

RESEARCH ARTICLE

# Development of Genetic Markers in *Eucalyptus* Species by Target Enrichment and Exome Sequencing

Modhumita Ghosh Dasgupta<sup>1\*</sup>, Veeramuthu Dharanishanthi<sup>1</sup>, Ishangi Agarwal<sup>2</sup>, Konstantin V. Krutovsky<sup>3,4,5,6</sup>

**1** Division of Plant Biotechnology, Institute of Forest Genetics and Tree Breeding, P.B. No. 1061, R.S. Puram, Coimbatore–641002, India, **2** Genotypic Technology Private Limited, #2/13, Balaji Complex, Poojari Layout, 80, Feet Road, R. M. V. 2nd Stage, Bangalore-560094, India, **3** Department of Forest Genetics and Forest Tree Breeding, Bösgen Institute, Georg August University of Göttingen, Bösgenweg 2, D-37077 Göttingen, Germany, **4** Department of Ecosystem Science and Management, Texas A&M University, 2138 TAMU, College Station, TX 77843-2138, United States of America, **5** N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119333, Russia, **6** Genome Research and Education Center, Siberian Federal University, 50a/2 Akademgorodok, Krasnoyarsk 660036, Russia

\* [modhumitaghosh@hotmail.com](mailto:modhumitaghosh@hotmail.com)



**OPEN ACCESS**

**Citation:** Dasgupta MG, Dharanishanthi V, Agarwal I, Krutovsky KV (2015) Development of Genetic Markers in *Eucalyptus* Species by Target Enrichment and Exome Sequencing. PLoS ONE 10(1): e0116528. doi:10.1371/journal.pone.0116528

**Academic Editor:** Swarup Kumar Parida, National Institute of Plant Genome Research (NIPGR), INDIA

**Received:** July 31, 2014

**Accepted:** December 8, 2014

**Published:** January 20, 2015

**Copyright:** © 2015 Dasgupta et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All aligned sequence files are deposited in NCBI Short Read Archive with the accession number SRP045253 for *E. tereticornis* (SRX747331), *E. camaldulensis* (SRX669390) and *E. grandis* (SRX747330).

**Funding:** The funding for the research work was provided to MDG by Department of Biotechnology, Government of India under the DBT-CREST Awardship with grant number BT/IN/CREST-Awards/15/MDG/2010–11 and under the research project with grant number BT/PR/10055/PBD/16/772/2007. The funding support as research fellowship was provided to VD by Department of Biotechnology, Government of India

## Abstract

The advent of next-generation sequencing has facilitated large-scale discovery, validation and assessment of genetic markers for high density genotyping. The present study was undertaken to identify markers in genes supposedly related to wood property traits in three *Eucalyptus* species. Ninety four genes involved in xylogenesis were selected for hybridization probe based nuclear genomic DNA target enrichment and exome sequencing. Genomic DNA was isolated from the leaf tissues and used for on-array probe hybridization followed by Illumina sequencing. The raw sequence reads were trimmed and high-quality reads were mapped to the *E. grandis* reference sequence and the presence of single nucleotide variants (SNVs) and insertions/ deletions (InDels) were identified across the three species. The average read coverage was 216X and a total of 2294 SNVs and 479 InDels were discovered in *E. camaldulensis*, 2383 SNVs and 518 InDels in *E. tereticornis*, and 1228 SNVs and 409 InDels in *E. grandis*. Additionally, SNV calling and InDel detection were conducted in pair-wise comparisons of *E. tereticornis* vs. *E. grandis*, *E. camaldulensis* vs. *E. tereticornis* and *E. camaldulensis* vs. *E. grandis*. This study presents an efficient and high throughput method on development of genetic markers for family– based QTL and association analysis in *Eucalyptus*.

## Introduction

The genus *Eucalyptus* belongs to family Myrtaceae and consists of over 700 species [1] that occupy a broad range of environmental conditions. Most of the species are native to Australia and have been introduced to India, France, Chile, Brazil, South Africa and Portugal in the first

under the research project with grant number BT/PR10055/PBD/16/772/2007. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Ms. Ishangi Agarwal employed in Genotypic Technology Private Limited, Bangalore, India was associated with the analysis of the sequence data and has no potential competing interest and financial disclosure. The authors declare that her authorship does not alter their adherence to all PLOS ONE policies on sharing data and materials as detailed online in the journal homepage.

quarter of 1800s [2]. It is one of the most widely planted hardwood crop in the world because of its superior growth, adaptability and wood properties and occupies 20.07 M hectares globally. India ranks second in area under *Eucalyptus* plantation (3.943 M ha) after Brazil (4.259 M ha) [3]. In tropical and subtropical regions, *E. grandis*, *E. urophylla* and their hybrids are highly preferred for pulp production and solid wood, while *E. globulus* is favored in the temperate regions [4]. Six species including *E. camaldulensis*, *E. grandis*, *E. globulus*, *E. pellita*, *E. tereticornis* and *E. urophylla* are reported to be suitable for Indian agro-climatic conditions and widely planted in the subcontinent [5–6].

*Eucalyptus* is a potential out-croser and due to unlimited free natural hybridizations, the populations are highly heterozygous. Hence, extensive studies were conducted to determine genetic diversity at species and population levels using different marker systems [7–16].

Linkage maps in different species of Eucalypts have been widely reported [17–21]. QTL mapping in this genus has been conducted tagging important traits like wood properties, vegetative propagation, response to biotic and abiotic stress, juvenile traits, stem growth, water stress tolerance and frost tolerance [22–27]. QTL studies in *Eucalyptus* species was recently reviewed in detail by Grattapaglia *et al.* [28]. Population based association studies were reported for *E. nitens* and *E. globulus* targeting wood property traits [29–31]. Recently, the first experimental study of Genomic Selection was reported by Resende and co workers [32] in two *Eucalyptus* populations for growth and wood property traits.

The genomic data in *Eucalyptus* species are well-documented and available in public databases, private collections and consortia as EST resources [33–34] and transcriptome resources [16, 35–42]. Several dedicated databases are available for *Eucalyptus* genome research, such as EUCANEXT, *EucalyptusDB*, Eucpresso [38], EUCATOUL, EUCAWOOD [33], EucaCold [34], EucGenIE [43] and Phytozome10.

Subsequently, the *Eucalyptus* genome sequencing project was initiated independently for *E. grandis* at the US Department of Energy Joint Genome Institute, USA and *E. camaldulensis* at Kazusa DNA Research Institute in Japan. Recently, the complete genome sequence of *E. grandis* ('BRASUZ1') was published [44] and the assembled non-redundant chromosome-scale reference (v1.0) was released with 640 Mb (94%) genome coverage organized into 11 pseudomolecules. It was also reported that 34% of the protein-coding genes occur as tandem duplication and 84% share similarity to rosid lineages.

The draft genome sequence of *E. camaldulensis* sequenced in Japan had a total length of 655,922,307 bp of non-redundant genomic sequences consisting of 81,246 scaffolds and 121,194 singlets. These sequences accounted for approximately 92% of the gene-containing regions. A total of 77,121 complete and partial structures of protein-encoding genes were annotated [45]. The database containing the draft sequence can be accessed at <http://www.kazusa.or.jp/eucaly>.

In the last decades several generic DNA markers have been employed for molecular breeding. These markers are usually effective but their development is labor-intensive and time consuming. However, with the advent of 'next generation' sequencing technologies, a paradigm shift has occurred in DNA sequencing approach, resulting in high throughput and cost effective sequencing methods [46–47]. Nevertheless, sequencing of large number of genomes is still not feasible due to the substantial cost, time, management and storage of the enormous informatics data. Hence, considerable effort has been directed towards sequencing of genome sub-regions by 'target enrichment' methods. Re-sequencing of these enriched genomic regions is time and cost effective and the data analysis is less complex [48].

In the present study, we conducted target enrichment of exomes for 94 genes involved in xylogenesis and re-sequenced them in three *Eucalyptus* species, which were used in developing mapping pedigrees. Presence of SNVs and InDels across different species in pair-wise

comparisons and in comparison to the *E. grandis* reference genome was documented. This study presents an efficient and high throughput method on development of genetic markers for family – based QTL and Association analysis in *Eucalyptus*.

## Materials and Methods

### Plant Material and DNA Isolation

Three genotypes from *Eucalyptus camaldulensis*, *E. tereticornis* and *E. grandis* were selected for target enrichment. *E. camaldulensis* (Ec111) belonging to Kennedy River Provenance from Queensland, Australia is a selection from the Provenance Resource Stand, Pudukkottai, Tamil Nadu, India while *E. tereticornis* (Et86) is a selection from Seed Production Area, Pudukkottai, Tamil Nadu, India. *E. grandis* (Eg9) is a selection from the Lorne provenance trial at Hossammund, Ootacamund, Tamil Nadu, India. These genotypes were used as parents for development of mapping populations targeting wood property traits.

The leaf tissues from the three genotypes were harvested and immediately frozen at  $-80^{\circ}\text{C}$ . Genomic DNA was isolated from the leaf tissues using the GenElute Plant Genomic DNA isolation kit (Sigma Aldrich, USA) and quantified using NanoDrop ND1000 spectrophotometer (Thermo Scientific, USA).

### Selection of Genes and Probe Design for Sequence Capture Array

Genes involved in different steps of secondary xylem formation including cell division, cell expansion, cell wall thickening, cell wall proteins, lignin biosynthesis and programmed cell death in *Arabidopsis*, *Populus*, *Zinnia* and *Eucalyptus* spp. were short-listed from literature and 94 genes were selected for target enrichment and re-sequencing. Their respective gene orthologs were downloaded from *E. grandis* genome database hosted by Phytozome portal (<http://www.phytozome.net/cgi-bin/gbrowse/Eucalyptus>). The sequences were functionally annotated and their position in chromosome, protein domains, biological pathways and gene ontology were defined based on the recent assembly of *E. grandis* using Phytozome v10 [44].

Hundred and twenty bp long hybridization probes (“baits”) were designed with 1bp tiling using SureSelect eArray software (Agilent Technologies, Santa Clara, California, USA) targeting exons and UTRs in 94 genes. A total of 169,700 baits were designed to capture the exons and UTRs in the three species. Using this design, a customized array was synthesized at Agilent Technologies.

### Library Preparation, Target Enrichment and Validation

Ten micrograms of DNA from each sample in 100  $\mu\text{l}$  of nuclease free water were sonicated to fragment DNA to size range of 100 to 500 bp. The size distribution was checked on the Agilent 2100 Bioanalyzer, and the DNA was cleaned using the Agencourt AMPure XP SPRI beads (Beckman Coulter, Australia). The libraries for each sample were prepared using the Illumina TruSeq DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA). The sheared DNA was subjected to a series of enzymatic reactions that repair frayed ends, phosphorylated the fragments, added a single nucleotide overhang to code the libraries and ligated adaptors using manufacturer’s protocol for the Illumina TruSeq DNA sample preparation kit. Subsequently, PCR enrichment (10 cycles) was performed to amplify the library. The three barcoded libraries were pooled in equimolar amounts and approximately 20mg of DNA was hybridized on the Agilent 244Kmicroarray (AMADID: EA560-037734) following manufacturers’ protocol. The hybridization was carried out at  $65^{\circ}\text{C}$  for 65 hrs as described by Hodges *et al.* [49]. After standard washing procedures, DNA was eluted in nuclease free water by incubating the array at

95°C for 10 min. The captured library was PCR amplified for 18 cycles and purified using the Agencourt AMPure XP SPRI beads (Beckman Coulter, Australia).

The enriched library was quantified using a NanoDrop Spectrophotometer and the quality was checked on the Agilent High Sensitivity Bioanalyzer Chip. RT-qPCR was conducted on pre- and post-captured library using primer pairs designed for the target (*EtCesA1*, *EtCesA2* and *EtCesA5*) and non-target (*EtelF4* and *EtH2B*) genes (S1 Table) to confirm enrichment of the targeted regions. The qRT-PCR data was analyzed using the  $\Delta\Delta\text{CT}$  method described by Livak and Schmittgen [50].

## Sequencing and Analysis

The three pooled barcoded libraries were subjected to cluster generation and  $2 \times 100\text{bp}$  paired end sequencing was conducted using the Illumina GAII Analyzer. High Quality (HQ) reads were filtered from raw data using SeqQC\_V2.2 (a proprietary QC tool of Genotypic Technologies Ltd., Bangalore, India) with cutoff Phred quality scores (Q) of 20 (the probability of 1 in 100 bases sequenced may be due to an error). Further, the quality passed sequencing reads were trimmed for Adapter, B Block and low quality end sequences with 50bp cut off using Raw Data Processing Script. The trimmed reads were aligned (gapped alignment) to the *E. grandis* reference sequence using bowtie 2-2.0.0-beta5 [51] with affine read gap penalty and affine reference gap penalty of 5 for gap open and 3 for gap extension. The un-gapped alignment was done using bowtie version 0.12.7 [52]. The variations across the aligned sequences were taken into account from both gapped and un-gapped alignments to overcome the possibilities of false variations induced by allowing gaps. Variations reported in both alignments are expected to be of higher confidence. SNV calling and InDel detection was done using SAMtools version 0.1.7a (<http://samtools.sourceforge.net>) with default parameters [53]. The cut off thresholds of 3 and 10 were set for the minimum number of reads showing variation and for the minimum RMS mapping quality for SNVs, respectively. The same tool was used to generate the consensus sequence of the aligned reads, while multiple alignments were done using ClustalW version 2.0.12. Pair wise comparison of the sequence data for the three species was conducted to identify SNVs and InDels based on their positions using R Bioconductor code. The ambiguous SNVs generated due to genetic divergence of the three species were not considered for analysis.

## Results and Discussion

### Selection of Candidate Genes

Ninety four xylogenesis-related genes involved in different stages of wood formation including biosynthesis of lignin, cellulose, pectin, monoterpene, xyloglucan, cell wall related genes, genes involved in carbohydrate metabolism, programmed cell death, phyto-hormone signaling, transcription factors and regulatory proteins were selected for the present study (Table 1). The position of the genes in chromosomes and their biological functions in respect to the *E. grandis* reference genome are presented in S2 Table. As many as 14 genes were localized on chromosome 7, while only 4 genes localized on chromosome 8. Two genes, monoterpene glucosyl transferase and IAA binding domain were not assigned to any chromosome.

The formation of the secondary cell wall is driven by the coordinated expression of numerous genes involved in the biosynthesis of cellulose and hemicellulose, lignin, pectin, cell wall proteins and minor soluble and insoluble compounds [54–59], [33, 38–39]. Expressed wood-formation genes show high functional conservation across plant genera and up to 90% of genes expressed in loblolly pine have homologs in *Arabidopsis* [60]. Similarly, a high proportion of poplar ESTs appear to have homologs in the *Arabidopsis* genome [61–62].

**Table 1. Genes selected for target enrichment.**

S. No.	Gene ID	Gene Product	CDS length (bp)	Transcript length (bp)	Biological Function	Xylogenesis-related function
1	<i>4CL</i>	4-Coumarate CoA Ligase	1635	2068	Provides activated thioester substrates for phenylpropanoid natural product biosynthesis	Lignin biosynthesis
2	<i>ACO1</i>	Aminocyclopropane-1-carboxylate oxidase	963	1296	Conversion of ACC to the gaseous hormone ethylene	Ethylene signaling
3	<i>ADH</i>	Alcohol dehydrogenase	1896	2173	Catalyzes the NAD <sup>+</sup> -dependent oxidation of alcohols	Alcohol fermentation
4	<i>AP2L</i>	APETALA TF	738	1052	Key regulators of several developmental processes like floral development	Regulate secondary wall biosynthesis
5	<i>ARF1</i>	Auxin response factor	2364	3228	Transcription factors that bind to TGTCTC auxin response elements in promoters of early auxin response genes	Auxin Signaling. Key regulator of cambium activity and wood formation
6	<i>ARF2</i>	Auxin response factor	2520	3949	TF that bind to auxin response elements in promoters of early auxin response genes	Auxin signaling
7	<i>ASP</i>	Aspartyl protease	1218	1218	Proteolytic enzyme	Programmed cell death
8	<i>BFN1</i>	S1/P1 Nuclease induced during senescence	912	1205	Degradation of RNA and single-stranded DNA	Programmed cell death
9	<i>BP</i>	KNAT Knotted like Homeobox TF	1164	1820	Regulates secondary cell wall biosynthesis	Secondary cell wall biogenesis
10	<i>bZIP</i>	Basic region/leucine zipper motif TF	858	858	Regulate pathogen defence, light and stress signalling	Stem development and xylem fibre differentiation
11	<i>C3H</i>	Coumarate 3- hydroxylase	1530	2048	Hydroxylation of p-coumarate to form caffeate	Lignin biosynthesis
12	<i>C4H</i>	Cinnamate 4- hydroxylase	1518	1928	Catalyzes the conversion of cinnamate into 4-hydroxy-cinnamate	Lignin biosynthesis
13	<i>CAD1</i>	Cinnamyl alcohol dehydrogenase	1158	1583	Conversion of coniferaldehyde to coniferyl alcohol	Lignin biosynthesis
14	<i>CAld5H</i>	Coniferaldehyde 5- hydroxylase	1590	2026	Conversion of coniferaldehyde to hydroxyconiferaldehyde	Lignin biosynthesis
15	<i>CCAAT</i>	CBF TF	453	453	Cis acting element with diverse functions	
16	<i>CCoAOMT1</i>	Caffeoyl-CoA-O-methyl transferase 1	741	1266	Methylation of the 3-hydroxyl group of caffeoyl CoA	Lignin biosynthesis
17	<i>CesA1</i>	Cellulose synthase 1	2937	2937	Cellulose deposition in developing secondary xylem	Cellulose biosynthesis
18	<i>CesA2</i>	Cellulose synthase 2	2775	3136	Cellulose deposition in developing secondary xylem	Cellulose biosynthesis
19	<i>CesA3</i>	Cellulose synthase 3	3126	3126	Cellulose deposition in developing secondary xylem	Cellulose biosynthesis
20	<i>CesA4</i>	Cellulose synthase 4	3243	3951	Cellulose deposition in primary cell wall	Primary cell wall formation
21	<i>CesA5</i>	Cellulose synthase 5	1710	2136	Cellulose deposition in primary cell wall	Primary cell wall formation
22	<i>CesA6</i>	Cellulose synthase 6	2691	3208	Cellulose deposition in primary cell wall	Primary cell wall formation
23	<i>CKX</i>	Cytokinin Oxidase	1386	1386	Cytokinin signaling	Hormonal regulators of cambial development

(Continued)

Table 1. (Continued)

S. No.	Gene ID	Gene Product	CDS length (bp)	Transcript length (bp)	Biological Function	Xylogenesi s-related function
24	<i>COMT1</i>	Caffeic acid O- methyl transferase	1101	1966	Catalyzes the conversion of caffeic acid to ferulic acid and of 5-hydroxyferulic acid to sinapic acid	Lignin biosynthesis
25	<i>CRE</i>	Cytokinin receptor 1	2994	3392	Cytokinin regulation	Cytokinin signal transduction
26	<i>DHN</i>	Dehydrin	513	918	Hydrophilic LEA proteins and accumulate during cellular dehydration	Expressed during dehydration
27	<i>DIR1</i>	Dirigent- like protein	498	498	Lignan biosynthesis process	Template for lignin polymerization
28	<i>DOF1</i>	Plant specific DNA-binding with one finger domain proteins.	1014	1864	Transcriptional Regulators in plant growth and development	Regulates Interfascicular Cambium Formation and Vascular Tissue Development
29	<i>DREB1</i>	Dehydration responsive element binding protein	762	1355	Belong to AP2 TFs and induced during abiotic stress	Stress tolerance
30	<i>DUF1</i>	Domain with unknown function	372	372	Unknown	Predicted function in fibre cell wall development
31	<i>ERF</i>	Ethylene responsive Transcription factor	681	824	Ethylene signaling	Ethylene signaling
32	<i>EXPA</i>	Alpha Expansin	753	1295	Cell wall proteins involved in plant cell growth and developmental processes where cell wall loosening occurs	Plasticize the cellulose-hemicellulose network of primary walls
33	<i>EXPB</i>	Beta Expansin	825	1122	Cell wall proteins involved in plant cell growth and developmental processes where cell wall loosening occurs	Cell wall related
34	<i>F5H</i>	Ferrulate 5-hydroxylase	1590	2026	Hydroxylation of ferulate to 5-hydroxyferulate	Lignin biosynthesis
35	<i>FLA1</i>	Fasciclin like Arabinogalactan protein	945	1281	Diverse developmental roles like differentiation, cell-cell recognition, somatic embryogenesis and PROGRAMMED CELL DEATH	Expressed during xylem differentiation
36	<i>GA20</i>	Gibberellin 20-oxidase	1158	1716	Key oxidase enzyme in the biosynthesis of gibberellin	GA signaling
37	<i>GATA1</i>	<i>GATA1</i> transcription factor	1002	1002	Nitrogen metabolism, blue-light-regulated morphogenesis and circadian rhythm	Unknown
38	<i>GLU</i>	Endo glucanase	1506	1938	Catalyzes the hydrolysis of cellulose	Cellulose biosynthesis
39	<i>GRAS1</i>	GRAS family TF	1485	1972	Play diverse roles in root and shoot development, gibberellic acid (GA) signaling and phytochrome A signal transduction	Vascular differentiation
40	<i>GT</i>	Monoterpeneglucosyltransferase	1383	1467	Monoterpene biosynthesis	Monoterpene biosynthesis
41	<i>GT 1</i>	Beta 1–4 xylosyltransferase/ Glycosyltransferase	1008	1262	Involved in the synthesis of the hemicellulose glucuronoxylan	Secondary cell wall biogenesis

(Continued)

Table 1. (Continued)

S. No.	Gene ID	Gene Product	CDS length (bp)	Transcript length (bp)	Biological Function	Xylogenesi s-related function
42	<i>HB</i>	Homeodomain TF	951	1774	Plant development, including maintenance of the biosynthesis and signaling pathways of different hormones.	Xylogenesi s
43	<i>HB1 Class III</i>	Homeodomain Transcription factor Class III	2535	4141	Regulates meristem function	Regulates vascular development
44	<i>HBI class II</i>	Homeodomain TF	759	1205	Phototropism and auxin response	Auxin Signaling
45	<i>HCT</i>	Hydroxycinnamoyl CoA shikimate	1494	2316	Insertion of the 3-hydroxyl group into monolignol precursors	Lignin biosynthesis
46	<i>HDKNOX1</i>	Homeobox Knotted 1-like 7-like TF	930	1434	Repression of progression into specific differentiation steps	Formation and maintenance of shoot apical meristem
47	<i>HYD</i>	Predicted alpha/beta hydrolase fold protein	960	1603	Common to several hydrolytic enzymes with diverse functions	Regulate Xylem Cell differentiation
48	<i>IAA</i>	IAA binding domain	258	736	Mediators of the auxin signal transduction pathway	IAA signaling
49	<i>KNOX2</i>	Class-I KNOTTED1-like homeobox(KNOX) TF	1305	2164	Growth of shoot meristem	Promote meristem function
50	<i>KOR</i>	Korrigan/ Endo glucanase	1872	2910	Catalyzes the hydrolysis of cellulose	Cellulose biosynthesis
51	<i>LAC</i>	Carbohydrate binding module 48/ Dextrinase	2451	2804	Hydrolysis of starch	Cellulose biosynthesis
52	<i>LAC2</i>	Laccase	1659	2255	Oxidative coupling of lignols	Lignin biosynthesis
53	<i>LBD</i>	LATERAL ORGAN BOUNDARY domain TF in inflorescences	735	1029	Involved in position of axillary meristem formation	Regulated by Vascular related NAC Domain TFs
54	<i>LEAFY</i>	Floricaula / Leafy protein	1080	1080	Floral meristem identity proteins	Express in floral and vegetative meristems
55	<i>LIM1</i>	Homeodomain TF	567	1421	Developmental regulators in basic cellular processes such as organizing of cytoskeleton	Lignin biosynthesis
56	<i>MAN</i>	1,4 beta Mannanendohydrolase	1308	2052	Depolymerization of these cell wall mannan polysaccharides	Cell wall components/ carbohydrate metabolic pathway
57	<i>MAX</i>	MORE AXILLARY GROWTH Gene	2094	2659	Regulate auxin transport	Stigolactone related auxin-dependent stimulation of secondary growth
58	<i>MIBP1</i>	Metal (copper) ion binding protein	1425	1823	Unknown	Predicted function in xylogenesi s
59	<i>MTS</i>	Monoterpene synthase	1749	2328	Monoterpene biosynthesis	Monoterpene biosynthesis
60	<i>MUR3</i>	XyloglucangalactosyltransferaseExostosin family	1854	2347	Xyloglucan biosynthesis	Xyloglucan biosynthesis
61	<i>MYB1</i>	Transcription Factor	768	1629	Second-level master regulators in secondary cell wall biosynthesis	Lignification
62	<i>MYB2</i>	Transcription Factor	666	807	Second-level master regulators in secondary cell wall biosynthesis	Lignification
63	<i>NAM1</i>	No apical meristem protein	1971	2982	NAC TF involved in development of shoot apical meristem	Vascular differentiation and Signaling

(Continued)

Table 1. (Continued)

S. No.	Gene ID	Gene Product	CDS length (bp)	Transcript length (bp)	Biological Function	Xylogenesi s-related function
64	<i>PAAPA</i>	Hydroxyproline-rich glycoprotein (HRGP) and 'PAAPA' motif	519	1189	Probable role in cell wall development	No function assigned
65	<i>PAE</i>	Pectin acetyl esterase	1272	2136	Deacetylation of pectin, a major compound of primary cell walls	Pectin biosynthesis
66	<i>PAL</i>	Phenyl alanine ammonia lyase	2148	3044	Participates in phenylpropanoid biosynthesis	Lignin biosynthesis
67	<i>PG</i>	Polygalacturanase	1530	2288	Degrades polygalacturonan	Cell wall degradation
68	<i>PIP1</i>	Aquaporin	864	1268	Membrane intrinsic protein for water channelling	Transport of water and/or small neutral solutes
69	<i>PL</i>	PectateLyase	1023	1023	Cleavage of pectate	Pectin biosynthesis
70	<i>POX1</i>	Peroxidase	951	1343	Hemoprotein catalyzing the oxidation by hydrogen peroxide	Lignin biosynthesis
71	<i>PTM5</i>	MAD box TF	279	279	Flower development	Vascular development
72	<i>RAB</i>	Ras-related protein	624	1122	Protein transport. Regulator of membrane traffic from the Golgi apparatus towards the endoplasmic reticulum	Activation of autophagy during wood formation
73	<i>RNS</i>	Ribonulcase T2 family	687	856	Hydrolyse RNA	Programmed cell death
74	<i>ROP1</i>	RAC- like small GTPase	594	1290	Regulate cellular processes ranging from vesicle trafficking to hormone signaling	Signaling protein during secondary xylem formation
75	<i>SAMS</i>	S-adenosylmethionine synthase	1182	1769	Catalyzes the formation of S-adenosylmethionine	Methylation of lignin precursors
76	<i>SBP1</i>	Squamosa promoter binding protein TF	1656	2407	Involved in the vegetative to reproductive phase transition; Expression is regulated by MIR156b.	Meristem activation
77	<i>SCD</i>	Short chain dehydrogenase	975	1449	Participates in secondary metabolism, stress responses and phytosteroid biosynthesis	Hormone biosynthesis
78	<i>SND1</i>	Wood-associated NAC domain transcription factor 1A (WND1A)	1200	1684	Plant developmental process	Key Regulator of secondary wall synthesis in fibres
79	<i>STM</i>	Shoot meristemless TF	1128	2348	Meristem formation and maintenance	Regulator of vascular cambium
80	<i>SuSy1</i>	Sucrose synthase	2418	2869	Starch and sucrose metabolism	Cellulose biosynthesis
81	<i>TUA1</i>	Alpha tubulin	1356	2011	Globular cytoskeleton proteins	Component of microtubules
82	<i>UBI LIG</i>	Ubiquitin Ligase	969	1595	Protein ubiquitination. Targets specific protein substrates for degradation by the proteasome	Programmed cell death
83	<i>UGDH</i>	UDP-glucose dehydrogenase	1443	2141	Oxidizes UDP-Glc (UDP-D-glucose) to UDP-GlcA (UDP-D-glucuronate)	Carbohydrate metabolism
84	<i>UGT</i>	UDP glucose glucosyltransferase	1410	1687	Catalyze the conjugation of glucose from sugar nucleotides to various substrates	Carbohydrate biosynthesis
85	<i>UXS1</i>	UDP-D-Glucuronatecarboxylase	615	689	Catalyzes the conversion of UDP-d-glucuronate to UDP-d-xylose	Carbohydrate metabolism

(Continued)



Table 1. (Continued)

S. No.	Gene ID	Gene Product	CDS length (bp)	Transcript length (bp)	Biological Function	Xylogenesis-related function
86	VND6	Vascular-related NAC-domain TF	1047	3267	Master regulator of xylem vessel differentiation	Master regulator of xylem vessel differentiation
87	VND7	Vascular-related NAC-domain TF	963	1531	Master regulator of xylem vessel differentiation	Master regulator of xylem vessel differentiation
88	WND1	Wood associated NAC TF	1152	1527	Activating the entire secondary wall biosynthetic program	Regulation of secondary wall biosynthesis pathway
89	WRKY1	TFs involved in biotic and abiotic stress responses	2256	2870	Plant responses to biotic and abiotic stresses	Cell wall lignification
90	WUS1	Homeodomain TF	685	726	A typical Homeodomain TF involved in lateral organ formation and meristem function	Shoot apical meristem formation and maintenance
91	XCP	Xylem specific cysteine protease	1131	1564	Cellular autolysis	Programmed cell death
92	XTH	Xyloglucanendo-transglycosylase/ hydrolase	894	1223	Cell wall extensibility	Regulates cell growth by strengthening or weakening xyloglucan-cellulose microfibril network
93	XYL	Endo 1–4 beta Xylanase	2796	3444	Degrades the linear polysaccharide beta-1,4-xylan into xylose	Carbohydrate active enzymes in secondary cell wall biogenesis. Decreases cellulose crystallinity in cell walls
94	Znf1	Zinc finger C3HC4 type (RING)	1113	1542	Cysteine rich domain involved in mediating protein-protein interactions	Ubiquitination

doi:10.1371/journal.pone.0116528.t001

The role of transcription factors as master switches in vascular and xylem development has been investigated in detail in poplar, eucalypts, pine and *Arabidopsis*. Highly expressed transcription factors like MYB and NAC families are implicated as critical regulators of vascular differentiation, phenylpropanoid metabolism, xylem differentiation and secondary wall formation. The other important regulators include the homeodomain superfamily of transcription factors (HD-Zip, WOX, KNOX, and ZF-HD), ethylene responsive elements (AP2/ERF domain), bZIP, WRKY and LIM [63–70].

Hormonal regulation of wood formation is well documented and major phyto-hormones playing pivotal role in cambial activity and wood formation include auxin, cytokinin, gibberellic acid, brassinosteroids and ethylene. The receptors of hormone responsive genes and transcription factors are reported to be expressed during cambial development and wood formation [71–74].

The selection of genes in the present study was based on the literature survey as described above and major functional and regulatory genes presumably involved in cambial development and wood formation were selected.

## Validation of Target Enrichment

The array based hybridization enrichment was conducted to capture the 94 xylogenesis-related genes in three species of *Eucalyptus*. The enrichment of the targeted regions after hybridization was validated using the RT-qPCR on pre- and post-capture libraries for target genes *EtCesA1*, *EtCesA2* and *EtCesA5* and non target genes *EteIF4* and *EtH2B*. The comparison of pre and post hybridization data demonstrated 64 fold, 165 fold and 59 fold enrichments of the target genes, *EtCesA1*, *EtCesA2* and *EtCesA5* respectively, while no enrichment was observed for the non target genes, *EteIF4* and *EtH2B*.

## Read and Alignment Statistics

The  $2 \times 100$  bp paired end raw reads were subjected to quality checking using SeqQC\_V2.2. In *E. camaldulensis* (Ec111), a total of 15.75 million reads were generated and the total number of HQ reads were 13.86 million (88.02%), while in *E. tereticornis* (Et 86), the total number of reads were 17.07 million and the number of HQ reads were 15.14 million (88.69%). In *E. grandis* (Eg9), the total number of reads was 11.41 million with 10.22 million HQ reads (89.59%).

The HQ reads from all the three species were aligned with the *E. grandis* reference sequence using both gapped and un-gapped alignment tools. In *E. camaldulensis*, 170866bp (98.43% read coverage) were aligned with the reference sequence, which had a total sequence length of 173593bp, while in *E. tereticornis*, 170825bp sequence length was aligned with reference with 98.41% coverage. Similarly, in *E. grandis*, 170671bp was aligned with the reference sequence with coverage of 98.32%. The total percent of reference covered with at least 5X depth was 97.71%, 97.86% and 97.12% in *E. camaldulensis*, *E. tereticornis* and *E. grandis*, respectively, while reference covered with at least 10X read depth was 96.99%, 97.36% and 95.67%, respectively. Similarly, the alignment statistics for reference covered with 20X depth was 95.9%, 96.34% and 93.53% in *E. camaldulensis*, *E. tereticornis* and *E. grandis*, respectively. The optimized average read depth in *E. camaldulensis* was  $\sim 223X$ , while in *E. tereticornis* it was calculated as  $\sim 227X$ . The optimized average read depth in *E. grandis* was  $\sim 199X$ . The aligned sequence data was deposited in NCBI Short Read Archive with the accession number SRP045253 for *E. tereticornis* (SRX747331), *E. camaldulensis* (SRX669390) and *E. grandis* (SRX747330).

Next generation sequencing platforms produce robust sequence output making high throughput DNA marker discovery feasible and cost effective [75–76]. It was reported that considering all available NGS platforms, Illumina was preferred for *de novo* sequencing, re-sequencing and high-throughput SNP discovery, due to generation of high read depth leading to reference based contig assembly with high confidence [75–77]. The efficiency of this platform in SNP discovery has been well documented in *E. camaldulensis* [78]; *Arabidopsis* [79]; wheat [80–82]; olive [83]; *Solanum* spp. [84]; Douglas—fir [85]; soybean [86–87]; apple [88] and pine [89].

Another important consideration while conducting target enrichment and re-sequencing is the read depth to reliably detect SNPs. It was reported that a minimum of 8X coverage [90] and up to 200X [91] was optimal for SNP calling. In the present study, the read depth was significantly high at  $\sim 223X$  in *E. camaldulensis*,  $\sim 227X$  in *E. tereticornis* and  $\sim 199X$  in *E. grandis*. Similar studies in *Fragaria vesca* documented the average depth as 120X [92], while in *E. camaldulensis*, the average read depth for all the bases was 6124X [78].

Specificity (the number of reads that map to the targeted sequence) is an important aspect of target enrichment experiments. The present study documented high read coverage with *E. camaldulensis* showing 98.43% coverage, *E. tereticornis* with 98.41% coverage and *E. grandis* with coverage of 98.32% with reference sequence, suggesting high specificity of the hybridized

probes to the target sequences. Similarly, in an earlier study in *E. camaldulensis*, 94.2% coverage was reported with reference genome of *E. grandis* [78]. In the wheat, NimbleGen array with genomic DNA derived from eight wheat varieties was used for target enrichment and exome sequencing and an average of 38.1% (22%–44.5%) was aligned to the reference sequence [80], while Saintenac and co workers [82] reported an increase in specificity of reads on target to 60% and the number of covered target bases reported was 92%. In *Populus trichocarpa*, an average of 86.8% of base pairs in the bait regions was mapped on the reference sequence [93]. Hence, the high read depth and coverage achieved in the present investigation can be considered optimal for identification of variation with high confidence.

## Identification of Variants (Snvs And Indels) in Three Eucalyptus Species across E. Grandis Reference Genome

The SNVs and InDels present in the sequences aligned with the reference were individually determined for each species. A total of 5905 SNVs were discovered in all three species, which included 2294 SNVs in *E. camaldulensis* (604 and 299 SNVs from gapped and un-gapped alignments, respectively and 1391 SNVs common for both gapped and un-gapped alignments), 2383 SNVs in *E. tereticornis* (636 and 303 SNVs from gapped and un-gapped alignments, respectively and 1444 SNVs common for both alignments), and 1228 SNVs in *E. grandis* (460 and 122 SNVs from gapped and un-gapped alignments, respectively and 646 SNVs common for both alignments) (Table 2).

The presence of SNVs in UTRs and exons were also identified and maximum number of SNVs was recorded in the exon region (4187), while 1226 SNVs were documented in the 3'UTR. A total number of 492 SNVs were identified in the 5'UTR across all the three species (Table 3, 4 & 5). In *E. tereticornis*, the maximum number of SNVs was recorded in *SuSy1* (85), while only one SNV was observed in *PTM5* (S3a Table). In *E. camaldulensis*, a similar trend was observed with maximum of 72 SNVs identified in *SuSy1* and only one SNV recorded in *PTM5* (S4a Table). However, when the *E. grandis* sequences were compared with the reference genome, a maximum of 60 SNVs was observed in *C3H* while a single SNV was documented in several genes, including *AP2L*, *ARF*, *ARF2*, *EXPA*, *GATA1*, *LAC2*, *PTM5*, *VND6*. No SNVs were detected in *CCAAT*, *FLA1*, and *LBD* (S5a Table).

The SNV frequency was calculated for exon and the UTR regions individually in each species. The SNV frequency in 5'UTR of *E. tereticornis*, *E. camaldulensis* and *E. grandis* was 1/78.49bp, 1/101.11bp and 1/170.42 respectively, while SNV frequency in the exon region was 1/126.78, 1/125.61 and 1/306.72 for *E. tereticornis*, *E. camaldulensis* and *E. grandis* respectively. In 3'UTR, the SNV frequency was 1/86.61, 1/100.23 and 1/176.08 for *E. tereticornis*, *E. camaldulensis* and *E. grandis* respectively (Table 3, 4 & 5).

Further, the presence of SNVs in pair-wise combination between the three *Eucalyptus* species was also conducted. The gene-wise presence of ambiguous nucleotides was not considered and SNV with no ambiguity was mapped on the candidate genes (S6 Table). When *E. camaldulensis* and *E. tereticornis* were compared, a total of 317 SNVs were documented with a minimum of one SNV in *4CL*, *bZIP*, *CCoAOMT1*, *CesA3*, *EXPA*, *GRAS1*, *NAM1*, *PIP1*, *PTM5*, *SBP1*, *SND1*, *STM*, *SuSy1*, *TUA1*, *VND7* and a maximum of 25 SNVs in *LAC*. Larger number of SNVs were recorded when *E. grandis* was compared with *E. tereticornis* and *E. camaldulensis* with 875 and 1014 SNVs respectively. In both pair-wise combinations, the maximum number of SNVs was observed in *LAC* with 53 SNVs when compared across *E. camaldulensis* and 46 SNVs when compared across *E. tereticornis*.

The presence of InDels were also detected when the sequences of 94 genes were compared individually across the reference and a total of 1406 InDels were discovered with the size range

**Table 2. SNVs and InDels across 94 genes in three *Eucalyptus* species.**

S. No.	Gene Name	Gene ID	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
			SNVs	InDels	SNVs	InDels	SNVs	InDels
1	4-coumarate-CoA ligase	<i>4CL</i>	41	5	43	2	5	5
2	Aminocyclopropane-1-carboxylate oxidase	<i>ACO1</i>	36	6	34	3	30	5
3	Alcohol dehydrogenase	<i>ADH</i>	32	7	31	7	26	8
4	APETALA TF	<i>AP2L</i>	7	4	6	4	1	4
5	Auxin response factor	<i>ARF</i>	26	11	30	12	1	7
6	Auxin response factor	<i>ARF2</i>	29	14	22	15	1	12
7	Aspartyl protease	<i>ASP</i>	36	0	32	0	33	0
8	S1/P1 nuclease induced during senescence	<i>BFN1</i>	25	3	21	3	4	3
9	KNAT knotted like homeobox TF	<i>BP</i>	11	7	12	7	2	7
10	Basic region / leucine zipper motif TF	<i>bZIP</i>	15	2	13	2	11	2
11	P-coumarate 3-hydroxylation	<i>C3H</i>	68	6	68	5	60	5
12	Cinnamate 4-hydroxylase	<i>C4H</i>	11	4	12	4	3	1
13	Cinnamyl alcohol dehydrogenase	<i>CAD1</i>	14	6	18	5	6	6
14	Coniferaldehyde 5-hydroxylase	<i>CAld5H</i>	19	0	19	0	9	0
15	CBF TF	<i>CCAAT</i>	2	1	2	2	0	1
16	Caffeoyl-CoA-O-methyltransferase	<i>CCoAOMT1</i>	14	9	7	4	4	5
17	Cellulose synthase 1	<i>CesA1</i>	22	11	16	10	28	9
18	Cellulose synthase 2	<i>CesA2</i>	19	8	29	8	17	6
19	Cellulose synthase 3	<i>CesA3</i>	35	11	36	11	27	10
20	Cellulose synthase 4	<i>CesA4</i>	51	15	59	15	30	11
21	Cellulose synthase 5	<i>CesA5</i>	55	8	62	10	57	8
22	Cellulose synthase 6	<i>CesA6</i>	30	13	29	13	25	12
23	Cytokinin oxidase	<i>CKX</i>	30	5	27	4	26	5
24	Caffeic acid-O-methyltransferase	<i>COMT1</i>	35	7	37	9	26	3
25	Cytokinin receptor 1	<i>CRE</i>	31	13	37	14	7	12
26	Dehydrin	<i>DHN</i>	20	4	20	3	4	1
27	Dirigent like protein	<i>DIR1</i>	2	0	2	1	3	0
28	DNA binding with one finger	<i>DOF1</i>	17	0	19	0	4	0
29	Dehydration-Responsive Element-Binding protein	<i>DREB1</i>	11	3	16	3	2	2
30	Domain of Unknown function 1	<i>DUF1</i>	7	1	13	1	7	1
31	Ethylene responsive transcription factor	<i>ERF</i>	19	1	21	1	20	3
32	Alpha expansin	<i>EXPA</i>	13	3	17	7	1	3
33	Beta expansin	<i>XPB</i>	29	3	30	2	29	2
34	Ferulate-5-hydroxylase	<i>F5H</i>	19	0	17	0	8	0
35	Fasciclin like arabinogalactan	<i>FLA1</i>	22	0	17	0	0	1
36	Gibberlin 20-oxidase	<i>GA20</i>	12	3	10	4	12	3
37	<i>GATA1</i> transcription factor	<i>GATA1</i>	7	2	9	0	1	0
38	Endo glucanase	<i>GLU</i>	21	4	16	5	16	4
39	GRAS family TF	<i>GRAS1</i>	16	3	14	1	4	2
40	Monoterene glycosyl transferases	<i>GT</i>	33	2	35	1	14	2
41	Beta 1–4 xylosyltransferase/glycosyl transferase	<i>GT_1</i>	15	3	13	3	10	4
42	Homeodomain TF	<i>HB</i>	23	10	23	8	11	4
43	Homeodomain TF	<i>HB1ClassIII</i>	32	20	22	18	16	17
44	Homeodomain TF	<i>HB1classII</i>	12	3	13	4	3	4
45	Hydroxycinnamoyl CoA shikimate	<i>HCT</i>	34	5	37	5	20	1
46	Homeobox knotted 1-like 7-like TF	<i>HDKNOX1</i>	9	6	11	5	5	4

(Continued)

Table 2. (Continued)

S. No.	Gene Name	Gene ID	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
			SNVs	InDels	SNVs	InDels	SNVs	InDels
47	Predicted alpha/beta hydrolase fold protein	<i>HYD</i>	17	6	15	4	7	4
48	IAA binding domain	<i>IAA</i>	13	3	9	1	8	1
49	Class-I KNOTTED 1 like homeobox (KNOX)TF	<i>KNOX2</i>	22	8	20	8	10	13
50	KORRIGAN /endo glucanase	<i>KOR</i>	60	9	54	11	28	11
51	Carbohydrate binding module 48/ dextrinase	<i>LAC</i>	26	10	25	13	6	8
52	Laccase	<i>LAC2</i>	29	5	23	7	1	5
53	Lateral organ boundary domain TF in inflorescences	<i>LBD</i>	8	5	9	2	0	1
54	Floricaula/ leafy protein	<i>LEAFY</i>	31	4	34	3	22	2
55	Homeodomain TF	<i>LIM1</i>	17	7	13	6	3	9
56	1,4 beta mannan endohydrolase	<i>MAN</i>	15	7	19	6	7	6
57	More axillary growth gene	<i>MAX</i>	48	8	56	4	39	2
58	Metal (copper) ion binding	<i>MIBP1</i>	55	4	44	2	33	3
59	Monopterene synthase	<i>MTS</i>	49	7	44	8	24	4
60	Xyloglucan galactosyl transferase exostosin family	<i>MUR3</i>	30	1	28	1	4	1
61	Myeloblastosis TF	<i>MYB1</i>	21	5	18	5	2	3
62	Myeloblastosis TF	<i>MYB2</i>	10	1	9	3	11	1
63	No apical meristem family protein	<i>NAM1</i>	71	7	62	8	8	6
64	Hydroxy proline rich glycoprotein (HRGP)/ PAAPA motif	<i>PAAPA</i>	37	3	21	1	6	2
65	Pectin acetyl esterase	<i>PAE</i>	26	9	32	10	11	6
66	Phenylalanine ammonia-lyase	<i>PAL</i>	75	6	71	4	27	3
67	Poly galacturanase	<i>PG</i>	38	7	44	5	8	5
68	Plasma membrane intrinsic protein	<i>PIP1</i>	21	3	24	5	12	5
69	Pectate lyase	<i>PL</i>	32	1	22	0	20	0
70	Peroxidase	<i>POX1</i>	27	3	24	3	12	3
71	MAD box TF	<i>PTM5</i>	1	1	1	1	1	1
72	RAS related protein	<i>RAB</i>	12	6	11	5	12	6
73	Ribonuclease T2 family	<i>RNS</i>	8	3	9	3	13	3
74	RAC like small GTPase	<i>ROP1</i>	11	5	8	6	4	4
75	S-Adenosyl methionine synthetase	<i>SAMS</i>	37	3	31	3	21	2
76	Squamosapromoter binding protein TF	<i>SBP1</i>	33	6	30	6	22	5
77	Sitosterol cello dextrin	<i>SCD</i>	17	5	8	7	11	5
78	Wood associated NAC domain TF 1(WND1)	<i>SND1</i>	6	7	5	7	4	6
79	Shoot meristemless TF	<i>STM</i>	18	20	17	12	3	7
80	Sucrose synthase	<i>SuSy1</i>	85	9	72	9	54	8
81	$\alpha$ -Tubulin	<i>TUA1</i>	33	7	21	7	5	4
82	Ubiquitin Ligase	<i>UBILIG</i>	9	5	9	4	2	5
83	UDP glucose glucosyl dehydrogenase	<i>UGDH</i>	21	6	18	3	12	3
84	UDP glucose glucosyl transferase	<i>UGT</i>	40	2	63	1	44	2
85	UDP-D-glucuronate carboxylase	<i>UXS1</i>	9	1	12	0	9	0
86	Vascular related NAC domain TF	<i>VND6</i>	22	8	21	5	1	4
87	Vascular related NAC domain TF	<i>VND7</i>	9	5	9	4	4	4
88	Wood associated NAC TF	<i>WND1</i>	19	2	16	4	5	2
89	TF involved in biotic and abiotic stress response	<i>WRKY1</i>	32	10	25	9	15	7
90	Homeodomain TF	<i>WUS1</i>	3	1	9	1	5	0
91	Xylem-specific papain-like Cysteine Peptidase	<i>XCP</i>	21	4	20	3	6	2

(Continued)

Table 2. (Continued)

S. No.	Gene Name	Gene ID	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
			SNVs	InDels	SNVs	InDels	SNVs	InDels
92	Xyloglucan transglycosylase	<i>XTH</i>	22	2	28	4	12	2
93	Endo 1,4 beta xylanase	<i>XYL</i>	48	9	41	8	16	10
94	Zinc finger (C3HC4-type ring finger) TF protein	<i>Znf1</i>	22	10	13	6	9	8
<b>Total</b>			<b>2383</b>	<b>518</b>	<b>2294</b>	<b>479</b>	<b>1228</b>	<b>409</b>

doi:10.1371/journal.pone.0116528.t002

of 1–24 nucleotides (Table 2). The position of InDels in exons and UTRs was also determined and the total number documented was 843, 309 and 254 in exons, 3'UTR and 5'UTR, respectively (Table 6). In *E. tereticornis*, a total of 518 InDels were detected and a maximum of 20 InDels was recorded in the transcription factor *HB1 Class III*, while a single InDel was documented in several genes including *CCAAT*, *DUF1*, *ERF*, *MUR3*, *MYB2*, *PL*, *PTM5*, *UXS1* and *WUS1*. No InDels were recorded in *ASP*, *Cald5H*, *DOF1*, *F5H*, *DIR1*, and *FLA1* (S3b Table). In *E. camaldulensis*, a total of 479 InDels were recorded and the maximum number of InDels was discovered in *HB1ClassIII* (18), while only a single InDel was identified in *DIR1*, *DUF1*, *ERF*, *GRAS1*, *GT*, *IAA*, *MUR3*, *PAAPA*, *PTM5*, *UGT* and *WUS1*. InDels were not detected in *ASP*, *Cald5H*, *DOF1*, *F5H*, *FLA1*, *GATA1*, *PL* and *UXS1* (S4b Table). In *E. grandis*, a total of 409 InDels were discovered and a maximum of 17 InDels was documented in *HB1ClassIII*, while only a single InDel was identified in *FLA1*, *DUF1*, *IAA*, *MUR3*, *PTM5*, *CCAAT*, *LBD*, *DHN*, *MYB2*, *C4H* and *HCT*. InDels were not found in *ASP*, *Cald5H*, *DOF1*, *F5H*, *GATA1*, *PL*, *UXS1*, *DIR1* and *WUS1* (S5b Table). The InDel frequency was calculated for each species (Table 6). The InDel frequency (bp/InDel) was the highest in the exon region for all the three species with 411.14, 446.38 and 482.58 in *E. tereticornis*, *E. camaldulensis* and *E. grandis*, respectively. The total InDel frequency was 332.05, 359.08 and 420.54 bp per InDel in *E. tereticornis*, *E. camaldulensis* and *E. grandis* respectively, across the all the genes selected (Table 6).

Similarly, the presence of InDels was also documented in pair-wise combination and a total of 731 and 699 InDels were detected across *E. grandis* & *E. tereticornis* and *E. grandis* & *E. camaldulensis*, respectively. A total of 702 InDels were detected between *E. camaldulensis* and *E. tereticornis*. Maximum number of InDels across all combinations was observed in *HB1 Class III* transcription factor with 26 InDels when compared between *E. grandis* and *E. tereticornis*, 27 InDels between *E. grandis* and *E. camaldulensis* and 27 InDels between *E. camaldulensis* and *E. tereticornis*. A minimum of one InDel was documented across several genes like *FLA1*; *DIR1*, *EXPB*, *FLA1*, *WUS1* and *DIR1*, *DUF1*, *PL*, *UXS1* in *E. grandis* & *E. tereticornis*; *E. grandis* & *E. camaldulensis* and *E. camaldulensis* & *E. tereticornis* respectively (S7a,b,c Table).

The abundance of SNPs / SNVs in plant genome and the availability of cost effective technologies for genotyping has made high-throughput SNP genotyping pivotal for genetic mapping, gene discovery, germplasm characterization and population genomics [94]. NGS based SNP discovery is reported in several crop like wheat [80], [81], [82]; *Eucalyptus* [95]; rice [96]; barley [97]; cotton [98]; soybean [86]; potato [99]; *Arabidopsis* [100]; maize [101] and several other species. Use of SNP marker panels for genetic analysis has been widely explored in less domesticated crop [102] and trees [103–105]. SNP genotyping in *Eucalyptus* species is reported from *E. grandis* [35], *E. globulus*, *E. nitens*, *E. camaldulensis* and *E. loxophleba* [16], inter-specific hybrids of *Eucalyptus* [106], *E. pilularis* [107], *E. globulus* [108] and *E. camaldulensis* [41,78].

The SNP frequency in *Eucalyptus* species is considered to be one of the highest in woody species due to its recent domestication, large population size and outbred mating system [94].

**Table 3. SNV frequency in three *Eucalyptus* species in 5'UTR region.**

GENE ID	5'UTR_length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
4CL	126	3	42.0	3	42.0	1	126.0
ACO1	107	2	53.5	1	107.0	1	107.0
ADH	0	na	na	na	na	na	Na
AP2L	59	1	59.0	1	59.0	-	
ARF	413	4	103.3	3	137.7	1	
ARF2	1004	8	125.5	9	111.6	-	
ASP	0	na	na	na	na	na	Na
BFN1	0	na	na	na	na	na	Na
BP	267	-		-		-	
bZIP	0	na	na	na	na	na	Na
C3H	86	2	43.0	1	86.0	2	43.0
C4H	125	3	41.7	1	125.0	-	
CAD1	127	2	63.5	5	25.4	-	
CAId5H	39	-		-		-	
CCAAT	0	na	na	na	na	na	na
CCoAOMT1	97	1	97.0	-		-	
CesA1	0	na	na	na	na	na	na
CesA2	361	-		-		-	
CesA3	0	na	na	na	na	na	na
CesA4	340	8	42.5	2	170.0	4	85.0
CesA5	118	-		1	118.0	1	118.0
CesA6	16	-		-		-	
CKX	0	na	na	na	na	na	na
COMT1	98	1	98.0	-		2	49.0
CRE	0	na	na	na	na	na	na
DHN	112	4	28.0	1	112.0	1	112.0
DIR1	0	na	na	na	na	na	na
DOF1	227	1	227.0	1	227.0	-	
DREB1	298	5	59.6	8	37.3	-	
DUF1	0	na	na	na	na	na	na
ERF	0	na	na	na	na	na	na
EXPA	43	-		1	43.0	-	
EXPB	44	1	44.0	1	44.0	1	44.0
F5H	39	-		-		-	
FLA1	1	-		-		-	
GA20	105	-		-		-	
GATA1	0	na	na	na	na	na	na
GLU	41	1	41.0	1	41.0	-	
GRAS1	213	3	71.0	3	71.0	-	
GT	54	3	18.0	3	18.0	-	
GT 1	187	-		-		-	
HB	305	1	305.0	1	305.0	1	305.0
HB1 ClassIII	1090	8	136.3	8	136.3	5	218.0
HBI class II	102	-		-		-	

(Continued)

Table 3. (Continued)

GENE ID	5'UTR_length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
HCT	298	4	74.5	9	33.1	3	99.3
HDKNOX1	103	-		-		-	
HYD	262	2	131.0	1	262.0	2	131.0
IAA	91	-		-		-	
KNOX2	182	3	60.7	7	26.0	3	60.7
KOR	465	9	51.7	9	51.7	7	66.4
LAC	0	na	na	na	na	na	na
LAC2	52	-		-		-	
LBD	35	-		-		-	
LEAFY	0	na	na	na	na	na	na
LIM1	408	12	34.0	10	40.8	2	204.0
MAN	218	-		1	218.0	-	
MAX	35	-		-		-	
MIBP1	0	na	na	na	na	na	na
MTS	230	6	38.3	5	46.0	5	46.0
MUR3	493	8	61.6	9	54.8	1	493.0
MYB1	429	8	53.6	6	71.5	1	429.0
MYB2	79	1	79.0	1	79.0	3	26.3
NAM1	307	10	30.7	5	61.4	1	307.0
PAAPA	131	-		2	65.5	-	
PAE	536	12	44.7	15	35.7	7	76.6
PAL	194	6	32.3	5	38.8	-	
PG	242	6	40.3	6	40.3	4	60.5
PIP1	67	-		-		-	
PL	0	na	na	na	na	na	na
POX1	42	1	42.0	-		-	
PTM5	0	na	na	na	na	na	na
RAB	158	2	79.0	2	79.0	1	158.0
RNS	15	-		-		-	
ROP1	241	-		-		-	
SAMS	271	13	20.8	9	30.1	2	135.5
SBP1	451	4	112.8	2	225.5	1	451.0
SCD	181	6	30.2	1	181.0	1	181.0
SND1	218	2	109.0	1	218.0	3	72.7
STM	669	5	133.8	4	167.3	2	334.5
SuSy1	140	2	70.0	1	140.0	-	
TUA1	318	11	28.9	9	35.3	-	
UBILIG	434	6	72.3	4	108.5	1	434.0
UGDH	258	2	129.0	2	129.0	-	
UGT	11	-		-		-	
UXS1	0	na	na	na	na	na	na
VND6	422	3	140.7	4	105.5	-	
VND7	117	1	117.0	1	117.0	-	
WND1	77	-		-		-	
WRKY1	390	8	48.8	5	78.0	-	

(Continued)



Table 3. (Continued)

GENE ID	5'UTR_length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
<i>WUS1</i>	41	-		-		-	
<i>XCP</i>	30	-		-		-	
<i>XTH</i>	119	1	119.0	-		1	119.0
<i>XYL</i>	362	6	60.3	3	120.7	2	181.0
<i>Znf1</i>	180	1	180.0	1	180.0	1	180.0
<b>Total</b>	<b>16246</b>	<b>223</b>	<b>78.49*</b>	<b>195</b>	<b>101.11*</b>	<b>74</b>	<b>170.42*</b>

na: Not applicable

\* denotes average SNV frequency

doi:10.1371/journal.pone.0116528.t003

Kulheim and coworkers [16] reported that the SNP density in *E. nitens* was 1/33bp, 1/31 bp in *E. globulus*, while in *E. camaldulensis* and *E. loxophleba* it was significantly high at 1/16bp and 1/17bp respectively. However, a later study showed that the SNP frequency was 1/83.9bp in *E. camaldulensis* [78]. In the present study, the SNV frequency ranged from 1/78.49bp to 1/306.72bp across different genic regions of *E. camaldulensis*, *E. tereticornis* and *E. grandis*. Recently, the SNP frequency in inter-specific hybrids of Eucalypts was documented as 1/133bp [109], suggesting that the SNP frequency was depended on the target region. In heterozygous species, the SNP frequency is generally high as documented in pine with 1/102.6bp [110], grapevine with 1/64bp [111], maize with 1/60bp [112] and rye which registered a SNP frequency of 1 SNP at 52bp interval [113].

Insertion and deletion polymorphisms (InDels) are an important source of genomic variation in plant and animal genomes. Mechanisms such as insertion and excision of transposable elements, slippage in simple sequence replication, errors in DNA synthesis and repair, recombination and unequal crossover can result in the formation of InDels [114–115]. However, accurate genotyping from low-coverage sequence data can be challenging [116]. Further, polymorphism in short InDels is increasingly being used as an important marker in humans [117], *Drosophila melanogaster* [118] and *G. gallus* [119]. Report on InDel genotyping in plants are limited to rice [120], *Arabidopsis thaliana* [121], *Citrus clementina* [122] and *Phaseolus vulgaris* [123]. In tree species, InDel discovery is reported from *Salix* spp. [124] and *Populus* spp. [125–126]. InDel markers for species discrimination have been reported in *E. grandis* and *E. gunnii* [39] and *Populus* spp. [125,127].

In the present study, high number InDels in the size range of 1–24 nucleotides were documented in the three *Eucalypts* species at a frequency of 332.05, 359.08 and 420.54 bp per InDel in *E. tereticornis*, *E. camaldulensis* and *E. grandis*, respectively. This is higher than the earlier reported InDel frequency of 1.5 InDel/1000 bp [115] in *Eucalyptus* genome and 1/2756bp in inter-specific hybrid population [109]. Similarly, in *Pinus taeda*, Kong et al. [128] reported that InDels were infrequent with only 0.67% frequency in targeted regions. The probable reason for this variance in the present investigation could be due to the highly divergent genotypes selected in the present study, indicating that InDels could be a useful marker for genetic analysis in *Eucalyptus* species.

## Conclusion

The NGS platforms have brought in paradigm shift in understanding the different aspects of plant biology especially in model species and plants with small genome. Its downstream

Table 4. SNV frequency in three *Eucalyptus* species in Exon region.

GENE ID	Exon length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
4CL	1635	37	44.2	38	43.0	4	408.8
ACO1	963	28	34.4	29	33.2	26	37.0
ADH	1896	18	105.3	23	82.4	20	94.8
AP2L	738	2	369.0	2	369.0	-	
ARF	2364	19	124.4	22	107.5	-	
ARF2	2520	17	148.2	12	210.0	1	2520.0
ASP	1218	36	33.8	32	38.1	33	36.9
BFN1	912	16	57.0	10	91.2	3	304.0
BP	1164	7	166.3	10	116.4	2	582.0
bZIP	858	15	57.2	13	66.0	11	78.0
C3H	1530	55	27.8	50	30.6	51	30.0
C4H	1518	6	253.0	8	189.8	3	506.0
CAD1	1158	12	96.5	13	89.1	6	193.0
CAId5H	1590	17	93.5	17	93.5	9	176.7
CCAAT	453	2	226.5	2	226.5	-	
CCoAOMT1	741	5	148.2	4	185.3	-	
CesA1	2937	22	133.5	16	183.6	28	104.9
CesA2	2775	19	146.1	29	95.7	17	163.2
CesA3	3126	35	89.3	36	86.8	27	115.8
CesA4	3243	38	85.3	50	64.9	21	154.4
CesA5	1710	45	38.0	48	35.6	47	36.4
CesA6	2691	29	92.8	28	96.1	23	117.0
CKX	1386	30	46.2	27	51.3	26	53.3
COMT1	1101	28	39.3	28	39.3	23	47.9
CRE	2994	27	110.9	33	90.7	5	598.8
DHN	513	10	51.3	11	46.6	2	256.5
DIR1	498	2	249.0	2	249.0	3	166.0
DOF1	1014	10	101.4	11	92.2	4	253.5
DREB1	762	4	190.5	7	108.9	-	
DUF1	372	7	53.1	13	28.6	7	53.1
ERF	681	18	37.8	19	35.8	19	35.8
EXPA	753	7	107.6	8	94.1	-	
XPB	825	19	43.4	19	43.4	23	35.9
F5H	1590	17	93.5	15	106.0	8	198.8
FLA1	945	20	47.3	14	67.5	-	
GA20	1158	7	165.4	6	193.0	7	165.4
GATA1	1002	7	143.1	9	111.3	1	1002.0
GLU	1506	16	94.1	11	136.9	14	107.6
GRAS1	1485	11	135.0	6	247.5	4	371.3
GT	1383	29	47.7	31	44.6	14	98.8
GT 1	1008	14	72.0	12	84.0	9	112.0
HB	951	13	73.2	12	79.3	4	237.8
HB1 ClassIII	2535	15	169.0	9	281.7	7	362.1
HBI class II	759	7	108.4	7	108.4	3	253.0
HCT	1494	21	71.1	25	59.8	14	106.7

(Continued)

Table 4. (Continued)

GENE ID	Exon length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
<i>HDKNOX1</i>	930	5	186.0	4	232.5	3	310.0
<i>HYD</i>	960	6	160.0	6	160.0	3	320.0
<i>IAA</i>	258	4	64.5	4	64.5	5	51.6
<i>KNOX2</i>	1305	5	261.0	3	435.0	1	1305.0
<i>KOR</i>	1872	40	46.8	31	60.4	6	312.0
<i>LAC</i>	2451	14	175.1	15	163.4	6	408.5
<i>LAC2</i>	1659	19	87.3	16	103.7	-	
<i>LBD</i>	735	6	122.5	7	105.0	-	
<i>LEAFY</i>	1080	31	34.8	34	31.8	22	49.1
<i>LIM1</i>	567	1	567.0	2	283.5	1	567.0
<i>MAN</i>	1308	6	218.0	9	145.3	5	261.6
<i>MAX</i>	2094	40	52.4	45	46.5	34	61.6
<i>MIBP1</i>	1425	40	35.6	31	46.0	28	50.9
<i>MTS</i>	1749	34	51.4	30	58.3	13	134.5
<i>MUR3</i>	1854	22	84.3	19	97.6	3	618.0
<i>MYB1</i>	768	7	109.7	5	153.6	1	768.0
<i>MYB2</i>	666	9	74.0	8	83.3	8	83.3
<i>NAM1</i>	1971	46	42.8	43	45.8	6	328.5
<i>PAAPA</i>	519	17	30.5	12	43.3	6	86.5
<i>PAE</i>	1272	9	141.3	10	127.2	4	318.0
<i>PAL</i>	2148	43	50.0	47	45.7	19	113.1
<i>PG</i>	1530	14	109.3	17	90.0	3	510.0
<i>PIP1</i>	864	14	61.7	15	57.6	7	123.4
<i>PL</i>	1023	32	32.0	22	46.5	20	51.2
<i>POX1</i>	951	16	59.4	16	59.4	9	105.7
<i>PTM5</i>	279	1	279.0	1	279.0	1	279.0
<i>RAB</i>	624	4	156.0	3	208.0	4	156.0
<i>RNS</i>	687	5	137.4	5	137.4	7	98.1
<i>ROP1</i>	594	3	198.0	3	198.0	1	594.0
<i>SAMS</i>	1182	17	69.5	16	73.9	13	90.9
<i>SBP1</i>	1656	23	72.0	23	72.0	12	138.0
<i>SCD</i>	975	6	162.5	4	243.8	8	121.9
<i>SND1</i>	1200	3	400.0	3	400.0	1	1200.0
<i>STM</i>	1128	6	188.0	2	564.0	-	
<i>SuSy1</i>	2418	72	33.6	68	35.6	49	49.3
<i>TUA1</i>	1356	18	75.3	8	169.5	1	1356.0
<i>UBILIG</i>	969	1	969.0	3	323.0	1	969.0
<i>UGDH</i>	1443	12	120.3	11	131.2	2	721.5
<i>UGT</i>	1410	35	40.3	57	24.7	40	35.3
<i>UXS1</i>	615	7	87.9	10	61.5	7	87.9
<i>VND6</i>	1047	6	174.5	3	349.0	1	1047.0
<i>VND7</i>	963	6	160.5	7	137.6	4	240.8
<i>WND1</i>	1152	15	76.8	11	104.7	3	384.0
<i>WRKY1</i>	2256	23	98.1	19	118.7	14	161.1
<i>WUS1</i>	685	3	228.3	9	76.1	5	137.0

(Continued)

Table 4. (Continued)

GENE ID	Exon length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
<i>XCP</i>	1131	6	188.5	7	161.6	4	282.8
<i>XTH</i>	894	16	55.9	22	40.6	10	89.4
<i>XYL</i>	2796	13	215.1	33	84.7	11	254.2
<i>Znf1</i>	1113	21	53.0	12	92.8	7	159.0
<b>Total</b>	<b>124987</b>	<b>1621</b>	<b>126.78*</b>	<b>1618</b>	<b>125.61*</b>	<b>948</b>	<b>306.72*</b>

na: Not applicable

\* denotes average SNV frequency

doi:10.1371/journal.pone.0116528.t004

usefulness in linkage map construction, genetic diversity analyses, association mapping, and marker—assisted selection has been demonstrated in several plants [129]. However, sequencing of complete genomes cannot be regularly employed due to high cost and computational limitations in handling robust informatics data. With availability of complexity reduction strategies, sequencing of sub-genomic regions by on-array/in-solution target enrichment technology has provided an efficient alternate strategy to amplicon re-sequencing for SNP/ SNV discovery [130]. In the present study, this strategy was implemented in re-sequencing ninety four genes across three *Eucalyptus* species. This study has also revealed that target enrichment strategy can be successfully used for identification of markers (SNVs and InDels) for future use in QTL and association mapping studies in *Eucalyptus* species.

### Supporting Information

**S1 Table. Primer pairs used for RT-qPCR to confirm enrichment of targeted genes.**  
(DOC)

**S2 Table. Functional Annotation of selected genes across *E. grandis* genome sequence using Phytozome v10.**  
(XLSX)

**S3 Table. A, Details of SNVs documented in *E. tereticornis* across reference sequence.**  
B, Details of InDels documented in *E. tereticornis* across reference sequence.  
(XLS)

**S4 Table. A, Details of SNVs documented in *E. camaldulensis* across reference sequence.**  
B, Details of InDels documented in *E. camaldulensis* across reference sequence.  
(XLS)

**S5 Table. A, Details of SNVs documented in *E. grandis* across reference sequence.** B, Details of InDels documented in *E. grandis* across reference sequence.  
(XLS)

**S6 Table. Presence of SNVs in Pair-wise comparison across three *Eucalyptus* species.**  
(XLS)

**S7 Table. A, Presence of InDels in Pair-wise comparison across *E. grandis* and *E. tereticornis*.**  
B, Presence of InDels in Pair-wise comparison across *E. grandis* and *E. camaldulensis*.

Table 5. SNV frequency in three *Eucalyptus* species in 3' UTR region.

GENE ID	3'UTR_length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
4CL	307	1	307.0	2	153.5	-	
ACO1	226	6	37.7	4	56.5	3	75.3
ADH	277	14	19.8	8	34.6	6	46.2
AP2L	255	4	63.8	3	85.0	1	255.0
ARF	451	3	150.3	5	90.2	-	
ARF2	425	4	106.3	1	425.0	-	
ASP	0	na	na	na	na	na	na
BFN1	293	9	32.6	11	26.6	1	293.0
BP	389	4	97.3	2	194.5	-	
bZIP	0	na	na	na	na	na	na
C3H	432	11	39.3	17	25.4	7	61.7
C4H	285	2	142.5	3	95.0	-	
CAD1	298	-		-		-	
CAld5H	397	2	198.5	2	198.5	-	
CCAAT	0	na	na	na	na	na	na
CCoAOMT1	428	8	53.5	3	142.7	4	107.0
CesA1	0	na	na	na	na	na	na
CesA2	0	na	na	na	na	na	na
CesA3	0	na	na	na	na	na	na
CesA4	368	5	73.6	7	52.6	5	73.6
CesA5	308	10	30.8	13	23.7	9	34.2
CesA6	501	1	501.0	1	501.0	2	250.5
CKX	0	na	na	na	na	na	na
COMT1	767	6	127.8	9	85.2	1	767.0
CRE	398	4	99.5	4	99.5	2	199.0
DHN	293	6	48.8	8	36.6	1	293.0
DIR1	0	na	na	na	na	na	na
DOF1	623	6	103.8	7	89.0	-	
DREB1	295	2	147.5	1	295.0	2	147.5
DUF1	0	na	na	na	na	na	na
ERF	143	1	143.0	2	71.5	1	143.0
EXPA	499	6	83.2	8	62.4	1	499.0
EXPB	253	9	28.1	10	25.3	5	50.6
F5H	397	2	198.5	2	198.5	-	
FLA1	335	2	167.5	3	111.7	-	
GA20	453	5	90.6	4	113.3	5	90.6
GATA1	0	na	na	na	na	na	na
GLU	391	4	97.8	4	97.8	2	195.5
GRAS1	274	2	137.0	5	54.8	-	
GT	30	1	30.0	1	30.0	-	
GT 1	67	1	67.0	1	67.0	1	67.0
HB	518	9	57.6	10	51.8	6	86.3
HB1 ClassIII	516	9	57.3	5	103.2	4	129.0
HBI class II	344	5	68.8	6	57.3	-	
HCT	524	9	58.2	3	174.7	3	174.7

(Continued)

Table 5. (Continued)

GENE ID	3'UTR_length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
<i>HDKNOX1</i>	401	4	100.3	7	57.3	2	200.5
<i>HYD</i>	381	9	42.3	8	47.6	2	190.5
<i>IAA</i>	387	9	43.0	5	77.4	3	129.0
<i>KNOX2</i>	677	14	48.4	10	67.7	6	112.8
<i>KOR</i>	573	11	52.1	14	40.9	15	38.2
<i>LAC</i>	353	12	29.4	10	35.3	-	
<i>LAC2</i>	544	10	54.4	7	77.7	1	544.0
<i>LBD</i>	259	2	129.5	2	129.5	-	
<i>LEAFY</i>	0	na	na	na	na	na	na
<i>LIM1</i>	446	4	111.5	1	446.0	-	
<i>MAN</i>	526	9	58.4	9	58.4	2	263.0
<i>MAX</i>	530	8	66.3	11	48.2	5	106.0
<i>MIBP1</i>	398	15	26.5	13	30.6	5	79.6
<i>MTS</i>	349	9	38.8	9	38.8	6	58.2
<i>MUR3</i>	0	na	na	na	na	na	na
<i>MYB1</i>	432	6	72.0	7	61.7	-	
<i>MYB2</i>	62	-		-		-	
<i>NAM1</i>	704	15	46.9	14	50.3	1	704.0
<i>PAAPA</i>	539	20	27.0	7	77.0	-	
<i>PAE</i>	328	5	65.6	7	46.9	-	
<i>PAL</i>	702	26	27.0	19	36.9	8	87.8
<i>PG</i>	516	18	28.7	21	24.6	1	516.0
<i>PIP1</i>	337	7	48.1	9	37.4	5	67.4
<i>PL</i>	0	na	na	na	na	na	na
<i>POX1</i>	350	10	35.0	8	43.8	3	116.7
<i>PTM5</i>	0	na	na	na	na	na	na
<i>RAB</i>	340	6	56.7	6	56.7	7	48.6
<i>RNS</i>	154	3	51.3	4	38.5	6	25.7
<i>ROP1</i>	455	8	56.9	5	91.0	3	151.7
<i>SAMS</i>	316	7	45.1	6	52.7	6	52.7
<i>SBP1</i>	300	6	50.0	5	60.0	9	33.3
<i>SCD</i>	293	5	58.6	3	97.7	2	146.5
<i>SND1</i>	266	1	266.0	1	266.0	-	
<i>STM</i>	551	7	78.7	11	50.1	1	551.0
<i>SuSy1</i>	311	11	28.3	3	103.7	5	62.2
<i>TUA1</i>	337	4	84.3	4	84.3	4	84.3
<i>UBILIG</i>	192	2	96.0	2	96.0	-	
<i>UGDH</i>	440	7	62.9	5	88.0	10	44.0
<i>UGT</i>	266	5	53.2	6	44.3	4	66.5
<i>UXS1</i>	74	2	37.0	2	37.0	2	37.0
<i>VND6</i>	1798	13	138.3	14	128.4	-	
<i>VND7</i>	451	2	225.5	1	451.0	-	
<i>WND1</i>	298	4	74.5	5	59.6	2	149.0
<i>WRKY1</i>	224	1	224.0	1	224.0	1	224.0
<i>WUS1</i>	0	na	na	na	na	na	na

(Continued)

Table 5. (Continued)

GENE ID	3'UTR_length (bp)	<i>E. tereticornis</i>		<i>E. camaldulensis</i>		<i>E. grandis</i>	
		No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)	No. of SNVs	SNV frequency (bp/SNV)
<i>XCP</i>	403	15	26.9	13	31.0	2	201.5
<i>XTH</i>	210	5	42.0	6	35.0	1	210.0
<i>XYL</i>	286	29	9.9	5	57.2	3	95.3
<i>Znf1</i>	249	-	-	-	-	1	249.0
<b>Total</b>	<b>30768</b>	<b>539</b>	<b>86.61*</b>	<b>481</b>	<b>100.23*</b>	<b>206</b>	<b>176.08*</b>

na: Not applicable

\* denotes average SNV frequency

doi:10.1371/journal.pone.0116528.t005

C, Presence of InDels in Pair-wise comparison across *E. camaldulensis* and *E. tereticornis*. (XLSX)

### Acknowledgments

The authors acknowledge Dr. V. Sivakumar and Shri D.R.S. Sekar, Scientists, Institute of Forest Genetics and Tree Breeding, Coimbatore, India for providing the plant material for the study. The authors are grateful to Genotypic Technologies Private Limited, Bangalore, India for array design, library construction and analysis of the data. MGD acknowledges the funding support by Department of Biotechnology, Government of India under the DBT-CREST Awardship. VD acknowledges the Department of Biotechnology, Government of India for research fellowship.

Table 6. InDel frequency in three *Eucalyptus* species.

Region	No. of InDels	Length(bp)	InDel frequency(bp/InDel)
<b><i>E. tereticornis</i></b>			
5'UTR	99	16246	164.10
EXON	304	124987	411.14
3'UTR	115	30768	267.55
<b>Total</b>	<b>518</b>	<b>172001</b>	<b>332.05*</b>
<b><i>E. camaldulensis</i></b>			
5'UTR	89	16246	182.54
EXON	280	124987	446.38
3'UTR	110	30768	279.71
<b>Total</b>	<b>479</b>	<b>172001</b>	<b>359.08*</b>
<b><i>E. grandis</i></b>			
5'UTR	66	16246	246.15
EXON	259	124987	482.58
3'UTR	84	30768	366.29
<b>Total</b>	<b>409</b>	<b>172001</b>	<b>420.54*</b>

\* denotes average InDel frequency

doi:10.1371/journal.pone.0116528.t006

## Author Contributions

Conceived and designed the experiments: MGD KVK. Performed the experiments: MGD VD. Analyzed the data: MGD VD IA. Contributed reagents/materials/analysis tools: MGD IA KVK. Wrote the paper: MGD KVK.

## References

1. Brooker MIH (2000) A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Aust Syst Bot* 13: 79–148.
2. Doughty RW (2000) *The Eucalyptus*. A natural and commercial history of the gum tree. London, UK: Johns Hopkins University Press. 256 p.
3. Iglesias I, Wiltermann D (2009) Eucalyptologics Information Resources on Eucalypt Cultivation. GIT Forestry Consulting, Available: <http://www.git-forestry.com>. Accessed 2009 March 29.
4. Potts BM (2004) Genetic improvement of eucalypts. In: Burley J, Evans J, Youngquist JA, editors. *Encyclopedia of forest science*. Oxford, UK: Elsevier Science. pp. 1480–1490.
5. Kallarackal J, Somen CK (1997) An ecophysiological evaluation of the suitability of *Eucalyptus grandis* for planting in the tropics. *For Ecol Manag* 95:53–61. doi: [10.1016/S0378-1127\(97\)00004-2](https://doi.org/10.1016/S0378-1127(97)00004-2)
6. Kallarackal J, Somen CK, Rajesh N (2002) Studies on water use of six tropical eucalypt species in Kerala. In: Bagchi SK, Varghese M, Siddappa, editors. *Recent Eucalypt Research in India*. Coimbatore: Inst. Forest Genetics and Tree Breeding. pp. 94–115.
7. Grattapaglia D, O' Malley D, Sederoff R (1992) Multiple applications of RAPD markers to genetic analysis in *Eucalyptus* sp. In: *Resolving Tropical Forest Resources Concerns through Tree Improvement, Gene Conservation and Domestication of New Species*. Proceedings of the IUFRO meeting, Cali, Colombia: 436–450.
8. Gaiotto FA, Grattapaglia D (1997) Estimation of genetic variability in a breeding population of *Eucalyptus urophylla* using AFLP (Amplified Fragment Length Polymorphism) markers. *Silviculture and Improvement of Eucalyptus*. Proc. IUFRO Conference on Salvador, Colombo EMBRAPA-CNPQ. 2: 46–52.
9. Byrne M, Parrish TL, Moran GF (1998) Nuclear RFLP diversity in *Eucalyptus nitens*. *Heredity* 81: 225–233. doi: [10.1046/j.1365-2540.1998.00386.x](https://doi.org/10.1046/j.1365-2540.1998.00386.x)
10. Poltri SNM, Zelener N, Traverso JR, Gelid P, Hopp HE (2003) Selection of a seed orchard of *Eucalyptus dunnii* based on diversity criteria calculated using molecular markers. *Tree Physiol* 23: 625–632. doi: [10.1093/treephys/23.9.625](https://doi.org/10.1093/treephys/23.9.625) PMID: [12750055](https://pubmed.ncbi.nlm.nih.gov/12750055/)
11. Balasarayanan T, Chezian P, Kamalakannan R, Ghosh M, Yasodha R, et al. (2005) Determination of inter- and intra-species genetic relationships among six *Eucalyptus* species based on inter-simple sequence repeats (ISSR). *Tree Physiol*. 25: 1295–1302. doi: [10.1093/treephys/25.10.1295](https://doi.org/10.1093/treephys/25.10.1295) PMID: [16076778](https://pubmed.ncbi.nlm.nih.gov/16076778/)
12. Muro-Abad JI, Rocha RB, Cruz CD, Araujo EFD (2005) Obtainment of *Eucalyptus* spp. hybrids aided by molecular markers-SSR analysis. *Scientia Forestalis* 67: 53–63.
13. Chezian P, Yasodha R, Ghosh M (2010) Genetic diversity analysis in a seed orchard of *Eucalyptus tereticornis*. *New Forests* 40:85–99.
14. Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, et al. (2010) A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* 6: 16. doi: [10.1186/1746-4811-6-16](https://doi.org/10.1186/1746-4811-6-16) PMID: [20587069](https://pubmed.ncbi.nlm.nih.gov/20587069/)
15. Arumugasundaram S, Ghosh M, Veerasamy S, Ramasamy Y (2011) Species discrimination, population structure and linkage disequilibrium in *Eucalyptus camaldulensis* and *Eucalyptus tereticornis* using SSR markers. *PLoS ONE* 6(12): e28252. doi: [10.1371/journal.pone.0028252](https://doi.org/10.1371/journal.pone.0028252) PMID: [22163287](https://pubmed.ncbi.nlm.nih.gov/22163287/)
16. Kulheim C, Yeoh SH, Maintz J, Foley WJ, Moran GF (2009) Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* 10: 1–11. doi: [10.1186/1471-2164-10-1](https://doi.org/10.1186/1471-2164-10-1) PMID: [19121221](https://pubmed.ncbi.nlm.nih.gov/19121221/)
17. Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137:1121–1137. PMID: [7982566](https://pubmed.ncbi.nlm.nih.gov/7982566/)
18. Verhaegen D, Plomion C (1996) Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. *Genome* 39:1051–1061. doi: [10.1139/g96-132](https://doi.org/10.1139/g96-132) PMID: [18469954](https://pubmed.ncbi.nlm.nih.gov/18469954/)
19. Marques CM, Araujo JA, Ferreira JG, Whetten R, O'Malley DM, et al. (1998) AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. *Theor Appl Genet* 96:727–737. doi: [10.1007/s001220050795](https://doi.org/10.1007/s001220050795)



20. Thamarus K, Groom K, Murrell J, Byrne M, Moran G (2002) A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. *Theor Appl Genet* 104:379–387. doi: [10.1007/s001220100717](https://doi.org/10.1007/s001220100717) PMID: [12582710](https://pubmed.ncbi.nlm.nih.gov/12582710/)
21. Myburg AA, Griffin RA, Sederoff RR, Whetten RW (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. *Theor Appl Genet* 107:1028–1042. doi: [10.1007/s00122-003-1347-4](https://doi.org/10.1007/s00122-003-1347-4) PMID: [12838392](https://pubmed.ncbi.nlm.nih.gov/12838392/)
22. Verhaegen D, Plomion C, Gion JM, Poitel M, Costa P, et al. (1997) Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers. 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theor Appl Genet* 95: 597–608. doi: [10.1007/s001220050601](https://doi.org/10.1007/s001220050601) PMID: [12838392](https://pubmed.ncbi.nlm.nih.gov/12838392/)
23. Marques CM, Vasquez JK, Carocha VJ, Ferreira JG, O'Malley DM, et al. (1999) Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. *Theor Appl Genet* 99: 936–946. doi: [10.1007/s001220051400](https://doi.org/10.1007/s001220051400)
24. Shepherd M, Chaparro JX, Teasdale R (1999) Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. *Theor Appl Genet* 99: 1207–1215. doi: [10.1007/s001220051326](https://doi.org/10.1007/s001220051326)
25. Junghans D, Alfenas AC, Brommonschenkel SH, Oda S, Mello EJ, et al. (2003) Resistance to rust in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. *Theor Appl Genet* 108: 175–180. doi: [10.1007/s00122-003-1415-9](https://doi.org/10.1007/s00122-003-1415-9) PMID: [14504745](https://pubmed.ncbi.nlm.nih.gov/14504745/)
26. Kirst M, Myburg AA, De Leon JPG, Kirst ME, Scott J, et al (2004) Coordinated genetic regulation of growth and lignin revealed by Quantitative Trait Locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. *Plant Physiol* 135: 2368–2378 doi: [10.1104/pp.103.037960](https://doi.org/10.1104/pp.103.037960) PMID: [15299141](https://pubmed.ncbi.nlm.nih.gov/15299141/)
27. Teixeira J, Missiaggia A, Dias D, Scarpinati E, Viana J, et al. (2011) QTL analyses of drought tolerance in *Eucalyptus* under two contrasting water regimes. *BMC Proc.* 5 (Suppl 7): P40. doi: [10.1186/1753-6561-5-s7-p40](https://doi.org/10.1186/1753-6561-5-s7-p40)
28. Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, et al. (2012) Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. *Tree Genetics & Genomes* 8: 463–508. doi: [10.1007/s11295-012-0491-x](https://doi.org/10.1007/s11295-012-0491-x)
29. Thumma BR, Nolan MR, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171:1257–1265. doi: [10.1534/genetics.105.042028](https://doi.org/10.1534/genetics.105.042028) PMID: [16085705](https://pubmed.ncbi.nlm.nih.gov/16085705/)
30. Southerton SG, MacMillan CP, Bell JC, Bhuiyan N, Dowries G, et al. (2010) Association of allelic variation in xylem genes with wood properties in *Eucalyptus nitens*. *Australian Forestry* 73: 259–264. doi: [10.1080/00049158.2010.10676337](https://doi.org/10.1080/00049158.2010.10676337)
31. Kulheim C, Yeoh SH, Wallis IR, Laffan S, Moran GF, et al. (2011) The molecular basis of quantitative variation in foliar secondary metabolites in *Eucalyptus globulus*. *New Phytologist* 191: 1041–53. doi: [10.1111/j.1469-8137.2011.03769.x](https://doi.org/10.1111/j.1469-8137.2011.03769.x) PMID: [21609332](https://pubmed.ncbi.nlm.nih.gov/21609332/)
32. Resende MDV, Resende MFR, Sansaloni CP, Petrolí CD, Missiaggia AA, et al. (2012) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol* 194: 116–128. doi: [10.1111/j.1469-8137.2011.04038.x](https://doi.org/10.1111/j.1469-8137.2011.04038.x) PMID: [22309312](https://pubmed.ncbi.nlm.nih.gov/22309312/)
33. Rengel D, Clemente HS, Servant F, Ladouce N, Paux E, et al. (2009) A new genomic resource dedicated to wood formation in *Eucalyptus*. *BMC Plant Biology* 9: 36. doi: [10.1186/1471-2229-9-36](https://doi.org/10.1186/1471-2229-9-36) PMID: [19327132](https://pubmed.ncbi.nlm.nih.gov/19327132/)
34. Keller G, Marchal T, San Clemente H, Navarro M, Ladouce N, et al. (2009) Development and functional annotation of an 11,303-EST collection from *Eucalyptus* for studies of cold tolerance. *Tree Genetics & Genomes* 5: 317–327. doi: [10.1186/1471-2229-9-36](https://doi.org/10.1186/1471-2229-9-36) PMID: [19327132](https://pubmed.ncbi.nlm.nih.gov/19327132/)
35. Novaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, et al. (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 9:312–326. doi: [10.1186/1471-2164-9-312](https://doi.org/10.1186/1471-2164-9-312) PMID: [18590545](https://pubmed.ncbi.nlm.nih.gov/18590545/)
36. Rasmussen-Poblete S, Valdes J, Gamboa MC, Valenzuela PDT, Krauskopf E (2008) Generation and analysis of an *Eucalyptus globulus* cDNA library constructed from seedlings subjected to low temperature conditions. *Electronic Journal of Biotechnology* 11: p1
37. Paux E, Carocha V, Marques C, Mendes de Sousa A, Borralho N, et al. (2005) Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New Phytol* 167: 89–100. doi: [10.1111/j.1469-8137.2005.01396.x](https://doi.org/10.1111/j.1469-8137.2005.01396.x) PMID: [15948833](https://pubmed.ncbi.nlm.nih.gov/15948833/)
38. Mizrachi E, Hefer C, Ranik M, Joubert F, Myburg AA (2010) *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* plantation tree produced by Illumina mRNA-Seq. *BMC Genomics* 11:681. doi: [10.1186/1471-2164-11-681](https://doi.org/10.1186/1471-2164-11-681) PMID: [21122097](https://pubmed.ncbi.nlm.nih.gov/21122097/)

39. Paiva JA, Prat E, Vautrin S, Santos MD, San-Clemente H, et al. (2011) Advancing *Eucalyptus* genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. *BMC Genomics* 12:137. doi: [10.1186/1471-2164-12-137](https://doi.org/10.1186/1471-2164-12-137) PMID: [21375742](https://pubmed.ncbi.nlm.nih.gov/21375742/)
40. Villar E, Klopp C, Noirot C, Novaes E, Kirst M, et al. (2011) RNA-Seq reveals genotype-specific molecular responses to water deficit in *Eucalyptus*. *BMC Genomics* 12:538. doi: [10.1186/1471-2164-12-538](https://doi.org/10.1186/1471-2164-12-538) PMID: [22047139](https://pubmed.ncbi.nlm.nih.gov/22047139/)
41. Thumma BR, Sharma N, Southerton SG (2012) Transcriptome sequencing of *Eucalyptus camaldulensis* seedlings subjected to water stress reveals functional single nucleotide polymorphisms and genes under selection. *BMC Genomics* 13:364. doi: [10.1186/1471-2164-13-364](https://doi.org/10.1186/1471-2164-13-364) PMID: [22853646](https://pubmed.ncbi.nlm.nih.gov/22853646/)
42. Klocko AL, Vining K, Amarasinghe V, Romanel E, Alves-Ferreira M, et al. (2013) Floral transcriptome of *Eucalyptus grandis*. In: Proceedings of Plant and Animal Genome XXI, held from January 11–16, 2013 at San Diego, CA.
43. Hefer C, Mizrachi E, Joubert F, Myburg A (2011) The *Eucalyptus* genome integrative explorer (EucGenIE): a resource for *Eucalyptus* genomics and transcriptomics. *BMC Proc.* 5 (Suppl 7): O49. doi: [10.1186/1753-6561-5-s7-o49](https://doi.org/10.1186/1753-6561-5-s7-o49)
44. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, et al. (2014) The genome of *Eucalyptus grandis*. *Nature* doi: [10.1038/nature13308](https://doi.org/10.1038/nature13308) PMID: [24919147](https://pubmed.ncbi.nlm.nih.gov/24919147/)
45. Hirakawa H, Nakamura Y, Kaneko T, Isobe S, Sakai H, et al. (2011) Survey of the genetic information carried in the genome of *Eucalyptus camaldulensis*. *Plant Biotechnology* 28: 471–480. doi: [10.1038/nature13308](https://doi.org/10.1038/nature13308) PMID: [24919147](https://pubmed.ncbi.nlm.nih.gov/24919147/)
46. Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet* 24: 133–141. doi: [10.1016/j.tig.2007.12.007](https://doi.org/10.1016/j.tig.2007.12.007) PMID: [18262675](https://pubmed.ncbi.nlm.nih.gov/18262675/)
47. Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135–1145. doi: [10.1038/nbt1486](https://doi.org/10.1038/nbt1486) PMID: [18846087](https://pubmed.ncbi.nlm.nih.gov/18846087/)
48. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, et al. (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7:111–118. doi: [10.1038/nmeth.1419](https://doi.org/10.1038/nmeth.1419) PMID: [20111037](https://pubmed.ncbi.nlm.nih.gov/20111037/)
49. Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, et al. (2009) Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat Protocols* 4: 960–974. doi: [10.1038/nprot.2009.68](https://doi.org/10.1038/nprot.2009.68) PMID: [19478811](https://pubmed.ncbi.nlm.nih.gov/19478811/)
50. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  Method. *Methods* 25: 402–408. doi: [10.1006/meth.2001.1262](https://doi.org/10.1006/meth.2001.1262) PMID: [11846609](https://pubmed.ncbi.nlm.nih.gov/11846609/)
51. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
52. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10:R25. doi: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25) PMID: [19261174](https://pubmed.ncbi.nlm.nih.gov/19261174/)
53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
54. Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, et al. (2001) A transcriptional roadmap to wood formation. *Proc Natl Acad Sci USA* 98: 14732–14737. doi: [10.1073/pnas.261293398](https://doi.org/10.1073/pnas.261293398) PMID: [11724959](https://pubmed.ncbi.nlm.nih.gov/11724959/)
55. Harakava R (2005) Genes encoding enzymes of the lignin biosynthesis pathway in *Eucalyptus*. *Genet Mol Biol* 28: 601–607. doi: [10.1590/S1415-47572005000400015](https://doi.org/10.1590/S1415-47572005000400015)
56. Shi R, Sun YH, Li Q, Heber S, Sederoff R, et al. (2010) Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol* 51:144–163. doi: [10.1093/pcp/pcp175](https://doi.org/10.1093/pcp/pcp175) PMID: [19996151](https://pubmed.ncbi.nlm.nih.gov/19996151/)
57. Yang X, Ye CY, Bisaria A, Tuskan GA, Kalluri UC (2011) Identification of candidate genes in *Arabidopsis* and *Populus* cell wall biosynthesis using text-mining, co-expression network analysis and comparative genomics. *Plant Sci* 181: 675–687. doi: [10.1016/j.plantsci.2011.01.020](https://doi.org/10.1016/j.plantsci.2011.01.020) PMID: [21958710](https://pubmed.ncbi.nlm.nih.gov/21958710/)
58. Ruprecht C, Mutwil M, Saxe F, Eder M, Nikoloski Z, et al. (2011) Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Front Plant Sci* 2: 23. doi: [10.3389/fpls.2011.00023](https://doi.org/10.3389/fpls.2011.00023) PMID: [22639584](https://pubmed.ncbi.nlm.nih.gov/22639584/)
59. Wong MM, Cannon CH, Wickneswari R (2011) Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via *de novo* transcriptome sequencing. *BMC Genomics* 12:342. doi: [10.1186/1471-2164-12-342](https://doi.org/10.1186/1471-2164-12-342) PMID: [21729267](https://pubmed.ncbi.nlm.nih.gov/21729267/)

60. Kirst M, Johnson AF, Baucom C, Ulrich E, Hubbard K, et al. (2003) Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. Proc Natl Acad Sci USA 100: 7383–7388. doi: [10.1073/pnas.1132171100](https://doi.org/10.1073/pnas.1132171100) PMID: [12771380](https://pubmed.ncbi.nlm.nih.gov/12771380/)
61. Park S, Oh S, Han KH (2004) Large-scale computational analysis of poplar ESTs reveals the repertoire and unique features of expressed genes in the poplar genome. Mol Breed 14: 429–440. doi: [10.1007/s11032-005-0603-5](https://doi.org/10.1007/s11032-005-0603-5)
62. Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, et al. (1998) Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. Proc Natl Acad Sci USA 95:13,330–13,335. doi: [10.1073/pnas.95.22.13330](https://doi.org/10.1073/pnas.95.22.13330) PMID: [9789088](https://pubmed.ncbi.nlm.nih.gov/9789088/)
63. Lenhard M, Jürgens G, Laux T (2002) The WUSCHEL and SHOOTMERISTEMLESS genes fulfill complementary roles in *Arabidopsis* shoot meristem regulation. Development 129: 3195–3206. PMID: [12070094](https://pubmed.ncbi.nlm.nih.gov/12070094/)
64. Cseke LJ, Zheng J, Podila GK (2003) Characterization of *PTM5* in aspen trees: a MADS-box gene expressed during woody vascular development. Gene 318: 55–67. doi: [10.1016/s0378-1119\(03\)00765-0](https://doi.org/10.1016/s0378-1119(03)00765-0) PMID: [14585498](https://pubmed.ncbi.nlm.nih.gov/14585498/)
65. Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, et al. (2005) Transcription switches for protoxylem and metaxylem vessel formation. Genes Dev 19: 1855–1860. doi: [10.1101/gad.1331305](https://doi.org/10.1101/gad.1331305) PMID: [16103214](https://pubmed.ncbi.nlm.nih.gov/16103214/)
66. Zhong R, Demura T, Ye ZH (2006) *SND1*, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. Plant Cell 18: 3158–3170. doi: [10.1105/tpc.106.047399](https://doi.org/10.1105/tpc.106.047399) PMID: [17114348](https://pubmed.ncbi.nlm.nih.gov/17114348/)
67. Hu R, Qi G, Kong Y, Kong D, Gao Q, et al. (2010) Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. BMC Plant Biology 10:145 doi: [10.1186/1471-2229-10-145](https://doi.org/10.1186/1471-2229-10-145) PMID: [20630103](https://pubmed.ncbi.nlm.nih.gov/20630103/)
68. Ohashi-Ito K, Fukuda H (2010) Transcriptional regulation of vascular cell fates. Curr Opin Plant Biol 13: 670–676. doi: [10.1016/j.pbi.2010.08.011](https://doi.org/10.1016/j.pbi.2010.08.011) PMID: [20869293](https://pubmed.ncbi.nlm.nih.gov/20869293/)
69. Jensen JK, Kim H, Cocuron JC, Orler R, Ralph J, et al. (2011) The DUF579 domain containing proteins IRX15 and IRX15-L affect xylan synthesis in *Arabidopsis*. Plant J 66: 387–400. doi: [10.1111/j.1365-313X.2010.04475.x](https://doi.org/10.1111/j.1365-313X.2010.04475.x) PMID: [21288268](https://pubmed.ncbi.nlm.nih.gov/21288268/)
70. Li E, Bhargava A, Qiang W, Friedmann MC, Forneris N, et al. (2012) The Class II KNOX gene *KNAT7* negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus*. New Phytol 194: 102–115. doi: [10.1111/j.1469-8137.2011.04016.x](https://doi.org/10.1111/j.1469-8137.2011.04016.x) PMID: [22236040](https://pubmed.ncbi.nlm.nih.gov/22236040/)
71. Andersson-Gunnerås S, Hellgren JM, Björklund S, Regan S, Moritz T, et al. (2003) Asymmetric expression of a poplar ACC oxidase controls ethylene production during gravitational induction of tension wood. Plant J 34: 339–349. doi: [10.1046/j.1365-313x.2003.01727.x](https://doi.org/10.1046/j.1365-313x.2003.01727.x) PMID: [12713540](https://pubmed.ncbi.nlm.nih.gov/12713540/)
72. Kalluri UC, DiFazio SP, Brunner AM, Tuskan GA (2007) Genome-wide analysis of *Aux/IAA* and *ARF* gene families in *Populus trichocarpa*. BMC Plant Biology 7:59 doi: [10.1186/1471-2229-7-59](https://doi.org/10.1186/1471-2229-7-59) PMID: [17986329](https://pubmed.ncbi.nlm.nih.gov/17986329/)
73. Nieminen K, Immanen J, Laxell M, Kauppinen L, Tarkowski P, et al. (2008) Cytokinin signaling regulates cambial development in poplar. Proc Natl Acad Sci USA 105: 20032–20037. doi: [10.1073/pnas.0805617106](https://doi.org/10.1073/pnas.0805617106) PMID: [19064928](https://pubmed.ncbi.nlm.nih.gov/19064928/)
74. Mauriat M, Moritz T (2009) Analyses of GA20ox- and GID1-over-expressing aspen suggest that gibberellins play two distinct roles in wood formation. Plant J 58: 989–1003. doi: [10.1111/j.1365-313X.2009.03836.x](https://doi.org/10.1111/j.1365-313X.2009.03836.x) PMID: [19228336](https://pubmed.ncbi.nlm.nih.gov/19228336/)
75. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, et al. (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 10: R32 doi: [10.1186/gb-2009-10-3-r32](https://doi.org/10.1186/gb-2009-10-3-r32) PMID: [19327155](https://pubmed.ncbi.nlm.nih.gov/19327155/)
76. Paszkiewicz K, Studholme DJ (2012) High-throughput sequencing data analysis software: current state and future developments. In: Rodriguez-Ezpeleta N, Hackenberg M, Aransay AM, editors. Bioinformatics for high throughput sequencing. New York: Springer Science. pp. 231–248.
77. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53–59. doi: [10.1038/nature07517](https://doi.org/10.1038/nature07517) PMID: [18987734](https://pubmed.ncbi.nlm.nih.gov/18987734/)
78. Hendre PS, Kamalakannan R, Varghese M (2012) High-throughput and parallel SNP discovery in selected candidate genes in *Eucalyptus camaldulensis* using Illumina NGS platform. Plant Biotechnol J 10: 646–656. doi: [10.1111/j.1467-7652.2012.00699.x](https://doi.org/10.1111/j.1467-7652.2012.00699.x) PMID: [22607345](https://pubmed.ncbi.nlm.nih.gov/22607345/)
79. Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. Trends Plant Sci. 16: 282–288. doi: [10.1016/j.tplants.2011.02.006](https://doi.org/10.1016/j.tplants.2011.02.006) PMID: [21439889](https://pubmed.ncbi.nlm.nih.gov/21439889/)

80. Winfield MO, Wilkinson PA, Allen AM, Barker GL, Coghill JA, et al. (2012) Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnol J* 10:733–742. doi: [10.1111/j.1467-7652.2012.00713.x](https://doi.org/10.1111/j.1467-7652.2012.00713.x) PMID: [22703335](https://pubmed.ncbi.nlm.nih.gov/22703335/)
81. Trick M, Adamski NM, Mugford SG, Jiang CC, Febrer M, et al. (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biology* 12:14. doi: [10.1186/1471-2229-12-14](https://doi.org/10.1186/1471-2229-12-14) PMID: [22280551](https://pubmed.ncbi.nlm.nih.gov/22280551/)
82. Saintenac C, Jiang D, Akhunov ED (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biology* 12: R88. doi: [10.1186/gb-2011-12-9-r88](https://doi.org/10.1186/gb-2011-12-9-r88) PMID: [21917144](https://pubmed.ncbi.nlm.nih.gov/21917144/)
83. Kaya HB, Cetin O, Kaya H, Sahin M, Sefer F, et al. (2013) SNP Discovery by Illumina-based transcriptome sequencing of the olive and the genetic characterization of Turkish olive genotypes revealed by AFLP, SSR and SNP markers. *PLoS ONE* 8: e73674. doi: [10.1371/journal.pone.0073674](https://doi.org/10.1371/journal.pone.0073674) PMID: [24058483](https://pubmed.ncbi.nlm.nih.gov/24058483/)
84. Jupe F, Witek K, Verweij W, Iiwka J, Pritchard L, et al. (2013) Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J* 76: 530–544. doi: [10.1111/tpj.12307](https://doi.org/10.1111/tpj.12307) PMID: [23937694](https://pubmed.ncbi.nlm.nih.gov/23937694/)
85. Howe GT, Yu J, Knaus B, Cronn R, Kolpak S, et al. (2013) A SNP resource for Douglas-fir: *de novo* transcriptome assembly and SNP detection and validation. *BMC Genomics* 14:137 doi: [10.1186/1471-2164-14-137](https://doi.org/10.1186/1471-2164-14-137) PMID: [23445355](https://pubmed.ncbi.nlm.nih.gov/23445355/)
86. Wu X, Ren C, Joshi T, Vuong T, Xu D, et al. (2010) SNP discovery by high-throughput sequencing in soybean. *BMC Genomics* 11:469 doi: [10.1186/1471-2164-11-469](https://doi.org/10.1186/1471-2164-11-469) PMID: [20701770](https://pubmed.ncbi.nlm.nih.gov/20701770/)
87. Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, et al. (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics* 11:38 doi: [10.1186/1471-2164-11-38](https://doi.org/10.1186/1471-2164-11-38) PMID: [20078886](https://pubmed.ncbi.nlm.nih.gov/20078886/)
88. Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, et al. (2012) Genome-Wide SNP detection, validation, and development of an 8K SNP array for Apple. *PLoS ONE* 7(2): e31745. doi: [10.1371/journal.pone.0031745](https://doi.org/10.1371/journal.pone.0031745) PMID: [22363718](https://pubmed.ncbi.nlm.nih.gov/22363718/)
89. Neves LG, Davis JM, Barbazuk WB, Kirst M (2014) A High-Density gene map of Loblolly Pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3 (Bethesda)* 4: 29–37. doi: [10.1534/g3.113.008714](https://doi.org/10.1534/g3.113.008714) PMID: [24192835](https://pubmed.ncbi.nlm.nih.gov/24192835/)
90. Kenny EM, Cormican P, Gilks WP, Gates AS, O'Dushlaine CT, et al. (2011) Multiplex target enrichment using DNA indexing for ultra-high throughput SNP detection. *DNA Res* 18: 31–38. doi: [10.1093/dnares/dsq029](https://doi.org/10.1093/dnares/dsq029) PMID: [21163834](https://pubmed.ncbi.nlm.nih.gov/21163834/)
91. Mokry M, Feitsma H, Nijman IJ, de Bruijn E, van der Zaag PJ, et al. (2010) Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. *Nucleic Acids Res* 38:e116. doi: [10.1093/nar/gkq072](https://doi.org/10.1093/nar/gkq072) PMID: [20164091](https://pubmed.ncbi.nlm.nih.gov/20164091/)
92. Tennessen JA, Govindarajulu R, Liston A, Ashman TL (2013) Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, *Fragaria vesca* ssp. *bracteata* (Rosaceae). *G3 (Bethesda)* 3:1341–1351. doi: [10.1534/g3.113.006288](https://doi.org/10.1534/g3.113.006288) PMID: [23749450](https://pubmed.ncbi.nlm.nih.gov/23749450/)
93. Zhou L, Holliday JA (2012) Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC Genomics* 13:703. doi: [10.1186/1471-2164-13-703](https://doi.org/10.1186/1471-2164-13-703) PMID: [23241106](https://pubmed.ncbi.nlm.nih.gov/23241106/)
94. Grattapaglia D, Silva-Junior OB, Kirst M, de Lima BM, Faria DA, et al. (2011) High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC Plant Biol* 11:65. doi: [10.1186/1471-2229-11-65](https://doi.org/10.1186/1471-2229-11-65) PMID: [21492434](https://pubmed.ncbi.nlm.nih.gov/21492434/)
95. Allen AM, Barker GL, Berry ST, Coghill JA, Gwilliam R, et al. (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnol J* 9:1086–1099. doi: [10.1111/j.1467-7652.2011.00628.x](https://doi.org/10.1111/j.1467-7652.2011.00628.x) PMID: [21627760](https://pubmed.ncbi.nlm.nih.gov/21627760/)
96. Feltus AF, Wan J, Schulze SR, Estill JC, Jiang N, et al. (2004) An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. *Genome Res* 14: 1812–1819. doi: [10.1101/gr.2479404](https://doi.org/10.1101/gr.2479404) PMID: [15342564](https://pubmed.ncbi.nlm.nih.gov/15342564/)
97. Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. *Curr Opin Plant Biol* 12:218–222. doi: [10.1016/j.pbi.2008.12.007](https://doi.org/10.1016/j.pbi.2008.12.007) PMID: [19185530](https://pubmed.ncbi.nlm.nih.gov/19185530/)

98. Byers RL, Harker DB, Yourstone SM, Maughan PJ, Udall JA (2012) Development and mapping of SNP assays in allotetraploid cotton. *Theor Appl Genet* 124: 1201–1214. doi: [10.1007/s00122-011-1780-8](https://doi.org/10.1007/s00122-011-1780-8) PMID: [22252442](https://pubmed.ncbi.nlm.nih.gov/22252442/)
99. Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, et al. (2011) Single nucleotide polymorphism discovery in elite north American potato germplasm. *BMC Genomics* 12:302. doi: [10.1186/1471-2164-12-302](https://doi.org/10.1186/1471-2164-12-302) PMID: [21658273](https://pubmed.ncbi.nlm.nih.gov/21658273/)
100. Zhang X, Borevitz JO (2009) Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics* 182: 943–954. doi: [10.1534/genetics.109.103499](https://doi.org/10.1534/genetics.109.103499) PMID: [19474198](https://pubmed.ncbi.nlm.nih.gov/19474198/)
101. Jones E, Chu WC, Ayele M, Ho J, Bruggeman E, et al. (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.) germplasm. *Molecular Breeding* 24: 165–176. doi: [10.1007/s11032-009-9281-z](https://doi.org/10.1007/s11032-009-9281-z)
102. Muchero W, Diop N, Bhat P, Fenton R, Wanamaker S, et al. (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. *Proc Natl Acad Sci USA* 106:18159–18164. doi: [10.1073/pnas.0905886106](https://doi.org/10.1073/pnas.0905886106) PMID: [19826088](https://pubmed.ncbi.nlm.nih.gov/19826088/)
103. Pavy N, Pelgas B, Beauseigle SP, Blais S, Gagnon F, et al. (2008) Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC genomics* 9(1):21. doi: [10.1186/1471-2164-9-21](https://doi.org/10.1186/1471-2164-9-21) PMID: [18205909](https://pubmed.ncbi.nlm.nih.gov/18205909/)
104. Eckert A, Pande B, Ersoz E, Wright M, Rashbrook V, et al. (2009) High throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genetics & Genomes* 5: 225–234. doi: [10.1007/s11295-008-0183-8](https://doi.org/10.1007/s11295-008-0183-8)
105. Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, et al. (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. *New Phytol* 188: 515–532. doi: [10.1111/j.1469-8137.2010.03415.x](https://doi.org/10.1111/j.1469-8137.2010.03415.x). PMID: [20831625](https://pubmed.ncbi.nlm.nih.gov/20831625/)
106. Lima BM, Silva-Junior OB, Faria DA, Mamani EMC, Pappas GJ, et al. (2011) Assessment of SNPs for linkage mapping in *Eucalyptus*: construction of a consensus SNP/microsatellite map from two unrelated pedigrees. *BMC Proc.* 5 (Suppl 7): P31. doi: [10.1186/1753-6561-5-s7-p31](https://doi.org/10.1186/1753-6561-5-s7-p31)
107. Sexton T (2011) Candidate gene SNP discovery, genotyping and association with wood quality traits in *Eucalyptus pilularis* (blackbutt). PhD thesis, Southern Cross University, Lismore, NSW. Available: <http://epubs.scu.edu.au/theses/285>. Accessed 2014 June 30.
108. Thavamanikumar S, McManus, L J, Tibbits JFG, Bossinger G (2011) The significance of Single Nucleotide Polymorphisms (SNPs) in 'Eucalyptus globulus' Breeding Programs. *Australian Forestry* 74: 23–29. doi: [10.1080/00049158.2011.10676342](https://doi.org/10.1080/00049158.2011.10676342)
109. Singh P, Mizrahi E, Myburg Z (2014) Genetic load and allelic imbalance estimated in *Eucalyptus* hybrids using RNAseq. *Proc Plant & Animal Genome XXII*, held on January 10–15, 2014 at San Diego, CA.
110. Dantec LL, Chagne D, Pot D, Cantin O, Garnier-Gere P, et al. (2004) Automated SNP detection in expressed sequence tags: statistical considerations and application to maritime pine sequences. *Plant Mol Biol* 54: 461–470. doi: [10.1023/B:PLAN.0000036376.11710.6f](https://doi.org/10.1023/B:PLAN.0000036376.11710.6f) PMID: [15284499](https://pubmed.ncbi.nlm.nih.gov/15284499/)
111. Lijavetzky D, Cabezas JA, Ibañez A, Rodríguez V, Martínez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* 8: 424. doi: [10.1186/1471-2164-8-424](https://doi.org/10.1186/1471-2164-8-424) PMID: [18021442](https://pubmed.ncbi.nlm.nih.gov/18021442/)
112. Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, et al. (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* 3: 19. doi: [10.1186/1471-2156-3-19](https://doi.org/10.1186/1471-2156-3-19) PMID: [12366868](https://pubmed.ncbi.nlm.nih.gov/12366868/)
113. Li Y, Haseneyer G, Schon C-C, Ankerst D, Korzun V, et al. (2011) High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biol* 11: 6. doi: [10.1186/1471-2229-11-6](https://doi.org/10.1186/1471-2229-11-6) PMID: [21219606](https://pubmed.ncbi.nlm.nih.gov/21219606/)
114. Britten RJ, Rowen L, Williams J, Cameron RA (2003) Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci USA* 100: 4661–4665. doi: [10.1073/pnas.0330964100](https://doi.org/10.1073/pnas.0330964100) PMID: [12672966](https://pubmed.ncbi.nlm.nih.gov/12672966/)
115. Singh TR, Gupta A, Riju A, Mahalaxmi M, Seal A, et al. (2011) Computational identification and analysis of single nucleotide polymorphisms and insertions/deletions in expressed sequence tag data of *Eucalyptus*. *J Genet* 90: e34–e38. PMID: [21873771](https://pubmed.ncbi.nlm.nih.gov/21873771/)
116. Shao H, Bellos E, Yin H, Liu X, Zou J, et al. (2013) A population model for genotyping indels from next-generation sequence data. *Nucleic Acids Res* 41(3):e46. doi: [10.1093/nar/gks1143](https://doi.org/10.1093/nar/gks1143) PMID: [23221639](https://pubmed.ncbi.nlm.nih.gov/23221639/)

117. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, et al. (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* 16(9):1182–1190. doi: [10.1101/gr.4565806](https://doi.org/10.1101/gr.4565806) PMID: [16902084](https://pubmed.ncbi.nlm.nih.gov/16902084/)
118. Ometto L, Stephan W, De Lorenzo D (2005) Insertion/deletion and nucleotide polymorphism data reveal constraints in *Drosophila melanogaster* introns and intergenic regions. *Genetics* 169:1521–1527. PMID: [15654088](https://pubmed.ncbi.nlm.nih.gov/15654088/)
119. Brandström M, Ellegren H (2007) The genomic landscape of short insertion and deletion polymorphisms in the chicken (*Gallus gallus*) genome: a high frequency of deletions in tandem duplicates. *Genetics* 176:1691–1701. doi: [10.1534/genetics.107.070805](https://doi.org/10.1534/genetics.107.070805) PMID: [17507681](https://pubmed.ncbi.nlm.nih.gov/17507681/)
120. Hayashi K, Yoshida H, Ashikawa I (2006) Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. *Theor Appl Genet* 113: 251–260. doi: [10.1007/s00122-006-0290-6](https://doi.org/10.1007/s00122-006-0290-6) PMID: [16791691](https://pubmed.ncbi.nlm.nih.gov/16791691/)
121. Salathia N, Lee HN, Sangster TA, Morneau K, Landry CR, et al. (2007) Indel arrays: an affordable alternative for genotyping. *Plant J* 51: 727–737. doi: [10.1111/j.1365-3113x.2007.03194.x](https://doi.org/10.1111/j.1365-3113x.2007.03194.x) PMID: [17645438](https://pubmed.ncbi.nlm.nih.gov/17645438/)
122. Ollitrault F, Terol J, Martin AA, Pina JA, Navarro L, et al. (2012) Development of indel markers from *Citrus clementina* (Rutaceae) BAC-end sequences and interspecific transferability in *Citrus*. *Am J Bot* 99: E268–E273. doi: [10.3732/ajb.1100569](https://doi.org/10.3732/ajb.1100569) PMID: [22733984](https://pubmed.ncbi.nlm.nih.gov/22733984/)
123. Moghaddam SM, Song Q, Mamidi S, Schmutz J, Lee R, et al. (2014) Developing market class specific InDel markers from next generation sequence data in *Phaseolus vulgaris* L. *Front Plant Sci* 5:185. doi: [10.3389/fpls.2014.00185](https://doi.org/10.3389/fpls.2014.00185) PMID: [24860578](https://pubmed.ncbi.nlm.nih.gov/24860578/)
124. Perdereau AC, Douglas GC, Hodkinson TR, Kelleher CT (2013) High levels of variation in *Salix* lignocellulose genes revealed using poplar genomic resources. *Biotechnol Biofuels* 6(1):114. doi: [10.1186/1754-6834-6-114](https://doi.org/10.1186/1754-6834-6-114) PMID: [23924375](https://pubmed.ncbi.nlm.nih.gov/23924375/)
125. Meirmans PG, Lamothe M, Pierre P, Nathalie I (2007) Species-specific single nucleotide polymorphism markers for detecting hybridization and introgression in poplar. *Canadian Journal of Botany* 85: 1082–1091. doi: [10.1139/b07-069](https://doi.org/10.1139/b07-069)
126. Chu Y, Huang Q, Zhang B, Ding C, Su X (2014) Expression and molecular evolution of two *DREB1* genes in black poplar (*Populus nigra*). *PLoS One* 9(6):e98334. doi: [10.1371/journal.pone.0098334](https://doi.org/10.1371/journal.pone.0098334) PMID: [24887081](https://pubmed.ncbi.nlm.nih.gov/24887081/)
127. Schroeder H, Höltken A, Fladung M (2011) Chloroplast SNP-marker as powerful tool for differentiation of *Populus* species in reliable poplar breeding and barcoding approaches. *BMC Proc.* 5(Suppl 7): P56. doi: [10.1186/1753-6561-5-s7-p56](https://doi.org/10.1186/1753-6561-5-s7-p56)
128. Kong F, Wang X, Chen Y, Bian A, Xu J, et al. (2013) Analyzing the nucleotide variations within the Expressed Sequence Tags of loblolly Pine (*Pinus taeda*). *J Plant Biochem Physiol* 1:2.
129. Kumar S, Banks TW, Cloutier S (2012) SNP Discovery through Next-Generation Sequencing and its applications. *International Journal of Plant Genomics*. doi: [10.1155/2012/831460](https://doi.org/10.1155/2012/831460) PMID: [23227038](https://pubmed.ncbi.nlm.nih.gov/23227038/)
130. Nordborg M, Weigel D (2008) Next-generation genetics in plants. *Nature* 456: 720–723. doi: [10.1038/nature07629](https://doi.org/10.1038/nature07629) PMID: [19079047](https://pubmed.ncbi.nlm.nih.gov/19079047/)