red arrows) were added to the graph as supplementary variables (not involved in calculation) to show the relative repartition of phyla/classes and functional gene categories at the different sites. See Supplementary Table S1 for more details about the relative abundance of taxa for the different sites. Axis 1 = 65.5%, axis 2 = 12.8%.

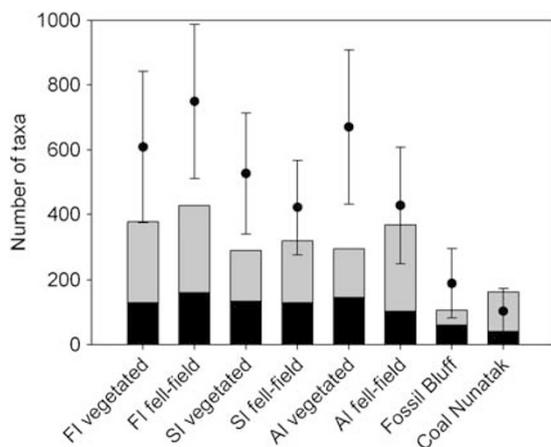


Figure 2 Number of taxa retrieved in Falkland (FI), Signy (SI) and Anchorage (AI) Islands and Fossil Bluff and Coal Nunatak soil samples estimated from cloning-sequencing using a 97% sequence identity cut-off (in black, from Yergeau *et al.*, 2007c) or PhyloChip-based identification (black + gray) compared with the estimated richness (Chao 1) calculated from clone libraries (black dots). The error bars show 95% confidence intervals of the Chao 1 index.

and *Verrucomicrobia* decreased significantly with increasing latitude, whereas *Cyanobacteria* increased with increasing latitude.

Community composition compared with clone libraries and real-time PCR

The relative signal intensity on the PhyloChip was summed at the phylum or family level and

compared with similar data from clone libraries (number of clones associated with a taxa/total number of clones) and from real-time PCR (number of 16S rRNA genes for a taxa/total number of bacterial 16S rRNA genes). At the phylum level (Supplementary Table S1), the PhyloChip and the clone libraries were significantly correlated ($r_s = 0.515$, $P < 0.0001$, $N = 120$), as were the clone library and the real-time PCR data ($r_s = 0.660$, $P < 0.0001$, $N = 36$), but no significant correlation was found between the PhyloChip and the real-time PCR data ($r_s = 0.032$, $P = 0.855$, $N = 36$). When comparing data from each sampling site/vegetation cover combination separately, the PhyloChip and the clone library data were still significantly correlated (r_s from 0.541 to 0.716 and P from 0.003 to 0.037, $N = 15$). At the family level, it was only possible to compare the PhyloChip with the clone libraries, and that comparison also yielded highly significant results ($r_s = 0.212$, $P < 0.0001$, $N = 344$).

Number of taxa and families detected

The number of taxa detected using the PhyloChip was significantly and inversely correlated with latitude ($r = -0.791$, $P = 0.019$, $N = 8$). The number of taxa detected by the PhyloChip was also significantly correlated with the taxon numbers recovered in the clone libraries ($r = 0.835$, $P = 0.010$, $N = 8$), as well as the Chao 1 richness estimations calculated from these data ($r = 0.832$, $P = 0.010$, $N = 8$). However, the number of different taxa

Table 1 Number of families uniquely detected in clone libraries and on the PhyloChip and shared families that were detected using both methods in soil samples from Falkland, Signy and Anchorage Islands, Fossil Bluff and Coal Nunatak

	Clone only (% of total)	Shared (% of total)	PhyloChip only (% of total)	Total
FI vegetated	2 (1.5)	14 (10.5)	117 (88.0)	133
FI fell-field	3 (2.2)	20 (14.9)	111 (82.8)	134
SI vegetated	8 (6.7)	10 (8.4)	101 (84.9)	119
SI fell-field	8 (6.8)	13 (11.1)	96 (82.1)	117
AI vegetated	10 (8.5)	17 (14.5)	90 (76.9)	117
AI fell-field	4 (3.2)	11 (8.9)	109 (87.9)	124
FB	16 (27.1)	3 (5.1)	40 (67.8)	59
CN	8 (11.4)	3 (4.3)	59 (84.3)	70
All	12 (6.8)	39 (22.2)	125 (71.0)	176

Abbreviations: AI, Anchorage Island; CN, Coal Nunatak; FB, Fossil Bluff; FI, Falkland Islands; SI, Signy Island.

detected using the PhyloChip was generally much higher than in the clone libraries (Figure 2). Although Chao 1 estimates were higher for a majority of samples, the taxon numbers detected on the PhyloChip were within the 95% confidence interval of the estimated richness (Chao1) calculated from the clone libraries (Figure 2). Similarly, the number of families detected was always higher on the PhyloChip than in the clone libraries (Table 1). However, for Fossil Bluff and Coal Nunatak soils, a relatively higher proportion of families unique to the clone libraries was present, compared with the other soils. The pooling of technical replicates increased the number of taxa detected in Falkland, Signy and Anchorage Islands samples by an average of 56 taxa (with a range of 33–77 taxa). On an average, the Fossil Bluff and Coal Nunatak samples harbored 171 and 116 fewer taxa, respectively, than other samples (before pooling). Therefore, the fact that these two samples did not have technical replication most probably did not influence the trends in diversity observed here.

Relationship between functional gene and 16S rRNA gene microarray data

To determine whether phylogenetic community structure, based on the PhyloChip analysis, was related to the distribution of microbial genes involved in nutrient cycling, we compared PhyloChip data with those gathered earlier from the same sites with the GeoChip (Yergeau *et al.*, 2007b). High-order functional and taxonomic information was first used to determine the general trends in the datasets. A simplified representation of the relationships is shown in Figure 1. Noteworthy associations observed in this figure include: chitinase and mannanase—*Bacteroidetes*; CH₄-oxidation genes—*Alphaproteobacteria*, and cellulase—*Actinobacteria*. Furthermore, Mantel tests showed that communities with more similar taxon compositions were also more closely related in their functional genes. Significant correlations were found between the similarities calculated from the taxon relative abundance and the relative abun-

dance of functional genes related to the N-cycle ($r_m = 0.745$, $P = 0.0050$), C-cycle ($r_m = 0.677$, $P = 0.0220$) and CH₄ transformations ($r_m = 0.887$, $P = 0.0010$).

To gain further insight into the relationships between environments, functional genes and taxa, RCCorA were performed. When using all the sampling sites, a clear dichotomy between Fossil Bluff and Coal Nunatak vs all other sites appeared (similar to Figure 1). To gain more detailed insight into taxon-functional genes–environment relationships, additional RCCorA analyses were performed excluding these two sites (Figure 3). Positive relationships can be visualized as the proximity of functional genes and taxa, and the most significant relationships are normally further away from the origin, thus outside the small, central circle in Figure 3. The top panel of Figure 3 is useful to identify sites for which particular functional genes–taxa relationships are strongest. This can be done by comparing the relative position of the sites in the top panel with the position of the taxa and functional genes in the bottom panel. As a large number of positive relationships are visible in Figure 3, two indices were used to identify the strongest relationships (Table 2). Bray–Curtis distances identified functional gene–taxa pairs that occurred at similar relative abundances in different samples, discarding double absences, whereas Pearson's correlation identified linear correlations between genes and taxa, including double-absence data. Depending on the region where they occurred in Figure 3, associations were arbitrarily categorized into nine groups (circled with letters A–J) to facilitate discussion (Table 2). By comparing the top and bottom panels of Figure 3, these groups could be associated to different environments: A to Falkland Islands fell-field sites, B to both Falkland Islands environments, C to Falkland Islands vegetated environments and Signy and Anchorage Islands fell-field environments, D and E to Signy and Anchorage Islands fell-field environments, F and H to Signy and Anchorage Islands vegetated environments, G to Anchorage Islands vegetated sites and J to Signy Islands vegetated sites (Figure 3).

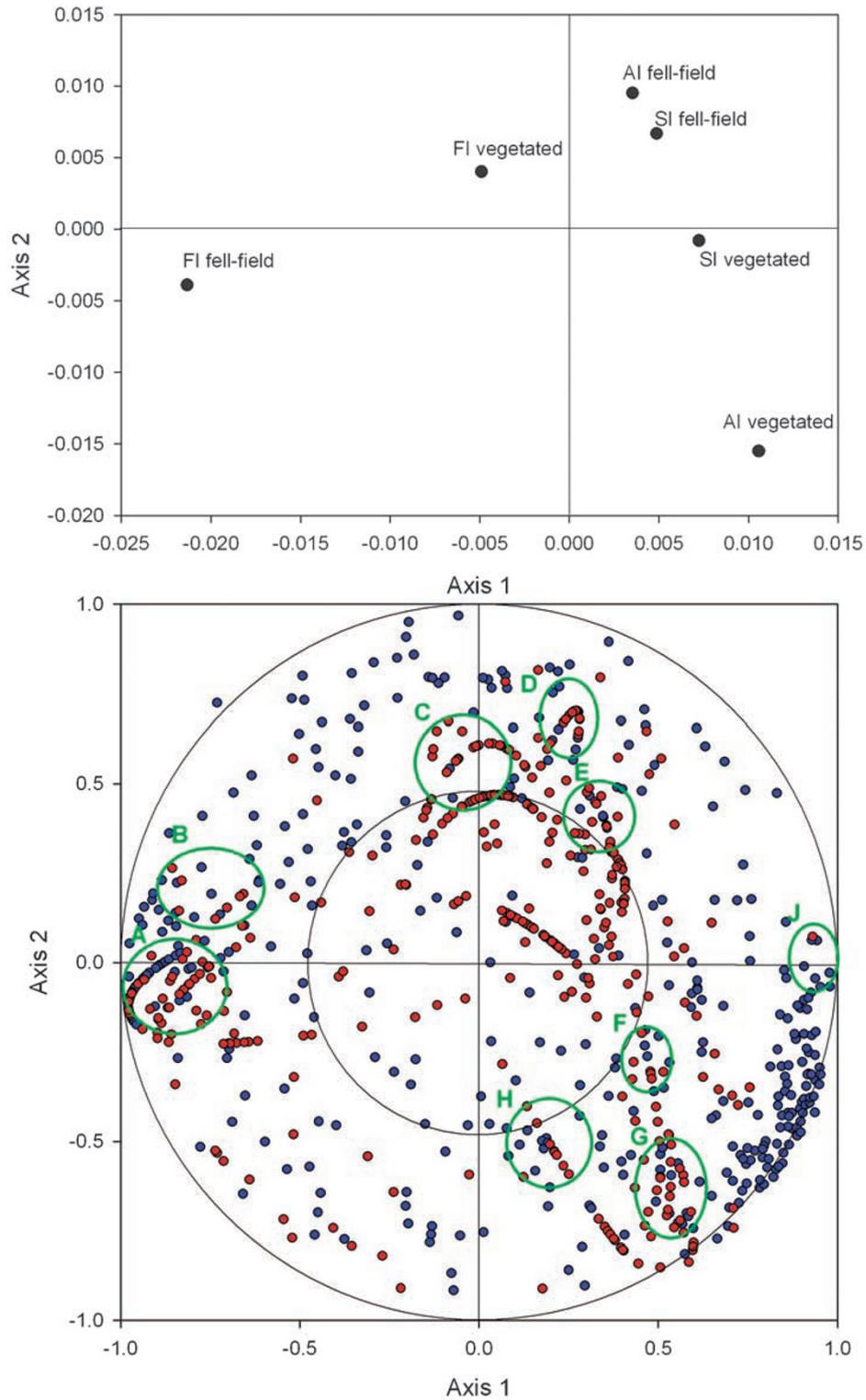


Figure 3 Regularized canonical correlation analysis of functional gene–taxon relationships for Falkland, Signy and Anchorage Islands. The upper panel of the figure depicts relationships between the different sampling sites, and the lower panel shows relationships between functional genes and taxa. Coordinates used to plot genes and taxa in the lower panel are correlation coefficients (between initial variables and canonical variables) and, consequently, data points are normally inside a circle of radius 1. A smaller circle (of radius 0.5) was added to highlight the most relevant relationships, which should occur in the zone between the two circles. Genes and taxa that were found to be most highly correlated to each other were further circled in green, lettered and reported in Table 2. Red dots: functional genes, Blue dots: taxa.

Table 2 Twenty-five highest Pearson correlations and twenty-five lowest Bray–Curtis distances between the relative abundance of individual genes from the GeoChip and individual taxa from the PhyloChip for samples from Falkland, Signy and Anchorage Islands

Rank	Bray–Curtis				Correlation			
	Distance	GeoChip	PhyloChip	Group	r	GeoChip	PhyloChip	Group
1	0.005	<i>nirK</i>	<i>Euryarcheota</i>	A	>0.999	<i>nifH</i>	<i>Actinobacteria</i>	E
2	0.006	<i>mmoA</i>	<i>Alphaproteo.</i>	E	>0.999	<i>gdh</i>	<i>Actinobacteria</i>	A
3	0.014	<i>mcrA</i>	<i>Actinobacteria</i>	G	>0.999	Laccase	<i>Euryarchaeota</i>	A
4	0.016	<i>amoA</i>	<i>Betaproteo.</i>	D	>0.999	Unknown	<i>Firmicutes</i>	A
5	0.019	<i>nifH</i>	<i>Deltaproteo.</i>	E	>0.999	<i>nifH</i>	<i>Bacteroidetes</i>	F
6	0.022	Chitinase	<i>Bacteroidetes</i>	B	>0.999	Chitinase	<i>Acidobacteria</i>	A
7	0.023	<i>narG</i>	<i>Actinobacteria</i>	A	>0.999	<i>gdh</i>	<i>Chloroflexi</i>	A
8	0.028	<i>amoA</i>	<i>Acidobacteria</i>	D	>0.999	Cellulase	<i>Bacteroidetes</i>	D
9	0.028	Urease	Unclassified	C	>0.999	<i>amoA</i>	<i>Gammaproteo.</i>	C
10	0.029	<i>nirS</i>	<i>Gammaproteo.</i>	F	>0.999	<i>nirK</i>	Unclassified	A
11	0.030	<i>narG</i>	<i>Actinobacteria</i>	A	>0.999	Chitinase	<i>Acidobacteria</i>	A
12	0.033	Laccase	<i>Bacteroidetes</i>	F	>0.999	Laccase	<i>Actinobacteria</i>	A
13	0.034	Polygalac.	<i>Gammaproteo.</i>	C	>0.999	Laccase	<i>Bacteroidetes</i>	F
14	0.036	Chitinase	unclassified	D	>0.999	Chitinase	<i>Acidobacteria</i>	A
15	0.037	<i>nifH</i>	<i>Actinobacteria</i>	G	>0.999	C-fixation	<i>Bacteroidetes</i>	D
16	0.038	Laccase	<i>Betaproteo.</i>	D	>0.999	Polygalac.	<i>Actinobacteria</i>	A
17	0.043	C-fixation	<i>Alphaproteo.</i>	C	>0.999	Cellulase	<i>Actinobacteria</i>	G
18	0.044	Urease	<i>Bacteroidetes</i>	na	>0.999	Urease	<i>Alphaproteo.</i>	D
19	0.047	Cellulase	<i>Actinobacteria</i>	H	>0.999	C-fixation	<i>Gammaproteo.</i>	A
20	0.047	Urease	<i>Firmicutes</i>	na	>0.999	Laccase	unclassified	A
21	0.047	Mannanase	<i>Actinobacteria</i>	B	>0.999	Chitinase	<i>Gammaproteo.</i>	B
22	0.047	Cellulase	<i>Actinobacteria</i>	D	>0.999	<i>nifH</i>	<i>Deltaproteo.</i>	E
23	0.049	Laccase	<i>Acidobacteria</i>	D	>0.999	Cellulase	<i>Firmicutes</i>	A
24	0.049	<i>nosZ</i>	<i>Chloroflexi</i>	J	>0.999	Cellulase	<i>Alphaproteo.</i>	E
25	0.050	Cellulase	<i>Deltaproteo.</i>	E	>0.999	<i>amoA</i>	<i>Gammaproteo.</i>	C

Groups refer to the groups identified in Figure 3.

Discussion

Distribution of bacterial and archaeal taxa in Antarctic soils

High-quality hybridization patterns were observed across all study sites, including low-biomass samples, such as Fossil Bluff and Coal Nunatak. However, the number of bacterial and archaeal taxa detected in individual samples was lower than that reported earlier for PhyloChip analyses of temperate soil environments (Brodie *et al.*, 2006; DeSantis *et al.*, 2007). The number of bacterial and archaeal taxa detected on the PhyloChip significantly decreased with increasing latitude, with a large reduction in the southernmost sites (Fossil Bluff and Coal Nunatak). This pattern agrees well with diversity estimates based on 16S rRNA gene libraries (Yergeau *et al.*, 2007c), as well as reported decreases in the diversity of other Antarctic terrestrial organisms with increasing latitude (Smith, 1992; Wynn-Williams, 1996; Sohlenius and Bostrom, 2005; Peat *et al.*, 2006). This pattern is thought to be related not only to decreases in temperature at higher latitudes, but also to concomitant decreases in water and nutrient availability (Kennedy, 1993). Interestingly, studies of northern latitudinal gradients have not shown such latitudinal patterns in bacterial diversity, suggesting that other environmental factors were more important in steering soil

bacterial diversity (Neufeld and Mohn, 2005; Fierer and Jackson, 2006). Decreasing biodiversity with latitude is one of ecology's most fundamental patterns (Willig *et al.*, 2003), and it would be interesting to examine more closely whether these observations are indeed indicative of true differences in general patterns of microbial diversity between the southern and northern hemispheres.

The southernmost sites (Fossil Bluff and Coal Nunatak) were clearly distinct from all other sites (Figure 1) with respect to bacterial and archaeal community composition. The main influencing factor in the community composition dataset was latitude or location, as confirmed by multivariate tests (db-RDA and Mantel tests). This dichotomy between the southernmost sites and the other study sites was also observed in earlier studies along comparable Antarctic latitudinal gradients for soil bacterial community composition (using PCR-DGGE and PLFA), abundance (using real-time PCR, PLFA and CFU counts), diversity (using cloning–sequencing) and functional gene distribution (using functional gene microarrays and real-time PCR) (Yergeau *et al.*, 2007a,b,c). Similarly, PhyloChip results revealed that several phyla (for example, *Actinobacteria*, *Cyanobacteria*, *Epsilonproteobacteria*, *Euryarcheota* and *Verrucomicrobia*) were correlated with latitude, being associated with either the northernmost or southernmost sites.

The general presence of vegetation did not exert a significant direct effect on community structure as determined by PhyloChip analysis at the taxon level. This might be related to the fact that the vegetation cover and environmental conditions at the Falkland Islands are quite different than the ones found at Signy and Anchorage Islands (Bokhorst *et al.*, 2007b). Indeed, similarly vegetated environments from Anchorage and Signy Islands grouped together in ordinations, separate from the vegetated environments from the Falkland Islands (Figures 1 and 3), suggesting a location-dependant vegetation effect on community structure. Although vegetation in general was not having a strong effect on total community composition, some phyla (for example, *Chloroflexi* and *Planctomycetes*) were positively correlated to vegetation-related soil factors like soil water and organic matter content.

Combining data from functional gene and 16S rRNA gene microarray analyses

To the best of our knowledge, this study is the first attempt to combine functional gene and 16S rRNA gene microarray data. The novel use of Mantel tests, ordination and RCorA allowed several interesting relationships to be gleaned from the data, including a relationship between *Bacteroidetes* and decomposition-related genes like chitinase and mannanase, which seemed to be associated with the Fossil Bluff and Coal Nunatak environments. Members of this phylum are recognized for their ability to degrade polymers (Multiple Authors, 2006) and were found frequently in Antarctic clone libraries, particularly in the most extreme, bare, nutrient-poor soils (Aislabie *et al.*, 2006; Yergeau *et al.*, 2007c). CH₄-related genes and *Alphaproteobacteria* also followed similar distribution patterns, being most prevalent at the Fossil Bluff and Coal Nunatak sites. Interestingly, clone libraries indicated that these two sites were dominated by pink-pigmented methylo-trophic bacteria from the genus *Methylobacterium*, members of the *Alphaproteobacteria* (Yergeau *et al.*, 2007c). Using Mantel tests, significant correlations were observed between the site similarity on the basis of taxon relative abundance from the PhyloChip and the similarity based on C- or N-cycle gene relative abundances from the GeoChip. This supports the notion that the functional genes detected in soils are strongly linked to community composition as determined by 16S rRNA gene-based analyses.

RCorA was used to obtain more detailed information about associations between particular phylogenetic taxa and functional genes. At this level of analysis, the amount of information involved precludes succinct, yet comprehensive, interpretation of all the data. Two different indices were therefore used to calculate association strengths and highlight the most significant relationships. Using this approach, three different taxa belonging to the

Actinobacteria were found to be associated with cellulase genes (Table 2, Bray–Curtis ranks 19 and 22; correlation rank 17), and each of these associations occurred across a range of environments (mainly on Signy and Anchorage Islands, groups D, G, H; see Figure 3). Several members of the *Actinobacteria* are indeed known to be able to degrade cellulose, and these data suggest that different Actinobacterial taxa may be involved in this process depending on the environment.

For the cases discussed above, it is likely that the coupled PhyloChip and GeoChip signals were derived from the same microbial populations. However, for several associations this was clearly not the case. For instance, *Bacteroidetes* taxa were related to different genes involved in decomposition, (Table 2, Bray–Curtis ranks 6 and 12; correlation ranks 8 and 13), which is in agreement with the high-level information reported above. However, these associations were probably not indicating that all these genes were found in *Bacteroidetes* taxa, especially in the case of laccase, which is almost exclusively found in fungi and plants (Mayer and Staples, 2002). Similarly, it is well established that the *Gammaproteobacteria* and *Acidobacteria* do not contain terrestrial ammonia-oxidizing bacteria (Kowalchuk and Stephen, 2001); yet, taxa from these two phyla were strongly associated with different bacterial *amoA* genes. Apparently, there is some overlap in the demonstrated environmental preferences of different ammonia oxidizer species (Kowalchuk *et al.*, 2000) and members of these two unrelated groups.

Phylogenetic inferences can often be drawn from the probes incorporated into the GeoChip, but functional gene phylogeny does not always match the 16S rRNA gene phylogeny, thereby hampering complete genetic comparison across these two microarray platforms. Even with such restrictions, the statistical methods used allowed us to link these different datasets at the functional gene/taxon level.

Comparison of the PhyloChip data with other molecular microbial community data

The PhyloChip has already been compared with clone library data from soil as a proof of concept (DeSantis *et al.*, 2007) or with a limited number of samples (Brodie *et al.*, 2006). We attempted here to compare the data retrieved using the PhyloChip with clone libraries and real-time PCR data from a range of soils. One of the major advantages of the PhyloChip was that in a single hybridization it revealed significantly broader diversity than clone libraries composed of almost 200 clones (Figure 2). The numbers of taxa detected by the PhyloChip were often in the range of the total richness estimated from clone library analyses. This indicates that the PhyloChip provided a more complete view of bacterial and archaeal diversity in the Antarctic soils than modestly sized 16S rRNA gene libraries, similar to what has been reported recently

for other environments (Brodie *et al.*, 2006, 2007; DeSantis *et al.*, 2007). At the family level, the PhyloChip also detected a larger number of families than the clone libraries (Table 1). However, compared with earlier reports (Brodie *et al.*, 2006; DeSantis *et al.*, 2007), we found a relatively large proportion of families that were uniquely detected in the clone libraries, especially for the southernmost sites (Fossil Bluff and Coal Nunatak). This is probably because of the fact that such microarray platforms are based on previously recovered sequence information, and resulting databanks have a poorer coverage of microbial groups resident to seldom-studied, extreme environments. In such environments, gene discovery methods, such as clone libraries, represent a necessary complement to phylogenetic microarray analyses, at least until more studies evaluate the microbial diversity present in these environments. It should also be stressed that even though the microarray platform offered a more complete view of the microbial diversity in our soils, the sensitivity of the method limited our analysis to the detection of the most dominant community members. Furthermore, potential biases associated with DNA extraction and PCR amplification still exist.

It was earlier reported for aerosol samples that there was a poor correlation between the proportion of clones recovered from a particular taxon and the intensity of the fluorescent signal for that given taxon (Wilson *et al.*, 2002). However, we found highly significant correlations between the relative abundance at the phylum and the family level in Antarctic soil samples when comparing clone library and PhyloChip data (relative intensity), supporting the quantitative potential of the PhyloChip recently shown for simple mixtures of bacterial species (Brodie *et al.*, 2007). The values of these correlations were, however, relatively low, suggesting that biases and error still influence the results of one or both of the approaches. We observed that the agreement between the PhyloChip and the clone library data was strongest for those bacterial and archaeal groups that are most studied (Supplementary Table S1), which would be expected given the aforementioned reliance on information present in public databases. For instance, values obtained from the PhyloChip and the clone libraries for the *Proteobacteria* were very similar, whereas large differences were observed for less-well-studied groups like *Verrucomicrobia* and *Acidobacteria* (Supplementary Table S1).

We found no significant correlation between real-time PCR and PhyloChip data for the microbial groups analyzed by both methods. This might not be surprising as the fundamental detection mechanisms and the associated limitations on quantitative analyses differ between these methods. It should also be emphasized that the real-time PCR assays relied on different primer binding sites than those used for probing on the array, making the analyses

independent. Furthermore, probe signals were a summation of signals derived from multiple specific signals within a phylum, whereas real-time PCR results were generated by the use of group-specific primers. Even when using probes and primers targeting the exact same site, correlations between real-time PCR and microarray intensity are not perfect (for example, $r=0.87$ in Rhee *et al.*, 2004). Here again, both methods obviously rely on available sequence data for primer and probe design and may either miss some members of the target phyla (incomplete coverage) or overlap with the related phyla (incomplete specificity). The former case (incomplete coverage) might be especially true for Antarctic soils, which potentially contain many novel organisms. Interestingly, when applying the same approach in temperate grasslands, significant correlations were often found (r between 0.324 and 0.867, EE Kuramae *et al.*, unpublished results). Incomplete primer specificity could also play a role, as this has been shown for some primers used in this study (Fierer *et al.*, 2005).

Conclusions

PhyloChip analyses across a range of Antarctic soils yielded ecological conclusions that were highly consistent with earlier literature based on other, more traditional molecular methods (that is, PCR-DGGE, cloning–sequencing, real-time PCR). However, the level of detail realized using the PhyloChip was much higher. Combined analysis of the PhyloChip and the GeoChip data uncovered several relevant associations between taxa and functional genes, with a strong coupling between functional gene distribution and taxonomic composition of the bacterial and archaeal community. Polyphasic strategies, including use of PhyloChip and GeoChip microarrays, offer the opportunity to unravel microbial community structure and function in the unique and vulnerable habitats of the Antarctic.

Acknowledgements

This study was supported by NWO grant 851.20.018 to Rien Aerts and GA Kowalchuk. Part of this work was performed under the auspices of the U.S. DOE's Office of Science, Biological and Environmental Research Program, and by the University of California, LBNL under contract no. DE-AC02-05CH11231. E Yergeau was partly supported by a FQRNT postgraduate scholarship. Stef Bokhorst, Merlijn Janssens and Kat Snell are gratefully acknowledged for sampling at Fossil Bluff, Coal Nunatak and Signy Islands. Comments from Eiko Kuramae significantly improved this paper. We thank Pete Convey and the British Antarctic Survey for insightful discussions and logistical support. This is NIOO-KNAW publication #4400.

References

- Aislabie JM, Chhour K-L, Saul DJ, Miyauchi S, Ayton J, Paetzold RF *et al.* (2006). Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica. *Soil Biol Biochem* **38**: 3041–3056.
- Bokhorst S, Huiskes AHL, Convey P, Aerts R. (2007a). Climate change effects on organic matter decomposition rates in ecosystems from the Maritime Antarctic and Falkland Islands. *Glob Change Biol* **13**: 2642–2653.
- Bokhorst S, Huiskes AHL, Convey P, Aerts R. (2007b). The effect of environmental change on vascular plant and cryptogam communities from the Falkland Islands and the Maritime Antarctic. *BMC Ecol* **7**: 15.
- Bokhorst S, Huiskes AHL, Convey P, van Bodegom PM, Aerts R. (2008). Climate change effects on soil arthropod communities from the Falkland Islands and the Maritime Antarctic. *Soil Biol Biochem* **40**: 1547–1556.
- Brinkmann M, Pearce DA, Convey P, Ott S. (2007). The cyanobacterial community of polygon soils at an inland Antarctic nunatak. *Polar Biol* **30**: 1505–1511.
- Brodie EL, DeSantis TZ, Joyner DC, Baek SM, Larsen JT, Andersen GL *et al.* (2006). Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl Environ Microbiol* **72**: 6288–6298.
- Brodie EL, DeSantis TZ, Parker JPM, Zubietta IX, Piceno YM, Andersen GL. (2007). Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci USA* **104**: 299–304.
- Casgrain P, Legendre P. (2001). *The R Package for Multivariate and Spatial Analysis*. Département de sciences biologiques, Université de Montréal: Montréal, Canada.
- Cole J, Chai B, Farris R, Wang Q, Kulam S, McGarrell D *et al.* (2005). The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* **33**: D294–D296.
- Convey P. (2001). Antarctic ecosystems. In: Levin SA (ed). *Encyclopedia of Biodiversity*. Academic Press: San Diego, CA, pp 171–184.
- Davis RC. (1981). Structure and function of two Antarctic terrestrial moss communities. *Ecol Monogr* **51**: 125–143.
- DeSantis TZ, Brodie EL, Moberg JP, Zubietta IX, Piceno YM, Andersen GL. (2007). High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* **53**: 371–383.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microb* **72**: 5069–5072.
- Fierer N, Jackson JA, Vilgalys R, Jackson RB. (2005). Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl Environ Microbiol* **71**: 4117–4120.
- Fierer N, Jackson RB. (2006). The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103**: 626–631.
- González I, Déjean S, Martin PGP, Baccini A. (2008). CCA: an R package to extend canonical correlation analysis. *J Stat Soft* **23** (<http://www.jstatsoft.org/v23/i12>).
- He ZL, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC *et al.* (2007). GeoChip: A comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1**: 67–77.
- Heal OW, Block W. (1987). Soil biological processes in the North and South. *Ecol Bull* **38**: 47–57.
- Kennedy AD. (1993). Water as a limiting factor in the Antarctic terrestrial environment: a biogeographical synthesis. *Arct Alp Res* **25**: 308–315.
- Kennedy AD. (1996). Antarctic fellfield response to climate change: a tripartite synthesis of experimental data. *Oecologia* **107**: 141–150.
- Kowalchuk GA, Stephen JR. (2001). Ammonia-oxidizing bacteria: a model for molecular microbial ecology. *Annu Rev Microbiol* **55**: 485–529.
- Kowalchuk GA, Stienstra AW, Heilig GHJ, Stephen JR, Woldendorp JW. (2000). Changes in the community structure of ammonia-oxidizing bacteria during secondary succession of calcareous grasslands. *Environ Microbiol* **2**: 99–110.
- Lawley B, Ripley S, Bridge P, Convey P. (2004). Molecular analysis of geographic patterns of eukaryotic diversity in Antarctic soils. *Appl Environ Microbiol* **70**: 5963–5972.
- Legendre P, Anderson MJ. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr* **69**: 1–24.
- Legendre P, Legendre L. (1998). *Numerical Ecology*, 2nd English edn. Elsevier Science BV: Amsterdam, 853pp.
- Masuda N, Church GM. (2002). Escherichia coli gene expression responsive to levels of the response regulator EvgA. *J Bacteriol* **184**: 6225–6234.
- Mayer AM, Staples RC. (2002). Laccase: new functions for an old enzyme. *Phytochemistry* **60**: 551–565.
- Multiple Authors (2006). Bacteroides and Cytophaga group. In: Dworkin M (ed). *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community*, 3rd edn. Springer-Verlag: New York.
- Neufeld JD, Mohn WW. (2005). Unexpectedly high bacterial diversity in Arctic tundra relative to boreal forest soils, revealed by serial analysis of ribosomal sequence tags. *Appl Environ Microbiol* **71**: 5710–5718.
- Peat HJ, Clarke A, Convey P. (2006). Diversity and biogeography of the Antarctic flora. *J Biogeogr* **34**: 132–146.
- Rhee SK, Liu XD, Wu LY, Chong SC, Wan XF, Zhou JZ. (2004). Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70**: 4303–4317.
- Smith HG. (1992). Distribution and ecology of the testate rhizopod fauna of the continental Antarctic zone. *Polar Biol* **12**: 629–634.
- Sohlenius B, Boström S. (2005). The geographic distribution of metazoan microfauna on East Antarctic nunataks. *Polar Biol* **28**: 439–448.
- ter Braak CJF, Šmilauer P. (2002). *CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination* (version 4.5) Microcomputer Power: Ithaca, NY, 500pp.
- Willig MR, Kaufman DM, Stevens RD. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu Rev Ecol Evol Syst* **34**: 273–309.

- Wilson KH, Wilson WJ, Radosevich JL, DeSantis TZ, Viswanathan VS, Kuczmariski TA *et al.* (2002). High-density microarray of small-subunit ribosomal DNA probes. *Appl Environ Microbiol* **68**: 2535–2541.
- Wynn-Williams DD. (1996). Antarctic microbial diversity: the basis of polar ecosystem processes. *Biodivers Conserv* **5**: 1271–1293.
- Yergeau E, Bokhorst S, Huiskes AHL, Boschker HTS, Aerts R, Kowalchuk GA. (2007a). Size and structure of bacterial, fungal and nematode communities along an Antarctic environmental gradient. *FEMS Microbiol Ecol* **59**: 436–451.
- Yergeau E, Kang S, He Z, Zhou J, Kowalchuk GA. (2007b). Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *ISME J* **1**: 163–179.
- Yergeau E, Kowalchuk GA. (2008). Responses of Antarctic soil microbial communities and associated functions to temperature and freeze–thaw cycle frequency. *Environ Microbiol* **10**: 2223–2235.
- Yergeau E, Newsham KK, Pearce DA, Kowalchuk GA. (2007c). Patterns of bacterial diversity across a range of Antarctic terrestrial habitats. *Environ Microbiol* **9**: 2670–2682.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)