Soil Biology & Biochemistry xxx (2013) 1–5

Contents lists available at SciVerse ScienceDirect

Soil Biology & Biochemistry

journal homepage: www.elsevier.com/locate/soilbio

Short communication

Illumina metabarcoding of a soil fungal community

Philipp-André Schmidt [a,b], Miklós Bálint [a,**,1], Bastian Greshake [a], Cornelia Bandow [a,c], Jörg Römbke [c], Imke Schmitt [a,b,*,1]

[a] Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt/Main, Germany
[b] Goethe Universität Frankfurt, Institut für Ökologie, Evolution und Diversität, Max-von-Laue-Str. 13, 60438 Frankfurt/Main, Germany
[c] ECT Oekotoxikologie GmbH, 65439 Flörsheim/Main, Germany

ARTICLE INFO

Article history:
Received 8 January 2013
Received in revised form
6 May 2013
Accepted 10 May 2013
Available online xxx

Keywords:
Illumina MiSeq
Soil fungi
Metabarcoding
Next-generation sequencing
Annealing temperature
Community richness
Distance decay
PCR bias

ABSTRACT

Next generation metabarcoding is becoming an indispensable tool in fungal community ecology. Here we tested Illumina metabarcoding, a method that generates shorter reads but achieves deeper sequencing than 454 metabarcoding approaches. We found that paired-end Illumina MiSeq data cover the full ITS1 in many fungal lineages and are suitable for environmental fungal community assessment. There was substantial read loss during data cleanup (78.6%), which, however, did not impede the analyses, because of the large number of initial sequences (over 4Mio). We observed a high stochasticity in individual PCR reactions. Comparing three repeated sets of PCRs products showed that 58.5% of the total fungal operational taxonomic units (OTUs) found were not recovered by any single set of PCR reactions. Similarly, comparing three annealing temperatures showed that 63.6% of all fungal OTUs were not recovered using any single annealing temperature. These findings suggest that sampling of soil fungal communities is more exhaustive, if we combine repeated PCR products, and PCR products generated at various annealing temperatures. To analyze the above issues we sampled 16 soil cores along a 270 cm transect in a meadow. In total we recovered 3320 fungal OTUs (based on a 95% similarity threshold). Distance decay analysis indicated that community similarity decreased slightly with geographical distance.

High throughput metabarcoding has been recognized as a powerful tool to study fungal communities. To date the majority of studies using next generation sequencing of soil fungi are based on 454 pyrosequencing (e.g. Öpik et al., 2009; Buée et al., 2009; Dumbrell et al., 2011). Modern 454 reads cover >400 base pairs of fungal barcode markers such as ITS rDNA or partial 18S rDNA. However, most fungal 454 metabarcoding studies use either ITS1 or ITS2 (Buée et al., 2009; Jumpponen and Jones, 2009; Jumpponen et al., 2010; Cordier et al., 2012; Danielsen et al., 2012; Xu et al., 2012; Zimmerman and Vitousek, 2012). Even if the full ITS is amplified, the most common 454 chemistry recovers either ITS1 or ITS2 in full length, while the sequenced fraction of the other ITS subfragment is too short for inferences (Bálint et al., 2013; Bazzicalupo et al., 2013). Fragments in the size range of ITS1 or ITS2 can be readily sequenced on the Illumina MiSeq platform. The Illumina platform provides sequencing at greater depth for a considerably lower price compared to 454, and this promises a deeper characterization of fungal communities.

While the potential of high throughput metabarcoding is undisputed for studying complex fungal communities, our ability to understand the ecology of these communities has been hampered by insufficient sequencing depth and the high cost of 454 sequencing. For example, OTU rarefaction curves were often not saturated (e.g. Jumpponen et al., 2010), or processing of large numbers of replicates was not feasible due to the unfavorable throughput/price ratio (Bálint et al., 2013). To improve our understanding of the community composition and distribution of complex fungal communities we need to achieve deeper sequencing, and analyze larger numbers of samples per study (Caporaso et al., 2012). Increasing the number of replicates allows a more thorough evaluation of methodological biases inherent in metabarcoding, e.g. stochasticity of individual PCR reactions, and improves the statistical power of downstream data analyses. The aim of this study was therefore to test a method that generates a greater number of fungal metabarcodes at a lower cost. We specifically addressed the following questions: 1) What proportion of

* Corresponding author. Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt/Main, Germany. Tel.: +49 6979824767; fax: +49 6979824771.
** Corresponding author. Tel.: +49 6979824798; fax: +49 6979824771.
E-mail addresses: mbalint@senckenberg.de (M. Bálint), ischmitt@senckenberg.de, imke.schmitt@senckenberg.de (I. Schmitt).
[1] Authors contributed equally.

0038-0717/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.soilbio.2013.05.014

Illumina raw data is lost due to methodological issues and primer non-specificity 2) Do replicate PCRs produce similar diversity estimates in metabarcoding studies? 3) Do different annealing temperatures affect the number and identity of OTUs recovered in Illumina metabarcoding? We performed the evaluations on a simple system: soil fungal communities along a short (270 cm) transect in a meadow. We show that Illumina metabarcoding is feasible for analyzing spatial community structure. The high sequencing depth may help to better understand the fine-scale complexity in soil fungal communities.

We collected 16 soil samples along a 270 cm transect from a low input meadow located at Flörsheim, Germany (N50° 0′ 26.482″ E8° 23′ 58.502″, see Supplementary Fig. S1 for transect design). The soil is alluvial silty clay, with a pH ($CaCl_2$) of 6.9, and an organic matter content of 2.9% (see also Supplementary Table T1). The site has not been treated with fertilizers or pesticides for at least 10 years. Surface soil cores were 1.6 cm in diameter and 10 cm deep. We removed the vegetation cover before sampling, homogenized the cores, and kept samples at −20 °C until DNA extraction. After drying at 45 °C we used 300 mg from each core for DNA extraction with the FastDNA SPIN Kit for Soil (MP Biomedicals, USA). We amplified the ITS1 region using the newly developed primer ITS1FI2, 5′-GAACCWGCGGARGGATCA-3′ and ITS2 (White et al., 1990a). ITS1FI2 overlaps in six positions with ITS1F (White et al., 1990b), but is located closer to the end of the 18S. We used combinatorial primer labeling to identify samples after multiplexed sequencing (Gloor et al., 2010). Amplifications were carried out in a total volume of 20 μl using 50 ng of DNA, 4 μl of HOT MOLPol Blend Master Mix (Molegene, Germany), and 0.5 μM of each of forward and reverse primers. PCR conditions were 15 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at either 52 °C, 55 °C or 58 °C, and 30 s at 72 °C. The PCR with 52 °C annealing temperature was repeated three times. Final elongation was done at 72 °C for 5 min. Amplicons from the five parallel PCR runs (3 × 52 °C, 1 × 55 °C, 1 × 58 °C) were individually labeled to estimate the effect of repeated PCRs and annealing temperatures on richness recovery. Purification was done with Agencourt AMPure XP SPRI magnetic beads. PCR products were normalized and pooled. We normalized PCR products after quantifying them with a Qubit 2.0 Fluorometer (Invitrogen), with Qubit dsDNA HS Assay Kit (Invitrogen). Paired-end sequencing (2 × 150 bp) was carried out on an Illumina MiSeq sequencer at the Biomedical Genomics Center of the University of Minnesota, U.S.A.

We assembled paired-end reads using PandaSeq (Masella et al., 2012), and filtered out all sequences containing "N"s. If there were mismatches between the overlapping fragments of the forward and reverse reads, these were corrected according to the base call with the higher sequencer-assigned quality score. The quality of the reads was checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, accessed on 1 August 2012). After demultiplexing the reads using fqgrep (https://github.com/indraniel/fqgrep, accessed on 1 August 2012) we kept only those sequences starting with a perfectly matching labeled primer sequence (omitting the first base pair, N-1). Initial denoising was performed with a 97% similarity clustering with the heuristic clustering algorithm UCLUST 2.1, implemented in USEARCH v.6.0.203 (Edgar, 2010). The longest sequences served as seeds for the preclustering. *De novo* chimera detection was performed with the UCHIME algorithm (Edgar et al., 2011). OTU picking was performed at 95% sequence similarity with a modified version of the OTUpipe 1.1.9 wrapper for UCLUST (http://drive5.com/otupipe/, accessed on 15 August 2012). The most abundant sequence types served as clustering seeds. Most sister species differ by less than 2–3% in fungi (Schoch et al., 2012). However, we opted for a more conserved clustering threshold to account for intragenomic ITS variability, which is present in many fungal lineages (e.g. Simon and Weiss, 2008; Kovács et al., 2011), and which may exceed the similarity thresholds commonly employed for OTU delimitation (Lindner and Banik, 2011). We excluded OTUs with less than 10 reads following a recommendation in the manual of the program OTUpipe wrapper (http://drive5.com/otupipe/otupipe_manual1.1.pdf). The centroid sequence of each cluster (a representative sequence from the most common sequence type in each OTU) was used for the annotation of fungal reads. Centroid sequences are provided as FASTA files in Supplementary Material 2. We used only fungal clusters for downstream analyses. We blasted (Altschul et al., 1997) the OTU-representative sequences against the entire GenBank nucleotide database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/nt*, downloaded on 18 October 2012). We parsed the BLAST outputs in MEGAN 4 (Huson et al., 2011) for taxonomic assignment (min. support: 1, min. score: 200, top percent: 5), and we retained clusters with supported fungal origin. The OTU-representative sequences and the abundance of OTUs in each PCR reaction are provided as Supplementary Material 3.

We estimated sample richness by computing rarefaction curves for samples amplified in five parallel PCR series (3 × 52 °C, 1 × 55 °C, 1 × 58 °C) and for all samples combined. Rarefaction curves were calculated in MOTHUR V.1.22.2 (Schloss et al., 2009). We combined the reads obtained from the parallel PCRs, and randomly sampled 20,000 reads for each of the 16 samples to account for differences in sequencing depth. We analyzed the spatial structure of fungal assemblages from the 16 soil cores in R v.2.15.2 (R Development Core Team, 2012). We calculated a Bray–Curtis community distance matrix using the vegdist command from the vegan v2.0-3 package (Oksanen et al., 2012). We correlated the community distance matrix with the Euclidean distances of the 16 cores with a Mantel test (9999 permutations). To test whether community similarities significantly decrease with increasing spatial distance we fitted a linear regression on the logarithmized Sørensen similarities of the fungal communities against their spatial distance (Nekola and White, 1999). Given the small sample size, we estimated the standard errors in the intercept and slope by jackknifing the linear model with a script modified from Millar et al. (2011).

Sequence data was deposited in the European Nucleotide Archive (ENA) as [accession number will be provided after acceptance]. Q1

We received a total of 4,280,264 raw reads from the sequencer. Read number decreased to 3,790,739 after paired-end assembly, to 2,640,085 after demultiplexing, and to 2,528,650 (using a 97% preclustering threshold) after chimera detection. Clustering at 95% sequence similarity successfully clustered 2,304,935 reads into OTUs. Taxonomic assignments of OTUs picked at 95% sequence similarities rendered 917,269 reads of fungal origin. Of the non-fungal sequences, 46.1% (627,160) were of plant origin, 48.4% (658,198) were not assignable, and 5.4% (73,664) belonged to other eukaryotes and bacteria. In each of the 16 soil samples we found 60,905 to 237,052 total reads, 20,658 to 97,377 fungal reads, and 1550 to 2963 fungal OTUs (Supplementary Table 2). Overall, we recorded 3320 fungal OTUs from the 16 soil cores combined. These counts exclude OTUs recovered with <10 reads. The 3320 OTUs include Ascomycota (78.14%), Basidiomycota (10.24%), Glomeromycota (3.61%) Chytridiomycota (0.36%), Blastocladiomycota (0.03%), and other fungal lineages (7.62%).

The three repeated sets of PCR amplifications using 52 °C annealing temperature yielded comparable numbers of sequencing reads (240,511, 182,624, 143,211) and fungal OTUs (Fig. 1). The identity of the OTUs deviated greatly between the three sets of PCR reactions. Only 1219 (41.5%) of the total 2937 OTUs in this PCR series were recovered in each of the three sets of PCR reactions (Fig. 2A).

Please cite this article in press as: Schmidt, P.-A., et al., Illumina metabarcoding of a soil fungal community, Soil Biology & Biochemistry (2013), http://dx.doi.org/10.1016/j.soilbio.2013.05.014