

Molecular Evolution of Regulatory Genes in Spruces from Different Species and Continents: Heterogeneous Patterns of Linkage Disequilibrium and Selection but Correlated Recent Demographic Changes

Marie-Claire Namroud · Carine Guillet-Claude · John Mackay · Nathalie Isabel · Jean Bousquet

Received: 2 July 2009 / Accepted: 8 March 2010 / Published online: 31 March 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Genes involved in transcription regulation may represent valuable targets in association genetics studies because of their key roles in plant development and potential selection at the molecular level. Selection and demographic signatures at the sequence level were investigated for five regulatory genes belonging to the *knox-I* family (*KN1*, *KN2*, *KN3*, *KN4*) and the *HD-Zip III* family (*HB-3*) in three *Picea* species affected by post-glacial recolonization in North America and Europe. To disentangle neutral and selective forces and estimate linkage disequilibrium (LD) on a gene basis, complete or nearly complete gene sequences were analysed. Nucleotide variation within species, haplotype structure, LD, and neutrality tests, in addition to coalescent simulations based on Tajima's *D* and Fay and Wu's *H*, were estimated. Nucleotide diversity was generally low in all species (average $\pi = 0.002\text{--}0.003$) and much heterogeneity was seen in LD

and selection signatures among genes and species. Most of the genes harboured an excess of both rare and frequent alleles in the three species. Simulations showed that this excess was significantly higher than that expected under neutrality and a bottleneck during the Last Glacial Maximum followed by population expansion at the Pleistocene/Holocene boundary or shortly after best explains the correlated sequence patterns. These results indicate that despite recent large demographic changes in the three boreal species from two continents, species-specific selection signatures could still be detected from the analysis of nearly complete regulatory gene sequences. Such different signatures indicate differential subfunctionalization of gene family members in the three congeneric species.

Keywords Bottleneck · *HD-Zip III* · *knox-I* · Linkage disequilibrium · *Picea* · Population expansion · Selection signatures

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession nos. DQ257690–DQ260227.

Electronic supplementary material The online version of this article (doi:10.1007/s00239-010-9335-1) contains supplementary material, which is available to authorized users.

M.-C. Namroud · C. Guillet-Claude · J. Mackay · N. Isabel · J. Bousquet (✉)
Arborea and Canada Research Chair in Forest and Environmental Genomics, Forest Research Centre and Institute for Systems and Integrative Biology, Université Laval, Québec, QC G1V 0A6, Canada
e-mail: jean.bousquet@sbf.ulaval.ca

M.-C. Namroud
e-mail: marie-claire.namroud@sbf.ulaval.ca

N. Isabel
Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Québec, QC G1V 4C7, Canada

Introduction

Patterns of DNA variation are often used to identify the evolutionary forces shaping species genetic structure. DNA polymorphism is generally governed by the interplay of various forces, some having locus-specific effects such as mutation, recombination, and selection, others having a genome-wide effect such as bottlenecks, migration, and population expansion (Charlesworth et al. 2003; Luikart et al. 2003). A departure of allele distribution from neutral expectations indicates that one or more of these forces are predominantly driving the evolution of the studied genes. Identifying the effect of these forces and understanding their interactions at the genome level have become a

cornerstone of modern evolutionary biology to address key questions such as the gene's contribution to species fitness, local adaptation, species divergence, and population demography (Storz 2005; Wright and Gaut 2005; Namroud et al. 2008; Stinchcombe and Hoekstra 2008).

In species like conifers, identifying the effect of each of these forces remains particularly challenging. On one hand, conifers have a large genome size (Wakamiya et al. 1993; Murray 1998; Siljak-Yakovlev et al. 2002) and linkage disequilibrium (LD) decays rapidly within their genes (Brown et al. 2004; Neale and Savolainen 2004; Heuertz et al. 2006), which makes the detection of locus-specific effects difficult without using a very large number of markers. On the other hand, conifers have a long generation time and as a result, demographic events may change the frequency of their ancestral and newly derived mutations in a way similar to that produced by locus-specific forces, one illustrating example being the excess of rare-derived variants resulting from both population expansion and positive selection (Nielsen 2005; Wright and Gaut 2005). Studies that tackled such questions often relied on a series of parameters and tests to identify the presence of natural selection, mainly: the level of nucleotide diversity in DNA sequences, haplotypic diversity, allele frequency spectrum, and more commonly, the extent of LD as a direct evidence of positive selection. However, in conifers, most of these tests have been performed on a limited number of nonregulatory genes with relatively short sequences, which reduced the number of informative sites and the detection power of selection signatures, especially when demographic forces may have eroded selection signatures (e.g. Heuertz et al. 2006; Pyhäjärvi et al. 2007). Because of the limited number of informative sites, LD is also usually estimated as an average over many genes which can hinder the utility of LD as an indicator of gene-specific processes (e.g. Brown et al. 2004; Neale and Savolainen 2004; González-Martínez et al. 2006; Heuertz et al. 2006).

In this study, our objective was to assess the extent to which the analysis of complete or nearly complete sequences of nuclear genes involved in transcription regulation allows the identification of natural selection in the presence of potential demographic effects related to the glacial and post-glacial history of the taxa investigated. Also, longer gene sequences carry a larger number of informative sites, and therefore provide an interesting opportunity to estimate LD on a gene-by-gene basis and assess heterogeneity among genes and species. Four of the genes investigated in this study belonged to the *knox-I* family (*KN1*, *KN2*, *KN3*, and *KN4*) and one to the homeodomain-leucine zipper III family (*HB-3*), and their sequences ranged from 2,101 to 5,886 bp. *Knox-I* genes are commonly involved in plant development and architecture (for review, Hake et al. 2004) and the control of shoot

apical meristem during embryogenesis (Ito et al. 2002). Homeodomain-leucine zipper (*HD-Zip*) genes are important for vascular development (Zhong and Ye 1999; Baima et al. 2001) and the establishment of functional apical meristems (Emery et al. 2003). In particular, *knox-I* genes have been shown to be under diversifying selection during their duplication history (Guillet-Claude et al. 2004). As a result, they might be of special interest for investigations on the implication of regulatory genes in adaptive natural variation in plants and trees.

To assess the consistency of our results, we further compared the genetic patterns for these genes among three largely distributed boreal species from two different continents: white spruce (*Picea glauca* [Moench] Voss) and black spruce (*Picea mariana* (Mill.)) in eastern North America, and Norway spruce (*Picea abies* L.) in Central Europe. The three species are taxonomically and phylogenetically distant (Bouillé and Bousquet 2005; Ran et al. 2006) and have been displaced during the Pleistocene glaciations and the ensuing Holocene recolonization. In addition to being prevalent species in their ecosystem, they are also characterized by a highly outcrossing mating system, wind pollen dispersal, large population sizes, and a long generation time (Bouillé and Bousquet 2005). Our specific objectives were to: (1) determine the levels of nucleotide polymorphism, haplotype structure, and LD on a gene basis in different congeneric species, (2) identify potential selection signatures and whether they are correlated among species, and (3) determine whether demographic forces are simultaneously affecting the genome of the three boreal species.

Materials and Methods

Plant Material

Needles and seeds were collected from 96 mature trees of *P. glauca* grown in a breeding orchard and representing various seed sources from Québec and Ontario in Canada (Canadian Forest Service, Beaulieu 1996), 26 mature trees of *P. mariana* grown in a test plantation and representing various seed sources across eastern Canada (Canadian Forest Service; Beaulieu et al. 1989), and 23 mature trees of *P. abies* grown in a test plantation and representative of several seed sources from central European countries, mainly Poland (Canadian Forest Service; Corriveau et al. 1988). *P. abies* samples were all from the same regional population to avoid the reported large-scale population structuring in various geographical groups across its natural range (Collignon et al. 2002; Heuertz et al. 2006). Population differentiation estimates for nuclear genes are usually weak and statistically nonsignificant for *P. glauca* and

P. mariana from the region sampled in central-eastern Canada (e.g. Isabel et al. 1995; Perry and Bousquet 2001; Jaramillo-Correa et al. 2001; Gamache et al. 2003), and for *P. abies* from Central Europe where the sampling for this study was conducted (e.g. Goncharenko et al. 1994; Collignon et al. 2002; Acheré et al. 2005). To further assess a priori the presence of possible geographical structure in our samples, pairwise two-parameter substitution rates (Kimura 1980) were estimated from concatenated sequence data of the five genes analysed for each haplotype, and neighbour-joining analysis (Saitou and Nei 1987) with 500 bootstrap replicates was conducted for each species using MEGA v.4 software (Tamura et al. 2007). No significant geographical structuring was found within the samples for each of the three taxa (results not shown), in agreement with published findings. Similar findings were found using a Bayesian approach implemented in the software BAPS v.5.1 (Corander and Tang 2007; Corander et al. 2008).

DNA Extraction, Primer Design, and PCR Amplification

For each individual, DNA was isolated from 60 to 70 mg of needles and from a haploid seed megagametophyte using a Dneasy Plant Mini Kit (Qiagen, Mississauga, Ontario). The primer sequences and PCR conditions used to amplify the coding regions of the *knox-1* genes are described by Guillet-Claude et al. (2004). A walking PCR procedure was used to amplify the 5'-flanking regions of *KN2* and *KN4* (800 and 500 bp of the promoter regions, respectively). It is described in the Supplementary Information Box. The eight overlapping-specific primer pairs used to amplify *HB-3* are described in Supplementary Table 1. PCR reactions for *HB-3* genes were performed in 30 µl containing 5–20 ng of genomic DNA, 20-mM Tris-HCl (pH 8.4), 50-mM KCl, 1.5–2.0-mM MgCl₂, 200 µM of each dNTP, 200 µM of both 5' and 3' primers, and 1.0 unit of Platinum *Taq* DNA Polymerase (Invitrogen, Carlsbad, California). Amplification was performed in a peltier thermal cycler (DYADTM DNA Engine, MJResearch, Waltham, Massachusetts) with the following thermal cycling profile: 4 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at annealing temperature optimized between 58 and 60°C for each pair of primers, and 1 min at 72°C, followed by 10 min at 72°C.

DNA Sequencing

Both haploid (megagametophytes) and diploid (needles) DNA were amplified for each individual and gene. The PCR products were sequenced in both directions with a Perkin-Elmer ABI 3730 XL DNA sequencer (Applied Biosystems, Foster City, California) using

BigDye Terminator Cycle Sequencing Kits Version 3.1. Contigs were constructed with Windows 32 SeqMan 5.05 (DNASTAR Inc., Madison, Wisconsin) and BioEdit v. 5.0.9 (Tom Hall, Department of Microbiology, North Carolina State University). The sequence of the two alleles amplified from diploid needle tissue was deduced using the allele sequenced from the haploid megagametophyte tissue. The sequences containing singletons were checked by re-amplifying and partially re-sequencing around the polymorphisms detected. Sequences were aligned using ClustalX (Thompson et al. 1997). If complex and large indels were present, the final alignment was adjusted by eye using BioEdit (Hall 1999).

Structure of the *Picea knox-1* and *HB-3* genes

The genes *KN1* and *KN2* are located on linkage group III of *Picea* spp., *KN3* on linkage group II, *KN4* on linkage group VI, and *HB-3* on linkage group IX (Pavy et al. 2008). While *KN4* was the first duplicate of the conifer *Knox-1* gene family to diverge after the split between gymnosperms and angiosperms, *KN3* diverged later, followed by *KN1* and *KN2* (Guillet-Claude et al. 2004). Supplementary Figs. 1 and 2 show the structure of each gene and the observed polymorphisms for each of the three studied *Picea* species. For each gene, we determined the sequence of 192 (or 120 for *HB-3*) alleles from the 96 (or 60 for *HB-3*) *P. glauca* individuals, 52 alleles from the 26 *P. mariana* individuals, and 46 alleles from the 23 *P. abies* individuals, for a total of 290 haploid complements. This was equivalent to about 2.2 kb of sequence in *KN1*, 3.2 kb in *KN2*, 2.5 kb in *KN3*, 2.5 kb in *KN4*, and 5.8 kb in *HB-3*, for a total of 16.3 kb for each haploid complement sampled. The total length of sequences determined for this study was 48.9 Mb. The coding regions of the four *knox-1* genes contained four introns (Supplementary Fig. 1). Because of the large size of the third intron (up to 5 kb; Vollbrecht et al. 1991; Sato et al. 2001), its sequence could not be determined and each *knox-1* gene was amplified in two fragments using gene-specific primers. The sequences of the two fragments were then concatenated in each individual, first by concatenating the sequences from the haploid megagametophyte, then by deducing the other allele from the corresponding sequences of the diploid tissue. *HB-3* had 17 short introns. Only the second intron was long and extended over 990 bp. The first intron was located in the 5'-untranslated region, just 46 nucleotides before the start codon (Supplementary Fig. 2).

Sequence Data Analyses

Unless mentioned otherwise, sequence data were analysed with the DnaSP4.00 software (Rozas et al. 2003) after

excluding insertions/deletions (indels). Nucleotide diversity was estimated by calculating the mean pairwise differences (π) (Tajima 1983) and theta-Watterson estimates based on the number of segregating sites (Watterson 1975). LD was measured as the squared allele frequency correlations (r^2) between polymorphic sites (Hill and Robertson 1968). Only informative polymorphic sites were considered in LD calculations. The significance of r^2 was tested by using Fisher's exact test and Bonferroni correction (Sokal and Rohlf 1995). LD decay was estimated with a nonlinear regression function (PROC NLIN, SAS software) of r^2 versus the distance between polymorphic sites in base pairs (Hill and Robertson 1968). The r^2 expectation was adjusted to take into account the sample size according to Hill and Weir's formula (1988). Given the length of our genes, LD decay and mean r^2 could be calculated for each gene separately. They were also calculated by pooling polymorphic sites from all genes together as is usually done in studies analysing shorter gene sequences (e.g. Brown et al. 2004; González-Martínez et al. 2006). Intra-genic recombination was estimated by calculating the minimum number of intra-genic recombination events (Rm) per informative site (Hudson and Kaplan 1985) and the maximum-likelihood estimator (ρ) based on independent linked pairs of sites (Hudson 2001) as implemented in the LDhat v. 2.0 software (McVean et al. 2002).

Haplotype structure was investigated by estimating the haplotype number (K) and diversity (Hd) for each gene based on the number of segregating sites (Fu 1997; Depaulis and Veuille 1998; Depaulis et al. 2001). The observed value of Hd was multiplied by $(n - 1)/n$ where n is the allele sample size as defined by Depaulis and Veuille (1998). The statistical significance of K and Hd values was tested by running 10,000 coalescent simulations conditional on the number of segregating sites and assuming: a neutral infinite-sites model, a large constant population size, and a recombination rate equal to the maximum-likelihood estimator (ρ) of Hudson (2001).

Neutrality Tests and Demographic Models

Departure from neutrality was estimated for each gene with the Tajima's D (1989) and Fay and Wu's H (2000) tests. Under neutrality, these parameters are expected to be null. A negative D indicates an excess of rare variants resulting from positive selection, recent population expansion, or background selection (Charlesworth et al. 1993; Fay and Wu 2000), while a negative H indicates an excess of high-frequency-derived variants resulting from a recent selective sweep (Fay and Wu 2000; Przeworski 2002) or a recent bottleneck (Przeworski 2002). In all cases, the outgroup was a highly matching sequence of *Pinus taeda*.

Despite the breadth of our data set with multiple species and loci, we could not assess the departure from neutrality with two traditional tests: (1) the multilocus Hudson–Kreitman–Aguadé (HKA) test (Hudson et al. 1987), which examines the neutral expectation of correlated levels of polymorphism and divergence across loci; and (2) the McDonald–Kreitman (MK) test which compares the ratio of nonsynonymous to synonymous substitutions within species to that between species (McDonald and Kreitman 1991). The HKA is a conservative statistical test that assumes no intra-genic recombination (Hudson et al. 1987), while an important recombination was observed in our genes (see “Results”). The MK test has little power to reject neutrality (Charlesworth and Eyre-Walker 2008) and in many of our genes, the number of fixed nonsynonymous differences between species was null or not large enough to conduct the test.

In an attempt to distinguish the effects of natural selection from those of demographic factors, we used several approaches. First, we compared the polymorphic patterns observed at the gene level, mainly nucleotide diversity (π), Tajima's D , and Fay and Wu's H with those observed in the noncoding regions of each gene. The rationale is that natural selection is expected to act on expressed gene loci and their linked regions, resulting in a polymorphic pattern different from that observed in non-linked regions. By contrast, similar patterns among different gene regions were expected if only demographic changes had affected the whole genome. This is particularly worth considering in this study, given the length of the genes analysed and the generally low LD within conifer genes. Second, multilocus coalescent simulations were conducted to examine whether the means and variances of D and H across loci (after pooling the five genes of each species) deviated significantly from those expected under a standard neutral equilibrium (SNM). The latter was constructed by running coalescent simulations conditional on θ using the ms software, assuming an infinite-site mutation model, and a constant population size (Hudson 2002). For each species, we simulated 10,000 replicates of the five genes with their respective sample sizes and number of sites, and we assumed that the population mutation and recombination parameters were constant among the genes and equal to the averages per site of Watterson theta (θ_w) and Hudson's (2001) estimates (ρ), respectively. We then multiplied these averages by the length of each gene to take into account the differences in the sequenced gene length. Average π was also considered to determine the fit of the model, and θ was modified in some cases to maintain a simulated average π value (per gene) similar to the observed one (within 5% of the empirical value). The significance of the multilocus estimates (means and variances of D and H) was tested for each species using the

Analysers HKA6 and multitest_pop1 programs, respectively. These programs were kindly provided by Peter Andolfatto and are available at http://genomics.princeton.edu/AndolfattoLab/program_analyser.html.

In addition, we tested the fit of our data to different bottleneck models followed by exponential population expansion. These models were constructed and tested as described for the SNM model, but they were simulated over a grid of parameter values for the bottleneck's severity and age. The severity was measured in units of the current population size N_e . It varied from $0.0004N_e$ to $0.5N_e$ which was equivalent to a population size varying between 40 and 50,000 individuals. The age of the bottleneck (from present to the time the bottleneck started) was measured in units of $4N_e$ generations and varied from 0.0005 to 0.005, which was equivalent to a time span of 10,000 to 100,000 years. The length of the bottlenecks varied from 2,000 to 83,000 years. In all these simulations, we assumed that the generation time was 50 years (we took into account the mortality in a stand and the longevity of the species), the divergence time from the common ancestor at 15 million years, and the population size N_e equal to 100,000 (Bouillé and Bousquet 2005).

Results

Intraspecific Nucleotide Diversity

Nucleotide diversity of the five *knox-I* and *HB-3* genes was lower at nonsynonymous than at synonymous sites in all genes and species (Tables 1, 2, 3), except for *KN4* of *P. abies* where the ratio of nonsynonymous to synonymous nucleotide diversity (π_a/π_s) reached 1.45, which could be indicative of selection. When comparing the five genes in each species, total nucleotide diversity was lowest in *HB-3* of *P. glauca* and *P. mariana*. In *P. abies*, although the lowest total nucleotide diversity was observed in *KN3*, the lowest nonsynonymous nucleotide diversity was also in *HB-3*. In general, nucleotide diversity was comparable between *P. glauca* and *P. mariana*, but lower in *P. abies*.

The largest proportions of indels (87%) and substitutions (55% to 87%) were located in the noncoding regions (Supplementary Figs. 1 and 2). Among the 60 indels identified in the four *knox-I* genes, 21 were singletons, of which 50% were single-base. The remaining indels were relatively large and ranged from 25 to 54 bp, mostly in introns. Ten indels (3–9 bp) were located in the coding regions but none of them induced a frameshift (Tables 1, 2, 3). Most of them corresponded to the deletion or insertion of a repeat motif. In *HB-3*, all of the 17 indels identified were located outside the exons. Four of them were singletons and their size ranged from one to six bases.

Coalescent simulations confirmed a significant excess of singletons in *KN3* of *P. glauca*, *KN1* and *KN2* of *P. mariana*, and in all *P. abies* genes except *KN4*, which harboured a high nonsynonymous to synonymous substitutions ratio (Tables 1, 2, 3).

Intraspecific Nonsynonymous Changes

Knox-I proteins had a structural organization similar to that found in other plants. It included a conserved KNOX domain, a highly conserved ELK domain, and a homeodomain (HD) responsible for DNA binding (Ito et al. 2002; Supplementary Fig. 1). The *HB-3* protein was also similar to that observed in other plant class III HD-Zip proteins. It had an N-terminal homeodomain/leucine zipper (HD-ZIP) followed by a region with a sequence similar to that in the mammalian sterol/lipid-binding domains (START domain). Most of the C-terminus was well conserved among the HD-Zip III proteins although its function remains unknown (Sessa et al. 1998; Prigge et al. 2005).

A total of 66 amino acids replacements were identified in the *knox-I* genes. Most of these replacements (71.2%) were in the N-terminal nonconserved regions. Some (19.7%) were located in the KNOX, but none or very few (7.6%) in the ELK and HD domains, respectively. The largest number of nonsynonymous replacements was in the *KN1* gene of *P. glauca* that contained 16 amino acid replacements including 12 in the N-terminal nonconserved region, one in the KNOX and three in the HD domains. The least polymorphic gene was *KN3* of *P. abies* that harboured three singleton nonsynonymous substitutions and one singleton as a 3-bp indel in its N-terminal nonconserved region. In all species and genes, the number of nonsynonymous substitutions was far lower than the number of synonymous substitutions, except in *KN4* of *P. abies* which harboured five nonsynonymous and two synonymous substitutions in its exons.

Linkage Disequilibrium and Haplotype Structure

The proportion of significant LD estimates ranged from 1.6 to 22.0% for *knox-I* genes, and from 1.0 to 16.4% for *HB-3* genes. The *P*-level ranged from 0.001 to 0.05 following Fisher's exact test with Bonferroni's correction. Mean squared allele frequency correlation r^2 was about 0.45 in all genes and species and declined rapidly to half between a few base pairs to around 2,000 bp depending on the gene and species (Fig. 1a–c). However, *HB-3* of *P. mariana* maintained a relatively high LD, up to 2,000 bp (Fig. 1b), which was consistent with its relatively low recombination rates (ρ) and haplotype number and diversity (Table 4). Overall, much heterogeneity was noted among congeneric species for the same gene. For instance, LD in *KN1* of

Table 1 Estimates of nucleotide and indel diversity for the *knox-1* and *HB-3* genes of *P. glauca*

Gene	Coding region		Noncoding region	Silent sites ^a	Total
	Synonymous	Nonsynonymous			
<i>KN1</i> (<i>n</i> = 192) ^b					
No of sites ^c	244.2	889.8	1129	1373.2	2263
SNPs (singletons) ^d	7	16 (4)	29 (8)	36 (8)	52 (12)
Indels (size) ^e	0	0	6 (Σ = 66 bp: 41, 19, 3, 1, 1, 1)	6 (Σ = 66 bp: 41, 19, 3, 1, 1, 1)	6 (Σ = 66 bp: 41, 19, 3, 1, 1, 1)
π	0.00150	0.00060	0.00461	0.00403	0.00264
θ	0.00492	0.00308	0.00468	0.00472	0.00406
<i>KN2</i> (<i>n</i> = 192)					
No of sites	247.7	889.3	2363	2610.7	3500
SNPs (singletons)	13	8 (5)	67 (13)	80 (13)	88 (18)
Indels (size)	0	1 (3 bp)	14 (Σ = 26 bp: 7, 5, 3, 2, 8 \times 1, 1)	14 (Σ = 26 bp: 7, 5, 3, 2, 8 \times 1, 1)	15 (Σ = 29 bp: 7, 5, 2 \times 3, 2, 8 \times 1, 1)
π	0.00909	0.00060	0.00382	0.00433	0.00338
θ	0.00900	0.00155	0.00492	0.00531	0.00435
<i>KN3</i> (<i>n</i> = 192)					
No of sites	273.7	989.3	841	1114.7	2104
SNPs (singletons)	11 (3)	3 (1)	24 (9)	35 (12)	38 (13*)
Indels (size)	0	3 (Σ = 15 bp: 6, 6, 3)	4 (Σ = 41 bp: 25, 8, 7, 1)	4 (Σ = 41 bp: 25, 8, 7, 1)	7 (Σ = 56 bp: 25, 8, 7, 6, 6, 3, 1)
π	0.00349	0.00006	0.00366	0.00361	0.00192
θ	0.00689	0.00053	0.00514	0.00559	0.00318
<i>KN4</i> (<i>n</i> = 192)					
No of sites	245.6	942.4	1400	1645.6	2588
SNPs (singletons)	7 (2)	5 (2)	52 (10)	59 (12)	64 (14)
Indels (size)	0	2 (Σ = 9 bp: 3, 6)	4 (Σ = 32 bp: 28, 2, 2 \times 1)	4 (Σ = 32 bp: 28, 2, 2 \times 1)	6 (Σ = 41 bp: 28, 9, 2, 2 \times 1)
π	0.00349	0.00016	0.00543	0.00513	0.00331
θ	0.00489	0.00092	0.00652	0.00626	0.00431
<i>HB-3</i> (<i>n</i> = 120)					
No of sites	611.2	1914.8	3354	3965.2	5880
SNP (singletons)	13	2 (2)	80 (21)	93 (21)	95 (23)
Indel (size)	0	0	9 (Σ = 18 bp: 6, 4, 2, 6 \times 1)	9 (Σ = 18 bp: 6, 4, 2, 6 \times 1)	9 (Σ = 18 bp: 6, 4, 2, 6 \times 1)
π	0.00456	0.00002	0.00351	0.00368	0.00248
θ	0.00397	0.00019	0.00447	0.00440	0.00302
Total (<i>n</i> = 888)					
No of sites	1622.4	5625.6	9087	10709.4	16335
SNPs (singletons)	51 (5)	34 (14)	252 (61)	303 (66)	337 (80)
Indels (total size)	0	6 (27 bp)	37 (183 bp)	37 (183 bp)	43 (210 bp)
Weighted average					
π	0.00445	0.00023	0.00404	0.00410	0.00275
θ	0.00551	0.00104	0.00499	0.00507	0.00367

^a Silent sites include synonymous and noncoding sites

^b *n* is the number of sequences in the sample

^c The number of sequenced sites includes gaps induced by indels

^d The total number of substitutions with the number of singletons indicated in *parentheses*

^e The total number of indels with the size of indels indicated in *parentheses* and indels corresponding to singletons in *italics*

* $P < 0.05$ corresponding to the P value for obtaining a number of SNP singletons (under the neutral coalescent process) equal or higher than that observed

P. abies was about six and three times higher than that for *P. glauca*, and *P. mariana*, respectively; LD in *KN2* of *P. mariana* was about 150 and 50 times higher than that observed for *P. glauca* and *P. abies*, respectively; *KN3* had very small LD in *P. abies* but was three times higher for *P. glauca* than for *P. mariana*; LD in *KN4* was almost similar between *P. glauca* and *P. mariana*, but was about three times higher in *P. abies*; and LD of *HB-3* was about five and ten times higher for *P. mariana* than for *P. abies* and *P. glauca*, respectively. When pooling polymorphic sites from all genes together, LD was highest in *P. mariana* (half-decay of $r^2 \sim 430$ bp; mean $r^2 = 0.18$), but when excluding *HB-3*, LD was highest in *P. abies* (half-decay of $r^2 \sim 92$ bp; mean $r^2 = 0.11$). For all species, *knox-I* genes showed a more rapid decay of LD than *HB-3* genes (Fig. 1). In concordance with its LD estimates, *HB-3* had lower recombination rates and haplotype number and diversity than *knox-I* genes in *P. glauca* and *P. mariana* (Table 4). In *P. abies*, *HB-3* had the lowest recombination rate, but it was *KN3* that harboured the lowest haplotype number and diversity (Table 4). Averaged over the five genes, recombination was lowest in *P. glauca*.

Neutrality and Demography Tests

Neutrality tests reaching statistical significance involved one gene in *P. glauca* and *P. mariana* but several genes in *P. abies* (Table 4). In addition, all genes (except *KN2* in *P. mariana*) had negative Tajima's *D* and Fay and Wu's *H* values, reflecting an excess, although not always significant, of both rare- and high-frequency-derived variants in the three species. When comparing the three species, both *D* and *H* values were generally more negative in *P. abies* genes. Negative *D* and *H* values and comparable patterns of nucleotide diversity were observed in the noncoding (promoters, introns, 5'- and 3'-UTRs) regions of each gene (results not shown), suggesting that all gene regions were affected by similar processes.

In concordance with these observations, coalescent simulations revealed that mean Tajima's *D* and Fay and Wu's *H* were negative and significantly ($P < 0.05$) deviated from expectations under the neutral model, thus indicating a significant excess of both rare- and high-frequency-derived alleles at the multilocus level of each species (Table 5). From all the models simulated, only those based on a bottleneck that occurred during the Last Glacial Maximum (LGM) followed by a population expansion at the beginning of the Holocene or shortly after were able to adequately explain the polymorphic patterns in the three *Picea* species. For the three species, the bottleneck was signalled at around 25,000 years ago and reduced the population effective sizes to less than 1% and as low as 0.1% (best model). The beginning of the

expansion phase may have varied among the three species: the best models showed that the bottleneck ended and exponential expansion started at about 17,000 years before present for the three *Picea* species, but near-best models showed that the impact of the bottleneck may have lasted longer and the beginning of expansion could have been delayed by many thousands of years for *P. glauca* (around 10,000 years ago) and *P. abies* (between 15,000 and 5,000 years ago).

Discussion

Patterns of Nucleotide Diversity and Selection

Despite the limited diversity of genes surveyed, the nucleotide diversity of regulatory genes was generally comparable to that reported in other conifer studies using shorter sequences for a larger number of genes such as for *Pinus radiata* and *Pinus pinaster* (Pot et al. 2005), *Pinus taeda* (Brown et al. 2004; Neale and Savolainen 2004; González-Martínez et al. 2006), and Douglas fir (Krutovsky and Neale 2005). Nucleotide diversity was also comparable to that previously reported for shorter stretches of other nuclear genes in the same three *Picea* species (Bouillé and Bousquet 2005; Chen et al. 2010) and in *P. abies* (Heuertz et al. 2006), and for two nearly complete genes in *Pinus sylvestris* (García-Gil et al. 2003). Some recent studies argued against low mutation rates to explain the relatively low to moderate nucleotide diversity in conifers; they suggested that low mutation rates are often calculated based on overestimated divergence times (Heuertz et al. 2006; Pyhäjärvi et al. 2007). In this study, the limited number of genes analysed did not permit to confirm these observations. However, it is worth mentioning that the lower nucleotide diversity noted in *P. abies* compared to the two North American species might reflect a sampling difference. Indeed, samples for *P. abies* were from populations from central Europe, mainly from Poland, thus covering a smaller part of the species natural range as compared to that for the North American species. Moreover, nucleotide diversity was quite comparable among the three species in a recent study, where *P. abies* sampling represented much of the species natural range (Chen et al. 2010).

The generally much lower nucleotide diversity at nonsynonymous compared to synonymous sites (Tables 1, 2, 3) was consistent with the findings of Guillet-Claude et al. (2004) who analysed the same *knox-I* genes and attributed such a pattern to purifying selection. Purifying selection appears particularly high for *HB-3* of *P. glauca* and *P. abies* that exhibited a relatively low π_a/π_s ratio, a pattern often attributed to purifying selection (Yang and

Table 2 Estimates of nucleotide and indel diversity for the *knox-1* and *HB-3* genes of *P. mariana*

Gene	Coding region		Noncoding region	Silent sites ^a	Total
	Synonymous	Nonsynonymous			
<i>KN1</i> (<i>n</i> = 52) ^b					
No of sites ^c	243.8	890.2	1068	1311.8	2202
SNPs (singletons) ^d	6 (1)	8 (6)	20 (8)	26 (9)	34 (15*)
Indels (size) ^e	0	0	2 (Σ = 4 bp: 3, 1)	2 (Σ = 4 bp: 1, 3)	2 (Σ = 4 bp: 1, 3)
π	0.00654	0.00093	0.00229	0.00308	0.00221
θ	0.00545	0.00199	0.00416	0.00440	0.00342
<i>KN2</i> (<i>n</i> = 52)					
No of sites	248.0	889.0	2350	2598.0	3487
SNPs (singletons)	6 (1)	4 (4)	32 (12)	38 (13)	42 (17*)
Indels (size)	0	0	5 (Σ = 8 bp: 2 \times 1, 3, 2, 1)	5 (Σ = 8 bp: 2 \times 1, 3, 2, 1)	5 (Σ = 8 bp: 2 \times 1, 3, 2, 1)
π	0.00351	0.00017	0.00280	0.00286	0.00218
θ	0.00535	0.00100	0.00302	0.00325	0.00267
<i>KN3</i> (<i>n</i> = 52)					
No of sites	275.0	985.0	841	1116.0	2101
SNPs (singletons)	8 (1)	2 (1)	14 (4)	22 (5)	24 (6)
Indels (size)	0	1 (6 bp)	1 (2 bp)	1 (2 bp)	2 (8 bp: 6, 2)
π	0.00580	0.00012	0.00344	0.00402	0.00219
θ	0.00644	0.00045	0.00370	0.00437	0.00254
<i>KN4</i> (<i>n</i> = 52)					
No of sites	245.5	942.5	1400	1645.5	2588
SNPs (singletons)	10 (2)	4 (2)	30 (8)	40 (10)	44 (12)
Indels (size)	0	2 (Σ = 6 bp: 2 \times 3)	3 (Σ = 4 bp: 2, 1, 1)	3 (Σ = 4 bp: 2, 1, 1)	5 (Σ = 10 bp: 2 \times 3, 2, 1, 1)
π	0.00836	0.00036	0.00333	0.00408	0.00273
θ	0.00901	0.00095	0.00476	0.00539	0.00377
<i>HB-3</i> (<i>n</i> = 52)					
No of sites	611.2	1914.8	3352	3963.2	5878
SNP (singletons)	9 (1)	1	64 (15)	73 (16)	74 (16)
Indel (size)	0	0	1 (6 bp)	1 (6 bp)	1 (6 bp)
π	0.00155	0.00014	0.00295	0.00251	0.00174
θ	0.00326	0.00012	0.00443	0.00408	0.00279
Total (<i>n</i> = 260)					
No of sites	1623.4	5621.6	9011	10634.4	16256
SNPs (singletons)	39 (6)	19 (13)	160 (47)	199 (53)	218 (66)
Indels (total size)	0	3 (12 bp)	12 (24 bp)	12 (24 bp)	15 (36 bp)
Weighted average					
π	0.00435	0.00030	0.00294	0.00307	0.00211
θ	0.00532	0.00075	0.00401	0.00415	0.00297

^a Silent sites include synonymous and noncoding sites

^b *n* is the number of sequences in the sample

^c The number of sequenced sites includes gaps induced by indels

^d The total number of substitutions with the number of singletons indicated in *parentheses*

^e The total number of indels with the size of indels indicated in *parentheses* and indels corresponding to singletons in *italics*

* $P < 0.05$ corresponding to the P value for obtaining a number of SNP singletons (under the neutral coalescent process) equal or higher than that observed

Table 3 Estimates of nucleotide and indel diversity for the *knox-1* and *HB-3* genes of *P. abies*

Gene	Coding region		Noncoding region	Silent sites ^a	Total
	Synonymous	Nonsynonymous			
<i>KN1</i> (<i>n</i> = 46) ^b					
No of sites ^c	244.2	889.8	1121	1365.2	2255
SNPs (singletons) ^d	10 (7)	3 (2)	35 (14)	45 (21)	48 (23*)
Indels (size) ^e	0	0	3 (Σ = 56 bp: 54, 1, 1)	3 (Σ = 56 bp: 54, 1, 1)	3 (Σ = 56 bp: 54, 1, 1)
π	0.00395	0.00019	0.00438	0.00430	0.00264
θ	0.00932	0.00077	0.00748	0.00782	0.00496
<i>KN2</i> (<i>n</i> = 46)					
No of sites	247.9	889.1	2351	2598.9	3488
SNPs (singletons)	5 (2)	5 (2)	34 (18)	39 (20)	44 (22***)
Indels (size)	0	0	9 (Σ = 49 bp: 25, 11, 5, 2 × 2, 4 × 1)	9 (Σ = 49 bp: 25, 11, 5, 2 × 2, 4 × 1)	9 (Σ = 49 bp: 25, 11, 5, 2 × 2, 4 × 1)
π	0.00536	0.00088	0.00193	0.00226	0.00191
θ	0.00459	0.00128	0.00336	0.00348	0.00291
<i>KN3</i> (<i>n</i> = 46)					
No of sites	275.0	982.0	842	1117.0	2099
SNPs (singletons)	6 (4)	3 (3)	16 (8)	22 (12)	25 (15**)
Indels (size)	0	1 (3 bp)	5 (Σ = 7 bp: 2, 2, 3 × 1)	5 (Σ = 7 bp: 2, 2, 3 × 1)	6 (Σ = 10 bp: 3, 2, 2, 3 × 1)
π	0.00140	0.00013	0.00150	0.00148	0.00085
θ	0.00496	0.00070	0.00436	0.00451	0.00272
<i>KN4</i> (<i>n</i> = 46)					
No of sites	246.5	938.5	1402	1648.5	2587
SNPs (singletons)	2	5 (2)	39 (10)	41 (10)	46 (12)
Indels (size)	0	0	5 (Σ = 13 bp: 5, 4, 2, 2 × 1)	5 (Σ = 13 bp: 5, 4, 2, 2 × 1)	5 (Σ = 13 bp: 5, 4, 2, 2 × 1)
π	0.00069	0.00100	0.00587	0.00509	0.00359
θ	0.00185	0.00121	0.00639	0.00570	0.00406
<i>HB-3</i> (<i>n</i> = 46)					
No of sites	611.2	1914.8	3360	3971.2	5886
SNP (singletons)	8 (5)	1 (1)	63 (28)	71 (33)	72 (34*)
Indel (size)	0	0	7 (Σ = 11 bp: 4, 2, 1, 1, 3 × 1)	7 (Σ = 11 bp: 4, 2, 1, 1, 3 × 1)	7 (Σ = 11 bp: 4, 2, 1, 1, 3 × 1)
π	0.00102	0.00002	0.00182	0.00170	0.00115
θ	0.00298	0.00012	0.00428	0.00408	0.00279
Total (<i>n</i> = 230)					
No of sites	1624.7	5614.3	9076	10700.7	16315
SNPs (singletons)	31 (18)	17 (10)	187 (78)	218 (96)	235 (106)
Indels (total size)	0	1 (3 bp)	29 (136 bp)	29 (136 bp)	30 (139 bp)
Weighted average					
π	0.00214	0.00037	0.00276	0.00267	0.00186
θ	0.00434	0.00069	0.00477	0.00470	0.00331

^a Silent sites include synonymous and noncoding sites

^b *n* is the number of sequences in the sample

^c The number of sequenced sites includes gaps induced by indels

^d The total number of substitutions with the number of singletons indicated in *parentheses*

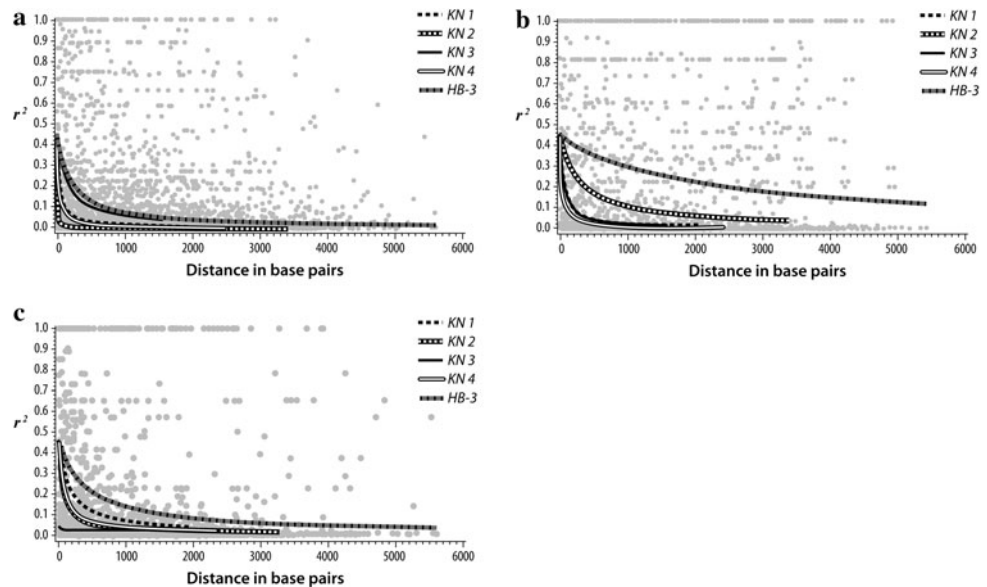
^e The total number of indels with the size of indels indicated in *parentheses* and indels corresponding to singletons in *italics*

* *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001 corresponding to the *P* value for obtaining a number of SNP singletons (under the neutral coalescent process) equal or higher than that observed

Bielawski 2000; Barrier et al. 2003; Palmé et al. 2008). Apparently, *HB-3* has a higher level of conservation and undergoes stronger functional constraints when compared

to *knox-1* genes. Constraints against deleterious amino acid changes are consistent with the fundamental roles of class III *HD-Zip* genes in meristem initiation, organ polarity, and

Fig. 1 Linkage disequilibrium (LD) decay over distance measured by the squared correlations of allele frequencies (r^2) between polymorphic sites plotted against the distance in base pairs in the five regulatory genes *KN1* to *KN4* and *HB-3* of: **a** *P. glauca*, **b** *P. mariana*, and **c** *P. abies*. In *P. glauca*, LD decay of *KN1* overlaps to a great extent with that of *KN4*, while in *P. mariana* it overlaps with that of *KN3*



vascular development (Baima et al. 2001; Zhong and Ye 1999, 2004; Emery et al. 2003; Green et al. 2005; Prigge et al. 2005).

Evidence for positive and balancing selection was also found. In *P. glauca*, *KN3* harboured the lowest total nucleotide (π) and haplotype (significant deficit) diversity (H_d) among all genes. It also harboured a significantly negative D value and a strongly negative (although non-significant) H in addition to a relatively high LD (half-decay of $r^2 \sim 150$ bp; Fig. 1b) which could all be interpreted as the signature of positive selection (Fay and Wu 2000). In *P. mariana*, *KN2* harboured a large positive and significant H value and a relatively high LD (half-decay of $r^2 \sim 300$ bp), supporting the hypothesis of a recurrent hitchhiking in this gene. Interestingly, the hypothesis of a positive selection could not clearly be confirmed based on neutrality tests alone for *HB-3* in *P. mariana*. This gene exhibited the most negative H and D values among all genes for the two North American species but these values were not significant. At the same time, it exhibited the highest LD among the genes of *P. mariana* (half-decay of $r^2 \sim 2,000$ bp; Fig. 1b) and among all the genes of all the species. Consequently, we argue that this gene may be under recent positive selection that could not be detected by the two standard neutrality tests, Tajima's D and Fay and Wu's H . In *P. abies*, both *KN2* and *KN4* reflected the marks of a strong natural selection. Although such a hypothesis is not congruent with the relatively low LD in *KN2* (Fig. 1c), the latter had the most negative and significant H value (Table 4) in addition to a significantly negative D value (Table 4) and a relatively high π_a/π_s ratio (0.16). In a recent study that compared nonsynonymous to synonymous substitution ratios in conifers, Palmé et al.

(2008) suggested that genes with a nonsynonymous to synonymous substitutions ratio (K_a/K_s) lower than 1 but in the range of 0.20–0.52 could be undergoing some effects of positive selection. This range is marginally higher than the observed *KN2* π_a/π_s ratio for *P. mariana* (0.16), suggesting that this gene is a possible candidate for positive selection. As for *KN4*, it harboured the highest π_a/π_s ratio (1.45) among all genes and species (Table 3). Such a high ratio (>1) is often interpreted as an evidence for positive or balancing selection (Wu et al. 1998; Yang and Bielawski 2000; Barrier et al. 2003; Palmé et al. 2008). In the present case, it is more likely that balancing selection is implicated, because the gene had the highest nonsynonymous diversity among all genes and species, harboured the highest haplotype number and diversity among *P. abies* genes, and displayed a low nonsignificant Tajima's D (Table 4) indicative of little or no excess of rare alleles. Despite its relatively low nucleotide diversity (Table 3), its significant deficit in haplotype number and diversity and its significant negative Tajima's D (Table 4), the hypothesis of positive selection could not be confirmed in *KN3* of *P. abies*; this gene showed no significant decay of LD (Fig. 1c) and its Fay and Wu's H was the highest among the genes of *P. abies*.

While showing that different types of selection can affect different genes and different taxa within the same plant genus, our results showed that it is possible to detect selection effects in genes of boreal tree taxa even if their genomes have been affected by recent large demographic effects. These findings suggest that there is no simple rule to identify genes under selection or to transfer information between species, even congeneric ones with similar ecology, life history, and boreal distribution. One possible

Table 4 Recombination, haplotype structure, and neutrality tests based on haplotype frequency distributions

	R_M per informative site ^a	ρ/bp^b	Haplotype number (K)	Haplotype diversity (Hd)	Tajima's D	Fay and Wu's H
<i>P. glauca</i>						
<i>KN1</i>	0.18	0.0049	63†	0.95	-1.05	-4.19
<i>KN2</i>	0.20	0.0034	123†	0.99†	-0.69	-2.49
<i>KN3</i>	0.11	0.0228	53	0.93**	-1.15*	-4.70
<i>KN4</i>	0.17	0.0046	103†	0.97	-0.70	-1.81
<i>HB-3</i>	0.12	0.0017	47	0.92	-0.83	-2.80
<i>P. mariana</i>						
<i>KN1</i>	0.09	0.0064	30†	0.97	-1.18	-4.83
<i>KN2</i>	0.10	0.0034	33†	0.97	-0.63	1.43
<i>KN3</i>	0.13	0.0252	27	0.96	-0.44	-0.97
<i>KN4</i>	0.18	0.0386	43	0.99	-0.95**	-3.41
<i>HB-3</i>	0.08	0.0005	22	0.78***	-1.32	-4.84
<i>P. abies</i>						
<i>KN1</i>	0.15	0.0266	25**	0.90***	-1.64***	-6.28
<i>KN2</i>	0.14	0.0169	36	0.99	-1.14**	-8.50*
<i>KN3</i>	0.04	0.0467	22***	0.79***	-2.28***	-3.13
<i>KN4</i>	0.16	0.0251	37	0.99	-0.40	-5.96
<i>HB-3</i>	0.10	0.0024	31	0.98	-2.08***	-6.85

^a R_M is the minimum number of recombination events (Hudson and Kaplan 1985) divided by the number of informative sites

^b ρ is the maximum-likelihood estimator of the recombination parameter based on independent linked pairs of sites (Hudson 2001)

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ that observed values are lower than expected based on 1,000 coalescent simulations using the DnaSP program

† $P < 0.05$ that observed values are higher than expected based on 1,000 coalescent simulations using the DnaSP program

Table 5 Mean values of Tajima's D and Fay and Wu's H for the three *Picea* species observed in the analysed data and expected under a standard neutral model (SNM) and a bottleneck model followed by population expansion (BottExp)

	<i>P. glauca</i>		<i>P. mariana</i>		<i>P. abies</i>	
	Mean Tajima's D	Mean Fay and Wu's H	Mean Tajima's D	Mean Fay and Wu's H	Mean Tajima's D	Mean Fay and Wu's H
Observed	-0.88	-3.20	-0.90	-2.52	-1.49	-6.14
SNM	-0.02 ($P = 0$)	0.01 ($P = 0.02$)	-0.02 ($P = 0.001$)	0.04 ($P = 0.02$)	-0.01 ($P = 0$)	0.04 ($P = 0$)
BottExp ^a	-0.92 ($P = 0.55$)	-6.89 ($P = 0.44$)	-0.76 ($P = 0.52$)	-5.50 ($P = 0.50$)	-0.75 ($P = 0.23$)	-5.25 ($P = 0.33$)

A model is considered to fit empirical data if the P values of all the parameters considered (mean and variances of D and H) are higher than 0.05 (no significant difference)

^a The bottleneck model that best fit the observed data for each of the three species occurred during the LGM around 25,000 years ago and ended about 17,000 years ago when populations started expanding. Other near-best models suggested that the expansion phase might have been delayed to the Holocene for *P. glauca* and *P. abies* (see "Results")

explanation for this may be related to the evolution of genes in conifers. In plant gene families such as *knox-I*, sub-functionalization was shown to occur more commonly than neo-functionalization following gene duplication, at least for genes recently duplicated during the evolution of conifers (Guillet-Claude et al. 2004). Recent duplications in regulatory genes were also noticed in the *MYB* gene family of conifers with differential expression of family

members (Bedon et al. 2007), and in the *HD-Zip* family in barley (Sakuma et al. 2009). Both gene duplication and function redundancy offer to gene duplicates the possibility to adapt and evolve differently in different species (Van de Peer et al. 2001). Another possible explanation may be related to the nature of the genes selected in this study. On one hand, a priori knowledge about the function and duplication history of the regulatory genes surveyed likely

helped to recover more frequently potential cases of selection at the molecular level. On the other hand, determining complete or nearly complete gene sequences instead of partial sequences may have constituted a decisive factor in detecting selection in this study by improving the coverage and the detection power in the context of boreal taxa affected by recent and large-scale demographic changes.

Linkage Disequilibrium

Given the length of the sequence determined for each gene in each species (up to 5,886 bp), an interesting aspect of this study is that it provided a clear estimate of LD on a gene-by-gene basis and evidence that an important heterogeneity in LD may exist among genes and among congeneric species for the same gene. While valuable, previous estimates of LD in conifers were most often calculated by pooling data from partial gene sequences to obtain a statistically acceptable number of polymorphic sites for the calculation of the correlations between the number of polymorphic sites and the distance in base pairs (e.g. Brown et al. 2004; González-Martínez et al. 2006; Heuertz et al. 2006; Pyhäjärvi et al. 2007). The calculation of LD for individual genes was possible in this study because of the long gene sequences analysed. LD half-decay for individual genes ranged from a few base pairs to around 2,000 thus confirming previous reports of low LD in coniferous genes by pooling data from several genes such as *Pinus taeda* (~800 bp; González-Martínez et al. 2006), Douglas fir (~500 bp; Krutovsky and Neale 2005), *Picea abies* (~100 bp, Rafalski and Morgante 2004; Heuertz et al. 2006), *Pinus taeda* (~2,000 bp; Brown et al. 2004), and *Pinus sylvestris* in northern populations (~1,400 bp; Pyhäjärvi et al. 2007). However, differences in LD among genes and between the three congeneric species for the same gene were notable, with differences sometimes reaching two orders of magnitude for the same gene.

These findings suggest that the generalization of low LD in conifer genes should be taken with caution: some genes may exhibit different patterns depending on different types of selection regime. For instance, *HB-3* of *P. mariana* had a high LD (half-decrease of ~2,000 bp; Fig. 1b) potentially indicative of positive selection and hitchhiking, though *D* and *H* test values were not significant. However, the noted higher LD for *HB-3* of *P. mariana* was correlated with lower values of haplotype diversity and number of haplotypes. Apparently, neutrality tests may not always capture the signature of positive selection, and other indicators such as the extent of LD and haplotype diversity of individual genes may be more informative. In a study that used simulations to assess the power of different tests to detect

selective sweeps, Depaulis et al. (2005) reported that haplotype diversity tests may be more powerful than frequency spectrum statistics such as Tajima's *D* when a selective sweep is mild, i.e. that recombination occurs between the selected site and the marker. Furthermore, our findings suggest that caution should be exercised in extrapolating LD at the multilocus level, because a multilocus estimate may hide the effect of selective forces one or few genes may experience. Interestingly, the *HB-3* gene nonlinear regression model was able to significantly explain ($P < 0.05$) the LD pattern at the multilocus level in *P. glauca* and *P. abies*, but not in *P. mariana* ($P = 0.37$) when the *HB-3* gene was included. The limited number of genes analysed could also be a plausible source for this bias, but the large size of the sequences analysed may have counterbalanced this effect, especially that a good fit of the regression model was observed for *P. glauca* and *P. abies* even when *HB-3* was included.

Correlated Demography

The negative *D* and *H* values obtained for all genes and the congruent patterns of polymorphism (nucleotide diversity, *D* and *H*) in noncoding gene regions support the hypothesis of demographic changes shaping DNA variation in the three *Picea* taxa analysed. Moreover, multilocus analysis indicated a significant excess of both rare- and high-frequency-derived variants in the three species, which is also considered as the signature of a bottleneck followed by population expansion (e.g. Haddrill et al. 2005; Heuertz et al. 2006; Pyhäjärvi et al. 2007). Hitchhiking may induce an excess of rare alleles in some genes, but it is unlikely that such a pattern would be observed for all genes analysed. Although our conclusions are only based on five genes, they do agree with other studies that detected the signature of bottlenecks followed by expansion in *Picea abies* gene sequences (Heuertz et al. 2006; Pyhäjärvi et al. 2007). In addition, this study brings relatively recent estimates for the bottleneck at/or around LGM and expansion events likely at different times after LGM and during the Holocene. Paleocological evidence brings further support to our findings as described below.

The pollen and macrofossil records indicate that white spruce and black spruce from North America survived in the Beringian glacial refugium (Anderson and Lozhkin, 2001; Brubaker et al. 2005) and in refugia south of the ice sheet, for instance in the Mississippi Valley (Davis and Shaw 2001), about 25,000 years ago during the LGM when the Laurentide and Cordilleran ice sheets were still covering most of North America. Following abrupt warming and the initiation of glacial retreat, black spruce and white spruce expanded more or less gradually from southern refugia to occupy their current range in North America

(Jaramillo-Correa et al. 2004; Anderson et al. 2006), but it was not until after the late glacial (approximately 17,000 years ago) that important concentrations of *Picea* pollen could be found in North America (Fig. 1a; Davis and Shaw 2001). These paleoecological data corroborate our coalescent simulations, especially the date of the bottleneck (around 25,000 years ago) and the beginning of expansion in the two North American species at the end of the LGM about 17,000 years ago. The observation of a longer bottleneck impact (up to 10,000 years before present) in some near-best models for *P. glauca* could have resulted from recolonization delays linked to differences in reproduction and adaptation between *P. glauca* and *P. mariana* in response to varying climatic conditions during the Holocene (Lindbladh et al. 2003, 2007; Pisaric et al. 2003). The competitive advantage of *P. mariana* in colder climates (Messaoud et al. 2007) and the absence of competition may have enabled *P. mariana* to expand rapidly at the end of the LGM, likely faster than *P. glauca* whose expansion could have largely occurred with the relatively warmer phases around 10,000 years ago (Pisaric et al. 2003).

Similar to *P. glauca* and *P. mariana*, our results indicate that *P. abies* would have expanded from putative refugia maintained through a late Pleistocene bottleneck. However, the relatively stronger (more negative) mean H and D values observed for this taxon (Table 5) suggested more severe bottleneck effects and/or a more recent expansion than those observed in the two North American species. Based on pollen and fossil records, Ravazzi (2002) estimated that *P. abies* populations in Central Europe underwent a phase of general reduction in population size during the LGM (~25,000–18,000 years before present). He also demonstrated that the species underwent another more severe decline during the Dry maxima that coincided with a phase of severe aridity between the end of the LGM and the onset of the late glacial interstadial (17,500–15,000 years before present). With the subsequent warmer phases, the species experienced an expansion that extended well into the Holocene (Ravazzi 2002; Latalowa and van der Knaap 2006). These paleobiogeographic observations fit well some of our near-best simulation results where the population expansion of *P. abies* could have been delayed well after the beginning of the Holocene, at least for Central Europe. Our findings are also in line with the observations of Lagercrantz and Ryman (1990) that: (1) Central European populations of *P. abies* exhibit a low level of polymorphism and an excess of rare alleles, distinct from those observed in other European countries; and (2) the genetic patterns in Central European populations reflect a bottleneck and a rapid expansion about 10,000 years ago with an ongoing process of population adaptation and differentiation.

More ancient bottlenecks of hundreds of thousands or million years old were inferred in *P. abies* by Heuertz et al. (2006) and Chen et al. (2010). However, their model assumptions and parameters fit criteria were slightly different and their populations were mostly from the Baltico-Nordic and Alpine refugia, contrary to the present *P. abies* populations which descended most likely from the more eastern and continental Carpathian refugium (Collignon et al. 2002). Moreover, in an extensive study of the pollen and fossil records in Central and Northern Europe, Latalowa and van der Knaap (2006) showed remarkable regional differences in the Holocene spruce expansion patterns related to a wide spectrum of interactions between environmental factors and spruce ecology. As a result, it is possible that these differences in demographic history simply reflect variation in genetic diversity or adaptive potential among glacial populations at the onset of LGM and/or spatio-temporal heterogeneity in the dynamics of the last glaciation and ensuing Holocene colonization. Additional sequence signature studies in different ancestral lineages for these species and for accompanying taxa in the same refugia would help confirm if geographically and genetically distinct glacial populations experienced differential historic demography, as is likely the case for *P. abies*.

Acknowledgments We are grateful to F. Larochelle (Centre for Bioinformatics and Computational Biology, Univ. Laval) for helping with the design of an application for the graphical representation of DNA sequences and their polymorphisms (Supplementary Figs. 1 and 2), P.-L. Poulin, S. Larose et J. Laroche for their help in setting softwares and running the coalescent simulations, F. Gagnon and S. Blais for their sustained contribution to the PCR experiments and sequencing, N. Pavy and H. Bérubé for assistance with programming, and J. P. Jaramillo-Correa for useful paleobotanical references and discussions. We are also grateful to Peter Andolfatto for his helpful instructions about neutrality multilocus analysis, and Julio Rozas for his helpful insights about some of the DnaSP procedures. This work was supported by grants from Genome Québec, Genome Canada, and the National Sciences and Engineering Research Council of Canada.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Acheré V, Favre JM, Besnard G, Jeandroz S (2005) Genomic organisation of molecular differentiation in Norway spruce (*Picea abies*). Mol Ecol 14:3191–3201
- Anderson PM, Lozhkin VA (2001) The Stage 3 interstadial complex (Karginskii/middle Wisconsinan interval) of Beringia: variations in paleoenvironments and implications for paleoclimatic interpretations. Quat Sci Rev 20:93–125

- Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proc Natl Acad Sci USA* 103:12447–12450
- Baima S, Possenti M, Matteucci A, Wisman E, Altamura MM, Ruberti I, Morelli G (2001) The *Arabidopsis* ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. *Plant Physiol* 126:643–655
- Barrier M, Bustamante CD, Yu J, Purugganan MD (2003) Selection on rapidly evolving proteins in the *Arabidopsis* genome. *Genetics* 163:723–733
- Beaulieu J (1996) Breeding program and strategy for white spruce in Québec Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Québec, QC, Canada, Information Report LAU-X-117E
- Beaulieu J, Corriveau A, Daoust G (1989) Phenotypic stability and delineation of black spruce breeding zones in Québec. Canadian Forestry Service, Laurentian Forest Research Centre, Québec, Canada, Information Report LAU-X-85E
- Bedon F, Grima-Pettenati J, Mackay J (2007) Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). *BMC Plant Biol* 7:17
- Bouillé M, Bousquet J (2005) Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. *Am J Bot* 92:63–73
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc Natl Acad Sci USA* 101:15255–15260
- Brubaker LB, Anderson PM, Edwards ME, Lozhkin AV (2005) Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *J Biogeogr* 32:833–848
- Charlesworth J, Eyre-Walker A (2008) The McDonald–Kreitman test and slightly deleterious mutations. *Mol Biol Evol* 25:1007–1101
- Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303
- Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic structure on neutral variation. *Annu Rev Ecol Syst* 34:99–125
- Chen J, Källman Gyllenstrand N, Lascoux M (2010) New insights on the speciation history and nucleotide diversity of three boreal spruce species and a tertiary relict. *Heredity* 104:3–14
- Collignon A-M, Van de Sype H, Favre J-M (2002) Geographical variation in random amplified polymorphic DNA and quantitative traits in Norway spruce. *Can J For Res* 32:266–282
- Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Math Biosci* 205:19–31
- Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinf* 9:539
- Corriveau A, Beaulieu J, Daoust G (1988) Phenotypic stability and productivity of central European Norway spruce provenances in Québec, Canada. Proceedings meeting of the IUFRO working party S-202-11, Tjörnarp, Sweden
- Davis M, Shaw R (2001) Range shifts and adaptive responses to quaternary climate change. *Science* 292:673–679
- Depaulis F, Veuille M (1998) Neutrality tests based on the distribution of haplotypes under an infinite-site model. *Mol Biol Evol* 15:1788–1790
- Depaulis F, Mousset S, Veuille M (2001) Haplotype tests using coalescent simulations conditional on the number of segregating sites. *Mol Biol Evol* 18:1136–1138
- Depaulis F, Mousset S, Veuille M (2005) Detecting selective sweeps with haplotype tests. In: Nurminsky D (ed) Hitchhiking and haplotype tests. Springer, New York, pp 34–54
- Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker N, Izhaki A, Baum S, Bowman J (2003) Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. *Curr Biol* 13:1768–1774
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925
- Gamache I, Jaramillo-Correa JP, Payette S, Bousquet J (2003) Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: imprint of a founder effect associated with postglacial colonization. *Mol Ecol* 12:891–901
- García-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Mol Ecol* 12:1195–1206
- Goncharenko GG, Zadeika IV, Birgelis JJ (1994) Genetic structure diversity and differentiation of Norway spruce (*Picea abies* (L.) Karst) in natural populations of Latvia. *Forest Ecol Manag* 72:31–38
- González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172:1–12
- Green KA, Prigge MJ, Katzman RB, Clark SE (2005) CORONA, a member of the Class III Homeodomain Leucine Zipper gene family in *Arabidopsis*, regulates stem cell specification and organogenesis. *Plant Cell* 17:691–704
- Guillet-Claude C, Isabel N, Pelgas B, Bousquet J (2004) The evolutionary implications of *knox-I* gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence. *Mol Biol Evol* 21:2232–2245
- Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005) Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res* 15:790–799
- Hake S, Smith HM, Holtan H, Magnani E, Mele G, Ramirez J (2004) The role of *knox* genes in plant development. *Annu Rev Cell Dev Bi* 20:125–151
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Heuertz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway Spruce [*Picea abies* (L.) Karst]. *Genetics* 174:2095–2105
- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* 38:226–231
- Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibrium in finite populations. *Theor Popul Biol* 33:54–78
- Hudson RR (2001) Two-locus sampling distributions and their application. *Genetics* 159:1805–1817
- Hudson RR (2002) Generating samples under a Wright-Fisher neutral of genetic variation. *Bioinformatics* 18:337–338
- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159
- Isabel N, Beaulieu J, Bousquet J (1995) Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. *Proc Natl Acad Sci USA* 92:6369–6373

- Ito M, Sato Y, Matsuoka M (2002) Involvement of homeobox genes in early body plan of monocot. *Int Rev Cytol* 218:1–35
- Jaramillo-Correa JP, Beaulieu J, Bousquet J (2001) Contrasting evolutionary forces driving population structure at expressed sequence tag polymorphisms, allozymes and quantitative traits in white spruce. *Mol Ecol* 10:2729–2740
- Jaramillo-Correa JP, Beaulieu J, Bousquet J (2004) Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Mol Ecol* 13:2735–2747
- Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness and wood quality-related candidate genes in Douglas fir. *Genetics* 171:2029–2041
- Lagercrantz U, Ryman N (1990) Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution* 44:38–53
- Latalowa M, van der Knaap WO (2006) Late Quaternary expansion of Norway spruce *Picea abies* (L) Karst in Europe according to pollen data. *Quat Sci Rev* 25:2780–2805
- Lindbladh M, Jacobson GL, Schaffner M (2003) The postglacial history of three *Picea* species in New England, USA. *Quat Res* 59:61–69
- Lindbladh M, Oswald WW, Foster DR, Faison EK, Hou J, Huang Y (2007) A late-glacial transition from *Picea glauca* to *Picea mariana* in southern New England. *Quat Res* 3:502–508
- Luikart G, England PR, Tallmon D, Jordan S, Tabarlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 4:981–994
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654
- McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160:1231–1241
- Messaoud Y, Bergeron Y, Asselin H (2007) Reproductive potential of balsam fir (*Abies balsamea*), white spruce (*Picea glauca*), and black spruce (*P. mariana*) at the ecotone between mixedwood and coniferous forests in the boreal zone of western Quebec. *Am J Bot* 94:746–754
- Murray BG (1998) Nuclear DNA amounts in gymnosperms. *Ann Bot* 82:3–15
- Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol Ecol* 17:3599–3613
- Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends Plant Sci* 9:325–330
- Nielsen R (2005) Molecular signatures of natural selection. *Annu Rev Genet* 39:197–218
- Palmé AE, Wright M, Savolainen O (2008) Patterns of divergence among conifer ESTs and polymorphism in *Pinus sylvestris* identify putative selective sweeps. *Mol Biol Evol* 25:2567–2577
- Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, Lamothe M, Isabel N, Bousquet J (2008) Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* 9:21
- Perry DJ, Bousquet J (2001) Genetic diversity and mating system of post-fire and post-harvest black spruce: an investigation using codominant sequence-tagged-site (STS) markers. *Can J For Res* 31:32–40
- Pisaric M, Holt C, Szeicz JM, Karst T, Smol JP (2003) Holocene treeline dynamics in the mountains of northeastern British Columbia, Canada, inferred from fossil pollen and stomata. *Holocene* 13:161–173
- Pot D, McMillan L, Echt C, Le Provost G, Garnier-Géré P, Cato S, Plomion C (2005) Nucleotide variation in genes involved in wood formation in two pine species. *New Phytol* 167:101–112
- Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE (2005) Class III Homeodomain-Leucine Zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *Plant Cell* 17:61–76
- Przeworski M (2002) The signature of positive selection at randomly chosen loci. *Genetics* 160:1179–1189
- Pyhäjärvi T, Garcia-Gil MR, Knutt T, Mikkonen M, Wachowiak W, Savolainen O (2007) Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* 177:1713–1724
- Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet* 20:103–111
- Ran J-H, Wei X-X, Wang X-Q (2006) Molecular phylogeny and biogeography of *Picea* (Pinaceae): Implications for phylogeographical studies using cytoplasmic haplotypes. *Mol Phylogenet Evol* 41:405–419
- Ravazzi C (2002) Late Quaternary history of spruce in southern Europe. *Rev Palaeobot Palyno* 120:131–177
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sakuma S, Pourkheirandish M, Matsumoto T, Koba T, Komatsuda T (2009) Duplication of a well-conserved homeodomain-leucine zipper transcription factor gene in barley generates a copy with more specific functions. *Funct Integr Genomics*. doi: 10.1007/s10142-009-0134-y
- Sato Y, Fukuda Y, Hirano HY (2001) Mutations that cause amino acid substitutions at the invariant positions in homeodomain of OSH3 KNOX protein suggest artificial selection during rice domestication. *Genes Genet Syst* 76:381–392
- Sessa G, Steindler C, Morelli G, Ruberti I (1998) The *Arabidopsis Athb-8, -9* and *-14* genes are members of a small gene family coding for highly related HD-ZIP proteins. *Plant Mol Biol* 38:609–622
- Siljak-Yakovlev S, Cerbah M, Coulaud J, Stoian V, Brown SC, Zoldos V, Jelenic S, Papes D (2002) Nuclear DNA content, base composition, heterochromatin and rDNA in *Picea omorika* and *Picea abies*. *Theor Appl Genet* 104:505–512
- Sokal RR, Rohlf FJ (1995) *Biometry*, 3rd edn. WH Freeman, New York
- Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100:158–170
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14:671–688
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Van de Peer Y, Taylor JS, Braasch I, Meyer A (2001) The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *J Mol Evol* 53:436–446

- Vollbrecht E, Veit B, Sinha N, Hake S (1991) The developmental gene *Knotted-1* is a member of a maize homeobox gene family. *Nature* 350:241–243
- Wakamiya I, Newton R, Johnston JS, Price HJ (1993) Genome size and environmental factors in the genus *Pinus*. *Am J Bot* 80:1235–1241
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276
- Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol* 22:506–519
- Wu J, Saupé SJ, Glass NL (1998) Evidence for balancing selection operating at the *het-c* heterokaryon incompatibility locus in a group of filamentous fungi. *Proc Natl Acad Sci USA* 95:12398–12403
- Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15:496–503
- Zhong R, Ye ZH (1999) IFL1, a gene regulating interfascicular fiber differentiation in *Arabidopsis*, encodes a homeodomain-leucine zipper protein. *Plant Cell* 11:2139–2152
- Zhong R, Ye ZH (2004) Amphivasal vascular bundle 1, a gain-of-function mutation of the IFL1/REV gene, is associated with alterations in the polarity of leaves, stems and carpels. *Plant Cell Physiol* 45:369–385