Supplementary Information

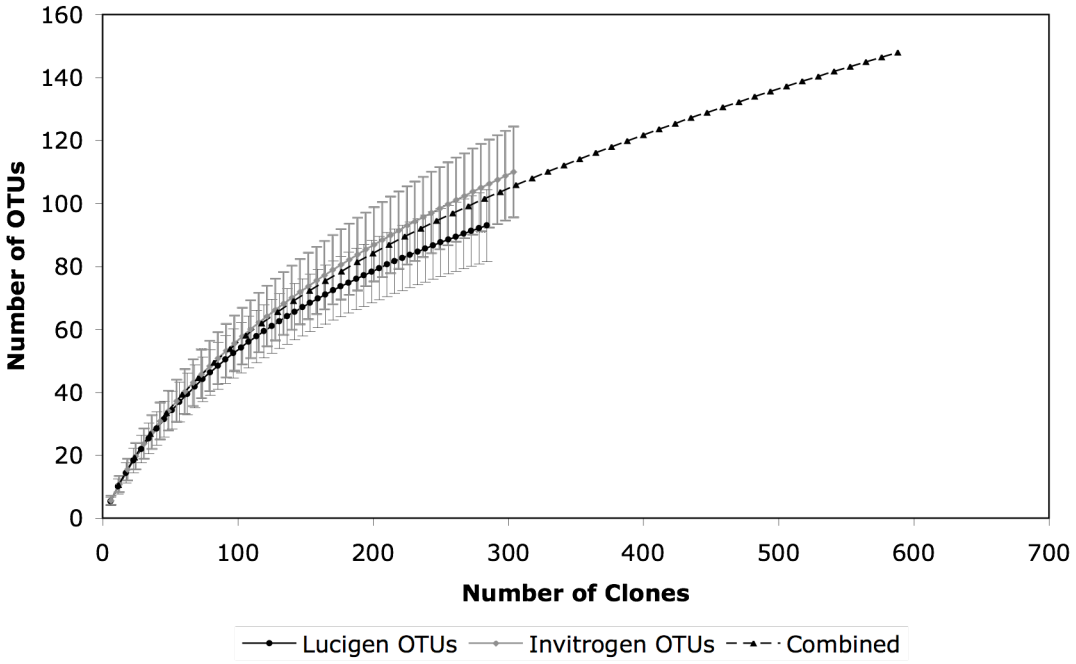

Figure S1. Species accumulation curves. The numbers of OTUs detected relative to the number of clones sequenced, are statistically equivalent, showing that the two methods detect similar levels of diversity. Error bars represent 95% confidence intervals for numbers of species (Mao Tao) generated by EstimateS 7.0 (Colwell and Coddington, 1994). To assess the phylogenetic representation across the two libraries, sequences were clustered into OTUs using the program Cap3 (Huang and Madan, 1999). Quality scores exported from Aligner were used in Cap3 to clip bases with scores below 10, and to consider only bases with combined qualities above 40 in determining mismatches between sequences. All parameters were set to lenient values favoring assembly of pairs of sequences except the minimum percent identity in the overlapping region, which was set to 90%, and the maximum overhang percent length, which was set to 20%. To create species (OTU) accumulation curves for each library, the observed individuals were randomly distributed to create 50 mock samples for each library, with 50 repetitions of this randomization process.

Table S1. Summary of clones used for analyses.

	Lucigen	Invitrogen
Passing Assemblies	308	312
Short Inserts	18	1
Chimeras	4	1
Non-fungal Clones	2	6
Remaining Clones	284	304

Ninety-two 1.8 cm diameter soil cores were collected from a 150 X 150 meter plot in a floodplain black spruce (*Picea mariana*) stand (N64.715dd, W148.141dd) in the Bonanza Creek Long Term Ecological Research site near Fairbanks, Alaska. Subsamples of the litter, humic and mineral horizons were dissected from each core within four days of collection. In the present study, only the humic horizon samples are considered. Environmental DNAs were extracted using the Ultraclean Soil DNA extraction kit (MoBio Laboratories, Carlsbad, CA, USA). A region spanning the nuclear ribosomal ITS1-5.8S-ITS2 regions and approximately 700 bp of the large subunit was amplified using Ready-to-Go PCR beads (Amersham Biosciences, Piscataway, NJ, USA) and the fungal-specific primer ITS 1F (Gardes et al., 1991) together with the universal eukaryote primer TW13 (T.J. White, unpublished). For the Invitrogen library, standard oligonucleotide primers were used, while for the Lucigen library, 5' phosphorylated primers were used, both from Integrated DNA Technologies (Coralville, IA, USA). PCR cycling parameters were as follows: 96C for 2 min, 94C for 30 sec, 55C for 40sec, 72C for 3 min for 28 cycles, followed by a final extension at 72C for 10 min. PCR products were purified through DNA Clean and Concentrator-25 columns (Zymo Research, Orange, CA, USA). To minimize the cloning of primer-dimers and other short inserts, the pooled fragments were size fractionated over Chroma Spin 400 columns (BD Biosciences, Palo Alto, CA, USA). For the Lucigen library, end repair to

remove overhanging adenines was performed on 2 ugs of purified PCR product following manufacturer instructions (Lucigen Corporation, Middleton, WI, USA). The end-repair product was then purified through Zymo DNA Clean and Concentrator-5 columns. The Lucigen ligation reaction utilized 200 ngs of insert DNA. The TOPO-TA for Sequencing kit with the pCR4-TOPO vector was used for construction of the TA library (Invitrogen, Carlsbad, CA, USA) in a reaction containing 100 ngs of insert DNA. At each step, DNA concentrations were measured on a Nanodrop spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) and adjusted to recommended concentrations. GC10 Thunderbolt™ electrocompetent cells (Sigma-Aldrich, St. Louis, MO, USA) were transformed and 384 colonies were picked from each library. Templiphi amplifications were carried out (Nelson et al., 2002), followed by cycle-sequencing with M13F and M13R primers (Invitrogen) or SR2 and CL3 primers (Lucigen) along with Ctb6 and ITS4 (White et al., 1990) as internal primers for both libraries, followed by ethanol precipitation and separation in an ABI 9700 capillary sequencer. Dirty ends of sequence reads were trimmed, vector sequences were removed and the four reads from each clone were assembled using Assembler 5.1 for Mac OSX (Codoncode, Boston, MA, USA). Reads with fewer than 300 quality base calls or with evidence of multiple sequences resulting from mixed colonies were discarded. If the same OTU was encountered in both libraries, meaning that very similar sequences were generated in separate PCR reactions, we reasoned that this OTU was unlikely to have a chimeric origin. We evaluated possible chimerism in every OTU which was encountered in only one library as follows. Clone assembly consensus sequences were exported in FASTA format and the ITS 1, ITS 2 and LSU regions isolated from each assembly. The three regions were then compared separately to comprehensive databases of fungal ITS and LSU regions using the FASTA matching algorithm implemented on our fungus identification web site (Geml et al., 2005). High sequence identities (i.e. above 90%) of the three portions of a clone to distantly related fungal families or orders was taken as evidence for chimerism, and these clones were discarded. Finally, the chimera detection feature of Cap3 was used to search for chimeras between sequences within our amplicon pool. In addition, clones were discarded if they 1) had the same primer at both ends, 2) were short inserts, as evidenced by vector sequence prior to the expected primer sequence, or 3) appeared not to be of fungal origin based on discontinuous Mega-BLAST searches of Genbank (Altshul et al., 1997) and phylogenetic analyses of the LSU regions.

Table S2. Diversity and similarity statistics for the Invitrogen and Lucigen clone libraries.

	Lucigen	Invitrogen	Combined/Compared
Mao Tao	93 (5.8)	110 (7.4)	148 (8.2)
ACE	145	221	276
ICE	150	221	280
Chao 1	147	226	272
Chao 2	147	219	272
Jackknife 1	135	170	221
Jackknife 2	160	212	271
Shannon	4.08	4.22	4.34
Simpson	42.57	50.39	47.59
Classic Jaccard			0.372
Chao-Jaccard Est			0.822 (.085)
Classic Sorenson			0.542
Chao-Sorenson Est			0.902 (.05)
Morista-Horn			0.819

Diversity and similarity statistics were calculated using EstimateS 7.0 (Colwell and Coddington, 1994). Mao-Tao through Simpson indices describe species richness in each library, and in the combined libraries; Jaccard through Morista-Horn indices describe similarity of the two libraries. The versions of the Chao-Jaccard and Chao-Sorenson statistics shown here are those which take species abundances, as well as estimated undetected species, into account. Values in parentheses are analytical estimates of standard deviations.

Altshul, S.F., Madden, T.L., Schäffer, A.A., Zhang, Z., Zheng, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI_BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Gardes, M., White, T.J., Fortin, J.A., Bruns, T.D., and Taylor, J.W. (1991) Identification of indigenous and introduced symbiotic fungi in ectomycorrhizae by amplification of nuclear and mitochondrial ribosomal DNA. *Can J Bot* **69**: 180-190.

Geml, J., Via, Z., Long, J., Huston, S., Marr, T., and Taylor, D.L. (2005) The Fungal Metagenomics Project. [web page] June 200. http://iab-devel.iab.alaska.edu/metagenomics [Accessed April 2006].

Nelson, J.R., Cai, Y.C., Giesler, T.L., Farchaus, J.W., Sundaram, S.T., Ortiz-Rivera, M. et al. (2002) TempliPhi, phi 29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *Biotechniques*: 44-47.

White, T.J., Bruns, T., Lee, S., and Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications.* Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. (eds). San Diego: Academic Press, Inc., pp. 315-322.