Transcriptome analysis in maritime pine using laser capture microdissection and 454 pyrosequencing

Rafael A. Cañas, Javier Canales, Josefa Gómez-Maldonado, Concepción Ávila and Francisco M. Cánovas

Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Instituto Andaluz de Biotecnología, Universidad de Málaga, Campus Universitario de Teatinos s/n, Málaga 29071, Spain; Corresponding author (canovas@uma.es)

Received July 30, 2013; accepted November 19, 2013; published online January 3, 2014; handling Editor Ron Sederoff

Maritime pine (Pinus pinaster Aiton) is one of the most advanced conifer models for genomics research. Conifer genomes are extremely large and major advances have recently been made in the characterization of transcriptomes. The combination of laser capture microdissection (LCM) and next-generation sequencing is a powerful tool with which to resolve the entire transcriptome of specific cell types and tissues. In the current work, we have developed a protocol for transcriptomic analyses of conifer tissue types using LCM and 454 pyrosequencing. Tissue sections were isolated using non-fixed flash-frozen samples processed by LCM. Complementary DNA synthesis and amplification from tiny amounts of total RNA from LCM samples was performed using an adapted protocol for Conifer RNA Amplification (CRA+). The cDNA amplification yield and cDNA quality provided by CRA+ were adequate for 454 pyrosequencing. Furthermore, read length and quality results of the 454 runs were near the optimal parameters considered by Roche for transcriptome sequencing. Using the CRA+ protocol, non-specific amplifications were prevented, problems derived from poly(A:T) tails in the 454 sequencing technology were reduced, and read length and read number considerably enhanced. This technical approach will facilitate global gene expression analysis in individual tissues of conifers and may also be applied to other plant species.

Keywords: cDNA amplification, conifers, Pinus pinaster, trees.

Introduction

Laser capture microdissection (LCM) is a technique that allows the isolation of DNA, RNA, proteins (Decarlo et al. 2011) or metabolites (Hölscher and Schneider 2007) from specific cell types or tissues. Thanks to LCM, it is possible to study compartmentalization of processes and to deepen our understanding of rare or specifically localized processes. The combination of LCM with high-throughput techniques is a very powerful tool for global gene expression analysis in plants and facilitates our understanding of complex processes localized in specialized plant cells. Trees are complex organisms with long life cycles, and many different tissues, physiological processes and symbiotic interactions, which require the expression of determinate genes in a very specific form. The combination of LCM with high-throughput expression analysis are is an excellent tool to identify transcriptional networks and key regulatory genes particularly associated with unique tissues and cell types in woody plants.

Next-generation sequencing (NGS) technology is one of the most important new techniques to obtain massive data (Mutz et al. 2013). Next-generation sequencing generates millions of DNA fragments in a single sequencing run. There are several approaches used in NGS technologies yielding different read length and read number per run, but the most often used platforms are Roche 454, Illumina-Solexa and SOLID (Mutz et al. 2013).

As the total RNA amount isolated from LCM samples is very small, it is necessary to amplify the cDNA to obtain enough for NGS. There are at least two major strategies for cDNA
amplification: ‘exponential amplification’ based on polymerase chain reaction (PCR) and multiple rounds of isothermal ‘linear amplification’ based on in vitro transcription with T7 RNA polymerase. Most reports that combine LCM with NGS have used ‘linear amplification’ methods (Emrich et al. 2007, Ohtsu et al. 2007, Scanlon et al. 2009, Li et al. 2010, Takacs et al. 2012, Thiel et al. 2012, Torti et al. 2012). The advantages and inconveniences of the different methods have been previously discussed (Kurimoto and Saitou 2010).

The amplification methods are based on oligo-dT adapters which maintain the poly(A:T) tails, and consequently generate a bias of sequence alignments toward the 3’ end of the transcripts (Emrich et al. 2007, Li et al. 2010, Takacs et al. 2012). Poly(A:T) tails generate serious problems in Roche 454 sequencing, diminishing the number of acceptable reads (Emrich et al. 2007, Ohtsu et al. 2007).

In the Roche 454 (pyrosequencing) system, the light produced by the firefly luciferase during the release of pyrophosphate in the incorporation of a nucleotide by DNA polymerase is used for sequencing (Mardis 2008). Roche 454 cannot properly interpret long stretches (>6 bp), such as polyA tails, leading to errors in the read sequences (Mardis 2008). The Illumina system utilizes a sequencing-by-synthesis approach that incorporates fluorescent nucleotides into template strands. The Roche 454 GS FLX+ System is able to yield read up to 1000 bp against the 150 bp of Illumina HiSeq 2500, although Roche 454 yields up to 1 million reads per run and the Illumina HiSeq 2500 up to 3 billion reads per run.

Our laboratory is interested in molecular and functional studies of maritime pine (Pinus pinaster Aiton), a coniferous tree. Advances in the genomics of conifers have been hampered by the large size of their genomes. Conifer genomes are extremely large in size ranging from 20 to 40 Gb, which is on average more than 200-fold the Arabidopsis genome and nearly sevenfold the human genome (Mackay et al. 2012, Ritland 2012). As the conifer genomes are extremely large, major research efforts have been concentrated in the characterization of transcriptomes (Fernández-Pozo et al. 2011, Rigault et al. 2011, Chen et al. 2012). Recently, massive parallel sequencing data have been used for de novo assembling of the maritime pine transcriptome (Canales et al. 2013). Several initiatives have been launched to sequence the pine genome and the genome drafts of Picea glauca (Moench) Voss and Picea abies (L.) Karst have been recently published (Birol et al. 2013, Nystedt et al. 2013).

In the current work, we have developed an improved protocol for transcriptomic analyses of conifer tissue types isolated by LCM. The tissue sections were isolated using non-fixed flash-frozen samples processed by LCM. The amount of total RNA from LCM samples was extremely limited, but we were able to perform cDNA synthesis and amplification from tiny amounts of total RNA, modifying the conditions of a commercial kit from Evrogen (Moscow, Russia). This protocol is based on a 5’ RACE-PCR designed to synthesize full-length-enriched double-stranded (ds) cDNA from total or poly(A)-enriched RNA, and it has been adapted to the main NGS platforms including Roche 454 sequencing (Babik et al. 2010, Kumar and Blaxter 2010, Xiong et al. 2011, Ekblom et al. 2012). The cDNA amplification yield and quality were optimal for 454 pyrosequencing. Using our improved protocol, the read length and quality results of the Roche 454 runs were near the optimal parameters considered by Roche for transcriptome sequencing. This technical approach will facilitate global gene expression analysis in individual tissues of conifers and may also be applied to other plant species.

Materials and methods

Laser capture microdissection of pine tissue types

Pinus pinaster seeds from the Oria provenance were germinated and grown at 20/24 °C with a 16:8 h photoperiod. Germinating seeds were watered twice a week with distilled water. One-month-old seedlings were sampled and 0.5-cm tissue sections were processed for LCM. Tissue sections were treated for paraffin embedding or cryosectioning.

The paraffin-embedded samples were fixed in ethanol: acetic acid (75 : 25 v/v) with vacuum pulses. The samples were then sequentially treated with 75% ethanol for 1 h, 100% ethanol for 1 h, ethanol : HistoHemol (1:1 v/v) for 1 h and pure HistoHemol (Carlo Erba, Val-de-Reuil, France) for 1 h. Five to six pearls of Paraplast X-tra (Leica, Wetzlar, Germany) were added to the HistoHemol and incubated for 1 h at room temperature. Later on, an equal volume of molten Paraplast X-tra was added to the samples at 58 °C and incubated for 2 h. Finally, the HistoHemol : Paraplast X-tra mix was replaced by pure liquid Paraplast X-tra at 58 °C. The liquid Paraplast X-tra was replaced four times before forming the blocks over the course of 1 day. The embedded samples were stored at 4 °C before sectioning. The samples were cut with a rotary microtome and the sections (10 µm thick) were mounted on polyethylene terephthalate (PET)-membrane 1.4 µm steel frames (Leica) and dried at 37 °C for 1–2 h. Dry slides were deparaffinized twice for 5 min each in HistoHemol. Subsequently, the samples were incubated for 5 min in 100% ethanol and air dried. The microdissection was made using a laser microdissector (LMD700, Leica). The microdissection samples were placed into the caps of 0.5 ml tubes containing 10 µl lysis buffer from the RNAqueous-Micro RNA Isolation Kit (Ambion, Austin, TX, USA). Samples used after were stored at −80 °C. All the RNA extractions from the microdissection of paraffin-embedded samples were made using the LCM protocol of the RNAqueous-Micro RNA Isolation Kit (Ambion). RNA quality was assessed using the RNA Pico Assay for the 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA).

The tissue sections for cryosectioning were mounted in a specimen holder with embedding medium optimal cutting temperature (OCT) (Tissue-Tek, Torrance, CA, USA)
mix was added to each tube. This RT mix included 2 μl of RT master mix. Incubation proceeded at 50 °C for 60 min. After this time 10 ng of total RNA were added to a mix with 1 μl of dNTP (10 mM) and 0.5 μl Mint RT (Evrogen) in a final volume of 10 μl. The 10 μl mixture was added to each tube and mixed by gently pipetting. The reaction tubes were maintained in the thermal cycler except for the time necessary to add the RT master mix. The tubes were incubated at 42 °C for 1.5 h.

The synthesized single-stranded (ss) cDNA was purified with the NucleoSpin® Gel and PCR Clean-up kit (Macheray-Nagel, Düren, Germany) according to the manufacturer’s manual. The ss cDNA was eluted from the columns with 25 μl of preheated water at 70 °C. The column stood for 2 min with the caps open and subsequently the tubes were closed and centrifuged at 14,000 rpm for 1 min to elute the sample. This operation was repeated once again.

To establish the optimal PCR conditions for the ds cDNA synthesis and amplification four PCR of 25, 28, 31 and 34 cycles were set. Each PCR tube contained 11.5 μl of purified ss cDNA, 4 μl M1 primer (10 μM) and 12.5 μl iProof™ HF Master Mix (BioRad, Hercules, CA, USA) in a total volume of 25 μl. The reaction parameters were a denaturing step at 95 °C for 1 min, followed by 25, 28, 31 or 34 cycles of 95 °C for 15 s, 63 °C for 20 s and 72 °C for 3 min. Ten microliters of each reaction were run in a 1.2% agarose gel to verify the PCR conditions in order to carry out the subsequent amplification. The selected PCRs had 25 or 28 cycles. The PCRs with the proper appearance were diluted 10 times in nuclease-free water and used for the amplification.

The ds cDNA amplification was carried out by nested PCR. Five reactions of 50 μl were prepared including the following components: 22 μl nuclease-free water, 25 μl iProof™ HF Master Mix, 2 μl 454-PCR primer mix and 1 μl diluted (1 : 10) ds cDNA from the first PCR step in a final volume of 50 μl. The reaction parameters were a denaturing step at 95 °C for 1 min, followed by three cycles of 95 °C for 15 s, 50 °C for 20 s and 72 °C for 3 min plus 11 cycles of 95 °C for 15 s, 63 °C for 20 s and 72 °C for 3 min. Finally, the five reactions were pooled and purified with the NucleoSpin® Gel and PCR Clean-up kit (Macheray-Nagel) according to the manufacturer’s manual. The ds cDNA was eluted from the columns with 25 μl of preheated water at 70 °C. The column stood for 2 min with the caps open and, subsequently, the tubes were closed and centrifuged at 14,000 rpm for 1 min to elute the sample. This operation was repeated once more.

The quantity of the amplified ds cDNA was determined using the QuantiT™ PicoGreen® dsDNA Kit (Invitrogen, Paisley, UK). The quality of the amplified ds cDNA was determined using the Agilent 7500 DNA Kit in the 2100 Bioanalyzer (Agilent).

454 pyrosequencing
Transcriptome sequencing was performed at the Universidad de Málaga ultrasequencing facility using the GS-FLX+ platform with
a GS-FLX Titanium kit, Roche Applied Sciences (Indianapolis, IN, USA). Each sample was run in one-half of a 454 PicoTiterPlate following the manufacturer’s sequencing protocol.

The quantity of the cDNA libraries was determined using the Quant-iT™ Green dsDNA Kit (Invitrogen). The quality of the cDNA libraries was determined using the Agilent High Sensitivity DNA Kit in the 2100 Bioanalyzer (Agilent).

The runs were analysed using the Roche GS-FLX+ software. The resulting reads were trimmed and analysed using the SeqTrimNext software (Falgueras et al. 2010).

Quantitative PCR

The expression analyses were carried out by quantitative PCR (qPCR) in LCM samples. Total RNA (2.5 ng) was used to synthesize cDNA using the iScript™ Reverse Transcription Supermix for RT-qPCR (BioRad), which is able to work with a broad linear dynamic range of total RNA input (1 µg–1 pg). Following the instruction manual, the reverse transcription conditions were: 5 min at 25 °C, 30 min at 42 °C and finally 5 min at 85 °C. The reverse transcription was carried out in a thermal cycler DNA Engine® (BioRad).

The primers used for the qPCRs were as follows:

- Actin-Forward (5′ATCTCTCAGCACATTCCACAG3′)
- Actin-Reverse (5′GCTTGGAGATGGTGGAATATA3′)
- EF1A-Forward (5′TGCTGTTGGAGTCTATAGG3′)
- EF1A-Reverse (5′CTCGTGCATCAGATAATGAG3′)
- GS1a-Forward (5′ATCGAGGACTTCTCAGTATGAGGT3′)
- GS1a-Reverse (5′GGTGTGTCCTCAGAATTACATAGAAT3′)
- GS1b-Forward (5′CAGCTTCAAACATGAGCGCTAT3′)
- GS1b-Reverse (5′TGGGGATCTCCCTGCATCAATG3′)
- AS1-Forward (5′CATACAGAAATGGGTGGG3′)
- AS1-Reverse (5′CCCATCACAATAAACCAG3′)
- ASPG-Forward (5′AGTAATGGGACAGCTTGGTC3′)
- ASPG-Reverse (5′TCCAACAAGACACAGGGTG3′)

The qPCR was carried out using SsoFast™ EvaGreen® Supermix (BioRad). The reaction mix was prepared as follows: 5 µl of 2x SsoFast™ EvaGreen® Supermix, 0.5 pmol of Forward primer, 0.5 pmol of Reverse primer and 1 µl of the diluted one-half cDNA (125 pg from the original total RNA sample). The qPCR were carried out in a thermal cycler CFX384 (BioRad) under the following conditions: 3 min at 95 °C (1 cycle), 1 s at 95 °C and 5 s at 60 °C (50 cycles) and a melting curve from 60 to 95 °C.

The raw fluorescence data from each reaction were fitted to the MAK2 model, which requires no assumptions about the amplification efficiency of a qPCR assay (Boggy and Woolf 2010). The initial target concentrations (D0 parameter) for each gene were deduced from the MAK2 model using the qpcR package for the R environment (Ritz and Spiess 2008) and normalized to the geometric mean of two reference genes (actin and EF1A). For the qPCR analysis, three biological replicates and three technical replicates per sample were made.

Read alignment and read count

The read obtained in the 454 sequencing were mapping against the SustainpineDB contigs as reference (Canales et al. 2013, http://www.scbi.uma.es/sustainpinedb/home_page) using the maximal exact matches (MEM) option of the Burrows–Wheeler Aligner (BWA) software for long reads (Li and Durbin 2010; H. Li and R. Durbin, unpublished). The scripts used for BWA-MEM analysis were as follows:

- bwa index -p sustainpine_v3.0_unigenes -a bwtsw sustainpine_v3.0_unigenes.fasta
- bwa mem sustainpine_v3.0_unigenes sample.fastq > sample.sam

The number of reads for each gene was counted using the Tablet assembly visualization software (Milne et al. 2013). The contigs considered for counting the reads of the genes were the following: for actin sp_v3.0_unigene17257 and sp_v3.0_unigene18113; for EF1A sp_v3.0_unigene1012; for GS1a sp_v3.0_unigene11788 and sp_v3.0_unigene10612; for GS1b sp_v3.0_unigene10503; for AS1 sp_v3.0_unigene14147; and for ASPG sp_v3.0_unigene4029.

Results and discussion

Setting up a working protocol for LCM in pine

We determined the most appropriate method to preserve RNA integrity for LCM. With regard to conifers, a single report is found in the literature describing the isolation of tissue-specific RNA by LCM from Picea glauca (Abbott et al. 2010). We initially started to assay LCM from paraffin-embedded samples (32 RNA extractions) and also from flash-frozen samples embedded in OCT medium for cryostats (>50 RNA extractions) (Figure 1a and b). The total RNA obtained from paraffin-embedded sections had in all cases a suboptimal quality for 454 pyrosequencing (Figure 1c) with an RIN <7, and even in many cases <5, which is considered to be the minimum RIN value to quantify transcript levels using real-time PCR (qPCR) (Fleige and Pfaffl 2006, Abbott et al. 2010). In contrast, the RIN values observed in the flash-frozen samples were always >5, and in most cases >7 (Figure 1d).

Although the paraffin-embedded samples were fixed with ethanol: acetic acid (75 : 25 v/v), they displayed problems by RNase re-activation after microtome cutting and/or in the time exposed to the air during the LCM process. This could be related to the ambient humidity (60–70% in our laboratory), as was previously observed for paraffin-embedded samples (Ordway et al. 2010). In contrast, the flash-frozen samples were not affected by the ambient humidity although the tissue sections were only fixed in absolute ethanol after cryostat cutting. In both treatments, the amount of isolated total RNA depended primarily on the number of sections used. However, the amount of total RNA, ranging between 15 and 60 ng, was insufficient for 454 pyrosequencing. Owing to low RNA yield, it was necessary to synthesize and amplify cDNA from all tissue-type samples. According to these
Figure 1. Laser capture microdissection and RNA isolation from maritime pine tissues. Images of LCM in root tips (a and b). In (a), the microdissection areas are selected by lines. (b) The microdissection areas are cut. Black scale bars in the pictures represent 400 \( \mu \)m. Analysis of isolated RNA from LCM samples (c and d). Virtual electropherograms from Agilent RNA 6000 Pico chip runs performed in an Agilent’s 2100 Bioanalyzer. (c) RNA sample isolated from LCM samples using the paraffin-embedded protocol. (d) Total RNA was isolated from LCM samples using the cryosectioning protocol. The RIN is shown in the image.
results, we decided to select the flash-freezing protocol as the best source for LCM, in close agreement with the data previously reported in *Picea glauca* (Abbott et al. 2010).

**Complementary DNA synthesis and amplification: evaluation of pre-existing and improved protocols**

The synthesis and amplification of cDNA from small quantities of total RNA or poly(A)+ RNA is the usual method in molecular biology and the lower limits of input RNA depend on the method chosen but can be just a few picograms (Kurimoto and Saitou 2010). To amplify pine RNA isolated from LCM, we decided to use the Mint2 kit from Evrogen because it is a 5'-RACE-PCR-based method able to amplify full-length mRNA from total RNA and poly(A)+ RNA (Matz et al. 1999, Schmidt and Mueller 1999). The main reason for using this method in our laboratory is that it is especially well adapted for Roche 454 sequencing as we previously confirmed with conventional RNA samples (see Table S1 available as Supplementary Data at *Tree Physiology* Online). This method is able to reduce and modify the poly(A:T) tails favouring the performance of 454 sequencing. This protocol for conifer RNA amplification and 454 sequencing was named CRA.

In the current work, the CRA protocol was adapted for the amplification of total RNA samples isolated by LCM from tissue types of maritime pine, and the improved protocol was named CRA+. In the present work we show data from the following samples: cotyledon parenchyma (CP), hypocotyl pith (HM.1 and HM.2) and root vascular bundles (RV). The total RNA yield of our samples was very low, ranging from 0.5 to 2 ng/µl. In the CRA+ protocol, the volume of total RNA used for the amplification was 5 µl corresponding to 2.5–10 ng of total RNA, which is much lower than the minimal total RNA amount accepted by the CRA original protocol (250 ng) (Evrogen Mint2 cDNA synthesis manual, http://www.evrogen.com/kit-user-manuals/Mint-2.pdf, 17 December 2013, date last accessed). In the CRA protocol, the ss cDNA synthesis was developed at 42 °C and in the CRA+ at 50 °C, and we divided the first-strand synthesis and the PlugOligo-Adapter incorporation into two steps, following the recommendations by Pinto and Lindblad (2010). Although the yield of ds cDNA synthesis from ss cDNA in the samples at 42 °C was higher (Figure 2, CRA lanes) than in the samples incubated at 50 °C (Figure 2, CRA+ lanes), the increased temperature in the first-strand cDNA synthesis reduced the problems related to the non-specific ss cDNA synthesis (Table 2).

Another parameter modified in the CRA protocol was the number of PCR cycles used for ds cDNA synthesis. The recommended number of cycles (from 15 to 24) in the CRA protocol was insufficient to obtain a visible amplification. We found that the optimal number of cycles to amplify our samples was between 25 and 34 cycles, and 25 or 28 cycles were usually selected in CRA+ (Figure 2). However, the samples incubated at 50 °C during the first-strand synthesis had very low yields in this step even if we included in the PCR mix all the ss cDNA generated in the previous step. This is a critical point because it can result in the need to begin the protocol again due to poor ss cDNA or ds cDNA synthesis. DNA polymerase was also changed in CRA+ to improve the protocol. The high-fidelity iProof™ HF enzyme supplied by BioRad is faster (2 versus 1 kb min⁻¹) and able to amplify longer sequences (37 versus 15 kb) than the Encyclo polymerase mix. The characteristics of the iProof™ HF allow the protocol to diminish the non-specific amplifications in reactions with a supraoptimal number of cycles (Figure 2). As in regular PCR amplifications, the products with a smaller length are promoted against the higher-length products in the course of the reaction.

Table 1. Summary of the sample 454 sequencing.

<table>
<thead>
<tr>
<th>Sample</th>
<th>cDNA protocol</th>
<th>Library length average (bp)</th>
<th>Reads number</th>
<th>Read length average (bp)</th>
<th>Median read length (bp)</th>
<th>Total bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM.1</td>
<td>CRA</td>
<td>1355</td>
<td>141,862</td>
<td>300</td>
<td>308</td>
<td>42,552,106</td>
</tr>
<tr>
<td>RV</td>
<td>CRA</td>
<td>1540</td>
<td>448,120</td>
<td>521</td>
<td>546</td>
<td>233,457,069</td>
</tr>
<tr>
<td>Mean</td>
<td>CRA</td>
<td>1448</td>
<td>294,991</td>
<td>411</td>
<td>427</td>
<td>138,004,588</td>
</tr>
<tr>
<td>HM.2</td>
<td>CRA+</td>
<td>1083</td>
<td>631,059</td>
<td>632</td>
<td>712</td>
<td>399,055,868</td>
</tr>
<tr>
<td>CP</td>
<td>CRA+</td>
<td>1274</td>
<td>776,974</td>
<td>653</td>
<td>737</td>
<td>507,257,184</td>
</tr>
<tr>
<td>Mean</td>
<td>CRA+</td>
<td>1179</td>
<td>704,017</td>
<td>643</td>
<td>725</td>
<td>453,156,526</td>
</tr>
</tbody>
</table>

Figure 2. Double-stranded cDNA first-round amplification from total RNA of LCM tissue samples. Double-stranded cDNA amplifications performed with the original CRA and modified CRA+ protocols. Lane numbers 25, 28, 31 and 34 indicate the PCR cycle numbers. Ten microliters of PCR product were loaded per lane. Lane L corresponds to 0.5 µg of a 1 kb ladder (GeneRuler™ 1 kb Plus, Fermentas). DNA electrophoresis was carried out in a 1.2% agarose/EtBr gel run in 1× TAE buffer.
The amplification of cDNA was performed in the CRA+ protocol after the ds cDNA synthesis without the purification performed in CRA. This step was omitted to avoid the loss of cDNA in the samples. In both cases, the PCRs were run with a pre-cycling step of three cycles and a second cycling step of 11 cycles. The number of cycles of this second cycling can be modified through a test PCR as carried out for the synthesis of ds cDNA with a number of optimal cycles from 11 to 17. The cDNA amplification was performed from a 250 µl PCR mix divided into five tubes and the products were purified together for subsequent use. Depending on the obtained yield, it is possible to increase the volume of PCR mix to reach the appropriate cDNA amount. The final cDNA yield was > 1 µg.

The quality of the amplified cDNA with CRA and CRA+ protocols was examined and the observed profiles were quite different (Figure 3). Samples amplified with the CRA protocol showed DNA smearing with a maximal intensity < 1000 bp and very intense discrete bands (Figure 3a, HM.1, RV). In contrast, the samples amplified with the CRA+ protocol showed a homogeneous DNA smearing with a maximal intensity ~1400 bp (Figure 3a, HM.2, CP). The isolated peaks observed in the CRA protocol samples suggest that the 42 °C ss cDNA synthesis promotes the non-specific reverse transcription of abundant RNAs (Figure 3b), while the ss cDNA synthesis at 50 °C is much more restrictive and linear (Figure 3c). All these differences are important to determine whether the cDNA synthesis and cDNA amplification were appropriate for GS-FLX+ sequencing.

GS-FLX+ sequencing
In the 454 sequencing, the conventional Roche protocols were followed without changes. Sequencing runs were carried out with samples generated with CRA and CRA+ protocols. The ds

<table>
<thead>
<tr>
<th>cDNA protocol</th>
<th>HM.1</th>
<th>RV</th>
<th>HM.2</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRA</td>
<td>14,112</td>
<td>44,812</td>
<td>63,1059</td>
<td>776,974</td>
</tr>
<tr>
<td>CRA+</td>
<td>3599</td>
<td>3616</td>
<td>459,228</td>
<td>591,958</td>
</tr>
<tr>
<td>Rejected</td>
<td>137,200</td>
<td>444,235</td>
<td>135,809</td>
<td>132,935</td>
</tr>
<tr>
<td>Low complexity</td>
<td>313</td>
<td>269</td>
<td>36,022</td>
<td>52,081</td>
</tr>
<tr>
<td>Mean input reads</td>
<td>694</td>
<td>800</td>
<td>916</td>
<td>962</td>
</tr>
<tr>
<td>Mean output reads</td>
<td>163</td>
<td>272</td>
<td>368</td>
<td>372</td>
</tr>
<tr>
<td>Low quality, %</td>
<td>1.83</td>
<td>0.33</td>
<td>22.06</td>
<td>23.38</td>
</tr>
<tr>
<td>Contaminants, %</td>
<td>0.88</td>
<td>0.58</td>
<td>4.45</td>
<td>3.22</td>
</tr>
<tr>
<td>Adapters, %</td>
<td>3.94</td>
<td>1.89</td>
<td>3.56</td>
<td>3.15</td>
</tr>
<tr>
<td>Indeterminations, %</td>
<td>2.08</td>
<td>0.75</td>
<td>1.16</td>
<td>1.09</td>
</tr>
<tr>
<td>Poly(AT), %</td>
<td>0.03</td>
<td>0.01</td>
<td>0.89</td>
<td>0.80</td>
</tr>
<tr>
<td>Low complexity, %</td>
<td>0.04</td>
<td>0.01</td>
<td>0.52</td>
<td>0.64</td>
</tr>
<tr>
<td>Nucleotide useful, %</td>
<td>0.65</td>
<td>0.30</td>
<td>31.55</td>
<td>32.04</td>
</tr>
<tr>
<td>Contaminant, %</td>
<td>1.54</td>
<td>0.86</td>
<td>6.93</td>
<td>4.75</td>
</tr>
<tr>
<td>Low complexity, %</td>
<td>0.01</td>
<td>0.00</td>
<td>2.47</td>
<td>3.02</td>
</tr>
<tr>
<td>Short inserts, %</td>
<td>76.81</td>
<td>51.17</td>
<td>0.49</td>
<td>0.43</td>
</tr>
<tr>
<td>Repeated sequences, %</td>
<td>6.20</td>
<td>43.00</td>
<td>11.51</td>
<td>8.79</td>
</tr>
<tr>
<td>Indeterminations, %</td>
<td>0.21</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Empty inserts, %</td>
<td>1.08</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>No valid inserts, %</td>
<td>11.38</td>
<td>4.05</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Rejected sequences, %</td>
<td>97.23</td>
<td>99.13</td>
<td>21.52</td>
<td>17.11</td>
</tr>
</tbody>
</table>
cDNA samples were nebulized before the generation of the 454 FLX+ libraries due to the length of the amplified cDNAs (Figure 3). All the constructed libraries had a similar appearance independent of whether the cDNA preparations were obtained using CRA or CRA+ protocols (Figure 4). In all the cases the libraries showed a wide band with a maximal peak between 1083 and 1540 bp (Table 1). Consequently, the lengths and profiles observed in the libraries were not useful parameters to determine whether the samples were suitable or not for 454 sequencing. We would recommend therefore to carefully examine the quality of the amplified cDNA rather than the profile of the library prior to 454 sequencing.

Table 1 shows the raw data of 454 sequencing runs from two LCM samples amplified with CRA (HM.1 and RV) and CRA+ (HM.2 and CP) protocols. Table 1 also shows the accumulative data obtained using both protocols. Each library was sequenced in one-half of a 454 PicoTiterPlate. The average number of reads was much higher in the sequencing of CRA+ libraries (704,017) compared with CRA libraries (294,991). The same was also true for other parameters such as read-length average and total number of sequenced bases. Thus, the read-length average of CRA+ libraries was 643 bp, whereas a much lower value was observed in CRA libraries (411 bp). The total number of sequenced bases using the CRA+ protocol was 453,156,526, whereas this value was significantly reduced when using the CRA protocol. Taken together, all these sequencing data clearly indicate that the number of reads, the mean length and the median length of reads were higher in the samples amplified using the CRA+ protocol, and subsequently also the total number of sequenced bases (Table 1). The observed values of the above-mentioned parameters using CRA+ were in the range of the optimal values recommended for 454 sequencing. This suggests that the CRA+ protocol promotes the linear synthesis of ss cDNA, avoiding the preferential synthesis of small abundant and non-specific sequences, improving the cDNA quality that is essential for accurate and high-quality 454 sequencing.

This conclusion is further supported by the frequency analysis of the distribution profiles of read lengths resulting from GS-FLX Titanium pyrosequencing (Figure 5). In the CRA libraries, the number of reads was not distributed homogeneously throughout the profile, with a large number of single peaks (Figure 5a). In the CRA+ libraries the number of reads was distributed homogeneously through the profiles without single peaks of read accumulation. Figure 5b clearly shows that the frequency of distribution of read lengths in the CRA+ libraries is close to a normal distribution, between 600 and 1000 bp lengths, with a long tail in the low read lengths, between 1 and 600 bp. These profiles are characteristic of the optimal 454 sequencing.

**Quality of the 454 sequencing**

Table 2 shows the quality analyses of 454 sequencing runs performed with two LCM samples amplified using CRA and CRA+ protocols. For these analyses, the trimming software SeqTrimNext (Falgueras et al. 2010) was used. The percentage of rejected sequences was higher in the CRA (98.18%) than in

![Figure 4. Electropherograms of the 454 libraries synthesized from the amplified ds cDNA. Virtual gel images and electropherogram from Agilent High Sensitivity DNA chip runs in an Agilent 2100 Bioanalyzer. Libraries constructed with ds cDNA amplified using the CRA protocol (CRA panel). Libraries constructed with ds cDNA amplified using the CRA+ protocol (CRA+ panel). Library length averages of the samples are indicated in parentheses.](https://academic.oup.com/treephys/article-abstract/34/11/1278/1678210)
the CRA+ samples (19.32%). Consequently, the percentages of useful nucleotides were much lower in the CRA (0.48%) than in the CRA+ (31.80%) samples. The mean length of the input reads in the CRA samples (747 bp) was lower than that in the CRA+ samples (939 bp). Consequently, the mean lengths of the output sequences were much lower in CRA than in CRA+ samples. These data show that the modifications introduced to the CRA+ protocol improved the quality of cDNA for subsequent 454 pyrosequencing. When we analyse the quality indexes of the reads, we observe that the CRA protocol was unable to avoid reverse transcription of non-specific RNA sequences (Table 2). The percentage level of no valid inserts, short inserts and repeated reads were usually lower in the CRA+ samples. Likely, this problem is due to the low amount of input mRNA that promotes the non-specific reverse transcription of abundant RNA as observed in the CRA samples. Non-specific amplification was prevented in the CRA+ samples by increasing the incubation temperature in the synthesis of first-strand cDNA.

The percentage of the poly(A:T) nucleotide regions eliminated was very low in all the sequenced samples (Table 2).

This shows that the RNA amplification protocols used here are able to effectively modify the poly(A:T) tails. This result supports the choice of the current protocol to process total RNA samples from LCM for 454 sequencing, avoiding the problems found in other studies that combine LCM and 454 sequencing (Emrich et al. 2007, Ohtsu et al. 2007). Another important problem of cDNA synthesis and amplification kits based on poly(T) primers/adapter is the bias in the 3′ end transcript sequences. In our case we observed this tendency, with sequence alignments towards the 3′ end of the transcripts in all the samples, no matter which protocol was employed (R. Cañas, L. Sterck and F.M. Cánovas, unpublished data), as has previously been described (Emrich et al. 2007, Babik et al. 2010, Li et al. 2010, Takacs et al. 2012).

Validation of 454 sequencing results by qPCR

With the aim of confirming the robustness of the 454 sequencing data obtained from the CRA+ samples, we performed qPCR analysis to compare the expression of different known genes. We analysed the expression of four genes that have been previously studied in our laboratory (glutamine synthetase 1a, GS1a; glutamine synthetase 1b, GS1b; asparagine synthetase 1, AS1; asparaginase, ASPG) and two reference genes regularly...
used in our laboratory (actin; elongation factor 1α, EF1A). Complementary DNA was synthesized and qPCR data (Figure 6a) were compared with the normalized number of reads for each gene in the 454 sequencing (Figure 6b). The qPCR analysis supports that the CRA+ is able to avoid gene bias in the cDNA amplification despite the high number of PCR cycles used in the protocol. Considering the number of useful reads in the 454 sequencing (HM.2, 459,228 reads; CP, 591,958 reads), qPCR- and CRA+-derived data showed high similarity. The observed differences could be explained by the lack of biological replicates in the 454 sequencing. The analysis is much more difficult in the CRA samples with a low number of useful reads (HM.1, 3599 reads; RV, 3616 reads). The reads for each gene were normalized, dividing by the total number of reads in the sequencing run. Although the normalization was able to recuperate some data for the expression analysis, the low number of reads in the CRA samples indicates that the CRA protocol is not suitable for expression analyses.

Conclusions

In the current work, we have successfully adapted pre-existing protocols for LCM and 454 transcriptome sequencing in the conifer tree P. pinaster. We have been able to obtain RNA preparations with good quality indexes from isolated tissue sections by LCM employing snap-frozen samples. Using an improved protocol for CRA+, we were able to obtain enough high-quality cDNA for 454 sequencing. The 454 runs generated quality parameters very similar to those obtained in runs performed with conventional RNA samples. Using our CRA+ protocol, we mainly avoid non-specific amplifications, reduction of read length and read number, and finally problems derived from poly(A:T) tails in 454 pyrosequencing.

Supplementary data

Supplementary data for this article are available at Tree Physiology Online.

Acknowledgments

We thank Prof. Jose Maria Pomares for the use of the cryostat HM 525. We thank Zhen Li from VIB (Ghent, Belgium) for the analysis of the bias in the 3′ end transcript sequences.

Conflict of interest

None declared.

Funding

This work was supported by the European Commission Seventh Framework grant PROCOCEN (FP7-KBBE-2011-5).

References


