This supplementary file contains:

Materials and methods

Fig. S1. Distribution and diversity of bacterial OTUs (97% cutoff) after rarefying the data set to 2015 sequences per sample. Error bars depict standard errors. PD, phylogenetic diversity.

Fig. S2. Collector's curve (mean ± 95% confidence interval) of bacterial OTU (97% cutoff) richness vs. number of samples.

Fig. S3. Taxonomic distribution of bacterial OTUs (97% cutoff). Affiliation was performed using the Ribosomal Database Project Classifier with a confidence threshold of 80%. The sequence data generated in this study were deposited in the NCBI Sequence Read Archive and are available under the project number SRP050278. A supplementary table containing sample metadata conforming to MIMARKS standards is also provided (Table S1).

Fig. S4. Frequency histogram: number of bacterial OTU (97% cutoff) that occurred in each sample (of 15 possible).

MATERIALS AND METHODS

Water sampling

Fifteen water samples were collected at the margins of five (3 samples each) bog pools (54°44'S, 67°49'W) situated in Rancho Hambre peat bog, Tierra del Fuego, Argentina, in November of 2012. The climate in the system is cold-temperate and the central area has a characteristic dome shape. The vegetation is dominated by the moss *Sphagnum magellanicum*. A detailed description of sampled pools is provided in González Garraza *et al.* (2012). Sampled water (250 ml) was passed through 0.22-µm sterile nitrocellulose membranes (Nalgene, Rochester, NY, USA) and membranes placed in RNAlater (Sigma-Aldrich, St. Louis, Mo, USA) and stored at 4ºC until further processing.

Water temperature, pH and conductivity were measured *in situ* using a Hach Sension 156 multiparametric probe (Hach, Loveland, CO, USA). Ammonium and phosphate concentration were determined, using a Hach DR2800 spectrophotometer and their corresponding reagent kits, according to the salicylate and ascorbic acid methods, respectively. Total nitrogen and phosphorus were determined by acid digestion with

potassium persulphate and boric acid (APHA, 2005). Dissolved organic carbon (DOC) was determined using the high temperature Pt catalyst oxidation method (Shimadzu analyzer TOC-5000A, SM 5310B technique) following the recommendations of Sharp *et al*. (1993) (Table 1).

DNA extraction and 16S rRNA amplicon pyrosequencing

DNA was extracted from half of each filter, cut into small pieces with a sterile blade, using a PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). PCR was performed in a single-step PCR using HotStarTaq Plus Master Mix Kit (Qiagen, Valencia, CA) with primer pairs 27F (5'-AGRGTTTGATCMTGGCTCAG-3') and 519R (5'- GTNTTACNGCGGCKGCTG-3'). The following conditions were used: 94ºC for 3 minutes; 28 cycles of 94ºC for 30 seconds, 53ºC for 40 seconds and 72ºC for 1 minute; a final elongation step at 72ºC for 5 minutes. All amplicon products from different samples were mixed in equal concentrations and purified using Agencourt Ampure beads (Agencourt Bioscience Corporation, MA, USA). Pyrosequencing was carried out on a Roche FLX 454 at the Molecular Research LP next generation sequencing service (http://www.mrdnalab.com). Sequences were analysed in MOTHUR (Schloss *et al.*, 2009), following a previously established pipeline (Schloss *et al*., 2011), using the Silva core set (http://www.arb-silva.de) for alignment and FastTree (Price *et al.*, 2010) to estimate a phylogeny containing all OTUs observed across all samples.

Reads were removed from further analysis if at least one of the following criteria was met: (i) reads shorter than 200 bp, (ii) number of ambiguous bases greater than 5, and (iii) presence of homopolymers with more than 8 bp. One mismatch to the sample-specific barcode and two mismatches to the target-specific primer were allowed. After removal of chimeras and poor quality reads, a total of 81,929 sequences (2,015-11,839 sequences per sample) were obtained. Sequences were grouped into OTUs, defined using a 97% sequence similarity cut-off. Each sample was rarefied to 2015 sequences (the lowest number of sequences in any sample) and 897 OTUs (30,225 sequences) were retained for analysis.

The sequence data generated in this study were deposited in the NCBI Sequence Read Archive and are available under the project number SRP050278. A supplementary table containing sample metadata conforming to MIMARKS standards is also provided (Table S1)

Data analysis

OTU richness and rarefaction statistics were calculated using MOTHUR and the vegan package (Oksanen *et al.*, 2013) for R (R Development Core Team, 2013). A paired t-test, due to spatial autocorrelation, was used to determine significant differences between ombrotrophic and minerotrophic pools. The resulting data matrix was Hellinger-transformed and the Bray-Curtis distance measure was used to generate a dissimilarity matrix. To assess the significance of community similarity between *versus* within sites, ANOSIM (Clarke, 1993) and the inverse of Whittaker's multiplicative measure (Anderson *et al.*, 2011) were used. The later provides an index of among-site diversity that ranges from 0 to 1. Results close to 1 indicate most diversity is within sites whereas results close to 0 indicate the importance of between-site diversity.

Bacterial community similarities were visualised using non-metric multidimensional scaling (nMDS). Environmental variables were incorporated into the analysis through the use of bi-plot ordinations, in which variables were plotted as vector fits against bacterial community ordinations. Vector fitting of variables within ordinations were performed using the 'envfit' functions in the vegan package. Permutation tests (n=1000) were used to determine the significance of vector fits with ordination axes, and significant ($P<0.05$) variables were included in the resulting biplots.

To quantify the contribution of environmental variables in shaping peat bog bacterial communities, we used PERMANOVA, as implemented in the 'adonis' function in vegan. Before applying PERMANOVA, we performed variable clustering to assess the redundancy of the environmental variables (Martiny *et al.*, 2011) with the 'varclus' function in the Hmisc R package (Harrell, 2014). pH, ammonium, phosphate and total N were selected to be included in the models. The evidence for non-neutral community assembly of microbial communities in peat bog pools was quantified using null models (Kembel, 2009). We estimated, using picante (Kembel *et al.*, 2010) for R, the abundance-weighted mean pairwise distance (MPD) and the mean nearest phylogenetic taxon distance (MNTD) among sequences in each sample and calculated a standardized effect size (SES.MPD and SES.MNTD) by comparing the observed values to the value expected if communities were assembled at random from the pool of all OTUs observed in all water bodies (Kembel, 2009). The null model algorithm we used was 'independentswap' with 999 randomized null communities. 'Independentswap' retains column and row totals for null model analysis of species co-occurrence (Gotelli, 2000). This approach is particularly suited because it accounts for variations in diversity and

richness between communities.

We compared our results to those generated from presence-absence distance metrics (Chao and Jaccard), the probabilistic Raup-Crick metric, and the weighted UniFrac metric, an abundance-weighted measure of the phylogenetic differentiation among bacterial communities (Lozupone *et al.*, 2011); and to results calculated with all singletons removed, all singletons and doubletons removed, and all singletons, doubletons, and tripletons removed (Zhan *et al.,* 2014). All distance measures provided very similar nMDS ordination plots, ANOSIM and PERMANOVA results.

References

Anderson M.J., Crist T.O., Chase J.M., Vellend M., Inouye B.D., Freestone A.L.*, et al.* (2011) Navigating the multiple meanings of beta-diversity: A roadmap for the practicing ecologist. *Ecology Letters* **14:** 19-28.

APHA (2005) Standard Methods for Examination of Water and Wastewater, 21st edn. WEF, Washington, DC.

Clarke K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18:** 117-143.

González Garraza G., Mataloni G., Iturraspe R., Lombardo R., Camargo S., and Quiroga M.V. (2012) The limnological character of bog pools in relation to meteorological and hydrological features. *Mires and Peat* **7:** 1-14.

Gotelli N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology* **81:** 2606-2621.

Harrell F.E. (2014) Hmisc: Harrell Miscellaneous. R package version 3.14-3. http://CRAN.R-project.org/package=Hmisc ed.

Kembel S.W. (2009) Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecology Letters* **12:** 949-960.

Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D.*, et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26:** 1463-1464.

Lozupone C., Lladser M.E., Knights D., Stombaugh J., and Knight R. (2011) UniFrac: an effective distance metric for microbial community comparison. *ISME Journal* **5:** 169-172.

Martiny J.B.H., Eisen J.A., Penn K., Allison S.D., and Horner-Devine M.C. (2011) Drivers of bacterial beta-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 7850-7854.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B.*, et al.* (2013) Vegan: Community Ecology Package. R package version 2.0-2, http://CRAN.R-project.org/package=vegan ed.

Price M.N., Dehal P.S., and Arkin A.P. (2010) FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. *Plos One* **5**.

R Development Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; http://wwwR-projectorg/.

Sharp J.H., Peltzer E.T., Alperin M.J. *et al*. (1993) DOC procedures subgroup report.

Mar Chem **41:** 37-49.
Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B.*, et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75:** 7537-7541.
Schloss P.D., Gevers D., and Westcott S.L. (2011) Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. *Plos One* **6**.
Zhan A., Song H., Brown E.A., et al. (2014) Reproducibility of pyrosequencing data for biodiversity assessment in complex communities doi: 10.1111/2041-210X.12230

Fig. S1

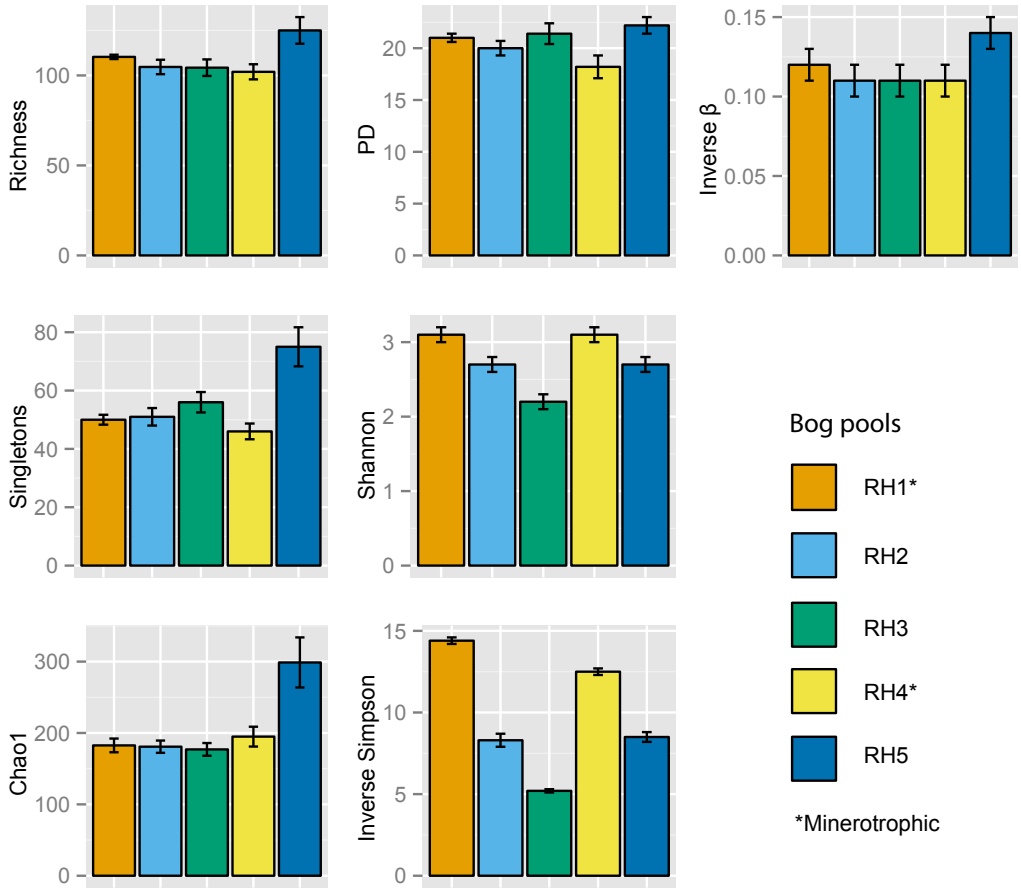

Richness
PD
Inverse β
Singletons
Shannon
Chao1
Inverse Simpson
Bog pools
RH1*
RH2
RH3
RH4*
RH5
*Minerotrophic

Fig. S2

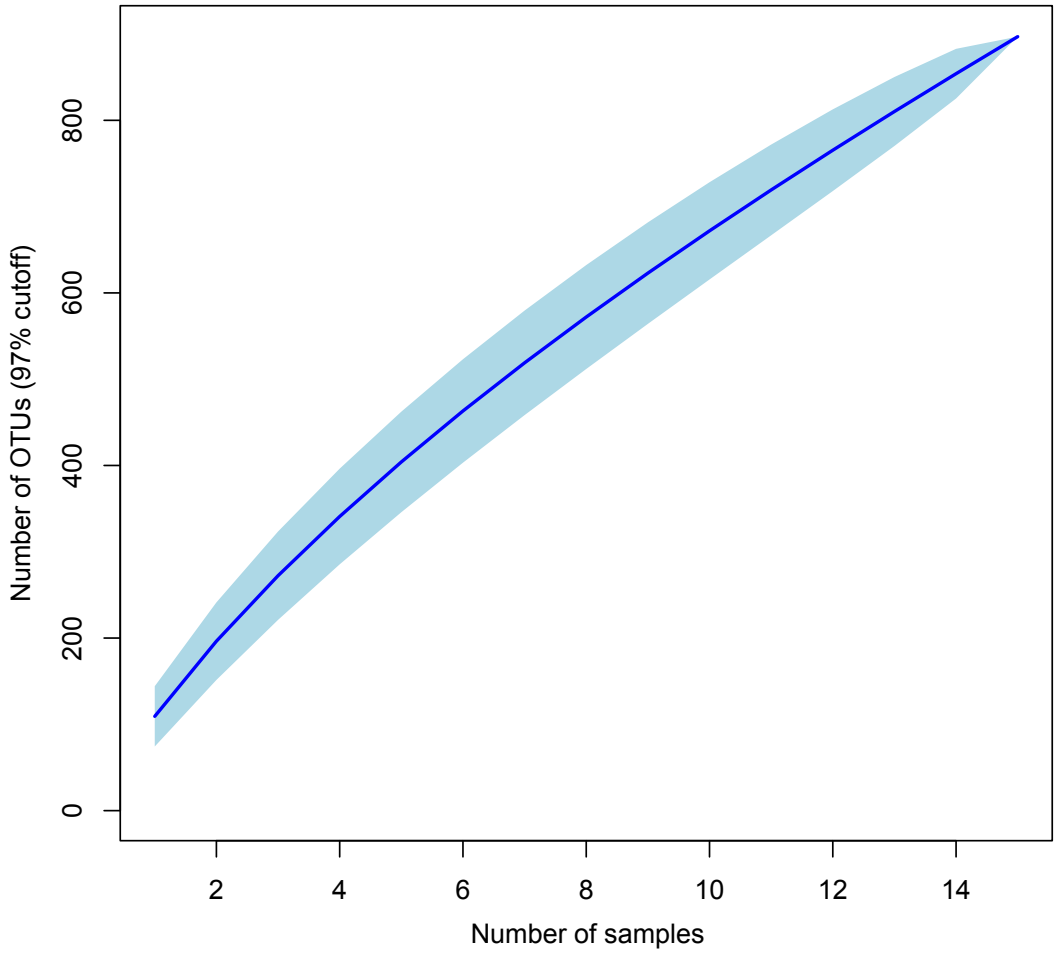

Number of OTUs (97% cutoff)
0
200
400
600
800
2
4
6
8
10
12
14
Number of samples

Fig. 3

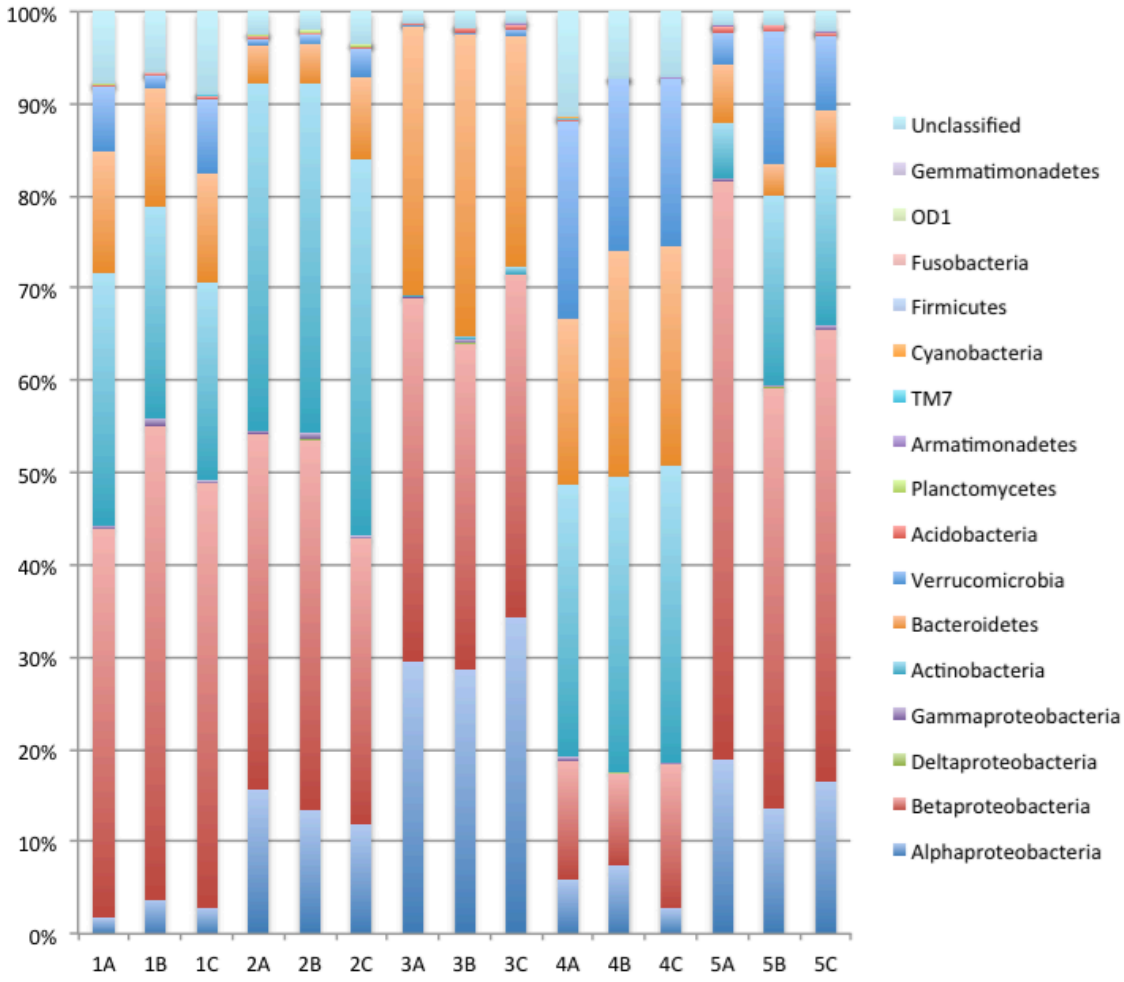
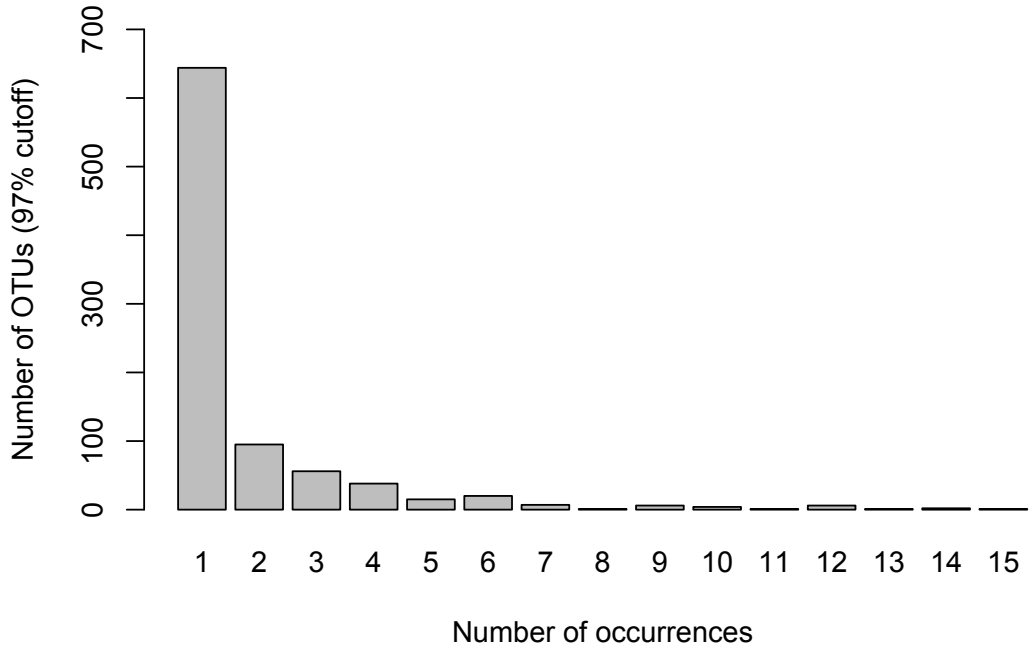

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1A
1B
1C
2A
2B
2C
3A
3B
3C
4A
4B
4C
5A
5B
5C
Unclassified
Gemmatimonadetes
OD1
Fusobacteria
Firmicutes
Cyanobacteria
TM7
Armatimonadetes
Planctomycetes
Acidobacteria
Verrucomicrobia
Bacteroidetes
Actinobacteria
Gammaproteobacteria
Deltaproteobacteria
Betaproteobacteria
Alphaproteobacteria

Fig. S4


Number of OTUs (97% cutoff)
0
100
300
500
700
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Number of occurrences