Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence

T.-M. Yin[1]*, S. P. DiFazio[1]*, L. E. Gunter[1], S. S. Jawdy[1], W. Boerjan[2] and G. A. Tuskan[1]

[1]Environmental Sciences Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37830-6422, USA; [2]Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Technologiepark 927, B-9000 Gent, Belgium

Summary

Author for correspondence:
Stephen P. DiFazio
Tel: +1 865 574 7844
Fax: +1 865 576 9939
E-mail: difazios@ornl.gov

• In an attempt to elucidate the molecular mechanisms of *Melampsora* rust resistance in *Populus trichocarpa*, we have mapped two resistance loci, *MXC3 and MER*, and intensively characterized the flanking genomic sequence for the *MXC3* locus and the level of linkage disequilibrium (LD) in natural populations.
• We used an interspecific backcross pedigree and a genetic map that was highly saturated with AFLP and SSR markers, and assembled shotgun-sequence data in the region containing markers linked to *MXC3*.
• The two loci were mapped to different linkage groups. Linkage disequilibrium for *MXC3* was confined to two closely linked regions spanning 34 and 16 kb, respectively. The *MXC3* region also contained six disease-resistance candidate genes.
• The *MER* and *MXC3* loci are clearly distinct, and may have different mechanisms of resistance, as different classes of putative resistance genes were present near each locus. The suppressed recombination previously observed in the *MXC3* region was possibly caused by extensive hemizygous rearrangements confined to the original parent tree. The relatively low observed LD may facilitate association studies using candidate genes for rust resistance, but will probably inhibit marker-aided selection.

Key words: disease-resistance genes, linkage disequilibrium (LD), *Melampsora*, *Populus* (poplar), recombination repression.

Introduction

One of the most economically important diseases in *Populus* plantations is caused by the rust fungus *Melampsora spp.* As a foliar disease, *Melampsora* causes premature defoliation, reduction in diameter growth, predisposition to other insect and disease pests, and plant death (Thielges & Adams, 1975; Newcombe *et al.*, 1994; Steenackers *et al.*, 1996). Developing genetically resistant plant materials has been the most effective means of controlling the economic impact of *Melampsora*.

A successful disease-resistance breeding program relies on an understanding of genetic interactions between hosts and pathogens (Hsiang & Van Der Kamp, 1985; Han *et al.*,

2000). Quantitative and molecular genetic approaches have been initiated to identify key genetic loci conferring disease resistance in forest trees, including resistance to leaf rust (*Melampsora* spp.) in poplar (Hsiang & Van Der Kamp, 1985; Cervera *et al.*, 1996; Newcombe & Bradshaw, 1996; Lefevre *et al.*, 1998). Molecular markers have been identified that are tightly linked to several disease-resistance loci (Devey *et al.*, 1995; Wilcox *et al.*, 1996; Lefevre *et al.*, 1998; Stirling *et al.*, 2001), but resistance genes have yet to be cloned and functionally characterized for a forest tree.

Tremendous advances have been made in recent years in understanding the molecular mechanisms of disease resistance in model plants. Martin *et al.* (2003) divided known genes conferring major gene resistance to pathogens into five main categories. Genes in category 1 contain a serine/threonine

---

*These authors contributed equally to this work.

protein kinase (STK) domain. Categories 2 and 3 each consist of the nucleotide-binding site/leucine-rich repeat (NBS/LRR) families. These categories differ in their N-terminal domain, a coiled-coil in the case of category 2, and a Toll-Interleukin 1 (TIR) receptor domain in category 3. While these three types of resistance gene are typically localized intracellularly, genes in categories 4 and 5 contain a *trans*-membrane domain, and each contains an LRR located extracellularly. Category 4 has no known intracellular domains, while category 5 contains an STK domain (Meyers *et al.*, 1999; Ellis *et al.*, 2000; Martin *et al.*, 2003). Plant resistance genes typically occur in clusters in plant genomes, and may account for as much as 2% of the protein-coding genes in *Arabidopsis* (Young, 2000; Meyers *et al.*, 2003).

Populus possesses a relatively small genome, 480–520 Mbp (International *Populus* Genome Consortium, unpublished data), similar in size to the rice genome, only three times larger than *Arabidopsis*, and 40 times smaller than most conifers. Positional cloning should therefore be a viable approach for obtaining disease-resistance genes in *Populus*. Rust-resistance genes have been studied intensively using genetic maps (Cervera *et al.*, 1996; Villar *et al.*, 1996; Lefevre *et al.*, 1998), with particular emphasis on the *MXC3* and *MER* loci (Stirling *et al.*, 2001; Zhang *et al.*, 2001). *MXC3* confers resistance to poplar leaf rust caused by *Melampsora × columbiana* (pathotype 3) (Stirling *et al.*, 2001), whereas *MER* confers resistance to *M. larici-populina* races E1, E2 and E3. These two loci were identified independently using a local mapping strategy based on bulked segregant analysis (Michelmore *et al.*, 1991). Interestingly, the genomic regions surrounding each of these loci demonstrated repressed recombination and segregation distortion (Stirling *et al.*, 2001; Zhang *et al.*, 2001). However, these two loci were not previously placed on a comprehensive *Populus* genetic map, and it was not clear if they were clustered on the same chromosome, or even if the phenotypes arose from pleiotropic effects of a single gene. In this study we located the *MER* and *MXC3* loci on a dense *Populus trichocarpa* genetic map, employing markers and sequence tags derived from sequences of the bacterial artificial chromosomes (BACs) where the two loci are located. We further characterized the genomic region surrounding *MXC3* to determine the extent of linkage disequilibrium and to identify putative resistance genes.

Materials and Methods

Plant materials

We mapped the two resistance loci using a high-density comprehensive genetic map constructed in family 13 based on a two-way pseudotestcross strategy (Yin *et al.*, 2004). Family 13 was created by crossing a hybrid female, clone 52-225 (*P. trichocarpa*, 93–968 × *P. deltoides*, ILL-101) with an alternative *P. deltoides* male, clone D109. DNA samples from 171 progeny were used for linkage analysis. For the LD analysis, a set of 85

unrelated *P. trichocarpa* individuals from natural populations was used. These trees were haphazardly sampled from *P. trichocarpa* populations along a 400 km east–west transect through Oregon and Washington, USA, and were confirmed to be unique genets using microsatellite markers.

Mapping of the *MXC3* and *MER* loci

Markers linked to the *MXC3* locus were derived from end- and subclone-sequences of a BAC library derived from *P. trichocarpa* clone 383-2499 (Nisqually 1), a tree that is heterozygous for the original resistance allele (Stirling *et al.*, 2001). Nisqually 1 is also the source of the template that was used for whole-genome shotgun sequencing by the Department of Energy (Tuskan *et al.*, 2004a). Four microsatellite (SSR) markers (O_348, O_349, O_356, O_394) developed from BAC 41g18 (Tuskan *et al.*, 2004b) and five sequence-tagged site (STS) markers from BAC clones 41g18, 99a06, 47m20, 31e15 and 19m15 (Table 1; Stirling *et al.*, 2001) were used to map the *MXC3* locus.

Markers linked to the *MER* locus were developed from a BAC library derived from *Populus × euramericana* cv. Ghoy, an F_1 hybrid from a controlled cross between *P. deltoides* cv. s9-2 × *P. nigra* cv. Ghoy. Two SSRs (R7 and R8) were developed for this study from a BAC located 0.4 CentiMorgan (cM) away from the *MER* locus (Table 1; Lescot *et al.*, 2004). The above markers were added to an existing genetic map for the female parent in family 13, consisting of 77 SSRs and *c.* 500 AFLP markers (Yin *et al.*, 2004).

SSRs and STSs were analyzed with fluorescent dye-labeled primers (HEX and FAM, Applied Biosystems, Foster City, CA, USA) using an ABI 3700 capillary sequencer, as described previously (Yin *et al.*, 2004). Multipoint linkage analysis was performed using the program MAPMAKER/EXP ver. 3.0b, with threshold LOD values of 10.0 and a recombination fraction of 0.3. CentiMorgan (cM) distances were calculated using the Kosambi function with corrections for segregation distortion using the program MAPDISTO (Lorieux *et al.*, 1995). Linkage group numbers were assigned according to Cervera *et al.* (2001).

### Assessing intermarker linkage disequilibrium

We tested intermarker LD in the *MXC3* region using two STSs and four SSRs in 82 unrelated individuals randomly sampled from wild populations of *P. trichocarpa* (Supplementary Material Table S1). No recombination events have previously been reported in the interspecific pedigree family 545 for the region from which these markers were derived (Stirling *et al.*, 2001). The genetic distances between markers were derived from family 13 (Fig. 1). As the haplotype phase is unknown in our data set, the significance of LD between markers A and B was tested using a likelihood ratio approach, using an empirical distribution obtained by a permutation procedure (Slatkin, 1996). The distribution of −2 times the natural logarithm of

Table 1 Parental genotypes and primer sequences for the sequence-tagged site (STS) (Stirling *et al.*, 2001) and SSR (Tuskan *et al.*, 2004b) loci linked to the *MXC3* or *MER* genes

Locus name	Loci[a]	Maternal genotype[b]	Paternal genotype[b]	Forward primer	Reverse primer
STS1_A	2	430/–	–/–	CATGCAAGCTTCGGGACTC	AGTAGCAAAGGTATTGACAG
STS3	1	380/–	–/–	TAAGTGTTGACTGAATGATG	ACTTAGATTTCTCATTCCTAC
O_348	1	–/–	170/–	CTTCAAGCTCAGCAATGCAA	CCTTGTAGTTGTCAGCGATCC
O_349	1	199/262	193/262	GAGCATGAAGCATGAGCAGA	TTTTCAGAACCAGGGGAAAA
O_356	1	145/160	108/108	CCACGCTCGACAACATTTTA	AATCGTCCAATAAAAGCCACA
O_394	1	189/189	189/189	AAAAAGCCCCACAATTATTCA	GCAAGTTGCAATTGATGTCC
R7	1	273/–	–/–	CACCCGGGTTCCAGATCTAC	CAATGTTATCAGACGGCCCA

[a]Number of loci amplified with primers in family 13. Only one locus was mapped to the rust-resistance regions for each primer set.
[b]Genotypes indicate the sizes (bp) of mapped alleles. '–', null alleles.
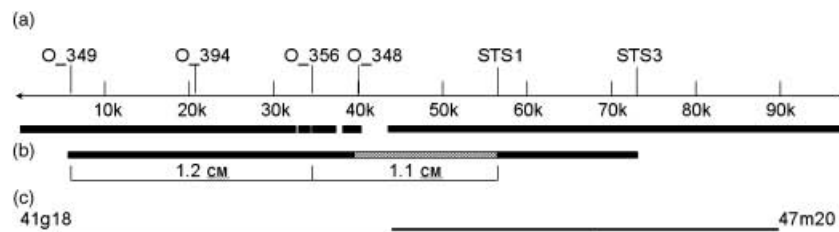
Fig. 1 Physical and genetic map of the *MXC3* region of linkage group IV. (a) Physical map derived from shotgun sequence assembly. Positions of SSR (markers beginning with 'O_') and STS markers are indicated above the line; physical length is below the line. Black bar, contiguous sequence. (b) Regions of significant linkage disequilibrium (black bar) and linkage equilibrium (hashed bar) between mapped markers. Mapping distance is indicated below the bar. There were no recombinations observed between STS1 and STS3. (c) Approximate positions of the BAC clones originally used to study this region (Stirling *et al.*, 2001).

this likelihood ratio asymptotically follows a χ^2 distribution, with degrees of freedom equal to $(u-1)(v-1)$, where u and v are the number of alleles at markers A and B, respectively. To reduce the effects of rare alleles and to preserve the power to reject the null hypothesis of linkage equilibrium by limiting the degrees of freedom of the test statistic, we pooled alleles with frequency < 5% (Mohlke *et al.*, 2001; Ohashi & Tokunaga, 2003). In addition, haplotype states and frequencies were estimated by maximum-likelihood using an expectation-maximization method, as implemented by the ARLEQUIN program (Excoffier & Slatkin, 1995). The significance of pairwise LD was also assessed by Fisher's exact probability test of contingency tables on the basis of a Monte Carlo Markov chain algorithm (Guo & Thompson, 1992).

Segregation distortion

We detected departures from Mendelian segregation using χ^2 tests. In addition, we estimated the selection intensity (S) of markers segregating 1 : 1 as:

$$S = (o - e)/e \qquad \text{Eqn 1}$$

which can also be obtained from:

$$S = \sqrt{(\chi^2/4N)} \qquad \text{Eqn 2}$$

where o is the observed number of individuals with the allele, e is the expected number of individuals with the allele, N is the total number of individuals. Thus the χ^2 of a marker with recombination rate of r would be:

$$\chi^2 = 4N[S - (0.5 + S)r]^2 \qquad \text{Eqn 3}$$

Construction of sequence scaffolds

We used sequence scaffolds from the draft *Populus* genome assembly (12 February 2004 build), which was accomplished using the JAZZ shotgun sequence assembler (Aparicio *et al.*, 2002) at the US Department of Energy (DOE)'s Joint Genome Institute at an average depth of 8.1X and an average PHRED quality score of 53.3 (Gordon *et al.*, 1998). In addition, we built mini-scaffolds from 384 of our own shotgun sequences from BACs 41g18 and 47m20 (Stirling *et al.*, 2003) and 5738 individual reads from the *Populus* genome sequence database from JGI (http://genome.jgi-psf.org/), using the PHRED/PHRAP/CONSED suite of programs (Gordon *et al.*, 1998). SSR and STS marker positions were determined by BLASTN (Altschul *et al.*, 1997) searches against the sequence scaffolds using the primer sequences as queries. We required that primers have an 80% or greater match over 70% of their length, and that they be correctly oriented and separated by the distance expected, based on the size of the products amplified.
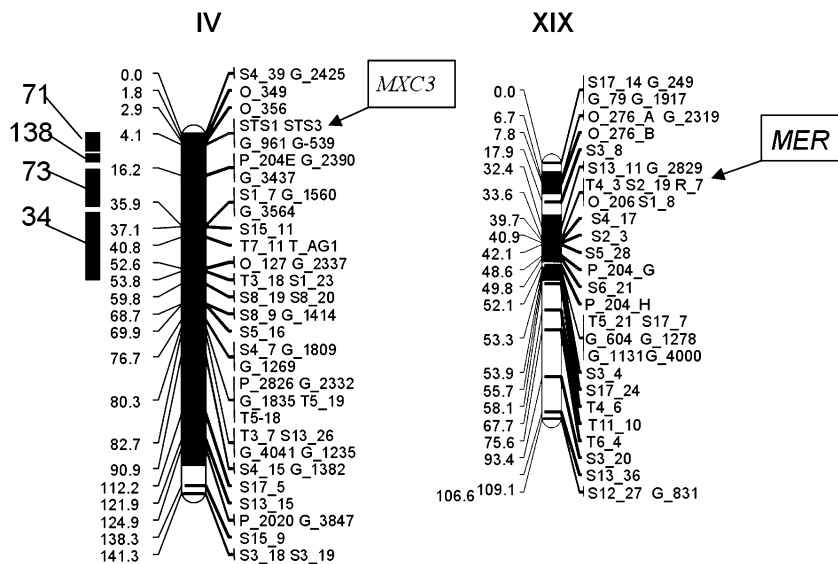
Fig. 2 Linkage groups on which markers linked to the *MXC3* and *MER* loci were mapped. Approximate positions of *MXC3* and *MER* loci are indicated, based on mapping positions of linked markers. Vertical bars connecting marker names indicate markers mapped to the same bin. Solid shading on linkage groups indicates chromosomal regions with significant segregation distortion in favor of *Populus trichocarpa* alleles. Vertical bars to left of linkage group IV represent sequence scaffolds. Scaffolds and gaps are not drawn to physical scale, but encompass known marker positions. AFLP marker names begin with T and S; microsatellite marker names begin with O, P, G or R. See Yin *et al*. (2004) for marker details.

Draft gene annotation

We explored the gene content of four sequence scaffolds covering *c.* 6.65 Mb in the region of the *MXC3* locus. Draft annotations of the sequence scaffolds were accomplished using the *Arabidopsis*-trained *ab initio* gene caller Genscan (Burge & Karlin, 1997), followed by hand corrections for predicted proteins with homology to known disease-resistance proteins. We derived putative functional characterizations by BLASTP searches of predicted proteins against the *Arabidopsis* protein database of July 2002 obtained from the Arabidopsis Information Resource (www.arabidopsis.org) and the non-redundant protein database of GenBank as of February 2004. We identified shared motifs using MEME/MAST (Bailey & Gribskov, 1998); delineated known protein motifs using INTERPROSCAN (InterPro Consortium, 2001); and predicted membrane-spanning regions using the dense alignment surface method (Cserzo *et al.*, 1997). Alignments between *Populus* genes and the best-matching disease resistance homologs were accomplished using CLUSTALX (Thompson *et al.*, 1997). These annotations should be considered initial drafts, pending release of *ab initio* gene predictions that are currently being accomplished with gene-calling software trained specifically for *Populus* (Tuskan *et al.*, 2004a).

Results

Map position

Two of the five tested STS segregated in family 13. STS1 is from the T7 end of BAC 41g18, and STS3 is from the SP6 end of BAC 47m20 (Stirling *et al.*, 2001). STS1 amplified two loci, STS1_A (430 bp) and STS1_B (284 bp), while STS3 amplified a single locus. No recombinants were discovered

between the loci STS1_A and STS3, and STS1_A and STS3 both mapped to linkage group IV. STS1_B mapped on linkage group VIII and was probably not derived from the *MXC3* region. Among the four SSR loci, O_394 generated an allele shared by both parents and showed no segregation (homozygous in one or both parents); O_348 generated a dominant paternally informative marker; and O_349 and O_356 revealed fully informative codominant markers. Therefore, O_349 and O_356 were placed on the genetic map. Two recombinants were detected between O_349 and O_356, and the two markers were mapped adjacent to the STS markers with one recombinant detected between the STS markers and SSR O_356 and three recombinants between the STS markers and SSR O_349. Based on these data, the *MXC3* locus, STS1_A, STS3, O_349 and O_356 have been placed at the end of linkage group IV and encompass a distance of 2.3 cM (Fig. 2).

We also placed the *MER* locus (Zhang *et al.*, 2001) on our map via two SSRs. One of the SSRs, R8, revealed a homozygous locus in our mapping pedigree, while the other, R7, generated segregating alleles originating from *P. trichocarpa*. The latter marker was 1.21 cM from previously mapped AFLP markers S2-19 and S13-11, and 1.80 cM from AFLP marker T4-3 in a two-point analysis. SSR R7 was thus located 33.6 cM from one end of linkage group XIX. When multipoint linkage analysis was performed with error detection on, the four closely linked markers were mapped in a single bin (Fig. 2).

Linkage disequilibrium

We assessed LD among markers in the *MXC3* region using 82 putatively unrelated *P. trichocarpa* genotypes collected from natural populations. We observed between four and 15 alleles among these genotypes for the microsatellite and STS markers (Table 2; Supplemental Table 1). Strong intermarker LD

Table 2 Intermarker linkage disequilibrium (LD) analysis at the *MXC3* locus for 85 native *Populus trichocarpa* genotypes sampled from eight sites along the Columbia River in southern Washington and northern Oregon, USA

Locus	Pairwise distances – exact *P* value for LD						Position	Observed[a]	Analyzed[b]
	O_349	O_394	O_356	O_348	STS1	STS3			
O_349		< 0.001	< 0.001	< 0.001	0.52148	0.30931	1458098	11	6
O_394	14642*		< 0.001	< 0.001	0.03931	0.06040	1472740	6	2
O_356	24438*	9796*		< 0.001	0.21555	0.58901	1482536	15	8
O_348	34052*	19410*	9614*		0.15346	0.72347	1492150	9	6
STS1	50496	35854	26058	16444		< 0.001	1508594	5	4
STS3	66560	51918	42122	32508	16064*		1524658	4	3

Lower diagonal represents intermarker distances; upper diagonal represents *P* values from Fisher's exact test for LD.
[a]Number of alleles observed per locus.
[b]Number of alleles per marker after alleles with a frequency below 5% were pooled.
*Significance below 0.01 level by Fisher's exact test for LD.

extended over 34 kb among markers O_348, O_349, O_356 and O_394. Linkage disequilibrium was also detected over 16 kb between markers STS1 and STS3 (Table 2; Fig. 1). However, LD was not significant between the microsatellite and STS markers, which were separated by *c.* 16 kb.

Segregation distortion

We observed significant segregation distortion in family 13 over extensive regions of both linkage groups to which the *MXC3* and *MER* loci mapped. The distorted region of linkage group XIX, on which the *MER* locus was located, accounted for 31.1% of the chromosome; whereas for linkage group IV, on which *MXC3* was located, segregation distortion occurred across 92.1% of the chromosome, covering a region of roughly 130.1 cM (Fig. 2). Together these two regions account for 56% of the segregation distortion observed in this pedigree (Yin *et al.*, 2004). Assuming recombination frequency is the same across the genome, the physical length of genomic regions showing segregation distortion in our pedigree is *c.* 26 Mb on linkage group IV and 6.78 Mb on linkage group XIX (assuming a genome size of 500 Mb). Using the χ^2 values to estimate *S*, we infer that a minimum of five loci would be needed to cause the observed contiguous distortion over such a large region.

Physical assembly

Assembly of shotgun sequence scaffolds in this region demonstrated that the markers span *c.* 66 kb (Fig. 1). However, this scaffold contains four gaps that range in size from *c.* 250 to 3250 bp, and a complex region of insertions and deletions near marker STS1 that resulted in considerable divergence between chromosomal haplotypes (Supplemental Figs 1 and 2). The approximate size of the gaps is inferred from the average insert size of the clone libraries used for end sequencing, so the distance estimates between markers contain some uncertainty. For example, a random selection of 60 clones

from one of the sequenced 8 kb libraries had a range of insert sizes from 5 to 19 kb, with a standard deviation of 2.3 kb.

The average physical : genetic distance ratio for this region is 29 kb/cM. By contrast, the average physical : genetic distance ratio in the *Populus* genome is *c.* 200 kb/cM (500 Mb/2500 cM). Examined another way, the three observed recombinants (b_{ij}) in this region were significantly more than would be expected (x_{ij}) among the 171 progeny surveyed ($P_{(b_{ij} > x_{ij})} = 0.000117$).

Identification of candidate genes for disease resistance

Genomic scaffolds covering *c.* 6.65 Mb of sequence and 40 cM of map distance on linkage group IV (Fig. 2) contained a total of 1530 genes predicted by Genscan. Six of the predicted proteins had significant homology to genes that have previously been demonstrated to be involved in disease resistance in other species. Expectation scores from BLASTP were less than $1E − 20$, and protein similarity ranged between 23 and 62% (Table 3; Supplemental Fig. 3). However, in contrast to the findings of Zhang *et al.* (2001), we did not encounter clusters of putative NBS-LRR resistance genes in the *MXC3* region. We identified three predicted peptides with substantial homology to class 5 resistance genes (an STK domain, a transmembrane domain, and an LRR domain), and one with homology to class 4 genes (a transmembrane domain and an LRR) (Table 3). However, these were located nearly 30 cM from the region to which markers linked to *MXC3* mapped in family 13. We also identified two putative genes within 1.9 cM of markers linked to *MXC3* with similarity to pathogenesis-related proteins containing thaumatin-like domains.

Discussion

Recombination suppression

Stirling *et al.* (2001) reported suppressed recombination around the *MXC3* locus based on lack of recombinants within

Table 3 Summary of putative disease-resistance candidate genes found in sequence scaffolds linked to the *MXC3* region

Scaffold number[a]	Peptide number[b]	Length (aa)[c]	Position (kb)[d]	Marker[e]	Marker position (kb)[d]	Map distance (cM)[f]	Motifs[g]	Annotation	E[h]	Similarity (%)[i]	Top disease resistance homolog hit[j]
138	68	334	665	G_961	414	0.0	Thau STK	Thaumatin-related	5E-69	62	NP_195325.1 thaumatin family, pathogenesis-related protein [*Arabidopsis thaliana*]
138	70	659	690	G_961	414	0.0	Thau STK	Thaumatin-related	4E-50	42	NP_173432.2 thaumatin family [*Arabidopsis thaliana*]
34	60	1066	578	T_AG1	1957	36.7	LRR TM STK	Class 5	5E-92	39	NP_197963.1 disease-resistance protein family [*Arabidopsis thaliana*]
34	171	455	1506	T_AG1	1957	36.7	TM LRR	Class 4	4E-21	23	NP_177296.1 disease-resistance protein family (LRR) [*Arabidopsis thaliana*]
34	240	1158	2117	T_AG1	1957	36.7	TIR LRR TM STK	Class 5	2E-88	45	T10504 disease-resistance protein Cf-2.1 [*Lycopersicon pimpinellifolium*]
34	273	1088	2341	T_AG1	1957	36.7	LRR TM STK	Class 5	8E-31	34	T10504 disease-resistance protein Cf-2.1 [*Lycopersicon pimpinellifolium*]

[a]Scaffold identifier assigned by JGI, based on 12 February 2004 *Populus* genome assembly.
[b]Identifier for predicted peptide assigned by Genscan. Numbers are assigned consecutively (e.g. peptide 68 is the 68th predicted peptide on that scaffold).
[c]Length in amino-acid residues of the predicted peptide.
[d]Position within the scaffold – scaffolds are oriented toward the telomere where possible, except where only one marker is available (e.g. scaffold 138).
[e]Name of closest mapped marker. Derived from Yin *et al*. (2004; unpublished data).
[f]Distance in centimorgans of closest mapped marker from the STS1/STS3 markers in the *MXC3* region.
[g]Protein motifs, presented in the order in which they appear in the predicted peptide. Thau, thaumatin-like domain; STK, serine–threonine protein kinase; LRR, leucine-rich repeat; TM, *trans*-membrane domain; TIR, Toll and Interleukin-1 receptor domain.
[h]Expectation score from BLASTP search.
[i]Similarity between *Populus* predicted protein and disease-resistance protein as determined by CLUSTALX alignment using Blosum62 matrix.
[j]Description of top BLASTP hits for genes known to be involved in disease resistance, beginning with GenBank accession numbers. All peptides also had more significant hits to other genes with no known function in disease resistance.

a *c.* 300 kb contig for 1902 F_1 progeny. Recombination was also suppressed 10-fold in the vicinity of the *MER* locus (W. Boerjan, unpublished). Zhang *et al.* (2001) also reported a significantly higher than expected number of markers in the coupling phase, compared with those in the repulsion phase, associated with the *MER* locus. Zhang *et al.* (2001) proposed that these results might be caused by a large chromosomal deletion in one of the parental genotypes, which would cause a hemizygous state in their F_1 pedigree. In apparent support of the hemizygosity hypothesis, all of the markers in the *MER* region (10 markers covering 24.2 cM) and the *MXC3* region (13 markers covering 16.7 cM) were linked in the coupling phase in our pedigree. However, this observation could also be an artifact resulting from the two-way pseudotestcross mapping approach used in our study, which favors mapping of visible AFLP alleles derived from *P. trichocarpa*. In fact, 69% of the mapped AFLP markers were maternally informative in family 13 (Yin *et al.*, 2004), and the visible allele was derived from *P. trichocarpa* in 71% of these cases.

The Nisqually 1 sequence assembly in the *MXC3* region provided direct support for the hemizygosity hypothesis. We observed substantial differences between chromosomal haplotypes in the vicinity of the STS1 marker, apparently caused by complex insertion/deletion events. In addition, there is an interscaffold gap of undetermined size within 60 kb of STS3, and this could be caused by complex rearrangements or length polymorphisms between Nisqually 1 haplotypes. Furthermore, given that recombination rates were significantly higher than expected for the same region in family 13, it is possible that the parents of this pedigree lack these rearrangements. This is somewhat surprising given that family 13 is an inter-specific pedigree, and rearrangements and recombination suppression might therefore be expected to be elevated relative to pure species (Rieseberg, 2001). However, *P. trichocarpa* and *P. deltoides* are closely related species with a long history of interbreeding (Eckenwalder, 1984), so substantial regions of synteny and sequence conservation are to be expected.

Linkage disequilibrium

Linkage disequilibrium refers to the nonrandom occurrence of allelic combinations in haplotypes. Unlike linkage analysis based on a two- or three-generation pedigree structure, LD analysis can monitor historical recombination events. Therefore, if strong recombination suppression is a ubiquitous phenomenon in the *MXC3* region at the population level, the locus will be expected to harbor unusually extensive LD across the region, a phenomenon that is commonly referred to as a haplotype block (Wall & Pritchard, 2003). Our observation of two blocks of LD in the *MXC3* region is consistent with this expectation. It remains to be determined if this pattern of regions of LD separated by regions of higher recombination is common within the *Populus* genome, similar to the LD structure in humans (Wall & Pritchard, 2003).

The observation of relatively high recombination rates between the microsatellite markers in our pedigree apparently contradicts the observation of LD among these markers on a population level. On closer examination, it is not entirely surprising that the observed recombination rates in a hybrid female tree under controlled conditions would differ from average rates in a pure species growing in the wild. Furthermore, the contribution of recombination to the level of LD can be negligible compared with other population-level forces, including population admixture, mutation and selection (Shifman & Darvasi, 2001; Nordborg & Tavaré, 2002). Population admixture is a distinct possibility, as our tested trees were sampled from widely separated sites. Theoretically, when populations with different allele frequencies are pooled together, LD is generated in the pooled population (Wahlund's principle; Hartl & Clark, 1997). Therefore the mixture of samples from different geographic sites may alone account for the LD detected in the sampled population. However, an initial analysis of population substructure using five unlinked microsatellites and a subset of these samples did not reveal significant differentiation of our sampled populations (not shown), so it is unlikely that this effect can account entirely for the observed LD.

Segregation distortion in the *MXC3* region

We hypothesize that differential disease susceptibility of the *P. trichocarpa* (resistant) and *P. deltoides* (susceptible) parents of the pedigree was partially responsible for the large-scale segregation distortion observed in the vicinity of both disease-resistance loci. Host–pathogen coevolution may have played an important role in the evolution of these portions of the *Populus* genome. *Melampsora occidentalis* and *Melampsora medusae* are the two major poplar leaf rust species native to the US. *Melampsora occidentalis* occurs in the Pacific northwest and is a primary pathogen of *P. trichocarpa*, whereas *M. medusae* is native to the eastern US and is a major pathogen of *P. deltoides* (Hsiang & Van Der Kamp, 1985; Pinon, 1992). The offspring of our pedigree were planted in northern Minnesota, which is within the natural range of *P. deltoides* and *M. medusae* and outside the natural range of *P. trichocarpa*. It is possible that a nonhost resistance system may be operating in *P. trichocarpa* relative to *M. medusae*. Although the phenomenon of nonhost resistance – the inability of a pathogen to cause disease in particular plant species – still remains poorly understood (Heath, 2000; Kamoun, 2001), it is beginning to appear that nonhost resistance is mediated by some of the same mechanisms involved in host-specific resistance (Jones & Takemoto, 2004).

An alternative explanation is that the observed segregation distortion is caused by deleterious interactions of interspecific alleles and other genomic incompatibilities between the parental species (Li *et al.*, 1997; Fishman *et al.*, 2001; Yin *et al.*, 2004). However, it is difficult to explain

why a heterospecific interaction would favor the *P. trich-ocarpa* haplotype over the recurrent species for such a large number of linked loci (Yin *et al.*, 2004). It is more likely that positive selection for *P. trichocarpa* alleles is responsible for the observed distortion in our pedigree, possibly mediated in part by alleles conferring nonhost disease resistance.

Michelmore & Meyers (1998) proposed that divergent selection might be a major factor in the evolution of resistance genes. Under this hypothesis, the accumulation of resistance genes would trigger a 'snowball effect', leading to clustered gene distributions within haplotypes and creating segregation distortion beyond the regions containing the resistance genes (Kruglyak, 1999). Resistance genes typically occur in large clusters in plant genomes (Young, 2000). For example, at least 10 *R* genes conferring resistance to downy mildew (*Bremia lactucae*) cluster together and span several megabases (Meyers *et al.*, 1998). Similarly, in flax more than 30 different *R* genes have been mapped to five linkage groups (Islam & Shepherd, 1991). The *Arabidopsis* genome contains 149 NBS-LRR *R* genes, two-thirds of which occur in 43 clusters containing up to seven genes each (Meyers *et al.*, 2003). The clustered occurrence of disease-resistance genes appears to play a central role in the generation and maintenance of the tremendous diversity observed in these gene families, as domains are shuffled within and between clusters because of large-scale insertion/deletion events and unequal crossing-over and gene conversion (Young, 2000; Meyers *et al.*, 2003). Assuming similar clustering occurs in *Populus*, it is possible that clusters of resistance genes have seeded divergence of linkage groups IV and XIX in *Populus*. It does appear that linkage group XIX may contain such clusters (Zhang *et al.*, 2001; Lescot *et al.*, 2004), but evidence is much weaker for linkage group IV.

Disease-resistance candidate genes

We identified six genes in scaffolds linked to the *MXC3* region that were strong candidates for involvement in disease-resistance responses. The class 4 resistance genes have been shown to be involved in major gene resistance in tomato to the leaf mold pathogen *Cladosporium fulvum* (Dixon *et al.*, 1998), while class 5 genes confer resistance to *Xanthomonas oryzae* in rice (Song *et al.*, 1995). However, all four of the genes from these two classes were located at least 20 cm from the markers linked to the *MXC3* region in family 13 (the exact distance is unknown because of gaps in the sequence assembly and a lack of markers in close proximity to the genes). It is theoretically possible that the region of suppressed recombination observed in family 545 (Stirling *et al.*, 2001) could extend over this distance such that the causative genes are located several megabases from the associated markers. Although this seems unlikely, our data are not sufficient to assess this hypothesis, which would have to be explored directly in family 545.

We also identified predicted proteins closely linked to *MXC3* on an adjacent scaffold with high homology to the thaumatin-like pathogenesis-related (PR-5) proteins. These proteins are implicated in plant defensive responses to fungal infections, and appear to play a role in disrupting fungal cell-wall integrity (Selitrennikoff, 2001). Thaumatin-like proteins display antifungal activity *in vitro* against a wide variety of pathogens (Krebitz *et al.*, 2003; Van Loon & Van Strien, 1999), and enhance fungal resistance when overexpressed in transgenic plants (Campbell *et al.*, 2002; Velazhahan & Muthukrishnan, 2003). However, there are no reports of direct implication of PR-5 proteins in major gene resistance, and these are more likely to be generalized secondary effectors (Martin *et al.*, 2003; Monteiro *et al.*, 2003). Benko-Iseppon *et al.* (2003) recently identified PR-5-like sequences in the vicinity of a fusarium wilt-resistance locus in chickpea (*Cicer arietinum*), but they hypothesized that this reflects non-random localization of genes encoding pathogenesis-related proteins such as PR-5 with disease-resistance proteins, as has been observed in *Arabidopsis* (Benko-Iseppon *et al.*, 2003). However, it is still possible that the thaumatin-like genes we have identified are directly involved in conferring resistance. This hypothesis is supported by the presence of an STK domain in the predicted proteins, which is known to be involved in pathogen-related signal transduction in other characterized disease-resistance genes such as *Pto* from tomato and *Xa21* from rice (Martin *et al.*, 2003).

Given the high levels of segregation distortion and recombination repression observed in family 545, it is surprising that we failed to identify clusters of class 2 and 3 NBS-LRR-type resistance genes in the vicinity of the *MXC3* locus in Nisqually 1, and that the only typical resistance genes identified were located implausibly far from the locus. By contrast, the *MER* locus does appear to contain such clusters (Zhang *et al.*, 2001; Lescot *et al.*, 2004). There are at least two possible explanations for this discrepancy. First, the NBS-LRR genes may be present in the gap between scaffold 71 and scaffold 138 (Fig. 2), which is within 60 kb of STS3. Large length mutations that differentiate haplotypes can cause substantial problems for shotgun sequence-assembly programs such as JAZZ (N. Putnam, personal communication) and PHRAP (Green, 2001). Currently more than one-third of the 8.1*X Populus* genome shotgun assembly database occurs in scaffolds smaller than one megabase (Populus Genome Sequencing Consortium, unpublished), and much of this misassembly is probably caused by hemizygosity in the sequenced tree. Another possibility is that the *MXC3* resistance phenotype is caused by a somewhat novel mechanism. The two thaumatin-like STK genes are excellent candidates. In addition, this region contained a substantial number of genes with one or more resistance-protein domains, but lacking in typical resistance protein structure (data not shown). These genes warrant closer examination as potential rust-resistance candidates.

Application of marker information

Linkage disequilibrium-based mapping can position a quantitative trait locus to small chromosomal segments, and this method has been widely used for fine-scale mapping of genes associated with human diseases (Ardlie *et al.*, 2002) and complex traits in animals (Haley, 1999). The efficacy of LD analysis depends on (1) the level of LD in the population studied; (2) its distribution and heterogeneity across the genome; and (3) its relationship with genetic or physical distances. The occurrence of LD is more frequent in a population that has undergone recent admixture, has a small effective population size, and/or has experienced intense selection. Lower levels of LD are expected to exist in panmictic populations. The lack of LD is one of the main problems hampering the application of marker-aided selection in practical tree-breeding programs (Strauss *et al.*, 1992; Haley, 1999). Unlike linkage analysis, which usually is used to associate markers that are reasonably close to the target gene, LD analysis can be used to detect markers within several kilobases. Linkage disequilibrium in open-pollinated forest trees with a long evolutionary history might be constrained to extremely short genomic regions, and thus require a large number of markers to find any association with a trait. From the limited data related to LD within the *MXC3* region of the *Populus* genome, pedigree-based linkage analysis may still be the most efficient approach for identifying target genes. However, high map resolution will require extremely large pedigrees. Our results demonstrate that even very tightly linked makers derived from a pedigree are not applicable for marker-assisted selection in unrelated pedigrees. The extent of haplotype blocks is a double-edged sword (Wall & Pritchard, 2003): more markers are needed to detect associations if LD extends only to a small region, but this also allows one to get closer to the candidate gene. For *Populus*, with the advent of the whole genome database (Tuskan *et al.*, 2004a), abundant information on candidate genes will be available for most traits. Linkage disequilibrium analysis combined with candidate gene information may be an efficient approach for clarification of gene function and gene isolation, especially for genes that are physically separated but closely linked, such as plant disease-resistance genes.

Conclusions

Genetics studies have revealed qualitative and quantitative interactions between *Populus* spp. and *Melampsora* spp. (Thielges & Adams, 1975; Lefevre *et al.*, 1994; Newcombe & Bradshaw, 1996; Newcombe *et al.*, 1996; Legionnet *et al.*, 1999). There are currently at least eight recognized species of *Melampsora*, with a large number of distinct physiological races varying in their specificity to individual poplar genotypes (Prakash & Thielges, 1987; Pinon, 1992). Therefore a number of genes may have evolved in the host to defend against the pathogen, and many of these resistance genes may be clustered together on chromosomes (Meyers *et al.*, 1998; Michelmore & Meyers, 1998; Graham *et al.*, 2002; Trevor *et al.*, 2002). The *MER* gene appears to be embedded in such a cluster, making it difficult to identify the particular resistance gene of interest from among the many closely linked candidates. On the other hand, the *MXC3* gene may well exist outside such clusters, facilitating functional characterization. Further characterization of the *MXC3* candidates could be readily accomplished by transformation of susceptible clones, or by association studies using single nucleotide polymorphisms in or near the candidate genes in populations for which rust resistance is segregating. This is among the first of many examples of how sequencing of the *Populus* genome is revolutionizing tree biotechnology and forest ecology research, and we anticipate a flurry of such studies to follow the publication of the sequence this year.

Acknowledgements

Supplementary Material

The following material is available as Supplementary material at http://www.blackwellpublishing.com/products/journals/suppmat/NPH/NPH1161/NPH1161sm.htm

Table S1 Genotypes of trees from wild *Populus trichocarpa* populations used for calculation of linkage disequilibrium (LD)
Figure S1 Assembly of two divergent haplotypes from the *MXC3* region containing marker STS1
Figure S2 Alignment of portions of two haplotypes from the *MXC3* region
Figure S3 Alignments between predicted *Populus* proteins from the *MXC3* region and the closest matching disease-resistance gene homologs.

References

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia J, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MDS, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJK, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**: 1301–1310.

Ardlie KG, Kruglyak L, Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**: 299–309.

Bailey TL, Gribskov M. 1998. Combining evidence using *P*-values: application to sequence homology searches. *Bioinformatics* **14**: 48–54.

Benko-Iseppon AM, Winter P, Huettel B, Staginnus C, Muehlbauer FJ, Kahl G. 2003. Molecular markers closely linked to fusarium resistance genes in chickpea show significant alignments to pathogenesis-related genes located on *Arabidopsis* chromosomes 1 and 5. *Theoretical and Applied Genetics* **107**: 379–386.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268**: 78–94.

Campbell MA, Fitzgerald HA, Ronald PC. 2002. Engineering pathogen resistance in crop plants. *Transgenic Research* **11**: 599–613.

Cervera MT, Gusmão J, Steenackers M, Peleman J, Storme V, Vanden Broeck A, Van Montagu M, Boerjan W. 1996. Identification of AFLP molecular markers for resistance against *Melampsora larici-populina* in *Populus*. *Theoretical and Applied Genetics* **93**: 733–737.

Cervera MT, Storme V, Ivens B, Gusmão J, Liu BH, Hostyn V, Van Slycken J, Van Montagu M, Boerjan W. 2001. Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* **158**: 787–809.

Cserzo M, Wallin ME, Simon I, von Heijne G, Elofsson A. 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Engineering* **10**: 673–676.

Devey ME, Delfino-Mix A, Kinloch BB Jr, Neale DB. 1995. Random amplified polymorphic DNA markers tightly linked to a gene for resistance to white-pine blister rust in sugar pine. *Proceedings of the National Academy of Sciences, USA* **92**: 2066–2070.

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JDG. 1998. The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell* **10**: 1915–1925.

Eckenwalder JE. 1984. Natural intersectional hybridization between North American species of *Populus* (Salicaceae) in sections *Aigeiros* and *Tacamahaca*. III. Paleobotany and Evolution. *Canadian Journal of Botany* **62**: 336–342.

Ellis J, Dodds P, Pryor T. 2000. Structure, function, and evolution of plant disease resistance genes. *Current Opinion in Plant Biology* **3**: 278–294.

Excoffier L, Slatkin M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Molecular Biology and Evolution* **12**: 921–927.

Fishman L, Kelly AJ, Morgan E, Willis JH. 2001. A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* **159**: 1701–1716.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Research* **8**: 195–202.

Graham MA, Marek LF, Shoemaker RC. 2002. PCR sampling of disease resistance-like sequences from a disease resistance gene cluster in soybean. *Theoretical and Applied Genetics* **105**: 50–57.

Green ED. 2001. Strategies for the systematic sequencing of complex genomes. *Nature Reviews Genetics* **2**: 573–583.

Guo SW, Thompson EA. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48**: 361–372.

Haley C. 1999. Advances in quantitative trait locus mapping. In: Dekkers JCM, Lamont SJ, Rothschild MF, eds. *From JL Lush to genomics: visions for animal breeding and genetics*. Ames, IA, USA: Iowa State University Press, 47–59.

Han ZM, Yin TM, Li CD, Huang MR, Wu RW. 2000. Host effect on genetic variation of *Marssonina brunnea* pathogenic to poplars. *Theoretical and Applied Genetics* **100**: 614–620.

Hartl DL, Clark AG. 1997. *Principles of population genetics*. Sunderland, MA, USA: Sinauer Associates.

Heath MC. 2000. Nonhost resistance and nonspecific plant defenses. *Current Opinion in Plant Biology* **3**: 315–319.

Hsiang T, Van Der Kamp BJ. 1985. Variation in rust virulence and host resistance of *Melampsora* on black cottonwood. *Canadian Journal of Plant Pathology* **7**: 247–252.

InterPro Consortium. 2001. The InterPro database: an integrated documentation resource for protein families domains and functional sites. *Nucleic Acids Research* **29**: 37–40.

Islam MR, Shepherd KW. 1991. Present status of genetics of rust resistance in flax. *Euphytica* **55**: 255–267.

Jones DA, Takemoto D. 2004. Plant innate immunity – direct and indirect recognition of general and specific pathogen-associated molecules. *Current Opinion in Immunology* **16**: 48–62.

Kamoun S. 2001. Nonhost resistance to *Phytophthora*: novel prospects for a classical problem. *Current Opinion in Plant Biology* **3**: 295–300.

Krebitz M, Wagner B, Ferreira F, Peterbauer C, Campillo N, Witty M, Kolarich D, Steinkellner H, Scheiner O, Breiteneder H. 2003. Plant-based heterologous expression of Mald2, a thaumatin-like protein and allergen of apple (*Malus domestica*), and its characterization as an antifungal protein. *Journal of Molecular Biology* **329**: 721–730.

Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22**: 139–144.

Lefevre F, Gouemourier MC, Faivre-Rampant P, Villar M. 1998. A single gene cluster controls incompatibility and partial resistance to various *Melampsora larici-populina* races in hybrid poplars. *Phytopathology* **88**: 156–163.

Lefevre F, Pichot C, Pinon J. 1994. Intra- and interspecific inheritance of some components of the resistance to leaf rust (*Melampsora larici-populina* Kleb.) in poplars. *Theoretical and Applied Genetics* **88**: 501–507.

Legionnet A, Muranty H, Lefevre F. 1999. Genetic variation of the riparian pioneer tree species *Populus nigra*. II. Variation in susceptibility to the foliar rust in *Melampsora larici-populina*. *Heredity* **82**: 318–327.

Lescot M, Rombauts S, Zhang J, Aubourg S, Mathé C, Jansson S, Rouzé P, Boerjan W. 2004. Annotation of a 95 kb *Populus deltoides* genomic sequence reveals a disease resistance gene cluster and novel class-I and class-II transposable elements. *Theoretical and Applied Genetics* **109**: 10–22.

Li Z, Pinson SRM, Paterson AH, Park WD, Stansel JW. 1997. Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (O*ryza sativa* 1) population. *Genetics* **145**: 1139–1148.

Lorieux M, Goffinet B, Perrier X, Deleon DG, Lanaud C. 1995. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. *Theoretical and Applied Genetics* **90**: 73–80.

Martin GB, Bogdanove AJ, Sessa G. 2003. Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology* **54**: 23–61.

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW. 1998. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* **10**: 1817–1832.

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant Journal* **20**: 317–332.

Meyers BC, Kozik A, Griego A, Kuang HH, Michelmore RW. 2003. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**: 809–834.

Michelmore RW, Meyers BC. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research* **8**: 1113–1130.

Michelmore RW, Paran I, Kesseli RV. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences, USA* **88**: 9828–9832.

Mohlke KL, Lange EM, Valle TT, Ghosh S, Magnuson VL, Silander K, Watanabe RM, Chines PS, Bergman RN, Tuomilehto A, Collins FS, Boehnke M. 2001. Linkage disequilibrium between microsatellite markers extends beyond 1 cm on chromosome 20 in Finns. *Genome Research* **11**: 1221–1226.

Monteiro S, Barakat M, Picarra-Pereira MA, Teixeira AR, Ferreira RB. 2003. Osmotin and thaumatin from grape: a putative general defense mechanism against pathogenic fungi. *Phytopathology* **93**: 1505–1512.

Newcombe G, Bradshaw HD. 1996. Quantitative trait loci conferring resistance in hybrid poplar to *Septoria populicola*, the cause of leaf spot. *Canadian Journal of Forest Research* **26**: 1943–1950.

Newcombe G, Bradshaw HD Jr, Chastagner GA, Stettler RF. 1996. A major gene for resistance to *Melampsora medusae* f. sp. *deltoidae* in a hybrid poplar pedigree. *Phytopathology* **86**: 87–94.

Newcombe G, Chastagner GA, Schuette W, Stanton BJ. 1994. Mortality among hybrid poplar clones in a stool bed following leaf rust caused *Melampsora medusae* f. sp. *deltoidae*. *Canadian Journal of Forest Research* **24**: 1984–1987.

Nordborg M, Tavaré S. 2002. Linkage disequilibrium: what history has to tell us. *Trends in Genetics* **18**: 83–90.

Ohashi J, Tokunaga K. 2003. Power of genome-wide linkage disequilibrium testing by using microsatellite markers. *Journal of Human Genetics* **48**: 487–491.

Pinon J. 1992. Variablility in the genus *Populus* in sensitivity of *Melampsora* rusts. *Silvae Genetica* **41**: 25–34.

Prakash CS, Thielges BA. 1987. Pathogenic variation in *Melampsora medusae* leaf rust of poplars. *Euphytica* **36**: 563–570.

Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution* **16**: 351–358.

Selitrennikoff CP. 2001. Antifungal proteins. *Applied and Environmental Microbiology* **67**: 2883–2894.

Shifman S, Darvasi A. 2001. The value of isolated populations. *Nature Genetics* **28**: 309–310.

Slatkin M. 1996. Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity* **76**: 377–383.

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* **270**: 1804–1806.

Steenackers J, Steenackers M, Steenackers V, Stevens M. 1996. Poplar disease consequences on growth and wood quality. *Biomass and Bioenergy* **10**: 267–274.

Stirling B, Newcombe G, Vrebalov J, Bosdet I, Bradshaw HD Jr. 2001. Suppressed recombination around the *MXC3* locus, a major gene for resistance to poplar leaf rust. *Theoretical and Applied Genetics* **103**: 1129–1137.

Stirling B, Yang ZK, Gunter LE, Tuskan GA, Bradshaw HD. 2003. Comparative sequence analysis between orthologous regions of the *Arabidopsis and Populus* genomes reveals substantial synteny and microcollinearity. *Canadian Journal of Forest Research* **33**: 2245–2251.

Strauss SH, Lande R, Namkoong G. 1992. Limitations of molecular-marker-aided selection in forest tree breeding. *Canadian Journal of Forest Research* **22**: 1050–1061.

Thielges BA, Adams JC. 1975. Genetic variation and heritability of *Melampsora* leaf rust resistance in Eastern Cottonwood. *Forest Science* **21**: 278–282.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876–4882.

Trevor L, Kunkel BN, Richards EJ. 2002. Epigenetic variation in *Arabidopsis* disease resistance. *Genes and Development* **16**: 171–182.

Tuskan GA, DiFazio SP, Teichmann T. 2004a. Poplar genomics is getting popular: the impact of the poplar genome project on tree research. *Plant Biology*. **6**: 2–4.

Tuskan GA, Gunter LE, Yang ZK, Yin TM, Sewell MM, DiFazio SP. 2004b. Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa*. *Canadian Journal of Forest Research* **34**: 85–93.

Van Loon LC, Van Strien EA. 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiological and Molecular Plant Pathology* **55**: 85–97.

Velazhahan R, Muthukrishnan S. 2003. Transgenic tobacco plants constitutively overexpressing a rice thaumatin-like protein (PR-5) show enhanced resistance to *Alternaria alternata*. *Biologia Plantarum* **47**: 347–354.

Villar M, Lefevre F, Bradshaw HD Jr, Ducros ET. 1996. Molecular genetics of rust resistance in poplars (*Melampsora larici-populina* Kleb./*Populus* sp.) by bulked segregant analysis in a 2 × 2 factorial mating design. *Genetics* **143**: 531–536.

Wall JD, Pritchard JK. 2003. Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics* **4**: 587–597.

Wilcox PL, Amerson HV, Kuhlman EG, Liu B-H, O'Malley DM, Sederoff RR. 1996. Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. *Proceedings of the National Academy of Sciences, USA* **93**: 3859–3864.

Yin TM, DiFazio SP, Gunter LE, Riemenschneider D, Tuskan GA. 2004. Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theoretical and Applied Genetics* **doi**: 10.1007/500122-004-1653-5.

Young ND. 2000. The genetic architecture of resistance. *Current Opinion in Plant Biology* **3**: 285–290.

Zhang J, Steenackers M, Storme V, Neyrinck S, Van Montagu M, Gerats T, Boerjan W. 2001. Fine mapping and identification of nucleotide binding site/leucine-rich repeat sequences at the *MER* locus in *Populus deltoides* s9-2. *Phytopathology* **91**: 1069–1073.