

Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa*

ERIN J. GILCHRIST,* GEORGE W. HAUGHN,* CHENG C. YING,† SARAH P. OTTO,‡ JUN ZHUANG,§ DOROTHY CHEUNG,¶ BJÖRN HAMBERGER,* FARIBA ABOUTORABI,* TATYANA KALYNYAK,¶ LEE JOHNSON,* JOERG BOHLMANN,§ BRIAN E. ELLIS,§ CARL J. DOUGLAS* and QUENTIN C. B. CRONK¶

*Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4, †British Columbia Ministry of Forests, Research Branch, Victoria, BC, Canada V8W 1N1, ‡Department of Zoology, University of British Columbia, Vancouver, BC, Canada V6T 1Z4, §Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada V6T 1Z4, ¶Centre for Plant Research and Botanical Garden, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Abstract

Ecotilling was used as a simple nucleotide polymorphism (SNP) discovery tool to examine DNA variation in natural populations of the western black cottonwood, *Populus trichocarpa*, and was found to be more efficient than sequencing for large-scale studies of genetic variation in this tree. A publicly available, live reference collection of *P. trichocarpa* from the University of British Columbia Botanical Garden was used in this study to survey variation in nine different genes among individuals from 41 different populations. A large amount of genetic variation was detected, but the level of variation appears to be less than in the related species, *Populus tremula*, based on reported statistics for that tree. Genes examined varied considerably in their level of variation, from *PoptrTB1* which had a single SNP, to *PoptrLFY* which had more than 23 in the 1000-bp region examined. Overall nucleotide diversity, measured as π_{Total} was relatively low at 0.00184. Linkage disequilibrium, on the other hand, was higher than reported for some woody plant species, with mean r^2 equal to 0.34. This study reveals the potential of Ecotilling as a rapid genotype discovery method to explore and utilize the large pool of genetic variation in tree species.

Keywords: linkage disequilibrium, natural variation, nucleotide diversity, *Populus trichocarpa*, TILLING, tree genomics

Received 21 June 2005; revision accepted 23 September 2005

Introduction

The first tree for which a genome sequence is available is the western black cottonwood, *Populus trichocarpa* (G. Tuskan *et al.* in preparation; <http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>). *P. trichocarpa* is the largest deciduous tree indigenous to western North America and its contribution to the ecology and general health of this biome is significant. Over large parts of the Northern Hemisphere, poplars are vital to carbon sequestration and watershed protection in addition to providing habitat for numerous species of animals and birds. *P. trichocarpa* and its hybrids are also important plantation species for pulp and paper industries, are a renewable energy resource,

and have been used for phytoremediation and as indicators of environmental damage (reviewed in Bradshaw *et al.* 2000).

Populus is emerging as a model for woody plant biology (Strauss & Martin 2004), but in spite of its economic and environmental importance, little is known about genetic variation within this genus. *P. trichocarpa*, the species whose genome has been sequenced, has a natural range that spans from Alaska to southern California, and from the Pacific Coast into interior mountain ranges. As a first step towards analysing genetic variation in this species, a live reference collection of *P. trichocarpa* has been established at the University of British Columbia (UBC). This collection includes trees from more than 140 different populations (Fig. 1) that exhibit great phenotypic variation even when grown in a common garden (Gilchrist & Cronk, unpublished).

The ability of individuals within a species to adapt to different environments resides in their genetic diversity.

Correspondence: Carl Douglas, Fax: 604-822-6089; E-mail: cdouglas@interchange.ubc.ca



Fig. 1 Geographic distribution of trees from the *Populus trichocarpa* reference collection established at the UBC Botanical Garden. Populations used in this study are shown as blue circles. The origin of the reference tree, Nisqually 1, is highlighted in red. Map was drawn using MAPPAD 2.0 (NOAA Palaeoclimatology Program).

This diversity, most commonly manifested as simple nucleotide polymorphisms (SNPs), can provide clues to the adaptive strategies and population histories that have played roles in species' evolution. In addition, natural DNA polymorphisms form the basis for the intraspecific variation that is of great relevance to breeders. Even in model plant systems such as *Arabidopsis thaliana* that are amenable to both forward and reverse genetics, the study of natural variation has allowed the elucidation of characteristics that

would not have been identified using only induced genetic mutations (Gazzani *et al.* 2003). A resource database of SNPs in *P. trichocarpa* would provide researchers and breeders with a tool for answering questions concerning population structure or adaptation and allow comparison of this species with others with which it has been hybridized for commercial purposes.

A number of different techniques for identifying SNPs have been developed but all have their limitations. Assays

based on gel mobility, like denaturing gradient gel electrophoresis (DGGE) and single strand conformation polymorphism (SSCP), do not pinpoint the location or the type of polymorphism present in the DNA fragment (DeFrancesco & Perkel 2001). Techniques that rely on denaturation kinetics and quantitative polymerase chain reaction (PCR) only work for small fragments of DNA (Gundry *et al.* 2003), and array hybridization techniques are efficient only in discovering approximately 50% of SNPs (Borevitz *et al.* 2003). Sequencing of candidate genes from multiple genotypes is a more accurate alternative to these methods, but is relatively expensive when applied to multiple loci in large numbers of individuals. TILLING (targeting induced local lesions in genomes) is a technique for detecting DNA polymorphisms using a mismatch-specific endonuclease (reviewed in Gilchrist & Haughn 2005) that has been used for reverse genetics (Till *et al.* 2003) and also for looking at natural variation in *A. thaliana* (Comai *et al.* 2004). This latter use of the technique is termed Ecotilling.

We have used Ecotilling for the first time as an SNP discovery tool in a species that is long-lived, dioecious and genetically heterogeneous. SNP variation was examined at nine different loci in individuals from 41 different populations distributed throughout most of the *P. trichocarpa* range (Fig. 1). Variation was analysed both within a single tree (heterozygosity) as well as between individual trees and a reference, *P. trichocarpa* 383–2499 (Nisqually 1), whose genome has been sequenced. The availability of a sequenced genome made it possible to direct our attention to candidate genes of interest, providing an unprecedented view of genetic variation at multiple loci in this species. This pilot study shows that the level of nucleotide diversity in *P. trichocarpa* makes it theoretically possible to examine regions of 1000–1500 base pairs (bp) in 96 individuals or more on a single gel. All SNPs discovered by Ecotilling have been positioned on the *P. trichocarpa* physical genomic map, and it will be straightforward to integrate these DNA markers into available genetic maps. Data also indicate that linkage disequilibrium (LD) is higher in this species than in some

other woody plants, implying that SNP information can be more efficiently used for association mapping. In addition, we have shown that Ecotilling can be used to identify species-specific SNPs which are useful for genotyping or hybrid identification.

Materials and methods

Collection of samples

Trees used in this project came from a collection of *Populus trichocarpa* provenances at the Botanical Garden of UBC drawn from a sample set of trees initially established by C. Ying, under the auspices of the British Columbia Forest Service (Research Branch) in Surrey, British Columbia, Canada. These samples represent populations from northern British Columbia, in Canada, to southern Oregon, in the USA, and are freely available to other researchers upon request. Nisqually 1 cuttings were obtained from T. Bradshaw (University of Washington, Seattle, Washington), and were also planted at the UBC Botanical Garden. One tree was selected from each of the populations shown in blue in Fig. 1. Accession numbers and geographic locations are provided as supplementary material (Table S1).

Candidate gene selection and primer design

Candidate genes, listed in Table 1, were selected either based on data from evolutionary studies in other organisms, or on their relevance to forestry concerns such as wood quality or pathogen defence. *PoptrTB1*, *PoptrLFY*, *PoptrMP* and *PoptrKNAT1* encode transcription factors, and their putative homologues in other species are integral to morphogenesis (*TB1*, Cubas *et al.* 1999; *LFY*, Rottmann *et al.* 2000), vascular differentiation (*MP*, Hardtke *et al.* 2004) or lignin deposition (*PoptrKNAT1*, Rogers & Campbell 2004). The *PoptrF5H* gene is a P450 monooxygenase whose homologues have been shown to affect lignin composition in both tobacco and poplar (Franke *et al.* 2000), and *PoptrCHI* and *Poptr4CL3*

Table 1 List of genes used for Ecotilling

Gene name	<i>Populus trichocarpa</i> linkage group*	Function
<i>PoptrTB1</i> (Teosinte branched-like-1)	VIII (rev. 7349961)	TCP-domain transcription factor
<i>PoptrLFY</i> (Leafy)	XV (rev. 7933490)	Meristem identity transcription factor
<i>PoptrMP</i> (Monopteros)	II (fwd. 1544569)	Auxin response factor
<i>PoptrF5H</i> (Ferulate 5-hydroxylase)	VII (fwd. 11484731)	Cytochrome P450
<i>PoptrCHI</i> (Chalcone flavanone isomerase)	X (fwd. 18469186)	Flavonoid biosynthesis enzyme
<i>Poptr4CL3</i> (4-Coumarate:CoA ligase 3)	XIX (fwd. 4083762)	Flavonoid biosynthesis enzyme
<i>PoptrNPR1</i> (Non-expresser of PR genes)	VI (fwd. 8443884)	Signalling protein
<i>PoptrMPK4</i> (MAP kinase-4)	II (rev. 12430699)	MAP kinase
<i>PoptrKNAT1</i> (Knotted 1-like)	II (fwd. 8527521)	Homeobox transcription factor

*The positions of the genes on linkage groups is based on JGI *Populus trichocarpa* version 1.0 sequence release (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>).

are probable homologues of genes that encode enzymes in the flavonoid biosynthesis pathway (Ehltling *et al.* 1999; Winkel-Shirley 2001) and may affect seed dispersal, plant-microbe interactions, or fertility. The last two genes, *PoptrNPR1* (Pieterse & VanLoon 2004) and *PoptrMPK4* (Petersen *et al.* 2000), are putative homologues of genes that are involved in Systemic Acquired Resistance in response to plant-pathogen interactions.

The sequences for *PoptrTB1* and *PoptrLFY* were obtained from GenBank (Accession nos AF309093 and U93196, respectively). All other sequences were drawn from the unassembled, whole genome, shotgun sequence database of the *P. trichocarpa* Nisqually 1 genome that was available at that time from the Joint Genome Institute (<http://genome.jgi-psf.org/poplar0/poplar0.home.html>). These sequences are provided as supplementary material and, for the purposes of this study, are numbered starting from the 'A' at the beginning of the predicted initiation codon. Gene models were predicted using FGENESH (Salamov & Solovyev 2000) through the Softberry website (<http://www.softberry.com>). Only genes that appeared to be single copy or that were easily distinguishable from other family members at the genomic sequence level were selected for analysis. Primers were designed using the Web-based program CODDLE (www.proweb.org/input) to amplify a product of approximately 1000 bp, and their sequences are listed in the Supplementary material. Sequencing was used to confirm that the primers amplified a single locus.

DNA extraction and Ecotilling

DNA was extracted using the DNeasy Plant DNA purification kit (QIAGEN) according to their protocol from leaf or bud tissue from individual trees that was stored at -70°C .

Polymerase chain reaction (PCR) and Ecotilling were performed as described by Comai *et al.* (2004). Fluorescently tagged primers were purchased from MWG Biotech. Samples were amplified individually and also mixed 1:1 with Nisqually 1 DNA to generate heteroduplexes for identification of homozygous variant alleles. Crude celery juice extract (CJE) was prepared as described by Till *et al.* (2004). Aliquots were stored at -70°C and were spun at approximately 2000 *g* for 1 min before use to remove any tissue debris. PCR products were digested with CJE by adding 20 μL of extract and buffer mix (100 mM MgSO_4 , 100 mM HEPES, 300 mM KCl, 0.02% Triton X-100, 0.002 mg/mL BSA, and 0.2% to 0.3% crude CJE) directly to the PCRs and incubating at 45°C for 15 min. Reactions were stopped by adding 2.5 μL of 0.5 M EDTA. The DNA was purified by passage through G50 Sephadex, and concentrated for 30 min at 90°C before running on a 25-cm LI-COR sequencing gel with a 0.4-mm, 96-well sharktooth comb. Analysis of the gel images was done using ADOBE PHOTOSHOP.

Identification and sequencing of SNPs

DNA amplified from individual trees was only cleaved by CJE if that individual was heterozygous for a particular SNP. Individuals homozygous for variant nucleotides were identified in CJE-treated heteroduplex mixtures generated by co-amplification of Nisqually 1 and target tree DNA. Once a polymorphism had been identified, representative alleles were sequenced to identify the nucleotide change relative to Nisqually 1. Wherever possible, each SNP was sequenced in more than one tree to confirm that there were not more than two alleles segregating at any individual site. At sites where Nisqually 1 was heterozygous, we were not able to determine the genotype of individuals carrying a homozygous variant unless they were sequenced since homozygous sites are not cleaved with CJE and therefore we could not establish which of the two Nisqually 1 alleles was present in these trees. In cases where two sequenced SNPs were found to be a single base pair apart, we were not able to consistently distinguish between these two SNPs on the Ecotilling gels, and these were not analysed further.

Sequence analysis was performed using SEQUENCHER 4.2 (Gene Codes) and the potential effect of the SNPs was predicted using project aligned related sequences and evaluate SNPs (PARSESNP; Taylor & Greene 2003). Once the alleles from representative trees had been sequenced, sequence files were generated for other trees using a Perl script that took the genomic sequence of Nisqually 1 and replaced the nucleotides at any polymorphic sites with either the alternate allele (in the case of homozygous variants), or with an ambiguous nucleotide code representing the two different alleles (in the case of heterozygous variants). Synthetic haplotype sequences were generated by independently assigning variant alleles to one of the two chromosomes in heterozygous individuals for analysis using DNASP 4.0 (Rozas *et al.* 2003) because this program does not recognize the ambiguous nucleotide codes used for diploid sequence analysis.

The number of haplotypes was estimated from diplo-type data using Bayesian methods as implemented by the program PHASE, version 2.1.1 (Stephens & Donnelly 2003).

Analysis of SNPs

Observed heterozygosity for each site (H_{OSite}) was determined from bands on Ecotilling gels. Expected heterozygosity for each site (H_{ESite}) was calculated based on the Hardy-Weinberg frequency of $2(pq)$, where p and q are the observed frequencies of the common and rare alleles, respectively, as determined from Ecotilling gels. It was not possible to calculate exact frequencies of p and q at sites that were heterozygous in our reference tree since we were not able to determine the genotype of homozygous individuals at these sites. In such cases, H_{ESite} was not calculated and

H_{OSite} was used instead. We calculated the expected allele frequency at these sites using the formula $H_{\text{OSite}} = 2p(1 - p)$ using QUICKMATH (www.quickmath.com).

Observed heterozygosity for each gene (H_{OGene}) was the frequency of individuals exhibiting one or more polymorphism in that gene. Nucleotide diversity per gene (π_{Gene}) was calculated by dividing the sum of H_{ESite} for that gene by the amplicon size (number of nucleotides). Silent site diversity (π_{Silent}) was calculated by dividing the sum of all synonymous or noncoding polymorphisms (those that do not affect the sequence of the protein product of the gene) by the number of synonymous and noncoding positions in the gene as calculated using DNASP 4.0 (Rozas *et al.* 2003). Average nucleotide diversity (π_{Total}) was the average of all π_{Gene} values.

Linkage disequilibrium and phylogenetic trees

Linkage disequilibrium was determined using JLIN (Java Linkage Disequilibrium) version 1.11 (K. W. Carter, P. A. McCaskie & L. J. Palmer, www.genepi.waimr.uwa.edu.au/projects/jlin). This program calculates various measures of LD based on unphased genotype data and uses an expectation-maximization (EM) algorithm to predict haplotype frequencies. We calculated r^2 using this program, selecting 1000 iterations for empirical simulation. The r^2 values for allele pairs were then plotted against distance. Both linear and log-linear regressions were performed using JMP version 4.0 statistical analysis software (SAS Institute). These gave similar results and were comparable to the Hill & Weir (1988) fitted curve that is shown in Fig. 5.

Dendrograms were generated using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method implemented in the program PAUP* 4.0 (Swofford 2003), using mean genetic distances and random break-ties. To obtain bootstrap values, 10 000 replicates were used.

Statistical significance

We determined whether or not genes varied from one another significantly in terms of SNP frequency, heterozygosity, nucleotide diversity or average LD by calculating Z scores using the formula [(observed – mean)/standard deviation]. *P* values for these scores were obtained by using the *P* Value Calculator at www.graphpad.com/quickcalcs (GRAPHPAD software). Bonferroni correction was used to determine significant *P* values (SISA online statistical analysis: <http://home.clara.net/sisa/bonfer.htm>).

Results

Experimental design and candidate gene selection

We screened for polymorphisms in nine different genes

from 46 trees representing different populations of western black cottonwood from the province of British Columbia and Washington and Oregon states (Fig. 1, see Supplementary material for list of accessions and their geographic locations). Five of these samples were not included in our analysis because of concerns about labelling or hybrid contamination; thus, the total number of trees included in this study was 41. The candidate genes selected for this project are listed in Table 1 and described in the Materials and methods. They are putative poplar orthologues of transcription factors, enzymes, or signalling intermediates known in other plants, and they represent a range of diverse genes suitable for giving an overall picture of genetic diversity in *Populus trichocarpa*.

For this pilot project, we screened for both heterozygous and homozygous variants to obtain an accurate estimate of nucleotide diversity. This means that DNA from each tree was tested individually and also heteroduplexed with DNA from the Nisqually 1 reference tree. As shown in Fig. 2, heterozygous alleles were detected by a cleavage product in the first lane of a sample set from an individual tree (indicated by white arrows in lane 2_i). Homozygous sites were not cleaved by CJE and therefore SNPs that were homozygous but different from the Nisqually 1 sequence were only detected in the second lane of the sample because the heteroduplexed DNA contained a mismatch at the site of the SNP (indicated by red arrows in lane 3_h of Fig. 2). For each gene surveyed in this study, we detected trees without SNP variation, although for all loci we also always detected trees with variant alleles.

Representative alleles identified by Ecotilling were sequenced in a subset of the trees, and *in silico* sequences were generated for all other individuals, thus greatly reducing the amount of sequencing required for genotype identification.

Analysis of SNPs

The amount of DNA analysed was approximately 358 kilobases (kb). Due to difficulties with individual PCRs, data for specific loci in some of the trees are missing, but our data set includes sequences for all nine loci in 37 trees. The number of polymorphisms identified for each locus is shown in Table 2, and the distribution of SNPs in the targeted region of each candidate gene is shown in Fig. 3. SNPs were detected in all of the nine genes tested although the number of polymorphisms varied from gene to gene with an average frequency in our sample set of 6.78 per gene. The least polymorphic locus, *PoptrTB1*, had a single SNP and the most polymorphic locus, *PoptrLFY*, had more than 23 SNPs.

A total of 63 *P. trichocarpa* SNPs were identified and characterized in 8191 bp of aligned DNA sequence and all but two were single base pair polymorphisms. A complete list

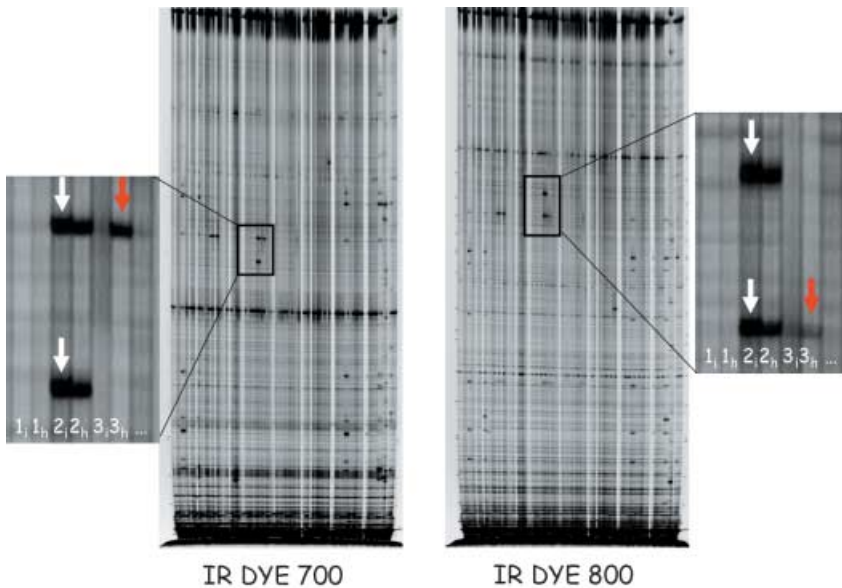


Fig. 2 Images of Ecotilling gels obtained from each of the two fluorescent channels of the LI-COR sequencing machine for the *PoptrMP* gene. The IR DYE 700 channel shows the PCR products that carry the left end primer and the IR DYE 800 channel shows the PCR products that carry the right end primer. Insets depict fragments that have been cleaved by CJE indicating an SNP in the amplicon. Cleavage by CJE is not to saturation, so multiple cleavage bands may be seen in each lane representing several polymorphic sites in one individual. DNA from each tree was run individually (lanes 1_i, 2_i, 3_i...) and heteroduplexed with Nisqually 1 DNA (1_h, 2_h, 3_h...) in adjacent lanes on the gel. White arrows show heterozygous alleles identified by the presence of a cleavage product when self-annealed. Red arrows show a homozygous SNP identified by a cleavage product only when heteroduplexed with Nisqually 1 DNA.

Table 2 Levels of polymorphism in *Populus trichocarpa*

Gene	No. of trees	No. of bp	Coding DNA	SNPs	Haplotypes*	$H_{\text{OGene}}^{\dagger}$	$\pi_{\text{Gene}}^{\ddagger}$	π_{Silent}^{\S}	$d_{\text{N}}/d_{\text{S}}^{\parallel}$
<i>PoptrTB1</i>	41	904	100%	1	2	0.098	0.00009	0.00000	N/A
<i>PoptrLFY</i>	41	906	36%	23**	25	0.878	0.00683**	0.00958**	0.00
<i>PoptrMP</i>	37	984	92%	8	8	0.162	0.00109	0.00236	0.44
<i>PoptrF5H</i>	38	812	100%	2	3	0.474	0.00056	0.00268	0.00
<i>PoptrCHI</i>	40	861	41%	8	10	0.550	0.00284	0.00261	0.60
<i>Poptr4CL3</i>	40	903	98%	4	7	0.775	0.00128	0.00349	0.11
<i>PoptrNPR1</i>	40	947	79%	3	6	0.525	0.00072	0.00142	0.16
<i>PoptrMPK4</i>	38	967	24%	10	10	0.447	0.00245	0.00306	0.00
<i>PoptrKNAT1</i>	37	907	63%	4	6	0.297	0.00072	0.00088	0.30
Total	352	8191		63					
Mean		910	70%	6.78		0.467	0.00184	0.00290	

*Number of haplotypes was estimated using PHASE (Stephens & Donnelly 2003).

†Observed heterozygosity per gene.

‡Nucleotide diversity per gene.

§Silent-site nucleotide diversity.

¶Ratio of nonsynonymous to synonymous changes.

**Statistically significant data.

of polymorphisms is provided as supplementary material. One of the exceptional SNPs was an insertion of 3 bp in the coding sequence of the *PoptrTB1* gene, and the other was a deletion of 6 bp in a predicted intron of the *PoptrCHI* gene. For statistical purposes these indels were treated as single site polymorphisms along with the other SNPs in our analysis. We identified 13 additional SNPs upon DNA sequencing that had not been detected through Ecotilling but, with the exception of those in the *PoptrLFY* gene, these were all near the ends of the amplified regions and could not consistently be resolved at the bottom of the Ecotilling gel image. For our statistical analysis we have excluded

these regions, thus reducing the size of the analysed region by 50 bp from each end. The *PoptrLFY* gene showed an extremely high number of SNPs: more than 30 bands on the Ecotilling gel and 29 SNPs by sequencing. Because the high level of nucleotide diversity in this region made it difficult to ensure that all of the bands on the gel were consistently scored and since our main goal was to develop a method for high-throughput SNP detection and genotyping, we have only analysed the 23 *PoptrLFY* polymorphisms consistently detected by both Ecotilling and sequencing. We did not experience similar difficulties with any of the other eight loci.

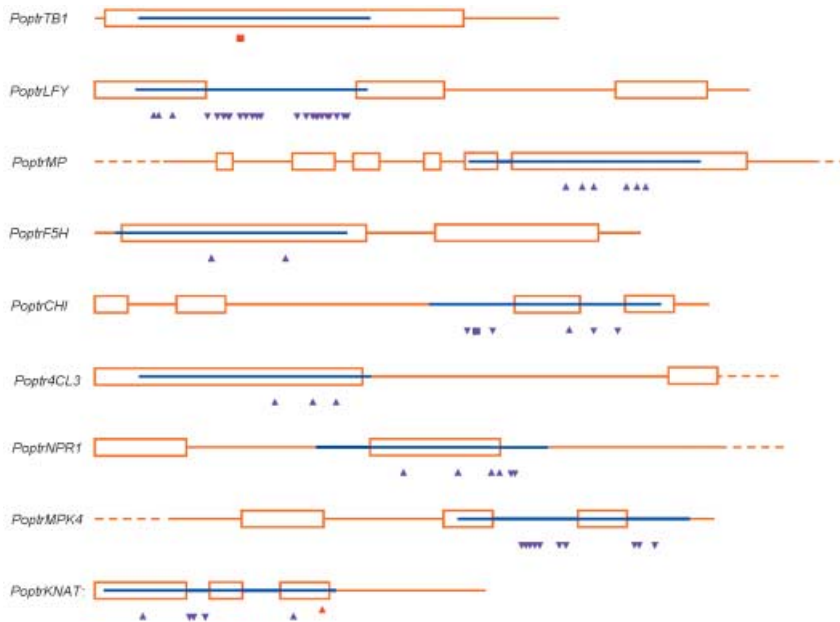


Fig. 3 Distribution of SNPs. The figure was designed from PARSESNP output files (Taylor & Greene 2003). Blue lines indicate the extent of the amplified region which was approximately 1 kb for each gene. Orange open boxes denote exons. Purple up arrows indicate changes in the DNA sequence that do not affect the amino acid sequence of the protein product. Purple down arrows indicate changes to noncoding regions of DNA. Black up arrows indicate changes that induce missense mutations in the predicted protein product. Red arrow indicates a premature stop codon in the coding region. Red or purple squares indicate insertions or deletions.

Nucleotide diversity

Nucleotide diversity is a measure of the genetic variation that is present in a species and is important because it is key to most phenotypic variation and can reflect evolutionary history. We observed an average of one SNP every 130 bp in *P. trichocarpa*. For noncoding DNA the rate of polymorphism was one in 64, whereas the rate for coding regions was one in 229. The range of diversity spanned from $\pi_{\text{Gene}} = 0.00009$ for *PoptrTB1*, where we saw only a single polymorphism in four trees, to the significantly higher value of $\pi_{\text{Gene}} = 0.00683$ for *PoptrLFY* with 23 polymorphisms distributed among 41 trees (Table 2). The average level of diversity (π_{Total}) was 0.00184 which is lower than reported in any other tree species to date, but not significantly less than that of *Populus nigra* (G. Zaina & M. Morgante, personal communication) or *Pinus sylvestris* (Garcia-Gil *et al.* 2003). Silent-site diversity was also significantly higher for *PoptrLFY* ($\pi_{\text{Silent}} = 0.00958$) compared with the other genes whose π_{Silent} values ranged from 0 to 0.00349. These data are included in Table 2, but because π_{Silent} variation follows the same pattern as overall π_{Gene} they are not discussed further.

Distribution of SNPs within different populations

The frequency of different SNPs in different populations may reflect distinctive histories or selective forces that are affecting individuals in that population. We thus reasoned that looking at variation in as many populations as possible would give us the best picture of genetic diversity in a species. The SNP frequency among the trees tested

from the 41 populations varied from 0.011 to 0.456 (Supplementary material). Almost 40% of the SNPs we quantified were present in less than five percent of our samples, and the most common SNP frequency was 0.011 which represents a polymorphism that was present heterozygously in only a single individual. This indicates that there are likely polymorphisms that have not yet been detected through our limited sampling.

Average H_{OSite} was 0.225, which was not significantly different from average H_{ESite} (0.236), indicating that although our samples do not represent a single interbreeding population, they do not deviate significantly from Hardy-Weinberg equilibrium. When Bonferroni correction was used, none of the individual trees in our sample deviated significantly from the mean in their expected heterozygosity, indicating that none of the populations have obviously undergone a recent population bottleneck or a selective sweep.

Identification of hybrid trees

In order to determine whether any of the Ecotilling alleles identified in this pilot study could be associated with environmental or geographic traits, we tested for correlations between all alleles and these variables. We identified four trees from central British Columbia (11-BULH-4-1, 9-KTWD-4, 7-IRVC-3-1 and 7-IRVD-5-1) that had identical and unique distributions of specific alleles of *PoptrMP*, *PoptrNPR1* and *PoptrKNAT1*. Many of these alleles induced changes to the predicted protein sequences of these genes. An additional tree, 28-CMBF-5, also had the same unusual alleles, but originated on Vancouver Island, a considerable distance away from the north-central populations so when

the data were analysed on these samples, the CMBF population was not considered because of concerns about the origin of this tree.

The BULH, KTWD, IRVC and IRVD populations occur at the boundary between the ranges of *P. trichocarpa* and *Populus balsamifera* ssp. *balsamifera*, where these two species are known to hybridize (Brayshaw 1996). The larger size and apparent hardiness of the trees from these locations compared to others in the UBC common garden (Gilchrist & Cronk, unpublished) was suggestive of hybrid vigour and because all four of the trees from these populations were heterozygous for each of the unusual alleles, we sequenced this region of the *MONOPTEROS* locus in *P. balsamifera* in order to determine whether or not these trees were *P. trichocarpa* × *P. balsamifera* hybrids. Neither of the two *P. balsamifera* trees we sequenced displayed any of the unusual alleles. However, sequencing of the *MONOPTEROS* locus from *Populus deltoides* ssp. *monilifera* indicated that all of the unusual alleles we identified through Ecotilling are present in this species. Because the location of the stool bed where these trees were grown adjoins a plantation of *P. trichocarpa* × *P. deltoides* hybrids, our molecular data appear to have identified hybrid contaminants in the original plantation of pure black cottonwoods. Copies of these clones from a different site in a common-garden test showed differences in growth rates and stem elongation that support the idea that the clones tested in this study were not the same as the native *P. trichocarpa* trees planted at the second site (C. Ying, unpublished). Some of the unusual alleles identified in the hybrids are also present in *P. tremuloides*, a more distantly related poplar species, indicating that they may be common to ancestors of both species.

Unweighted UPGMA analysis was done using all genes collectively to examine the relationship of the populations to one another, and showed that the putative *P. trichocarpa* × *P. deltoides* individuals cluster in a group by themselves on the dendrogram while trees from the other populations show no defined population structure (Fig. 4). UPGMA with different distance measures and the neighbour-joining (NJ) method were used to test the sensitivity of the results to different analytical conditions. The same major groupings were retrieved in all analyses so only UPGMA results are shown. This analysis indicates that these four trees are genetically similar to one another, but different from the pure *P. trichocarpa* trees, consistent with them being hybrids harbouring *P. deltoides* alleles of the genes investigated.

Data from the trees shown to carry the *P. deltoides* alleles were not included in any statistical analysis of polymorphism and linkage disequilibrium in *P. trichocarpa* because they would have inflated the true values for within-species calculations. However, identification of these hybrid contaminants demonstrates a very useful application for the Ecotilling technique in differentiating between purebred and hybrid individuals in a population.

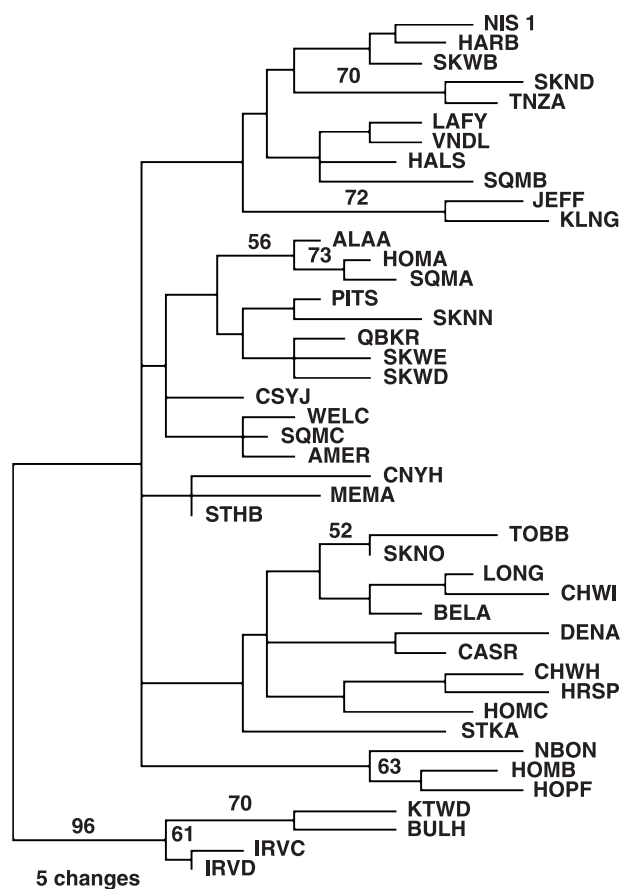


Fig. 4 Dendrogram constructed using alleles from all genes combined, showing absolute genetic distances mapped onto the UPGMA tree topology. Branch lengths represent numbers of substitutions. Numbers above the branches are bootstrap values (10 000 replicates).

Possible effects of selection

We looked at the ratio of nonsynonymous to synonymous changes in order to determine whether or not selection might be acting on any of the regions examined. *PoptrTB1*, *PoptrLFY*, *PoptrF5H* and *PoptrMPK4* could not be considered since there were no synonymous changes identified in *PoptrTB1* and no nonsynonymous SNPs identified in *PoptrLFY*, *PoptrF5H* or *PoptrMPK4*. All of the remaining loci had d_N/d_S ratios of less than one, indicating purifying selection in the regions of the genes we examined. Tajima's *D* (1989) was also calculated for each locus since this is another indicator of selection. Ecotilling data produce unphased genotype information for each locus, but DNASP 4.0 does not allow the use of ambiguous nucleotide codes, so haplotypes were assigned independently for all heterozygous sites for these calculations. The Tajima's *D* values were not significant and because this result was consistent with other results (above) we did not pursue this analysis further.

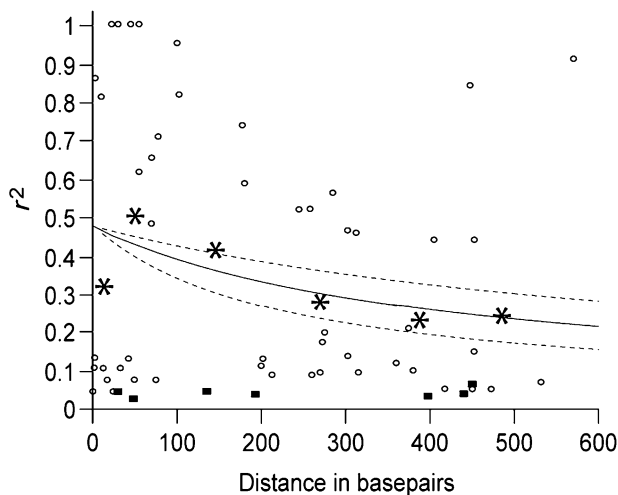


Fig. 5 Linkage disequilibrium in *Populus trichocarpa*. Calculations are for all pairs of alleles (excluding data from hybrid trees). The square of the correlated allele frequencies (r^2) is plotted as a function of distance between alleles. Regression line is based on Hill & Weir (1988) with the solid line illustrating the best fitting curve and the dashed curves representing the 95% bootstrap limits (resampling the data points). The method of Hill and Weir predicts an intercept of $0.455 + 0.744/n$, where n is the number of haplotypes (here, $n = 70$), regardless of the data; this reflects the expected effects of drift, mutation, and sampling. Open circles indicate r^2 values that are statistically significant ($P < 0.05$) and solid boxes indicate nonsignificant values. The asterisks indicate the means of adjacent points along the horizontal axis binned in groups of 10 (seven for the last asterisk).

Linkage disequilibrium

Linkage disequilibrium is used to measure the degree of association between two alleles among different individuals. It is important for making inferences about the history of a population or species and also forms the basis for association mapping where specific alleles are matched with specific phenotypic traits. The square of the correlation coefficient between alleles (r^2) was used as the measure of LD in this study because this is predicted to be more accurate than other measures when allele frequencies are low. The mean intragenic LD for our *P. trichocarpa* samples was $r^2 = 0.34$, and the regression remained substantially above the value of nonsignificant r^2 values over the 600 bp represented by our data (Fig. 5). This result was consistent whether log-linear regression, standard linear regression or the Hill and Weir regression (Hill & Weir 1988) were used. We recognize that the Hill and Weir model assumes no spatial heterogeneity and small mutation rates, assumptions that may not apply to western black cottonwood, but have chosen this representation because it allows direct comparison with other studies. Our data suggest that LD decays relatively slowly in *P. trichocarpa* compared to some other outbreeding plant species (Tenailon *et al.* 2001; Brown *et al.* 2004; Ingvarsson 2004). Indeed, genetic asso-

ciations between sites at different loci, while low (average $r^2 = 0.02$), were much more likely to be significant than expected by chance. Fifty-one out of 295 intergene r^2 values calculated were significant at the $P = 0.05$ level, which is substantially more than the 5% expected by chance ($P = 10^{-14}$ using exact probability test). Further support of significant LD in western black cottonwood comes from the fact that the estimated number of haplotypes (Table 2) is much lower than haplotype frequencies predicted by multiplying together the frequency of individual alleles.

Discussion

This pilot study provides a snapshot of the species-wide genetic diversity that underlies the phenotypic variation that has important biological consequences as well as practical implications for tree improvement in *Populus trichocarpa*, or western black cottonwood. The trees used in our study exhibit a wide spectrum of phenotypic differences including growth rate, timing of onset of bud flush and fall dormancy, and susceptibility to frost and disease (C. Ying, unpublished). We chose to look at trees from different populations, rather than examine the diversity within a single, or a few populations in order to obtain as broad as possible a picture of the variation that exists in this species. Ecotilling proves to be an effective medium-to-high throughput technique for the study of natural variation even in heterozygous, outbreeding species such as *P. trichocarpa*. This technique employs a mismatch-specific nuclease that cleaves amplified PCR fragments at the site of a nucleotide polymorphism. Our Ecotilling strategy allowed the genotypes of all 41 trees used in this study to be revealed on a single, 96-well gel for each locus, as shown in Fig. 2. We identified 63 novel SNPs in 8191 bp of aligned DNA sequence from nine *P. trichocarpa* genes using this technique. This is an average of one SNP every 130 bp for this species. Although we sequenced 54 loci in this study in order to ensure that our Ecotilling results were accurate, the minimum number of trees that would need to be sequenced in order to deduce the sequence of all nine genes in all of the 41 individuals we examined (352 loci) is 21. Once an example of each polymorphism has been sequenced it is easy to generate *in silico* sequence data for all of the trees carrying the same SNP. Ecotilling can also be useful for applications where direct sequencing is not necessary. For example, in our initial analysis of the Ecotilling data, we used synthesized sequences composed of 0, 1 and 2 (where 0 represents a site homozygous for the common allele, 1 a heterozygous site, and 2 a site homozygous for the rare allele) to model the distribution of SNPs and linkage disequilibrium. When we repeated this analysis after sequencing, we obtained very similar results, indicating that our prediction of diversity and LD levels through Ecotilling was very accurate. Ecotilling can also be used for

genotyping applications without subsequent sequencing when only the presence or absence of a band on the gel is needed as a marker.

A key objective of this study was to catalogue the level of DNA variation that is present in natural populations of *P. trichocarpa*. Although estimates of nucleotide diversity have been made for other woody plants, the availability of the poplar genome sequence makes it possible to obtain data for this species on a genomic scale such as has never before been possible in trees. Average nucleotide diversity in our study, measured as π_{Total} , was 0.0018 which is comparable to *Populus nigra* ($\pi_{\text{Total}} = 0.0024$; G. Zaina & M. Morgante, personal communication) and *Pinus sylvestris* ($\pi_{\text{Total}} = 0.0025$; Garcia-Gil *et al.* 2003), but considerably lower than reported for other trees such as *Cryptomeria japonica* ($\pi_{\text{Total}} = 0.0038$; Kado *et al.* 2003), *Pinus taeda* ($\pi_{\text{Total}} = 0.0039$; Brown *et al.* 2004), or *Populus tremula* ($\pi_{\text{Total}} = 0.011$; Ingvarsson 2004). Nucleotide diversity varied significantly between loci in our study, as it does in most species. It was lowest for the *PoptrTB1* locus ($\pi_{\text{Gene}} = 0.00009$), and highest for *PoptrLFY* ($\pi_{\text{Gene}} = 0.00683$). *Zea mays* ssp. *parviglumis*, another outbreeding, wind-pollinated wild species, shows an equally broad range of diversity with π_{Gene} ranging from 0.0036 to 0.0328 (White & Doebley 1999), and even in the inbreeding species *Arabidopsis thaliana* π ranges from 0.0007 (Olsen *et al.* 2002) to as high as 0.05 (Rose *et al.* 2004), depending on the locus. Animal systems are just as variable, with average π in humans being 0.00087, in gorillas, 0.00158 (Yu *et al.* 2004), and in *Drosophila melanogaster*, 0.007 (Oppentocht *et al.* 2002). Because nucleotide diversity depends on the rate of mutation, as well as the effects of selection and population history, locus-specific rates are probably more useful than species-specific averages for examining the meaning of this variation. The significantly high π_{Gene} value for *PoptrLFY* might result from balancing selection or from a locally high mutation rate. Although we did not detect any amino acid polymorphism in the sequences analysed, further sequencing of remaining coding regions and neighbouring genes is warranted to distinguish between these alternatives. In contrast, the unusually low level of variation at the *PoptrTB1* locus, while not significant after Bonferroni correction, is suggestive of a selective sweep or of a locally reduced mutation rate in this region. Sequence comparison between *P. trichocarpa* and related species is warranted to assess whether selectively important substitutions have occurred in western black cottonwood.

The strategy we employed for this study allowed discovery not only of heterozygous SNPs, but also of alleles that were different from the reference tree to provide a more accurate picture of overall nucleotide diversity. We selected Nisqually 1 as our reference because its genome had been sequenced and so results for each locus could be compared to the same standard. However, at sites where Nisqually 1 was heterozygous, we were unable to deter-

mine the genotype of homozygous trees from Ecotilling gel data because it was not possible to determine which of the two Nisqually 1 alleles was present in these trees. An advantageous alternative strategy would be to empirically select a homozygous tree as a reference for each individual locus even if it meant using a different reference tree for each gene. Another economical option would be to simply look for heterozygous markers and thereby eliminate the need for a standard reference.

In order to increase the efficiency of Ecotilling even more, DNA of up to eight individuals could be pooled before amplification, just as it is in reverse genetic screens (Till *et al.* 2003). Pooling is only feasible if genetic diversity is not too high because otherwise data analysis becomes complicated by the high number of bands on each gel. Nucleotide diversity would have to be below a threshold of about 0.07 in order for eightfold pooling to be practical, although lower fold pooling could still be used at higher levels of diversity. Average nucleotide diversity for *P. trichocarpa* is 0.00184, indicating that pooling would be a feasible approach for most loci in this species, and would not overly complicate analysis of gels such as that shown in Fig. 2. However, empirical assessment of nucleotide diversity at a given locus in *P. trichocarpa* or other wild species will be necessary to determine the optimal pooling strategy.

We used a combination of coding and noncoding sequences in our assessment of Ecotilling as a technique, and conclude that both types of DNA can provide useful data. Noncoding regions provide more variation and may therefore be more useful for mapping and genotyping. In the *PoptrLFY* gene, for example, we screened mostly noncoding DNA and detected significantly more variation than at other loci ($P < 0.0001$). However, we also identified a nonpolymorphic fragment of 150 nucleotides within the first intron (Fig. 3) that we speculate may demarcate a domain that is important for the gene's regulation since regions in this intron have been shown to be important for regulation in other species (Prasad *et al.* 2003; D. Weigel, personal communication). The data produced by this technique could be especially useful in a species like poplar whose genome has been sequenced, but for which gene annotation programs depend largely on similarities to other plant species and *ab initio* gene predictions. Identifying conserved regions amidst highly variable DNA could indicate regulatory domains or might identify coding regions that have been mislabelled by gene prediction programs.

SNPs in coding regions are less frequent, but investigation of these areas may allow identification of alleles that are under selection. In this study we observed very specific changes that are predicted to affect the protein sequence of the *PoptrMP* transcription factor in five trees. All of the unusual alleles at this locus were in linkage disequilibrium with one another and subsequent sequencing of the *MP* locus revealed that these alleles likely arose from

contaminating *P. trichocarpa* × *P. deltoides* hybrids in the original nursery collection. This set of polymorphisms that we have identified through Ecotilling appears to identify nucleotide differences that distinguish different species. This indicates that Ecotilling could be employed very efficiently by tree breeders and forest geneticists to track the pedigree of parent trees in breeding populations involving interspecific hybridization, or by taxonomists to establish high resolution geographic boundaries between species that hybridize naturally.

Linkage disequilibrium has significant implications for marker development and association mapping and average LD in *P. trichocarpa*, measured as r^2 , was 0.34 which is considerably higher than was recently reported for another *Populus* species, *P. tremula* (Ingvarsson 2004). This number is, however, comparable to *P. nigra* where r^2 also remains above 0.25 for several hundred base pairs (G. Zaina & M. Morgante, personal communication). The rate of decay of LD, as expected, is higher than in *A. thaliana*, an inbreeding species, where r^2 is maintained for up to 250 kb, and also higher than in animal models where, for example, in *Drosophila* it decays to negligible values only after approximately 1000 bp. In humans, LD varies considerably from region to region but extends for several hundred kilobases in some blocks of the genome. In species where LD has been shown to decay completely within 100–200 bp, only very closely linked markers will remain in disequilibrium with phenotypic characteristics that are of interest. In *P. trichocarpa* LD seems to be consistently maintained over distances of up to 600 bp, although direct comparison to the LD data of Ingvarsson (2004) for European aspen is difficult, given methodological and sampling difference.

In this study, we have characterized 63 new DNA polymorphisms that can now be used as markers in this species. Data on adaptive and growth traits are available for a large number of *P. trichocarpa* trees in our reference collection (C. Ying, unpublished), and once a clear geographic pattern of population structure has been established, association mapping of SNPs detected by Ecotilling could prove to be extremely fruitful.

In conclusion, we have shown that Ecotilling is a fast and efficient way of screening for genetic variation in a species for which mutagenesis and classical genetic analysis is difficult if not impossible. For small sample sizes in which there is a high level of polymorphism, we suggest that sequencing will remain the preferred option. However, Ecotilling appears to offer significant efficiency and cost gains over sequencing for SNP discovery when the sample size is large (> 100), particularly if the experimental design permits maximal pooling (eightfold), for instance for the discovery of rare SNPs. It may also be advantageous to perform a first round of Ecotilling to detect individuals that are homozygous for all SNPs in an amplicon. If homozygous DNA is used as a reference for Ecotilling it considerably simplifies data interpretation and minimizes the amount of

confirmatory sequencing required. Ecotilling is also likely to be much more cost-effective relative to sequencing in cases where raw polymorphism data is sufficient without base-pair determination, for example, in identifying species-specific polymorphisms that can then be used for species identification or to detect interspecific hybridization events.

Acknowledgements

This work was supported by funds from Genome BC/Genome Canada and the Province of British Columbia (to J.B., C.J.D., B.E.E.), Genome Prairie/Genome Canada and the UBC Michael Smith Laboratories (to G.W.H.), and the Faculty of Land and Food Systems (to Q.C.B.C.). We would like to thank the BC Ministry of Forests for making cuttings available to us for our reference collection and Toby Bradshaw for providing us with Nisqually 1 living samples, as well as Seane Trehearne and the UBC Botanical Garden and Centre for Plant Research for assistance with tree cultivation. We are very grateful to Stewart Rood, Sally Aitkin, Robert Guy, Collin Kelleher, and Toby Bradshaw for useful discussions; David Groden and Dana Aeschliman for statistical support; and Brad Till, Steve Henikoff, and Luca Comai for technical advice and support.

Supplementary material

The supplementary material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/MEC/EC2885/MEC2885sm.htm>

Table S1 *Populus trichocarpa* accessions used in this project

Table S2 Polymorphisms and their heterozygosity and effect on protein sequence

Table S3 Primer sequences

Figure S1 Sequences not obtained from GenBank.

References

- Borevitz JO, Liang D, Plouffe D *et al.* (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research*, **13**, 513–523.
- Bradshaw HD Jr, Ceulemans R, Davis J, Stettler RF (2000) Emerging model systems: poplar (*Populus*) as a model forest tree. *Journal of Plant Growth Regulators*, **19**, 306–313.
- Brayshaw TC (1996) *Catkin-Bearing Plants of British Columbia*. Royal British Columbia Museum, Victoria, British Columbia, Canada.
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences, USA*, **101**, 15255–15260.
- Comai L, Young K, Reynolds SH *et al.* (2004) Efficient discovery of nucleotide polymorphisms in populations by Ecotilling. *Plant Journal*, **37**, 778–786.
- Cubas P, Lauter N, Doebley J, Coen E (1999) The TCP domain: a motif found in proteins regulating plant growth and development. *Plant Journal*, **18**, 215–222.

- DeFrancesco L, Perkel JM (2001) In search of genomic variation: a wealth of technologies exists to find elusive genetic polymorphisms. *Scientist*, **15**, 24.
- Ehrling J, Buttner D, Wang Q *et al.* (1999) 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant Journal*, **19**, 9–20.
- Franke R, McMichael CM, Shirley AM *et al.* (2000) Modified lignin in tobacco and poplar plants overexpressing the *Arabidopsis* gene encoding ferulate 5-hydroxylase. *Plant Journal*, **22**, 223–234.
- Garcia-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology*, **12**, 1195–1206.
- Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiology*, **132**, 1107–1114.
- Gilchrist EJ, Haughn GW (2005) TILLING without a plough: a new method with applications for reverse genetics. *Current Opinion in Plant Biology*, **8**, 211–215.
- Gundry CN, Vandersteen JG, Reed GH *et al.* (2003) Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clinical Chemistry*, **49**, 396–406.
- Hardtke CS, Kukurshumova W, Vidaurre DP *et al.* (2004) Overlapping and non-redundant functions of the *Arabidopsis* auxin response factors MONOPTEROS and NONPHOTOTROPIC HYPOCOTYL 4. *Development*, **131**, 1089–1100.
- Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, **33**, 54–78.
- Ingvarsson PK (2004) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics*, October 16; [Epub ahead of print].
- Kado T, Yoshimaru H, Tsumura Y, Tachida H (2003) DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae *sensu lato*). *Genetics*, **164**, 1547–1559.
- Olsen KM, Womack A, Garrett AR, Suddith JJ, Purugganan MD (2002) Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. *Genetics*, **160**, 1641–1650.
- Oppentocht JE, vanDelden W, vandeZande L (2002) Isolation and characterization of the genomic region from *Drosophila kuntzei* containing the *Adh* and *Adhr* genes. *Molecular Biology and Evolution*, **19**, 1026–1040.
- Petersen M, Brodersen P, Naested H *et al.* (2000) *Arabidopsis* MAP kinase 4 negatively regulates systemic acquired resistance. *Cell*, **103**, 1111–1120.
- Pieterse CMJ, VanLoon LC (2004) NPR1: the spider in the web of induced resistance signaling pathways. *Current Opinion in Plant Biology*, **7**, 456–464.
- Prasad K, Kushalappa K, Vijayraghavan U (2003) Mechanism underlying regulated expression of RFL, a conserved transcription factor, in the developing rice inflorescence. *Mechanisms of Development*, **120**, 491–502.
- Rogers LA, Campbell MM (2004) The genetic control of lignin deposition during plant growth and development. *New Phytologist*, **164**, 17.
- Rose LE, Bittner-Eddy PD, Langley CH, Holub EB, Michelmore RW, Beynon JL (2004) The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in *Arabidopsis thaliana*. *Genetics*, **166**, 1517–1527.
- Rottmann WH, Meilan R, Sheppard LA *et al.* (2000) Diverse effects of overexpression of LEAFY and PTLF, a poplar (*Populus*) homolog of LEAFY/FLORICAULA, in transgenic poplar and *Arabidopsis*. *Plant Journal*, **22**, 235–246.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.
- Salamov A, Solovyev V (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Research*, **10**, 516–522.
- Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.
- Strauss SH, Martin FM (2004) Poplar genomics comes of age. *New Phytologist*, **164**, 1–4.
- Swofford DL (2003) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Taylor NE, Greene EA (2003) PARSESNP: a tool for the analysis of nucleotide polymorphisms. *Nucleic Acids Research*, **31**, 3808–3811.
- Tenaillon MI, Sawkins MC, Long AD *et al.* (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences, USA*, **98**, 9161–9166.
- Till BJ, Burtner C, Comai L, Henikoff S (2004) Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Research*, **32**, 2632–2641.
- Till BJ, Reynolds SH, Greene EA *et al.* (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Research*, **13**, 524–530.
- White SE, Doebley JF (1999) The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. *Genetics*, **153**, 1455–1462.
- Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology and biotechnology. *Plant Physiology*, **126**, 485–493.
- Yu N, Jensen-Seaman MI, Chemnick L, Ryder O, Li WH (2004) Nucleotide diversity in gorillas. *Genetics*, **166**, 1375–1383.

The authors have broad and complementary interests in *Populus* genomics, plant gene function, plant adaptation, population biology and evolution, and development of TILLING and EcoTILLING as analytical tools. Authors Gilchrist and Haughn are using TILLING as a reverse genetic tool to investigate gene function in agricultural crops, and to expand the repertoire of induced loss-of-function alleles in model systems (*Arabidopsis* and *C. elegans*). Authors Ellis, Bohlmann, Douglas and Cronk have general interests in *Populus* genomics, as well as specific interests in plant and tree metabolism, defense, development, and evolution. In addition, Cronk and Douglas are specifically interested in natural variation in *Populus*, and phenotype-genotype associations in this genus with respect to adaptive traits and traits of commercial value (e.g. wood quality).
