Transcriptome analysis of tree peony during chilling requirement fulfillment: Assembling, annotation and markers discovering

Shupeng Gai, Yuxi Zhang, Ping Mu, Chunying Liu, Shao Liu, Lei Dong, Guosheng Zheng *

College of Life Sciences, Qingdao Agricultural University, Key Lab of Plant Biotechnology in Universities of Shandong Province, Changcheng Road 700, Qingdao, China

ARTICLE INFO

Article history:
Accepted 7 December 2011
Available online xxxx
Received by Richard Cordaux

Keywords:
1. Introduction

Paeonia suffruticosa Andrews
Transcriptome
Endo-dormancy
Chilling treatment
454 pyrosequencing
SSR

Abstract

Tree peony (Paeonia suffruticosa Andrews) is a well-known horticultural and medicinal plant. The flower buds must go through a period of endo-dormancy before bud sprouting in winter, but very little information concerned with dormancy release is available. We obtained 625,342 sequencing reads with massive parallel pyrosequencing on the Roche 454 GS FLX platform (mean length: 358.1 bp). De novo assemblies yielded 23,652 contigs and singletons. 15,284 contigs longer than 300 bp were further annotated, among them 12,345 ESTs showed significant similarity with sequences present in public databases (with an E-value < 1e − 10). 484 putative transcription factors were obtained. In addition, 2253 potential Simple Sequence Repeats (SSR) loci were identified in the 454-ESTs. Total 149 pairs of primers were designed, and 121 pairs were amplified successfully in initial screening. In addition, 73 pairs of primers displayed polymorphism. This sequence collection provides a significant resource for gene discovery during endo-dormancy of tree peony.

© 2011 Published by Elsevier B.V.

1. Introduction

Tree peony (Paeonia suffruticosa Andrews) belongs to the Moutan subfamily of the genus Paeonia, Paeoniaceae, being one of the earliest and most well-known horticultural and medicinal plants (mainly P. ostii) in the world. Sect. Moutan comprises eight wild species, P. cathayana, P. jishanensis, P. qui, P. ostii, P. rockii, P. decomposita, P. delavayi and P. ludlowii (Hong and Pan, 2005a, 2005b, 2007). P. suffruticosa refers to the commonly cultivated tree peonies. Five wild species, P. ostii, P. rockii, P. jishanensis, P. qui, and P. suffruticosa sps. Yinpengmudan, are known as the progenitors of tree peony cultivars (Zhou et al., 2003). Among all the wild species, P. ostii was proved to share more genetic similarity and be distributed as a sister clade among cultivars of the Zhongyuan group (Han et al., 2008; Zhang et al., 2009, 2011). As a wild species and its close relative with cultivars, P. ostii can be a preferred candidate to perform genomic sequencing and analysis.

Like many perennial woody plants, the flower buds of tree peony which complete differentiation in July or August each year must go through a period of bud dormancy before bud sprouting in winter (Wang and Zhang, 1987). The dormancy is a major obstacle to the house culture of tree peony in winter. Such dormancy belongs to endo-dormancy which is controlled by internal bud signals, and prevents untimely growth during seasonal transitions when environmental conditions often fluctuate between permissive or inhibitory to growth (Lang and Martin, 1987). Previous studies indicated that light and temperature play significant roles in the induction and breaking of endo-dormancy (Anderson et al., 2005; Li et al., 2003). And, Wang et al. showed that insufficient chilling is one of the key factors that prevent tree peony from flowering (Wang et al., 1996).

So far, although many work has been flourished in the literature about the physiological and molecular aspects of bud dormancy in kiwifruit, apple, peach and grapes and other plants (Arora et al., 2003; Brunel et al., 2002; Faust et al., 1997; Faye et al., 2003; Halaly et al., 2008; Ophir et al., 2009; Or et al., 2000, 2002; Pang et al., 2007; Saure, 1985), and vast quantities of genes associated with dormancy have been screened and identified. In grape, Or (2009) suggested a working hypothesis for metabolic processes involved in artificially induced dormancy release, including tricarboxylic acid cycle (TCA cycle), ATP synthesis, oxidative phosphorylation and oxidative response, et al. (Or, 2009), but more detailed understanding about these processes might be regulated was still limited (Ophir et al., 2009; Or, 2009). Therefore, characterization of complex network of biochemical and cellular processes responsible for the regulation...
and execution of bud dormancy release is very important and essential. On the other hand, there might be some difference in the dormancy release mechanism between the tree peony and the above fruit plants. For example, fruit crops seem to be sensitive to hydrogen cyanamide (HC) which can promote the breaking of dormancy, but tree peony seems to be insensitive to HC according to our unpublished work (Gai et al., unpublished), while Gibberellic acid (GA3) can effectively stimulate the breaking of dormancy. Comparing with other fruit plants, the knowledge about dormancy release in tree peony is relatively limited. Huang et al. reported 31 unique genes that might be associated with dormancy release during chilling treatments in a subtractive cDNA library, but most of them were related to structural function (Huang et al., 2008a, 2008b). Currently only 2204 pieces of EST are found in National Center for Biotechnology Information (NCBI) database, and most of them were recently submitted with a purpose for bud differentiation in tree peony (Shu et al., 2009).

To have a better understanding on the complex biochemical network responsible for the regulation and execution of the dormancy process, more study is needed to gain novel insight into the coordinated induction (or repression) of metabolic pathways that act together during dormancy release in tree peony buds. This can be achieved by initiating a survey on expressed genes using genomic approaches, such as 454 pyrosequencing (Ellegren, 2008; Margulies et al., 2005; Vera et al., 2008). The 454-EST analysis provides a way for the identification of novel genes and allows characterization of the transcriptome in various tissues and development stages (Luo et al., 2010). At the same time, ESTs can be used as a valuable source for marker development of short sequence repeats (SSRs), which are most commonly used for the construction of linkage maps of nuclear genomes. EST-SSRs have more advantages for marker development than SSR located in non-coding region because they represent coding regions of genome and have also been quickly developed for many species (Tenných et al., 2001). In this study, we aim to establish a normalized cDNA library prepared from flower buds during chilling treatment and obtain a substantial EST dataset using massive parallel pyrosequencing technology. These data will provide a useful mean of gene discovery, genome characterization and global understand of dormancy release as well as novel insight into the expressed genes during dormancy release.

2. Materials and methods

2.1. Plant materials

Total 45 individuals (Paeonia ostii ‘Feng Dan’) were potted and moved to refrigeration house with temperature 0–4 °C from 5 November to 30 December 2009 in Qingdao, Shandong, China, and 9 individuals of them were as a treatment. Two buds for each individual were collected after 0, 6, 12, 18 (endo-dormancy release), 24 days (eco-dormancy) chilling requirement fulfilling (0–4 °C, 8-h-light/16-h-dark cycle) during the period of bud dormancy release, and all buds were mixed. The dormancy stages of mixed buds were ascertained according to the method of Huang et al. (Huang et al., 2008a, 2008b). All materials were immediately frozen in liquid nitrogen and stored in −80 °C refrigerator until use.

2.2. RNA extraction and cDNA synthesis

Total RNA was extracted using TRizol Reagent (Invitrogen) according to manufacturer’s protocols. RNA preparations were quantified photometrically, checked on agarose gel and stored at −80 °C until use. Contaminated DNA in RNA sample was excluded using TURBO DNA-free™ kit (Ambion). The cDNA was synthesized according to BD SMART PCR cDNA Synthesis Kit (Clontech), but using a revised 3’ primer (5’- AACGAGTGTGATACGCAGAGTCGAC T(12) VN’-3’) with BsgI recognition sites. The synthesized double stranded (ds) cDNA was normalized by Duplex-specific nuclease (Evrogen, EA003), in order to prevent over-representation of the most common transcripts, and then was purified by PureLink™ PCR Purification Kit (Invitrogen). Poly A was digested by Bsg I (NEB). The digested cDNA was recovered by QIAquick Gel Extraction Kit (Qiagen).

2.3. Sequencing and assembling

Approximately 1 μg of double-stranded cDNA was used for sequencing on 454/Roche GS FLX sequencer, performed by Shanghai Bio Corporation (Shanghai, China) following manufacturer’s protocols. All analyses of the sequencing data were performed with the GS FLX Software v2.0.0.01 (454 Life Sciences, Roche). After using a series of normalization and data quality-filtering check, the 454 data were filtered for weak signals and low-quality sequences, and the read ends were screened and trimmed for 454 adaptor sequences to yield high quality (HQ) sequences (>99.5% accuracy on single base reads). A subsequent filtering step included the masking of SMART PCR primer sequences (Clontech) and the removal of sequences shorter than 50 bp before assembling. Finally, these HQ reads were assembled into unique putative transcripts (including contigs and singletons) using GS De Novo Assembler Software, which is an application of the GS FLX Software. The assembling was performed using the default parameters. Contigs and singletons containing ORF (open reading frame) was predicted by EMBOSS.

2.4. Annotation

Merged UniGenes were annotated based on sequence similarity using BLASTX against the NCBI non-redundant database. Only significant BLAST results (E-value<1e−10) were considered. Each of the merged UniGene was used as an annotation for the description of the target protein that was associated with the lowest E-value. Functional category was assigned to each gene product according to the gene ontology (GO) terms from the GO annotated Arabidopsis database of TAIR. Pathway assignment was performed according to Arabidopsis database classification in the KEGG database (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg).
EDTA, distilled water) and underwent vertical electrophoresis in a 6% denaturing polyacrylamide gel and silver staining.

3. Results and discussion

3.1. Sequencing and assembling

A normalized cDNA pool was constructed from combined flower buds of tree peony (Paeonia ostii ‘Feng Dan’) with five chilling treatments. The cDNA was pyro-sequenced on a 454 GS-FLX platform, and a total of approximately 625,342 raw reads averaging 358.1 bp in length were obtained. After filtering with the default parameters, 223.9 Mb sequences were obtained. Assuming the number of genes and average gene length occurred in P. ostii was similar to that in A. thaliana (25,000) (Bevan and Walsh, 2005), average transcriptome coverage was estimated at 4.5×. After removal of low quality regions, adaptors and all possible contaminations, a total of 23,652 contigs/singletons were obtained by de novo assembling. Contigs and singletons longer than 300 bp were further analyzed, which accounts for 15,284 contigs/singletons and total length of 11.6 Mb with an average of 758.9 bp. Most contigs and singlettons were longer in length than those in the previous 454-EST studies (Fig. 1) (Li et al., 2010; Meyer et al., 2009; Sun et al., 2010). There were only 15 singlettons in the assembling sequences, so they were not analyzed separately.

The length distribution of these high quality ESTs indicated that 70% contigs/singletons were between 300 bp and 800 bp in length and the longest was 5361 bp long (Fig. 1). The proportions of contigs/singletons with lengths longer than 500 bp was 69.6% (10,634). Among them, 42.8% (6,542) were between 500 bp and 800 bp. The result of ORF prediction indicated that 10,327 contigs contained one or more ORF, among them 10,205 contigs had one or more ORF longer than 300 bp. Therefore, this study has identified a large number of transcripts in Paeonia ostii ‘Feng Dan’ flower bud according to the EST analysis.

3.2. Annotation

3.2.1. BLAST

The annotation was based on the sequence similarity search in the Kyoto Encyclopedia of Genes and Genomes database (KEGG) using BLASTX algorithm. UniGenes bigger than 300 bp were subjected to a BLASTX homology search against the plantRefSeq database. A total of 12,345 sequences matched known genes, which represented about 80.8% of the total UniGene set with a significant E-value threshold of 1e-10. UniGenes bigger than 300 bp have been deposited in NCBI’s Transcriptome Shotgun Assembly (TSA, http://www.ncbi.nlm.nih.gov/Genbank/TSA) and the Project was 65,217, and the results are listed in additional file 1, including GenBank ID, contig length, numreads, alignment GenBank ID, its description in plantRefSeq database, and E-value. The remaining 2939 sequences (~19%) and 147 no hit sequences might be unknown genes.

Among the annotated sequences corresponding to known plant proteins, 46.7% matched the Vitis vinifera sequences. This might be due to the similar endured dormancy-release cycle in Vitis vinifera. Meanwhile, a large number of sequences (13.3%) were similar to Populus trichocarpa, a commonly used model for woody plants. In addition, 14.1% sequences matched to Ricinus communis with a significant E-value threshold of 1e-10 (Table 1).

3.2.2. Gene ontology classification

GO Slim terms contain specified subsets of higher-level ontology categories that provide a broad profile for genome-genome comparison (Lomax, 2005). Using Blast2go, the unique genes from our data and TAIR (Arabidopsis thaliana annotated database) were mapped to respective TAIR GO Slim terms, aiming to compare the distribution of gene ontology annotations in our P. ostii 454 data to those of the Arabidopsis genome. For functional annotation, among sequences with E-value < 1e−10, 12, 345 were parsed out and used to search the gene ontology (GO) terms from the GO annotated Arabidopsis database of TAIR. Totally, 8709 sequences could be assigned to molecular function ontology, and 8315, 8290 sequences to cellular component and biological process ontology, respectively (Fig. 2). The well-represented molecular functions were metallochaperone activity and translation regulator activity accounting for 75% and 61.1%, respectively. In terms of cellular components, envelope and membrane-enclosed lumen were well-represented with percentage of 70.7% and 67.9%, respectively. In the biological process category, about 53.7% and 52.4% are related to reproduction and multi-cellular organism process, respectively (Fig. 2). The percentages of annotated P. ostii sequences assigned to GO Slim classes generally mirror those of A. thaliana genes, reflecting a similar distribution of genes in different functional categories, and further highlighting that a large diversity of P. ostii transcripts is represented by these sequences.

Biological processes involved in bud dormancy release were further analyzed using a division class approach described in Zhou et al. (2009). The most abundant category in the analysis was related to cellular component organization with 51.7% including components such as membrane-enclosed lumen proteins, envelop, macromolecular complex. The following categories were metabolic process, response to stimulus and biological regulation, etc. (Fig. 3). Our results differed from a recent study of tree peony (Shu et al., 2009). The most abundant category in the analysis of Shu et al. was protein synthesis, including ribosomal proteins, translation factors and RNA-binding proteins (Shu et al., 2009). The most possible reason for the
3.2. Biochemical pathways

Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis provided an alternative functional annotation based on genes associated with biochemical pathways. Overall, 4247 P. ostii UniGenes were assigned to biochemical pathways matching to 5914 Arabidopsis transcripts. The KEGG metabolic pathways were well-represented by 2461 Arabidopsis thaliana UniGenes (2.56%) than they did (2241 raw ESTs).

<table>
<thead>
<tr>
<th>Taxonomic category</th>
<th>Contigs</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>392</td>
<td>2.56%</td>
</tr>
<tr>
<td>Ricinus communis</td>
<td>2154</td>
<td>14.09%</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>7135</td>
<td>46.68%</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td>2028</td>
<td>13.27%</td>
</tr>
<tr>
<td>Chlamydomonas reinhardtii</td>
<td>30</td>
<td>0.20%</td>
</tr>
<tr>
<td>Zea may</td>
<td>72</td>
<td>0.47%</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td>168</td>
<td>1.10%</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>178</td>
<td>1.16%</td>
</tr>
<tr>
<td>Physcomitrella patens subsp. Patens</td>
<td>78</td>
<td>0.51%</td>
</tr>
<tr>
<td>Micromonas sp. RCC29</td>
<td>23</td>
<td>0.15%</td>
</tr>
<tr>
<td>Ostreococcus lucimarinus CCE9901</td>
<td>21</td>
<td>0.14%</td>
</tr>
<tr>
<td>Other plant</td>
<td>66</td>
<td>0.43%</td>
</tr>
<tr>
<td>nobit or e-value &gt; 1e-10</td>
<td>2939</td>
<td>19.23%</td>
</tr>
<tr>
<td>total</td>
<td>15284</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 1

Summary and taxonomic source of BLAST matches to 454 ESTs.

3.3. Transcription factors or regulators

Transcription factors (TF) play key roles in the regulation of gene expression in response to development and environmental stress in plants (Singh et al., 2002). Based on the BLAST results, many proteins could be identified as putative transcription factors or regulators belonging to different families APETELA2 (AP2), Auxin/Indole-3-Acetic acid (AUX/IAA), Myelocytomatosis viral oncogene homolog (MYC), basic helix-loop-helix (bHLH), WD40, F-box (a protein motif of approximately 50 amino acids that function as a site of protein–protein interaction), WRKY family proteins, Zinc finger (ZIP), Myeloblastosis viral oncogene homolog (MYB), MADS, F-box, Auxin Response Factor (ARF), Ethylene Response Factor (ERF) and so on (Table 2).

The most abundant family in our data set was the zinc finger protein family including CCH, C3H4, C2H2, RING, ZFWD1, DHHC and CSH2. Zinc finger proteins were involved in lateral shoot initiation, floral organ morphogenesis and gametogenesis (Huang et al., 2006; Yilmaz and Mittler, 2008). As most zinc finger proteins were plant specific, further analysis is needed to validate their putative function in tree peony.

Another highly expressed set of transcription factors was the F-box family in our analysis. F-box protein regulates various developmental processes in plants, such as photomorphogenesis, circadian...
rhythm, self-incompatibility and floral meristem and floral organ identity determination (Jain et al., 2007). Huang et al indicated that there was an obvious anatomical change of flower bud from tree peony during endo-dormancy release (Huang et al., 2008a). This may explain the frequent occurrence of F-box protein in the induction and breaking of endo-dormancy. Of course, the function of the F-box proteins in our data set still needs to be tested. In addition, MYB, MADS-box, NAM ATAF1/2 CUC2 (NAC), GRAS, AP2, ARF and WRKY families were also abundant in our study. MYB TFs played important roles in dehydrate response such as dryness, cold, chilling and salt stresses. Chilling requirement of the tree peony bud must be fulfilled to allow proper bud break. Maximum bud break improves with increased chilling exposure (Wang et al., 1996). Further study needs to be performed to understand whether MYB TFs play roles in chilling induced dormancy release. In particular, genes of MADS-box family were very important for floral development, such as flower bud meristem transition, flower organ identity and flower shapes (Theißen, 2001). Identification of these TFs would help to understand the floral development and its role in endo-dormancy release. DAM (Dormancy Associated MADS-BOX) genes were first shown to directly impact endo-dormancy in peach where a deletion of a series of DAM such as MIK-like MADS-box genes results in loss of endo-dormancy induction (Bielenberg et al., 2008). Leafy spurge DAM genes were preferentially expressed in shoot tips and buds in response to cold temperatures and day length in a manner that is relative to the level of endo-dormancy induced by various environmental conditions.

SSRs derived from ESTs essentially present expressed sequences, hence were potential candidate for the construction of markers for gene tagging and comparative genomic studies. In this study, we performed a general screen on the pyrosequencing dataset for the presence of SSRs using SSR Hunter (Li and Wan, 2005). A total of 2253 potential SSRs were found in 1,969 UniGenes, with a frequency of 12.9% in the EST dataset. Di-nucleotide repeats were by far the most common SSRs in our ESTs, with tri-nucleotide repeats being present at much smaller frequencies. Only 14 tetra-nucleotide repeats could be detected with a percentage of 0.62% (Table 3), which was coincide with the one reported by Shu in tree peony (Shu et al., 2009). The most frequent SSR motif was TC (340), followed by AG (323), GA (310), and CT (269) in di-nucleotide repeats, and GAA (36) was the most in tri-nucleotide repeats, followed by AGA, CCA and CAT (32)

### Table 2

<table>
<thead>
<tr>
<th>Factor family</th>
<th>No. of unique transcripts</th>
<th>Factor family</th>
<th>No. of unique transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZIP</td>
<td>54</td>
<td>WRKY</td>
<td>10</td>
</tr>
<tr>
<td>F-box</td>
<td>37</td>
<td>AUX/IAA</td>
<td>6</td>
</tr>
<tr>
<td>NAC</td>
<td>13</td>
<td>bHLH</td>
<td>5</td>
</tr>
<tr>
<td>GRAS</td>
<td>12</td>
<td>FHY3</td>
<td>5</td>
</tr>
<tr>
<td>AP2</td>
<td>10</td>
<td>bZIP</td>
<td>3</td>
</tr>
<tr>
<td>MYB</td>
<td>14</td>
<td>ERF</td>
<td>6</td>
</tr>
<tr>
<td>ARF</td>
<td>11</td>
<td>PHD finger</td>
<td>3</td>
</tr>
<tr>
<td>ERF</td>
<td>6</td>
<td>WD40</td>
<td>3</td>
</tr>
<tr>
<td>MADS</td>
<td>9</td>
<td>others</td>
<td>271</td>
</tr>
<tr>
<td>MYC</td>
<td>6</td>
<td>Total No. of TFs</td>
<td>484</td>
</tr>
</tbody>
</table>

primers successfully amplified the mixed DNA of the two wild species. In addition, comparing the amplified patterns between P. ostii and the bulked two wild species (P. rockii and P. qiui), 73 pairs of primers displayed polymorphic and 23 were monomorphic. Of 73 pairs of primers, 21 (28.8%) are di-nucleotide repeats, 43 (58.9%) are tri-nucleotide repeats, 9 (12.3%) are tetra-nucleotide and pentanucleotide. Among 21 dinucleotide repeat motifs, 20 (95.24%) are (GA/CT)n. Our results differed from recent studies of tree peony (Hou et al., 2011; Wang et al., 2009). The most abundant repeat motif in the analysis of Hou et al. and Wang et al. was di-nucleotide repeat motif, neither tri-nucleotide nor tetra-nucleotide. These polymorphic abundance shows that EST-based SSRs are very useful in population genetics, linkage mapping, comparative genomics, the studies aimed to understand the genetic control of adaptive traits. In addition, they also have excellent potential as genetic markers to assist future breeding strategies.

4. Conclusions

We established normalized cDNA library from flower buds during chilling fulfillment, which was sequenced with massive parallel pyrosequencing on the Roche 454 GS FLX platform. A total of 23,652 assembled UniGenes were obtained with 15,284 annotated UniGenes with a minimal length of 300 bp. To our knowledge, our results represent nearly 7.0-fold more ESTs than all Paeonia suffruticosa ESTs deposited in the public database (http://www.ncbi.nlm.nih.gov/) (as of July 2011). We have identified a large number of SSRs in our 454 EST collection, and designed 149 pairs of high quality PCR primer, among them nearly 80 percent of primer pairs were amplified successfully, at the same time, 73 pairs of primers displayed polymorphism between P. ostii and the two wild species (P. rockii and P. qiui). This study demonstrated that high-throughput transcriptome sequencing could be applied as a fast and cost-effective way to rapidly obtain information on coding region and develop EST-SSR markers which should contribute to population genomic and gene-based association analysis. We believe that all the analysis and information are valuable resources for better understanding of this important plant and also useful for functional study in Paeoniaceae, especially for understanding of dormancy release.

Acknowledgments

This work was supported by grants from National Natural Science Foundation of China (30871728 and 31071828) and National Natural Science Foundation of Shandong province (ZR2009DM002). We would also like to thank Drs Yuehua Cui, Kelian Sun, Chunhai Dong and Zhixia Niu for constructive suggestions in writing and revision of this manuscript.

Appendix A. Additional material

Additional file 1. The BLASTX results from the NCBI NR database for the UniGenes bigger than 300bp

UniGenes bigger than 300 bp were subjected to BLASTX homology search in the plantReSeq database, including contig length, number of reads, alignment GenBank id and its description in plantReSeq database, E-value.

Additional file 2. Predicted SSRs loci f using SSR Hunter

List of the detected SSRs loci including sequence ID, SSR motif, number of repeats, SSR start, SSR end, the length of sequence, the number of SSR motif and the corresponding description.

Additional file 3. Primer sequences for SSR loci

SSR primers were designed using Premier Primer 5.0 software. Information on the name of the sequences containing each SSR, length of primer, GC%, melting temperature, start position of each primer, expected length of PCR product, SSR motif, number of repeats, positions of forward and reverse primers for each primer, length of sequence including SSR motif, and the length of SSR sequence was included.

Additional file 4. Mapping of P. ostii unique putative transcripts to KEGG biochemical pathways

The distributions of biochemical pathways based on Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis aiming to compare the distribution of gene ontology annotations in our P. ostii 454 data to that of the A. thaliana genome.

References


