Gene 526 (2013) 251–258

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Transcriptomic analysis of purple leaf determination in birch

Lin Lin [a], Huaizhi Mu [a,b], Jing Jiang [a], Guifeng Liu [a,*]

[a] State Key Laboratory of Forest Genetics and Tree Breeding, Northeast Forestry University, 26, Hexing Road, Harbin 150040, China
[b] Forestry College, Beihua University, 3999, Binjiang East Road, Jilin 132013, China

ARTICLE INFO

Article history:
Accepted 7 May 2013
Available online 31 May 2013

Keywords:
Betula pendula
Leaf color
Anthocyanin
Transcriptome

ABSTRACT

'Purple Rain', a purple cultivar of *Betula pendula*, has dark purple leaves throughout the vegetative period. In this study, *B. pendula* 'Purple Rain' was found to have a higher anthocyanidin level compared with *B. pendula*, Transcriptome analysis revealed numerous changes in gene expression that could be attributed to color change, including the upregulation of 2467 unigenes and the downregulation of 2299 unigenes in 'Purple Rain'. Furthermore, anthocyanidin synthesis and transcriptional regulation were altered in 'Purple Rain', which may have contributed to phenotypic changes. These results provide unique molecular insights into the biochemical pathways and regulatory networks that function in a purple variety of *B. pendula*.

1. Introduction

Purple cultivars commonly exist in the plant kingdom. This trait is often used to distinguish among different plant cultivars (Anjani et al., 2007). Consequently, much research has focused on the formation of purple pigmentation in plants. Prior to advancements in biotechnology, researchers mainly focused on phenotypic and physiological variation between purple and green cultivars (Giusti et al., 1999; Phippen and Simon, 1998; Yoshitama et al., 1994). With the advent of advanced techniques, numerous color-related genes were discovered in model plants, which enhanced our understanding of the formation and characterization of purple pigmentation in plants (Guo et al., 2008; Himi and Noda, 2005; Melz and Thiele, 1990; Mou et al., 1995). Although some ligneous species have purple leaves, such as *Prunus cerasifera* var. atropurpurea, *Berberis thunbergii* var. atropurpure and *Cotinus coggygria* 'Purpureus' (Chen et al., 2008; Hao et al., 2006; Shi et al., 2007), the molecular mechanisms underlying the formation of purple leaves in woody plants has not been addressed due to the long growth periods and complex genomes of these plants.

Members of *Betula pendula* are widely distributed in Eurasia and North America. *B. pendula* 'Purple Rain', an intraspecies variety of *B. pendula*, is a valuable ornamental plant due to its purple leaves (Li et al., 2009a). In 2011, 35 *B. pendula* 'Purple Rain' saplings were obtained through micropropagation. These purple saplings exhibited phenotypic change compared with green saplings, but information concerning transcriptomic changes that occurred as a result of color change in *B. pendula* has been scant. Recently developed high-throughput sequencing technology, i.e., next generation sequencing (NGS), including sequencing employing the Roche/454 Genome Sequencer FLX, the ABI SOLiD System and the Illumina Genome Analyzer, is a powerful and cost-efficient tool for advanced research in many areas, such as de novo transcriptome sequencing of *Betula* (Mu et al., 2012; Wang et al., 2012). In the present study, we compared differences between the transcriptomes of *B. pendula* 'Purple Rain' and *B. pendula* plants using Illumina paired-end sequencing technology. By investigating changes in the expression of genes related to leaf color variation, this study deepens our understanding of variations in birch leaf color and identifies important genes for the genetic engineering of birch trees.

2. Materials and methods

2.1. Plant materials

Leaf buds from *B. pendula* 'Purple Rain' and *B. pendula* were collected from Heilongjiang Forest Botanical Garden (Harbin, China) in November 2010. The buds were disinfected in 10% (*w*/*v*) $Ca(ClO)_2$ for 10 min and rinsed in distilled water for 1 min. The buds were then transferred to WPM solid medium containing 1.0 mg/l 6-Benzyl Aminopurine, 0.5 mg/l Gibberellin A_3 and 0.8% (*w*/*v*) agar for seedling regeneration. Uniformly growing regenerated seedlings were divided to two groups. Shoot tips from one group of regenerated seedlings were stored at −80 °C for RNA extraction, and the other group was transplanted into

Abbreviations: NGS, Next generation sequencing; WPM, Woody plant medium; 6BA, 6-benzylamino purine; GA3, Gibberellin A3; CTAB, Hexadecyltrimethy Ammonium Bromide; KEGG, Kyoto Encyclopedia of Genes and Genomes; COG, Cluster of Orthologous Groups of proteins; GO, Gene ontology; DEUs, Differential expression of unigenes; RPKM, Reads per kb per million reads; FDR, False discovery rate.

* Corresponding author at: State Key Laboratory of Forest Genetics and Tree Breeding, Northeast Forestry University, 150040, Harbin, China. Tel.: +86 451 82190607 11; fax: +86 451 82191627 11.
E-mail address: guifengliu@yahoo.com.cn (G. Liu).

0378-1119/$ – see front matter.
http://dx.doi.org/10.1016/j.gene.2013.05.038

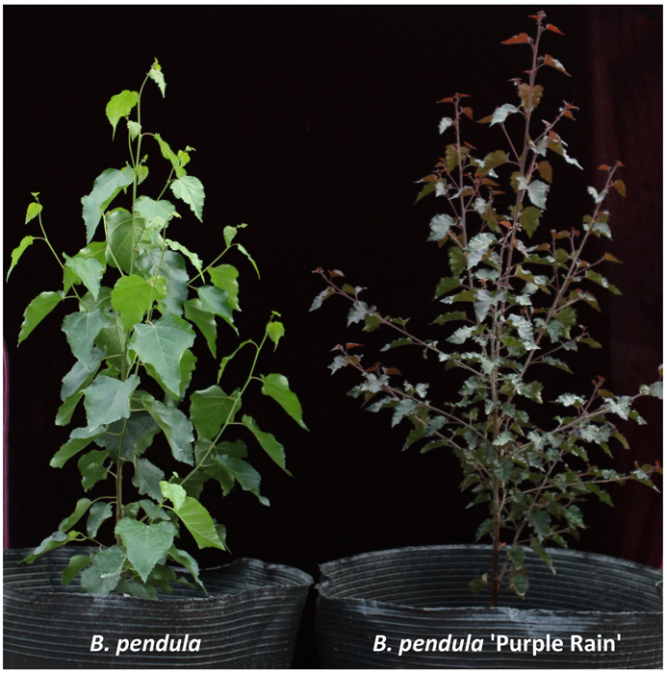

Fig. 1. Saplings of *B. pendula* 'Purple Rain' and *B. pendula*.

plastic pots in an intensive seed orchard in Harbin. In 2011, 35 *B. pendula* 'Purple Rain' saplings and 35 *B. pendula* saplings were obtained (Fig. 1).

2.2. Anthocyanidin measurement

Anthocyanidin was extracted from the leaves with methanol:HCl (99:1, *v*/*v*) at 4 °C in the dark. After leaf bleaching, the spectrum of anthocyanin in the extracts was scanned separately with a Vis–UV spectrophotometer (TU-1901, PERSEE, China). Total anthocyanin quantification was performed as described by Pirie and Mullins (1976).

2.3. RNA extraction, library construction and RNA-Seq

Total RNA was extracted by the CTAB method and treated with DNaseI (Promega, USA) to remove contaminating DNA. Enrichment of mRNA, fragment interruption, addition of adapters, size selection, PCR amplification and RNA-Seq were performed by staff at the Beijing Genome Institute (BGI; Shenzhen, China). The mRNA was isolated with oligo (dT) cellulose and broken into short fragments by adding fragmentation buffer. Using these short fragments as templates, first-strand cDNA and second-strand cDNA were synthesized. Sequencing adapters were ligated to the short fragments after purifying them using a QiaQuick PCR Extraction Kit; the sequencing adapters were used to distinguish among different sequencing samples. Fragments 200 ± 25 bp in length were then separated by agarose gel electrophoresis and selected for use as sequencing templates in PCR amplification. All experiments were conducted with three biological replicates for each variety. Finally, the six libraries were sequenced using Illumina HiSeq™ 2000, and the remaining RNA was used for real-time quantitative RT-PCR verification.

2.4. De novo assembly and functional annotation

To obtain accurate data, 43 transcriptomic libraries from *B. pendula*, *B. pendula* 'Dalecarlica', *B. pendula* 'Purple Rain', *B. platyphylla* and *B. halophila* were used for de novo assembly. The raw reads were first filtered by removing the adapter sequences and low-quality sequences, which included reads with N percentages (i.e., the percentage of nucleotides in a read that could not be sequenced) of over 5% and sequences containing more than 20% nucleotides with a *Q*-value ≤10. The *Q*-value represents the sequencing quality of related nucleotides. Clean reads were used in de novo assembly and read-mapping to the transcriptome. RNA-Seq data were de novo assembled using the Trinity assembling program (Grabherr et al., 2011). The short reads were first assembled into longer but gapless contigs, and the reads were then mapped back to contigs, taking the distance of paired-end reads as frame. The contigs were connected to access the sequence that could not be extended on either end, and the sequence of the unigene was then produced. Next, unigenes were further spliced and assembled to obtain maximum length nonredundant unigenes using TGICL clustering software (Pertea et al., 2003), with a minimum overlap length of 100 bp. After clustering, the unigenes were divided into two classes, i.e., clusters, with the prefix CL and singletons, with the prefix Unigene. Finally, unigenes were aligned with the NCBI Nr, Swiss-Prot, KEGG and COG protein databases using Blastx with an *E*-value $\leq 10^{-5}$. The best results were used to further determine

Table 1
Primers used for real-time quantitative PCR verification.

Gene-ID	Forward primer (5′–3′)	Reverse primer (5′–3′)	Product size (bp)
18S rRNA	ATCTTGGGTTGGGCAGATCG	CATTACTCCGATCCCGAAGG	223
α-Tublin	GCACTGGCCTCCAAGGAT	TGGGTCGCTCAATGTCAAGG	282
Unigene47948_All	CCCTACGACGGGAAAACAGT	TCCACCTCAAGCGTGCTAAT	224
Unigene79074_All	TTGTTTCTATGGCTGCTACGT	TTGGCTCAGAGCTTTGAGTATA	283
CL8897.Contig2_All	GTACTCAGCCTTCACCTCCC	TCCAACTAATAACCGATTCACA	208
CL3482.Contig3_All	CCTCCAAGCCATCGTCAAAG	CCAAACGGGATAAACTCAAAAC	255
CL4068.Contig5_All	TGTGGTTGACGGTTGGTTAG	GATTCCTTTAGTTGCCTTTCC	297
CL3134.Contig5_All	TTACATAGCGTCATGGCTGGTG	TCCCTTCAATGCTGGGTCAA	274
Unigene68713_All	AACTTCTGTCGGTAAACCTGG	TGTTGTCTTTGCGGACTTGA	213
CL198.Contig1_All	CAAAAGTCTCGGTCAGTCGC	TCATAAACAGAGTTCGCTCCA	272
Unigene17213_All	GATTCCAACCACGAAACCTAA	ATGCCTTCCTGAACATACGG	213
CL11232.Contig2_All	TCGTGCGGTTCATACATTTC	TGTTTTACTACAGATTCCCTTGC	297
Unigene16116_All	TGTCCACAGAATCGAGGAGTG	TGTCTACCCGAAGATCATCAAG	231
Unigene23470_All	CAAAACAATCAAGATGTGGAAGA	CAATAGACCAAGTGCTGAGTAGAA	241
Unigene76491_All	ACGACGGTTAGTAGCAGAGCA	TGTAAACAACAACTGGCAAAAG	218
CL257.Contig1_All	GTCGGATGTGATTCCTTGTCT	TGTATTTGCCTTGATGATGCT	244
CL8045.Contig3_All	GCTCTTTCACCCTAATGGACT	CAGTTGTGGTTGCCTCATCTC	291
CL10102.Contig1_All	TTGCCAGTTTAATTCCCATG	TGCCACCATTCAGGTTATTT	299
CL9627.Contig1_All	ATCTCCCGTGAATCCTCCTC	GGCCCTCCAATGATACAATC	225
Unigene56380_All	CGGAAGAACAGACAACGAGA	TTTGGAGATGGAGGAGGATC	247
Unigene65390_All	GAGACAAATGGGAAGAAGCC	TGTTACCAAGAAGATTATGGAGTC	269
CL10405.Contig2_All	CACCCAGTCACAATGAACCC	ACCGTCGTCAATAATCTTCTCC	254