



# Allelic variation in candidate genes associated with wood properties of cultivated poplars (*Populus*)

Zoltán Attila Köbölkuti<sup>1</sup>, Klára Cseke<sup>1</sup>, Attila Benke<sup>1</sup>,  
Mátyás Báder<sup>2</sup>, Attila Borovics<sup>1</sup> and Róbert Németh<sup>2</sup>

## Original Article

**Cite this article:** Köbölkuti ZA, Cseke K, Benke A, Báder M, Borovics A, and Németh R. 2019. Allelic variation in candidate genes associated with wood properties of cultivated poplars (*Populus*). *Biol. Fut.* 70, 286–294.

Received: 13 March 2019

Accepted: 26 October 2019

### Keywords:

SNP markers, *Populus*, wood property

<sup>1</sup>Department of Tree Breeding, National Agricultural Research and Innovation Centre, Forest Research Institute, Várkerület 30/A, Sárvár, 9600, Hungary

<sup>2</sup>Simonyi Karoly Faculty of Engineering, Wood Sciences and Applied Arts, University of Sopron, Bajcsy Zs. u. 4., 9400, Sopron, Hungary

DOI: [10.1556/019.70.2019.32](https://doi.org/10.1556/019.70.2019.32)

*Introduction:* Since *Populus* has veritable value as timber, plywood, pulp, and paper, genomic research should create the sound basis for further breeding toward desirable wood quality attributes. *Materials and methods:* In this study, we addressed the need for a research methodology that initially identifies and then characterize candidate genes encoding enzymes with wood property phenotypic traits, toward the aim of developing a genomics-based breeding technology. *Results:* On 23 different poplar species/hybrid samples, we successfully amplified 55 primers designed on *Populus trichocarpa* L. Considering the number of polymorphic sites, out of 73,206 bp, 51 SNPs and 31 indel events were found. Non-synonymous single base mutations could be detected in number of 30, 21 out of 164 sequences were the number of minimum recombination events and 41 significant pairwise comparisons between loci could be detected. *Discussion and conclusion:* Our results provide a roadmap for a future association genetic study between nucleotide diversity and precise evaluation of phenotype.

## INTRODUCTION

Poplars are of considerable economic value with a cultivation started in Europe, Asia, and the Middle East over hundreds of years ago. Breeding programs have focused on the production of interspecific hybrids, the first-generation hybrid vigor being the main purpose and aim, mostly because of the ease with which the species can be vegetatively propagated (Joshi et al., 2011 and references therein). As high-intensity cultivation of woody biomass crops has great economic potential, *Populus* is one of the first candidates for implementation of intensive plantation forestry. Hybrid poplars are important components of current energy portfolios for their significant amount of biomass, but they also provide a wide range of wood products, including industrial roundwood and poles, pulp and paper, reconstituted boards, plywood, veneer, sawn timber, packing crates, pallets, and furniture. The volume of high-density hardwoods on the market does not change significantly while demand is constantly increasing; thus, some additional material will be required soon. In the near future, densified poplars may be used for flooring or as furniture elements, complementing or replacing high-density hardwoods on the wood market (Rademacher et al., 2017).

Hungary lost 84% of its forest area after World War I. Mitigating the wood deficiency a massive afforestation program started after World War II, playing a significant role in the increase of its area. During this program, the foresters preferred, besides other fast-growing tree species like Scotch pine (*Pinus sylvestris* L.) and black locust (*Robinia pseudoacacia* L.), poplars primarily. Nowadays, hybrid poplars are the fifth in the ranking of forest tree species by covered area in the country; the total area of the poplar plantations being 112,970 hectares. The preferred clones are mostly Euramerican hybrids (hybrids between *Populus deltoides* and *Populus nigra* L.; *Populus* × *euramericana*), the importance of clones with other origin (e.g., Interamerican clones or balsam poplars) being considered as less significant. The most popular hybrid poplar clones are the following, in order by their covered area: Pannónia, I-214, Agathe-F.

Having significant economic interest, there is a demand for accelerated improvement strategies that aim at developing *Populus* species as wood-producing crops with

### Author for correspondence:

Zoltán Attila Köbölkuti

e-mail: [kobolkuti.zoltan@erti.naik.hu](mailto:kobolkuti.zoltan@erti.naik.hu)

Electronic supplementary material is available online at [www.akademai.com/doi/suppl/10.1556/019.70.2019.32](http://www.akademai.com/doi/suppl/10.1556/019.70.2019.32).

© 2019 The Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated. (SID\_1)

both improved trunk performance and specific exploitation characteristics. Thus, for innovative selection toward a specific structure, composition and properties of the raw material would be of great help the marker assisted breeding of the species.

Objectives of primary importance to the forestry and wood industry are the genetic control of traits, such as growth, adaptation to environmental stress, disease resistance, wood uniformity, specific gravity, and fiber quality. With classical breeding, these demands cannot be fulfilled within a reasonable time span because of the long generation time. Thus, to meet these needs, molecular breeding could provide the necessary tools to solve many of these problems. However, to apply genetic and molecular tools to optimize the production of woody materials, thorough understanding of the mechanisms that determine their properties is required.

In general, wood is manufactured by a succession of steps, including cell division, cell expansion, cell wall thickening (involving the biosynthesis and deposition of cellulose, hemicellulose, cell wall proteins, and lignin), programmed cell death, and heartwood formation. In the formation and modification of the matrix of wood cell walls, the xylem-specific glycosyltransferases belonging to carbohydrate enzyme families have an important role and *Populus* has approximately 1,600 genes encoding carbohydrate enzymes (Geisler-Lee et al., 2006).

Primary-walled wood cells of *Populus* are rich in pectins, xyloglucan, and cellulose, and the vascular cambium and adjacent radial expansion zone are the sites of highest expression of genes encoding wall-modifying enzymes (Mellerowicz & Sundberg, 2008).

Secondary wall deposition requires a reprogramming of wall biosynthesis by activating transcription factors responsible for inducing secondary wall programs (Zhong & Ye, 2007). The wood cell walls have multiple layers that differ in microfibril organization and ratios of cellulose to matrix (lignin, hemicellulose, and pectin) components. Expression analyses have shown that cellulose synthase proteins, different from those involved in primary wall biosynthesis, play major roles during secondary wall biosynthesis in xylem tissues (Mutwil et al., 2008). The microfibrils aggregate in the cell wall into macrofibrils, which are organized in tangential sectors. The macrofibril size seems to depend on the lignin and hemicellulose contents in the wall.

Important for mechanical properties of the wood fiber is the microfibril angle, which typically varies longitudinally and radially within the tree stem. An important role in determining microfibril angle in wood cells is the microtubules, this fact being supported by the highly expressed family of  $\alpha$ -tubulin and  $\beta$ -tubulin during xylem secondary wall formation (Oakley et al., 2007). Several microtubule-associated proteins putatively involved in a scaffolding between nascent microfibrils and the microtubule network through scaffold proteins has been identified, like the kinesin-like protein (Luo et al., 2016) and a katanin microtubule-severing protein (Sasaki et al., 2017). The fibers' and vessel elements' secondary cell wall composition is strongly determined by the hemicellulose biosynthesis (Ratke et al., 2015). The presence of hydrolytic and transglycosylating enzymes in secondary cell walls suggests that the matrix of secondary walls in some cases is also modified

(Mellerowicz & Sundberg, 2008). The tension wood development is also based on cell wall synthesis shifts between lignin and hemicellulose and cellulose biosynthesis (Sawada et al., 2018).

Poplar has a relatively small genome (550 Mb) (Bradshaw & Stettler, 1993). Genetic maps are available and over the years extensive quantitative trait loci mapping surveys have been conducted (Carletti et al., 2016). The expression of genes associated with some of the steps of the wood production process has been analyzed (Li et al., 2006), contributing to a better understanding of wood development. Still, the extensive breeding programs in poplar have led to a significant number of hybrids, and the characterization of different clones is necessary for the design of breeding strategies, but despite the ample genomic resources mostly microsatellite markers have been developed for this purpose. The number of studies concerning single nucleotide polymorphisms (SNPs) – variations in a single nucleotide, at a specific position in the genome, but only where are present to appreciable degree within a population – presumably in conjunction with the wood production process in poplar is not excessively large in the literature. For that reason, SNP markers connected with wood production encoding genes in our opinion would represent a meaningful tool to meet the breeding necessities.

Consequently, the main purposes of this study were: (a) to identify functional genes corresponding to the wood production process with use of available sequence data from public genome database; (b) To develop SNP markers for these gene fragments; (c) to obtain variable sequence sets from Hungarian poplar clones commonly used in breeding programs; and (d) to identify “clone specific” haplotypes in these target regions with possible use in further association genetic studies by correlating adequate levels of nucleotide diversity, linkage disequilibrium, and precise evaluation of phenotype from clonal or progeny testing.

## MATERIALS AND METHODS

### Marker development

For the development of markers putatively connected with the wood production process, a data mining was performed, from literature where proteins with possible role in the wood production process were selected. Table 1 presents these selected proteins.

Based on Table 1, a BLAST database was filled with sequences downloaded from NCBI database (<http://www.ncbi.nlm.nih.gov>, 07.03.2018), using the following criteria: first, conducting a search for EST sequences for these proteins that were annotated in the *Populus* genus. If such could not be found, the search was extended to “eudicots.” The contents of the BLAST database are presented in Supplementary Table S1.

The critical factor in developing PCR markers from EST sequences could be the presence of repetitive introns with different length in the genome. For the reason that these sequence repeats may disturb primer annealing during the PCR amplification of the selected sequences and in some cases the length of the sequence that should be amplified

Table 1. Selected proteins related to wood production based on literature data search

Nr.	Protein	Role	References
1	Caffeoyl-coenzyme A 0-methyltransferase	Lignification of the cell wall	Takabe et al. (2001)
2	Phenylcoumaran benzylic ether reductase	Protection against oxidative damage associated with lignifying cells	Ralph et al. (2009)
3	Caffeic acid/5 hydroxy-ferulic acid 0-methyltransferase	Role in modifying the composition of lignin; secondary xylem development	Halpin (2004)
4	s-Adenosyl-L-methionine synthetase	Catalyzes a key step in the biosynthesis of lignin	Vander Mijnsbrugge et al. (2000)
5	Lycine hydroxyl-methyltransferase	Role in secondary xylem differentiation	Villalobos et al. (2012)
6	K-transporters tetraethylammonium	Role in wood increment and the vessel size	Fromm (2010)
7	Ethylene high-affinity K <sup>+</sup> -transporters	Role in unloading K <sup>+</sup> from the xylem stream	Ache et al. (2010)
8	Shaker-like K channel protein (SKOR)	K <sup>+</sup> transporter, involved in the regulation of K <sup>+</sup> -dependent wood production	Arend et al. (2005)
9	KAT1-type channel protein	Associated with bud flush, stomatal guard cell movements, and control of transpirational water loss; role in K <sup>+</sup> -dependent xylogenesis	Langer et al. (2004)
10	HAK1 transporter protein	Controlling seasonal changes in wood formation	Ache et al. (2010)
11	PTORK channel protein	Regulation of xylem cell expansion, cambial reactivation after winter dormancy; control electrical properties of wood-forming cells	Fromm (2010)
12	PTK2 protein	Role in secondary xylem formation	Fromm and Lautner (2016)
13	PKUP transporter protein	Role in potassium nutrition; controlling seasonal changes in wood formation	Garcia and Zimmermann (2014)
14	Cyclophilin	Role in lignification during wood formation, in biotic and abiotic stress tolerance	Rajnikanth et al. (2007)
15	Blue copper protein	Role in lignin formation	Nersissian et al. (1998)
16	ADP-ribosylation factor I	Role in salinity stress tolerance; regulating leaf initiation, development, and shape	Drost et al. (2015)
17	ADP ribosylation factor II	Cell wall formation in the cambial region	Sterky et al. (1998)
18	Nodulin	Involved in microtubule-based movement in the developing xylem	Ranjan et al. (2010)
19	Laccase	Role in formation of cell wall structure, lignin content	Lu et al. (2013)
20	s-adenosyl-t-methionine	Role in developing xylem, lignin synthesis	Bedon and Legay (2011)
21	Elongation factor 1 $\alpha$	Function in primary cell wall synthesis; housekeeping, normalization of expression signals	Dharmawardhana et al. (2010)
22	14-3-3 like protein	Primary metabolism, hormone signaling, growth, and cell division	Denison et al. (2011)
23	Carbohydrate-active enzymes (GT2, GT8, GT14, GT31, GT43, GT47, and GT61)	Role in remodeling the cell wall matrix; role in xylogenesis	Geisler-lee et al. (2006)
24	Glicosylases (GH9, GH10, GH16, GH17, GH19, GH28, GH35, and GH51)	Role in xylem differentiation; auxin induced growth; secondary cell wall formation	Loziuk et al. (2017)

could be excessively long, the constructed EST BLAST database was searched using the BLAST Genomes toolkit (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the *Populus trichocarpa* (taxid 3694) database. After the search based on sequence homology, only those sequences were selected that had the highest BLAST hit by setting the thresholds lowest E-value at <0.001, greatest identity at >98%, and the high-scoring segment pair length between 100 and 1,000 bp. Following this step, in total, 55 sequences were selected for primer design. The design of the primers was performed with the Primer 3 Plus software (<https://primer3plus.com/cgi-bin/dev/primer3plus.cgi>) and the primer pairs were tested *in silico* using the Primer-BLAST toolkit (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>) to avoid simultaneous amplification of paralogous loci. The list of the selected sequences and designed primer pairs is presented in Supplementary Table S2.

#### Plant material

The experiment in the laboratory was conducted with a single sample/individual, collected in the NAIK ERTI *ex situ* gene collection plantation established in Sárvár-Bajti (Hungary). From several individuals, wood tissue samples were taken by an increment borer. In some cases, samples originated from leaf material, selected from the former NAIK ERTI DNA collection base. The following poplar species were involved into our investigation: *Populus nigra* L. (“Lassicsárda 7” and “Lébény 211”), *Populus deltoides* Marsh. (“S-1-54,” “S 307-24,” and “S 332-1”), *Populus alba* L. (“Villafranca”), *Populus tremula* L. (“105”), *Populus × canescens* Sm. (“65”), and *Populus grandidentata* Michx. (“227”). In addition to clones representing pure species or natural hybrids, cultivars bred and registered in Hungary with the highest aboveground biomass yield and wood quality (“Pannónia,” “Koltay,” and “Kopecky”); moreover, four promising new Hungarian clones (“Sv-656,” “Sv-761,” “Sv-874,” “Sv -890”) and several clones originated from Belgium, Italy, and Poland (“Beaupre,” “Raspalje,” “S-611-C,” “I-214,” “Kornik-21”) were collected. The name of each specimen, DNA identification code, the place of origin, type of tissue, and available parental combinations are presented in Supplementary Table S3.

#### DNA extraction and PCR amplification

About 1 cm<sup>2</sup> of a single leaf or 10 mg core was ground into powder in liquid nitrogen. Total DNA was extracted, following a modified alkyltrimethylammonium bromide protocol (Dumolin et al., 1995). As a result, 100 µl of high purity concentrated DNA solution was obtained. About 50 µl of the DNA solutions was stored as a reserve at –80 °C, for possible future studies. Oligonucleotide primers were synthesized by standard phosphoramidate chemistry at IDT (Integrated DNA Technologies, Bio-Science Ltd., Coralville, IA, USA).

For testing the molecular markers, with each primer pair on each DNA sample, a PCR test was carried out to evaluate the functionality of the designed primers. PCR was conducted by following a modified protocol of Isabel et al. (2013). The reactions contained 10 ng/µl template DNA,

1× reaction buffer (Promega GoTaq G2 Flexi, 5× reaction buffer with no magnesium), 2 mM of MgCl<sub>2</sub> (Promega), 15 µM of dNTPmix (Promega, 10 mM each), 0.5 polymerase (Promega GoTaq G2 Flexi, 5U/µl), and 0.13 µM of each primer (IDT) in a total volume of 15 µl. PCR was carried out in a Veriti Personal Thermocycler (Applied Biosystems, Waltham, MA, USA), with a pre-denaturation step at 95 °C for 3 min, followed by 35 cycles of 94 °C for 30 s, an annealing temperature for each primer combination at 50 °C for 45 s, extension on 72 °C for 90 s, and a final elongation at 72 °C for 10 min. The PCR conditions were considered optimal and the tested markers suitable for further sequence analysis in case of PCR-amplified sequential motifs appeared as a single band in the agarose gel. For markers with multiple bands, further optimization was performed by increasing the annealing temperature up to 55 °C.

PCR amplification products were analyzed on a 2% agarose gel (Roti Agarose, Roth GmbH, Karlsruhe, Germany), following electrophoresis with EPS 600 (Pharmacia Biotech, Rochester, NY, USA), in 1× TAE as electrophoresis buffer, stained with GelRed (Biotium Inc., Hayward, CA, USA) (Supplementary Figures S1–S3). From all products that appeared as a single band, eight selected by random were used to verify the previous annotation’s correctness and to detect potential SNP polymorphisms.

#### Sanger sequencing

Before sequencing, the purification of the products after PCR was performed by hydrolyzing the excess primers and dephosphorylated unincorporated dNTPs, in one step, with ExS-Pure enzymatic PCR cleanup kit (NimaGen BV, Nijmegen, The Netherlands), according to the manufacturer’s protocol. As a next step, the products resulted from the ExS-Pure process were sequenced, the direction of sequencing being chosen considering the best BLAST hit between the reverse and forward variant of each sequence. Considering the aforementioned direction, clone “Pannónia” was an exception, in case of which the process was performed in both directions. Sequencing was completed at Biomi Ltd. (Hungary).

#### Sequence analysis

Visual analysis and editing of the sequences with clearly defined set of regularly spaced peaks were performed using BioEdit Sequence Alignment Editor version 7.0.9.0 (Hall, 1999). A second analysis was performed in case of *ambiguity sequences*, where the analysis of the double traces was completed by Trace Recalling method (Tenney et al., 2007), with use of the CodonCode Aligner 8.0.2 (CodonCode Corporation, Centerville, MA, USA) software. After editing, the alignment of sequences was completed with CLC Genomic Workbench version 12.0-Viewing Mode (Qiagen Bioinformatics, Venlo, Netherlands), setting gap open cost at the value of 10, gap extension cost at the value of 1, and end gap cost “as any other,” and in a very accurate (slow) mode. Following sequence alignment, a final validation of all potential SNPs was performed visually on chromatograms again with use of BioEdit, and sites that differentiated *Populus* clones were identified. Number of

polymorphic sites, nucleotide diversity, number of insertions/deletions, the character of SNPs (synonymous or non-synonymous), minimum number of recombination events, identification of conserved DNA regions along the data set, and the number of significant pairwise comparisons between polymorphic sites of each sequence with statistical significance by Fisher's exact test were calculated using DNA Sequence Polymorphism version 6.10.01 (Rozas & Rozas, 1995) software. The chromosomal location of each gene fragment was determined by the Genome Data Viewer (<https://www.ncbi.nlm.nih.gov/genome/gdv/>) toolkit.

## RESULTS

The tested 55 primers on 23 different poplar DNA samples successfully amplified 49 sequences (success rate 89%), and 26 PCR products appeared as a single band in the electrophoresis gels (success rate: 47.27%; Supplementary Figures S1–S3). In most cases, the annealing temperature at 50 °C proved to be optimal, in case of four gene fragments (nos. 5, 9, 11, 39 in Supplementary Table S2), only by increasing the temperature up to 55 °C, the same result could be obtained. The following eight gene fragments were selected by random, for sequencing: blue copper binding protein 2, Kt 1, ptk 2/2, SAMS 1 in forward and COMT 3, COMT 4, CCoAOMT 4, SKOR 3 in reverse direction (for abbreviations, see Supplementary Table S4). Conclusive chromatograms were obtained in case of seven gene fragments, the “blue copper binding protein 2” marker being excluded from further analyses.

In case of CCoAOMT 4 marker's 13 samples (Pannónia, Kopecky, Koltay, Durvakérg, Sv-656, Sv-874, Sv-890, I-214, Raspalje, Beaupré, 227, 65, 105), considering the chromatograms, unfortunately not all traces' peaks were regularly spaced and of similar height, but clearly two bases were represented by each pair of overlapping peaks. To localize the insertion event with the aim to determine which sequence and where has been disrupted, we created a second ambiguity sequence, this ambiguity sequence being aligned to an assembled *Populus* genome sequence. The genomic sequence that best aligned to the ambiguity sequence was assumed to be the second sequence of the same sample. By this method, in case of these aforementioned 13 samples, two sequences were used as input for sequence alignment.

To verify the previous annotation's correctness, after editing and visual validation, we randomly selected a single sequence from each marker's amplified product, which we verified using BLASTN search. The putative similarity of the sequences was estimated according to the best BLAST hit, the results being summed in Supplementary Table S4.

By this method, on 22 poplar DNA samples each, fragments from seven different genes were successfully amplified and identified. After sequencing, COMT 3, ptk2/2, and SKOR 3 showed clear sequencing chromatograms only for PCR products of 21 samples each; therefore, totally, 164 sequences were analyzed. The results of the sequence analysis using DNA Sequence Polymorphism version 6.10.01 with the sequence alignment results (CLC Genomics Workbench version 12.0, Qiagen Bioinformatics) as inputs are presented in Table 2.

Table 2. Outputs of the sequence analysis with DnaSP v. 6.10.01 effectuated on seven sequences' 108 PCR products

Abbreviation	No. of sequences (samples) used	No. of sites	No. of polymorphic sites	No. of haplotypes	Nucleotide diversity	No. of indel events	No. of non-synonymous SNPs	No. of synonymous SNPs	Conserved DNA regions in the sequence	Minimum number of recombination events	Recombination has been detected between sites
COMT 3*	21	149	9	7	0.01531	14	0	9	No conserved region	1	(-2, 38)
COMT 4	22	114	12	13	0.04176	1	10	3	No conserved region	3	(15, 16) (16, 20) (20, 24)
Kt 1	22	164	7	8	0.01109	0	4	2	Regions 53-127	1	(-132, 164)
CCoAOMT 4**	34	166	17	10	0.01414	15	n.a.	n.a.	Regions 1-87	1	(57, 110)
ptk 2/2*	21	140	8	8	0.01439	1	4	4	No conserved region	1	(-78, 98)
SAMS 1	22	180	8	9	0.01502	0	7	1	No conserved region	4	(15, 58) (58, 82) (82, 97)
SKOR 3*	21	78	7	10	0.01831	0	7	0	No conserved region	10	(115, 139) (-53, 72)

Note. PCR: polymerase chain reaction; SNP: single nucleotide polymorphism.

\*Genes for which only 21 sequences were analyzed. \*\*The gene with duplicated samples due to double traces (for sequence abbreviation, see Supplementary Table S4).

In total, 51 SNPs were found out of 73,206 bp. Mutations ranged between 7 (Kt 1 and SKOR 3) and 17 (CCoAOMT 4). Nucleotide diversity values were between 0.04176 (COMT 4) and 0.01109 (Kt 1), on average 0.01857 (Table 2). In a second phase of the analysis, the number of indel events and the character of SNPs (synonymous or non-synonymous) were determined, the results being summed in Table 2.

Indel events were found in number of 31 out of 164 sequences. The detected value was the highest in case of CCoAOMT 4 and COMT 3, two sequences (COMT 4 and ptk 2/2) being characterized by one, whereas four gene fragments (Kt 1, SAMS 1, and SKOR 3) with no indel event. Synonymous SNPs were found in number of 21, ranging from 1 to 9 (for CCoAOMT 4, no data were available). The detected value was the highest in case of COMT 3 (9), while in case of SKOR 3, no synonymous SNP could be observed. COMT 3, COMT 4, ptk 2/2, SAMS 1, and SKOR 3 turned out to be with no conserved region, whereas Kt1 and CCoAOMT 4 presented conserved DNA regions in their sequence (Table 2). Non-synonymous single-base mutations could be detected ranging from 4 to 10, in number of 30, with the highest value of 10 in case of COMT 4 (for CCoAOMT 4, no data were available). All non-synonymous SNPs led to an amino acid exchange, no one caused an early stop codon. The minimum number of recombination events was in number of 21 out of 163 sequences, SKOR 3 being of special interest (Table 2). The number of significant pairwise comparisons with statistical significance by Fisher's exact test (significant at  $*.01 < p < .05$ ;  $**.001 < p < 0.01$ ;  $***p < .001$  level) was in number of 41, COMT 4 proved to be of special interest due to its highest value (16; Supplementary Table S5).

## DISCUSSION

Traditional breeding of poplar for its wood, a constitute of raw material for construction materials as well as many other products, can be enhanced by approaches based on marker-assisted selection. Advantages of the method include a reduced breeding cycle time, reduced cost of field testing, and increased intensity and efficiency of selection. Genetic markers are based on the variation of DNA sequences, and comparative sequencing is the ultimate method to detect variation within any DNA fragment. Certainly, today, it is already possible to analyze and to compare whole genomes or transcriptomes using high-throughput sequencing technologies. However, this is still with significant financial impact in examining a sufficient number of individuals. Considering these limitations, in this study, SNP markers were developed to obtain a start point of a further association genetic study, considering wood's and fibers' structural and chemical properties of several – already registered and also promising – clones cultivated in Hungary.

Our first goal was to identify SNP markers within candidate genes encoding cellulose and lignin biosynthetic enzymes, with an assumed role in wood property phenotypic traits. All candidate proteins have been chosen based on literature surveys suggesting an impact of the enzymes on either cell division, cell expansion, cell wall thickening involving cellulose, hemicellulose, cell wall proteins, lignin

biosynthesis and deposition, programmed cell death, and heartwood formation (Table 1). The NCBI EST database (<http://www.ncbi.nlm.nih.gov>) was used to determine corresponding sequences from the *Populus* genus. In several cases where sequences belonging to this genus could not be found, all available sequences of eudicots were verified by BLASTN search and used for primer design to amplify the corresponding genomic regions. From this point of view and also considering the fact that our amplified sequences belonging to *P. nigra*, *P. deltoides*, *P. alba*, and *P. tremula*, *P. grandidentata* species or *P. × canescens*, *P. × euramericana* and *P. × interamericana* hybrids could be identified by BLASTN search in *P. trichocarpa*, *P. kitakamiensis*, *P. fremontii* species, and the *P. tremula × P. tremuloides* hybrid (Supplementary Table S4), this study could be considered as a cross-species marker transferability test, and concerning this type of analyses, previous attempts were made. Transferability of EST markers among closely related species has been reported not only in crop species (Decroocq et al., 2003; Feng et al., 2009) but also in *Populus* (Du et al., 2013). Our – on *P. trichocarpa* sequences designed – 55 primers on 23 different poplar species/hybrid DNA samples successfully amplified sequences with a success rate of 89%, whereas PCR products with single band in the gel were registered in rate of 47.27%. These values could be considered somewhat congruent with the results of similar marker transferability tests. Peakall et al. (1996) used simple sequence repeat markers developed from soybean genome amplifying across *Glycine max* and the results showed that 65% of the markers can be amplified in cross-species, but only 3%–13% in cross-genera.

Our results suggest that such primers could work across a wider range of *Populus* species/hybrids. The persistence of homologous sequences in the genomic DNA of poplars provides further support for the use of *de novo* primers targeting important regions across different species from the genus. Therefore, these results seem to indicate that coding sequences functionally annotated can be amplified and utilized as genetic markers in species from the same genus, using heterologous primers. By the detection of polymorphic sites, these primers can now be more extensively tested for eventual polymorphisms that are more or less frequent in poplar hybrids both with improved trunk performance and specific exploitation characteristics.

In total, 51 SNPs were found out of 73,206 bp and nucleotide diversity values were between 0.04176 (COMT 4) and 0.01109 (Kt 1), on average 0.01857 (Table 2). According to the literature, nucleotide diversity provides valuable insights into the genomic imprint of selection in regions of the genome with different functional characteristics (Wright & Andolfatto, 2008), and into the genetic basis of wood formation, perenniality, and dormancy (Brunner et al., 2004). Reviewed by Savolainen and Pyhajarvi (2007) in other tree species, nucleotide diversity ranged from 0.0034 to 0.0247, in which our average result falls into this interval. In their analysis of 590 gene fragments from different populations from throughout the natural range of *Populus balsamifera*, Olson et al. (2010) found nucleotide diversity (0.0027) low, compared with the previously mentioned literature data. Accordingly, our results can be considered as high values. Nevertheless, this

discrepancy may be caused by the natural populations' gene flow versus the targeted breeding process toward specific exploitation characteristics of our different samples, the diverse origin of the investigated individuals, or the portion of the genome considered.

Indel events were found in number of 31 out of 164 sequences (Table 2). As elite poplar varieties are created through interspecific hybridization followed by clonal propagation, altered gene dosage relationships are believed to contribute to hybrid performance. Polyploids can arise in interspecific crosses (Bradshaw & Stettler, 1993), being expected to be tolerant of insertion/deletion events because their higher ploidy background that provides a buffer (Henry et al., 2015). According to the same author, indel events are not only frequent, being tolerated in hybrid poplars, but their presence is an important genetic basis of novel phenotypes, consequently in breeding strategies.

Non-synonymous single-base mutations could be detected ranging from 4 to 10, in number of 30 (Table 2), no one causing an early stop codon. Some of the non-synonymous SNPs detected in this study are of special interest because they might have an influence on the protein structure and function. For example, non-synonymous SNPs found in CCoAOMT 4 could lead to the modification the composition of lignin and secondary xylem development, or non-synonymous base changes in the potassium channel SKOR encoding sequence could change the regulation of K<sup>+</sup>-dependent wood production. From this aspect, our markers could provide a valuable examination tool of specific alleles that contribute to poplar wood traits, a very important aspect of current Hungarian poplar breeding programs.

The minimum number of recombination events was in number of 21 out of 164 sequences, SKOR 3 being of special interest (Table 2). In natural populations, recombination facilitates the removal of damaging alleles from the population, and genomic regions of low recombination accumulate more deleterious alleles as the efficiency of purifying selection is reduced (Charlesworth, 1990). In case of hybrid poplar varieties, the characterization of recombination may be effective at identifying very small genomic regions underlying phenotypic variation, perhaps even causative SNPs (Olson et al., 2010).

The number of significant pairwise comparisons was 41; COMT 4 proved to be of special interest due to its highest value (16) (Supplementary Table S5). As results of a test to measure the extent of linkage disequilibrium between pairs of loci, our data provide useful information about the forces governing those loci and allow selection for traits of economic importance. Selection of SNP markers that reflect the distribution and magnitudes of linkage disequilibrium in the poplar genome facilitates the design of optimal genetic association studies.

## CONCLUSION FOR FUTURE BIOLOGY

Poplars are the primary wood producers in countries lacking natural forests. The applications of their wood range from elements of construction and timber to the manufacture of paper, plywood, and reconstituted boards. Nevertheless, the success in the aforementioned applications results from the

breeding of poplar hybrids with both improved trunk performance and specific exploitation characteristics. To meet this need, as a first step, in this study, SNP markers were developed, tested, and analyzed to depict candidate genes encoding enzymes with wood property phenotypic traits in different clones with importance in Hungarian poplar cultivation. The use of these genes in a future association study based on candidate gene approach, in our opinion, will help to obtain a more complete characterization of the specific structure and wood composition. The analysis of the detected polymorphisms allows the effect on phenotype to be determined. Although an evaluation of the wood structural features in a larger study is warranted to estimate the interaction of the environment with the genotype, we consider that the information generated by our work represents a promising input to support a marker-assisted breeding strategy of improved poplars for wood industry.

---

*Acknowledgments:* The authors are grateful to Andrea Rumi for all the labworks during DNA isolation. They would like to acknowledge Roland Takács for his assistance in sample collection. They would also like to thank Biomi Ltd., Hungary, for sequencing the PCR products and to the Higher Education Restructuring Fund (FSA) for supporting the experiment.

*Ethical Statement:* The authors confirm that this work is not under consideration for publication elsewhere and all authors and institutions have approved the manuscript for submission.

*Funding Statement:* The experiment was supported by the Higher Education Restructuring Fund (FSA).

*Competing Interests:* The authors have no conflict of interest and they have agreed to submit this manuscript to *Biologia Futura*.

*Authors' Contributions:* ZAK made the database and literature data mining, developed the markers, contributed to the data analysis, and wrote the article. KC has significant contributions to the laboratory work, analysis of data, interpretation of the results, figure editing, and revising the article. AB has major contribution to the sampling design and sample collection, table editing, and paper revision. MB contributed to the revision of the manuscript. AB contributed in the cooperation between the two institutes and provided the institutional support indispensable to fulfill the experiment. RN besides revising the manuscript has significant contributions to the conceptualization and supervision of the experiment.

---

## REFERENCES

- Ache, P., Fromm, J., Hedrich, R. (2010) Potassium-dependent wood formation in poplar: seasonal aspects and environmental limitations. *Plant Biol.* 12, 259–267.
- Arend, M., Stinzinger, A., Wind, C., Langer, K., Latz, A., Ache, P., Hedrich, R. (2005) Polar-localised poplar K<sup>+</sup> channel capable

- of controlling electrical properties of wood-forming cells. *Planta* 223, 140–148.
- Bedon, F., Legay, S. (2011) Lignin synthesis, transcriptional regulation and potential for useful modification in plants. *Plant. Sci. Rev.* 6, 89.
- Bradshaw, H. D., Stettler, R. E. (1993) Molecular genetics of growth and development in *Populus*. I. Triploidy in hybrid poplars. *Theor. Appl. Gen.* 86, 301–302.
- Brunner, A. M., Busov, V. B., Strauss, S. H. (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends Plant Sci.* 9, 49–56.
- Carletti, G., Carra, A., Allegro, G., Vietto, L., Desiderio, F., Bagnaresi, P., Nervo, G. (2016) QTLs for woolly poplar aphid (*Phloeomyzus passerinii* L.) resistance detected in an interspecific *Populus deltoides* × *P. nigra* mapping population. *PLoS ONE* 11, 1–18.
- Charlesworth, B. (1990) Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* 55, 199–221.
- Decroocq, V., Fave, M. G., Hagen, L., Bordenave, L., Decroocq, S. (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theor. Appl. Genet.* 106, 912–922.
- Denison, F. C., Paul, A. L., Zupanska, A. K., Ferl, R. J. (2011) 14-3-3 Proteins in plant physiology. *Semin. Cell. Dev. Biol.* 22, 720–727.
- Dharmawardhana, P., Brunner, A. M., Strauss, S. H. (2010) Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. *BMC Genomics* 11, 150.
- Drost, D. R., Puranik, S., Novaes, E., Novaes, C. R., Dervinis, C., Gailing, O., Kirst, M. (2015) Genetical genomics of *Populus* leaf shape variation. *BMC Plant Biol.* 15, 1–10.
- Du, F. K., Xu, F., Qu, H., Feng, S., Tang, J., Wu, R. (2013) Exploiting the transcriptome of euphrates poplar, *Populus euphratica* (*Salicaceae*) to develop and characterize new EST-SSR markers and construct an EST-SSR database. *PLoS one* 8, 1–11.
- Dumolin, S., Demesure, B., Petit, R. J. (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theor. Appl. Genet.* 91, 1253–1256.
- Feng, S. P., Li, W. G., Huang, H. S., Wang, J. Y., Wu, Y. T. (2009) Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Mol. Breed.* 23, 85–97.
- Fromm, J. (2010) Wood formation of trees in relation to potassium and calcium nutrition. *Tree Physiol.* 30, 1140–1147.
- Fromm, J., Lautner, S. (2016) Abiotic stresses on secondary xylem formation. In: Kim, Y. S., Funada, R., Singh, A. P. (eds) *Secondary Xylem Biology*, Elsevier Academic Press, Amsterdam, Netherlands, pp. 59–71.
- Garcia, K., Zimmermann, S. D. (2014) The role of mycorrhizal associations in plant potassium nutrition. *Front. Plant Sci.* 5, 1–9.
- Geisler-Lee, J., Geisler, M., Coutinho, P. M., Segerman, B., Nishikubo, N., Takahashi, J., Sundberg, B. (2006) Poplar carbohydrate-active enzymes. Gene identification and expression analyses. *Plant Physiol.* 140, 946–962.
- Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acid Sci.* 41, 95–98.
- Halpin, C. (2004) Re-designing lignin for industry and agriculture. *Biotechnol. Genet. Eng.* 21, 229–248.
- Henry, I. M., Zinkgraf, M. S., Groover, A. T., Comai, L. (2015) A system for dosage-based functional genomics in poplar. *Plant Cell* 27, 2370–2383.
- Isabel, N., Lamothe, M., Thompson, S. L. (2013) A second-generation diagnostic single nucleotide polymorphism (SNP)-based assay, optimized to distinguish among eight poplar (*Populus* L.) species and their early hybrids. *Tree Genet. Genomes* 9, 621–626.
- Joshi, C. P., DiFazio, S. P., Kole, C. (eds) (2011) *Genetics, Genomics and Breeding of Poplar*. Science Publisher, Clemson.
- Langer, K., Levchenko, V., Fromm, J., Geiger, D., Steinmeyer, R., Lautner, S., Hedrich, R. (2004) The poplar K<sup>+</sup> channel KPT1 is associated with K<sup>+</sup> uptake during stomatal opening and bud development. *Plant J.* 37, 828–838. doi: 10.1111/j.0960-7412.2003.02008.x.
- Li, L., Lu, S., Chiang, V. (2006) A genomic and molecular view of wood formation. *Crit. Rev. Plant Sci.* 25, 215–233.
- Loziuk, P. L., Hecht, E. S., Muddiman, D. C. (2017) N-linked glycosite profiling and use of Skyline as a platform for characterization and relative quantification of glycans in differentiating xylem of *Populus trichocarpa*. *Anal. Bioanal. Chem.* 409, 487–497.
- Lu, S., Li, Q., Wei, H., Chang, M. J., Tunlaya-Anukit, S., Kim, H., Yeh, T. F. (2013) Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in *Populus trichocarpa*. *Proc. Natl. Acad. Sci. U. S. A.* 110, 10848–10853.
- Luo, R., Chen, P. W., Wagenbach, M., Jian, X., Jenkins, L., Wordeman, L., Randazzo, P. A. (2016) Direct functional interaction of the kinesin-13 family member kinesin-like protein 2A (Kif2A) and Arf GAP with GTP-binding protein-like, ankyrin repeats and PH domains 1 (AGAP1). *J. Biol. Chem.* 291, 25761.
- Mellerowicz, E. J., Sundberg, B. (2008) Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. *Curr. Opin. Plant Biol.* 11, 293–300.
- Mutwil, M., Debolt, S., Persson, S. (2008) Cellulose synthesis: a complex complex. *Curr. Opin. Plant Biol.* 11, 252–257.
- Nersissian, A. M., Valentine, J. S., Immoos, C., Hill, M. G., Hart, P. J., Williams, G., Herrmann, R. G. (1998) Uclacyanins, stellacyanins, and plantacyanins are distinct subfamilies of phytoeyanins: plant-specific mononuclear blue copper proteins. *Protein Sci.* 7, 1915–1929.
- Oakley, R. V., Wang, Y. S., Ramakrishna, W., Harding, S. A., Tsai, C. J. (2007) Differential expansion and expression of alpha- and beta-tubulin gene families in *Populus*. *Plant Physiol.* 145, 961–973.
- Olson, M. S., Robertson, A. L., Takebayashi, N., Silim, S., Schroeder, W. R., Tiffin, P. (2010) Nucleotide diversity and linkage disequilibrium in balsam poplar (*Populus balsamifera*). *New Phytol.* 186, 526–536.
- Rademacher, P., Báder, M., Németh, R., Rousek, R., Paril, P., Baar, J., Hornicek, S., Dejmál, A., Domény, J., Kudela, J., Kutnar, A., Neyses, B., Sandberg, D. (2017) European co-operation in wood research— from native wood to engineered materials. Part 2: densification modification in product development. In: Gurau, L., Campean, M., Ispas, M. (eds) *Proceedings of the International Conference Wood Science and Engineering in the Third Millennium (ICWSE 2017)*. Transilvania University, Brasov, Romania, pp. 469–478.



- Peakall, R., Gilmore, S., Keys, W., Morgante, M., Rafalski, A. (1996) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Mol. Biol. Evol.* *15*, 1275–1287.
- Rajinikanth, M., Harding, S. A., Tsai, C. J. (2007) The glycine decarboxylase complex multienzyme family in *Populus*. *J. Exp. Bot.* *58*, 1761–1770.
- Ralph, J., Schatz, P. F., Lu, F., Kim, H., Akiyama, T., Nelsen, S. F. (2009) *Quinone Methides in Lignification*. Wiley-Blackwell, Hoboken, NJ.
- Ranjan, P., Yin, T., Zhang, X., Kalluri, U. C., Yang, X., Jawdy, S., Tuskan, G. A. (2010) Bioinformatics-based identification of candidate genes from QTLs associated with cell wall traits in *Populus*. *Bioenergy Res.* *3*, 172–182.
- Ratke, C., Pawar, P. M. A., Balasubramanian, V. K., Naumann, M., Duncranz, M. L., Derba-Maceluch, M., Mellerowicz, E. J. (2015) *Populus* GT43 family members group into distinct sets required for primary and secondary wall xylan biosynthesis and include useful promoters for wood modification. *Plant Biotechnol. J.* *13*, 26–37.
- Rozas, J., Rozas, R. (1995) DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *CABIOS* *11*, 621–625.
- Sasaki, T., Fukuda, H., Oda, Y. (2017) Cortical microtubule disordering1 is required for secondary cell wall patterning in xylem vessels. *Plant Cell* *29*, 3123–3139.
- Savolainen, O., Pyhajarvi, T. (2007) Genomic diversity in forest trees. *Curr. Opin. Plant Biol.* *10*, 162–167.
- Sawada, D., Kalluri, U. C., O'Neill, H., Urban, V., Langan, P., Davison, B., Pingali, S. V. (2018) Tension wood structure and morphology conducive for better enzymatic digestion. *Biotechnol. Biofuels* *11*, 1–9.
- Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Van Montagu, M. (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 13330–13335.
- Takabe, K., Takeuchi, M., Sato, T., Ito, M., Fujita, M. (2001) Immunocytochemical localization of enzymes involved in lignification of the cell wall. *J. Plant Res.* *114*, 509–515.
- Tenney, A. E., Wu, J. Q., Langton, L., Klueh, P., Quatrano, R., Brent, M. R. (2007) A tale of two templates: automatically resolving double traces has many applications, including efficient PCR-based elucidation of alternative splices. *Genome Res.* *17*, 212–218.
- Vander Mijnsbrugge, K., Meyermans, H., Van Montagu, M., Bauw, G., Boerjan, W. (2000) Wood formation in poplar: identification, characterization, and seasonal variation of xylem proteins. *Planta* *210*, 589–598.
- Villalobos, D. P., Díaz-Moreno, S. M., Said, E. S. S., Cañas, R. A., Osuna, D., Van Kerckhoven, S. H., Cantón, F. R. (2012) Reprogramming of gene expression during compression wood formation in pine: coordinated modulation of s-adenosylmethionine, lignin and lignan related genes. *BMC Plant Biol.* *12*, 100.
- Wright, S. I., Andolfatto, P. (2008) The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annu. Rev. Ecol. Evol. Syst.* *39*, 193–213.
- Zhong, R., Ye, Z. H. (2007) Regulation of cell wall biosynthesis. *Curr. Opin. Plant Biol.* *10*, 564–572.