

# Metatranscriptomic analysis of ectomycorrhizal roots reveals genes associated with *Piloderma*–*Pinus* symbiosis: improved methodologies for assessing gene expression *in situ*

H.-L. Liao,<sup>1\*</sup> Y. Chen,<sup>2</sup> T. D. Bruns,<sup>3</sup> K. G. Peay,<sup>4</sup>  
J. W. Taylor,<sup>3</sup> S. Branco,<sup>3</sup> J. M. Talbot<sup>4</sup> and  
R. Vilgalys<sup>1</sup>

Departments of <sup>1</sup>Biology and

<sup>2</sup>Medicine, Duke University, PO box 90338, Biological Sciences Building, Durham, NC 27708, USA.

<sup>3</sup>Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA.

<sup>4</sup>Department of Biology, Stanford University, Stanford, CA, USA.

## Summary

**Ectomycorrhizal (EM) fungi form symbiotic associations with plant roots that regulate nutrient exchange between forest plants and soil. Environmental metagenomics approaches that employ next-generation sequencing show great promise for studying EM symbioses; however, metatranscriptomic studies have been constrained by the inherent difficulties associated with isolation and sequencing of RNA from mycorrhizae. Here we apply an optimized method for combined DNA/RNA extraction using field-collected EM fungal–pine root clusters, together with protocols for taxonomic identification of expressed ribosomal RNA, and inference of EM function based on plant and fungal metatranscriptomics. We used transcribed portions of ribosomal RNA genes to identify several transcriptionally dominant fungal taxa associated with loblolly pine including *Amphinema*, *Russula* and *Piloderma* spp. One taxon, *Piloderma croceum*, has a publically available genome that allowed us to identify patterns of gene content and transcript abundance. Over 1500 abundantly expressed *Piloderma* genes were detected from mycorrhizal roots, including genes for protein metabolism, cell signalling, electron transport, terpene synthesis and other extracellular**

**activities. In contrast, *Piloderma* gene encoding an ammonia transporter showed highest transcript abundance in soil samples. Our methodology highlights the potential of metatranscriptomics to identify genes associated with symbiosis and ecosystem function using field-collected samples.**

## Introduction

The symbiotic activities of ectomycorrhizal (EM) fungi are essential for pine forest establishment and sustainability (Hasselquist *et al.*, 2005; Wang and Qiu, 2006; Smith and Read, 2008). The composition of EM fungal communities can be affected by environmental factors (Cullings and Makhija, 2001; Cullings *et al.*, 2003; Wolfe *et al.*, 2008). The EM symbiosis is characterized by a hyphal mantle that forms over root tips (Bonfante and Genre, 2010) where both epidermal and cortical root cells are colonized by intercellular hyphae (the Hartig net). EM fungi, through cell–cell interaction of their hyphal network with root cells, enjoy direct access to their plant host's carbon and compensate the plant host by stimulating plant growth (Sung *et al.*, 1995), mineral nutrient uptake (Van den Driessche, 1991), water absorption (Smith and Read, 2008) and disease resistance (Zhao *et al.*, 2008).

Genomic studies of EM fungi are providing novel insights into mechanisms of symbiosis. In *Laccaria bicolor* (the first EM fungus to have its genome sequenced), approximately 20 000 protein-encoding genes were identified including ammonia transporters, GTPase and small-secreted proteins involved in symbiosis establishment (Martin *et al.*, 2008). Genome sequencing of the black truffle (*Tuber melanosporum*) has also revealed many genes that play a role in EM–fungal–plant interaction including plant cell adhesion, plant defence evasion, and plant cell wall modifications (Martin *et al.*, 2010).

Molecular dissection of mycorrhizal symbiosis is complicated by many challenges, including the difficulty of culturing most EM fungi and the requirements for inoculation and successful EM colonization of living plant roots. For most EM fungi, *in vitro* gene expression analyses are still not practical or possible. Fortunately, metagenomic

Received 8 May, 2014; accepted 27 August, 2014. \*For correspondence. E-mail sunny.liao@duke.edu; Tel. 9196607362; Fax 919-660-7293.

methods that follow gene expression in field-collected samples show much promise for understanding the mechanisms by which different EM fungal lineages respond to their natural environment. Next-generation DNA sequencing studies using high-throughput amplicon sequencing especially show great promise for functional studies of EM fungal communities (Gottel *et al.*, 2011; Gugerli *et al.*, 2013; Shakya *et al.*, 2013; Talbot *et al.*, 2014). For example, a recent study of the North American pine soil 'mycobiome' showed strong geographic patterning of EM fungal communities but high degrees of functional convergence across these communities (Talbot *et al.*, 2014). However, only few methods have been applied to study EM function and they are based on relatively coarse measures of community enzyme production or gross hyphal morphology (Courty *et al.*, 2005; Moeller *et al.*, 2014). Consequently, application of finer resolution metagenomic approaches shows great potential to advance the study of EM functional ecology. While transcriptomic tools are well developed for some model systems *in vitro* (*Laccaria*, *Tuber*) (Martin *et al.*, 2008; 2010; Plett *et al.*, 2011), metatranscriptomics has not been widely applied to study the function of ectomycorrhizae in the natural environment. This is due to the difficulty of obtaining high-quality RNA from a small amount of field tissues and the challenge of assembling and analysing metatranscriptomic data. To study the functional distribution of EM fungal hyphal network in both root and soil systems, we collected root clusters and soils from two *Pinus taeda* native forests. Here we propose a methodology to study the metatranscriptome of EM fungi–pine root symbiotic associations from small samples (pine root clusters) using next-generation sequencing technology and advanced assemblers that use data from the rapidly growing number of transcriptomic sequence databases. To test our methodology, we used the *Pinus–Piloderma* system because of the large amount of publicly available transcriptomic data from these two species (Grigoriev *et al.*, 2012; NCBI). Our methodology identifies a core set of *Piloderma* genes expressed under a variety of developmental/environmental conditions.

## Results

### *Sampling and RNA extraction for field-collected samples*

To study EM fungal–*Pinus* interactions *in situ*, we sampled at least 12 replicates of EM root clusters and soils from two *P. taeda* native forests 10 mile apart (Fig. S1). The organic soil horizons were collected using soil cores. In each plot, two soil cores were collected from two points 50 m apart from each other (i.e. four soil cores were collected in total). Intact root clusters containing fresh root tips were collected using forceps under a dissection microscope. Soil samples were also collected from the

same soil core from which roots were collected. At least three individual root and soil samples were collected from a single core as subreplicates.

Application of metagenomic approaches presents challenges due to the difficulty of obtaining high-quality RNA from a small amount of field tissues, especially from plant roots (Chang *et al.*, 1993; Martin *et al.*, 2004), and the challenge of assembly of whole-metatranscriptomic data. Because initial attempts to isolate RNA from single mycorrhizal root tip yielded insufficient RNA for HiSeq, we chose to focus on EM root clusters, which are larger and more suitable for obtaining high-quality RNA, next-generation sequencing and comparative metatranscriptomics at the single-species level. Because high-quality mRNA is critical for transcriptomics, choosing the right extraction method is critical for field-collected root cores with small amounts of usable tissue. Of the methods we compared (Table S1), including kit-based and conventional methods, the best results were obtained using a cetyltrimethyl ammonium bromide (CTAB)/chloroform extraction with LiCl precipitation of RNA. The CTAB/LiCl method described by Chang and colleagues (1993) was able to recover good quality RNA from root samples of *P. taeda*; however, the quantity of recovered RNA was relatively low (~ 6 ng RNA mg<sup>-1</sup> root cluster). Several CTAB methods modified from Chang and colleagues (1993) or Liao and colleagues (2004) were applied to extract RNA from larger amount of pine tissues (1–4 g) (Joosen *et al.*, 2006; Lorenz *et al.*, 2009; 2010; Canales *et al.*, 2011). In this study, we modified the RNA extraction method from Chang and colleagues (1993) combined with genome grinding/beading strategy to homogenize small amounts of root tissue. The improved methods enabled us to obtain good quantity of DNA and RNA from the same extraction in a single root cluster (~ 100 ng RNA mg<sup>-1</sup> tissues). RNA isolation from single root clusters yielded of 0.5 µg high-quality RNA, sufficient for the production of quality reads by Illumina RNAseq. Except RNA PowerSoil kit (MoBio, Carlsbad, CA, USA), all of the RNA extraction methods we examined were not adopted for soil samples collected in the field (Table S1).

### *Integrative assembly of cDNA sequences*

We performed Illumina HiSeq sequencing to analyse the transcriptomic activity for individual root clusters and soil samples respectively. We recovered an average of 40 million reads that passed quality control from a single EM root cluster (Table S2). The computational workflow sorted out the reads representing fungal ribosomal RNA (rRNA) using fungal rRNA databases (Fig. 1). *De novo* assembly using the Velvet package (Zerbino and Birney, 2008) was applied to assemble and identify the variable region (divergent domain 2, D2) of rRNA large subunit

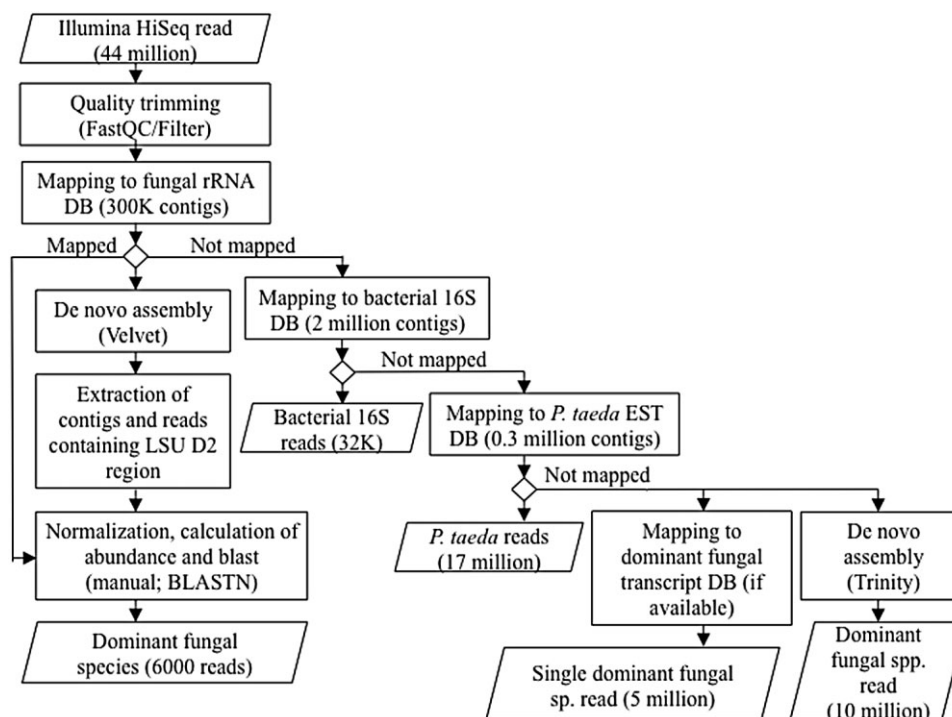


Fig. 1. Computational flowchart used to analyse Illumina HiSeq RNA reads of EM root clusters.

regions (LSUs) of fungal species. The assembled contigs containing D2 regions were used as a database for fungal rRNA reads aligning to calculate the abundance of particular fungi. Short rRNA reads representing the LSU D2 region (around 180 bp) were extracted, paired and normalized using RPKM (reads per kilobase per million). The fungal taxonomic identity (at the genus level) was determined using BLAST (Bayesian classifier and NCBI) and dominant fungal lineages identified. The unmapped reads were subsequently mapped to bacterial 16S rRNA and *P. taeda* transcript database (NCBI). Of the transcriptionally active EM fungi, *Piloderma* spp. were most often detected from root samples we collected (the details were described in the following sections). Thus, the *Piloderma croceum* database (Pilcr1 F 1598; Grigoriev *et al.*, 2012) was used to extract reads representing *Piloderma* genes from *Piloderma* dominated root clusters and soil samples. *De novo* assembly using Trinity package was applied to recover the fungal genes for the samples that are dominated by the fungal species without publicly available databases.

#### The distribution of reads in individual root clusters

Figure 2 shows the percentage of reads within a typical EM root cluster associated with different metatranscripts. Our approach was able to recover reads from functional genes, with 35% from fungi and 42% from *P. taeda*. In

addition, smaller amounts of ribosomal DNA reads generated during cDNA synthesis were obtained which could be attributed to fungi (0.9%), *P. taeda* (0.8%) and bacteria (0.1%), as well as 0.007% reads that mapped to the fungal internal transcribed spacer (ITS) region. Similar ratios were observed between the biological replicates in

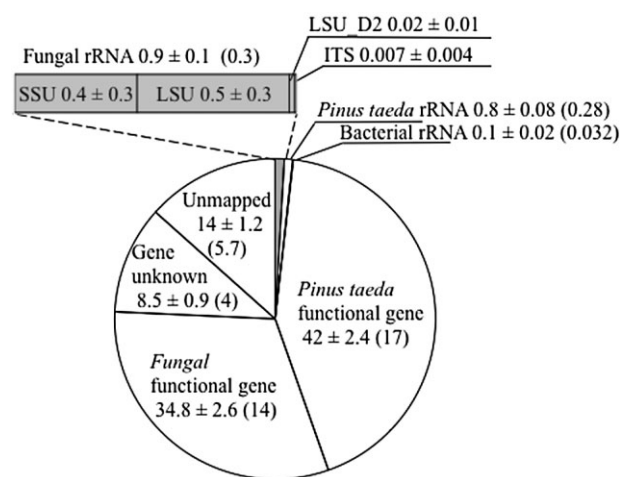


Fig. 2. Percent of read associated with fungal, bacterial and pine transcripts in EM root clusters detected by Illumina HiSeq. Numbers in parentheses indicate the number of reads (in millions) recovered from total reads for each group. ( $n = 16$ , total number of reads after quality trimming = 35 million).

all root samples (Table S2), which also contained an average of 9% unknown sequences, and 14% unqualified and unassembled reads.

#### Identification of fungal taxa within individual root clusters

To identify active fungi in EM root clusters, we recovered total DNA and RNA from the same extraction, and then compared fungi detected from stable DNA with those detected from more labile RNA, using the logic that the rRNA of active fungi would be overrepresented (Baldrian *et al.*, 2012; Liu *et al.*, 2012). Using total DNA with 454 amplicon sequencing of the ribosomal ITS region, we detected from 27 to 52 fungal operational taxonomic units (OTUs) per root cluster (Fig. 3A; Table S3A). In contrast, fungal transcripts coding for rRNA (LSU D2 region) from individual root clusters were dominated by only one to several fungal OTUs belonging to *Piloderma*, *Amphinema*, *Russulaceae* and *Cenococcum*, *Telephoraceae* and *Mycena* sp. (Fig. 3; Table S3). With exception of *Mycena*, all are known to be EM taxa. Of these, the most transcriptionally active EM fungi were *Piloderma* spp. that were detected in 7 of 16 root samples, comprising up to 93% of total LSU D2 reads from a single root cluster. Phylogenetic analysis revealed the presence of three distinct *Piloderma* species with 93–98% similarity in their D2 region (Figs S2 and S3). Comparison with sequences from the UNITE database identified these OTUs as *Piloderma byssinum*, *P. croceum* and a third unidentified *Piloderma* species (Fig. S3). In addition, two lineages of *Amphinema* and three lineages of *Russulaceae* were also detected with 85–98% sequence identity within the D2 region of each OTU respectively.

#### Mapping *Piloderma*-specific functional reads in soil and root clusters

To investigate *Piloderma* transcriptional activity in soil and root samples, we used genomic and transcriptomic data for *P. croceum* (Pilcr1 F 1598 v 1.0; Grigoriev *et al.*, 2012) to identify functional genes. Aligning all RNA-Seq reads against the *P. croceum* F 1598 transcript dataset (v1.0), we obtained an average of 100 000 and 1 000 000 reads in individual soil and root samples respectively. The most highly expressed transcripts could be readily assigned to Gene Ontology (GO) categories (Fig. 4). Most of the identified transcripts are related to gene functions in protein translation, protein degradation, protein folding, cell signalling, mitochondrial activity, terpene synthesis or to unknown proteins.

Functional category analysis reveals that *Piloderma* genes for nutrient transportation were predominantly expressed in soil samples, while genes for terpene biosynthesis were significantly expressed in root samples

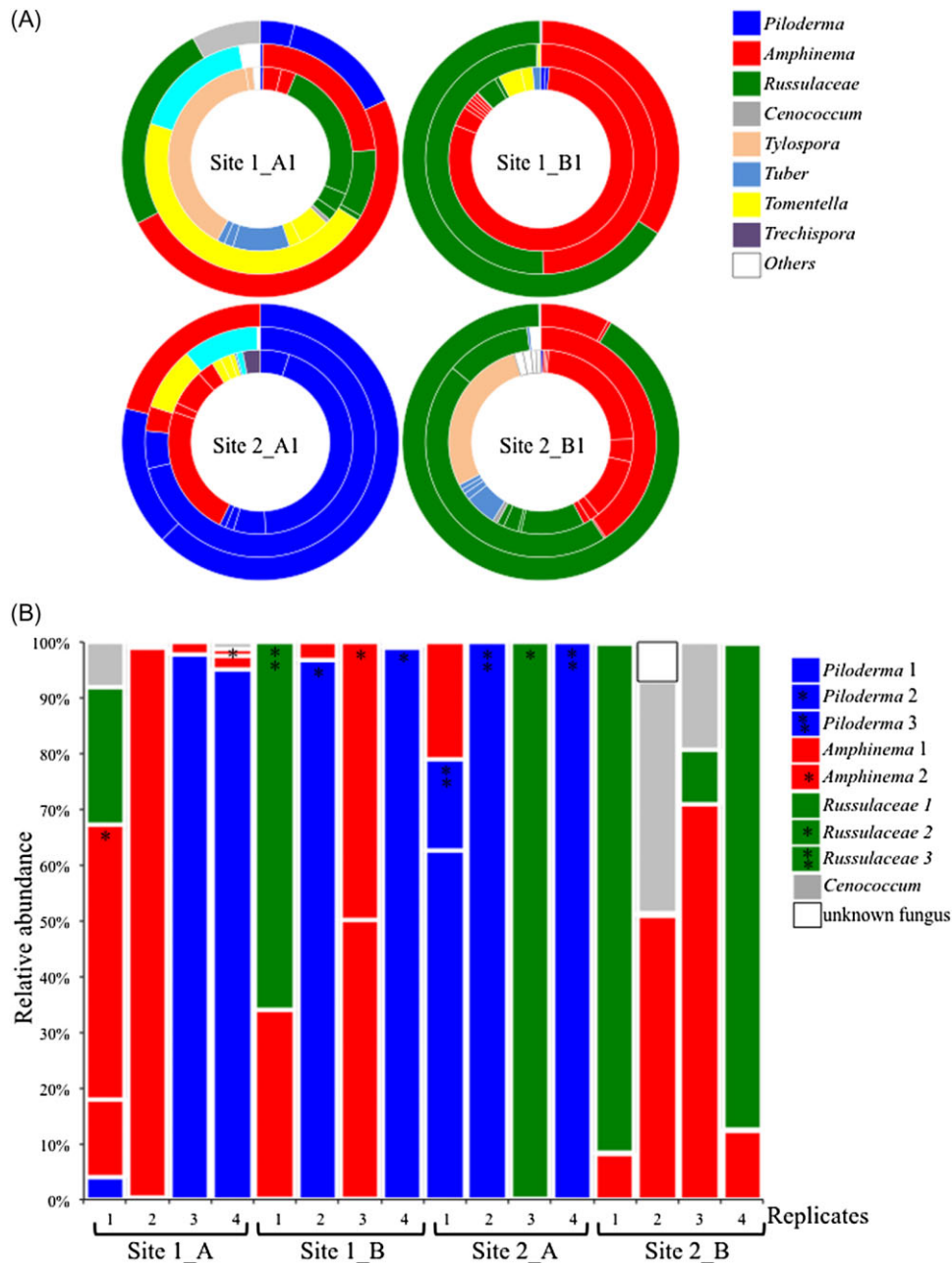
(Fig. 4). Sets of highly expressed genes (with percentage of reads >0.02%) were identified using BLASTX (Fig. 5; Table S4). Of 141 highly expressed functional-related gene groups, 124 gene groups were commonly expressed in both root and soil samples and are assumed to be directly associated with response of *Piloderma* hyphae to its environment (Fig. 5A). Only 17 unique gene groups were determined by comparison with false discovery rate < 0.1 (Fig. 5B). Eight *Piloderma* gene groups were found to be upregulated in soil compared with the EM root and nine gene groups displayed significantly higher expression within the EM root compared with soil. Figure 5B shows that similar patterns of differential gene expression in root versus soil *Piloderma* transcripts were observed across all sample locations.

#### Genomic and transcriptomic identification of *Piloderma* small secreted proteins

The small proteins secreted by plant pathogenic fungi or plant symbiotic fungi have been implicated in playing an important role in host compatibility (Chisholm *et al.*, 2006; Plett *et al.*, 2011). Four highly expressed genes encoding small secreted proteins were detected from *Piloderma*, and investigated further (Figs 5 and S4). These small secreted proteins included PiCr, PiBas, PiSs1 and PiEsp, which had sequences homologous to the cysteine-rich protein (gi392569350), fungal biotrophy-associated secreted protein (gi471559611), small secreted protein (gi328860126) and EM-associated small secreted protein (gi170100116). These small-secreted proteins have distinct primary and tertiary protein structures (Figs S4 and S5) indicating their functional independence. Like many other effectors associated with fungal virulence (Saunders *et al.*, 2012), these four small secreted proteins share at least one of the following properties: a signal polypeptide, small protein size and cysteine rich, implying their involvement in *Piloderma*–host symbiosis.

All the small secreted proteins encode a small polypeptide of less than 215 amino acids and were predicted as non-cytoplasmic proteins with no transmembrane domains (data not shown). PiCr, PiBas and PiSs1 contain a signal peptide at their N-terminals. A distinct sequence pattern was found in PiCr, which contains six cysteine residues arranged in an inverted symmetrical repeat. No specific domains and motifs were predicted from those small secreted proteins (data not shown). All four small secreted proteins show high sequence identities (over 99%) with representative contigs in *P. croceum* genome.

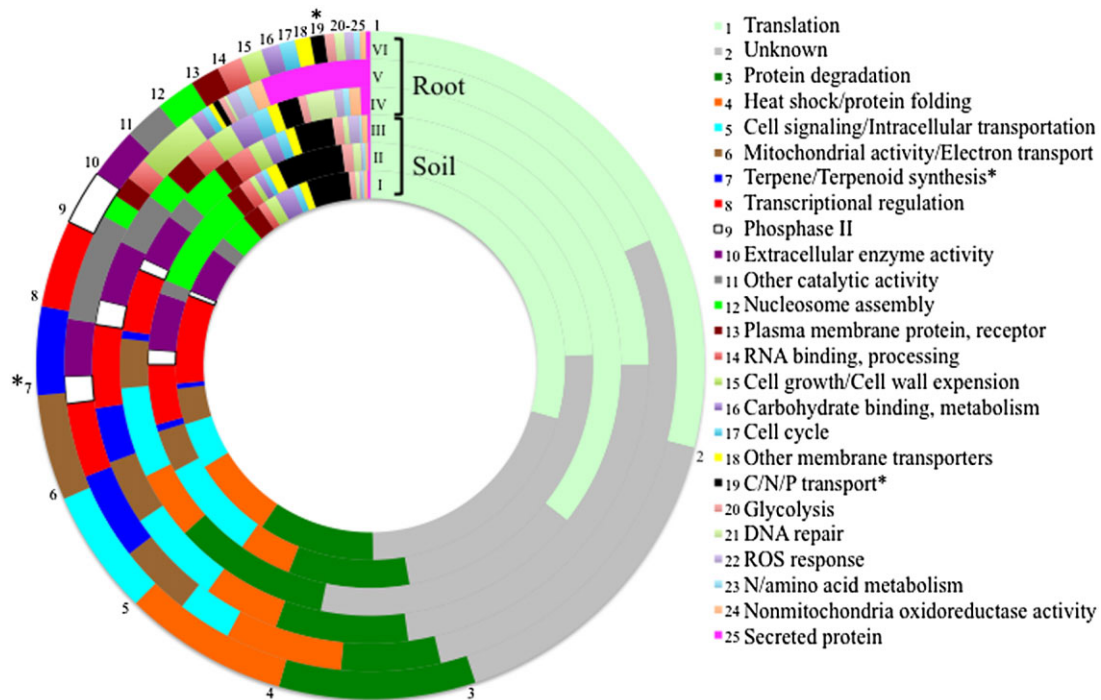
Sequence analysis of the *P. croceum* genome contigs indicated that the several highly expressed genes encoding functionally associated proteins are clustered with



**Fig. 3.** Taxonomic composition of EM root clusters based on: (A) ITS DNA amplicon sequences and ribosomal RNA sequences (ITS RNA and LSU D2 region). The individual circles indicate the %OTU of ITS DNA sequences (innermost), ITS RNA reads (middle) and the normalized abundance of D2 LSU reads (outermost) obtained from the same extraction of root clusters; (B) ribosomal RNA sequences (LSU D2 region). Asterisks ('\*' and '\*\*') indicate different lineages present within each taxa. The replicates indicate the reads recovered from different root clusters collected from the same soil cores. The percentages displayed pertain to relative values of paired reads. The relative values below 0.5 appear as zero. The absolute counts of reads are shown in Table S2. The % OTU of ITS DNA sequences and normalized abundance of D2 LSU reads: *Piloderma* in dark blue; *Amphinema* in red; *Russulaceae* in green; *Cenococcum* in grey; *Tylospora* in orange; *Tuber* in light blue; *Tomentella* in yellow; *Trechispora* in purple; and other fungi or unknown soil fungus in white.

either *PiCr*, *PiBas* and *PiSs1* respectively. Most genes belonging to the same individual clusters also shared similar RNA transcript levels (% reads >0.01) (Fig. 6). Based on time course analyses for individual root samples, it was determined that the genes in the clusters

were induced along with the small, secreted proteins (data not shown). Transcriptomic analysis shows that genes in the clusters of *PiCr*, *PiBas* and *PiSs1* are most likely involved in membrane trafficking. A set of genes encodes extracellular activity including secreted



**Fig. 4.** Functional categories of *Piloderma* genes from root clusters and comparison with *Piloderma* genes from soil samples. Nested circles from innermost to outermost represent transcript abundance (% reads) per sample ( $n = 3$ ): (I) Site1\_A\_soil; (II) Site1\_B\_soil; (III) Site2\_A\_soil; (IV) Site1\_A\_root; (V) Site1\_B\_root; (VI) Site2\_A\_root. The genes with % reads > 0.01 from the total *Piloderma* reads were assigned to the categories. The % reads for a gene group = number of reads mapped to a *P. croceum* gene group  $\times$  100/number of reads mapped to *P. croceum* database. Asterisks highlight on outermost circles indicate gene categories statistically different in root samples from soil samples.

phospholipase C, laccase and phospholipase-dependent transferases were also found located adjacent to the *PiBas* gene.

#### Identification of *P. taeda* functional genes in mycorrhizal root clusters

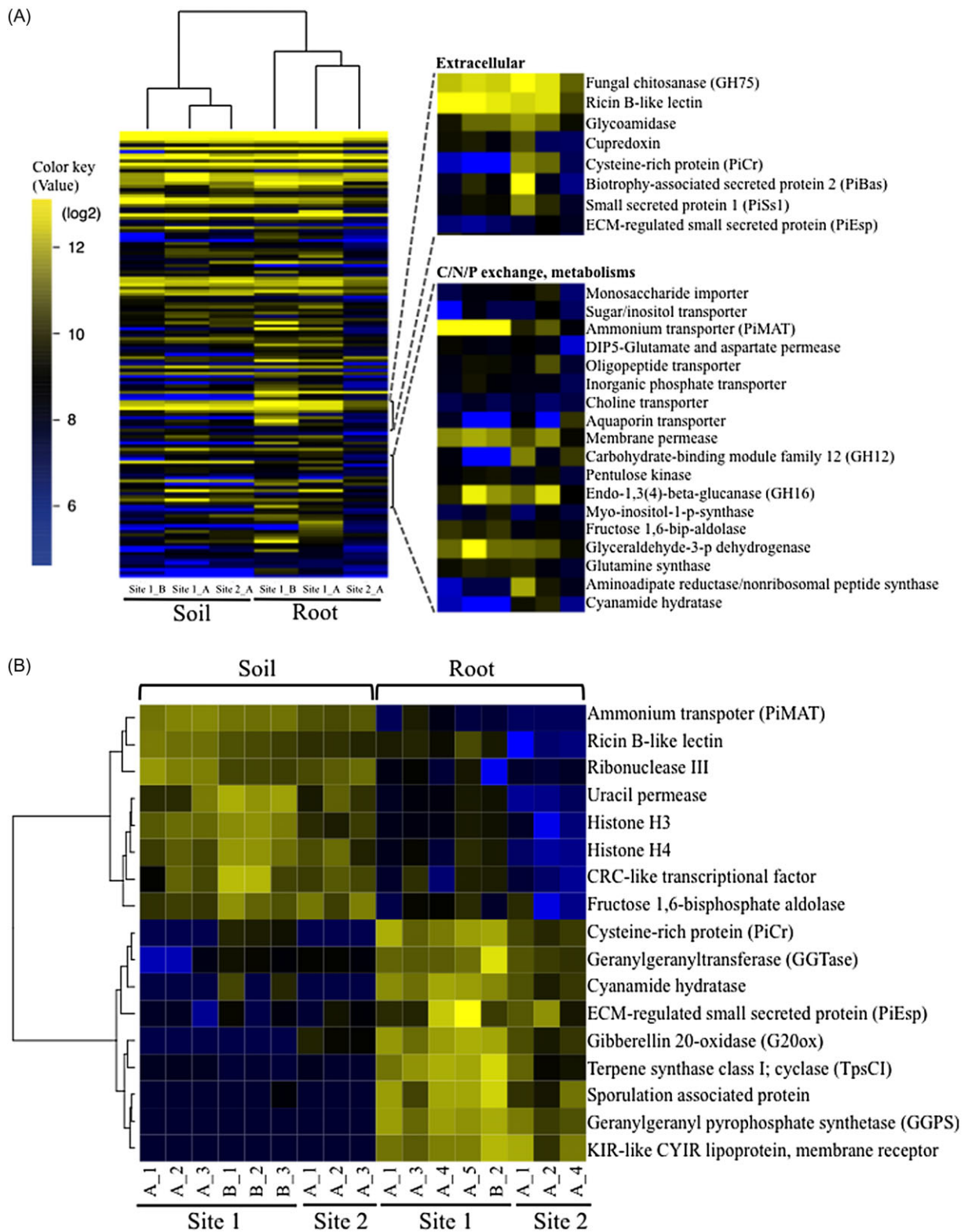
In addition to fungal genes, metatranscriptomic analysis of root clusters recovered over 40% of the reads representing functional plant genes from *P. taeda* that allowed us to study the function of both organisms from the same sample (Fig. 2). A total of 212 500 functional contigs from *P. taeda* were identified from EM root clusters. Of those, 2600 highly expressed contigs representing over 50 gene groups were shared among different sampled EM roots (Table S5).

## Discussion

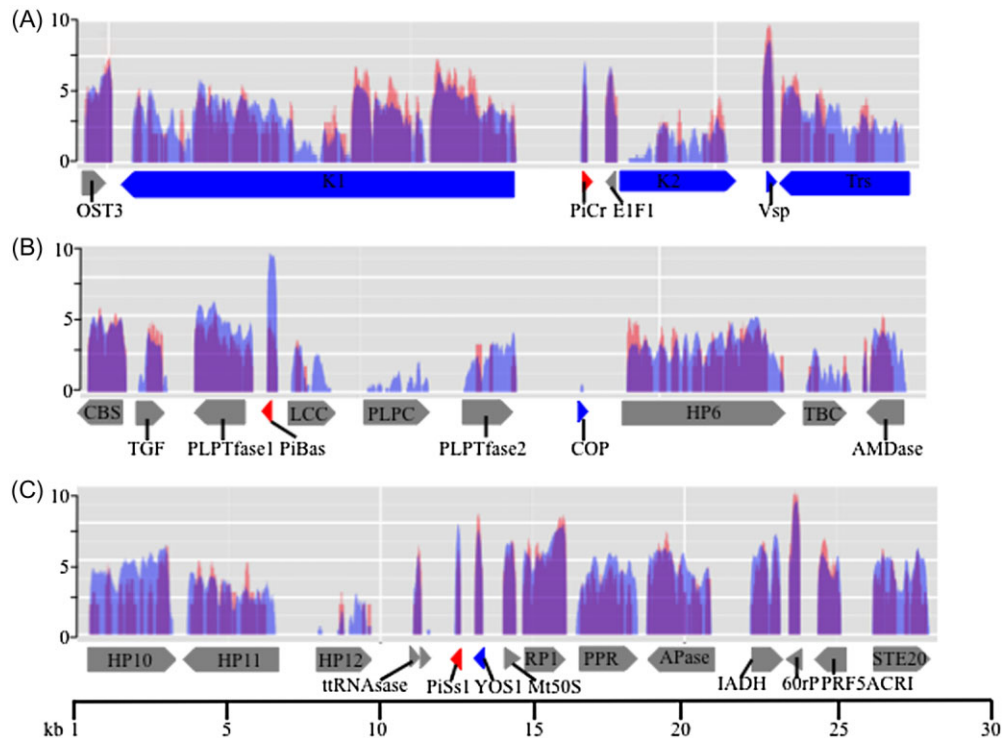
### Simultaneous identification of active fungal and plant transcriptomes using next-generation sequencing

To study EM root function *in situ*, we applied improved methodologies for obtaining high-quality DNA/RNA (from

the same extraction), poly-A enrichment, cDNA construction, data assembly and comparative metatranscriptomics at the single-species level. A workflow for bioinformatic analysis using advanced and popular software packages allowed us to sort and assemble the RNA-seq data (Fig. 1). The cDNA construction methods using poly-T primers were appropriate for isolating eukaryotic mRNAs and cDNA libraries, such that mRNA accounted for more than 80% of the total number of reads. This methodology still yields up to 1% rRNA, sufficient to taxonomically EM fungal species and other microbes that are present. Our method sorted and extracted the reads represent D2 regions from environmental samples for fungal identification, while at the same time, quantifying their abundance at the transcriptomic level (Fig. 1; Fig. S6). Based on LSU D2 RNA analysis, we found that most root clusters are dominated by a single fungal species (Fig. 3); in contrast, a much higher diversity of fungal taxa (>5) was observed from the same root samples using 454-sequencing of ITS amplicons (Fig. 3A), suggesting that only one or very few species of EM fungi are actively functioning within a given root cluster at the same time. A higher diversity of dominant fungal taxa (>5) was observed from soil samples (data not shown); however, RNA-seq again revealed that



**Fig. 5.** Cross-comparative expression of (A) highly expressed and (B) unique *Piloderma* gene groups in soil and root samples. The gene groups highly expressed at least in an individual sample (% reads >0.02 from all *Piloderma* reads were selected for clustering and heat map visualization). The color key represents RPKM normalized log<sub>2</sub> transformed counts of the genes. Wilcoxon signed-rank test (Bauer, 1972) was used to filter the data. Each gene family was significantly overexpressed in one of the group as determined by comparisons with false discovery rate (FDR) <0.1 using Benjamini-Hochberg test.



**Fig. 6.** The relative expressions of the cluster genes for secreted proteins: (A) PiCr1, (B) PiBas and (C) PiSs1. A pool of randomly selected samples from three sites ( $n = 3$ ) were used for mapping. Genes are depicted as arrows: red, secreted proteins; blue, proteins for the secretory pathway; grey, other proteins. The area chart quantifies the transcript abundances of *Piloderma* genes in root (blue) and soil (orange), and the overlap area shown in purple. The y-axis represents the  $\log_2$  for % reads corresponding to the total reads mapped to *Piloderma* database. OST3, oligosaccharyl transferase subunit 3; K1 and K2, kinesin motor protein; PiCr, cysteine-rich protein; E1F1, E1F1 transcription elongation factor; Vsp, Vsp55-domain trafficking-related protein vacuolar protein; Trs, TRS120 cis-Golgi transport protein; CBS, CBS domain containing protein; TGF, transforming growth factor-receptor; PLPTfase1 and 2, PLP-dependent transferase; PiBas, fungal biotrophy-associated secreted protein; LCC, laccase; PLPC, phospholipase C; COP, coatomer protein; HP, hypothetical protein; TBC, GTPase-activating TBC domain protein; AMDase, general amidase; tRNAase, tryptophanyl-tRNA synthetase; PiSs1, small-secreted protein; YOS1, ER-to-Golgi transport membrane protein; Mt50S, mitochondrial 50S ribosomal protein; RP1, Ribohporin I; PPR, pentatricopeptide repeat-containing protein; APase, acid protease; IADH, iron-containing alcohol dehydrogenase; 60rP, 60S ribosomal protein L35; PRF5ACRI, Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase; STE20, protein kinase ste 20.

these transcripts were also dominated by one or only a few fungal taxa.

Analysis of the EM metatranscriptome using poly-A selection following by gene mapping and de-novo assembly identified functional ‘symbiosis’ genes from both EM fungus and its plant host *P. taeda*. *Piloderma* spp. are consistently the dominant EM fungal taxon in root and soil samples from over 50 pine forests across 10 states in North America (Talbot *et al.*, 2014). By using the publically available *Piloderma* genome database, we focused on how genomic machinery operates in dominant taxa (*Piloderma*) in the investigated natural environment. The representative genes of interests and their functions are discussed in the following sections.

Previous studies on the *Piloderma*–*Quercus* EM symbioses identified a set of premycorrhiza-related host genes (*Quercus rober*) (Krüger *et al.*, 2004; Herrmann and Buscot, 2007). Because HiSeq provides much larger numbers of RNA reads (up 30 million reads per lane),

efficient mapping of transcript reads (Fig. 1) enables simultaneous analysis of metagenome expression even when multiple fungal and plant species are present.

#### *Unique metabolic genes are highly expressed in symbiotic tissues*

EM fungi develop communication networks using their hyphae as functional units. At the cellular level, EM hyphae that form the Hartig net establish the physical interaction with hosts; in contrast, hyphae that extend outward into the soil mostly contribute to resource uptake or interaction with biotic factors in the environment (Nehls, 2007). In addition to root clusters, we also collected soils from the same soil cores where the root samples were collected and generated the metatranscriptomic data to study microbial function in the soil samples. Regardless of which collecting site we sampled, comparative functional profiles for several gene groups differed significantly



between root and soil samples (Fig. 5), reflecting a high degree of functional variability between *Piloderma* growing in different parts of the rhizosphere.

Nine transcripts were significantly upregulated in EM root by comparison to mycelium growing in soil (Fig. 5B). These include two effectors (*PiCr* and *PiEsp*) mentioned in the result sections, as well as seven unique genes that include three gene groups for terpene synthesis (cyclase/TspCl, GGPTase, GGPS), and a gene for gibberellin 20 oxidase (G20ox) in symbiotic root indicated *Piloderma*, as other pathogenic (MacMillan, 1997) and EM fungi (Strzelczyk and Pokojaska-Burdziej, 1984), could produce gibberellin precursors and activate gibberellin to promote plant growth. Interestingly, fungal genes encoding cyclase, GGPS and G20ox are known to be functional clusters associated with gibberellin biosynthesis (Kawaide, 2006). As in other Agaricomycetes (Abraham, 2001), *Piloderma* may synthesize cytotoxic sesquiterpenoid metabolites to enhance antimicrobial activity. More studies are needed to determine the role of terpene-associated genes in the *Piloderma*–pine symbiosis.

Three other highly expressed ectomycorrhiza-associated transcripts were detected including genes encoding proteins with potential roles in detoxification of phytotoxin: cyanamide hydratase (Maier-Greiner *et al.*, 1991); fungal reproduction: cell growth-associated protein; and lipoprotein receptor: KIR-like CYIR protein.

#### *Highly expressed genes for carbon/nitrogen/phosphate (C/N/P) cycling and other extracellular activity*

As with other EM fungi, genes coding for C/N/P uptake, transportation and utilization must play important roles in *Piloderma* mutualistic function in the environment and communication with its pine host. Our study using comparative metatranscriptomics illustrates different patterns of gene expression of *Piloderma* growing in different parts of the rhizosphere (EM roots versus soil). Nine of 18 gene groups we detected for C/N/P-associated proteins are transporters, including one gene for ammonium (*PiAMT*), one for choline (ammonium salts), two for amino acids, two for monosaccharides, one for inorganic phosphate, one water channel and one general membrane permease, indicating the great activation of fluxes for nutrient exchange (Fig. 5A). One type of high-affinity ammonium transporter (*PiAMT*; amino acid sequence in Fig. S7) that shares high amino acid identity (61%) with *Aspergillus* MepA (Monahan *et al.*, 2006) (data not shown) was strikingly enriched in soil compared with root samples, suggesting that ammonium may be a preferential source of nitrogen in soil. Similarly, a high level of gene expression for glutamine synthase and glutamate-related transporters in both free mycelium and EM roots implies that glutamine pathway is the main biogeochemical

pathway for ammonium assimilation and nitrogen utilization in *Piloderma*. Glutamine could be assimilated using ammonium, or directly imported from soil and serve as the major molecule translocated to the roots.

Current studies showed that at least 44 of 132 glycoside hydrolases (GHs) contribute to plant cell wall deconstruction (Gilbert, 2010; <http://www.cazy.org/>). Of 141 highly expressed gene families (Fig. 5A), we detected only two GHs (GH12 and GH16); no other enzymes involved in degradation of cellulose, hemicellulose and pectin were identified by our metatranscriptome studies, suggesting adaptation of *Piloderma* to the EM symbiotic lifestyle. As with other EM fungi, we infer that *Piloderma* mainly absorbs monosaccharides from the host and environment via its monosaccharide/sugar transporter families.

Several other extracellular enzymes were highly expressed in both root and soil samples. These include fungal chitinase (GH75), which is involved in modification of fungal cell wall structure, and may be required for fungal development or to make cell more accessible for colonization. Ricin B-like lectin could activate fungal signal transduction through binding and clustering specific carbohydrates on fungal cell surfaces. Cupredoxin is involved in electron transfer. High expression of the gene for glycoamidase indicates active regulation of glycoprotein maturation (Suzuki *et al.*, 1994). Overall, these common extracellular enzymes are most likely involved in basic biological processes of fungal development and adaptation to environment.

#### *Fungal small secreted proteins present in EM pine roots*

Secreted effector-like proteins are involved in a variety of pathogenic and symbiotic fungal–host interactions (Martin *et al.*, 2008; Lowe and Howlett, 2012). However, the mechanisms for delivery and function of the effectors are less known. In general, effectors are often highly divergent in plant-pathogenic fungi and lack similarity in their sequences to proteins in most databases. Our study identified four highly expressed effector-like genes (*PiCr*, *PiBas*, *PiSs1* and *PiEsp*) that share high sequence similarity with other effector-like proteins. This implies that those proteins are likely to be involved in communication between *Piloderma* and pine roots. However, more studies on functional analysis are needed to determine the role of these small secreted proteins in *Piloderma* symbiosis.

Four sets of *Piloderma* genes for secretory pathway are also clustered in the genome with these four effector-like genes involved in the EM symbiotic interface (Fig. 6). The six cysteines of *PiCr* protein are arranged in a symmetrical structure, likely form three disulphide bridges. The disulphide bridges could contribute to protein stability

during secretion. There is no direct evidence to link such a structure with their function on symbiosis, though many small secreted proteins of fungi that contain four to eight cysteine residues have found function inside the plant cells to induce plant defence system (Rep, 2005; Stergiopoulos and de Wit, 2009). Gene expression analysis by mapping the transcriptomic reads to genome sequence of the *PiCr* cluster revealed a set of five genes including genes for cysteine-rich protein (PiCr), two motor proteins (K1 and K2), ER-to-Golgi transport (Vsp) and cis-Golgi transport (Trs) are induced exclusively in symbiotic tissues. One motor protein (K1) contains the multi-functional domains for kinesin heavy chain (K1Sc), transmembrane channel (MscS), protein kinase (HrpJ), RNA-binding (Pumilio), vesicle coat trafficking (Sec1) and a coatomer element (COP II) using conserved domain prediction server in NCBI (data not shown). Those results indicate the mediation of organelle-to-organelle secretory pathway is involved in PiCr secretion and fungal–host recognition. PiBas encodes a protein with high homology to the fungal biotrophy-associated secreted protein (Bas), a small protein secreted into plant cells during compatible mycorrhizal interactions (Mosquera *et al.*, 2009). Genomic features related to the biotrophic lifestyle in PiBas clustered genes include plant cell wall and membrane manipulation [e.g. laccase, phospholipase C and pyridoxal phosphate (PLP)-dependent transferases]. Expression analysis showed that most of the genes selected contained in the PiBas cluster were induced in root tissues and soil samples, suggesting that PiBas clusters could be involved in manipulating host cell structures and cell processes during EM symbiosis; the same gene cluster could play a role during growth in soil.

PiSs1 and PiEsp are homologous to effector-type small secreted proteins with unknown function. PiSs1 cluster encode proteins for transcription, translation, protein translocation and ER-to-Golgi transport (YOS1), indicating that a specific synthetic and delivery pathway is involved in function of PiSs1, whereas no functional-documented genes were detected adjacent to PiEsp (data not shown).

In this study, we developed an effective approach for studying EM fungal metatranscriptomes to identify specific molecular mechanisms operating to regulate fungal–plant interactions and fungal fitness. We were able to identify *Piloderma* spp. as the main functional EM fungal player in pine forest environments. Availability of publically available genome resources for *P. croceum* also made it also possible for us to identify key sets of genes which were expressed within ectomycorrhizae and soil. Unfortunately, genome data were not yet available at the present time for several other fungi identified by our study (*Russula*, *Amphinema*, *Thelephora*). Given the increasing availability of fungal genome assemblies and annotations,

exploitation of metatranscriptomic analysis shows great promise for profiling lineage-specific transcriptomics *in situ* in other systems and contribute for a better understanding of the functional roles fungi play in nature.

## Experimental procedures

### Sampling

Organic soil horizons were collected from two *P. taeda* plots located in the Duke Forest, Durhamco, NC, USA in April 2012 using soil cores (10 cm diameter; 15 cm height) (Fig. S1). At the time of sampling, the average temperature in these forests was 21°C. Root and soil samples were stored on ice and frozen in liquid nitrogen or dry ice within an hour after collection, and stored at –80°C prior to RNA extraction.

### RNA preparation and cDNA construction

Total RNA was isolated from root samples using a CTAB/chloroform extraction and LiCl precipitation method modified according to Chang and colleagues (1993). Frozen roots were ground using ten 2 mm zirconia beads (BSP, Bartlesville, OK, USA) in a GenoGrinder 2000 (OPS Diagnostics, Lebanon, NJ, USA) at 1500 strokes min<sup>-1</sup> for 30 s. Immediately after grinding, 1 ml DEPC/CTAB solution (2% CTAB and PVP; 100 mM Tris-HCl, pH 8.8; 25 mM EDTA, pH 8.0; 2 M NaCl; 2% fresh beta-mercaptoethanol and 1% DEPC, v/v) was added to the ground roots. The sample was then homogenized by vortexing and incubated at 65°C for 10 min. Equal volumes of a chloroform : isoamyl alcohol mixture (24:1) were added to the samples, mixed by vortex and centrifuged at 9000 g for 10 min. The supernatant was recovered and RNA in the supernatant precipitated with 2:1 volume 8 M LiCl at 4°C for 16 h. After centrifugation at 9000 g for 5 min, the supernatant was collected for DNA extraction and the pellet (RNA) was washed using 70% EtOH and eluted in DNase/RNase-free water. The supernatant collected in LiCl precipitation step was treated with 1 unit RNase at 37°C for 30 min, then precipitated with equal volumes of ice cold isopropanol. After centrifugation at 9000 g for 10 min, the pellet (DNA) was washed using 70% EtOH and eluted in DNase/RNase-free water.

RNA from soil was extracted using the RNA PowerSoil kit (MoBio, Carlsbad, CA, USA) as indicated in the manufacturer's instructions. Total RNA quality was quantified using a 2100 bioanalyser (Agilent, Santa Clara, CA, USA) and Qubit 2.0 fluorometer (Invitrogen, Grand Island, NY, USA) following the supplier's protocols. Only RNA with high integrity (RIN > 8.0) and showing well-defined peaks for 18S and 28S rRNAs on an electropherogram (RNA 6000 Nano LabChip, Agilent) was used for cDNA library construction. The 0.5 µg RNA was used for cDNA library construction using TruSeq RNA Sample Prep v2 (Illumina, San Diego, CA, USA) and poly-T primers to enrich poly-A containing mRNA. The cDNA samples were sequenced using HiSeq 2000 (Illumina) in Institute for Genome Sciences and Policy at Duke University. Twelve samples were sequenced using a lane of Illumina run to generate 38 Gb of data. The raw reads were submitted to Sequence Read Archive (<http://trace.ncbi.nlm.nih.gov/Traces/sra/>) (Accession No. SRP041643).

### Packages and databases used for bioinformatics analysis

The bioinformatic packages used for assembling and annotating the datasets were: the Galaxy web-based package (<https://main.g2.bx.psu.edu>) to check the quality of reads, FastQC for sequence trimming and quality filtering (cut-off with quality scores <28; Blankenberg *et al.*, 2010), Bowtie for reference-based mapping (<http://bowtie-bio.sourceforge.net/index.shtml>), Velvet (Zerbino and Birney, 2008) and Trinity (Grabherr *et al.*, 2011) for *de novo* assembly and a combination of Blastx (Tao, 2010), GO (Ashburner *et al.*, 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2012) and NCBI eukaryotic orthologous groups (KOG) (Tatusov *et al.*, 2003) for gene annotation. Amino acids predicted from nucleotide sequences was performed using ExPASy (<http://web.expasy.org/translate/>). EMBL-EBI and Signal-3L were used for signal-peptide prediction (<http://www.ebi.ac.uk/Tools/pfa/phobius/>; <http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/>). NCBI (Marchler-Bauer *et al.*, 2011) and Myhits were applied for production of conserved domains and motifs (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). In general, default parameters were applied for the bioinformatics analysis.

Databases used for mapping including: Fungal rRNA (~ 400 000 contigs; NCBI); Bacterial 16S (~ 2 1000 000 contigs; Ribosomal Database Project, <http://rdp.cme.msu.edu>); *P. taeda* (~ 320 000 contigs; NCBI), *P. croceum* (~ 340 000 contigs; Pilcr1 F 1598 v 1.0, Grigoriev *et al.*, 2012). The annotated sequences of coding genes downloaded from NCBI for over 30 fungal species (mostly EM fungi of trees or wood rot fungi) that were assumed to be best hits for the sequences of *Pinus*-associated EM fungi were pooled and used as references for fungal gene annotation using Blastx (Tao, 2010), including *Serpula lacrymans* (12 925 genes), *Coniophora puteana* (13 898 genes), *Laccaria bicolor* (18 315 genes), *Dichomitus squalens* (12 511 genes), *Gloeophyllum trabeum* (11 889 genes), *Stereum hirsutum* (14 455 genes) and others. The annotated sequences of coding genes downloaded from NCBI for 15 plant species were pooled and used as references for *P. taeda* gene annotation, including *Populus trichocarpa* (62 211 genes), *Glycine max* (62 511 genes), *Vitis vinifera* (56 152 genes), *Zea mays* (33 490 genes), *Ricinus communis* (32 094 genes) and others. The databases were quality filtered using FASTA manipulation of Galaxy web-based package.

### Statistical analysis

In total, nine root and nine soil samples that contain RNA-Seq reads of *Piloderma* taxa were analysed for Figs 4 and 5. In Fig. 4, we tested for statistical differences between factors using a Tukey test. Differences were considered significant at  $P < 0.05$ . The data of each sample were the mean for the biological replicates. The functional categories (Fig. 4) and heat map (Fig. 5) were generated using R package(s) (R Development Core Team, 2011).

### Acknowledgements

We would like to thank Igor Grigoriev, Francis Martin and François Buscot for access to genome data for *Piloderma*

*croceum*. Genome sequence data were produced by the US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user community. We also thank members of the Vilgalys, Peay and Bruns lab for assistance with soil/root sampling and Mr. Chih-Ming Hsu for help with bioinformatics data. This work is supported by NSF grant DBI-10-46052 to RV, DBI-10-45658 to KGP and DBI 10-46115 to TDB ([http://www.stanford.edu/~kpeay/DOB\\_Home.html](http://www.stanford.edu/~kpeay/DOB_Home.html)).

### References

- Abraham, W.R. (2001) Bioactive sesquiterpenes produced by fungi: are they useful for humans as well? *Curr Med Chem* **8**: 583–606.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.
- Baldrian, P., Kolařík, M., Stursová, M., Kopecký, J., Valášková, V., Větrovský, T., *et al.* (2012) Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J* **6**: 248–258.
- Bauer, D.F. (1972) Constructing confidence sets using rank statistics. *J Am Stat Assoc* **67**: 687–690.
- Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., and Galaxy Team (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**: 1783–1785.
- Bonfante, P., and Genre, A. (2010) Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat Commun* **1**: 48.
- Canales, J., Avila, C., Pacheco-Villalobos, D., Díaz-Moreno, S., Molina-Rueda, J.J., Ariza, D., *et al.* (2011) Gene expression profiling in the stem of young maritime pine trees: detection of ammonium stress-responsive genes in the apex. *Trees Struct Funct* **26**: 609–619.
- Chang, S., Puryear, J., and Cairney, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* **11**: 113–116.
- Chisholm, S.T., Coaker, G., Day, B., and Staskawicz, B. (2006) Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**: 803–814.
- Courty, P.E., Pritsch, K., Schloter, M., Hartmann, A., and Garbaye, J. (2005) Activity profiling of ectomycorrhiza communities in two forest soils using multiple enzymatic tests. *New Phytol* **167**: 309–319.
- Cullings, K., and Makhija, S. (2001) Ectomycorrhizal fungal associates of *Pinus contorta* in soils associated with a hot spring in Norris Geyser Basin, Yellowstone National Park, Wyoming. *Appl Environ Microbiol* **67**: 5538–5543.
- Cullings, K.W., New, M.H., Makhija, S., and Parker, V.T. (2003) Effects of litter addition on ectomycorrhizal associates of a lodgepole pine (*Pinus contorta*) stand in Yellowstone National Park. *Appl Environ Microbiol* **69**: 3772–3776.
- Gilbert, H.J. (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol* **153**: 444–455.
- Gottel, N.R., Castro, H.F., Kerley, M., Yang, Z., Pelletier, D.A., Podar, M., *et al.* (2011) Distinct microbial communities

- within the endosphere and rhizosphere of *Populus deltoids* roots across contrasting soil types. *Appl Environ Microbiol* **77**: 5934–5944.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* **29**: 644–652.
- Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., et al. (2012) The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res* **40**: D26–D32.
- Gugerli, F., Brandl, R., Castagneryol, B., Franc, A., Jactel, H., Koelewijn, H.P., et al. (2013) Community genetics in the time of next-generation molecular technologies. *Mol Ecol* **22**: 3198–3207.
- Hasselquist, N., Germino, M.J., McGonigle, T., and Smith, W.K. (2005) Variability of *Cenococcum* colonization and its ecophysiological significance to young conifers. *New Phytol* **165**: 867–873.
- Herrmann, S., and Buscot, F. (2007) Cross talks at the morphogenetic, physiological and gene regulation levels between the mycobiont *Piloderma croceum* and oak microcuttings (*Quercus robur*) during formation of ectomycorrhizas. *Phytochemistry* **68**: 52–67.
- Joosen, R.V., Lammers, M., Balk, P.A., Brønnum, P., Konings, M.C., Perks, M., et al. (2006) Correlating gene expression to physiological parameters and environmental conditions during cold acclimation of *Pinus sylvestris*, identification of molecular markers using cDNA microarrays. *Tree Physiol* **26**: 1297–1313.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res* **40**: D109–D114.
- Kawaide, H. (2006) Biochemical and molecular analysis of gibberellin biosynthesis in fungi. *Biosci Biotechnol Biochem* **70**: 583–590.
- Krüger, A., PeSkan-Berghöfer, T., Frettinger, P., Herrmann, S., Buscot, F., and Oelmüller, R. (2004) Identification of premycorrhiza-related plant genes in the association between *Quercus robur* and *Piloderma croceum*. *New Phytol* **163**: 149–157.
- Liao, Z., Chen, M., Guo, L., Gong, Y., Tang, F., Sun, X., and Tan, K. (2004) Rapid isolation of high-quality total RNA from taxus and ginkgo. *Prep Biochem Biotechnol* **34**: 209–214.
- Liu, K.-L., Porrás-Alfaro, A., Kuske, C.R., Eichorst, S.A., and Xie, G. (2012) Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. *Appl Environ Microbiol* **78**: 1523–1533.
- Lorenz, W.W., Yu, Y.-S., Simúes, M., and Dean, J.F.D. (2009) Processing the loblolly pine PtGen2 cDNA microarray. *J Vis Exp* **25**: 1182.
- Lorenz, W.W., Yu, Y.-S., and Dean, J.F.D. (2010) An improved method of RNA isolation from loblolly pine (*P. taeda* L.) and other conifer species. *J Vis Exp* **36**: e1751.
- Lowe, R.G., and Howlett, B.J. (2012) Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. *PLoS Pathog* **8**: e1002515.
- MacMillan, J. (1997) Biosynthesis of the gibberellin plant hormones. *Nat Prod Rep* **14**: 221–243.
- Maier-Greiner, U.H., Klaus, C.B.A., Estermaier, L.M., and Hartmann, G.R. (1991) Herbicide resistance in transgenic plants through degradation of the phytotoxin to urea. *Angew Chem Int Ed (English)* **30**: 1314–1315.
- Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39**: 225–229.
- Martin, F., Duplessis, S., and Kohler, A. (2004) Exploring the transcriptome of the ectomycorrhizal symbiosis. In *Molecular Genetics and Breeding of Forest Trees*. Kumar, S., and Fladung, M. (eds). New York, USA: Haworth's Food Products Press., pp. 81–109.
- Martin, F., Aerts, A., Ahren, D., Brun, A., Danchin, E.G., Duchaussoy, F., et al. (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88–92.
- Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P.M., Jaillon, O., et al. (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033–1038.
- Moeller, H.V., Peay, K.G., and Fukami, T. (2014) Ectomycorrhizal fungal traits reflect environmental conditions along a coastal California edaphic gradient. *FEMS Microbiol Ecol* **87**: 797–806.
- Monahan, B.J., Askin, M.C., Hynes, M.J., and Davis, M.A. (2006) Differential expression of *Aspergillus nidulans* ammonium permease genes is regulated by GATA transcription factor AreA. *Eukaryot Cell* **5**: 226–237.
- Mosquera, G., Giraldo, M.C., Khang, C.H., Coughlan, S., and Valent, B. (2009) Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. *Plant Cell* **21**: 1273–1290.
- Nehls, U. (2007) Mastering ectomycorrhizal symbiosis: the impact of carbohydrates. *J Exp Bot* **59**: 1097–1108.
- Plett, J.M., Kemppainen, M., Kale, S.D., Kohler, A., Legué, V., Brun, A., et al. (2011) A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr Biol* **21**: 1197–1203.
- R Development Core Team (2011) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, [WWW document]. URL <http://www.R-project.org/>.
- Rep, M. (2005) Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiol Lett* **253**: 19–27.
- Saunders, D.G.O., Win, J., Cano, L.M., Szabo, L.J., Kamoun, S., and Raffaele, S. (2012) Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS ONE* **7**: e29847. doi:10.1371/journal.pone.0029847.
- Shakya, M., Gittel, N., Castro, H., Yang, Z.K., Gunter, L., Labbé, J., et al. (2013) A multifactor analysis of fungal and bacterial community structure in the root microbiome of mature *Populus deltoides* trees. *PLoS ONE* **8**: e76382. doi:10.1371/journal.pone.0076382.
- Smith, S.E., and Read, D.J. (2008) *Mycorrhizal Symbiosis*, 3rd edn. San Diego, CA, USA: Academic Press.
- Stergiopoulos, L., and de Wit, P.J.G.M. (2009) Fungal effector proteins. *Annu Rev Phytopathol* **47**: 233–263.

- Strzelczyk, E., and Pokojska-Burdziej, A. (1984) Production of auxins and gibberellin-like substances by mycorrhizal fungi, bacteria and actinomycetes isolated from soil and the mycorrhizosphere of pine (*Pinus sylvestris* L). *Plant Soil* **81**: 185–194.
- Sung, S.-J.S., White, L.M., Marx, D.H., and Orosina, W.J. (1995) Seasonal ectomycorrhizal fungal biomass development on loblolly pine (*Pinus taeda* L.) seedlings. *Mycorrhiza* **5**: 439–477.
- Suzuki, T., Kitajima, K., Inoue, S., and Inoue, Y. (1994) Occurrence and biological roles of proximal glycanases in animal cells. *Glycobiology* **4**: 777–789.
- Talbot, J.M., Bruns, T.D., Taylor, J., Smith, D.P., Branco, S., Glassman, S.I., *et al.* (2014) Endemism and functional convergence across the North American soil mycobiome. *PNAS* **111**: 6341–6346.
- Tao, T. (2010) Standalone BLAST Setup for Unix. In: BLAST® Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2008-. URL <http://www.ncbi.nlm.nih.gov/books/NBK52640/>.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**: 1–14.
- Van den Driessche, R. (1991) *Mineral Nutrition of Conifer Seedlings*. Boca Raton, FL, USA: CRC press.
- Wang, B., and Qiu, Y.L. (2006) Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**: 299–363.
- Wolfe, B.E., Rodgers, V.L., Stinson, K.A., and Pringle, A. (2008) The invasive plant *Alliaria petiolata* (garlic mustard) inhibits ectomycorrhizal fungi in its introduced range. *J Ecol* **96**: 777–783.
- Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821–829.
- Zhao, B.G., Futai, K., Sutherland, J.R., and Takeuchi, Y. (eds) (2008) *Pine Wilt Disease*. Berlin, Germany: Springer.

## Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Fig. S1.** Location of the collection sites.

**Fig. S2.** Partial large-subunit ribosomal RNA sequences (D2 region, ~ 180 bp) for the most dominant fungal taxa detected from mycorrhizal root clusters (Fig. 3).

**Fig. S3.** Phylogenetic placement of LSU RNA sequences for three *Piloderma* spp. based on ribosomal large-subunit D2 region. Sequences were aligned against taxonomically identified reference sequences of *Piloderma* spp. from GenBank with ClustalW; phylogenetic analyses performed using parsimony criterion in PAUP 3.0. Bootstrap tree ('fast

bootstrap' with 300 replicates) shows placement of three *Piloderma* spp.

**Fig. S4.** Nucleotide and peptide sequences of *Piloderma* effectors: (A) PiCr1, (B) PiBas, (C) PiSs1 and (D) PiEsp. The putative signal sequences are shown in the peptide sequence (Bold and underlined). In (A), the symmetrical arrangement of cysteine (C) residues is shown in boldface.

**Fig. S5.** Ribbon diagrams showed the predicted-tertiary structures for (A) PiCr, (B) PiBas, (C) PiSs1 and (D) PiEsp. The helix (pink) and sheet structures (yellow) were shown. N and C terminal are labelled. The protein tertiary structures were predicted using I-TASSER v 3.0 (Zhang 2008; Roy *et al.*, 2010; Roy *et al.*, 2012). C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is in the range from -5 to 2, where a C-score of higher value signifies a model with a high confidence.

**Fig. S6.** Partial images of visualization for reads of root samples Bowtie-mapped to a 28S rRNA reference (gi300394395, *Piloderma fallax*) using Integrative Genomics Viewer 2.2× (Thorvaldsdóttir *et al.* 2013). Fewer and variable sequences were found in variable region (D1 and D2) compared with the conserved regions of 28S. The red box indicates the region (~ 180 bp) where the reads were extracted for fungal taxa identification in this study. Mapped bars in blue indicate the inferred insert size on the reference gene is smaller than expected given the actual insert size; red is the inferred insert size on the reference gene is larger than expected given the actual insert size.

**Fig. S7.** Predicted amino acid sequence of PiAMT.

**Table S1.** Comparison of RNA extraction methods for pine roots (EM) and needles.

**Table S2.** Read and Percent of read associated with fungal, bacterial and pine transcripts in EM root clusters detected by Illumina HiSeq.

**Table S3.** Taxonomic composition of EM root clusters based on (A) ITS DNA amplicon sequences (B) transcribed ITS RNA sequences, and (C) ribosomal RNA (LSU D2 region). (A–C) The absolute counts (read counts) and relative value (% reads) are shown.

**Table S4.** Highly expression of *Piloderma* gene groups in soil and root samples.

The gene families with % reads >0.02 in at least a sample were selected. Table S4A showed the number of reads of individual gene families obtained from root and soil samples. Table S4B showed the list *Piloderma* protein ID from PiCr1 database for each gene group.

**Table S5.** Relative expression of *Pinus taeda* genes of mycorrhizal root clusters. Deseq 1.14.0 package (Anders and Huber, 2010) was applied to normalize the reads mapped to *Pinus taeda* EST database (Deseq > 100 cutoff). Simon Anders and Wolfgang Huber (2010): Differential expression analysis for sequence count data. *Genome Biology* **11**:R106.