

- ROSSI, P., G. G. VENDRAMIN and R. GIANNI (1996): Estimation of mating system parameters in two Italian natural populations of *Fagus sylvatica*. *Can. J. For. Res.* **26**: 1187–1192.
- ROUSSET, F. (2008): GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Res.* **8**: 103–106.
- SCALFI, M., M. TROGGIO, P. PIOVANI, S. LEONARDI, G. MAGNASCHI, G. G. BENDRAMIN and P. MENOZZI (2004): A RAPD, AFLP and SSR linkage map, and QTL analysis in European beech (*Fagus sylvatica* L.). *Theor. Appl. Genet.* **108**: 433–441.
- STEINKELLNER, H., C. LEXER, E. TURETSCHKE and J. GLOSSL (1997): Conservation of (GA)_n microsatellite loci between *Quercus* species. *Mol. Ecol.* **6**: 1189–1194.
- STORME, V., A. V. BROECK, B. IVENS, D. HALFMAERTEN, J. VAN SLYCKEN, S. CSTIGLIONE, F. GRASSI, T. FOSSATI, J. E. COTTRELL, H. E. TABBENER, F. LEFEVRE, C. SAINTAGNE, S. FLUCH, V. KRSTUFEK, K. BURG, S. BORDACS, A. BOROVICS, K. GEBHARDT, B. VORNAM, A. POHL, N. ALBA, D. AGUNDEZ, C. MAESTRO, E. NOTIVOL, J. BOVENSCHEN, B. S. VAN DAM, J. VAN DER SCHOOT, B. VOSMAN, W. BOERJAN and M. J. M. SMULDERS (2004): *Ex-situ* conservation of black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development. *Theor. Appl. Genet.* **108**: 969–981.
- STREIFF, R. DUCOUSSO, C. LEXER, H. STEINKELLNER, J. GLOSSL and A. KREMER (1999): Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Molec. Ecol.* **8**: 831–841.
- TANAKA, K., Y. TSUMURA and T. NAKAMURA (1999): Development and polymorphism of microsatellite markers for *F. crenata* and the closely related species, *F. japonica*. *Theor. Appl. Genet.* **99**: 11–15.
- VICTORY, E. R., J. C. GLAUBITZ, O. E. RHODES JR. and K. E. WOESTE (2006): Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. *Amer. J. Bot.* **93**: 118–126.
- WANG, Y. (2008): Microsatellite transferability in Chestnut. *J. Amer. Soc. Hort. Sci.* **133**: 692–700.

Isolation, Characterization and Phylogenetic Analysis of Nucleotide Binding Site-encoding Disease-resistance Gene Analogues from European Aspen (*Populus tremula*)

By YONG ZHANG^{1,2,4}, SHOUGONG ZHANG³, LIWANG QI³, TAO ZHANG¹,
CHUNGUO WANG², CHENGBIN CHEN² and WENQIN SONG^{2,5}

(Received 5th February 2009)

Abstract

The majority of verified plant disease resistance genes (R genes) isolated to date was of the nucleotide binding site-leucine rich repeat (NBS-LRR) class. The conservation between different NBS-LRR R genes opens the avenue for the use of PCR based strategies in isolating and cloning other R gene family members or analogs (resistance gene analogue, RGA) using degenerate primers for these conserved regions. In this study, to better understand the R gene in European aspen (*Populus tremula*), a perennial tree, we used degenerate

primers to amplify RGA sequences from European aspen. Cloning and sequence characterization identified 37 European aspen RGAs, which could be phylogenetically classified into seven subfamilies. Deduced amino acid sequences of European aspen RGAs showed strong identity, ranging from 30.41 to 46.63%, to toll interleukin receptor (TIR) R gene subfamily. BLAST searches with reference to the genomic sequence of *P. trichocarpa* found 209 highly homologous regions distributed in 28 genomic loci, suggesting the abundance and divergence of NBS-encoding R genes in European aspen genome. Although, numerous studies have reported that plant R genes are under diversifying selection for specificity to evolving pathogens, non-synonymous to synonymous nucleotide substitution (d_N/d_S) ratio were <1 for NBS domains of European aspen RGA, showing the evidence of purifying selection in this perennial tree. In further analysis, many intergenic exchanges were also detected among these RGAs, indicating a probable role in homogenising NBS domains. The present study permits insights into the origin, diversification, evolution and function of NBS-LRR R genes in perennial species like European aspen and will be useful for further R gene isolation and exploitation.

Key words: *Populus tremula*, disease resistance gene, nucleotide binding site, resistance gene analogue, phylogenetic analysis.

¹ School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu 610054, P. R. China.

² College of Life Sciences, Nankai University, Tianjin 300071, P. R. China.

³ Laboratory of Cell Biology, The Research Institute of Forestry, The Chinese Academy of Forestry, Beijing 100091, P. R. China.

⁴ Corresponding author 1: YONG ZHANG, Laboratory of Molecular Cytogenetic Biology, School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu, 610054, P. R. China. Tel. 86-28-83206556, Fax 86-28-83206556. E-Mail: zhangyong916@uestc.edu.cn

⁵ Corresponding author 2: WENQIN SONG, Laboratory of Cell Biology, College of Life Sciences, Nankai University, Tianjin 300071, P. R. China. Tel. 86-22-23508241, Fax 86-22-23491710. E-Mail: songwq@nankai.edu.cn

Abbreviations: d_N , non-synonymous nucleotide substitution; d_S , synonymous nucleotide substitution; LRR, leucine rich repeat; NBS, nucleotide binding site; ORF, open reading frame; R gene, resistance genes; RGA, resistance gene analogue; TIR, toll interleukin receptor.

Introduction

Plants utilize a variety of strategies to withstand attack by a huge assortment of pathogens found in their habitats. The resistance genes (R genes) that confer resistance against pathogens that express matching avirulence genes in a “gene-for-gene” manner are one of the most effective and predominant strategies in plants (FLOR, 1971). Recently, over 30 disease resistance genes have now been cloned from a wide range of plant species, such as tomato, tobacco, *Arabidopsis*, flax and rice; they confer resistance to various plant pathogens, including viruses, bacteria, oomycetes, fungi, nematodes and insects (DEYOUNG and INNES, 2006; KUANG et al., 2008; MEYERS et al., 2005).

R genes are grouped into five classes based on the structure of their protein products. Most R genes encode the protein containing a nucleotide-binding site (NBS) domain and a stretch of leucine-rich repeats (LRR) at the C-terminus. Based on motifs located N-terminal to the NBS domain, the NBS-LRR super-family can be further subdivided into the TIR (*Drosophila* Toll and human Interleukin Receptor-like) subfamily and the non-TIR subfamily. The other four classes consist of: (1) cytoplasmic serine/threonine kinases; (2) extracellular LRRs anchored to a transmembrane domain; (3) receptor-like kinases (RLKs) with extracellular LRRs and an intracellular serine/threonine kinase domain; and (4) the HM1 gene encoding a toxin reductase (DEYOUNG and INNES, 2006; MEYERS et al., 2005; VAN DER HOORN and KAMOUN, 2008).

The NBS domain is characterized by several distinct motifs: P-loop/kinase-1a, kinase-2, kinase-3a, and GLPL, and all of them are highly conserved among species (MEYERS et al., 1999; PAN et al., 2000). These well-conserved regions within the NBS domain have been used to isolate resistance gene analogs (RGAs) in a broad range of plant species, such as soybean (KANAZIN et al., 1996), potato (LEISTER et al., 1996), rice (BAI et al., 2002), *Arabidopsis thaliana* (MEYERS et al., 1999; MEYERS et al., 2003), grapevine (DONALD et al., 2002), apple (LEE et al., 2003), strawberry (MARTINEZ ZAMORA et al., 2004), oats (IRIGOYEN et al., 2006), apricot (SORIANO et al., 2005), rose (HATTENDORF and DEBENER, 2007), banana (MILLER et al., 2008) and sugarcane (GLYNN et al., 2008). These RGAs are useful in physical mapping and as gene candidates in positional cloning. In addition, Conservation of the NBS domain has been used to study the genomic architecture of this gene family. R genes are unevenly distributed in plant genomes and many reside in local multigene clusters. The clustered distribution of R genes provides a reservoir of genetic variation from which new specificities can evolve. Mechanisms like duplication, unequal crossing over, ectopic recombination, gene conversion, and diversifying selection have been proposed to contribute to the structure of R gene clusters and the evolution of resistance specificities

(AMELINE-TORREGROSA et al., 2008; BAI et al., 2002; KUANG et al., 2004; PAN et al., 2000; YANG et al., 2008).

Despite the wealth of information on the structure and function of NBS-LRR RGA that has come from several elaborate analyses in herbaceous monocots and dicots, only fragmentary information is available on NBS-LRR genes in perennial species (KÖHLER et al., 2008; SORIANO et al., 2005; TUSKAN et al., 2006; ZHANG et al., 2008). Such long-lived species are more prone to attacks by pathogens before reproduction and their long generation time makes it impossible for them to match the evolutionary rates of a pathogen that goes through several generations every year (KÖHLER et al., 2008). *P. tremula* is a species of aspen (in section Leuce, subsection Trepidae) native to cool temperate regions of Europe and Asia, and is widely used for forestation, landscape enhancement and paper making (http://en.wikipedia.org/wiki/Populus_tremula). However, this aspen is susceptible to many diseases, which often severely affect the growth and yield. In contrast to tremendous progress in identification and characterisation of resistance loci in *P. deltoids* (in section Aegiros) (LESCOT et al., 2004) and RGA in *P. tomentosa* (in section Leuce, subsection Albidae) (ZHANG et al., 2008) as well as of putative disease R genes from genome sequences of *P. trichocarpa* (in section Tacamahaca) (KÖHLER et al., 2008; TUSKAN et al., 2006), little is currently known about disease resistance in *P. tremula*.

In the present study, we followed the PCR-based strategy using R gene specific degenerate primers to (i) isolate RGA from *P. tremula* to facilitate functional cloning of R gene from European aspen in the future, (ii) evaluate the extent of RGA sequences variation in European aspen genome, (iii) analyze the phylogenetic relationships of European aspen RGA sequences with the known R genes from other plant species, and (iiii) elucidate the evolution of European aspen RGA sequences and their patterns of genomic organization. These data permit insights into the origin, diversification, evolution and function of NBS-LRR resistance genes in perennial species like European aspen.

Materials and Methods

Plant material

Populus tremula was used as the experimental material in this study. The abscised branches from a randomly selected individual were picked from the collection in the Research Institute of Forestry, Chinese Academy of Forestry. The mature branches were surface-sterilized with 75% ethanol and were kept under hydroponic culture conditions. After sprouting, the branches were grown in an incubator at $25^{\circ}\text{C} \pm 0.5^{\circ}\text{C}$ under 16 h of artificial daylight and 8 h of darkness.

DNA extraction and PCR amplification of RGA

The *P. tremula* total genomic DNA was isolated from young leaf tissue using CTAB method according to MURRAY and THOMPSON (1980), with minor modifications as described by ZHANG et al. (2005). Degenerate primers used in this study were either designed on the basis of

Table 1. – Degenerate primers used for PCR amplification of RGAs from genomic DNA of *P. tremula*.

primer combination		oligonucleotide sequencea (5'→3')	conserved motif	reference
NBS-1	forward	ATGCCAGGARTNGGNAARACNAC	MGGP/VGKTT (P-loop)	this paper
	reverse	ACCTCAAGAGCNARNGGNARNCC	GLPLAL (hydrophobic domain)	
NBS-2	forward	GGIGGIGTIGGIAAIACIAC	GGVGKT (P-loop)	(KANAZIN et al., 1996)
	reverse	ARIGCTARIGGIARICC	GLPLAL (hydrophobic domain)	
NBS-3	forward	GGIGGIGTIGGIAAIACIAC	GGV/IGKTT (P-loop)	(LEISTER et al., 1996)
	reverse	IAGIGCIAGIGGIAGICC	GLPLAL (hydrophobic domain)	
NBS-4	forward	GGWATGGGWWGWRTHGGWAARACHAC	GMGGVGKT (P-loop)	(LEE et al., 2003)
	reverse	ARNWYYTTVARDGCVARWGGVARWCC	GLPLAL (hydrophobic domain)	
NBS-5	forward	GGIGGIGTIGGIAAIACIAC	GGVGKT (P-loop)	(HATTENDORF and DEBENER, 2007)
	reverse	ARIGCTARIGGIARICC	GLPLAL (hydrophobic domain)	
NBS-6	forward	GGNATGGGNGGNGTNGGNAARACNAC	GMGGVGKT (P-loop)	(ZHANG et al., 2008)
	reverse	NACYTTNAGNGCNAGNGGNAGNCC	GLPLALKV (hydrophobic domain)	

two highly conserved motifs in the NBS domain of products of tobacco *N* (U15605), *Arabidopsis RPS2* (U12860) and *RPS5* (AF074916), *flax L6* (U27081) and rice *Xa1* (AB002266) genes or obtained from previously published articles (Table 1).

PCR reaction was performed with 200 ng genomic DNA as template. Each 25 µl reaction volume consisted of the following reagents: 2.5 µl of 10xTaq buffer (Takara, China), 0.5 µl of 10mM dNTPs (Sigma-Aldrich, America), 1 µl of 10 µM each primer, 0.5U of Taq DNA polymerase (Takara, China). PCR amplifications were performed in a thermal cycler (MG 5331, Eppendorf, Germany) using the following procedures: after denaturation at 94°C for 2 min, amplification was performed for 35 cycles at 94°C for 0.5 min, at 58°C for 1 min, and at 72°C for 1 min, and a final step of elongation at 72°C for 10 min.

Cloning and sequencing of amplified RGA

PCR products were separated on a 2% agarose gel, and DNA fragments of the appropriate size were extracted from the gel using a PCR Purification Kit (Sangon, China). The target bands were ligated to the pMD18 T-vector (Takara, China) and then transferred into *Escherichia coli* strain DH5α competent cells by heat shock (SAMBROOK et al., 1989). Individual clones were distinguished by 4-base restriction enzyme digestion, such as *Sau3A* and *Taq I*. Clones showing the same restriction patterns were presumed to be identical and therefore sorted into the same class. One to four clones from each defined class were sequenced by Invitrogen (America). The RGA nucleotide sequences with identity of less than 99% were deposited in the GenBank under the Accession Nos. from DQ104366 to DQ104388, from EU816423 to EU816434, DQ452616 and DQ452617.

Database search

Nucleotide and deduced amino acid sequences derived from positive clones were searched for their similarity to cloned R genes and their corresponding loci using the BLAST algorithm (ALTSCHUL et al., 1997) against the

databases of the NCBI GenBank (<http://www.ncbi.nlm.nih.gov/>), the Swedish *Populus* EST collection (<http://www.populus.db.umu.se/>) and the genome sequence of *P. trichocarpa* (<http://genome.jgi-psf.org/Poptr1/>).

Multiple alignments and phylogenetic analysis

Phylogenetic analyses were performed on the European aspen RGA nucleotide sequences as well as the deduced amino acid sequences. Searches for introns were done using the NetPlantGene (<http://www.cbs.dtu.dk/services/NetPGene/>). Predicted amino acid sequences were generated using the ExPASy (Expert Protein Analysis System) (<http://www.expasy.org/>). Pairwise comparisons and multiple alignments were performed using the Clustal W 1.8 with default parameters (THOMPSON et al., 1997). A dissimilarity matrix and a phylogenetic tree were obtained from this alignment following the neighbor-joining method (SAITOU and NEI, 1987) with the Kimura's two-parameter distances for nucleotide sequences and poisson correction distance for deduced amino acid sequences (NEI and KUMAR, 2000) in MEGA 4.1 (TAMURA et al., 2007). The reliability of the tree was established by conducting 1000 neighbor-joining bootstrap sampling steps. The phylogenetic tree was constructed using an *AfsR* gene (P25941) as an outgroup (MEYERS et al., 2003). The RGA sequences in the TIR sub-group were aligned using the Clustal W 1.8 (THOMPSON et al., 1997), and the multiple sequence alignment editor program GeneDoc (<http://www.nrbsc.org/gfx/genedoc/>) was used to generate a consensus sequence from the sub-group alignments.

Detection of nucleotide substitution and gene conversion

The number of synonymous (d_s) and nonsynonymous (d_n) substitutions were calculated using the modified Nei and Gojobori method in MEGA 4.1 (TAMURA et al., 2007). The one-tailed Z-test and Fisher's exact test were used to detect deviation of the d_n/d_s ratios from neutrality ($d_n/d_s = 1$) (TAMURA et al., 2007). The gene conversion was detected using the Geneconv (SAWYER, 1989). Global and pairwise P values were calculated to assess the sta-

tistical significance of the observed fragment lengths. The P values for global fragments were based on 10,000 permutations of the sequence alignments, while pairwise fragment P values were achieved as described by Karlin and Altschul (KARLIN and ALTSCHUL, 1990). Only intergenic exchange events maintaining global Bonferroni corrected P values < 0.05 were considered as significant.

Analysis of conserved motif structures

Nucleotide diversity, π (NEI and KUMAR, 2000), of the RGA sequences in European aspen genome was calculated in shifting windows of 30 nucleotides with a step size of 2, considering all sequences as haplotypes of the same locus. This parameter was calculated using the DnaSP 4.0 (ROZAS et al., 2003). In order to investigate the structural divergence among the NBS RGA, their predicted amino acid sequences were subjected to domain and motif analyses. The NBS regions were analyzed individually using the MEME/MAST system (<http://meme.sdsc.edu/meme/website/>) to investigate protein motifs in more detail.

Results

Isolation and Identification of RGA sequences from *P. tremula*

Six degenerate primer combinations designed based on the P-loop and hydrophobic domains were used to obtain RGA from European aspen genome. Only a clear band of about 520bp was detected in the amplification products (Fig. 1). This was similar to the size expected based on the size of the targeted NBS domain from other plants (KANAZIN et al., 1996; LEISTER et al., 1996; SORIANO et al., 2005; GLYNN et al., 2008). The 520bp band was excised and cloned, and 230 individual clones were obtained and classified by digestion with 4-base restriction enzymes. One to four clones of every class were selected and sequenced. In total, 50 clones were sequenced, of which 37 clones showed similarity to known plant R genes or RGAs, while the other 13 clones presented no resemblance. Sequences analysis revealed that 26 European aspen RGAs contained a single uninterrupted open reading frame (ORF). The deduced RGA amino acid sequences showed not only the two motifs targeted by the primers used but also the inter-

nal motifs characteristic of the NBS-LRR gene class (MEYERS et al., 1999; PAN et al., 2000) (Fig. 5).

Alignment analysis

Searches of GenBank using BLAST algorithms (BLASTN and BLASTP) demonstrated that European aspen RGAs shared significant homologies with well-characterized R genes from other plants, as indicated by the E values ranging from $1.0e-13$ to $6.0e-39$. Comparison of the deduced amino acid sequence of European aspen RGA sequences with the NBS domain encoded by known plant R genes, such as tobacco *N* (U15605), flax *M* and *L6* (U73916 and U27081), and *A. thaliana* *RPP1*, *RPP5* and *RPP8* (AF098962, AF180942 and AF089710), showed identities ranging from 30.41% to 46.63%.

To estimate the genomic distribution and expression of NBS encoding R genes in *Populus* genome, the BLAST analysis against the genomic and transcript sequences of *P. trichocarpa* was performed. The results showed a total of 209 hits from the *Populus* genome sequences, with length over 450bp and identity of more than 90% (Appendix S1 in Supplementary material). These sequences were distributed in 28 loci, including 21 scaffolds and 7 linkage groups (LG). Similarly, search of *P. trichocarpa* transcript database also demonstrated 162 hits (Appendix S2 in Supplementary material), using the same parameter. Within each matching region in the poplar genome, at least one NBS-encoding disease resistance like gene was identified using the Genscan (data not shown).

Nucleotide divergence analysis

Divergence analysis of 37 NBS-RGAs, isolated from European aspen, was conducted using the neighbor-joining method (SAITOU and NEI, 1987) and a phylogenetic tree was constructed (Fig. 2). The RGA sequences were grouped into seven clusters or families (A-H), all of which were highly supported by bootstrap values. There were three members in cluster A and F, one member in cluster B, eight in cluster C and D, ten in cluster E and four in cluster G, respectively (Fig. 2). The results showed a high degree of divergence and considerable sequence variation among RGAs in *P. tremula* genome, the divergent distance for members of the European aspen RGA sequences ranged between 0.02 and 1.58 (Appendix S3 in Supplementary material). In accordance with this, the average d_s for *P. tremula* was 0.463, confirming the high divergence for European aspen RGA sequences. A maximum parsimony analysis produced a similar tree (data not shown), confirming the robustness of the tree.

The analysis of the nucleotide polymorphism and divergence of the European aspen RGA sequences showed that they were less polymorphic and highly conserved at the functionally defined NBS motifs by MEYERS et al. (1999) and PAN et al. (2000), such as the P-loop, RNBS-A-TIR, Kin-2, RNBS-B, RNBS-C and GLPL motifs (Fig. 3). The other more conserved region identified in *P. tremula* sequences was named RNBS-*P. tremula* motif (Fig. 3). This high special conservation at the RNBS-*P. tremula* motif might be, at least partial-

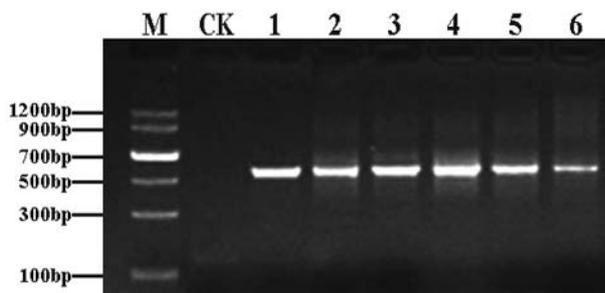


Figure 1. – Resistance gene analogs obtained from *P. tremula* with the degenerate primers: the DNA molecular-weight marker lane M; the negative control without DNA template (lane CK); using different primer combinations in table 1 represented in lanes 1 to 6.

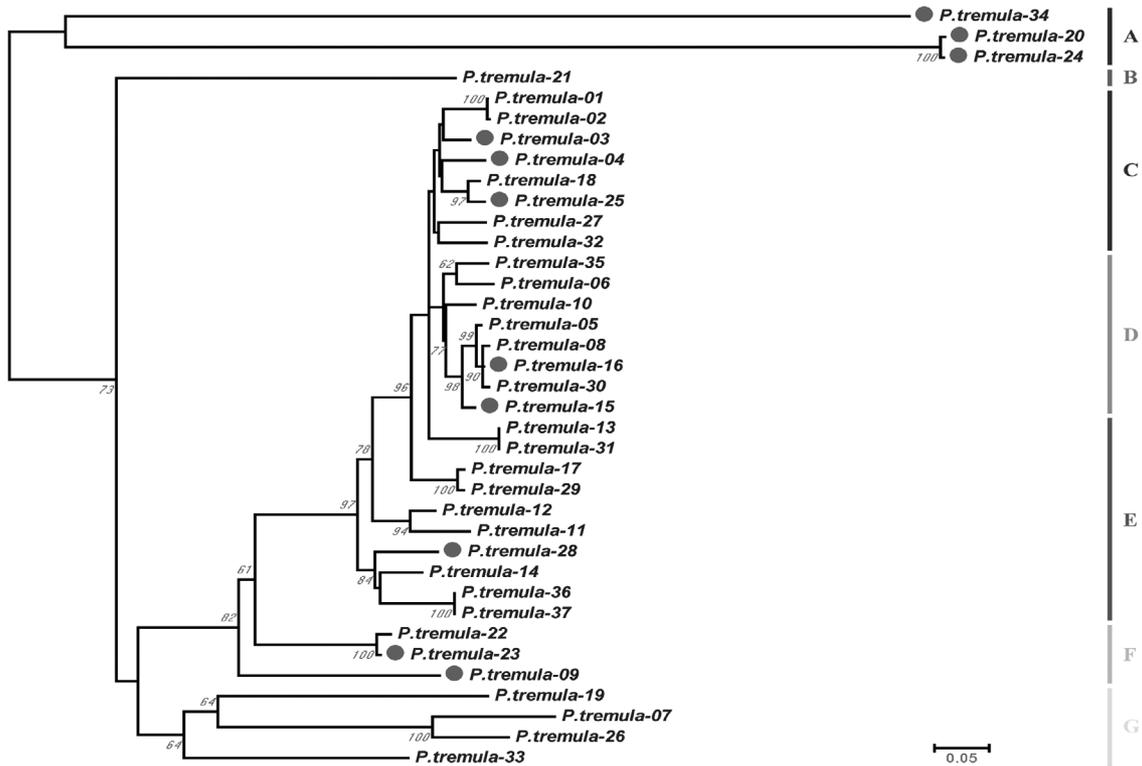


Figure 2. – Phylogenetic consensus tree for RGA nucleotide sequences isolated from the *P. tremula*. Seven RGA families (A-G) were obtained using a 0.150 divergence threshold value. Numbers (%) on the main branches represented bootstrap values (for 1000 iterations) and only bootstrap values of more than 60 were indicated. The red dot (●) indicates the sequences that do not have complete ORF and were not involved in the follow analysis in this paper. The scale bar shows a distance equal to 5% nucleotide divergence.

ly, because of the species-specific requirement for the disease resistance.

Phylogenetic analysis of NBS domain RGA from *P. tremula* with other cloned R genes

The Neighbor-Joining phylogenetic tree constructed from the deduced amino acid alignment of NBS domains of known plant R genes and European aspen RGAs was shown in Fig. 4. The tree had long branch lengths and closely clustered nodes, reflecting a high level of sequence divergence. The distinction between the different European aspen RGA families was clearly confirmed, which were consistent with the nucleotide analysis observed in the relationships between European aspen RGA families (Fig. 2). When the *AfsR* gene (P25941) was used as an outgroup, the phylogenetic tree showed that they could be split into two major branches with strong supports of bootstrap analysis (Fig. 4). With a bootstrap value of 94, the phylogram revealed clear separation between the TIR and non-TIR classes. Non-TIR RGA sequences were derived from monocotyledonous and dicotyledonous plants, whereas TIR RGA sequences exclusively came from dicots. This pattern indicates the divergence during evolution and an ancient origin for these RGA sequences (MEYERS et al., 1999; PAN et al., 2000). All European aspen RGAs identified in the current study were present in one branch containing the *N*, *L6*, *M*, *RPP1* and *RPP5* genes. According to the previously defined distinction between the

TIR and non-TIR class, all isolated European aspen RGAs seemed to belong to the TIR-NBS-LRR subfamily. Among the TIR group, two sub-clusters emerged, 26 European aspen sequences joining the TIR-B and the other plant R genes joining the TIR-A.

Comparative analysis of conserved motifs in NBS domain

Alignments of deduced amino acid sequences demonstrated that the European aspen RGA also contained the conserved motifs in NBS domains that were previously identified in R genes of other plants (MEYERS et al., 1999; PAN et al., 2000), although in some sequences with minor amino acid substitutions (Fig. 5). For example, the most common P-loop motif in European aspen was MPGVGKTT, nevertheless, a variation, MPG(I/M)GKTT, was also found, and this variant motif could represent the divergence of RGA sequences in *P. tremula* genome (Fig. 5). The non-TIR sequences had the characteristic tryptophan residue (W) at the final position of the kinase-2 motif that was absent in the TIR sequences (MEYERS et al., 1999; PAN et al., 2000). Instead of the tryptophan residue (W), the kinase-2 motifs found in European aspen NBS sequences had aspartic acid (D), alanine (A), serine (S), or threonine (T) residues at this last position (Fig. 5). This sequence feature was consistent with the result of phylogenetic analysis, revealed above in this paper. In accordance with PAN et al. (PAN et al., 2000), short conserved amino-acids were also

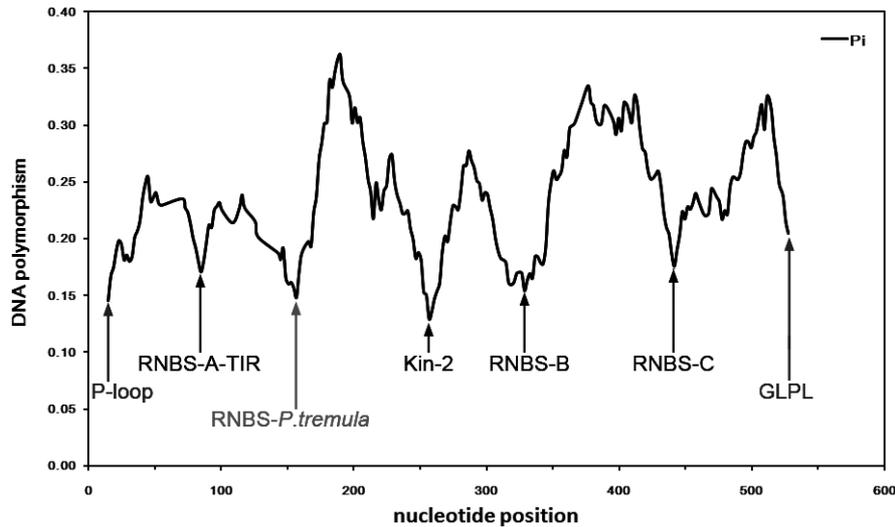


Figure 3. – Nucleotide polymorphism of *P. tremula* RGA sequences. Lower polymorphism rates indicated the conserved regions of NBS.

detected. In NBS-II (RNBS-A), a phenylalanine (F) was invariably present in both TIR and non-TIR groups and, 5 residues downstream, a phenylalanine (F) or a tryptophan (W) residue was highly conserved in the TIR and non-TIR group, respectively.

Source of variation and evolution of the *P. tremula* RGA sequences

In order to explore the evolutionary forces acting on European aspen RGA sequences, the d_N/d_S ratio in this NBS domain was calculated. The d_N/d_S ratio reveals whether sequence polymorphism supports either purifying selection ($d_N/d_S < 1$) or diversifying selection ($d_N/d_S > 1$) (NEI and GOJOBORI, 1986). The d_S and d_N values among nucleotide sequences encoding European aspen RGAs ranged from 0.028 to 1.244 and 0.003 to 0.653, respectively, and the respective d_N/d_S values ranged from 0.188 to 1.372 with the most values below 1.000 and the average value as 0.597 (Appendix S4 in Supplementary material).

Moreover, we tested the number of potential recombination events among these sequences using the Geneconv (SAWYER, 1989) which searched for the most likely candidates for gene conversion events between pairs of aligned sequences within and outside an alignment to reveal the contributions of genetic recombination for the evolution of European aspen RGA. Our results detected many significant genetic recombination events ($P < 0.05$), including 61 inner fragment and 9 outer fragments, suggesting that genetic recombination played a significant role in homogenising the RGA in European aspen genome.

Discussion

PCR amplification with degenerate primers targeted to the short conserved regions in the NBS domain has been proven to be an efficient method for identifying

RGAs from a wide variety of plant species (AMELINE-TORREGROSA et al., 2008; KANAZIN et al., 1996; KUANG et al., 2004; LEISTER et al., 1996; MILLER et al., 2008; YANG et al., 2008). In this study, using a similar approach, we have successfully identified 37 RGA sequences from European aspen genome. Searches of GenBank using BLAST demonstrated that European aspen RGAs shared significant homologies with well-characterized R genes from other plants, such as tobacco *N* (U15605), flax *M* and *L6* (U73916 and U27081), rice *Xa1* (AB002266), and *A. thaliana* *RPP1* (AF098962), *RPP5* (AF180942) and *RPP8* (AF089710). Divergence analysis results showed the considerable nucleotide acid sequence variation among RGAs in European aspen genome (Fig. 2). The BLAST searches also showed a total of 209 and 162 matching hits from the *Populus* genome and transcript sequences, respectively, which indicated that the RGA sequences are abundant in the genome of *Populus* species. The amino acid alignments showed that all of the crucial motifs of the NBS domain were highly conserved among European aspen RGAs and the known plant NBS-LRR R proteins (Fig. 5).

The last residue of kinase-2 motif of identified RGAs can be used to predict with 95% accuracy whether they belong to the TIR or non-TIR subclass of NBS-LRR R genes (MEYERS et al., 1999; PAN et al., 2000). Using this criterion, it was observed that all European aspen RGA sequences amplified by the six primers (Table 1) belonged to the TIR NBS-LRR subclass (Fig. 5). In accordance with PAN et al. (2000), short conserved amino-acid were also detected in NBS-II (RNBS-A). Based on these differences, the known non-TIR R proteins clustered into one group in the amino acid sequences phylogenetic tree, whereas all of 26 European aspen RGA sequences together with the known plant TIR R proteins formed a second group (Fig. 4). Interestingly, the 26 TIR European aspen RGA sequences were absolutely clustered into one subclass (TIR-B) (Fig. 4),

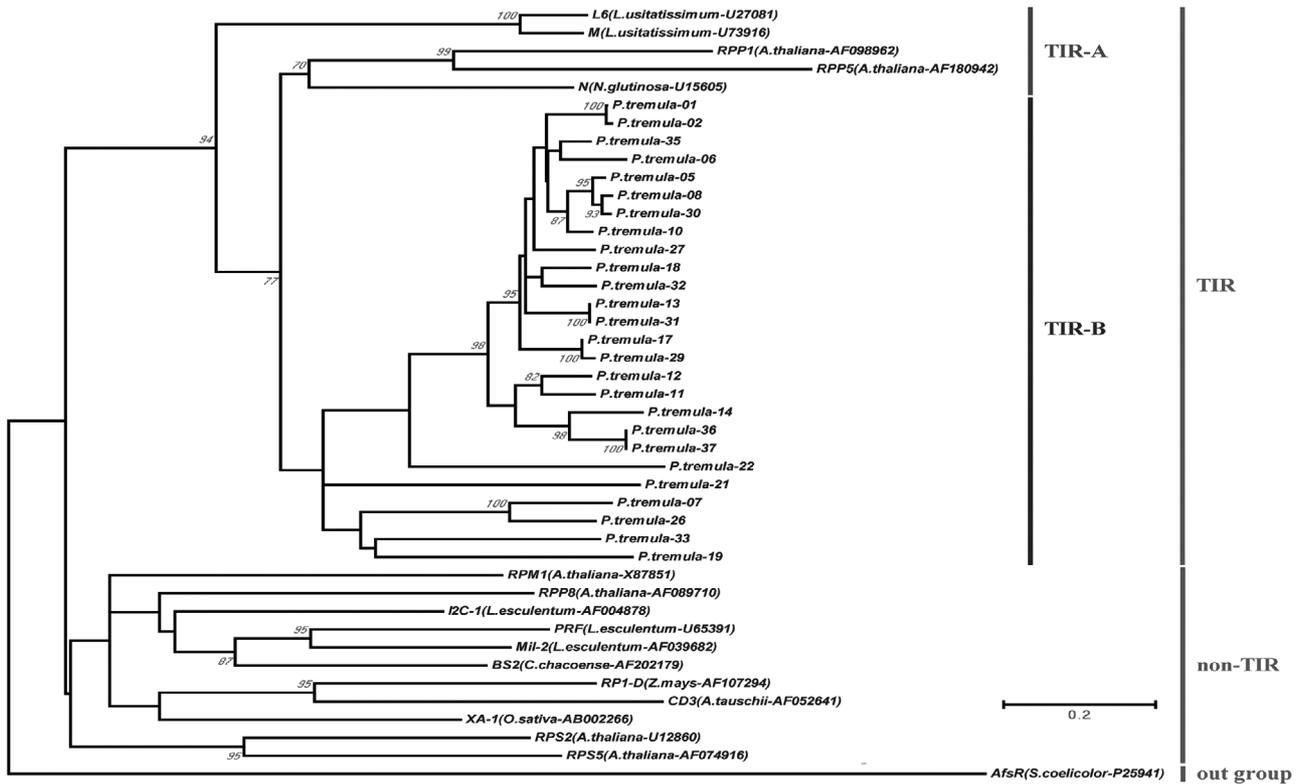


Figure 4. – Phylogenetic tree based on the alignment of the consensus amino acid sequences of *P. tremula* RGAs and the NBS domain of sixteen known plant R genes. Numbers (%) on the main branches represent bootstrap values (for 1000 iterations) and only bootstrap values of more than 60 were indicated. The scale bar in the bottom left corner displays a distance corresponding to 0.2 amino acid substitutions per site.

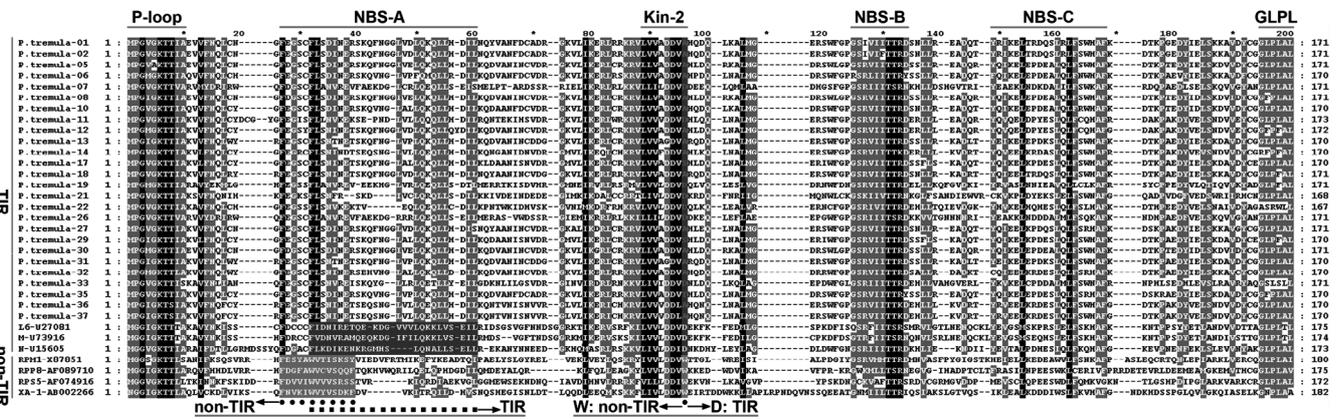


Figure 5. – Alignment of deduced amino-acid sequences of *P. tremula* RGAs, along with the NBS domain of known R genes from other plants. Consensus residues above 60% are shown by shading in grey. Gaps are indicated as dots. The six previously reported conserved motifs of the NBS domain (P-loop, NBS-A, Kinase2, NBS-B, NBS-C and GLPL) are boxed in black lines.

in which no known TIR R proteins were included, suggesting that the subclass TIR-B is probably a specific class for perennial species like European aspen.

The reported ratio of non-TIR:TIR-type NBS sequences in the complete *Arabidopsis* genome is 1:3 (MEYERS et al., 1999; MEYERS et al., 2003). However, this ratio can vary among different species. KANAZIN et al. (1996) in soybean and MARTINEZ ZAMORA et al. (2004) in strawberry identified mainly TIR-subfamily sequences,

but DONALD et al. (2002) found a high proportion of non-TIR:TIR sequences (6:1) in grapevine. In this study, the NBS encoding RGAs from European aspen showed considerable sequence variation (Fig. 2), however, all European aspen RGA sequences isolated fall into the TIR NBS-LRR resistance gene subfamily. One possible explanation for the absence of non-TIR RGA sequences in European aspen could be the loss of this type of sequence during evolution. It has been reported that the

absolute loss of TIR or non-TIR type NBS-LRR RGA sequences occurred both in monocot (BAI et al., 2002; GLYNN et al., 2008; IRIGOYEN et al., 2006; MILLER et al., 2008) and dicot plants (SORIANO et al., 2005; TIAN et al., 2004), but no mechanism has been described that can explain this kind of elimination. This finding is consistent with the assumption that different NBS-LRR genes could also have been lost from other dicot plant species (SORIANO et al., 2005; TIAN et al., 2004). Another interesting phenomenon is that the ratio of non-TIR:TIR type NBS sequences display a great variation among the different *Populus* species. For instance, the non-TIR RGA sequences were predominant in black cottonwood (*P. trichocarpa*) (KÖHLER et al., 2008; TUSKAN et al., 2006), which is in contrast to white poplar (*P. tomentosa* × *P. bolleana*) × *P. tomentosa* where the TIR sequences were predominant (ZHANG et al., 2008). The possible reason is that the black cottonwood was more dependent on non-TIR resistance genes for defense against pathogens and this resulted in their numbers amplified in the evolutionary course, conversely, the white poplar or European aspen was possibly dependent on TIR resistance genes in its evolutionary history and their numbers amplified during the evolutionary course. These data indicated different evolutionary history for the RGA superfamily in different species of *Populus*. PAN et al. (2000) proposed a hypothesis that at least two stages are involved in the evolution of NBS-LRR R genes in the plant kingdom: the presence of a few NBS-LRR with broad spectrum specificity (Stage I) and the presence of R genes shaped by divergent gene duplication followed by gene diversification after the monocot/dicot separation (Stage II). The diversity of distribution for TIR or non-TIR RGA sequences in different species within *Populus* supports the idea that sequences were indeed lost from those ancestral loci in plant genomes, regardless of the type of species.

The evolution of R gene remains largely unexplored, but useful information has recently been gained from molecular genetic analyses. Several genetic mechanisms, including point mutation, recombination, unequal crossing-over, and/or gene conversion, have been proposed to account for the evolution of R gene (AMELINE-TORREGROSA et al., 2008; ASHFIELD et al., 2004; KUANG et al., 2004; MICHELMORE and MEYERS, 1998). These analyses refer to crop and model plants with short life cycles, and so far little data have been obtained from perennial plants like European aspen. In this paper, the d_N/d_S ratio of European aspen RGA sequences ranged from 0.188 to 1.372 with the most values below 1.000 and the average value as 0.597, reflecting that these sequences were subject to purifying selection ($d_N/d_S < 1$) rather than diversifying selection ($d_N/d_S > 1$) and the purifying selection dominated the evolution of European aspen RGAs. This conclusion is consistent with the proposed function of these domains, and purifying selection was generally agreed to be the predominant form of natural selection involved in the patterning of most proteins (BUSTAMANTE et al., 2000). For long-lived species such as European aspen, with a long life span, the generation of new resistances via somatic mutations may be particularly advantageous

(MEYERS et al., 2005; MICHELMORE and MEYERS, 1998). Comparisons of variability in the R gene domains that determine pathogen specificity among long-lived individuals need to be performed. In conclusion, the evolution of NBS encoding sequences in European aspen appears to involve the gradual accumulation of mutations, and thus reflects the action of slowly acting divergence mechanisms within distinct R gene families, rather than a rapidly evolving process. As suggested by MEYERS (2005) and MICHELMORE and MEYERS (1998), R gene might not be evolving rapidly in order to keep pace with changes in the pathogen, but is rather evolving fairly slowly to provide resistance against pathogen populations that are heterogeneous in space and time. This could be particularly relevant in the case of perennial plants like European aspen.

Sequence analysis of different loss-of-function alleles in the R-gens *RPM1*, *RPS2* and *Prf* demonstrated the functional requirements underlying the conservation of the consensus motifs characterizing the NBS region of plant R genes (PAN et al., 2000). Specific signatures were revealed by comparing European aspen RGAs with the different closely related R genes isolated from monocot and/or dicot species that are distantly related to European aspen. Within families characterized by NBS sequences of hypothetically common origin, consensus motifs appeared extended and short additional conserved blocks were observed (Fig. 3, 5). Since these genes or RGAs result from an ancient divergence, this conservation suggests a functional family specificity. It is possible that some of the differences between the NBS sequences encoded by R genes influence the specificity of interactions either with a pathogen elicitor or with proteins downstream in the signal transduction cascade (DEYOUNG and INNES, 2006; MEYERS et al., 2005).

Further genetic and biochemical analyses are required in order to better define the structural and functional roles of particular motifs and the divergence within the NBS domain. Elucidating the evolution of R genes is of particular importance for the understanding of how plants maintain and adapt their defences to pathogens (PAN et al., 2000). Since highly contrasting situations hold for different plants, investigations of various species, including tree-species, appear necessary, in addition to studies on model plants such as *Arabidopsis* and rice. In this endeavour, European aspen could constitute an attractive model for perennial plants.

Acknowledgements

The authors would like to thank anonymous reviewers and editors for critical and constructive suggestions to improve this manuscript. This work was supported by the National Basic Research Program of China (Grant No. 2009CB119100), the National Natural Science Foundation of China (Grant No. 30571517) and the Youth Foundation of University of Electronic Science and Technology of China. The *Populus tremula* sample were morphological identified and kindly provided by Professor Qi in the Research Institute of Forestry, Chinese Academy of Forestry.

References

- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG, W. MILLER and D. J. LIPMAN (1997): Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389–402.
- AMELINE-TORREGROSA, C., B. B. WANG, M. S. O'BLENESS, S. DESHPANDE, H. ZHU, B. ROE, N. D. YOUNG and S. B. CANNON (2008): Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol* **146**(1): 5–21.
- ASHFIELD, T., L. E. ONG, K. NOBUTA, C. M. SCHNEIDER and R. W. INNES (2004): Convergent evolution of disease resistance gene specificity in two flowering plant families. *Plant Cell* **16**(2): 309–18.
- BAI, J., L. A. PENNILL, J. NING, S. W. LEE, J. RAMALINGAM, C. A. WEBB, B. ZHAO, Q. SUN, J. C. NELSON, J. E. LEACH and S. H. HULBERT (2002): Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res* **12**(12): 1871–84.
- BUSTAMANTE, C. D., J. P. TOWNSEND and D. L. HARTL (2000): Solvent accessibility and purifying selection within proteins of *Escherichia coli* and *Salmonella enterica*. *Mol Biol Evol* **17**(2): 301–8.
- DEYOUNG, B. J. and R. W. INNES (2006): Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat Immunol* **7**(12): 1243–9.
- DONALD, T. M., F. PELLERONE, A. F. ADAM-BLONDON, A. BOUQUET, M. R. THOMAS and I. B. DRY (2002): Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevine. *Theor Appl Genet* **104**(4): 610–618.
- FLOR, H. H. (1971): Current status of the gene-for-gene concept. *Annu Rev Phytopathol* **9**: 275–296.
- GLYNN, N. C., J. C. COMSTOCK, S. G. SOOD, P. M. DANG and J. X. CHAPARRO (2008): Isolation of nucleotide binding site-leucine rich repeat and kinase resistance gene analogues from sugarcane (*Saccharum* spp.). *Pest Manag Sci* **64**(1): 48–56.
- HATTENDORF, A. and T. DEBENER (2007): Molecular characterization of NBS-LRR-RGAs in the rose genome. *Physiol Plant* **129**: 775–786.
- IRIGOYEN, M. L., E. FERRER and Y. LOARCE (2006): Cloning and characterization of resistance gene analogs from *Avena* species. *Genome* **49**(1): 54–63.
- KANAZIN, V., L. F. MAREK and R. C. SHOEMAKER (1996): Resistance gene analogs are conserved and clustered in soybean. *Proc Natl Acad Sci USA* **93**(21): 11746–50.
- KARLIN, S. and S. F. ALTSCHUL (1990): Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc Natl Acad Sci USA* **87**(6): 2264–8.
- KOHLER, A., C. RINALDI, S. DUPLESSIS, M. BAUCHER, D. GELEN, F. DUCHAUSOY, B. C. MEYERS, W. BOERJAN and F. MARTIN (2008): Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol Biol* **66**(6): 619–36.
- KUANG, H., S. S. WOO, B. C. MEYERS, E. NEVO and R. W. MICHELMORE (2004): Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **16**(11): 2870–94.
- KUANG, H., K. S. CALDWELL, B. C. MEYERS and R. W. MICHELMORE (2008): Frequent sequence exchanges between homologs of RPP8 in *Arabidopsis* are not necessarily associated with genomic proximity. *Plant J* **54**(1): 69–80.
- LEE, S. Y., J. S. SEO, M. RODRIGUEZ-LANETTY and D. H. LEE (2003): Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species. *Mol Genet Genomics* **269**(1): 101–8.
- LEISTER, D., A. BALLVORA, F. SALAMINI and C. GEBHARDT (1996): A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nat Genet* **14**(4): 421–9.
- LESCOT, M., S. ROMBAUTS, J. ZHANG, S. AUBOURG, C. MATHE, S. JANSSON, P. ROUZE and W. BOERJAN (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. *Theor Appl Genet* **109**(1): 10–22.
- MARTINEZ ZAMORA, M. G., A. P. CASTAGNARO and J. C. DIAZ RICCI (2004): Isolation and diversity analysis of resistance gene analogues (RGAs) from cultivated and wild strawberries. *Mol Genet Genomics* **272**(4): 480–7.
- MEYERS, B. C., A. W. DICKERMAN, R. W. MICHELMORE, S. SIVARAMAKRISHNAN, B. W. SOBRAL and N. D. YOUNG (1999): Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J* **20**(3): 317–32.
- MEYERS, B. C., A. KOZIK, A. GRIEGO, H. KUANG and R. W. MICHELMORE (2003): Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**(4): 809–34.
- MEYERS, B. C., S. KAUSHIK and R. S. NANDETY (2005): Evolving disease resistance genes. *Curr Opin Plant Biol* **8**(2): 129–34.
- MICHELMORE, R. W. and B. C. MEYERS (1998): Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res* **8**(11): 1113–30.
- MILLER, R. N., D. J. BERTIOLI, F. C. BAURENS, C. M. SANTOS, P. C. ALVES, N. F. MARTINS, R. C. TOGAWA, M. T. SOUZA, JR. and G. J. PAPPAS, JR. (2008): Analysis of non-TIR NBS-LRR resistance gene analogs in *Musa acuminata* Colla: isolation, RFLP marker development, and physical mapping. *BMC Plant Biol* **8**: 15.
- MURRAY, M. G. and W. F. THOMPSON (1980): Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**(19): 4321–5.
- NEI, M. and T. GOJOBORI (1986): Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**(5): 418–26.
- PAN, Q., J. WENDEL and R. FLUHR (2000): Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J Mol Evol* **50**(3): 203–13.
- ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS (2003): DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**(18): 2496–7.
- SAITOU, N. and M. NEI (1987): The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**(4): 406–25.
- SAWYER, S. (1989): Statistical tests for detecting gene conversion. *Mol Biol Evol* **6**(5): 526–38.
- SORIANO, J. M., S. VILANOVA, C. ROMERO, G. LLACER and M. L. BADENES (2005): Characterization and mapping of NBS-LRR resistance gene analogs in apricot (*Prunus armeniaca* L.). *Theor Appl Genet* **110**(5): 980–9.

- TAMURA, K., J. DUDLEY, M. NEI and S. KUMAR (2007): MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**(8): 1596–9.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS (1997): The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**(24): 4876–82.
- TIAN, Y., L. FAN, T. THURAU, C. JUNG and D. CAI (2004): The absence of TIR-type resistance gene analogues in the sugar beet (*Beta vulgaris* L.) genome. *J Mol Evol* **58**(1): 40–53.
- TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV, U. HELLSTEN, N. PUTNAM, S. RALPH, S. ROMBAUTS, A. SALAMOV, J. SCHEIN, L. STERCK, A. AERTS, R. R. BHALERAO, R. P. BHALERAO, D. BLAUDEZ, W. BOERJAN, A. BRUN, A. BRUNNER, V. BUSOV, M. CAMPBELL, J. CARLSON, M. CHALOT, J. CHAPMAN, G. L. CHEN, D. COOPER, P. M. COUTINHO, J. COUTURIER, S. COVERT, Q. CRONK, R. CUNNINGHAM, J. DAVIS, S. DEGROEVE, A. DEJARDIN, C. DEPAMPHILIS, J. DETTER, B. DIRKS, I. DUBCHAK, S. DUPLESSIS, J. EHLTING, B. ELLIS, K. GENDLER, D. GOODSTEIN, M. GRIBSKOV, J. GRIMWOOD, A. GROOVER, L. GUNTER, B. HAMBERGER, B. HEINZE, Y. HELARIUTTA, B. HENRISSAT, D. HOLLIGAN, R. HOLT, W. HUANG, N. ISLAM-FARIDI, S. JONES, M. JONES-RHOADES, R. JORGENSEN, C. JOSHI, J. KANGASJARVI, J. KARLSSON, C. KELLEHER, R. KIRKPATRICK, M. KIRST, A. KOHLER, U. KALLURI, F. LARIMER, J. LEEBENS-MACK, J. C. LEPLE, P. LOCASCIO, Y. LOU, S. LUCAS, F. MARTIN, B. MONTANINI, C. NAPOLI, D. R. NELSON, C. NELSON, K. NIEMINEN, O. NILSSON, V. PEREDA, G. PETER, R. PHILIPPE, G. PILATE, A. POLIAKOV, J. RAZUMOVSKAYA, P. RICHARDSON, C. RINALDI, K. RITLAND, P. ROUZE, D. RYABOV, J. SCHMUTZ, J. SCHRADER, B. SEGERMAN, H. SHIN, A. SIDDIQUI, F. STERKY, A. TERRY, C. J. TSAI, E. UBERBACHER, P. UNNEBERG, J. VAHALA, K. WALL, S. WESSLER, G. YANG, T. YIN, C. DOUGLAS, M. MARRA, G. SANDBERG, Y. VAN DE PEER and D. ROKHSAR (2006): The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**(5793): 1596–604.
- VAN DER HOORN, R. A. and S. KAMOUN (2008): From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. *Plant Cell*.
- YANG, S., T. GU, C. PAN, Z. FENG, J. DING, Y. HANG, J. Q. CHEN and D. TIAN (2008): Genetic variation of NBS-LRR class resistance genes in rice lines. *Theor Appl Genet* **116**(2): 165–77.
- ZHANG, Q., Z. Y. ZHANG, S. Z. LIN, H. Q. ZHENG, Y. Z. LIN, X. M. AN, Y. LI and H. X. LI (2008): Characterization of resistance gene analogs with a nucleotide binding site isolated from a triploid white poplar. *Plant Biol (Stuttg)* **10**(3): 310–22.
- ZHANG, Y., S. G. ZHANG, L. W. QI, B. LIU, J. M. GAO, C. B. CHEN, X. L. LI and W. Q. SONG (2005): Chromosome microdissection, cloning and painting of the chromosome 1 in poplar (*Populus tremula*). *Silvae Genetica* **54**: 211–217.

Bio-economic Modelling as a Method for Determining Economic Weights for Optimal Multiple-Trait Tree Selection

By M. IVKOVIĆ^{a,c}, H. WU^a and S. KUMAR^b

(Received 6th April 2009)

Abstract

A bio-economic model provides a framework for simultaneously considering breeding, management, and production decisions. Such a model should result in optimal breeding (and silvicultural) objectives if main goals of a production system are well defined. Historically estimation of economic weights for breeding-objective traits has been based on partial regressions and profit functions relating only to certain parts of the production system. A bio-economic model includes effects of growth rate, branching, form, and wood quality on all produc-

tion system components and on overall profitability of an integrated production system. However, long rotation cycles in forestry make determination of relative economic values for the breeding-objective traits particularly difficult. When modelling complex systems under uncertainty about future production goals, there are necessary trade offs between the complexity of the model and the use of simplifying assumptions.

Key words: tree improvement, breeding objective, bio-economic modelling, production system analysis.

Introduction

Breeding Objectives and Multiple Trait Selection

The first step in the design of a breeding program is to set its objectives. Correctly set breeding objectives will

^a) CSIRO Plant Industry, PO Box E4008, Kingston ACT 2604, Australia.

^b) Scion, Private Bag 3020, Rotorua, New Zealand.

^c) Corresponding author. E-Mail: milosh.ivkovich@csiro.au