Canadian Journal of Forest Research
Revue canadienne de recherche forestière

Single nucleotide polymorphism (SNP) characterization of drought-responsive genes to estimate genetic variation of *Populus tremula* var. *daavidiana* and eight other *Populus* species

Journal:	<i>Canadian Journal of Forest Research</i>
Manuscript ID	cjfr-2017-0387.R2
Manuscript Type:	Article
Date Submitted by the Author:	19-Feb-2018
Complete List of Authors:	Kim, Yong-Yul ; National Forest Seed and Variety Center, Department of Seed and Seedling Management Kwon, Soon-Ho; Seoul National University, Collage of agriculture and lifesience Jo, Arna; National Forest Seed and Variety Center, Department of Seed and Seedling Management Kim, Yang-Gil; Seoul National University, Department of Forest Sciences Lee, Jei-wan; National Institute of Forest Science, Department of Forest Genetic Resources Kang, Kyu-Suk; Seoul National University, Department of Forest Sciences
Keyword:	drought tolerance, single-nucleotide polymorphism, genetic variation, <i>Populus</i> , haplotype
Is the invited manuscript for consideration in a Special Issue? :	N/A

SCHOLARONE™
Manuscripts

1 **Single nucleotide polymorphism (SNP) characterization of drought-**
2 **responsive genes to estimate genetic variation of *Populus tremula* var.**
3 ***daurica* and eight other *Populus* species**

4 **Yong-Yul Kim[¶], Soon-Ho Kwon[¶], Aruna Jo, Yang-Gil Kim, Jei-Wan Lee, and Kyu-Suk**
5 **Kang***

6
7 **Soon-Ho Kwon, Yang-Gil Kim and Kyu-Suk Kang.** Department of Forest Sciences, College
8 of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul
9 08826, Republic of Korea.

10 **Jei-Wan Lee.** Department of Forest Genetic Resources, National Institute of Forest Science,
11 Suwon 16631, Republic of Korea

12 **Yong-Yul Kim and Aruna Jo.** Department of Seed and Seedling Management, National Forest
13 Seed and Variety Center, 72 Suhyori-ro, Suanbo-myeo, Chung-ju 27495, Republic of Korea

14
15 *** Corresponding author:** Kyu-Suk Kang (email: kangks84@snu.ac.kr).

16
17 [¶] These authors contributed equally to this work.

18

Abstract

In order to reveal the genetic mechanisms behind drought tolerance, we attempted to identify the responsive genes on drought tolerance in *Populus tremula* var. *daurica* and eight other species in *Populus*. Nine drought-responsive genes were assayed by single nucleotide polymorphism (SNP) genotyping using *TaqMan* probe-based RT-PCR. A total of 346 SNPs were found from the 101 sequences of the nine genes. Among them, 57 primers were selected for SNP genotyping and haplotype determination. For the SNP genotyping, 56 assays were applicable for section *Leuce*, 53 for section *Aigeiros*, 50 for *P. maximowiczii*, and 52 for *P. simonii*. The expectation maximization algorithm was implemented to determine haplotypes among species and sections. A total of 78 haploid types were estimated from nine drought gene loci, and the average of haploid types per locus was 8.67. The analysis of genetic diversity revealed that *P. tremula* var. *daurica* had substantial level of genetic variation for the nine drought genes, which was higher than the eight other species. The means of observed number of allele (N_A) and heterozygosity (H_o) were 5.33 and 0.6609, respectively. The genetic differentiation of *P. tremula* var. *daurica* showed that 7.52% of genetic variation was among populations and the remaining was within a population.

Key words; drought tolerance, single-nucleotide polymorphism, genetic variation, haplotype, *Populus*

Introduction

Drought is one of the most common abiotic stresses responsible for yield loss in commercial tree species (Whitehead & Beadle 2004; Tavousi *et al.* 2015). Efforts have been given to investigate physiological mechanism and genetic basis of drought tolerance (Newton *et al.* 1991; Wang 2014). There has also been progresses on the identification of drought-response genes and the development of gene expression studies; they include a *dehydrin* gene for late embryogenesis abundant (LEA) proteins (Caruso *et al.* 2002; Pedrosa *et al.* 2015), transcription-factor genes for dehydration-responsive element-binding (DREB) protein family, ethylene responsive element binding factor (ERF) family (Zhao *et al.* 2014), the zinc-finger family (Liu *et al.* 2015), and WRKY family (Nanjo *et al.* 2004; Rushton *et al.* 2010).

Populus tremula var. *daurica* is a unique aspen species in the Korean peninsula, belonging to section *Lecuce* Duby in genus *Populus*. It also naturally occurs over a wide variety of habitats in Korea, China, Mongolia and Russia. The taxonomical studies using AFLP show that there is a high similarity between *P. tremula* and *P. tremuloides* genetically (Cervera *et al.* 2005). The genus *Populus* has been divided into six sections on the basis of leaf and flower characters (Noh *et al.* 1992). Some species have different relationships among the sections, which are indicated by their nuclear DNA (paternal inherit) and chloroplast DNA sequences (maternal inherit). Moreover, some ecotypes in the species significantly differ in

57 physiology and growth response to drought stress and exogenous ABA application (Li *et al.*
58 2004; Zhang *et al.* 2004).

59 A single nucleotide polymorphism (SNP) is a variation in a single nucleotide that occurs at
60 a specific position in the genome, where each variation is present to some appreciable degree
61 within a population. The SNP is the most common type of genetic variation among individuals
62 and species. Each species has nucleotide polymorphisms on the genetic loci that respond to a
63 drought stress and link the SNP at the gene with the already known properties of responding to
64 drought stresses of species. This may be a good indicator for early selection of individuals with
65 high resistance to stress in the future genome-wide association study (Brumfield *et al.* 2003;
66 Varshney *et al.* 2005; Tuberosa & Salvi 2006; Wang *et al.* 2015).

67 Large amounts of sequence data for many functional genes have been obtained from the
68 genome projects for poplar, aspen, cottonwood and their interspecific hybrids (Bhalerao *et al.*
69 2003; Kohler *et al.* 2003; Sterky *et al.* 2004). The public genome database can use to develop
70 SNP markers for finding biologically and functionally important genes.

71 The main purposes of this study were to identify SNPs occurring in partial genomic regions
72 of *Populus tremula* var. *daurica* corresponding to nine drought-responsive genes, to develop
73 the panel of SNP genotyping assays using *TaqMan* probe based real-time PCR for the genes,

74 and to apply the assays to estimate genetic variation of *P. tremula* var. *daurica* and its related

75 eight species from three sections in genus *Populus*.

76

Draft

77 **Material and method**

78 **Plant material and re-sequencing for nine drought genes**

79 Genomic DNA was extracted from leaf samples of 18 trees selected in a poplar germplasm-
80 collection garden at the National Institute of Forest Science in South Korea. The trees included
81 eight species from three sections in genus *Populus*: 4 species for section *Leuce* Duby (*P.*
82 *tremula* var. *davidiana*, *P. tremula* var. *glandulosa*, *P. alba* and *P. alba* x *P. tremula* var.
83 *glandulosa*), 1 for section *Tacamahaca* Spach (*P. maximowizii*), and 3 for section *Aigeiros*
84 Duby (*P. nigra*, *P. deltoides* and *P. euramericana*). Additionally, *P. tremula* var. *davidiana* was
85 subdivided into five populations based on the origin of the collection.

86 Nine drought-responsive genes were included in the study (Caruso et al. 2002, Nanjo et al.
87 2004). They are dehydrin (*dhn*), alternative oxidase 1a (AOX), amino acid permease I (AAP1),
88 galactinol synthase (GS), transcription factor like protein (TF), ethylene-responsive element-
89 binding factor 1 (ERF), 4-hydroxyphenylpyruvate dioxygenase (HPD), leucoanthocyanidin
90 dioxygenase-like protein (LAD) and putative zinc finger protein (ZF). ‘Primer3’ program
91 (Whitehead Institute for Biomedical Research) was used to design primers to amplify the
92 genomic regions of the 18 trees from 13 sequences of the 9 genes available in NCBI genbank
93 database; they are the pseudhn1 cDNA sequence of *P. euramericana* (GenBank accession no.
94 AJ300524) for *dhn*, and 2 full-length EST sequences (5’- and 3’-end EST) of *P. nigra* for each

of AAP1 (BP923026 and BP930782), HPD (BP922875 and BP930649), LAD (BP926606 and BP934164) and ZF (BP925347 and BP932979), and 5'-end EST for AOX (BP922653), ERF (BP924591) and GS (BP922497), and 3'-end EST for TF (BP930753)..

The PCR amplification was performed in 10 μ L of mixtures containing 20 ng template DNA, 0.2 μ M of each forward and reverse primer, 0.2 mM dNTPs, 1.2 U of HotStart Taq DNA polymerase (Qiagen) and 1 μ L of 10X PCR buffer including 15 mM $MgCl_2$. PCR cycling conditions consisted of an initial denaturation of 95 °C for 15 min, and 35 cycles of 94 °C for 30 sec, 55 °C for 30 sec and 72 °C for 1 min, and followed by a final extension step of 72 °C for 5 min. The amplified DNA fragment was isolated from 1.0 % agarose gel and purified with QIAquick Gel Extraction Kit (Qiagen), and then used in re-amplification reaction to generate template DNA for further sequencing works. Sequence data were obtained directly from the re-amplified products on the ABI 3100 genetic analyzer using the BigDye terminator v.3.1 cycle sequencing kit (Applied Biosystems). All samples were sequenced from both ends. Base calling and assembly of forward and reverse reads were conducted using the SeqScape v.2.0 program (Applied Biosystems). Low quality bases were excluded. However, those resulting from heterozygous sequence of which the lower peak height was greater than 50% of the higher one on the electropherogram were included.

113 **SNP identification and TaqMan probe development**

114 Multiple sequence alignment was performed using the Clustal X (1.81) program, which is
115 widely used and suitable for small sequences, where two DNA sequences of *P. euramericana*
116 (GenBank accession no. AJ30024) and *P. alba* x *P. tremula* var. *glandulosa* (DQ856592.1)
117 were used as reference for the *dhn* gene, and 12 full-length ESTs of *P. nigra* for the remaining
118 eight genes (Jeanmougin *et al.* 1998). Unmatched areas at both ends of the aligned sequences
119 were trimmed thoroughly. Single-base indels were crosschecked to the original
120 electropherogram for verifying that they were not sequencing artifacts missed during base
121 calling. The BLAST N searches were also performed to find best-matched sequences.

122 Single nucleotide polymorphism (SNP) markers are one of the useful tools for population
123 genetic studies and genotype/phenotype association mapping. SNPs were detected directly
124 from the multiple sequence alignments obtained by the MEGA 3 program (Kumar *et al.* 2004)
125 based on the definition as either nucleotide differences between the aligned samples at the
126 same site or the occurrence of double peaks at a site in one individual. Primer Express v3.0
127 program (Applied Biosystems) was used to design the primers and *TaqMan* probes for the
128 SNPs. Finally, Assay-by-Design service (Applied Biosystems) was used to create panels of
129 SNP genotyping assays using *TaqMan* probe-based real-time PCR.

130

131 **SNP genotyping assay**

132 The practical applicability of each SNP genotyping assay was tested for 100 DNA samples
133 from nine *Populus* species; 58 samples of *P. tremula* var. *daurica*, 26 of 3 species in section
134 *Leuce* (*P. alba*, *P. tremula* var. *glandulosa* and *P. alba* x *P. tremula* var. *glandulosa*), 8 of 3
135 species in section *Aigeiros* (*P. deltoides*, *P. euramericana* and *P. nigra* var. *italica*), and 7 of *P.*
136 *maximowiczii*, and 1 of *P. simonii* in section *Tacamahaca*. The assays were basically performed
137 in 384-well reaction plates, and two wells in each plate were used as NTCs (No Template
138 Controls). At least three samples representing each of three different SNP types (two
139 diplotypes and one heterozygote) at each assay were also included in the plate, which helped to
140 validate an automatic SNP allelic discrimination results based on two-dimensional ordination
141 plotting SNP type of each sample in a X-Y coordinate system.

142 Each reaction was conducted in a 5 μ L volume consisting of 2 μ L of the diluted template
143 DNA (2ng/ μ L), 2.5 μ L of 2 x *TaqMan* Universal PCR Master Mix (Applied Biosystems, P/N
144 4304437), 0.125 μ L of each 40 x SNP Genotyping Assay Mix, and 0.375 μ L of sterile-filtered
145 water. The 40 x Assay Mix contains 36 μ M of unlabelled sequence-specific forward and
146 reverse primers, and also 8 μ M of *TaqMan* MGB probes each labeled with either VIC- or
147 FAM-dye. The reaction was carried out by an ABI 7900HT fast real-time PCR system
148 (Applied Biosystems) under the following thermal cycling condition; an initial denaturation of

149 10 min at 95 °C, and 40 cycles of 92 °C for 15 sec and 60 °C for 1 min. The 7900HT sequence
150 detection system v2.2.1 (Applied Biosystems) was used to read the fluorescence generated
151 from each well, and to perform an automatic or a manual calling of SNP type for each assay.

152

153 **Haplotype determination and genetic diversity estimation**

154 Haplotypes of each gene were determined using an EM algorithm function (Excoffier &
155 Slatkin 1995) implemented in the PowerMarker v.3.25 program (Liu & Muse 2001), and their
156 frequencies by species and sections were also estimated. One thousands of bootstrapping were
157 performed across samples to calculate standard error and confidence interval for each
158 haplotype at a given gene locus.

159 Genetic diversity parameters such as the observed number of alleles (N_A), the effective
160 number of alleles (N_E), the observed heterozygosity (H_o), Nei's average heterozygosity (H_e)
161 (Nei 1978) were calculated using POPGENE v.1.32 program (Yeh & Boyle 1997), and the
162 polymorphism information content (PIC) (Botstein *et al.* 1980) using the PowerMarker
163 program. Wright's F-statistics (F_{IS} , F_{IT} and F_{ST}) to estimate the genetic differentiation among
164 populations within *P. tremula* var. *daurica* were also computed using the POPGENE
165 program.

166

167 **Result and discussion**

168 **Re-sequencing and SNP identification**

169 A total of 346 SNPs were found in the 101 sequences of the nine genes, ranging from 7 (GS)
170 to 69 (HPD) SNPs. A total of 59 pairs of primers and probes were firstly selected as SNP
171 genotyping assays. However, 57 of the candidate assays (96.7%) were finally created using the
172 Assay-by-Design service (Applied Biosystems); 17 assays were developed for *dhn*, 8 for
173 AAP1, 4 for AOX, 3 for ERF, 4 for GS, 7 for HPD, 7 for LAD, 2 for TF and 5 for ZF (Table 1).

174

175 (Insert Table 1, close here)

176

177 The finally verified 101 sequences obtained in the study are now available in NCBI
178 Genbank database; they are 18 sequences for *dhn* (Genbank accession nos. EF639391 to
179 EF639408), 12 for AAP1 (ER973770 to ER973778), 9 for AOX (ER973782 to ER973790), 8
180 for ERF (ER973791 to ER973798), 6 for GS (ER 973799 to ER973804), 16 for HPD
181 (ER973805 to ER973820), 14 for LAD (ER973821 to ER973834), 5 for TF (ER973835 to
182 ER973839) and 13 for ZF (ER973840 to ER973844, and ER973762 to ER973765).

183

184 SNP genotyping assay and haplotype determination

185 All of the 57 assays successfully generated interpretable SNP types for *P. tremula* var.
186 *daurica* (Table 1). For section *Leuce*, 56 assays (98.2%) were applicable for SNP genotyping,
187 53 for section *Aigeiros* (93.0%), 50 for *P. maximowiczii* (87.7%), and 52 for *P. simonii* (91.2%).
188 In overall, 47 assays (82.5%) were commonly applicable over the species and the sections. The
189 proportions of the dimorphic assays by species and sections ranged from 42 to 77.4%. Only the
190 28 assays were polymorphic in *P. tremula* var. *daurica* (49.1%), 21 in *Leuce* (37.5%), 13 in
191 *Aigeiros* (22.6%) and 29 in *P. maximowiczii* (58%).

192 By regarding the polymorphic SNP assays as individual gene loci, we computed the
193 observed heterozygosity (H_o) (Weir 1996), and polymorphic information content (PIC)
194 (Botstein *et al.* 1980) by species and sections. The values of H_o and PIC for *P. tremula* var.
195 *daurica* were 0.3723 and 0.2940, 0.1398 and 0.1108 for section *Leuce*, 0.1323 and 0.1096
196 for section *Aigeiros*, and 0.1223 and 0.1075 for *P. maximowiczii*, respectively (Table 2).

197

198 (Insert Table 2, close here)

199

200 A total of 78 haplotypes were estimated from nine gene loci over the species and the
201 sections. They were encoded with their own alphabetical ID, and defined as one of the alleles

for each gene loci. The number of haplotypes observed by gene loci varied from 3 (TF) to 24 (dhn), with an average of 8.67 haplotypes per locus. In *P. tremula* var. *daurica*, 48 haplotypes were determined from the nine genes; 5.33 per locus (Table 2). In section *Leuce*, 31 haplotypes were observed (3.44/locus), and 20 in section *Aigeiros* (2.22/locus), and 23 in *P. maximowiczii* (2.56/locus). We also found species-specific haplotypes in several gene loci. Twenty-four haplotypes in six genes were unique for *P. tremula* var. *daurica*, and six in three loci for *P. maximowiczii*. Among the gene loci, the dhn had the largest number of species-specific haplotypes; 21 out of 24 haplotypes (87.5 %) were specific for the species and the sections.

Genetic diversity among species and sections

The estimated parameters by genes and species/or sections were listed in Table 3. In *P. tremula* var. *daurica*, the averages of N_A and H_o for nine gene loci were 5.33 and 0.6609, which were higher than those of the trembling aspen (*P. tremuloides*) estimated by six SSR markers (4.2 and 0.36) reported by Rahman *et al.* (2000) and Rahman and Rajora (2002). The species seemed to have substantial level of genetic variation for the nine gene loci. In the *Aigeiros* section, the averages of N_A and PIC for the eight polymorphic gene loci were 2.22 and 0.3647. They were lower than those of *P. deltoides* (3.3 and 0.372 for 6 SSR markers), and

220 much lower than *P. x canadensis* (9.3 and 0.693 for 6 markers) (Fossati *et al.* 2005). In *P.*
221 *maximowiczii*, the averages of N_A and H_o for the seven polymorphic loci were estimated as
222 2.56 and 0.5918, which were the smallest in the study.

223

224

(Insert Table 3, close here)

225

226 We computed the Wright's F -statistics (F_{IS} , F_{IT} and F_{ST}) to estimate the genetic
227 differentiation among populations within *P. tremula* var. *daurica*. The averages of F_{IS} and F_{IT}
228 for nine genes were -0.2601 and -0.1007 . Most of the genes showed a high level of
229 heterozygote excess in terms of both within-population and in total population, while
230 extremely high level of heterozygote deficiency was observed in the TF locus (Table 3). The
231 F_{ST} values of the gene loci ranged from 0.0223 (ERF) to 0.2552 (ZF); on average 0.1317. In
232 particular, five genes (AAP1, AOX, dhv, TF and ZF) showed significant genetic differentiation
233 among populations. The proportion of genetic variance for among-population was 7.52%, and
234 the remaining was for within-population. This result occurred with the report of *P. tremula* by
235 Ingvarsson (2005), where the F_{ST} for five nuclear functional genes was 0.117.

236 It is generally known that the subdivided populations show a high level of linkage
237 disequilibrium (LD) among gene loci resulting from the limitation of migration between them

(Ohta 1982). In order to estimate a level of LD among the nine gene loci in *P. tremula* var. *daurica*, we performed two-locus (Ohta 1982) and multiple-locus analysis (Smouse *et al.* 1983). The average pairwise correlations (r^2) for within- and among-populations were estimated as 0.68390 and 1.04179, and they were highly significant ($\alpha = 0.001$). In addition, the overall means of two statistics ($D'_{IS}^2/D'_{ST}^2 > 1$, $D_{ST}^2/D_{IS}^2 > 1$) across pairs of the nine gene loci were 6.74 and 20.89 (both greater than 1). Therefore, it is evidently concluded that the five populations in *P. tremula* var. *daurica* are genetically subdivided for the nine genes, and the subdivision results in a level of LD (Ohta 1982). This results could be supported by that the species shows natural propagation of root-suckers rather than seeds.

We additionally calculated *F*-statistics for three-level sampling hierarchy, *i.e.* eight species from the three sections of genus *Populus* (*P. tremula* var. *daurica* was included in section *Leuce*), assuming that each species was considered as one of the random mating populations in genus *Populus*. Three statistics for measuring differentiation among species (*F*), sections (θ_S) and the whole genus (θ_P) were estimated as 0.2696, 0.4060 and 0.2926, respectively. Moreover, the average pairwise correlations (r^2) for within- and among-sections were significant (0.684 and 0.892). These results implied that the genetic differentiation on the nine drought-responsive gene loci was high among species and in particular very high among sections.

The NJ tree from genetic distances between species showed the coincident grouping with the conventional botanical classification and the AFLP marker analysis for a large set of *Populus* species (Cervera *et al.* 2005) (Fig 1). It might also be relevant to the differentiation for the nine drought-responsive genes among species within genus *Populus*.

Fig 1. Neighbor-joining tree for genetic relationships among the eight species of three sections in genus *Populus*. (pd: *P. tremula* var. *daurica*; pa: *P. alba*; ptg: *P. tremula* var. *glandulosa*; paptg: *P. alba* x *P. tremula* var. *glandulosa*; pm: *P. maximowiczii*; pdel: *P. deltoides*; pni: *P. nigra* var. *italica*; peu: *P. euramericana*)

Many SNP markers in genus *Populus* have been reported and high-quality reference sequences are available in public genome data base. In this respect, our results might encourage for further development of SNP markers of many functional genes in *P. tremula* var. *daurica* and other species in genus *Populus*. Recently, various research about drought-

269 response genes in genus *Populus* is underway (Cohen et al. 2010). Dehydration-responsive
270 element-binding (DREB) gene has been studied in the *Populus* for transcriptome analysis and
271 SNPs related drought (Chu et al. 2014). In particular, it could be expected that the SNP assays
272 developed from the night drought-responsive genes would deeply contribute to population
273 genetics and genotype/phenotype association mapping studies in genus *Populus*.

274

275 **Acknowledgements**

276 We thank all participants, especially members of the Forest Resources Genome Sequencing
277 project (project code S111414L070120) of Korea Forest Service, for contributing to this study
278 and also for many helpful suggestions for improving the manuscript of this article. This paper
279 was partially contributed from the research project (code: FG0601-2004-03-2017) of National
280 Institute of Forest Science, Republic of Korea.

281

References

- Bhalerao, R., Nilsson, O. and Sanberg, G. 2003. Out of the woods: forest biotechnology enters the genomic era. *Current Opinion in Biotechnology* 14: 206-213.
- Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314-331.
- Brumfield, R.T., Beerli, P., Nickerson, D.A. and Edwards, S.V. 2003. The utility of single nucleotide polymorphisms in references of population history. *Trends in Ecology and Evolution* 18: 249-256.
- Caruso, A., Morabito, D., Delmotte, F., Kahlem, K. and Carpin, S. 2002. Dehydrin induction during drought and osmotic stress in *Populus*. *Plant Physiology and Biochemistry* 40: 1033-1042.
- Cervera, M.T., Strome, V., Soto, A., Ivens, B., Montagu, M. and Rajora, O. 2005. Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. *Theoretical and Applied Genetics* 111: 1440-1456.
- Chu, Y., Huang, Q., Zhang, B., Ding, C. and Su, X. 2014. Expression and Molecular Evolution of Two DREB1 Genes in Black Poplar (*Populus nigra*). *Plos one*; e98334

- 299 Cohen, D., Bogeat-Triboulot, M.B., Tisserant, E., Balzergue, S., Martin-Magniette, M.L. and
300 Lelandais, G. 2010. Comparative transcriptomics of drought responses in *Populus*: a meta-
301 analysis of genome-wide expression profiling in mature leaves and root apices across two
302 genotypes. *BMC genomics* 11: 630.
- 303 Excoffier, L. and Slatkin, M. 1995. Maximum-likelihood estimation of molecular haplotypes
304 frequencies in a diploid population. *Molecular Biology and Evolution* 12: 921-927.
- 305 Fossati, T., Zapelli, H., Bisoffi, S., Micheletti, A., Vietto, L. and Sala, F. 2005. Genetic
306 relationships and clonal identity in a collection of commercially relevant poplar cultivars
307 accessed by AFLP and SSR. *Tree Genetics & Genomes* 1: 11-19.
- 308 Ingvarsson, P.K. 2005. Nucleotide Polymorphism and Linkage Disequilibrium Within and
309 Among Natural Populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics*
310 169: 945-953.
- 311 Jin, L. and Chakraborty, R. 1993. Estimation of genetic distance and coefficient of gene
312 diversity from single-probe multilocus DNA fingerprinting data. *Molecular Biology and*
313 *Evolution* 11: 120-127.
- 314 Kohler, A., Delaruelle, C., Martin, D., Encelot, N. and Martin, F. 2003. The poplar root
315 transcriptome: analysis of 7000 expressed sequence tags. *Federation of European*
316 *Biochemical Society Letters* 542: 37-41.

- 317 Kumar, S., Tamura, K. and Nei, M. 2004. MEGA3: Integrated software for molecular
318 evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 5: 150-
319 163.
- 320 Li, C., Yin, C. and Liu, S. 2004. Different responses of two contrasting *Populus davidiana*
321 populations to exogenous Absciscic acid application. *Environmental and Experimental*
322 *Botany* 51: 237-246.
- 323 Liu, Q., Wang, Z., Xu, X., Zhang, H. and Li, C. 2015. Genome-wide analysis of C2H2 zinc-
324 finger family transcription factors and their responses to abiotic stresses in Poplar (*Populus*
325 *trichocarpa*). Plos One; <https://doi.org/10.1371/journal.pone.0134753>
- 326 Nanjo, T., Futamura, N., Nishiguchi, M., Igasaki, T., Shinozaki, K. and Shinohara, K. 2004.
327 Characterization of full-length enriched expressed sequence tags of stress-treated poplar
328 leaves. *Plant Cell Physiology* 45: 1738-1748.
- 329 Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number
330 of individuals. *Genetics* 89: 583-590.
- 331 Newton, R.J., Funkhouser, E.A., Fong, F. and Tauer, CG. 1991. Molecular and physiological
332 genetics of drought tolerance in forest trees. *Forest Ecology and Management* 43: 225-250.
- 333 Noh, E.R., Koo, Y. and Shim, C. 1992. Breeding of aspen (*Populus davidiana* Dode) in Korea.
334 Proceeding of 19th session of the Internal Poplar Commission. Vol. 1: 403-412.

- 335 Liu, K. and Muse, S.V. 2005. PowerMarker: integrated analysis environment for genetic
336 marker data. *Bioinformatics* 21: 2128-2129.
- 337 Pedrosa, A.M., Martins, C.P.S., Gonçalves, L.P., Costa, M.G.C. 2015. Late embryogenesis
338 abundant (LEA) constitutes a large and diverse family of proteins involved in development
339 and abiotic stress responses in Sweet orange (*Citrus sinensis* L. Osb.). *Plos One*;
340 DOI:10.1371/journal.pone.0145785
- 341 Rahman, M.H., Dayanandan, S. and Rajora, O.P. 2000. Microsatellite DNA markers in
342 *Populus tremuloides*. *Genome* 43: 293-297.
- 343 Rahman, M.H. and Rajora, O.P. 2002. Microsatellite DNA fingerprinting, differentiation, and
344 genetic relationships of clones, cultivars, and varieties of six poplar species from three
345 sections of the genus *Populus*. *Genome* 45: 1083-1094.
- 346 Rushton, P.J., Somssich, I.E., Ringler, P. and Shen, Q.J. 2010. WRKY transcription factors.
347 *Trends in Plant Science* 15(5): 247-258.
- 348 Smouse, P.E., Neel, J.V. and Liu, W. 1983. Multiple-locus departures from panmictic
349 equilibrium within and between village gene pools of Amerindian tribes at different stages
350 of agglomeration. *Genetics* 104: 133-153.

- 351 Sterky, F., Bhalerao, R.R., Unneberg, P., Segerman, B., Nilsson, P. and Brunner, AM. 2004. A
352 Populus EST resource for plant functional genomics. *Proceedings of National Academic*
353 *Science USA* 101: 13951-13956.
- 354 Tavousi, M., Kaveh, F., Alizadeh, A., Babazadeh, H. and Tehranifar, A. 2015. Effects of
355 drought and salinity on yield and water use efficiency in Pomegranate tree. *J. Mater.*
356 *Environ. Sci.* 6: 1975-1980.
- 357 Jeanmougin F., Thomson, J.D., Gouy M., Higgins D.G., Gibson T.J. 1998. Multiple sequence
358 alignment with Clustal X. *Trends Biochem Sci.* 23(10): 403-5.
- 359 Tuberosa, R. and Salvi S. 2006. Genomics-based approaches to improve drought tolerance of
360 crops. *Trends in Plant Science* 11: 405-412.
- 361 Varshney, R.K., Graner, A. and Sorrells, ME. 2005. Genomics-assisted breeding for crop
362 improvement. *Trends in Plant Science* 10: 621-630.
- 363 Wang, B., Tan, H-W., Fang, W., Meinhardt, L., Mischke, S., Matsumoto, T. and Zhang, D. 2015.
364 Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences
365 for identification of longan (*Dimocarpus longan*) germplasm. *Horticulture Research* 2:14065.
366 doi: 10.1038/hortres.2014.65
- 367 Wang, L.F. 2014. Physiological and molecular responses to drought stress in rubber tree (*Hevea*
368 *brasiliensis* Muell. Arg.). *Plant Physiol Biochem.* 83: 243-249.

- 369 Weir, B.S. 1996. *Genetic Data Analysis II*. Sinauer Associates. Inc. Sunderland. MA
- 370 Whitehead, D. and Beadle, C.L. 2004. Physiological regulation of productivity and water use
- 371 in *Eucalyptus*: a review. *Forest Ecology and Management* 193: 113-140.
- 372 Yeh, F.C. and Boyle, T.J.B. 1997. Population genetic analysis of co-dominant and dominant
- 373 markers and quantitative traits. *Belgian Journal of Botany* 129: 157.
- 374 Zhang, X., Zang, R. and Li, C. 2004. Population differences in physiological and
- 375 morphological adaptations of *Populus davidiana* seedlings in response to progressive
- 376 drought stress. *Plant Science* 166: 791-797.
- 377 Zhao, T., Xia, H., Liu, J. and Ma, F. 2014. The gene family of dehydration responsive element-
- 378 binding transcription factors in grape (*Vitis vinifera*): genome-wide identification and
- 379 analysis, expression profiles, and involvement in abiotic stress resistance. *Molecular*
- 380 *Biology Reports* 41(3): 1577-1590.

Table 1. Summary of re-sequencing for nine drought responsive genes and SNP identification in genus *Populus*

Gene loci	Reference sequence	Number of samples Species (number of clone)	Read length (bp)	GenBank accession number	Nucleotide polymorphism	SNP assays created by Assay by design
dhn	<i>P. euramericana</i> Peudhn 1 (AJ300524)	18 samples			75	17 assays / 56 SNPs
		<i>P. tremula</i> var <i> davidiana</i> (13 clones)	848	EF639391 ~ EF639403	SNP: 56	Coding: 16
		<i>P. tremula</i> var. <i> glandulosa</i> (1)	845	EF639404	Coding: 41	Non-coding: 1
		<i>P. alba</i> (1)	852	EF639405	Non-coding: 15	
		<i>P. maximowiczii</i> (1)	842	EF639406	INDEL: 19	
		<i>P. nigra</i> (1)	842	EF639407		
		<i>P. euramericana</i> (I-476) (1)	842	EF639408		
AAP1	<i>P. nigra</i> Amino acid permease I 5'-end EST (BP923026)	12 samples			55	8 assays / 53 SNPs
		4 samples			28	4 assays / 26 SNPs
		<i>P. tremula</i> var <i> davidiana</i> (3)	523, 525	ER973778 ~ ER973780	SNP: 26	
		<i>P. nigra</i> var. <i> italica</i> (1)	523	ER973781	INDEL: 2	
		3'-end EST (BP930782)			27	4 assays / 27 SNPs
AOX	<i>P. nigra</i> Alternative oxidase 1a 5'-end EST (BP922653)	8 samples			SNP: 27	
		<i>P. tremula</i> var <i> davidiana</i> (7)	466 ~ 467	ER973770 ~ ER973776		
		<i>P. nigra</i> var. <i> italica</i> (1)	467	ER973777		
		9 samples			19	4 assays / 16 SNPs
		<i>Salix pseudolasioogyne</i> (1)	376	ER973782	SNP: 16	
		<i>P. alba</i> x <i>P. tremula</i> var. <i> glandulosa</i> (1)	376	ER973783	INDEL: 3	
		<i>P. tremula</i> var <i> davidiana</i> (4)	375 ~ 376	ER973784 ~ ER973786 , ER973789		
ERF	<i>P. nigra</i> Ethylene-responsive element-binding factor 1 5'-end EST (BP924591)	<i>P. euramericana</i> (Eco28) (1)	375	ER973787		
		<i>P. deltoides</i> (LUX 69/55) (1)	376	ER973788		
		<i>P. nigra</i> var. <i> italica</i> (1)	374	ER973790		
		8 samples			21	3 assays / 21 SNPs
		<i>P. tremula</i> var <i> davidiana</i> (7)	320	ER973791 ~ ER973797	SNP: 21	
GS	<i>P. nigra</i> Galactinol synthase 5'-end EST (BP922497)	<i>P. nigra</i> var. <i> italica</i> (1)	320	ER973798		
		6 samples			10	3 assays / 7 SNPs
		<i>Salix pseudolasioogyne</i> (1)	399	ER973799	SNP: 7	
		<i>P. alba</i> x <i>P. tremula</i> var. <i> glandulosa</i> (1)	399	ER973800	INDEL: 3	
		<i>P. tremula</i> var <i> davidiana</i> (4)	398 ~ 400	ER973801 ~ ER973804		

HPD	<i>P. nigra</i>	16 samples			73	7 assays / 71 SNPs
	4-Hydroxyphenylpyruvate dioxygenase					
	5'-end EST (BP922875)	8 samples			28	3 assays / 28 SNPs
		<i>Salix pseudolasioogyne</i> (1)	308	ER973813	SNP: 28	
		<i>P. alba</i> x <i>P. tremula</i> var. <i>glandulosa</i> (1)	308	ER973814		
		<i>P. tremula</i> var. <i>daurica</i> (5)	308	ER973815 ~ ER973819		
		<i>P. nigra</i> var. <i>italica</i> (1)	308	ER973820		
	3'-end EST (BP930649)	8 samples			45	4 assays / 43 SNPs
		<i>Salix pseudolasioogyne</i> (1)	386	ER973805	SNP: 43	
		<i>P. nigra</i> var. <i>italica</i> (1)	384	ER973806	INDEL: 2	
		<i>P. tremula</i> var. <i>daurica</i> (6)	384	ER973807 ~ ER973812		
LAD	<i>P. nigra</i>	14 samples			63	7 assays / 60 SNPs
	Leucoanthocyanidin dioxygenase-like protein					
	5'-end EST (BP926606)	8 samples			38	4 assays / 36 SNPs
		<i>P. nigra</i> var. <i>italica</i> (1)	252	ER973827	SNP: 36	
		<i>P. tremula</i> var. <i>daurica</i> (7)	252, 254	ER973828 ~ ER973834	INDEL: 2	
	3'-end EST (BP934164)	6 samples			25	3 assays / 24 SNPs
		<i>P. nigra</i> var. <i>italica</i> (1)	386	ER973821	SNP: 24	
		<i>P. tremula</i> var. <i>daurica</i> (5)	385, 386	ER973822 ~ ER973826	INDEL: 1	
TF	<i>P. nigra</i>	5 samples			16	2 assays / 13 SNPs
	Transcription factor like protein					
	3'-end EST (BP930753)	<i>P. alba</i> x <i>P. tremula</i> var. <i>glandulosa</i> (1)	401	ER973835	SNP: 13	
		<i>P. tremula</i> var. <i>daurica</i> (4)	399 ~ 401	ER973836 ~ ER973839	INDEL: 3	
ZF	<i>P. nigra</i>	13 samples			54	5 assays / 48 SNPs
	Zinc finger protein					
	5'-end EST (BP925347)	8 samples			35	3 assays / 30 SNPs
		<i>P. nigra</i> var. <i>italica</i> (1)	474	ER973762	SNP: 30	
		<i>P. tremula</i> var. <i>daurica</i> (7)	473 ~ 476	ER973763 ~ ER973769	INDEL: 5	
	3'-end EST (BP932979)	5 samples			19	2 assays / 18 SNPs
		<i>P. nigra</i> var. <i>italica</i> (1)	438	ER973840	SNP: 18	
		<i>P. tremula</i> var. <i>daurica</i> (4)	439	ER973841 ~ ER973844	INDEL: 1	

Table 2. Summary for SNP genotyping assay results using 99 samples of eight species from three sections in genus *Populus*. The numeric values in parenthesis are arithmetic means.

Species/section ^a	Statistics	AAP1	AOX	dhn	ERF	GS	HPD	LAD	TF	ZF	Total
<i>Populus</i> total	No. assay ^b	8	4	17	3	4	7	7	2	5	57
99 samples	Poly. assay ^c	8	3	16	3	4	7	7	2	5	55
8 species	SNP type ^d	20 (2.5)	10 (2.5)	48 (2.8)	9 (3.0)	12 (3.0)	17 (2.4)	19 (2.7)	6 (3.0)	14 (2.8)	155 (2.7)
3 sections	SNP allele ^e	16 (2.0)	7 (1.8)	34 (2.0)	6 (2.0)	8 (2.0)	14 (2.0)	14 (2.0)	4 (2.0)	10 (2.0)	113 (2.0)
	<i>Ho</i> ^f	0.3395	0.1995	0.1945	0.0448	0.2371	0.1651	0.2567	0.0817	0.1513	0.2066
	<i>He</i> ^g	0.3481	0.2210	0.3120	0.1683	0.2975	0.2849	0.3551	0.2192	0.2067	0.2916
	PIC ^h	0.2779	0.1792	0.2600	0.1512	0.2508	0.2329	0.2845	0.1883	0.1698	0.2397
<i>P. tremula</i>	No. assay	8	4	17	3	4	7	7	2	5	57
var. <i>davidiana</i>	Poly. assay	4	2	9	1	4	2	4	1	1	28
58 samples	SNP type	11 (1.4)	8 (2.0)	33 (1.9)	4 (1.3)	12 (3.0)	9 (1.3)	13 (1.9)	4 (2.0)	7 (1.4)	101 (1.8)
5 populations	SNP allele	12 (1.5)	6 (1.5)	27 (1.6)	4 (1.3)	8 (2.0)	9 (1.3)	11 (1.6)	3 (1.5)	6 (1.2)	86 (1.5)
	<i>Ho</i>	0.3319	0.2155	0.1968	0.0057	0.3836	0.2030	0.2771	0.0100	0.1098	0.2166
	<i>He</i>	0.2057	0.2328	0.1660	0.0057	0.4067	0.1411	0.2103	0.2139	0.0990	0.1829
	PIC	0.1595	0.1784	0.1351	0.0056	0.3223	0.1062	0.1695	0.1681	0.0745	0.1444
<i>Leuce</i>	No. assay	8	4	17	3	4	7	7	2	4	56
26 samples	Poly. assay	4	2	6	0	1	1	4	2	1	21
3 species	SNP type	10 (1.3)	6 (1.5)	24 (1.4)	3 (1.0)	5 (1.3)	7 (1.0)	11 (1.6)	5 (2.5)	4 (1.0)	75 (1.3)
Pa	SNP allele	12 (1.5)	6 (1.5)	24 (1.4)	3 (1.0)	5 (1.3)	8 (1.1)	11 (1.6)	4 (2.0)	5 (1.3)	78 (1.4)
Ptg	<i>Ho</i>	0.4087	0.2404	0.2131	0.0000	0.0200	0.1429	0.2403	0.1955	0.2500	0.2144
PaPtg	<i>He</i>	0.2333	0.1781	0.1455	0.0000	0.0192	0.0714	0.1948	0.2396	0.1250	0.1423
	PIC	0.1786	0.1451	0.1172	0.0000	0.0185	0.0536	0.1524	0.2106	0.0938	0.1128
<i>Aigeiros</i>	No. assay	8	4	14	3	4	6	7	2	5	53
8 samples	Poly. assay	2	1	1	2	1	0	1	1	4	13
3 species	SNP type	9 (1.1)	6 (1.5)	14 (1.0)	7 (2.3)	5 (1.3)	6 (1.0)	7 (1.0)	3 (1.5)	12 (2.4)	69 (1.3)
Pdel	SNP allele	10 (1.3)	5 (1.3)	15 (1.1)	5 (1.7)	5 (1.3)	6 (1.0)	8 (1.1)	3 (1.5)	9 (1.8)	66 (1.3)
Peu	<i>Ho</i>	0.1563	0.1250	0.0714	0.3333	0.0000	0.0000	0.1429	0.0625	0.3917	0.1289
Pni	<i>He</i>	0.0898	0.1250	0.0357	0.3333	0.0612	0.0000	0.0714	0.0586	0.3922	0.1046
	PIC	0.0712	0.0938	0.0268	0.2500	0.0537	0.0000	0.0536	0.0552	0.2960	0.0802
<i>Tacamahaca</i>	No. assay	6	4	15	3	4	6	6	2	4	50
<i>P. maximowiczii</i>	Poly. assay	5	0	9	2	0	5	6	1	1	29
7 samples	SNP type	10 (1.7)	4 (1.0)	23 (1.3)	6 (2.0)	4 (1.0)	11 (1.8)	11 (1.8)	3 (1.5)	5 (1.3)	77 (1.5)
	SNP allele	11 (1.8)	4 (1.0)	24 (1.6)	5 (1.7)	4 (1.0)	11 (1.8)	12 (2.0)	3 (1.5)	5 (1.3)	79 (1.6)
	<i>Ho</i>	0.2619	0.0000	0.1444	0.1905	0.0000	0.0754	0.2857	0.2143	0.0357	0.1410
	<i>He</i>	0.1718	0.0000	0.1204	0.1973	0.0000	0.1513	0.1939	0.1684	0.0332	0.1194
	PIC	0.1451	0.0000	0.1043	0.1592	0.0000	0.1364	0.1657	0.1400	0.0310	0.1026

^a Pa (*P. alba*); Ptg (*P. tremula* var. *glandulosa*); PaPtg (*P. alba* x *P. tremula* var. *glandulosa*); Pdel (*P. deltoides*); Peu (*P. euramericana*); Pni (*P. nigra*)

^b Number of SNP assays to complete *TaqMan* probe-based real-time PCR

^c Number of SNP assays to show polymorphisms

^d The observed number of SNP types (SNP genotypes)

^e The observed number of SNPs (SNP alleles)

^f The observed heterozygosity (Weir 1996); proportion of heterozygous individuals in a given population

^g The expected heterozygosity (Weir 1996); probability that two randomly chosen SNPs from a given population are different

^h The polymorphism information content (Bostein *et al.* 1980)

Table 3. Polymorphisms in *Populus tremula* var. *daurica* and three sections for nine drought-responsive gene loci as described by the observed number of alleles (N_A), the effective numbers of alleles (N_E) and genotypes (N_G), the observed heterozygosity (H_o), Nei's (1978) average heterozygosity (H_e), and the polymorphism information content (PIC).

Species/Section	Sample size	Gene loci	No. of assay	N_A	N_E	N_G	H_o	H_e	PIC	H-W ^a Prob	F_{IS}	F_{IT}	F_{ST} ^b
<i>P. tremula</i> var. <i>daurica</i>	58	AAP1	4	5	4.13	5	1.000	0.666	0.721	0.0000***	-0.5014	-0.2834	0.1452***
		AOX	2	4	3.34	7	0.724	0.628	0.643	0.0038***	-0.2086	-0.0553	0.1269***
Population 1	4	dhn	9	13	7.07	13	1.000	0.738	0.844	0.0000***	-0.3560	-0.1473	0.1539***
Population 2	18	ERF	1	2	1.02	2	0.017	0.011	0.017	1.0000NS	-0.0286	-0.0056	0.0223NS
Population 3	13	GS	4	6	3.04	12	0.638	0.619	0.625	0.1169NS	-0.0956	-0.0290	0.0608NS
Population 4	12	HPD	2	4	3.33	3	1.000	0.646	0.645	0.0000***	-0.5481	-0.4305	0.0759NS
Population 5	11	LAD	4	10	4.41	10	1.000	0.693	0.738	0.0000***	-0.4421	-0.2862	0.1081NS
		TF	1	2	1.75	3	0.020	0.313	0.336	0.0000***	0.9601	0.9672	0.1782**
		ZF	1	2	1.98	3	0.549	0.369	0.373	0.4795NS	-0.3432	-0.0004	0.2552***
Total Mean	9(9 ^c)		28	48									
			3.11	5.33	3.34	6.44	0.6609	0.5204	0.5491		-0.2601	-0.1007	0.1317
<i>Leuce</i>	26	AAP1	4	5	3.36	4	1.0000	0.5628	0.6508				
<i>P. alba</i>	6	AOX	2	3	1.90	3	0.6154	0.1970	0.4255				
<i>P. tremula</i> var. <i>glandulosa</i>	2	dhn	6	7	4.10	5	1.0000	0.5761	0.7191				
		ERF	3	1	-	1	-	-	-				
<i>P. alba</i> x <i>P. tremula</i> var. <i>glandulosa</i>	18	GS	1	2	1.08	2	0.0800	0.1067	0.0739				
		HPD	1	2	2.00	1	1.0000	0.5000	0.3750				
		LAD	4	6	3.98	5	1.0000	0.6564	0.7117				
		TF	2	3	1.59	3	0.3846	0.1584	0.3242				
		ZF	1	2	2.00	1	1.0000	0.5000	0.3750				
Total Mean	9(8 ^c)		24	31									
			2.67	3.44	2.50	2.78	0.7600	0.4072	0.4569				
<i>Aigeiros</i>	8	AAP1	2	3	2.46	2	1.0000	0.5417	0.5112				
<i>P. deltoides</i>	2	AOX	1	2	2.00	3	0.5000	0.1667	0.3750				
<i>P. x euramericana</i>	4	dhn	1	2	2.00	1	1.0000	0.5000	0.3750				
<i>P. nigra</i> var. <i>italica</i>	2	ERF	2	2	2.00	3	0.5000	0.1667	0.3750				
		GS	1	2	1.32	2	0.0000	0.0000	0.2149				
		HPD	6	1	-	1	-	-	-				
		LAD	1	2	2.00	1	1.0000	0.5000	0.3750				
		TF	1	2	1.13	2	0.1250	0.0729	0.1103				
		ZF	4	4	2.84	4	0.7500	0.3646	0.5815				
Total Mean	9(8 ^c)		19	20									
			2.11	2.22	1.97	2.11	0.6094	0.2891	0.3647				
<i>Tacamahaca</i>	7	AAP1	5	3	2.28	2	1.0000	0.5612	0.4650				
<i>P. maximowiczii</i>	7	AOX	4	1	-	1	-	-	-				
		dhn	9	5	3.06	3	1.0000	0.6735	0.6169				
		ERF	2	3	2.00	4	0.4286	0.5000	0.4275				
		GS	4	1	-	1	-	-	-				
		HPD	5	3	1.34	2	0.1429	0.2551	0.2401				
		LAD	6	3	2.28	2	1.0000	0.5612	0.4650				
		TF	1	2	1.51	2	0.4286	0.3367	0.2800				
		ZF	1	2	1.15	2	0.1429	0.1327	0.1239				
Total Mean	9(7 ^c)		37	23									
			4.11	2.56	1.95	2.11	0.5918	0.4315	0.3740				

^a Probability estimated from likelihood ratio (G^2) tests for Hardy-Weinberg equilibrium at each locus. Significance levels at $\alpha = 0.05, 0.01$ and 0.001 are indicated as “*”, “**”, and “***”, respectively. NS: non-significance.

^b Probability estimated from likelihood ratio (G^2) tests for population subdivision at each locus. Significance levels at $\alpha = 0.05, 0.01$ and 0.001 are indicated as “*”, “**”, and “***”, respectively. NS: non-significance.

^c The number of polymorphic gene loci