ORGINAL ARTICLE

Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in black poplar (*Populus nigra* L.) inbred lines using the candidate gene approach

Identificação de polimorfismos de nucleotídeo único e análise do desequilíbrio de ligação em linhagens consanguíneas de choupo preto (Populus nigra L.) usando a abordagem do gene candidato

Xihua Liu[1] , Hai Lin[1] , Changjun Ding[2] , Xiaohua Su[2]

[1]Fujian Provincial Key Laboratory of Resources and Environment Monitoring & Sustainable Management and Utilization, Sanming University, Sanming, Fujian, China

[2]Research Institute of Forestry, Chinese Academy of Forestry, Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration, Beijing, China

Abstract

Populus is an important model organism in forest biology, and *Populus nigra* is an important tree species in poplar breeding. This research studied the polymorphism pattern of nine candidate genes in regards to N absorption and assimilation in *P. nigra*. A total of 205 single nucleotide polymorphism (SNP) sites were found in nine genes. The average nucleotide diversity for the sequenced regions were calculated as θw=0.00525 and π_T=0.00380. The *Ka:Ks* ratios for all nine genes were lower than 1, indicating the action of purifying selection. Tests of neutrality for each gene revealed that the results reject neutral evolution at *PTNRT1.2* and *PttAMT1;2.* The nucleotide sequences of *PTNRT1.2* and *PttAMT1;2* likely accepted negative selection, causing a hitchhiking effect. A recombination analysis of the nine candidate genes revealed that the average reorganization event between SNPs in *P. nigra* r is 0.1951. Linkage disequilibrium (LD) of SNPs in five candidate genes was detected, and the result showed that LD declined rapidly within these gene regions. These outcomes were similar to those determined for other trees, suggesting that wide LD mapping in the whole genome of *Populus* might be impossible and unnecessary, but LD mapping based on candidate genes could be particularly useful in poplar breeding programs. The results showed that *P. nigra* was suitable for association analysis with candidate genes and for the development of SNP markers. This article adds knowledge on N absorption and assimilation in poplars.

Keywords: Single nucleotide polymorphism (SNP); *Populus nigra*; Candidate gene; Linkage disequilibrium.

Resumo

Populus é um importante organismo modelo em biologia florestal e *Populus nigra* é uma importante espécie arbórea para o cultivo de choupos. Esta pesquisa estuda o padrão de polimorfismo de nove genes candidatos em relação à absorção e assimilação de N em *P. nigra*. Um total de 205 sítios de polimorfismo de nucleotídeo único (SNP) foram encontrados em nove genes. A diversidade de nucleotídeos média para as regiões sequenciadas foi calculada como θw = 0,00525 e π_T = 0,00380. As razões *Ka: Ks* para todos os nove genes foram menores que 1, indicando a ação da seleção purificadora. Os testes de neutralidade para cada gene revelaram que os resultados rejeitam a evolução neutra em

PTNRT1.2 e *PttAMT1; 2*. As sequências de nucleotídeos de *PTNRT1.2* e *PttAMT1; 2 p*rovavelmente aceitaram a seleção negativa, causando um efeito de carona. Uma análise de recombinação dos nove genes candidatos revelou que o evento de reorganização médio entre SNPs em *P. nigra* r é 0,1951. O desequilíbrio de ligação (LD) de SNPs em cinco genes candidatos foi detectado, e o resultado mostrou que o LD diminuiu rapidamente dentro dessas regiões gênicas. Esses resultados foram semelhantes aos determinados para outras árvores, sugerindo que o mapeamento amplo de LD em todo o genoma de *Populus* pode ser impossível e desnecessário, mas o mapeamento de LD com base em genes candidatos pode ser particularmente útil em programas de reprodução de choupos. Os resultados mostraram que *P. nigra* foi adequada para análise de associação com genes candidatos e para o desenvolvimento de marcadores SNP. Este artigo adiciona aos nossos conhecimentos sobre a absorção e assimilação de N em choupos.

Palavras-chave: Polimorfismo de nucleotídeo único (SNP); *Populus nigra;* Gene candidato; Desequilíbrio de ligação.

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the most common form of sequence variation between individuals within the same species. Recently, the exponential increase in the number of available expressed sequence tags (ESTs) suggests that SNP-based marker systems are likely to become the marker system of choice for plant breeders and academics in the near future (Barker & Edwards, 2009). The genetic dissection of complex traits to their individual gene components is a topic of great interest in biomedical and agricultural science (Tanksley, 1993). Quantitative trait locus (QTL) mapping in crops (Tanksley, 1993) and forest trees (Sewell & Neale, 2000) have contributed greatly to our understanding of complex trait architecture, such as the number, position and size of the effect of QTLs (Neale & Savolainen, 2004). This knowledge of QTLs can be efficiently applied to marker-assisted breeding (MAB) in inbred crops, but in outcrossing forest trees, the situations in which MAB application can be used are more limited. Complex trait dissection in humans has largely been achieved by the selection of physiological functions of candidate genes or association analysis of alleles (Neale & Savolainen, 2004). Association analysis will be an effective measure in the genetic dissection of complex traits of polygenes, so the association mapping method has become an important tool in plant genetics (Simko et al., 2006). The power and resolution of association mapping depends on the extent of linkage disequilibrium (LD) in mapping populations (Simko et al., 2006).

LD is the non-random union of different sites of alleles that affects processes observed in population genetics, such as mating mode, reorganization rate and population evolution history. However, LD only exists between genetic markers that are nearby and does not extend very far. SNPs, as a type of stable, high density genetic markers, are the best genetic markers for LD studies. Information on population diversity can be obtained using SNP haplotypes and LD, which together contribute to understanding the evolution history, such as origin, migration, and genetic drift, and thus contribute to understanding the community's evolutionary course.

Nitrate (NO_3^-) and ammonium (NH_4^+) absorption are two efficient N absorption processes performed by higher plants. NH_4^+ acquisition by plant roots is thought to involve members of the NH_4^+ transporter family (AMT) that is found in plants, yeast, bacteria, and mammals (Kaiser et al., 2002). *PttAMT1.2* expression is root-specific, affected by N nutrition, and strongly increases in a N-independent manner upon ecto-mycorrhiza formation (Selle et al., 2005). Many findings indicated that in higher plants, the NO_3^- transportation protein family includes *NRT1* and *NRT2*, and their expressions have important functions in NO_3^- absorption (Orsel et al., 2002a; Von Wirén et al., 1996). Early researches had demonstrated that *NTR1* was a low-affinity transport protein and that *NRT2* was a high-affinity transport protein (Zhao et al., 2007). Regarding the utilization of inorganic nitrogen in plants, glutamine synthetase (GS) was a key enzyme that participates

in the assimilation process of nitrogen. Inorganic nitrogen must be assimilated in order to form organic nitrogen, such as glutamine and glutamate, which can be absorbed and utilized in plants (Miflin & Lea, 1980). Until now, studies on polymorphisms of the three types of genes involved in the absorption and utilization of nitrogen in *Populus nigra* have not been published.

At present, QTLs that control nitrogen in some crops had been discovered (Hirel et al., 2001), but linkage analysis with QTL mapping could identify large chromosomes regions (usually the genetic distance was approximately 5 cM away) associated with quantitative characteristics (González-Martínez et al., 2006b; Boerjan, 2005). Association analysis based on SNPs may examine specific bases in genes that are associated with the objective traits, which are directly involved in intra-family and inter-family selection in forest breeding. *P. nigra* has a large genetic variation (Legionnet et al., 1999); hybrid offspring from a cross between *Populus deltoides* Marsh (female) and *P. nigra* (male) showed particularly good stem forms and excellent growth (Csencsics et al., 2009); these hybrid offspring could be used to develop rapid, high-throughput molecular markers (Holderegger et al., 2005). Hundreds of *P. nigra* gene resources had be archived by the Chinese Academy of Forestry, which was an important genetic resource for breeding *P. nigra* with highly effective nitrogen traits. Research on SNPs in *P. nigra* has made some progress (Ding & Huang, 2008; Chu et al., 2009). Past research used 115 *P. nigra* plants and detected SNPs of genes linked with lignin and holocellulose (*4CL*, *PAL* and *CesA2*) using TaqMan to identify correlations between the SNPs (Ding & Huang, 2008). Twenty-seven SNPs were obtained in the genes, and the wood properties of the various genotypes of SNP1 were significantly different. Another study reported a total of 312 SNPs among 8,056 bp sequenced from nine drought-adaptation and photosynthesis-related gene loci and found that LD decayed rapidly (Chu et al., 2009) and was similar to the rate of decay reported in other poplars (Xu et al., 2009; Yang et al., 2009). SNP mapping of genes related to nitrogen use in *P. nigra* will also be a powerful tool for genetic mapping and phyletic evolution. In this paper, 25 *P. nigra* samples collected from Europe and Asia which were evaluated, and polymorphisms were studied in 9 genes for AMT, NRT, and GS, that are related to the absorption and utilization of nitrogen. We sequenced these 9 genes to identify single nucleotide polymorphisms and analysed the relationships between SNPs and nitrogen use efficiency to identify genetic resources resulting from high N application. This study could improve the nitrogen absorption and utilization abilities of the poplar.

MATERIALS AND METHODS

Plant materials

A set of 25 genotypes was selected for SNP discovery. These genotypes represented a wide range of genetic diversity in the *P. nigra* samples. These samples included contributions from Belgium, et al (Table 1). Young and healthy leaf tissues were sampled from all 25 genotypes in June 2008. The leaves were treated with liquid nitrogen and were quickly transferred to storage at -80°C.

Candidate gene selection

Table 2 shows the 9 candidate genes selected for SNP identification. Of these genes, *AMT1;2*, *AMT1;3* and *AMT1;5* were correlated with NH_4^+ absorption in higher plants (Gazzarrini et al., 1999; Ninnemann et al., 1994; Selle et al., 2005), *NRT1.2*, *PTNRT2.1*, *NRT2.3* and *NRT2.4* were correlated with NH_3^- absorption in higher plants (Huang et al., 1999; Orsel et al., 2002b; Quesada et al., 1993), and *GW1* and *GW2* were correlated with the assimilation process of nitrogen in plants (Man et al., 2005).

Table 1. *P.nigra* genotype used in SNP discovery

Sample	Country	Longitude(°')	Latitud(°')	Sample	Country	Longitud(°')	Latitude(°')
N181	Belgium	50 40	04 20	N8	Croatia	45 45	16 00
N185	Belgium	50 40	04 20	N40	Italy	45 43	10 13
N187	Belgium	50 40	04 20	N33	Italy	45 45	11 25
N164	Netherlands	52 20	04 45	N37	Italy	44 41	10 06
N6	U.K	51 30	00 00	N20	Romania	44 30	26 10
N49	Russia	52 00	38 00	N15	Yugoslavia	40 00	20 30
N85	Russia	55 00	85 00	N21	Spain	40 30	03 40
N17	Czech Republic	50 00	14 30	N86	Turkey	37 44	37 37
N116	Germany	52 30	13 20	N90	Turkey	37 02	37 22
N120	Germany	46 11	13 20	N93	Turkey	37 37	36 51
N109	Hungary	47 30	19 00	E001	China	48 00	87 48
N98	Hungary	47 30	19 00	W005	China	47 00	88 00
N13	Slovakia	49 00	20 00				

DNA extraction and PCR amplification

DNA was extracted from tender leaves using a plant genomic DNA extraction kit (Beijing TIANGEN) following the manufacturer's instructions. Primer pairs were designed from published sequences of homologous genes in *Populus trichocarpa* v3.0 (Tuskan et al., 2006) using the web-based program Primer 5, primer as shown in Table 2. DNA amplification was performed in a 50 ìl volume containing 100 ng genomic DNA, 10 mM (0.2 ìM) of each primer, 2.5 mM of each dNTP, 5 µl 10×Tag Buffer, and 2.5 U DNA Polymerase (TaKaRa, Japan). The reactions were incubated using a Perkin Elmer 9700 thermocycler (Applied Biosystems, USA) with the following cycling conditions: 94°C for 5 min, 35 cycles of 30 s at 94°C, 30 s at 52-68°C, and 1 min at 72°C, which was followed by a final extension of 7 min at 72°C. The PCR products were analysed for DNA quality and quantity on 1% agarose gels.

Purification and sequence analysis of PCR products

The purified PCR products that could not be sequenced directly were cloned into pGEMT-easy (Promega, TaKaRa, Japan), and at least two clones were sequenced with forward and reverse primers to eliminate PCR errors. The nucleotide sequences from both strands were obtained with an ABI 3730xl sequencer (Applied Biosystems, California, USA). The sequences were reported and scored by PHRED/PHRAP (Ewing & Green, 1998). Sets of alleles for each locus were aligned using Clustal W (Larkin et al., 2007). Polymorphic sites were identified and counted by visually inspecting each sequence alignment for each locus. The corresponding chromatogram files were assessed to confirm the absence of overlapping peaks and evaluate each polymorphic site.

Nucleotide diversity analysis and recombination

Sequence analysis of 25 genes was performed using DnaSP 4.0 (Rozas et al., 2003). Nucleotide diversity was estimated by Watterson's uw (Watterson, 1975), π and the average number of pairwise nucleotide differences among sequences in a sample (Nei & Li, 1979). The number of haplotypes h and the haplotype diversity (*Hd*) for each gene were also estimated as described by Nei & Li (1979). The minimum and number of historical recombination events (R_M) were also computed using a four-gamete test in DnaSP 4.0 (Nei & Li, 1979).

Table 2 Candidate genes involved in nitrate use efficiency

Gene	Linkage group	Total length (bp)	Coding region (bp)	Noncoding region (bp)	Indel	Primer sequence(5'-3')	Ann temperature
AMT1;2	XIX	776	570	206	0	CTTCGTTAGTGGTGCTCG TTATGCCTGGCTTCTTCT	54.0
						CCTTTGGGCACATCTGTCT ACCCGAACCACAACAAGAA	57.0
						ATATTCTAGTGCCGTACTCT CACCAGCAGCAGCATCAA	56.0
AMT1;3	VIII	1237	1237	-	2	TGAAGGTCCAAGAATCGG AATGAATCCAGGCACAGG	54.0
						TGAAGGTCCAAGAATCGG ATCCATACCCGCCATCTC	55.0
						GGCAGGCAATTTCAACGC TGGTGATAACGCCCGATT	55.0
AMT1;5	II	1277	1244	33	5	ACGAGCCAGAAATAGACC AAGCCACAATCACCAAAA	52.0
						GGCTGAGAAGTTCCATTA ACGCATCCTTGCCCTGAA	54.0
						AGCAGTAGCAGGAATCAC AATAAGAACCCAAGCAGA	52.0
NRT1.2	XIV	1987	1549	438	14	TTGTTTGCGGTTATAGTGC AGAGTTCTTGGGCTTGAT	53.0
						GGCAGTTGCTATCTTCGC GTCGCTTGTTCAATGGTG	55.0
NRT2.1	IX	1610	1451	159	7	CTACAGTCCCGACAGATAC AGACGACAGGGTGAAGAT	56.0
						ATCTTCACCCTGTCGTCT GTACTTGGTGGTGTTGGT	54.0
NRT2.3	XV	874	874	-	1	ATCCCACCGTGTTTCCCTT CCTGCAACGCCATTAGCC	58.0
						AGTTGGCACGGCTAATGG TTCCTGCTGTATGGAGTTTG	57.0
NRT2.4	IX	1624	1426	198	9	GAAGGGTGATGTTGCTAA TGGTACTTGGTGGTGTTG	53.0
						ACAGTCCCGACAGATACAA TGAGACGACAGGGTGAAG	55.0
GW1	-	305	149	156	2	TTGATGGTTCCAGCTTTG TTAATGTCAGCGGTTTCG	52.0
GW2	-	445	269	176	1	AGGGAGTTTGGAATTGAG ATTTCATGTGGAGGGTTG	52.0
		10,135	8,769	1,366			

Tests of neutrality

Statistical analyses were conducted to identify genes or amino acid sites departing from the standard neutral model of evolution. Tajima's *D*-statistic (Tajima, 1989) was computed for each locus for both the full sequence and for a sliding window (window length and step size of 100 and 25 sites, respectively) (Fu & Li, 1993). Tajima's *D*-statistic

reflects the difference between π and θw. At mutation–drift equilibrium, the expected value of *D* is close to 0. Fu-Li D test uses Coalescent simulation to compare the distribution of mutations at different evolution times. (Fu & Li, 1993). Both Tajima's *D*-statistic and Fu's Fs-test statistics can also reflect demographic changes. To compute tests that required data from an outgroup of the orthologous DNA sequence from *P. trichocarpa*, we performed Fay and Wu's H-test (Fay & Wu, 2000) based on the relative excess of high-frequency-derived alleles expected immediately after a selective sweep, the McDonald–Kreitman (MK) test (McDonald & Kreitman, 1991), and the H test (Hudson et al., 1987). The significance of these statistics was tested by 1,000 replicates of coalescent simulation based on Pi.

Estimation of linkage disequilibrium

The LD descriptive statistic *r2* value between pairs of informative SNP sites in candidate genes was calculated using Tassel 5.0. The statistical significance of r^2 was computed with a one-tailed Fisher's exact test, and Bonferroni corrections were applied for multiple comparisons. Only sites with a minimum frequency of 0.10 were considered in these calculations (Ingvarsson, 2005; Remington et al., 2001). The decay of LD with physical distance was estimated using nonlinear regression of LD between polymorphic sites, as estimated by r^2, and the distance, in base pairs, between sites. Under a simple drift-recombination model, the expected value of r^2 was E(r^2)=1/(1+ρ) (Hill & Weir, 1988), where ρ=4*Nc* was the scaled recombination rate for the gene region, *N* was the effective population size and *c* was the recombination rate. In the presence of mutations, the expectation changes to where n was the number of haplotype samples. In this method, Equation 1 was fit using pooled r^2 values with the Origin 8 package.

$$E\left(r^2\right)=\left[\frac{10+\rho}{(2+\rho)(11+\rho)}\right]\left[1+\frac{(3+\rho)(12+12\rho+\rho^2)}{n(2+\rho)(11+\rho)}\right],$$

(1)

RESULTS AND DISCUSSION

Nucleotide diversity

The total sequenced length of these regions was 10,135 bp, including 8,769 bp of coding and 1,366 bp of noncoding (5'-UTR and intron) sequences. Indels were found in all genes except *AMT1;2*. Indels were excluded from all further analyses. In total, 164 SNPs were found, and the average SNP frequency was one per 62 bp over the nine sequenced regions (Table 3). There was an average of one SNP per 81 bp within the coding sequences and one SNP per 34 bp within the noncoding sequences. In summary, a moderate level of DNA polymorphisms was found in *P. nigra* (Table 4). Genetic variation at the nucleotide level was estimated from mean nucleotide diversity (π_T=0.003 80) and from the number of segregating sites (θw=0.00525). The average silent-site diversity (π_{sil}=0.00460) and synonymous-site diversity (π_{syn}=0.00385) were higher than non-synonymous changes (π_{nonsyn}=0.00328). *Ka* and *Ks* represent the nonsynonymous and synonymous substitutions per available site, respectively. There was a striking difference between the genes. The *Ka:Ks* ratios for all nine gene were lower than 1, indicating the action of purifying selection. The *Ka:Ks* ratios for *AMT1;2*, *AMT1;3*, *AMT1;5* and *GW2* were between 0.5-1, so the evolutionary rate of these genes was moderate. The remainder had a maximum *Ka:Ks* ratio significantly lower than 0.5, so these genes are relatively conservative genes. The number of haplotypes per locus ranged from 9 to 23 among the candidate genes, and the average haplotype diversity was 0.894.

Table 3 Nucleotide variation in the nine candidate genes in *P. nigra*

Gene	Length	SNP site	Frequency (bp/SNP)	SNP in coding region	SNP in noncoding region	Silent SNPs	Non-synon SNPs
AMT1;2	776	15	51.7	15	0	3	12
AMT1;3	1237	22	56.2	22	0	3	19
AMT1;5	1277	8	159.6	8	0	2	6
NRT1.2	1987	45	44.4	20	25	30	15
NRT2.1	1610	24	67.1	19	5	10	14
NRT2.3	874	21	41.6	21	0	10	11
NRT2.4	1624	12	135.3	9	3	9	13
GW1	305	8	38.1	5	3	7	1
GW2	445	9	49.4	5	4	5	4
Total	10135	164	62.4	124	40	79	95

Silent SNPs: Synonymous and noncoding

Table 4 Estimates of nucleotide diversity in the nine candidate genes in *P. nigra*

Gene	SNP	θw	π_T	π_{sil}	π_{syn}	π_{nonsyn}	π_{nonsyn}/π_{syn}	*Ka*	*Ks*	*Ka/Ks*	Haplotypes	Haplotype diversity	RM	RM/SNP
AMT1;2	15	0.00469	0.00233	0.00187	0.00301	0.00219	0.72757	0.00219	0.002907	0.753121	11	0.701	2	0.133
AMT1;3	22	0.00494	0.00468	0.00351	0.00519	0.00451	0.86898	0.004519	0.005233	0.863421	19	0.986	4	0.182
AMT1;5	8	0.00174	0.00159	0.00081	0.00287	0.00125	0.43554	0.001268	0.002474	0.512416	13	0.929	2	0.250
NRT1.2	45	0.00955	0.00493	0.00976	0.00315	0.00185	0.58730	0.001844	0.009873	0.186772	23	0.993	8	0.178
NRT2.1	24	0.00358	0.00192	0.00443	0.00450	0.00066	0.14667	0.000665	0.004525	0.146961	12	0.853	2	0.083
NRT2.3	21	0.00774	0.00648	0.00749	0.01405	0.00379	0.26975	0.003805	0.014231	0.267352	18	0.968	7	0.333
NRT2.4	12	0.00174	0.00156	0.00436	0.00525	0.00018	0.03429	0.000180	0.004379	0.041105	13	0.911	3	0.250
GW1	8	0.00649	0.00500	0.00737	0.00357	0.00111	0.31092	0.001113	0.003646	0.305266	9	0.839	1	0.125
GW2	9	0.00611	0.00551	0.00745	0.00466	0.00347	0.74464	0.003491	0.004713	0.740717	11	0.862	2	0.222
Average		0.00525	0.00380	0.00460	0.00514	0.00211	0.45841	0.002119	0.005776	0.424126	14.3	0.894		0.1951

θw, π_T, nucleotide diversity; π_{sil}, diversity at silent sites; π_{syn}, diversity at synonymous sites; π_{nonsyn}, diversity at non-synonymous sites; *Ka*, average number of mutations at non-synonymous sites in coding regions; *Ks*, average number of mutations at synonymous sites in coding regions; and R_M, the minimum number of historical recombination events

Poplar has long growth periods in the wild and is highly heterozygous. In the long-term evolution process, poplars have formed many variation types under natural selective pressure, allowing them to adapt to changing climates, non-biology coercion, and violent habitats. Our findings show that the SNP quantity in the coding region was greater than in the noncoding region in 9 candidate genes in *P. nigra*. This finding was similar to those obtained for other poplars, indicating that coding regions of poplars had a higher conservation than other locations. However, the SNP frequency from those genes was different, indicating differences in conservation for each gene. This could occur because genes suffer different selective pressures during evolution. The nucleotide diversity of *P. nigra* (π=0.0038) was between that of *Populus tremula* (π=0.0042) (Ingvarsson, 2008) and *Populus trichocarpa*, (π=0.0018) (Gilchrist et al., 2006; Tuskan et al., 2006) due the distribution for these species. Therefore, the hereditary basis and the nucleotide diversity would be different. Our data show that *P. nigra* had similar distribution and nucleotide diversity as *P. tremula*. However, *P. trichocarpa* from North America, differed from *P. nigra*. Thus, these two species had greatly different nucleotide diversity. *Ka* and *Ks* represent nonsynonymous and synonymous substitutions per available site, respectively, and a striking difference was observed between the investigated genes. The *Ka:Ks* ratios of *P. nigra* of the nine genes were the same as those obtained for *P. tremula* (Ingvarsson, 2008), *Populus tomentosa* (Xu et al., 2009) and *Populus simonii* (Wei et al., 2009). This result explained purifying selection plays an influential role in the coding region of poplar genes.

Tests of neutrality

Tajima's *D* distributions for genes were skewed toward negative values, indicating a relative excess of low frequency variants compared with expectations under a stationary neutral model, as shown in Table 5. Both Tajima's *D* and Fu and Li's statistics were negative and significantly different from zero for *NRT1.2* and *NRT2.1*, revealing an excess of rare variants

and a greater number of haplotypes than expected, respectively. This pattern of polymorphisms was commonly associated with genetic hitchhiking or negative selection. Fay and Wu's *H* test revealed genetic hitchhiking involving *NRT1.2* with significant and negative selection involving *NRT2.1*. Fu and Li's *D* test rejected neutrality involving *AMT1;2* and Fay and Wu's *H* tests were not significant, revealing an excess of rare variants or negative selection. MK tests involving *NRT2.4* were significant, revealing interspecies missense mutations rather than the expected intra-species values. Therefore, this gene experienced selective inter-species action. However, none of the neutrality tests conducted using ZnS and HKA tests were significant, and evidence of natural selection acting on all genes were unclear.

Table 5 Neutrality tests for nine candidate genes loci in *P.nigra*

Gene	Neutrality tests					
	Tajima's *D*	Fu and Li's *D*	Fay and Wu's *H*	Kelly's Z_{nS}	MK tests *P*-value	HKA tests *P*-value
AMT1;2	-1.75750	**-2.38196***	0.17339	0.2159	0.362542	0.4914
AMT1;3	-0.20102	-0.56575	2.98095	0.0841	0.295217	0.3402
AMT1;5	-0.64164	-0.49949	0.18571	0.0661	0.505495	0.0520
NRT1.2	**-2.14380***	**-3.15087****	**-15.62667****	0.1190	0.469615	0.2894
NRT2.1	**-2.00595***	**-2.66857****	-0.70996	0.1953	0.067964	0.4155
NRT2.3	-0.73343	-0.25694	-6.79051	0.1173	0.550447	0.5229
NRT2.4	-0.65686	-0.45619	-3.13684	0.1928	**0.002646****	0.4589
GW1	-0.68743	-0.70709	-0.70709	0.0583	0.265734	0.3440
GW2	-0.33329	-0.55765	-1.03623	0.1088	0.169802	0.4924

Based on the history of evolution of the NH_4^+ transportation protein family, NH_4^+ transporters originated from a different ancestor. Therefore, in this research, the nucleotide diversity obtained for the three NH_4^+ transporters were not the same. The π_{nonsyn}/π_{syn} and *Ka*/*Ks* ratios were lesst than 1, demonstrating that nonsense mutations were favoured for Darwinian selection, implying that these genes accepted purifying selection and that the nucleotide sequences were conserved. *AMT1.2* was the only gene confirmed to perform N absorption in poplar at present, and its expression was root-specific, affected by N nutrition, and strongly increased in a N-independent manner upon ecto-mycorrhiza formation (Selle et al., 2005). Therefore, the nucleotide diversity of *AMT1:2* and other related genes can help us understand the evolutionary course of the poplar, specifically regarding polymorphisms in nitrogen absorption and utilization in plants.

Neutrality tests were conducted to identify genes or sites departing from standard neutral patterns, which included evaluating the *Ka*/*Ks* ratio and the different rates of synonymous mutations and missense mutations in genes, performing Tajima's *D*, MK and HKA tests for the presence of intra-species polymorphisms, and utilizing Fu and Li's D and Fay and Wu's *H* statistical tests evaluating coalescent theory. A selective sweep might have occurred at nine genes to reduced nucleotide variation, as shown by Tajima's *D* and Fu and Li's D tests. Both Tajima's *D* and Fu and Li's statistics were negative and significantly different for *NRT1.2* and *NRT2.1*, revealing an excess of rare variants and a greater number of haplotypes than expected, respectively. This pattern of polymorphisms was commonly associated with genetic hitchhiking or negative selection. Consistent with many other European countries, a massive introduction of genes of foreign species into the native *P. nigra* population could lower the effective population size and reduce the overall fitness of seedlings of the native *P. nigra* (Cagelli & LeEvre, 1995). It was not always possible to detect gene infiltration in the offspring of *P. nigra* on the basis of morphological traits alone (Heinze, 1997).

In the evolutionary processes of species, recombination is an important driving force to break LD, to enhance heterozygosity and to increase growth. Recombination analysis for the nine candidate genes revealed nine recombination events of SNPs in *P. nigra* at 0.1951, which was similar to the results obtained for *P.tomentosa* (Xu et al., 2009), *Arabidopsis thaliana* (0.08-0.15) (Kuittinen & Aguade, 2000) and *Pinus sylvestris* (0.25) (Dvornyk et al., 2002). Analysing the population's evolutionary history, distribution range, mating system and reproduction ability, outcrosses and broad distribution range revealed a high recombination rate in *P. sylvestris*, thus producing a high reorganization rate (Dvornyk et al., 2002). Selfing results in a low recombination rate in *Arabidopsis thaliana* (Kuittinen & Aguade, 2000). Recombination increases the diversity of SNPs and breaks LD, so with the extension of nucleotide sequences, LD in SNPs

should decay. Large effective population sizes in conifers would result in low LD due to high recombination rates at the population level (González-Martínez et al., 2006a). The low LD findings discussed above can be accounted for by the large effective population sizes of conifers (Neale & Savolainen, 2004); this result agrees with the results obtained for *P. tremula* (Ingvarsson, 2008), *P. tomentosa* (Xu et al., 2009; Yang et al., 2009) and *P. simonii* (Wei et al., 2009).

Decay of linkage disequilibrium

To estimate the LD decay, 2862 informative sites were evaluated using the r^2 measure. Significance tests using Fisher's exact test showed 270 sites were significant at the 0.05 level. A low level of LD was observed within five genes. LD decayed rapidly within genes, with r^2 dropping below 0.2 within 600 bp, as shown in Figure 1.

At least 9 haplotypes (Table 4) were found in each gene, revealing a wide range of variation that could be caused by the rearranging of polymorphic sites via recombination events. Recombination could explain the rapid LD decay (Zein et al., 2007). In this research, LD declines rapidly with distance in all five genes, average linkage disequilibrium declined to negligible levels ($r^2<0.2$) at <600 bp. The LD rate dictated the type of association analysis. The rate of LD decay determined if genome scanning or candidate gene evaluation was used to perform the association analysis. When the genome LD level was high, the number of markers associated with mapping was low, and the resolution was low, which indicated that scanning the genome was a suitable approach. Otherwise, LD declined rapidly, making it difficult to scan the genome. Therefore, high density and high resolution markers were required for association mapping (Rafalski & Morgante, 2004), the obtained result was similar to that obtained for other poplars. Therefore, in poplars, a candidate gene-based strategy for association mapping is favored.
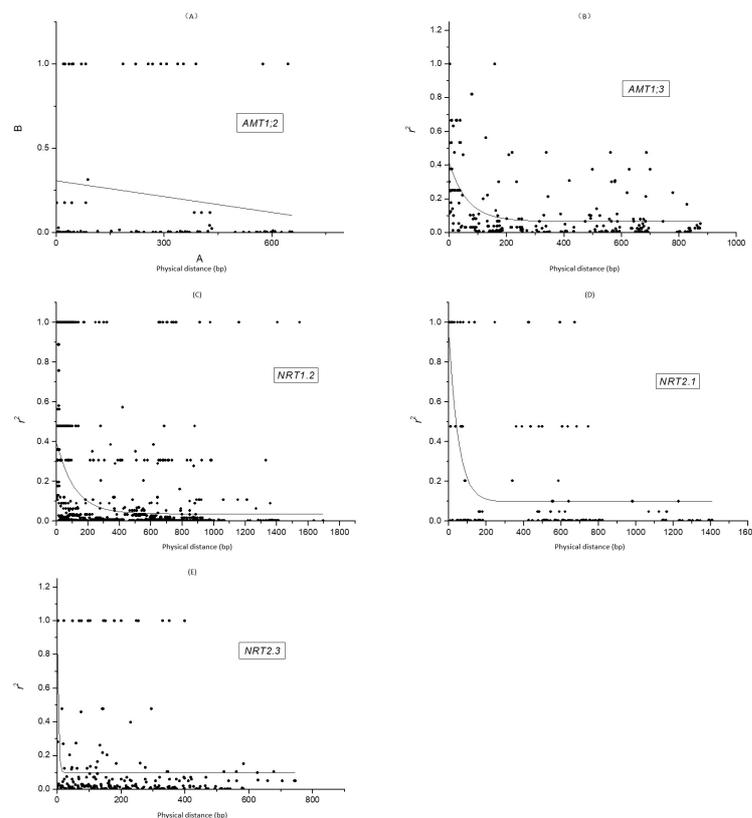

Figure 1. Decay of LD with distance in the base pairs between sites in five candidate genes. (A)Decay of LD with distance in the base pairs between sites in *AMT1;2*. *(B)* Decay of LD with distance in the base pairs between sites in *AMT1;3*. *(C)* Decay of LD with distance in the base pairs between sites in *NRT1;2*. *(D)* Decay of LD with distance in the base pairs between sites in *NRT2;1*. *(E)* Decay of LD with distance in the base pairs between sites in *NRT2;3*

ACKNOWLEDGMENTS

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest: The authors have no conflict of interest to declare.

Research involving human and animal rights: This research did not involve animal or human subjects.

REFERENCES

Barker, G. L. A., & Edwards, K. J. (2009). A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. *Plant Biotechnology Journal*, 7(4), 318-325. PMid:19386040. http://dx.doi.org/10.1111/j.1467-7652.2009.00412.x.

Boerjan, W. (2005). Biotechnology and the domestication of forest trees. *Current Opinion in Biotechnology*, 16(2), 159-166. PMid:15831381. http://dx.doi.org/10.1016/j.copbio.2005.03.003.

Cagelli, L., & LeEvre, F. (1995). The conservation of *Populus nigra* L. and gene flow with cultivated poplars in Europe. *Forest Genetics*, 2(3), 135-144.

Chu, Y., Su, X., Huang, Q., & Zhang, X. (2009). Patterns of DNA sequence variation at candidate gene loci in black poplar (*Populus nigra* L.) as revealed by single nucleotide polymorphisms. *Genetica*, 137(2), 141-150. PMid:19484494. http://dx.doi.org/10.1007/s10709-009-9371-1.

Csencsics, D., Angelone, S., Paniga, M., Rotach, P., Rudow, A., Sabiote, E., Schwab, P., Wohlhauser, P., & Holderegger, R. (2009). A large scale survey of *Populus nigra* presence and genetic introgression from non-native poplars in Switzerland based on molecular identification. *Journal for Nature Conservation*, 17(3), 142-149. http://dx.doi.org/10.1016/j.jnc.2009.01.004.

Ding, M. M., & Huang, Q. J. (2008). Analysis on SNPs linked with wood properties of *Populus nigra* L. gene resources. *Hereditas*, 30(6), 795-800. PMid:18550506.

Dvornyk, V., Sirvio, A., Mikkonen, M., & Savolainen, O. (2002). Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris. Molecular Biology and Evolution*, 19(2), 179-188. PMid:11801746. http://dx.doi.org/10.1093/oxfordjournals.molbev.a004070.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, 8(3), 186-194. PMid:9521922. http://dx.doi.org/10.1101/gr.8.3.186.

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, 155(3), 1405-1413. PMid:10880498.

Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133(3), 693-709. PMid:8454210. http://dx.doi.org/10.1093/genetics/133.3.693.

Gazzarrini, S., Lejay, L., Gojon, A., Ninnemann, O., Frommer, W. B., & Von Wirén, N. (1999). Three functional transporters for constitutive, diurnally regulated, and starvation-induced uptake of ammonium into *Arabidopsis* roots. *The Plant Cell*, 11(5), 937-948. PMid:10330477. http://dx.doi.org/10.1105/tpc.11.5.937.

Gilchrist, E. J., Haughn, G. W., Ying, C. C., Otto, S. P., Zhuang, J., Cheung, D., Hamberger, B., Aboutorabi, F., Kalynyak, T., Johnson, L., Bohlmann, J., Ellis, B. E., Douglas, C. J., & Cronk, Q. C. (2006). Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa. Molecular Ecology*, 15(5), 1367-1378. PMid:16626459. http://dx.doi.org/10.1111/j.1365-294X.2006.02885.x.

González-Martínez, S. C., Ersoz, E., Brown, G. R., Wheeler, N. C., & Neale, D. B. (2006a). DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in Pinus taeda L. *Genetics*, 172(3), 1915-1926. PMid:16387885. http://dx.doi.org/10.1534/genetics.105.047126.

González-Martínez, S. C., Krutovsky, K. V., & Neale, D. B. (2006b). Forest-tree population genomics and adaptive evolution. *The New Phytologist*, 170(2), 227-238. PMid:16608450. http://dx.doi.org/10.1111/j.1469-8137.2006.01686.x.

Heinze, B. (1997). A PCR marker for a *Populus deltoides* allele and its use in studying introgression with native European *Populus nigra. Belgian Journal of Botany*, 129, 123-130.

Hill, W. G., & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, *33*(1), 54-78. PMid:3376052. http://dx.doi.org/10.1016/0040-5809(88)90004-4.

Hirel, B., Bertin, P., Quillere, I., Bourdoncle, W., Attagnant, C., Dellay, C., Gouy, A., Cadiou, S., Retailliau, C., Falque, M., & Gallais, A. (2001). Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. *Plant Physiology*, *125*(3), 1258-1270. PMid:11244107. http://dx.doi.org/10.1104/pp.125.3.1258.

Holderegger, R., Angelone, S., Brodbeck, S., Csencsics, D., Gugerli, F., Hoebee, S., & Finkeldey, R. (2005). Application of genetic markers to the discrimination of European Black Poplar (*Populus nigra*) from American Black Poplar (*P.deltoides*) and Hybrid Poplars (*P. × canadensis*) in Switzerland. *Trees*, *15*(6), 742-747.

Huang, N.-C., Liu, K.-H., Lo, H.-J., & Tsay, Y.-F. (1999). Cloning and functional characterization of an arabidopsis nitrate transporter gene that encodes a constitutive component of low-affinity uptake. *The Plant Cell*, *11*(8), 1381-1392. PMid:10449574. http://dx.doi.org/10.1105/tpc.11.8.1381.

Hudson, R. R., Kreitman, M., & Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics*, *116*(1), 153-159. PMid:3110004. http://dx.doi.org/10.1093/genetics/116.1.153.

Ingvarsson, P. K. (2005). nucleotide polymorphism and linkage disequilibrium within and among natural populations of european aspen (*Populus tremula* L., Salicaceae). *Genetics*, *169*(2), 945-953. PMid:15489521. http://dx.doi.org/10.1534/genetics.104.034959.

Ingvarsson, P. K. (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula. Genetics*, *180*(1), 329-340. PMid:18716330. http://dx.doi.org/10.1534/genetics.108.090431.

Kaiser, B. N., Rawat, S. R., Siddiqi, M. Y., Masle, J., & Glass, A. D. M. (2002). Functional analysis of an Arabidopsis T-DNA"knockout" of the high-affinity NH transporter AtAMT1:1. *Plant Physiology*, *130*(3), 1263-1275. PMid:12427993. http://dx.doi.org/10.1104/pp.102.010843.

Kuittinen, H., & Aguade, M. (2000). Nucleotide variation at the CHALCONE ISOMERASE locus in *Arabidopsis thaliana. Genetics*, *155*(2), 863-872. PMid:10835405.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947-2948. PMid:17846036. http://dx.doi.org/10.1093/bioinformatics/btm404.

Legionnet, A., Muranty, H., & Lefevre, F. (1999). Genetic variation of the riparian pioneer tree species populus nigra. II. Variation In susceptibility to the foliar rust melampsora larici-populina. *Heredity*, *82*(3), 318-327. PMid:10336707. http://dx.doi.org/10.1038/sj.hdy.6884880.

Man, H. M., Boriel, R., El-Khatib, R., & Kirby, E. G. (2005). Characterization of transgenic poplar with ectopic expression of pine cytosolic glutamine synthetase under conditions of varying nitrogen availability. *The New Phytologist*, *167*(1), 31-39. PMid:15948827. http://dx.doi.org/10.1111/j.1469-8137.2005.01461.x.

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila. Nature*, *351*(6328), 652-654. PMid:1904993. http://dx.doi.org/10.1038/351652a0.

Miflin, B. J., & Lea, P. J. (1980). Ammonia assimilation. In B. J. Miflin (Ed.), *Amino acids and their derivatives* (pp. 169-202). New York: Academic Press. http://dx.doi.org/10.1016/B978-0-12-675405-6.50010-3.

Neale, D. B., & Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends in Plant Science*, *9*(7), 325-330. PMid:15231277. http://dx.doi.org/10.1016/j.tplants.2004.05.006.

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(10), 5269-5273. PMid:291943. http://dx.doi.org/10.1073/pnas.76.10.5269.

Ninnemann, O., Jauniaux, J. C., & Frommer, W. B. (1994). Identification of a high affinity NH4+ transporter from plants. *The EMBO Journal*, *13*(15), 3464-3471. PMid:8062823. http://dx.doi.org/10.1002/j.1460-2075.1994.tb06652.x.

Orsel, M., Filleur, S., Fraisier, V., & Daniel-Vedele, F. (2002a). Nitrate transport in plants: which gene and which control. *Journal of Experimental Botany*, *53*(370), 825-833. PMid:11912225. http://dx.doi.org/10.1093/jexbot/53.370.825.

Orsel, M., Krapp, A., & Daniel-Vedele, F. (2002b). Analysis of the NRT2 nitrate transporter family in *Arabidopsis*: structure and gene expression. *Plant Physiology*, *129*(2), 886-896. PMid:12068127. http://dx.doi.org/10.1104/pp.005280.

Quesada, A., Galvan, A., Schnell, R. A., Lefebvre, P. A., & Fernandez, E. (1993). Five nitrate assimilation-related loci are clustered in *Chlamydomonas reinhardtii. Molecular & General Genetics*, *240*(3), 387-394. PMid:8413188. http://dx.doi.org/10.1007/BF00280390.

Rafalski, A., & Morgante, M. (2004). Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics*, *20*(2), 103-111. PMid:14746992. http://dx.doi.org/10.1016/j.tig.2003.12.002.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler 4th, E. S. (2001). Structure of linkage disequilibriumand phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11479-11484. PMid:11562485. http://dx.doi.org/10.1073/pnas.201394398.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., & Rozas, R. (2003). DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*(18), 2496-2497. PMid:14668244. http://dx.doi.org/10.1093/bioinformatics/btg359.

Selle, A., Willmann, M., Grunze, N., Gessler, A., Weiss, M., & Nehls, U. (2005). The high affinity poplar ammonium importer *PttAMT1;2* and its role in ectomycorrhizal symbiosis. *The New Phytologist*, *168*(3), 697-706. PMid:16313651. http://dx.doi.org/10.1111/j.1469-8137.2005.01535.x.

Sewell, M. M., & Neale, D. B. (2000). Mapping quantitative traits in forest trees. In S. M. Jain & S. C. Minocha (Eds.), *Molecular biology of woody plants* (Forestry Sciences, No. 64, pp. 407-423). Dordrecht: Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-94-017-2311-4_17.

Simko, I., Haynes, K. G., & Jones, R. W. (2006). Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics*, *173*(4), 2237-2245. PMid:16783002. http://dx.doi.org/10.1534/genetics.106.060905.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595. PMid:2513255. http://dx.doi.org/10.1093/genetics/123.3.585.

Tanksley, S. D. (1993). Mapping polygenes. *Annual Review of Genetics*, *27*(1), 205-233. PMid:8122902. http://dx.doi.org/10.1146/annurev.ge.27.120193.001225.

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G. L., Cooper, D., Coutinho, P. M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J. C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D. R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C. J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y., & Rokhsar, D. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, *313*(5793), 1596-1604. PMid:16973872. http://dx.doi.org/10.1126/science.1128691.

Von Wirén, N., Gazzarini, S., & Frommer, W. B. (1996). Regulation of mineral nitrogen uptake in plants. *Plant and Soil*, *196*, 191-199.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*(2), 256-276. PMid:1145509. http://dx.doi.org/10.1016/0040-5809(75)90020-9.

Wei, Z., Guo, Q., Li, B., Zhang, J., & Zhang, D. (2009). Isolation, expression and single nucleotide polymorphisms (SNPs) analysis of *GA20Ox* gene in *Populus simonii. Linye Kexue*, *45*(4), 19-27.

Xu, B., Yang, X., Li, B., Zhang, Z., & Zhang, D. (2009). Isolation, expression and single nucleotide polymorphisms analysis of cellulose synthase gene (*PtCesA4*) from *Populus tomentosa. Linye Kexue*, *45*(5), 1-10.

Yang, X., Zhang, Y., Zhang, Z., Li, B., & Zhang, D. (2009). Isolation and single nucleotide polymorphisms analysis of stem cell organizer gene *wuschel* in *Populus tomentosa. Linye Kexue*, *45*(1), 43-49.

Zein, I., Wenzel, G., Andersen, J. R., & Lubberstedt, T. (2007). Low level of linkage disequilibrium at the COMT (Caffeic acid O-methyl transferase) locus in European maize (*Zea mays* L.). *Genetic Resources and Crop Evolution*, *54*(1), 139-148. http://dx.doi.org/10.1007/s10722-005-2637-2.

Zhao, S., Zhao, X., & Shi, W. (2007). Advance in research on molecular mechanism for nitrogen absorption in higher plants. *Soils*, *39*(2), 173-180.

Authors' contributions: XL: Data curation, Formal Analysis, Investigation, ethodology, Writing – original draft, Writing – review & editing, Project administration; HL: Data curation, Formal analysis; CD: Data curation, Formal analysis; XS: Conceptualization, Project administration