Ecography

ECOG-02314
Weiser, M. D., Michaletz, S., Buzzard, V., Deng, Y., He, Z., Shen, L., Enquist, B. J., Waide, R. B., Zhou, J. and Kaspari, M. 2017. Toward a theory for diversity gradients: the abundance-adaptation hypothesis. – Ecography doi: 10.1111/ecog.02314

Supplementary material

Appendix 1
Estimating Diversity of Trees, Invertebrates and Microbes

Trees

Tree species were identified to species using field guilds, regional species keys, and local experts. Each species was assigned to a plant family (using The Plant List http://www.theplantlist.org/) and a phylogenetic clade (following the Angiosperm Phylogeny assignment at http://www.mobot.org/MOBOT/research/APweb/.)

Soil Invertebrates

We quantified the diversity of meso and macro invertebrates (taxa >0.5 mm in length) in a variety of ways. For our two focal taxa—the ants and the oribatids—we endeavored to build species lists for each site from the 21 samples using our own curated collections, the taxonomic literature, and taxonomic experts. For the ants (Family Formicidae) we were able to build a list that allowed comparisons across the sites using our own reference collections and texts (Bolton, 1994, 1995; Ellison *et al.*, 2012). For the oribatids (Order Oribatida) we used a variety of text sources (Balogh, 1972; Krantz, 1978; Niedbała, 1992) and consultations with taxonomists to build dependable lists of species and morphospecies within sites, but that could not dependably be compared across the six forests of this study. Finally, we quantified the total invertebrate richness of individual m^2 samples using morphospecies concepts and characters gleaned from the literature (Voegtlin, 1982; Dindal, 1990). This allowed us to contrast our invertebrate diversity at the m2 scale with sequence-based microbial richness estimates at the m^2 scale across the six forests.

Bacteria, Archaea, and Fungi

We used 16S rRNA gene sequences to identify OTUs of archaea and bacteria, and ITS

sequences to identify fungal OTUs. The high degree of sequence variation in the fungal ITS region allowed us to resolve species level OTUs at the expense of poorer resolution at the phylum and family level (Schoch *et al.,* 2012). Soil DNA was extracted using the grinding-SDS-based DNA extraction method as previously described (Zhou *et al.*, 1996) DNA concentration was measured by a PicoGreen method using a FLUOstar OPTIMA fluorescence plate reader (BMG LABTECH, Jena, Germany). For 16S rRNA genes, the V4 region was amplified with the primer pair 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'- GGACTACHVGGGTWTCTAAT-3') combined with the Illumina adapter sequence, a pad and a linker of two bases, and a barcode sequences on the reverse primers (Caporaso *et al.*, 2011; Caporaso *et al.*, 2012). PCR amplification was performed in 25 μl reactions containing 2.5 μl 10×AccuPrime PCR buffer (including dNTPs) (Invitrogen, Grand Island, NY), 0.4 μM of both forward and reverse primers , 10ng of template DNA and 0.2μl AccuPrime High Fidelity Taq Polymerase. Three replicates were made for each sample and then mixed after PCR amplification to minimize potential biases from amplification (Zhou *et al.*, 2011). Thermal cycling conditions were as follows: initial denaturation at 94°C for 1 min, followed by 30 cycles of 94°C for 20 s, 53°C for 25 s, and 68°C for 45 s, with final extension at 68°C for 10 min. For *nifH* and fungal ITS sequencing, the phasing amplicon sequencing approach was used. For nifH, an amplicon of 302 bp (excluding the primers) was targeted using the primers: Pol115F, TGCGAYCCSAARGCBGACTC, and Pol457R, ATSGCCATCATYTCRCCGGA. For fungal ITS, an amplicon of 309 bp (not including the primers) in ITS2 region was targeted using the primers: gITS7F, GTGARTCATCGARTCTTTG, and ITS4R, TCCTCCGCTTATTGATATGC (Ihrmark

et al., 2012).

A two step-PCR was performed to avoid extra PCR bias which could be introduced by the added components in the long primers used for PCR library preparation. Phasing primers were designed and used in the second step of the two-step PCR. Spacers of different length (0-7 bases) were added between the sequencing primer and the target gene primer in each of the 8 forward and reverse primer sets to shift the sample DNA sequences among different samples. To ensure that the total length of the amplified sequences do not vary with the primer set used, the forward and reverse primers were used in a complementary fashion so that all of the extended primer sets have exactly 7 extra bases as the spacer for sequencing phase shift. Both forward and reverse phasing primers include the Illumina adapter, the Illumina sequencing primer, a spacer, and the target gene primer, and a barcode of 12 bases in the reverse primer between the sequencing primer and the adaptor. In the two-step PCR, the first round PCR was carried out in a 50 μl reaction containing 5 μl 10×PCR buffer II (including dNTPs), 0.5 U high fidelity AccuPrime™ Taq DNA polymerase (Life Technologies), 0.4 μM of both forward and reverse target only primers, 10 ng soil DNA. Reactions were performed in triplicate and the sample amplification program described above was used except that only 10 cycles were performed and the annealing temperature was 56^{o}C for ITS and 62^{o}C for *nifH*. The triplicate products from the first round PCR were combined, purified with an Agencourt® AMPure XP kit (Beckman Coulter, Beverly, MA, USA), eluted in 50 μl water, and aliquoted into three new PCR tubes (15 uL each). The second round PCR used a 25 μl reaction containing 2.5 μl 10×PCR buffer II (including dNTPs), 0.25 U high fidelity AccuPrime™ Taq DNA polymerase (Life Technologies), 0.4 μM of both forward

and reverse phasing primers, 15 µl aliquot of the first-round purified PCR product. The amplifications were cycled 20 times following the above program. Positive PCR products were confirmed by agarose gel electrophoresis. PCR products from triplicate reactions were combined and quantified with a PicoGreen method.

PCR products from samples to be sequenced in the same MiSeq run (generally 3×96=288 samples) were pooled at equal *molality*. The pooled mixture was purified with a QIAquick Gel Extraction Kit (QIAGEN Sciences, Germantown, MD, USA) and re-quantified with PicoGreen. Sample libraries for sequencing were prepared according to the MiSeq™ Reagent Kit Preparation Guide (Illumina, San Diego, CA, USA) as described previously (Caporaso *et al.*, 2012). First, the combined sample library was diluted to 2 nM. Then, sample denaturation was performed by mixing 10 µl of the diluted library and 10 µl of 0.2 N fresh NaOH and incubated 5 min at room temperature. 980 µL of chilled Illumina HT1 buffer was added to the denatured DNA and mixed to make a 20 pM library. Finally, the 20pM library was further adjusted to the desired concentration (about 12pM) for sequencing using chilled HT1 buffer. The library for sequencing was mixed with a certain proportion of a Phix library of the same concentration to achieve a 10% Phix spike.

A 300-cycle v1 (for 16S rDNA) or 500-cycle v2 (for ITS or nifH) MiSeq reagent cartridge (Illumina) was thawed for 1 h in a water bath, inverted ten times to mix the thawed reagents, and stored at 4 °C for a short time until use. For 16S rDNA sequencing, customized sequencing primers for forward, reverse, and index reads were added to the corresponding wells on the reagent cartridge prior to being loaded as described previously (Caporaso *et al.*, 2012). Sequencing was performed for 151, 12, and 151

cycles (for 16S rDNA), or 251, 12, and 251 cycles (for ITS and *nifH*) for forward, index, and reverse reads, respectively.

Sequence preprocessing

The raw reads of 16S, ITS and nifH genes were collected in Miseq sequencing machine in fastq format. Their forward and reverse directions and barcodes were generated into separated files. First, the spiked PhiX reads were removed by using BLAST against PhiX genome sequence in E value less than 10^{-5}. Second, the reads were assigned to samples according to the barcodes in the barcode file with up to one mismatch allowed. For both 16S and ITS, forward and reverse reads of same sequence with at least 10 bp overlap and lower than 5% mismatches were combined as single sequence by using FLASH program (Magoc & Salzberg, 2011), while the minimum overlap length for *nifH* was set to 50 bp. Any joined sequences with an ambiguous base, or less than 240 bp for 16S rRNA, less than 200 bp for ITS and less than 229 bp for *nifH* gene sequences were discarded. Besides, the Btrim program (Kong, 2011) with threshold of QC higher than 20 over five bp window size was used to further filter the unqualified sequences. Thereafter, U-CHIME (Edgar *et al.*, 2011) was used to remove chimeras by searching against GreenGene reference dataset (DeSantis *et al.*, 2006) for 16S rRNA gene sequence dataset, against UNITE/QIIME released ITS reference (http://qiime.wordpress.com/2012/11/27/uniteqiime-12_11-its-reference-otus-now-available-alpha-release/) for ITS dataset and against *nifH* database released by Zehr Laboratory (http://pmc.ucsc.edu/~wwwzehr/research/database/) for *nifH* dataset. OTUs were classified using UCLUST with different similarity levels (Edgar, 2010) for all 16S, ITS and nifH genes. Thereafter, the reads of OTUs were re-assigned back to their

samples and a matrix with 126 samples as columns and all OTUs as rows was generated for each dataset. Since reliable taxonomic assignments (OTU annotations) for both ITS and *nifH* were still unavailable, we only got taxonomic annotation for 16S rRNA gene sequence dataset through RDP classifier (Wang *et al.*, 2007) with minimal 50% confidence score. All of sequences were also reanalyzed using the recent program UPARSE, which was developed by the same author (Wang *et al.*, 2007). The main differences of these two programs are the algorithms for read quality filtering and the chimera filtering. For the 16S data, about 5 times of less OTUs were obtained with UPARSE than UCLUST. Although the OTU numbers obtained by these two programs were dramatically different, but relationships between the OTU richness and climate and soil variables were not changed significantly (data not shown), no matter what programs were used. Since UPARSE requires good reference database by using UPARSE-REF algorithm to remove chimeras, reliable OTU classification could not be obtained with *nifH* data because of lack of a comprehensive reference database. Thus, to be consistent, all of the sequence analysis results reported are based on the UCLUST approach in this study.

S1:

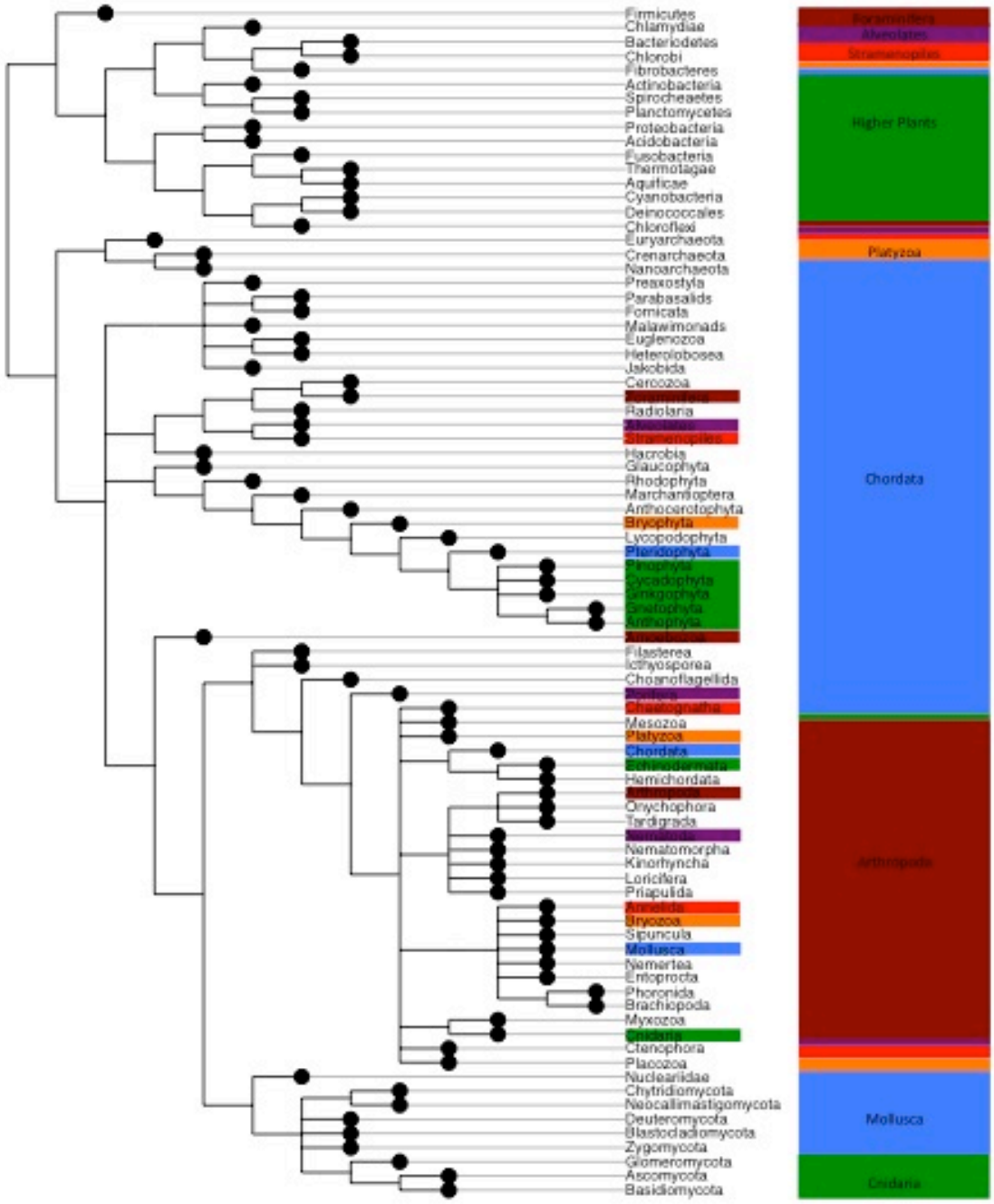

Figure A1: The phylogenetic distribution of studies of the Latitudinal Diversity Gradient. The dendrogram represents an outline of the phylum level relationships across the tree of life (www.tol.org/). Dendrogram labels with colors represent the clades for which the latitudinal diversity gradient has been examined(in the review by Hillenbrand Hillebrand, 2004). The stacked bar graph is the proportional representation of these clades in Hillenbrand's data. 77% of studies of the LDG are from higher plants (12.5%), Chordata (38%) and Arthropoda (26.7%).

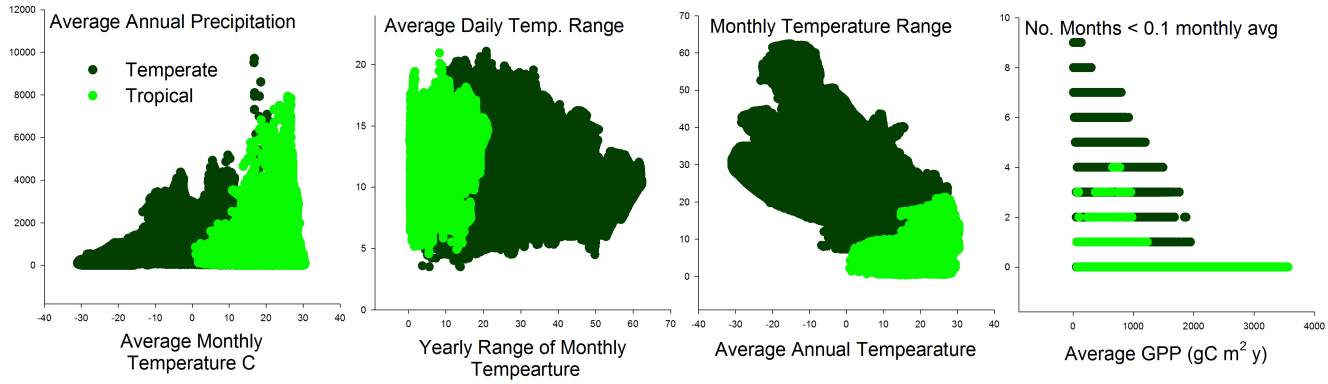

Figure A2. Average terrestrial abiotic conditions in and out of the tropics (within 23.4° latitude), at a 10 x 10 ‘ grid scale (New *et al.*, 2002). Dark green areas are recorded only from extra-tropical areas. Average GPP estimated at 0.5° grid scale over 5 years, and plotted against the number of “lean” months, those in which GPP falls below 10% of the monthly average.

We downloaded the monthly Gross Primary Production data from MOD17 version 055 from 2009 – 2013 (http://www.ntsg.umt.edu/project/mod17). The original data are delivered in GeoTIFF format with a spatial resolution of 0.05 degrees and represent 10* total grams of carbon per m2 for each month. To get to grams of carbon per m2 you should multiply the values by 0.1. We resized the data to 0.5 degrees and then calculated the five-year averages for each month.

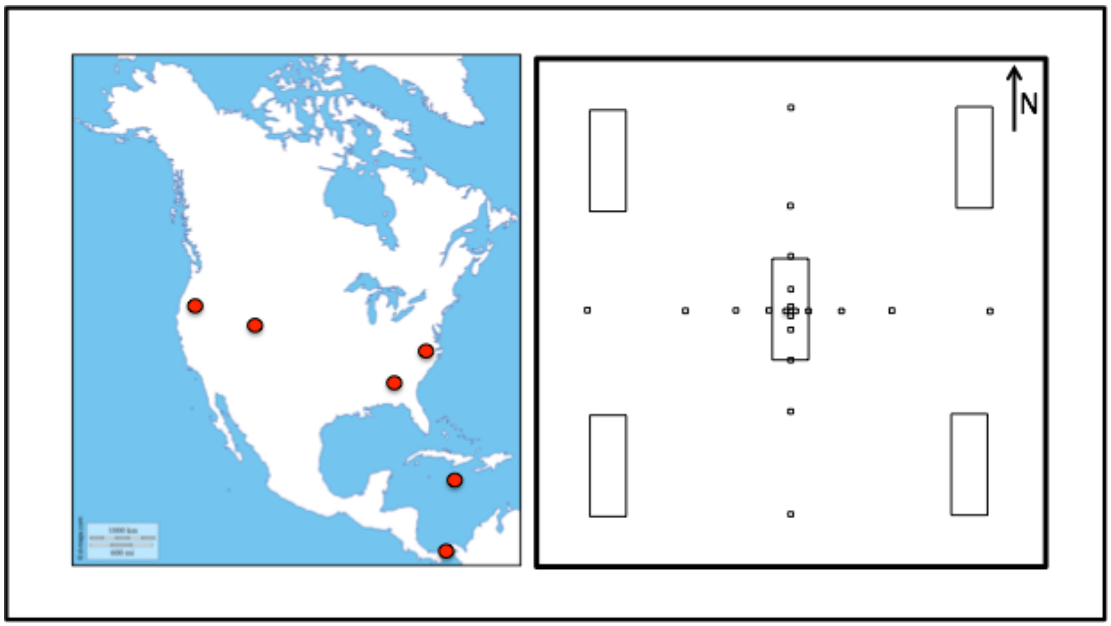

Figure A3. Map of 6 sample sites, and layout of sampling effort at each site. The small squares represent each of the 21 1m^2 plots for soil arthropods and microbes. The rectangles represent the 5 gentry plots sampled for tree diversity.

Barro Colorado Island (BCI), Barro Colorado National Monument, Panama- Barro Colorado Island (9.16N, 79.85W, 157masl) is a 1560 ha island formed by the rising waters of Lake Gatun due to the creation of the Panama Canal. BCI has semi-deciduous, lowland tropical moist forest with basaltic and sedimentary derived clay soils.

Luquillo LTER (LUQ), El Yunque National Forest, Puerto Rico, USA. The Luquillo array (18.32N, 65.82W, 386masl) overlaps with the Luquillo Forest Dynamics Plot in Luquillo Experimental Forest, near the El Verde Field Station. Soils are kaolonitic oxisols. (http://luq.lternet.edu/data/luqmetadata127).

Coweeta Hydrological Laboratory LTER (CWT), Nantahala National Forest, North Carolina, USA. Coweeta Hydrological Laboratory (35.05N, 83.43W, 864masl).

H.J. Andrews Experimental Forest LTER (AND), Willamette National Forest, Oregon, USA. The plots at H.J. Andrews Experimental Forest were located in old growth forest coniferous (Douglas Fir) forest (44.23N, 122.15W, 860masl).

Harvard Forest LTER (HFR), (42.54N, 72.18W, 356masl), Massachusetts. Harvard Forest is a second-growth deciduous hardwood forest with podzolic soils (Stout, 1952).

Niwot Ridge LTER (NWT), Mountain Research Station (40.04N, 105.56W, 3186 masl), Colorado. The forest at Niwot Ridge is Spruce/Fir.

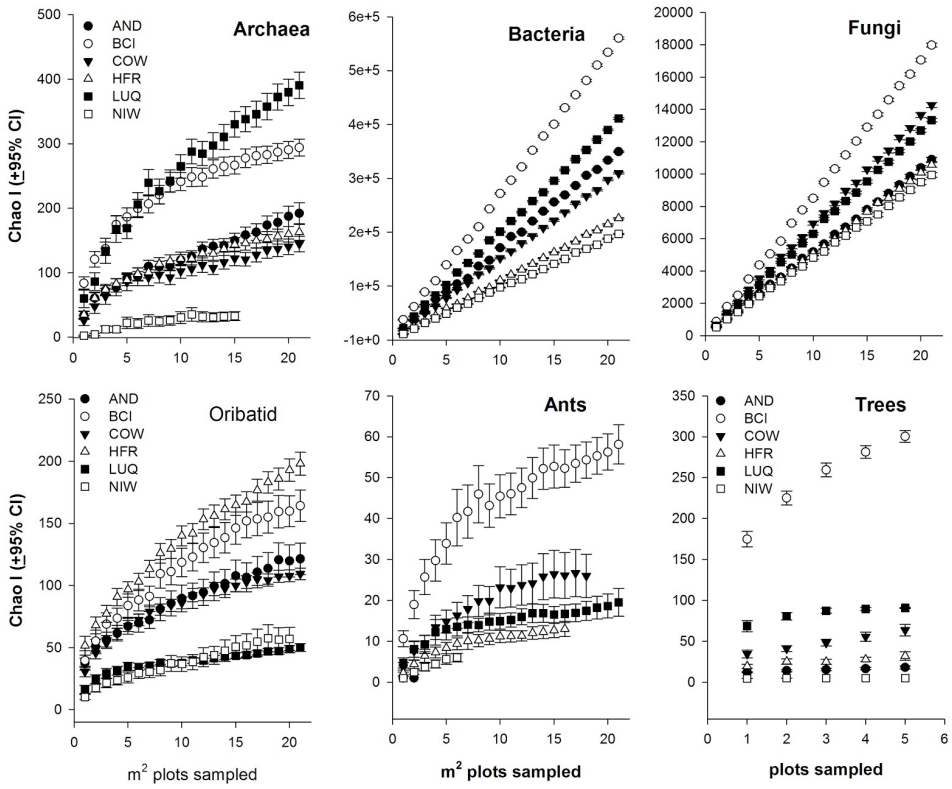

Figure A4: Chao1 accumulation curves for the six focal taxa samples across six forests.

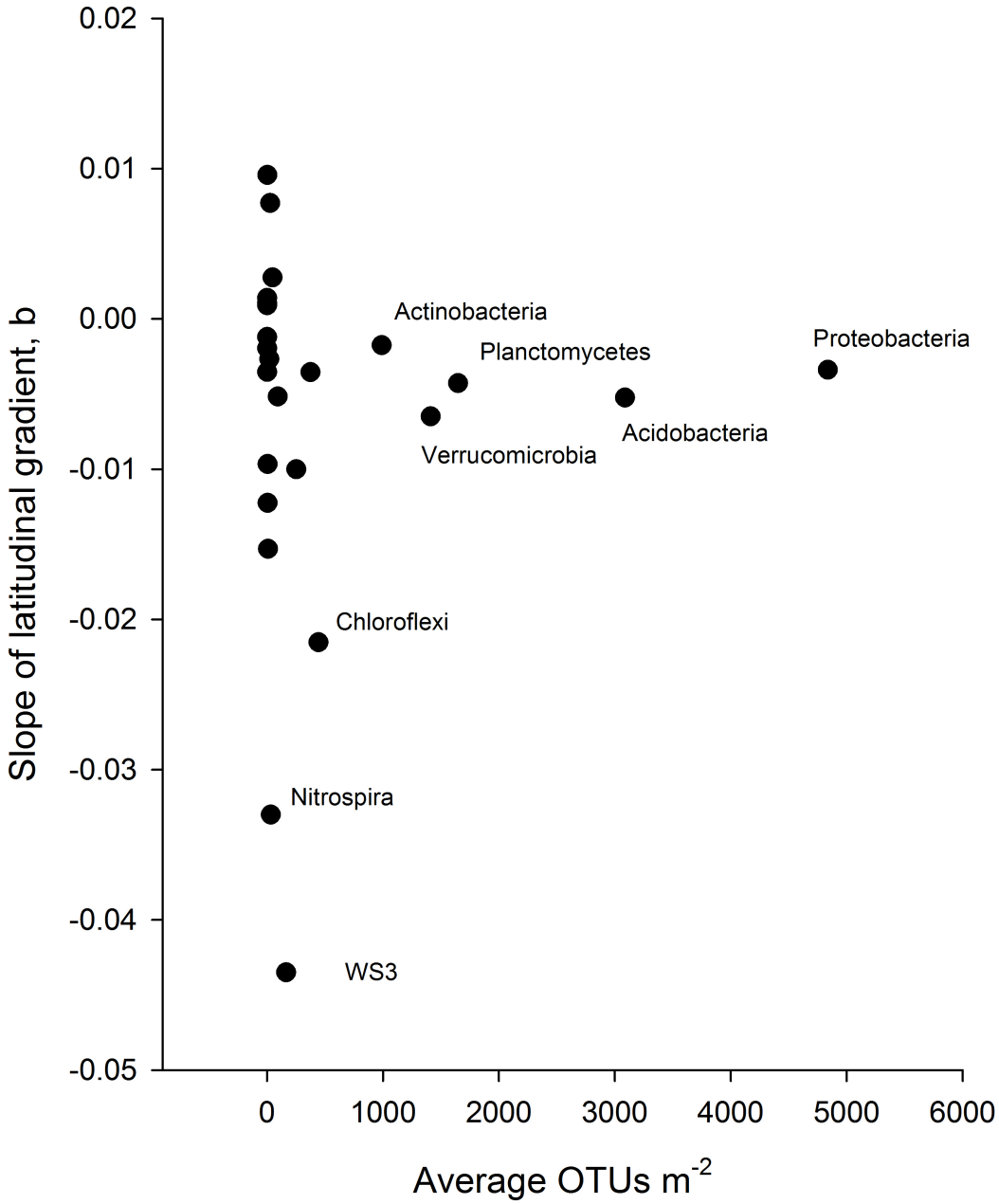

Figure A5: Distribution of the slopes b for latitudinal diversity gradients of 24 Eubacteria phyla, plotted against the average number of species per m^2. The 7 most OTU-rich phyla, plus two outliers, are labeled.

Table A6: Linear models of Log_{10} diversity per 200 m^2 (Trees), per site (Bats and Birds), and per m^2 (all else) versus latitude for major clades across six forests, as well as estimates of average mass for all eukaryotes except fungi.

Domain	Group	Intercept	exponent (slope)	Mass (g)	Mass References
archaea	Archaea	1.59	-0.015		
eubacteria	Acidobacteria	3.48	-0.005		
eubacteria	Actinobacteria	2.99	-0.001		
eubacteria	Armatimonadetes	1.68	0.002		
eubacteria	BRC1	0.98	-0.015		
eubacteria	Bacteroidetes	2.57	-0.003		
eubacteria	Chlamydiae	1.44	0.007		
eubacteria	Chlorobi	0.373	-0.003		
eubacteria	Chloroflexi	2.64	-0.021		
eubacteria	Cyanobacteria/ Chloroplast	1.29	-0.003		
eubacteria	Deinococcus-Thermus	0.10	-0.002		
eubacteria	Elusimicrobia	0.24	0.001		
eubacteria	Firmicutes	2.40	-0.010		
eubacteria	Fusobacteria	0.07	-0.001		
eubacteria	Gemmatimonadetes	1.97	-0.005		
eubacteria	Nitrospira	1.53	-0.033		
eubacteria	OD1	0.50	0.010		
eubacteria	OP11	0.64	-0.012		
eubacteria	Planctomycetes	3.21	-0.004		
eubacteria	Proteobacteria	3.68	-0.003		
eubacteria	Spirochaetes	0.64	-0.010		
eubacteria	TM7	0.06	0.001		
Domain	**Group**	**Intercept**	**exponent**	**Mass (g)**	**Mass References**
eubacteria	Tenericutes	0.03	0.001		
eubacteria	Verrucomicrobia	3.15	-0.007		
eubacteria	WS3	2.21	-0.044		
eukarya	Trees	2.77	-0.037	100000	
eukarya	Birds	2.65	-0.025	50	(Thibault *et al.*, 2011)
eukarya	Bats	1.9	-0.025	12.65	(Smith *et al.*, 2003)
eukarya	Formicidae	1.08	-0.021	0.01	(Kaspari & Weiser, 2007)
eukarya	Insecta	1.26	-0.014	0.00056	(Kaspari & Weiser, 2007)

eukarya	Myriapoda	0.49	-0.008	0.00185	(Kaspari & Weiser, 2007)
eukarya	Fungi	3.19	-0.005	.	
eukarya	Araneae	0.64	-0.004	0.00043	(Kaspari & Weiser, 2007)
eukarya	Collembola	0.75	-0.002	3.1715E-06	(Kaspari & Weiser, 2007)
eukarya	Acari	1.348	0.001	1.07934E-05	(Kaspari & Weiser, 2007)

Supplemental Information Literature Cited

Balogh, J. (1972) The oribatid genera of the world. *Budapest, Hungary, Akademiai Kiado*,

Bolton, B. (1994) *Identification guide to the ant genera of the world*. Harvard University Press, Cambridge Massachusetts.

Bolton, B. (1995) *A New General Catalogue of the Ants of the World*. Harvard University Press, Cambridge, Massachusetts.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N. & Knight, R. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Science USA*, **108**, 4516-4522.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G. & Knight, R. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platform. *ISME Journal*, **6**, 1621-1624.

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

Dindal, D.L. (1990) *Soil biology guide*. New York.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460-2461.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194-2200.

Ellison, A.M., Gotelli, N.J., Farnsworth, E.J. & Alpert, G.D. (2012) *A field guide to the ants of New England*. Yale University Press.

Hillebrand, H. (2004) On the generality of the latitudinal diversity gradient. *The American Naturalist*, **163**, 192-211.

Ihrmark, K., Bodejer, I.T.M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandstrom-Durling, M., Clemmensen, K.E. & Lindahl, B.D.

(2012) New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology*, **82**, 666-677.

Kaspari, M. & Weiser, M. (2007) The size-grain hypothesis: do macroarthropods see a fractal world? *Ecological Entomology*, **32**, 279-282.

Kong, Y. (2011) Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics*, **98**, 152-153.

Krantz, G.W. (1978) *A manual of acarology*. Oregon State University Book Stores, Inc., Corvallis OR, USA.

Magoc, T. & Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957-2963.

New, M., Lister, D., Hulme, M. & Makin, I. (2002) A high-resolution data set of surface climate over global land areas. *Climate research*, **21**, 1-25.

Niedbała, W. (1992) *Phthiracaroidea (Acari, Oribatida): systematic studies*. Elsevier.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K. & Crous, P.W. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, **109**, 6241-6246.

Smith, F.A., Lyons, S.K., Ernest, S.M., Jones, K.E., Kaufman, D.M., Dayan, T., Marquet, P.A., Brown, J.H. & Haskell, J.P. (2003) Body Mass of Late Quaternary Mammals: Ecological Archives E084-094. *Ecology*, **84**, 3403-3403.

Stout, B.B. (1952) Species distribution and soils in the Harvard Forest. *Harvard Forest Bulletin*, **24**, 1-29.

Thibault, K.M., White, E.P., Hurlbert, A.H. & Ernest, S. (2011) Multimodality in the individual size distributions of bird communities. *Global ecology and biogeography*, **20**, 145-153.

Voegtlin, D.J. (1982) Invertebrates of the HJ Andrews Experimental Forest, western Cascade Mountains, Oregon: a survey of arthropods associated with the canopy of old-growth Pseudotsuga menziesii.

Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261-5267.

Zhou, J., Bruns, M.A. & Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, **62**, 316-322.

Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y.-H., Tu, Q., Xie, J., Van Nostrand, J.D., He, Z. & Yang, Y. (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *ISME Journal*, **5**, 13031313.