Methods

454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases

Leho Tedersoo1,2,*, R. Henrik Nilsson1,3,*, Kessy Abarenkov1, Teele Jairus1, Ave Sadam1, Irja Saar1, Mohammad Bahram3, Eneke Bechem4, George Chuyong4 and Urmas Kõljalg1

1Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, 40 Lai St, 51005 Tartu, Estonia; 2Natural History Museum of Tartu University, 46 Vanemuise Street, 51005 Tartu, Estonia; 3Department of Plant and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Göteborg, Sweden; 4Department of Plant and Animal Sciences, University of Buea, PO Box 63, Buea, Cameroon

Author for correspondence:
Leho Tedersoo
Tel: +372 7376222
Email: leho.tedersoo@ut.ee
Received: 14 April 2010
Accepted: 25 May 2010

doi: 10.1111/j.1469-8137.2010.03373.x

Key words: 454 massively parallel pyrosequencing, artifacts, biodiversity, community structure, ectomycorrhizal fungi, Korup National Park, PCR bias.

Summary

• Compared with Sanger sequencing-based methods, pyrosequencing provides orders of magnitude more data on the diversity of organisms in their natural habitat, but its technological biases and relative accuracy remain poorly understood.

• This study compares the performance of pyrosequencing and traditional sequencing for species’ recovery of ectomycorrhizal fungi on root tips in a Cameroonian rain forest and addresses biases related to multi-template PCR and pyrosequencing analyses.

• Pyrosequencing and the traditional method yielded qualitatively similar results, but there were slight, but significant, differences that affected the taxonomic view of the fungal community. We found that most pyrosequencing singletons were artifactual and contained a strongly elevated proportion of insertions compared with natural intra- and interspecific variation. The alternative primers, DNA extraction methods and PCR replicates strongly influenced the richness and community composition as recovered by pyrosequencing.

• Pyrosequencing offers a powerful alternative for the identification of ectomycorrhizal fungi in pooled root samples, but requires careful selection of molecular tools. A well-populated backbone database facilitates the detection of biological and technical artifacts. The pyrosequencing pipeline is available at http://unite.ut.ee/454pipeline.tgz.

Introduction

The vast majority of soil microbes, including fungi, cannot be cultivated and propagated in axenic conditions, but DNA-based methods provide insight into their occurrence, relative abundance and ecological niches. Nuclear ribosomal genes and markers, particularly the small subunit (SSU), are widely used as barcoding targets in prokaryotes because of the ease of amplification, the presence of universal primer sites and the alignability across phyla and domains (Cole et al., 2009). For fungi, however, the highly conserved SSU gene provides little species-level resolution, resulting in the lumping of closely related phylotypes and even genera. As a result of improved species-level resolution, the internal transcribed spacer (ITS) region at the 3’-end of the SSU (in eukaryotes) has been accepted as an exploratory barcoding target for fungal identification (Kõljalg et al., 2005; Ryberg et al., 2009).

Accurate identification of microbial taxa is essential, especially for the detection of pathogens and mutualists which...
Contrary to expectations, rain forests support lower EcM fungal richness relative to temperate ecosystems (Tedersoo & Nara, 2010). Roots of tropical EcM host trees harbour several fungal taxa with experimentally unproven mycorrhizal status (Tedersoo et al., 2007, 2010b; Diedhiou et al., 2010; Peay et al., 2010). Molecular techniques have been widely used for the identification of EcM fungi on root tips and in soil because of a paucity of reliable morphological features. Traditionally, individual root tips are subjected to morphotyping, separate DNA extraction, amplification and ITS-based sequencing or fingerprinting techniques for identification. However, sample preparation and molecular analyses for hundreds of individual root tips separately are time consuming and costly. Our primary aim was to evaluate the suitability of pyrosequencing in large-scale environmental surveys of fungi, and to address the problems associated with potential technical artifacts, fungal taxonomy and assignment to lifestyle. Our second aim was to test the effects of alternative ITS primers (fungi-specific vs universal) on species’ recovery and view of the fungal community structure within the pyrosequencing dataset.

Materials and Methods

Sampling

The study was performed in a 12-ha primeval rain forest site within a previously established ‘P-plot’ in Korup National Park, south-west Cameroon (5°01'N; 8°48'E, c. 120 m above sea level; Newbery et al., 1988, 1998). This area receives mean annual precipitation of 5100 mm and has a mean annual temperature of 26°C. The soils are shallow and rest on nutrient-poor sand. The site supports particularly species-rich and abundant EcM hosts of the caesalpinioide legumes. In terms of basal area, the site is dominated by Microberlinia bisulcata A. Chev., Tetramerodendron bifoliolata (Harms) Hauman, T. korupensis Wieringa (previously considered as T. morelifera Aubrév.) (Caesalpinioideae, Fabaceae) and Oubanguia alata (Baker f.) Exell & Hillc., Aphanocalyx microphyllus (Harms) Wieringa, Berlingia bracteosa Benth., Didelotia africana Baill., Afzelia bipindensis Harms (all Caesalpinioideae) and Uapaca staudtii Pax. (Phyllanthaceae).

In April 2008 (early rainy season), 115 root samples (15 × 15 cm to 5 cm depth) were randomly collected from the forest floor soil by the use of a sharpened spade and knife. The distance between samples was at least 8 m to minimize the effect of spatial autocorrelation. Samples were collected and transported out of the rain forest camp in groups of 30–40 within 3 d of collection and processed in...
the nearby Mundemba village within the next 2 d. The roots were soaked in water and cleaned of adhering debris. EcM roots were placed in Petri dishes and mycorrhizas were separated into morphotypes according to Agerer (1991) by the use of a portable dissecting microscope. The color and texture of the root tips and the abundance of hyphae, cystidia and rhizomorphs served as the most important characters for morphotyping. To reduce sampling biases, morphotyping was restricted to each sample, and several root tips from each morphotype per sample were subjected to molecular analyses. EcM root tips were suspended in CTAB buffer (1% cetyltrimethylammonium bromide, 100 mM Tris-HCl (pH 8.0), 1.4 M NaCl and 20 mM EDTA) for shipping to a molecular laboratory.

Molecular identification

Initially, the DNA of a single root tip from each morphotype was extracted using a Qiagen MagAttract 96 DNA Plant Kit (Qiagen, Crawley, UK), following the manufacturer’s instructions. Depending on the amount of available mycorrhizas, one to four individual root tips from each morphotype per sample were subjected to separate DNA extraction, PCR and sequencing. To improve sequence quality, tomentelloid and other morphotypes were initially amplified using the primer pairs ITSOF-T (5’-cttttctttcccctacctagg-3’), and ITSOF-T and LB-W (5’-cttttctttcccctacctagg-3’), respectively. DNA yielding no PCR product was amplified with the primer ITSOF-T combined with universal primers ITS4 (5’-ctcggcataattgatagc-3’) or ITS2 (5’-gctgattcctactgc-3’). Nuclear rDNA large subunit (nuLSU) and mitochondrial rDNA large subunit (mtLSU) were amplified in DNA samples that consistently yielded no PCR product or readable sequence chromatogram in ITS analysis. Primer pairs LR0r (5’-accgctgaaactctctgctcc-3’) and TW13 (5’-ctcgtgctgctgctgctgct-3’) and ML5 (5’-ctcggcaattatcctcaaag-3’) and ML6 (5’-cagtagagctgactggtc-3’) were applied for nuLSU and mtLSU, respectively. The thermal cycling parameters, PCR purification and sequencing protocols were those of Tedersoo et al. (2003). Based on the merging patterns of EcM lineages in particular follow Hibbett et al. (2007) and Tedersoo et al. (2010a), respectively.

To compare the performance of the traditional method and pyrosequencing, we re-selected a single root tip from each morphotype and sample combination (n = 387), and pooled and pulverized these root tips using 3-mm tungsten carbide beads (Qiagen) with a Mixer Mill 301 (Retsch GmbH, Haan, Germany). Half of the mixture was stored at −80°C for further reference, and the DNA of an equal proportion of the remaining root tip mixture (c. 5 mg fresh weight) was extracted in three replicates using a PowerMax Soil DNA Isolation Kit (MoBio Laboratories Inc., West Carlsbad, CA, USA) and two replicates of Roche Kit for Mammalian Tissue. We have used the latter successfully in species. Thus, we refer to ‘species’ instead of OTUs only when relying on a 97.0% ITS1 barcoding threshold in the traditional method and respective error-corrected pyrosequencing dataset. Higher level taxonomy of fungi in general and EcM lineages in particular follow Hibbett et al. (2007) and Tedersoo et al. (2010a), respectively.

Fig. 1 Relative performance of the internal transcribed spacer 1 (ITS1) barcoding threshold values in the ectomycorrhizal fungal community as based on the traditional identification method. (a) Decay of matching pairs of individual sequences with increasing threshold. (b) Exponential growth of the number of operational taxonomic units (OTUs) with increasing threshold.
previous studies (see Tedersoo et al., 2007 for the protocol). For further comparison, the DNA extracted for Sanger sequencing (using a Qiagen MagAttract 96 DNA Plant Kit from 387 individual root tips) was pooled. The DNA of all six mixtures was amplified in two replicate reactions using the forward primer ITS1F or ITS5 and reverse primer ITS2. Thus, the pyrosequencing dataset has two replicates of the two fixed factors, DNA extraction method (three levels with one, two or three extraction replicates) and primer pair (two levels), totalling 24 tagged samples. The ITS1F and ITS5 primers were linked to the sequencing primer A. Barcode sequences of 6 bp were inserted between the A primer and ITS1F primer sequences. Thus, the composite forward primers were 5'–GCCCTCCCTCGGCCCCATCACGnnnnnnctggcatcataggagagta-3' (ITS1F) and 5'–GCCCTCCTCCGCCCCATCACGnnnnnggaagtaaaagtcgtaacaagg-3' (ITS5), where the A primer is in capital letters, the barcode as 'n' and ITS1F or ITS5 in lower case letters. The 6-bp barcodes were designed according to Parameswaran et al. (2007) and Opik et al. (2009) to avoid more than two successive occurrences of the same nucleotide.

The PCR conditions follow Opik et al. (2009): 20 µl of Qiagen HotStarTaq Master Mix, 0.23 µM of each of the primers and 2 µl of template DNA in a total volume of 40 µl. The reactions were run under the following conditions: 95°C for 15 min; five cycles of 42°C for 30 s, 72°C for 60 s and 92°C for 45 s; 35 cycles of 65°C for 30 s, 72°C for 60 s and 92°C for 45 s, followed by 65°C for 30 s and 72°C for 10 min. The PCR products were separated by electrophoresis through a 1.5% agarose gel in 0.5 × TBE (45 mM Tris Base, 45 mM Boric Acid, 1 mM EDTA (pH 8.0)) and purified from the gel using a Qiagen QIAquick Gel Extraction kit. The amount of DNA in the purified PCR products was measured using NanoDrop 1000 (Thermo Scientific, Wilmington, DE, USA). The products were mixed at equimolar concentrations. A total of 3 µg of these products was subjected to pyrosequencing on a Genome Sequencer FLX™ (454 Life Sciences Corp., Bradford, CT, USA) at GATC Biotech (Constance, Germany).

Bioinformatic processing

Pyrosequencing provided 64 171 individual sequences spanning the SSU gene, ITS1 region and sometimes extending into the 5.8S rDNA gene. Because these highly conserved ribosomal genes flanking the ITS1 marker may distort sequence clustering and similarity searches (Nilsson et al., 2009a), we removed these from the dataset using Fungal ITS Extractor 1.1 (Nilsson et al., 2010b). We also excluded sequences for which the tag, forward primer and neither of the flanking genes could be detected, and that remained below the quality threshold as revealed by Newbler software (454 Life Sciences Corp.). Therefore, all subsequent steps refer to the ITS1 sequence data only. In addition, sequences that were < 140 bp in length after trimming, or that contained > 3% ambiguous bases, were removed from further consideration. The remaining 44 411 sequences were partitioned according to their 6-bp barcode. The ITS1 sequences derived from the traditional approach were similarly trimmed to provide a database for the identification of EcM OTUs. The sequences were clustered in TGICL with the requirement of ≥ 90.0% coverage and ≤ 50.0% overhang at the 3'-end. The latter option was necessary to ensure that highly similar sequences of different length were correctly clustered. DNA barcoding thresholds of 90.0–99.0% (with an increment of 1.0%) were used to differentiate OTUs. Based on the results of the traditional approach, we performed all subsequent analyses using a 97.0% sequence similarity threshold. One of the longest sequences was randomly selected as a representative of each OTU. These sequences were subjected to a bulk BLASTN search against all ITS sequences in the International Nucleotide Sequence Databases (INSD) and UNITE (Abarenkov et al., 2010) for identification. In a few cases, manual BLASTN searches were performed separately for the 5'- and 3’-ends of ITS1 sequences to detect potential chimeras. Representative sequences of each fungal OTU were submitted to EMBL (accession numbers FN669299–FN669501). The raw sequence data are available from the UNITE database (http://unite.ut.ee/454_EcM_CMR.zip). The pyrosequencing pipeline that can be used to analyze any large fungal ITS-based dataset is available at http://unite.ut.ee/454_pipeline.tgz.

Statistical analysis

To confirm the BLASTN-based identification and to validate the pyrosequencing results, we subjected the representative sequence of each species (the ITS1 sequence from pyrosequencing and full ITS region from the traditional method) to a global alignment using MAFFT 6.717 (Katoh & Toh, 2008). We used the default options of the G-INS-i strategy of MAFFT (gap penalty = 1.0; offset value = 0.1). Separate alignments were prepared for the /amanita, /boletus, /russula–lactarius, /tomentella–thelephora, /marcellina–peziza gerardii and /sordariales lineages. For the last two groups, which were found to be problematic in terms of identification, assignment to lifestyle and/or species’ delimitation, all available ITS sequences in INS and UNITE were included. For the alignment of the /sordariales lineage, several closely related, putatively nonEcM taxa were also included as an outgroup. All other nonEcM taxa and EcM taxa that belong to other lineages were forced to align using the same parameters. Seaview 3.2 (Galtier et al., 1996) was used to manually correct the alignments and trim some overhanging ends. Based on each alignment, a maximum likelihood analysis with fast bootstrap calculation was
performed in RaXML 7.2.3 (Stamatakis et al., 2008). The alignments were used to detect the amount of transitions, transversions and indels within and among species of Sanger sequencing data, and in pyrosequencing singletons (i.e. OTUs represented by a single sequence), when compared with their closest species (neglecting comparisons with < 90% sequence similarity). In pyrosequencing singletons, insertions and deletions were considered separately relative to the sequence of the closest species. Homopolymers were defined as > 2-bp repeats of a single nucleotide. Differences in the proportions of indels, transitions and transversions were tested for significance using chi-squared and G-tests. Singletons were considered to be artifactual when chimeric, when erroneously unassigned to contigs or when possessing a long branch relative to a sister taxon with zero branch length as revealed in phylograms.

To assess the robustness of pyrosequencing, we studied the effects of the DNA extraction method, extraction replicate and PCR primer on the perceived species richness and community composition of EcM fungi. These variables also served to evaluate the differences among the identification methods. EstimateS 8.0 (Colwell, 2006) was used to generate sample-based rarefaction curves for the traditional approach and Coleman curves for the pyrosequencing dataset by sampling randomly without replacement. Coleman curves were computed for each PCR product, DNA extraction replicate and extraction method, as well as for the whole dataset. Separate Coleman curves were calculated for communities constructed by the use of different DNA barcoding thresholds. Because of the memory consumption of EstimateS, we were unable to compute the Coleman curves for the 98% and 99% barcoding thresholds.

Nested two-way ANOVAs were used to test the similarity in the number of sequences and OTUs recovered, the proportion of fungi and EcM species accumulation. In these analyses, the PCR primer pair, DNA extraction method and DNA extraction replicate (nested within the DNA extraction method) served as independent variables and the 24 PCR products represented replicates. The relative effect of these variables on the perceived EcM fungal community structure was further tested using the ADONIS routine, and visualized using nonmetric multidimensional scaling (NMS) of the Vegan package of R (R Core Development Team, 2007). The Wisconsin-standardized species abundance and Bray–Curtis distance were used for ADONIS, and species occurrence and Sørensen distance for NMS.

Results

Traditional identification method

Using one to four replicate root tips for sequencing, 77.8% of the morphotype and sample combinations were successfully identified. These yielded a single EcM fungal species in 90.1% of cases. When identification based on the ITS region failed in spite of using multiple primers and DNA extractions, mtLSU and/or nuLSU regions could sometimes be amplified and sequenced. Based on BLASTN queries, these non ITS sequences were assigned to the /boletus, /sebacina, /marcellina–peziza gerardii and /cantharellus lineages, the latter of which was absent from the ITS library.

Based on the ITS region, the traditional identification method revealed 111 species of EcM fungi on 326 successfully sequenced root tips from 106 root samples. Of these taxa, 54 occurred in a single sample and 40 on a single root tip. Clustering of individual sequences at various similarity levels revealed that most converged at > 97.0% sequence similarity, whereas a few merged at the level of 90.0–97.0% sequence similarity (Fig. 1), justifying the selection of the 97.0% DNA barcoding threshold.

The /russula–lactarius, /tomentella–thelephora, /boletus and /amanita lineages dominated in terms of species richness and comprised 44, 20, 12 and eight species, respectively. The individual species of /russula–lactarius sp. CAM19, /piloderma sp. CAM01 and /tomentella–thelephora sp. CAM12 were found in the greatest number of samples (Supporting Information Table S1).

Validation of pyrosequencing

A total of 44 411 sequences (median length, 195 bases; range, 141–233 bases) passed the various quality control steps and were subsequently assigned to OTUs based on 90.0–99.0% sequence similarity of the ITS1 region. In agreement with the traditional method, the number of OTUs recovered based on the 90.0–97.0% threshold produced relatively similar results with regard to the total species richness (Fig. 2). Largely as a result of the

![Fig. 2 Accumulation of operational taxonomic units (OTUs) with different internal transcribed spacer 1 (ITS1) barcoding threshold values in the ectomycorrhizal fungal community as based on the original pyrosequencing dataset.](image-url)
A considerable number of singletons, the 99.0% and 98.0% thresholds produced very high estimates of the OTU richness (Fig. 3), suggesting the presence of either a large number of rare, closely related taxa or sequencing artifacts. Using a 97.0% ITS1 similarity barcoding threshold, pyrosequencing revealed 312 OTUs. Of these, 87 (27.9%) were singletons and 19 (6.1%) were doubletons. The results of BLASTN searches suggested that these 312 OTUs represented 243 (71.8%) taxa of fungi (including three obvious laboratory contaminants), 20 (6.4%) taxa of plants (all matching closest to Caesalpinioideae host plants) and a single nematode taxon, whereas 48 (15.4%) were unassigned to any higher taxon. The alignment of taxonomically unassigned sequences with known fungal taxa derived from pyrosequencing and traditional methods, combined with phylogenetic analyses, suggested fungal affinities for 42 (87.5%) of these taxa, including 33 ECM OTUs that were found to belong to the /marcellina–peziza gerardii, /elaphomyces and /boletus lineages. The distribution of taxa among eukaryote kingdoms and fungi is illustrated in Fig. 4.

Pyrosequencing revealed the presence of 25 closely related ribotypes in the /marcellina–peziza gerardii lineage (Fig. S1a). This finding strongly contrasts with the results of Sanger sequencing which revealed a single species based on the ITS region and an additional taxon based on nuLSU. These 25 ribotypes were all phylogenetically closely related, but their internal phylogenetic structure was unresolved. This observation contrasts with that made for taxa from other regions and lineages (Fig. S1a–g). Nevertheless, the primer sites and flanking gene regions of these 25 ribotypes were well conserved, suggesting that these rDNA alleles are potentially functional. Despite this, the taxa were manually merged into two species (see Discussion for further explanation).

The assignment of a nutritional mode to another group of OTUs belonging to Sordariales was ambiguous on the basis of BLASTN matches. Phylogenetic analyses relying on additional sequence data from INSD and nonmycorrhizal root-associated fungi from the study site suggested the placement of most of these OTUs outside any ECM fungal lineage (Fig. S1b).

None of the pyrosequencing singletons corresponded to any species recovered based on the traditional approach at the 97.0% ITS1 similarity level. The subsequent quality check revealed that sequences of singletons (relative to their closest species) had 3.2 times higher proportions of indels compared with the inter- and intraspecific variation as revealed by Sanger sequencing (chi-squared test: \( \chi^2 = 130; \) df = 1; \( P < 0.001 \); Fig. 5). Within singletons, insertions were 9.6 times more common than deletions (G-test: \( G = 84.3; \) df = 1; \( P < 0.001 \)). However, there were no differences in indel accumulation in homopolymers vs monomers and dimers of pyrosequencing singletons compared with inter- and intraspecific variation (chi-squared test: \( \chi^2 = 1.20; \) df = 1; \( P < 0.274 \)). By contrast, transitions mostly contributed to intra- and interspecific variation, but were underrepresented in pyrosequencing singletons (Fig. 5). Among the 45 ECM singletons, we detected 29 (64.4%) of such artifactual sequences, a single putative chimera, a single reverse complementary sequence, three sequences that were erroneously unassigned to species and only 11 (24.4%) sequences of potentially high quality. In phylograms, artifactual singletons nearly always occupied
were represented by 33 (10.1%) root tips and 4435 revealed 17 (15.0%) and 30 (24.2%) unique species that were shared. The traditional method and pyrosequencing revealed 141 species of EcM fungi, 94 (66.7%) of which were recovered only by the use of the pyrosequencing method (Fig. S3). By contrast, the /boletus, /russula–lactarius, /amanita and /elaphomyces lineages contributed four, three, three and two species, respectively, found only by the traditional approach. The five most species-rich lineages differed significantly in the proportion of unique species recovered by the traditional approach (Fisher’s exact test: df = 4; P = 0.019), but the difference was marginally nonsignificant for pyrosequencing (df = 4; P = 0.080). Because six species were recovered more than once with the traditional approach only, we assessed whether the length of the ITS1 sequence or a mismatch with the ITS2 primer may cause this discrepancy. However, species recovery using pyrosequencing was independent of these variables, although five OTUs belonging to the /russula-lactarius, /amanita and /elaphomyces lineages featured one or more mismatches in the centre or the 5′-terminus of the ITS1 marker.

Biases within pyrosequencing data

Of the forward primers, ITS5 recovered significantly more sequences (F1,18 = 19.4; P < 0.001), including those assigned to EcM fungi (F1,18 = 18.5; P < 0.001), and revealed a greater proportion of high-quality sequences (F1,18 = 6.9; P = 0.017) and nonfungal sequences (F1,18 = 13.2; P = 0.002; Table S2) relative to the ITS1F primer. DNA extraction techniques had a strong effect on OTU recovery at the same sampling intensity of 700 sequences (nested ANOVA: F2,18 = 18.5; P < 0.001; Fig. S4), but the extraction replicate and primer choice had no effect. The DNA extraction method (ADONIS: F1,12 = 2.76; P = 0.002), primer choice (F1,12 = 5.12; P < 0.001) and their interaction term (F2,12 = 3.38; P < 0.001) had a strong effect on EcM fungal community structure, explaining 15.7%, 14.5% and 19.2%, respectively, of the residual variation in the nested multivariate model (Fig. 6). Individual PCRs varied greatly in the number and proportion of species recovered (Figs S4, S5). In addition, most rare taxa (n = 2–5) were recovered only in a single PCR. For example, 16 of 19 doubletons were recovered during a single PCR, which is significantly different from the expected ratio of 0.042 (G-test: G = 18.9; P < 0.001).

Discussion

Pyrosequencing vs traditional method

Pyrosequencing and the traditional identification method revealed a roughly similar phylogenetic structure in the EcM fungal community that resembled other tropical ecosystems (Tedersoo & Nara, 2010). The /russula–lactarius lineage

Fig. 5 Proportion of nucleotide substitutions and indels in the internal transcribed spacer 1 (ITS1) marker: (a) within species (at 97.0% similarity level; n = 74); (b) among species (at 90.0–96.9% similarity level, n = 141) based on the traditional identification method, (c) between sequences of putative artifacts and their respective parent species as revealed by pyrosequencing (n = 218); (d) expected proportions assuming equal representation of bases and equal probability of any nucleotide substitution and indel formation. Blue, C–T transitions; cyan, A–G transitions; green, transversions; orange, single nucleotide indels; dark orange, indels > 1 bp; red, single nucleotide deletions; brown, > 1 nucleotide deletions; yellow, single nucleotide insertions; dark yellow, > 1 nucleotide insertions. Hatched sectors, indels associated with homopolymers. Note that the within-species and among-species difference was not significant (χ2 = 2.48; df = 3; P = 0.478).

Traditional method vs pyrosequencing

Combining traditional and pyrosequencing approaches revealed 141 species of EcM fungi, 94 (66.7%) of which were shared. The traditional method and pyrosequencing revealed 17 (15.0%) and 30 (24.2%) unique species that were represented by 33 (10.1%) root tips and 4435 (10.0%) sequences, respectively. The /russula–lactarius, /sebacina, /amanita and /tomentella–telephora lineages comprised ten, five, four and three species, respectively, that were recovered only by the use of the pyrosequencing method (Fig. S3).
was the species-rich, followed by the /tomentella-thelephora, /boletus, /amanita and /sebacina lineages. Although the proportions of the dominant lineages remained similar, the species unique to either identification method tended to be nonrandomly distributed among lineages. The ITS region of several taxa was not captured by the traditional approach. Thus, compared with pyrosequencing, the traditional approach is likely to be more biased towards the representation of certain lineages. By contrast, the fact that some of the species were not picked up by pyrosequencing may be explained by the use of different root tips in the traditional and pyrosequencing approaches (material extracted using Mo Bio and Roche kits). However, the high frequency of some of these species in the Sanger sequencing dataset suggests that the pyrosequencing technique has technical biases other than primer mismatches and differential sequence length. As a result of PCR biases, pyrosequencing (Porazinska et al., 2009) and terminal restriction fragment length polymorphism (Avis et al., 2010; Dickie et al., 2010) underestimated species richness in the artificially pooled DNA mixtures.

Pyrosequencing artifacts

BLASTN searches against the INSD database provided an incomplete solution to distinguish EcM fungi from nonmycorrhizal fungi. Several EcM lineages (/boletus, /amanita, /inocybe, /piloderma and all Ascomycota) included species (7.3% of the final pyrosequencing dataset) that were not matched to any organism or members of their lineage because of the lack of published sequences from vouchered collections and the relatively high divergence of the ITS region. Full-length ITS sequences and information on EcM morphology provided the assignment of a nutritional mode for many of these taxa.

The comparison of pyrosequencing with the traditional identification method enabled us to address technical artifacts and biological outliers in the pyrosequencing dataset. Pyrosequencing revealed high richness of ribotypes (13.9% of the nonsingletons) within two species of the /marcelleina–peziza gerardii lineage. Their absence in the Sanger sequencing dataset and a careful analysis of the alignment and phylogram suggest that these additional ITS1 ribotypes are biological rDNA artifacts. We speculate that these ribotypes may have arisen concomitantly from the two most common sequence types probably as a result of polyploidization events or relaxation of concerted evolution. Multiple divergent ITS sequences are common in the polyploid plants and Glomeromycota (e.g. Stockinger et al., 2010), but uncommon in EcM fungi (Hebeloma velutipes; Aanen et al., 2001). The presence of multiple alleles may disable direct sequencing of the ITS amplicon in certain fungal taxa (den Bakker et al., 2004; Hughes et al., 2009). Both pyrosequencing and cloning, and fingerprinting techniques, seem to resolve these biases, at least partly.

Alarming, none of the pyrosequencing singletons were matched to species recovered by the traditional identification method. Singletons comprised 27.9% of the OTUs and accounted for the greatest source of bias in the original pyrosequencing dataset. Careful examination of all EcM singletons (excluding the biological outliers from the /marcelleina–peziza gerardii lineage) revealed that 81.0% were potentially technical artifacts that included many erroneously detected bases. Indel accumulation in homopolymers is considered to be the major technical problem of 454 pyrosequencing technology (Margulies et al., 2005; Huse et al., 2007; Kunin et al., 2009). Singletons contained an elevated proportion of insertions compared with the biological inter- and intraspecific variation and expected proportions. The accumulation of indels in homopolymers was similarly high in pyrosequencing artifacts and biological variation compared with the expected proportions (Fig. 5). Because our analyses were based on mixed environmental DNA, we cannot claim with 100% confidence that the anticipated artifacts are truly derived from pyrosequencing per se. Nevertheless, our results suggest that these biases merit attention in situations in which the templates are known (Kunin et al., 2009). In addition to their low quality, singletons harboured a few sequences that were chimeric, reverse complementary or mistakenly unassigned.

Fig. 6 Nonmetric multidimensional scaling (NMS) ordination of the ectomycorrhizal fungal community as revealed by the use of the traditional identification method (asterisk) or pyrosequencing (other symbols). Stress value = 0.201; $R^2 = 0.080$. Circles, communities recovered using the Mo Bio PowerMax Soil DNA Isolation Kit; triangles, Roche Kit for Mammalian Tissue; squares, Qiagen MagAttract 96 DNA Plant Kit. Colors represent different DNA extraction replicates. Open symbols, PCR amplification using internal transcribed spacer 1F (ITS1F) and ITS2 primers; closed symbols, PCR amplification using ITS5 and ITS2 primers.
to contigs (as a result of clustering errors). Because of the nonbiological nature of most, perhaps all, singletons, we excluded them from our analyses and encourage this conservative procedure for other studies in similar systems. Both sequence errors and chimeras may strongly inflate diversity estimates (Reeder & Knight, 2009) and contaminate sequence databases with nonbiological data (Harris, 2003; Nilsson et al., 2010a).

Technical biases

Within the pyrosequencing dataset, the choice of primers strongly affected the number of total and EcM fungal sequences recovered, but not species accumulation or community composition as based on the ITS1 marker. The ITS5 primer performed better than the ITS1F primer by providing both more raw sequences and a greater proportion of high-quality sequences. As expected, the fungi-specific primer ITS1F produced a greater proportion of fungal sequences compared with the universal ITS5 primer, but amplified trace amounts of plant DNA as well (as in Jumpponen & Jones, 2010). Because the PCR products of different primer pairs are of similar length (14 bp difference) and we used equal amounts of products for pyrosequencing, we speculate that the ITS5 primer recovered more raw sequences because of enhanced ‘competitive ability’ for binding the beads. In addition to primers, DNA extraction methods had a substantial effect on the number of sequences and species recovered, and on EcM fungal community composition, which could be related to the differential ability to degrade cell walls and/or remove inhibitors. The inclusion of multiple PCR replicates substantially enhanced OTU recovery, as 25.2% of nonsingleton OTUs of the original dataset proved to be unique to a single PCR, whereas all 24 PCR replicates shared only 3.1% of OTUs (Fig. S4, Table S1). Gomez-Alvarez et al. (2009) recently found that sequencing artifacts can be found among nonsingleton taxa as well because of the post-PCR replication of artificial molecules.

Methodological implications

Pyrosequencing is a powerful alternative to traditional identification techniques in terms of cost, time and throughput of samples (Fierer et al., 2008; Jumpponen et al., 2010). Pyrosequencing is likely to recover more species and provide a less biased qualitative picture of the microbial community composition (Sogin et al., 2006; Opik et al., 2009) but, because of PCR biases and technical errors, caution should be used to interpret quantitative aspects and the biological meaning of the data (Reeder & Knight, 2009; Medinger et al., 2010). Thus, a database of full-length sequences is especially important to recognize various artifacts and evaluate barcoding thresholds in previously unstudied ecosystems or taxa. Datasets comprising an order of magnitude more sequences are likely to contain a higher proportion of artifactual OTUs, because substantial sequencing errors tend to be unique, whereas the accumulation of biological data starts to level off (Reeder & Knight, 2009). Therefore, we recommend the removal or careful quality control of any sequences that prove to be singletons to avoid the release of numerous potential artifacts.

To overcome the PCR biases in environmental studies including pyrosequencing, it is recommended that DNA extraction protocols are optimized, multiple PCR replicates are used and the number of PCR cycles is reduced to 20–25 (Kanagawa, 2003; Medinger et al., 2010). To maximize species recovery, we recommend the performance of at least five PCR replicates which should be pooled into a common pyrosequencing analysis. Primers of choice should cover all target taxa (Taylor & McCormick, 2008); nontargeted organisms can be simply removed from datasets based on automated searches against databases or phylogenetic analyses. With the advancement of pyrosequencing technology, fungi can be better addressed by amplification of the entire ITS region using primers (e.g. ITS1 – 5‘-tcggagggacctgcg-3’ that avoid the co-amplification of the Group I intron at the 3‘-end of SSU which is particularly widespread in Ascomycota (Bhattacharya et al., 2000). Chimeric sequences may be a greater problem in analyses involving conserved genes (Porazinska et al., 2009) and full-length ITS sequences that cover the extremely conserved 5.8S gene (Nilsson et al., 2010a). In highly sensitive methods, such as pyrosequencing, trace laboratory contamination can be a common phenomenon and should be interpreted accordingly (Kunin et al., 2009).

Acknowledgements

We thank D. M. Newbery, L. Njume and J. Norghauer for site information and logistical support, M. Toots for help with R software and B. Lindahl, J. Norghauer, M. Opik, I.J. Alexander (the editor) and three anonymous referees for constructive comments on earlier drafts of the manuscript. This project was funded from ESF grants 6606, 7434, JD92 and FIBIR.

References


**Supporting Information**

Additional supporting information may be found in the online version of this article.

**Fig. S1** Phylogenetic placement of species and artifacts recovered by the traditional method and pyrosequencing.

**Fig. S2** Frequency distribution of the quality scores of 44,411 sequences that passed the quality control.

**Fig. S3** Species richness of 16 lineages of ectomycorrhizal fungi in the study site as revealed by the traditional identification method and pyrosequencing.

**Fig. S4** Operational taxonomic unit (OTU) accumulation curves of all organisms as recovered from PCR replicates of pyrosequencing, DNA extraction replicates, DNA extraction methods and the entire dataset.

**Fig. S5** Relative abundance of ectomycorrhizal fungal species as recovered by the four PCR replicates of a single DNA extract.

**Table S1** Identification and quantitative estimates of ectomycorrhizal fungi and other organisms as revealed by the use of the traditional method and pyrosequencing.

**Table S2** The effects of DNA extraction method, extraction replicate and primer pair on the performance of pyrosequencing.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.