Annotation and Re-Sequencing of Genes from De Novo Transcriptome Assembly of Abies alba (Pinaceae)

Authors: Roschanski, Anna M., Fady, Bruno, Ziegenhagen, Birgit, and Liepelt, Sascha

Source: Applications in Plant Sciences, 1(1)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1200179

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne’s Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.
To gain insights into the molecular level of adaptation, attention has turned to the investigation of adaptively relevant genes (candidate genes). For nonmodel organisms, access to candidate genes is limited and the transfer of primers, e.g., from expressed sequence tag (EST) libraries, if available, requires high labor costs. For instance, the resequencing of 800 genes selected from more than 7000 ESTs from *Pinus taeda* L. yielded only 70 candidate genes for *Abies alba* Mill. (Mosca et al., 2012). Because sequencing costs are decreasing rapidly, de novo sequencing in nonmodel organisms is now achievable. For the identification of candidate genes in de novo–sequenced organisms, the use of differential expression profiling (e.g., Street et al., 2006; Huang et al., 2012) can be performed, but it requires the sequencing of several samples. The sequencing of a single transcriptome, in contrast, is very cost-effective. However, the reduction of the data remains challenging. Blasting against available databases is the standard method, which results in outputs of large quantities and is therefore mainly used for annotation only (e.g., Parchman et al., 2010). Here, we present a protocol for the efficient reduction of transcriptomic data down to 283 candidate gene sequences that were used for immediate primer development. The protocol is applicable for species that lack genomic resources. It combines a standard and a specific annotation approach and led to the resequencing of 88 gene fragments in *A. alba*.

METHODS AND RESULTS

A normalized transcriptome of a 1-yr-old *A. alba* seedling from the Black Forest (Forest District Calw, Germany; voucher MB-P-001007, Herbarium Marburgense, University of Marburg) was sequenced on a 454 GS FLX Titanium platform (cDNA library preparation: Veridis Biotechnology AG, Freising, Germany; sequencing: Genoscreen, Lille, France). The 454 run yielded 1521 698 reads with an average length of 359 nucleotides (nt). Trimming and de novo assembly of the raw reads into contigs using Newbler software version 2.3 (454 Life Sciences, Branford, Connecticut, USA) resulted in 25 149 contigs consisting of 381 808 complete and 619 615 partially assembled reads. The contig length was between 100 nt and 2394 nt, with an average length of 498 nt. A total of 484 576 reads remained as singletons (Table 1). Contigs were submitted to the Transcriptome Shotgun Assembly database (TSA) at the National Center for Biotechnology Information (NCBI) (accession no.: JV134525–JV157085).

In the specific approach (Fig. 1), we tested a novel annotation protocol: After a literature survey with key words “adaptation,” “candidates,” “drought,” “evolution,” “RT-PCR,” and “selection” in various combinations using the Web of Science database, we selected 5349 unique proteins and downloaded them from UniProt or NCBI (downloaded in November 2011). The proteins were subsequently searched against the contigs coming from the de novo transcriptome sequencing that were formatted as the reference database using the BLAST+ 2.2.24 toolkit (tblASTx parameters: softmasking = threshold 15 max_target_seqs 10000). To increase reliability of alignments and to avoid too-short amplicons, only alignments with a length of at least 100 amino acids and an identity of at least 90% were considered further. From the contigs that passed the filter, 157 were selected for primer design. In the standard approach (Fig. 1), contigs were searched against the refseq_protein database (downloaded from NCBI 14 June 2011) with strict BLAST-settings (BLASTx parameters: threshold old 999, window-size 4, gapopen 32767, gapextend 32767, E-value 1e −20 ) (Altschul et al., 1990). Gene ontologies (Ashburner et al., 2000) were assigned to contig-protein hits using Blast2GO 2.5.0 (Conesa et al., 2005) and subsequently filtered as described above. To select for well-described proteins, contig sequences were used for primer design if they could be assigned to enzyme IDs with the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata et al., 1999) in the final annotation step. Primers were developed specifying the amplified range according to the contig-protein alignment boundaries using default standard PCR settings of PeriPrimer (version 1.11; Marshall, 2004). Primers were tested in a 30 μL PCR reaction with 17.28 μL double-distilled water, 3 μL 10× PCR buffer with MgCl₂ (20 mM), 1.2 μL MgCl₂ (25 mM), 3 μL Primermix (forward and reverse each 2 μM), 1.44 μL dNTPs (each 5 mM), and 0.12 μL Taq polymerase (Invitrogen, Carlsbad, California). Primers were tested for specificity by running 3% agarose gel electrophoresis. Amplification products were sequenced on a 3730xl DNA Analyzer (Applied Biosystems).
Table 1. Statistics of the 454 transcriptome sequencing run and metrics of the Newbler assembly software.

| Sequence type          | Number | %   | Nucleotides | Average (nt) | 0%   | 25%   | 50%   | 75%   | 100%  |
|------------------------|--------|-----|-------------|--------------|------|-------|-------|-------|-------|
| Reads trimmed           | 1521698| 100 | 546346058   | 359.0        | <21 | <303  | <395  | <444  | <1088 |
| Reads assembled         | 381808 | 25.1|             |              |      |       |       |       |       |
| Reads partial           | 619615 | 40.7|             |              |      |       |       |       |       |
| Reads singleton         | 484576 | 31.8| 175198711   | 361.6        | <50 | <307  | <397  | <443  | <876  |
| Reads repeat            | 1617   | 0.1 |             |              |      |       |       |       |       |
| Reads outlier           | 20389  | 1.3 |             |              |      |       |       |       |       |
| Reads too short         | 13693  | 0.9 |             |              |      |       |       |       |       |
| Contigs                 | 25149  |     | 12511848    | 498          | <100| <365  | <468  | <601  | <2394 |
| N50 Contig*             |        |     |             |              | 704 |       |       |       |       |

*Half of all bases are assembled in contigs of this size or longer.

0.24 μL bovine serum albumin (BSA) (20 mg mL⁻¹), 0.24 μL Dream Taq polymerase (5 U μL⁻¹, Fermentas, St. Leon-Rot, Germany), and 3.6 μL DNA (10 ng μL⁻¹). The PCR was performed with 5 min initial denaturation at 94°C followed by 35 repetitions of 45 s denaturation at 94°C, 45 s annealing at 52–59°C, 45 s elongation at 72°C, and a 10 min final elongation at 72°C. For the amplification test, four samples were randomly chosen for each gene from a set of 80 different silver fir trees that were sampled in May 2011 in Mont Ventoux (44°10′44.35″N, 5°14′32.29″E, France). Amplification was evaluated by electrophoresis in 1% agarose gels. When amplification was too weak, the volume of MgCl₂ was increased to 1.8 μL. When faint ancillary bands appeared, no additional magnesium was added to the mastermix. If PCR products occurred as a single band, one sample was chosen for sequence analysis in each case to ensure that the region of interest was amplified (LLC Genomics GmbH, Berlin, Germany). Gene sequences were aligned to the corresponding contigs using the CodonCode Aligner software (default large gap settings) to reveal the location of the introns. The gene sequences were searched against the nr nucleotide database of NCBI (default contiguous megaBLAST settings, web application).

In the specific approach, BLASTn and subsequent sorting led to 321 contigs. For primer development, 185 contigs were picked. In the standard approach, the initial number of contigs was decreased to one third after the BLASTx step. Approximately half of the hits could be further annotated with Gene Ontologies. After filtering, 126 contigs were successfully assigned to enzyme-IDs and used for primer design (Fig. 1). In combination, 283 different contigs were annotated and only 28 were annotated with both approaches. Primer testing and sequencing resulted in 88 gene sequences (Table 2). Fifty-seven genes were annotated using the specific approach, and 42 using the standard approach. Eleven were annotated by both approaches. The assembly of the gene sequences and the corresponding cDNA contigs revealed 43 introns in 26 genes. The length of the gene sequences ranged from 262 to 1486 nt. All gene sequences aligned to sequences from the nr nucleotide database (NCBI) where the highest E-value was 5.00e⁻¹². Twelve gene sequences hit organelle DNA (10 chloroplast, one mitochondrial, and one ribosomal). The remaining 76 are involved in the biosynthesis of different compounds (21), regulation (20), primary metabolism (14), growth (11), stress response (8), and water transport (2). In the biosynthesis group, enzymes from the auxin pathways, the phenylpropanoid pathways, and the tetrapyrrol pathways were dominant. With the exception of the primary metabolism group, all groups included candidates for the analysis of adaptation at gene level that had been investigated in previous studies of conifers (e.g., González-Martínez et al., 2006).

**CONCLUSIONS**

The two approaches of the workflow are complementary, each contributing approximately half of the annotations in the final set of sequences. The standard approach can be run rapidly, but targets only well-known genes. The specific approach based on a review of the relevant literature is novel and provided a substantial amount of nonredundant annotations. As an advantage, it can be easily adjusted and extended freely to the researcher’s interest. The quality-tested primers can be used for assessing the degree of gene polymorphism in ecological genetics studies.

**LITERATURE CITED**

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.

ASHBURNER, M., C. A. BALL, J. A. BLAKE, D. BOSTEIN, H. BUTLER, J. M. CHERRY, A. P. DAVIS, ET AL. 2000. Gene Ontology: Tool for the unification of biology. Nature Genetics 25: 29–29.

CONEA, A., S. GÖTZ, J. M. GARCÍA-GÓMEZ, J. TEROL, M. TALÓN, and M. ROBLES. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics (Oxford, England) 21: 3674–3676.

GONZÁLEZ-MARTÍNEZ, S. C., E. ERESOZ, G. R. BROWN, N. C. WHEELER, and D. B. NEALE. 2006. DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in Pinus taeda L. Genetics 172: 1915–1926.

HUANG, H.-R., P.-C. YAN, M. LASCOUX, and X.-J. GE. 2012. Flowering time and transcriptome variation in Capsella bursa-pastoris (Brassicaceae). New Phytologist 194: 676–689.

MARSHALL, O. 2004. PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite, and real-time PCR. Bioinformatics 20: 2471–2472.

MOSCA, E., A. J. ECKERT, J. D. LEECHTY, J. L. WEGRZYN, N. LA PORTA, G. G. VENDRIMAN, and D. B. NEALE. 2012. Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. Evolutionary Applications 5: 762–775.

OGATA, H., S. GOTO, K. SATO, W. FUJIBUCHI, H. BONO, and M. KANEHISA. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 27: 29–34.

PARCHMAN, T. L., K. S. GEIST, J. A. GRAHNEN, C. W. BANKMAN, and C. A. BUEKLE. 2010. Transcriptome sequencing of an ecologically important tree species: Assembly, annotation, and marker discovery. BMC Genomics 11: 180.

STREET, N. R., O. SKOOGSTROM, A. SIODIN, J. TUCKER, M. RODRÍGUEZ-ACOSTA, P. NELSON, S. JANSSON, and G. TAYLOR. 2006. The genetics and genomics of the drought response in Populus. Plant Journal 48: 321–341.
Fig. 1. Workflow of the annotation protocol. Numbers of the output after each step are given. The standard approach starts with 25,149 contigs. The specific approach uses them as the reference database for the tBLASTn step.
| Gene Locus ID | Primer sequences (5′–3′) | Ta (°C) | No. of introns | Intron length (nt) | Total length (nt) | Annotation approach | BLASTn of gene sequences against nr nucleotide database (E-value) | GO-Id biological process |
|--------------|---------------------------|---------|---------------|-------------------|------------------|-------------------|---------------------------------------------------------------|--------------------------|
| 95           | F: ACGAAACTAAAGCTAGTGCG  | 57      | 0             | —                 | 696              | 1                  | *Keteleeria davidiana* chloroplast DNA, complete sequence (0) | reductive pentose-phosphate cycle, photosynthesis, oxidation reduction response to UV, response to wounding, phenylpropanoid metabolic process, response to fungus glycolysis |
| 215          | F: CCAGAGCTCTGATCGAATC   | 56      | 2             | 411               | 1486             | 2                  | *Abies firma* clone 1 4-coumarate-CoA ligase (4CL) gene, partial cds (0) | *Abies firma* chloroplast DNA, complete sequence (0) |
| 241          | F: AAGCTCCGTGTAATAGCTCGG | 56      | 3             | 256               | 1370             | 2                  | *Arabidopsis thaliana* fructose-bisphosphate aldolase, class I (FBA1) mRNA, complete cds (1E-125) | *Arabidopsis thaliana* fructose-bisphosphate aldolase, class I (FBA1) mRNA, complete cds (1E-125) |
| 323          | F: AAGCAGGCTCTGTAAATC    | 53      | 2             | 278               | 804              | 1                  | *A. thaliana* plasma membrane H+-ATPase gene, partial cds (1E-90) | *A. thaliana* plasma membrane H+-ATPase gene, partial cds (1E-90) |
| 1362         | F: GCCAGGCTCTGATAGCTCGG  | 59      | 0             | —                 | 871              | 1                  | *Ricinus communis* processing-spooling factor, putative, mRNA (0.0) | *Ricinus communis* processing-spooling factor, putative, mRNA (0.0) |
| 1704         | F: CAACTCTCCGAGAGCAGC    | 52      | 2             | 327               | 858              | 2                  | *Pinus taeda* mitogen-activated protein kinase 13 (MAPK13) mRNA, complete cds (2E-84) | *Pinus taeda* mitogen-activated protein kinase 13 (MAPK13) mRNA, complete cds (2E-84) |
| 2387         | F: TAAGTGGCTCAGTTCTCAG   | 61      | 1             | 128               | 624              | 1                  | *Medicago truncatula* Alpha-1,4-galacturonosyltransferase (MTR_7g075840) mRNA, complete cds (8E-99) | *Medicago truncatula* Alpha-1,4-galacturonosyltransferase (MTR_7g075840) mRNA, complete cds (8E-99) |
| 2565         | F: GTGTCTGGAGAGGAAACAG   | 58      | 0             | —                 | 432              | 1                  | PREDICTED: *Vitis vinifera* adenosinehomocysteinase-like, transcript variant 1 (LOC100253872), mRNA (1E-109) | PREDICTED: *Vitis vinifera* adenosinehomocysteinase-like, transcript variant 1 (LOC100253872), mRNA (1E-109) |
| 2774         | F: GTTACAGGAGATATGCCCTG  | 55      | 0             | —                 | 502              | 2                  | *Citrus sinensis* lectin mRNA, complete cds (5E-32) | *Citrus sinensis* lectin mRNA, complete cds (5E-32) |
| 2937         | F: TGAGCTCTGTGAATGCGG    | 58      | 0             | —                 | 622              | 2                  | *Solanum tuberosum* clone 15D06 fructose-bisphosphate aldolase-like mRNA, complete cds (5E-120) | *Solanum tuberosum* clone 15D06 fructose-bisphosphate aldolase-like mRNA, complete cds (5E-120) |
| 2986         | F: TCGTCTGGAGAGGCACCTAC  | 57      | 0             | —                 | 355              | 1                  | *Populus trichocarpa* aragonate/prephenate dehydratase (PDT1), mRNA (1E-52) | *Populus trichocarpa* aragonate/prephenate dehydratase (PDT1), mRNA (1E-52) |
| 3421         | F: CTACCTCTCCGAGAGACG    | 55      | 0             | —                 | 324              | 2                  | *Picea sitchensis* isolate CR201 phenylalanine ammonia lyase-like protein mRNA, partial cds (0.0) | *Picea sitchensis* isolate CR201 phenylalanine ammonia lyase-like protein mRNA, partial cds (0.0) |
| 3593         | F: AGGACCTGAATACCTGCT    | 56      | 0             | —                 | 337              | 2                  | *Abies firma* chloroplast, partial genome (6E-170) | *Abies firma* chloroplast, partial genome (6E-170) |
| 3689         | F: CAGTCTGGCTGAGCTGAG    | 58      | 0             | —                 | 619              | 2                  | *Pseudospora menziesii* var. menziesii haplootype Pm-TBE_412m2 thiazole biosynthetic enzyme (TBE) gene, complete cds (0.0) | *Pseudospora menziesii* var. menziesii haplootype Pm-TBE_412m2 thiazole biosynthetic enzyme (TBE) gene, complete cds (0.0) |
| 3918         | F: TTTCAAGGTCTTCTCAAGGT  | 55      | 0             | —                 | 502              | 2                  | *Pinus taeda* cellulose synthase catalytic subunit (CosA1) mRNA, complete cds (0.0) | *Pinus taeda* cellulose synthase catalytic subunit (CosA1) mRNA, complete cds (0.0) |
| 3942         | F: GTATGATACCGGTGAGGCA   | 55      | 0             | —                 | 273              | 2                  | *Ricinus communis* proteasome subunit alpha type, putative, mRNA (8E-48) | *Ricinus communis* proteasome subunit alpha type, putative, mRNA (8E-48) |
| 3981         | F: GCAGGCGTCGTCCTCCAG    | 54      | 0             | —                 | 918              | 1                  | *Pinus radiata* UDP-glucose dehydrogenase gene, partial sequence (0.0) | *Pinus radiata* UDP-glucose dehydrogenase gene, partial sequence (0.0) |
| 4103         | F: ATGCCACCTTACGAGAAC    | 57      | 0             | —                 | 698              | 1                  | *Pinus pinaster* mRNA for S-adenosylmethionine synthase 1 (samS1 gene) (0.0) | *Pinus pinaster* mRNA for S-adenosylmethionine synthase 1 (samS1 gene) (0.0) |
| 4492         | F: TGGTGGCAGCTAGAGTGA    | 57      | 0             | —                 | 698              | 1                  | *Medicago truncatula* magnesium-chelatase subunit chll (MTR_2g015390) mRNA, complete cds (4E-160) | *Medicago truncatula* magnesium-chelatase subunit chll (MTR_2g015390) mRNA, complete cds (4E-160) |
Table 2. Continued.

| Gene Locus ID | Primer sequences (5′–3′) | $T_c$ (°C) | No. of introns | Intron length (nt) | Total length (nt) | Annotation approach | BLASTn of gene sequences against nr nucleotide database (E-value) | GO-ID biological process |
|---------------|--------------------------|------------|----------------|-------------------|------------------|----------------------|--------------------------------------------------|-------------------------|
| 4921          | F: GAAGGTGCCTTAATACGCTT | 56         | 0              | —                 | 664              | 2                    | PREDICTED: Glycine max proteasome subunit alpha type-4-like, transcript variant 1 (LOC100786457), mRNA (2E-147) | response to cadmium ion, ubiquitin-dependent protein catabolic process |
|               | R: AGCTTAGAGAGAGACTCTG |            |                |                   |                  |                      |                                                                 |                         |
| 5004          | F: CAGATGTAAGCCATACCTG    | 57         | 0              | 461               | 523              |                      | Picea sitchensis isolate VD401 magnesium chelatase H-like protein mRNA, partial cds (0) | chlorophyll biosynthetic process |
|               | R: AGCTTAGACAGAGACTCGG   |            |                |                   |                  |                      |                                                                 |                         |
| 5823          | F: TGGTTGATATACGTCGG     | 57         | 0              | 293               | 2                |                      | Picea sitchensis isolate VD401 phytochrome A-like protein mRNA, partial cds (0) | regulation of transcription, photomorphogenesis, tryptophan biosynthetic process |
|               | R: CTAGACTCTGGATAGAGACC  |            |                |                   |                  |                      |                                                                 |                         |
| 5945          | F: CTGTACCTGAGATCTTC     | 55         | 0              | 339               | 1                |                      | P. abies (L.) Karst. Lhcb1*2-2 mRNA for light-harvesting chlorophyll a/b-binding protein (0.0) | photosynthesis, light harvesting, protein-chromophore linkage |
|               | R: CAACCTCTGAATATAGCTGC |            |                |                   |                  |                      |                                                                 |                         |
| 6119          | F: AGAGATGTGAGCCATATTAG  | 57         | 0              | 567               | 1                |                      | Picea mariana pyruvate dehydrogenase E1 beta subunit (Sb68) mRNA, partial cds (0.0) | pollen tube development, oxidation reduction |
|               | R: CAATCAATGTATCTTACAGC  |            |                |                   |                  |                      |                                                                 |                         |
| 6594          | F: TGGCTTTTAATCTGGAGG     | 58         | 1              | 348               | 712              | 1                    | P. communis phosphatidylinositol 4-kinase, putative, mRNA (5E-51) | phosphoinositide biosynthetic process, phosphoinositide phosphorylation, signal transduction, phosphoinositide-mediated signaling |
|               | R: GATAGTCGATATGAGGACG   |            |                |                   |                  |                      |                                                                 |                         |
| 6757          | F: TATACATTCCGGTTCAAGCTG | 58         | 5              | 177               | 939              | 1                    | Arabidopsis thaliana ribonucleoside-diphosphate reductase subunit MI (RRN1) mRNA, complete cds (1E-39) | dTDP-rhamnose biosynthetic process, D-xylose metabolic process |
|               | R: ACCATCCCAAGAGAGACTCT  |            |                |                   |                  |                      |                                                                 |                         |
| 7098          | F: GTTTTCTGTCTCTGATCTCC  | 55         | 0              | —                 | 782              | 1                    | Arabidopsis thaliana UDP-glucuronic acid decarboxylase (AUD1) mRNA, complete cds (1E-153) | transport, ATP synthesis coupled electron transport |
|               | R: GTTTCTGTTGATCTCTCC    |            |                |                   |                  |                      |                                                                 |                         |
| 7208          | F: GTTACATCGAGAGGATCTTG  | 54         | 0              | —                 | 326              | 1                    | Pinus thunbergii NADH dehydrogenase subunit 5 (nad5) gene, partial cds; mitochondrial (0) | terpenoid biosynthetic process, thiamin biosynthetic process |
|               | R: AATGTCCTGAAGAGACTGGC  |            |                |                   |                  |                      |                                                                 |                         |
| 7324          | F: ATCCCAGTGAATGGCTGAC   | 57         | 0              | —                 | 471              | 1                    | Picea abies 1-deoxy-D-xylulose 5-phosphate synthase type I (DXS1) mRNA, complete cds (0) | response to water deprivation, water transport, transmembrane transport, response to salt stress |
|               | R: TCTTGCCTGATTGGCTAAC   |            |                |                   |                  |                      |                                                                 |                         |
| 8248          | F: CAAGATTCCCGAGTTAACAG  | 57         | 1              | 610               | 1128             | 2                    | P. abies mRNA for perin Mip1 (3E-154) | phenylpropanoid metabolic process, phenylalanine ammonia-lyase-like protein mRNA, partial cds (5E-162) |
|               | R: ACAAGGTCACCCCAATCTCC  |            |                |                   |                  |                      |                                                                 |                         |
| 8583          | F: TCTTCTAATCCATTGACGAT  | 56         | 0              | 393               | 926              | 2                    | Picea sitchensis isolate VA301 phenylalanine ammonia-lyase-like protein mRNA, partial cds (5E-162) | photosynthesis, light harvesting |
|               | R: CCACATCAGGAGCTGGAGG   |            |                |                   |                  |                      |                                                                 |                         |
| 8855          | F: TATTGCTCTGTGCTGGGATCC | 58         | 2              | 275               | 792              | 2                    | P. sylvestris Lhca4*1-2 mRNA encoding Lhca4 protein (type 4 protein of light-harvesting complex of photosystem I) (partial) (7E-179) | cell redox homeostasis, oxidation reduction |
|               | R: CTGCACTATGGGTCTGAAC   |            |                |                   |                  |                      |                                                                 |                         |
| 9366          | F: AGTGAAAGCCACCAACTCTG  | 53         | 0              | 598               | 561              | 1                    | Tamarix hispida peroxiredoxin1 in 2 (Prx2) mRNA, complete cds (5E-139) | regulation of transcription, DNA-dependent |
|               | R: TCTTCTGCTGCTGGGATCC   |            |                |                   |                  |                      |                                                                 |                         |
| 9512          | F: GAACTGGATGAGTCCAGCA   | 59         | 1              | 99                | 415              | 2                    | Cycas revoluta class III HD-Zip protein HDZ32 gene, partial cds (4E-52) | cell redox homeostasis, oxidation reduction |
|               | R: TACGAGATCTGTGCTGAGG   |            |                |                   |                  |                      |                                                                 |                         |
| 9652          | F: TGCAAGAAGATTCAAGCAGG  | 58         | 2              | 418               | 913              | 2                    | Pinus pinaster COBRA-like protein gene, partial cds (0) | malate metabolic process, oxidation reduction, tricarboxylic acid cycle, glycolysis |
|               | R: CCCATCGGTTTTAATGCTTG  |            |                |                   |                  |                      |                                                                 |                         |
| 11301         | F: GTGTTTCTGTGCTGGGAGC   | 54         | 0              | 490               | 490              | 2                    | Pinus pinaster mRNA for malate dehydrogenase (MDH gene) (0.0) | malate metabolic process, oxidation reduction, tricarboxylic acid cycle, glycolysis |
|               | R: GGCAATACATGCTCTCCCTC  |            |                |                   |                  |                      |                                                                 |                         |
| 13329         | F: GATACGTGCGCAAAGACACTG | 57         | 0              | 350               | 350              | 2                    | PREDICTED: Glycine max probable thiamine biosynthetic enzyme 1-like (LOC100789909), mRNA (7E-87) | regulation of transcription, DNA-dependent |
|               | R: CCTTGCAGCTTCCAGAAGG   |            |                |                   |                  |                      |                                                                 |                         |
| 13536         | F: CTCATTGTTCTGCTCATCC   | 56         | 0              | 369               | 369              | 2                    | Pinus thunbergii PrANTL1 mRNA for AINTEGUMENTA-like protein, complete cds (0) | regulation of transcription, DNA-dependent |
|               | R: TCCTCAAATCCAACTATGAGC |            |                |                   |                  |                      |                                                                 |                         |
| Gene Locus ID | Primer sequences (5′–3′) | \( T_r \) (°C) | No. of introns | Total length (nt) | Annotation approach | BLASTn of gene sequences against nr nucleotide database (E-value) | GO-ID biological process |
|--------------|--------------------------|----------------|---------------|-----------------|-------------------|-------------------------------------------------|--------------------------|
| 14455        | F: GAACAGACGTGGACTTCTGCC | 56             | 0             | 834             | 2                 | *Pinus taeda* mRNA for alpha-1, 6-xylosyltransferase (x34.1 gene) (0.0) | root hair elongation, xyloglucan biosynthetic process |
|              | R: TTGTGGCTGCAAAAGCAG    |                |               |                 |                   | *Picea abies* mRNA for translation elongation factor-1 alpha, partial (0.0) | translation elongation    |
| 14479        | F: CCACCTCAAGCTTCAAGAGG  | 57             | 0             | 588             | 1                 | *Metasequoia glyptostroboides* fructose-1,6-diphosphate aldolase mRNA, complete cds (2E-74) | pentose-phosphate shunt, response to salt stress, glycolysis, response to cadmium ion |
| 14514        | F: GGTGGCTATCTTCTCAAAGG  | 56             | 0             | 322             | 2                 | *Picea glauca* isolate D8411049-162 cellulose synthase family gene, partial sequence (0.0) | cellulose biosynthetic process, cellular cell wall organization |
| 14585        | F: TTCCTGAACCTCTATCTCCACAG | 57             | 1             | 193             | 1                 | *Vitis vinifera* galacturonosyltransferase 8-like (LOC100258818), mRNA (6E-119) | homogalacturonan biosynthetic process |
| 14887        | F: GGTAGGACCTGCTAACCAGAGG | 53             | 0             | 1156            | 2                 | *Glycine max* elongation factor 2-like (LOC100788357), mRNA (0.0) | l-serine metabolic process, one-carbon metabolic process, glycine metabolic process |
| 15135        | F: TTGGAGGCTCTTTAATGGTG   | 53             | 0             | 657             | 2                 | *Picea glauca* heat shock protein, putative, mRNA (0.0) | oxidation reduction, response to stress, auxin biosynthetic process |
| 15337        | F: TTATATTGATCCTCTTGGAGGCAAGG | 57             | 1             | 232             | 1                 | *Pinus pinaster* phloem alanine ammonia-lyase (pal) mRNA, complete cds (0.0) | phenylpropanoid metabolic process, biosynthetic process, l-phenylalanine catabolic process |
| 15484        | F: TTCGACGGGTTATATGCTTG   | 58             | 0             | 663             | 2                 | *Pinus pinaster* mRNA for cystolic serine hydroxy methyl transferase (chsmt gene) (2E-138) | l-serine metabolic process, one-carbon metabolic process, glycine metabolic process |
| 15727        | F: CACCTGAGGCTTGGAGGAGG   | 58             | 0             | 325             | 2                 | *Pinus pinaster* mRNA for polyubiquitin (3E-116) | auxin biosynthetic process, one-carbon metabolic process |
| 15811        | F: CTGCTGACCTCTTGGAGAGG   | 57             | 0             | 438             | 2                 | *Populus EST from severe drought-stressed opposite wood (0.000000003) | lipid transport |
| 15969        | F: GGAACCTCTCTGTCTACCTTG  | 57             | 0             | 990             | 1                 | *Pinus contorta* S-adenosylmethionine synthetase (sams2) mRNA, complete cds (0.0) | auxin biosynthetic process, one-carbon metabolic process |
| 16246        | F: CATCTTGCGCTTGATGCTTG   | 57             | 3             | 132             | 2                 | *Pseudotsuga menziesii* class III homeodomain-leucine zipper (CHDZ1) gene, complete cds (0.0) | auxin biosynthetic process, one-carbon metabolic process |
| 16727        | F: GGTCTGACCTGAGGAAAGT    | 58             | 0             | 331             | 2                 | *Ricinus communis* chloroplast DNA, complete sequence (0) | transport, photosynthesis, electron transport chain |
| 16883        | F: CTCTCGACGGGGGTGAGAGG   | 58             | 0             | 710             | 1                 | *Pinus contorta* S-adenosylmethionine synthetase (sams2) mRNA, complete cds (0.0) | auxin biosynthetic process, one-carbon metabolic process |
| 16979        | F: CTCCTCTGGCTGCTTGAATGGG | 55             | 0             | 535             | 1                 | *Keteleeria davidi ana* chloroplast DNA, complete sequence (0) | auxin biosynthetic process |
| 17340        | F: CTGTGTATTCTCTGGCTCTTG  | 54             | 0             | 281             | 2                 | *Abies firma* chloroplast, partial genome (0) | auxin biosynthetic process |
| 17637        | F: TGCCTGGAAGGTATGTTCTTG  | 56             | 0             | 424             | 1                 | *Ricinus communis* transference, transferring glycosyl groups, putative, mRNA (2E-69) | auxin biosynthetic process |
| 17975        | F: CAAACATTGCTGGAAGGCTC   | 56             | 2             | 547             | 1                 | *Ricinus communis* cysteine synthase, putative, mRNA (1E-65) | cysteine biosynthetic process from serine biosynthetic pathway |
| 18135        | F: GAGAGTCTCTGATAAGATCC   | 55             | 1             | 132             | 2                 | *Picea abies* mRNA for photosynthetic chlorophyll II A-B binding protein, (pPA0001 gene) (0) | photosynthesis, light harvesting in photosystem 1 |
| 18444        | F: AAGAGGGCGAAAATATAGTG   | 54             | 0             | 313             | 1                 | *P. sylvestris* mRNA for polyubiquitin (3E-116) | oxidation reduction |
| 18599        | F: GGATCTCAATGGTATGTTCTTG | 55             | 0             | 678             | 1                 | *S. tuberosum* mRNA for NADH dehydrogenase, NADH-binding subunit (complex I) (0.0) | regulation of transcription |
| 18680        | F: TTCGAGAGGCTTATTCACCT   | 55             | 1             | 214             | 2                 | *Picea glauca* isolate D761009-28 myb family protein gene, partial sequence (1E-140) | regulation of transcription |
| Gene | Locus ID | Primer sequences (5’–3’) | T_c (°C) | No. of introns | Intron length (nt) | Total length (nt) | Annotation approach | BLASTn of gene sequences against nr nucleotide database (E-value) | GO-ID biological process |
|------|----------|--------------------------|----------|--------------|--------------------|------------------|-------------------|---------------------------------------------------------------|--------------------------|
| 19005 | F: GGAGATTGACACGAGGAGG | 56 | 0 | 368 | 1 | Abies firma chloroplast, partial genome (0) | auxin biosynthetic process |
| 19173 | F: AGAAGCAACCATTGCCATCAGCG | 55 | 0 | 343 | 2 | Ricinus communis proteasome subunit alpha type, putative, mRNA (2E-84) | defense response to bacterium, ubiquitin-dependent protein catabolic process, response to zinc ion |
| 19540 | F: ACCATGGTTGCGATCCGG | 55 | 0 | 634 | 1 | Cedrus deodara chloroplast DNA, complete sequence (0) | plasma membrane ATP synthesis coupled proton transport, auxin biosynthetic process RNA processing, chlorophyll biosynthetic process |
| 20156 | F: AATGATCCATCAGCTTC | 55 | 0 | 386 | 1 | Picea sitchensis isolate VD401 magnesium chelatase H-like protein mRNA, partial cds (3E-110) | root hair elongation, cellulose biosynthetic process, response to cold, cellular cell wall organization, plant-type cell wall biogenesis cell wall modification, regulation of transcription |
| 20318 | F: ACAGTTGCTAATGACCTGAG | 55 | 0 | 356 | 1 | PREDICTED: Glycine max cellulose synthase-like protein D3-like (LOC100783983), mRNA (6E-69) | RNA processing, chlorophyll biosynthetic process, response to cold, cellular cell wall organization, plant-type cell wall biogenesis cell wall modification, regulation of transcription |
| 20694 | F: TCGGACACATGGAAGAGGAGG | 56 | 0 | 346 | 1 | PREDICTED: Vitis vinifera zinc finger CCCH domain-containing protein 49-like (LOC100259323), mRNA (6E-43) | auxin biosynthetic process, oxidation reduction, chlorophyll biosynthetic process, photosynthesis |
| 21136 | F: AGAAGCGTTGATTCAATTGC | 57 | 1 | 229 | 535 | P. taeda gene for protochlorophylide reductase (3E-168) | response to hypoxia, sucrose biosynthetic process, regulation of apoptosis, transcription, DNA-dependent metabolic process |
| 21165 | F: ATGCCATGGTCTCTTATGACG | 57 | 2 | 204 | 644 | PREDICTED: Glycine max pre-mRNA-processing-splicing factor 8-like (LOC100804026), mRNA (4E-46) | cell wall macromolecule catabolic process, chitin catabolic process |
| 21173 | F: ACAGTTGCTAATGACCTGAG | 56 | 0 | 333 | 2 | Picea sitchensis isolate VA100 basic endochitinase-like protein mRNA, partial cds (1E-138) | RNA processing |
| 21890 | F: GAAGAGCTTCACGGGAAGGAGG | 55 | 1 | 358 | 607 | PREDICTED: Sisymbrium officinale domain-containing protein-like protein mRNA, partial cds (2E-116) | auxin biosynthetic process, oxidation reduction, photosynthesis, dark reaction |
| 21957 | F: AACACTTCTCACAGTTCTC | 54 | 0 | 292 | 2 | Abies firma chloroplast, partial genome (2E-157) | auxin biosynthetic process, oxidophylic biosynthetic process, oxidation reduction, photosynthesis, dark reaction |
| 22174 | F: AGGAGTCCGGTTGCAATGCC | 55 | 0 | 334 | 1 | Abies firma chloroplast, partial genome (6E-157) | auxin biosynthetic process, oxidation reduction, photosynthesis, dark reaction |
| 23660 | F: AGGAGTCCGGTTGCAATGCC | 58 | 1 | 781 | 1232 | P. abies mRNA for porin Mip1 (6E-157) | response to water deprivation, water transport, transmembrane transport |
| 23809 | F: ATGCCACTTACAGCTACAGCA | 55 | 0 | 1058 | 2 | Abies firma chloroplast, partial genome (9E-168) | oxidation reduction, chlorophyll biosynthetic process, photosynthesis, dark reaction |
| 23850 | F: AGAAGCCAGACATGCACTG | 52 | 1 | 449 | 695 | Pseudomonas aeruginosa elongation factor 1-alpha mRNA, complete cds (4E-65) | auxin biosynthetic process, protein acid phosphorylation, conjugation, mitosis, cell division |
| 23982 | F: TGAGACTTCTCTCCAGGAAAAGG | 57 | 2 | 586 | 921 | Pisum sativum non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (gapN) gene, complete cds (4E-33) | metabolic process |
| 24523 | F: TTACACGTGGCTACTCACCA | 58 | 0 | 449 | 2 | Ginkgo biloba catalase mRNA, complete cds (3E-98) | hydrogen peroxide catabolic process, oxidation reduction |
| 24699 | F: AAGATGAACGCTTCTTGCTGA | 56 | 0 | 262 | 1 | Agaricinidae adenophora heat shock protein 70.58 mRNA, complete cds (2E-81) | auxin biosynthetic process, response to stress |
| 24902 | F: CCGGCTCTGCTTACAGCTACAGCA | 58 | 1 | 240 | 662 | Arabidopsis thaliana ferredoxin–NAPD+ reductase (RFNR2) mRNA, complete cds (3E-54) | electron transport chain |
| 25060 | F: CTGAGACTTTCTGGACATC | 58 | 2 | 163 | 624 | PREDICTED: Glycine max ATP-citrate synthase beta chain protein 1-like (LOC100809904), mRNA (2E-74) | acetyl-CoA biosynthetic process, cellular carbohydrate metabolic process |
| 26089 | F: GATTATGGATCTTCAACCCGGAGA | 55 | 1 | 281 | 1233 | Rosa multiflora elongation factor 1-alpha mRNA, complete cds (0.0) | translational elongation |
## Table 2. Continued.

| Gene Locus ID | Primer sequences (5′-3′) | $T_a$ (°C) | No. of introns | Intron length (nt) | Total length (nt) | Annotation approach | BLASTn of gene sequences against nr nucleotide database (E-value) | GO-ID biological process |
|---------------|--------------------------|------------|----------------|-------------------|-------------------|---------------------|---------------------------------------------------------------|--------------------------|
| 26764         | F: GGGAAATTGGCTCGATCTGG  | 58         | 0              | —                 | 359               | 1                   | *Cucumis sativus* 6-phosphogluconate dehydrogenase (6PGDH) mRNA, complete cds (7E-55) | response to glucose stimulus, response to sucrose stimulus, response to fructose stimulus, response to salt stress, pentose-phosphate shunt, oxidation reduction, response to cadmium ion |
|               | R: GTTCTGCCTAGCAATGCTCTTTC  |            |                |                   |                   |                     |                                                               |                          |
| 27033         | F: TTTACTCCACCCATTAGAGG  | 55         | 0              | —                 | 948               | 2                   | *Medicago truncatula* heat shock protein (MTR_7g024390) mRNA, complete cds (0) | response to virus, auxin biosynthetic process, protein folding, response to heat, response to bacterium, response to cadmium ion, response to high light intensity, response to hydrogen peroxide, protein amino acid phosphorylation |
|               | R: TTCGCAATGATAGGATTTGA  |            |                |                   |                   |                     |                                                               |                          |
| 27963         | F: TAGGCCCATAAGCTAACAACC  | 57         | 0              | —                 | 318               | 1                   | *Keteleeria davidi ana* chloroplast DNA, complete sequence (0) | transcription, DNA-dependent |
|               | R: TCGAATGTTGCCATCTTCCCA  |            |                |                   |                   |                     |                                                               |                          |
| 28203         | F: GTGGACGGAGGATATTCCG  | 56         | 0              | —                 | 315               | 2                   | *Pinus pinaster* mRNA for cytosolic serine hydroxymethyltransferase (cshmt gene) (1E-128) | L-serine metabolic process, one-carbon metabolic process, glycine metabolic process |
|               | R: TCCAGAAGGGGCTGTTAGG  |            |                |                   |                   |                     |                                                               |                          |
| 28456         | F: GTTTCGAGAGCTGATCTCCC  | 58         | 0              | —                 | 853               | 1                   | *Ricinus communis* oligosaccharyl transferase, putative, mRNA (7E-69) | protein amino acid glycosylation |
|               | R: AGCTGTCGTTGATATGCTG  |            |                |                   |                   |                     |                                                               |                          |
| 28639         | F: GTGAAATAGGGGACCCGTT  | 57         | 0              | —                 | 438               | 2                   | *Abies fabri* 26S ribosomal RNA gene, partial sequence (0) | protein | |
|               | R: ATACGAAGAGGGGACATCGA  |            |                |                   |                   |                     |                                                               |                          |
| 29437         | F: CTTCCGAGCTGATCTCCGT  | 56         | 0              | —                 | 403               | 2                   | *Populus trichocarpa* argonaute protein group (AGO911), mRNA (2E-99) | protein | |
|               | R: TCAACTGAAACCCTTAGTTC  |            |                |                   |                   |                     |                                                               |                          |

**Note:** — = not available; $T_a$ = annealing temperature.

*Values are based on the sequence of one sample randomly chosen from a sample set of 80 trees from a population at Mont Ventoux (France).