Plant Mol Biol (2008) 66:619–636
DOI 10.1007/s11103-008-9293-9

Genome-wide identification of *NBS* resistance genes in *Populus trichocarpa*

Annegret Kohler · Cécile Rinaldi · Sébastien Duplessis · Marie Baucher · Danny Geelen · Frédéric Duchaussoy · Blake C. Meyers · Wout Boerjan · Francis Martin

Received: 2 August 2007 / Accepted: 11 January 2008 / Published online: 5 February 2008

Abstract As the largest class of disease resistance *R* genes, the genes encoding nucleotide binding site and leucine-rich repeat proteins ("*NBS-LRR* genes") play a critical role in defending plants from a multitude of pathogens and pests. The diversity of *NBS-LRR* genes was examined in the *Populus trichocarpa* draft genome sequence. The *NBS* class of genes in this perennial tree is large and diverse, comprised of ~400 genes, at least twice the complement of *Arabidopsis*. The NBS family can be divided into multiple subfamilies with distinct domain organizations. It includes 119 *Coiled-Coil-NBS-LRR* genes, 64 *TIR-NBS-LRR* genes, 34 *BED-finger-NBS-LRR*, and both truncated and unusual *NBS*- and *NBS-LRR*-containing genes. The transcripts of only 34 *NBS-LRR* genes were detected in rust-infected and non-infected leaves using a whole-genome oligoarray. None showed an altered expression two days post inoculation.

Keywords *Populus* · NBS-LRR · Resistance

Electronic supplementary material The online version of this article (doi:10.1007/s11103-008-9293-9) contains supplementary material, which is available to authorized users.

Annegret Kohler and Cécile Rinaldi contributed equally to this work.

A. Kohler (✉) · C. Rinaldi · S. Duplessis · F. Duchaussoy · F. Martin
Unité Mixte de Recherche INRA/UHP 1136 Interactions Arbres/Microorganismes, Institut National de la Recherche Agronomique, Centre de Recherches de Nancy, 54280 Champenoux, France
e-mail: kohler@nancy.inra.fr

M. Baucher
Laboratoire de Biotechnologie Végétale, Université Libre de Bruxelles, 8 rue Adrienne Bolland, 6041 Gosselies, Belgium

D. Geelen
Department of Plant Production, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Gent, Belgium

B. C. Meyers
Department of Plant and Soil Sciences, Delaware Biotechnology Institute, Delaware Technology Park, Newark, DE 19711, USA

W. Boerjan
Department of Plant Systems Biology, Flanders Institute for Biotechnology, 9052 Gent, Belgium

W. Boerjan
Department of Molecular Genetics, Ghent University, 9052 Gent, Belgium

Introduction

Plants have a variety of disease-resistance (*R*) proteins that are involved in the detection of pathogens and pests, and subsequent triggering of defense responses, including localized cell death (the hypersensitive response) and systemic acquired resistance. Based on their predicted domain structure, R proteins encoded by the *R* genes have been classified into different classes: intracellular kinases, extracellular receptors, extracellular receptors coupled to kinases, and intracellular receptors (McDowell and Woffenden 2003). Collectively, they confer resistance to a broad spectrum of pathogens, including viruses, bacteria, fungi and nematodes (Dangl and Jones 2001; McDowell and Woffenden 2003; Meyers et al. 2005). The largest two classes of *R* genes encode putative intracellular receptors (Dangl and Jones 2001), which contain either a coiled-coil (CC) or a Toll/Interleukin-1 receptor (TIR) domain at their N-terminal end, followed by a nucleotide binding site (NBS) having homology to eukaryotic cell death effectors. At the

C-terminal end, these proteins consist of a series of leucine rich repeats (LRRs). The functions of the CC, TIR, and NBS domains are not fully known, but all similar proteins identified in animal systems play roles in protein–protein interactions and signal transduction (Dangl and Jones 2001; McDowell and Woffenden 2003). The function of LRR domains is clearer because recent data suggest that LRRs in *Arabidopsis R* proteins mediate direct or indirect interaction with pathogen molecules (Belkhadir et al. 2004). *NBS-LRR* genes appear to function mainly in defense responses, but for ADR1, an *Arabidopsis* CC-NBS-LRR and for At5g17880, an *Arabidopsis* TIR-NBS-LRR, a role in drought tolerance (Chini et al. 2004) and in photomorphogenic development (Faigon-Soverna et al. 2006) has been shown, respectively.

Despite the wealth of information on the structure and function of *NBS-LRR* genes that has come from several genome-wide analyses in herbaceous monocots and dicots (Meyers et al. 1999; Bai et al. 2002; Mondragón-Palomino et al. 2002; Baumgarten et al. 2003; Meyers et al. 2003; Ayliffe and Lagudah 2004; Zhou et al. 2004, Ameline-Torregrosa et al. 2007), only fragmentary information is available on *NBS-LRR* genes in perennial species (Zhang et al. 2001; Liu and Ekramoddoullah 2003; Lescot et al. 2004; Liu and Ekramoddoullah 2007). 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. These features may have favoured *R* gene family expansion. Here, we characterize the complete set of *R* gene-related NBS-encoding genes in the recently sequenced *Populus trichocarpa* Nisqually-1 genome (Tuskan et al. 2006). This includes cataloguing the conserved domains of all of the NBS-LRR proteins, conducting phylogenetic analyses to investigate their relationships, and surveying their level of transcripts in various tissues and environmental conditions.

Materials and methods

Identification of *NBS-LRR* genes

Using the Blast, Advanced Search and Gene Ontology tools, we identified gene models with a predicted NB-ARC (IPR002182) and disease resistance protein (IPR000767) domains in the draft genome of *P. trichocarpa* (v.1.1) at the JGI *Populus* Genome database (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html). Gene prediction was performed using four methods: GENEWISE, FGENESH, GRAILEXP6, and EUGENE, and gene models were selected by the JGI annotation pipeline (Tuskan et al. 2006). Selection of the models was based on EST support, completeness, and homology to a curated set of proteins. All detected *NBS* gene models were inspected manually, and the automatically selected gene model of the JGI *Populus* Genome database was modified by manual curation if needed. To determine whether they encoded TIR, CC, NBS, or LRR motifs, we used the protein families Pfam (http://pfam.wustl.edu/hmmsearch.shtml) and PRODOM (http://protein.toulouse.inra.fr/prodom/current/html/home.php) databases, as well as the COILS and MacStripe programs (Lupas et al. 1991) (http://www.ch.embnet.org/software/COILS_form.html). The latter programmes were used with a threshold of 0.9 to specifically detect CC domains. For non-ambiguous sequences, pcoils (http://toolkit.tuebingen.mpg.de/pcoils), PAIRCOIL2 (http://paircoil2.csail.mit.edu/paircoil2.html) and MARCOIL1.0 (http://www.isrec.isb-sib.ch/webmarcoil/webmarcoilC1.html) were used in addition. These detailed data on protein motifs and domains were used to classify the *NBS* genes into subgroups (Table 1).

Analysis of conserved motif structures

In order to investigate the structural diversity among the *NBS* genes, their predicted amino acid sequences were subjected to domain and motif analyses. For this purpose, the NBS domain was defined as the region of ~300 amino acids spanning from the P-loop to the MHDV motif (Meyers et al. 2003). The fragments upstream of the P-loop and downstream of the MHDV were designated as the N-terminal motif and the LRR-C-terminal domain, respectively. The NBS, N-terminal, and LRR-C-terminal regions were analyzed individually using the MEME/MAST system (http://meme.sdsc.edu/meme/website/intro.html) (Bailey and Elkan 1995) to investigate protein motifs in more detail. The MEME motif analyses were performed on each subgroup separately (e.g., TNL, CNL, BNL, NL) and with settings designed to identify 15, 20, 30 or 40 motifs. MAST (Bailey and Gribskov 1998) was used to assess correlations between MEME motifs and the distance matrix.

Sequence alignment and phylogenetic tree construction

Selected sequences from the present survey, as well as selected sequences of *Arabidopsis* and rice *NBS* gene sequences were aligned using the programme Clustal X (version 1.83.1) (Jeanmougin et al. 1998) using default settings. The alignment was carried out using either the NBS regions or the whole predicted protein sequences. The aligned sequences were exported to a NEXUS file and analysed using PAUP*4.08b (PPC/Altivec) (Swofford 1999).

Neighbour-joining (NJ) was the primary method to infer phylogenetic relationships between the *NBS* genes, and all final results were compared to Maximum Parsimony (MP)

Table 1 Numbers of genes that encode domains similar to plant *R* proteins in *P. trichocarpa*, *A. thaliana* and *Oryza sativa*

Predicted protein domains[a]	*Populus*	*Arabidopsis*[c]	Rice[d]
Non-TIR-NBS-LRR	224	55	480
CC-NBS-LRR (CNL)	119	51	159
BED[b]*-NBS-LRR (BNL)*	34	–	3
$NBS_{CC/BED}$*-LRR (NL)*	71	4	40
Other *Non-TIR-NBS-LRR*	–	–	278
TIR-NBS-LRR	91	94	–
TIR-NBS-LRR (TNL)	64	83	–
TIR-NBS-LRR-TIR (TNLT)	13	–	–
TIR-NBS-LRR-NBS (TNLN)	1	–	–
NBS-LRR-TIR (NLT)	1	–	–
NBS_{TIR}*-LRR (NL)*	12	2	–
Other *TIR-NBS-LRR*	–	9	–
MIXED NBS-LRR	2	–	–
TIR-CC-NBS-LRR (TCNL)	2	–	–
Total NBS-LRR	317	149	480
Non-TIR-NBS	53	6	52
CC-NBS (CN)	19	4	7
BED[b]*-NBS (BN)*	6	–	–
NBS-BED[b] *(NB)*	1	–	–
NBS_{CC}	27	1	45
Other *Non-TIR-NBS*	–	1	–
TIR-NBS	32	23	3
TIR-NBS	13	21	3
NBS_{TIR}	19	–	–
Other *TIR-NBS*	–	2	–
Total *NBS*	85	29	55
Total	402	178	535

Genes were identified by automated annotation, manual assessment of the genomic sequence, and predicted protein domains

[a] PFAM domains in the predicted proteins. NBS domains from CNL and TNL are distinct

[b] BED finger and/or DUF1544 domain (SMART00614)

[c] Meyers et al. (2003)

[d] Zhou et al. (2004)

(Swofford 1999) to detect any algorithm-specific results. MP trees were identified using heuristic searches based on 1000 random sequence additions, including clades compatible with the 50% majority-rule in the bootstrap consensus tree, using the Nearest-Neighbour Interchange (NNI) branch-swapping option and saving no more than 100 trees with >1,000 length for each replication and MULPART in effect (Swofford 1999). All characters were treated as unordered and with equal weight.

Search for homologs of *Populus NBS-LRR* genes in rice and *Arabidopsis*

The protein sequences for the complete genomes of *Arabidopsis* and rice were taken from the following web-pages:

Arabidopsis thaliana

ftp://ftp.tigr.org/pub/data/a_thaliana/ath1/SEQUENCES;

Oryza sativa

ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_4.0/all_chrs. Formatdb (ftp://ftp.ncbi.nlm.nih.gov/blast/db/) was used to create an *Arabidopsis* and a Rice protein database out of these files. A Unix script was used for the automatic execution of BLASTP of the predicted protein sequences of the 402 *Populus R* genes and the predicted NBS part of these poplar genes against the two databases. The Unix script used, to evaluate similarity between Sequence and Database, the stand-alone NCBI BLAST (v2.2.10) with executables for Macintosh downloaded from ftp://ftp.ncbi.nlm.nih.gov/blast/executables/.

A parsing PERL script was used to retrieve best matches from each output BLAST html file and to generate text files indicating the *Arabidopsis* and rice best match as well as the e-value for this match (Supplemental table 2).

Microarray experiments

All experiments were performed on rooted cuttings of the hybrid poplar *P. trichocarpa* × *P. deltoides* 'Beaupré'. Fully expanded leaves from Leaf Plastochrony Index (LPI) 5 to 9 were detached from several 'Beaupré' plants, grown for 12 weeks in a greenhouse from dormant cuttings, and spray-inoculated on their abaxial surface with an urediniospore suspension of *M. larici-populina* strain 93ID6 in water-agar (0.1 g l^{-1}) adjusted to 100,000 urediniospores.ml^{-1} or with water-agar as a control (mock-inoculated leaves). Total RNA extraction was performed with the RNeasy Plant Mini Kit (Qiagen) from 100 mg of pooled (−80°C) foliar disks coming from leaves of various LPI and various individual Beaupré plants for each treatment considered. The *P. trichocarpa* whole genome expression array (NimbleGen Systems Inc, Madison, WI, USA) consisted of 65,965 probe sets corresponding to 55,970 gene models predicted on the *P. trichocarpa* genome sequence v1.0 and 9,995 aspen cDNA sequences; all probe sets being duplicated on array except for 264 aspen probes (Tuskan et al. 2006; Brunner, DiFazio and Dharmawardhana, unpublished data).

For hybridization, three replicates from mock-inoculated *Populus* leaves (control) as well as from leaves infected with an incompatible strain (93ID6) of *M. larici-populina* at 48 hpi were used. Preparation of Cy3 labelled cDNA samples from total RNA, hybridization procedures, and

data acquisition and normalization were performed at the NimbleGen facility (NimbleGen Systems, Reykjavik, Iceland) following the manufacturer's procedures. Gene annotation and array design files for platform NimbleGen *Populus* 2.0; experimental design; datasets; and gene profiles are available in the GEO database at http://www.ncbi.nlm.nih.gov/geo/ (GEO Accession# GSE7098).

Construction of GFP reporter plasmids and transformation of BY-2 tobacco cells

To construct the *RGLR4* reporter plasmid, the *P. tremula* EST *ub48dpf09* was amplified by PCR (1 min 94°C, 1 min 60°C, 1 min 30 s 68°C for 30 cycles, followed by 10 min at 68°C) with the 5′-primer, 5′-AAAAAGCAGGCTTAATGC AGAAAGAAAAA-3′, and the 3′-primer, 5′-AGAAAGCT GGGTATAGATAAAAGGCATA-3′, yielding a 578 bp fragment. This fragment contains the putative N-terminal nuclear localization signal (NLS) and the six TIR domains (TIR I to TIR VI), flanked by 12 nt partial *attB1* and *attB2* recombination sites. To construct the *RGLR2* reporter plasmid, the same EST was amplified by PCR (1 min 94°C, 1 min 50°C, 1 min 30 s 68°C for 30 cycles, followed by 10 min at 68°C) with the 5′-primer, 5′-AAAAAGCAGGCT ACCACATGCCAGATTCT-3′, and the 3′-primer, 5′-AGA AAGCTGGGTATAGATAAAAGGCATA-3′, yielding a 504 bp fragment without the putative NLS and flanked by 12 nt partial *attB1* and *attB2* recombination sites.

For both constructs, a second PCR with the 5′-primer 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3′ and the 3′-primer 5′-GGGGACCACTTTGTACAAGAAAGCT GGGT-3′ was performed, creating the entire attB1 and attB2 recombination sites (1 min 94°C, 1 min 60°C, 1 min 30 s 68°C for 30 cycles, followed by 10 min at 68°C). PCR reactions were performed using the Platinum *Pfx* Polymerase (Invitrogen). The two PCR fragments were cloned into the GATEWAY™ compatible binary T-DNA destination vector *pK7FWG2*, allowing the fusion of the GFP coding region at the C-terminal part of the resistance gene (Karimi et al. 2002) in reactions mediated by the GATEWAY™ BP and LR Clonase™ Enzyme Mix (Invitrogen). Expression was driven by the constitutive 35S promoter.

Growth, transformation and fluorescence microscopy

Tobacco BY-2 cells were grown according to Nagata et al. (1992). The *RGLR4* and *RGLR2* plasmids were introduced into the *Agrobacterium* strain LBA4404 according to Geelen and Inzé (2001) yielding the *Agrobacterium* strains RGNLS-GFP and RGNONLS-GFP, respectively. Transgenic calli producing GFP fusion proteins were identified by epifluorescence microscopy. Confocal images were taken with a scanning confocal microscope 100 M with software package LSM 510 version 3.2 (Zeiss) equipped with a 63× water corrected objective (numerical aperture of 1.2). GFP fluorescence was imaged with argon laser illumination at 488 nm and 500–530 nm band emission cut off. Images were taken with 25% laser power. Samples were prepared for life cell recordings, and applied to a chambered cover glass (Lab-Tek, Naperville, IL, USA) coated with 5 mg/ml poly-L-lysine (Sigma-Aldrich). For transmission light images, differential interference contrast optics were used.

Results

Identification and classification of *NBS* genes

The initial stages of *NBS-LRR* gene sequence collection consisted of mining gene models from the Joint Genome Institute *Populus* Genome Sequence Database (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) (Tuskan et al. 2006). Searches for sequences in version 1.1 of the whole-genome shotgun assembly of 410 Mb of sequence were performed using a variety of predicted sequences from monocot and dicot *NBS-LRR* genes as BLAST queries and conserved INTERPRO, KOG and PFAM domains (e.g., apoptotic ATPase, NB-ARC, disease resistance protein, LRR, TIR) in automated annotations. TIR only and TIR-X sequences were not included in our analysis. Several softwares (Coils, MacStripe, pcoils, Paircoil2 and Marcoil 1.0) were used to identify R proteins with a N-terminal coiled-coil domain (named CN/CNL). In addition, to ensure that there were no additional related genes missing from the automated annotation, we performed TBLASTN analyses using subfamily sequences selected to represent the diversity of *Populus* NBS-LRR proteins to search the entire *Populus* genomic sequence. A total of 402 NBS-encoding sequences was identified in the current draft of the genome of *P. trichocarpa* Nisqually-1 (Table 1, Supplemental table 1). For 194 NBS-coding sequences, the "best-model" selected by the automated annotation JGI pipeline was incorrect and we proposed to use a model prediction from a different annotation algorithm (Supplemental table 1). Among the automatically annotated genes, 189 were truncated, biased by sequence gaps or missed the START and/or STOP codon. After manual editing, the predicted amino acid sequence was still incomplete for 86 gene models, due to sequencing errors leading to insertion of false introns and frameshifts by the automatic gene predictors. Some of these gene models are probably pseudogenes. The initial sequence comparisons allowed the identification of several known families of *R* genes,

including the *TIR-NBS-LRR* (*TNL*) and *CC-NBS-LRR* (*CNL*) genes that are abundant in the *Arabidopsis* and rice genomes (Table 1). The different predicted protein sequences formed 15 distinct subfamilies depending on their N-terminal and C-terminal PFAM/INTERPRO domains (Table 1). Predicted NBS proteins without any N-terminal domain or with an ambiguous coiled-coil domain were combined in the *NBS/NBS-LRR* (*N/NL*) subfamily, which was then further divided, according to the NBS signature, in non-TIR-NBS related, non-TIR-NBS-LRR related, TIR-NBS related and TIR-NBS-LRR related proteins. For *Populus*, we identified in total 224 *non-TIR-NBS-LRR* genes, compared to 55 in *Arabidopsis* and 480 in rice, and 91 *TIR-NBS-LRR* genes compared to 94 in *Arabidopsis* and none in rice. In addition, 53 *non-TIR-NBS* genes were identified compared to six in *Arabidopsis* and 52 in rice and 32 *TIR-NBS* genes compared to 23 in *Arabidopsis* and three in rice (Table 1). Taking all subgroups together, we found 317 *NBS-LRR* genes in *Populus*, compared to 149 in *Arabidopsis* and 480 in rice, and 85 *NBS* genes compared to 29 in *Arabidopsis* and 55 in rice (Table 1).

In order to learn more about the evolution of the different sub-groups, the predicted amino acid sequences of the 402 *R* gene homologs, as well as the predicted NBS part of these *R* genes, were compared to the whole genome predicted amino acid sequences from *Arabidopsis* and rice by Blast search (Supplemental table 2). Interestingly, all *Populus* predicted NBS proteins containing a BED finger domain (IPR003656) at their N-terminal part (named BN/BNL) had their top Blast match with *Arabidopsis* CNLs from type B (At4g27220, At4g10780, At4g27190 and At4g26090/RPS2). From the 129 *Populus* genes classified as N/NLs, i.e., lacking a N-terminal TIR or a non-ambiguous coiled-coil domain, 98 were related to *Arabidopsis* CN/CNLs and 31 to *Arabidopsis* TN/TNLs. Amongst the NLs related to *Arabidopsis* CNLs, we identified two, well conserved (56% amino acid identity) ADR1-like *Populus* R proteins (Poptr1_1:710696 and Poptr1_1:826770). The five *Populus* TNLs with a N-terminal NLS motif best matched with At5g36930, a type D TNL. All TIR-NBS-LRR-TIR (TNLT) had their top Blast match with At5g17680, another type D TNL. They are not closely related to the *TNLT-X* genes identified in *Arabidopsis* (Meyers et al. 2003). All *Populus* predicted TNs were closer to *Arabidopsis* TNLs than to *Arabidopsis* TN and vice versa, except two unusual TNs (Poptr1_1:753803 and Poptr1_1:554389), which are the highly conserved homologs of At4g23440 and At5g56220 (69% amino acid identity; Meyers et al. 2002). Table 2 summarizes the comparison of the *Arabidopsis* and *Populus* subclasses (for classes see Meyers et al. 2003) as revealed by the Blast searches. The increase from 55 non-TIR-NBS-LRR to 224 non-TIR-NBS-LRR and from six non-TIR-NBS to 53 non-TIR-NBS in *Populus* refers mainly to a striking higher number of members of the CNL-B class (4×, acquisition of the BED-finger domain) and the CNL-C class (16×) in the *Populus* genome. The total number of TIR-NBS and TIR-NBS-LRR related proteins in *Populus* and *Arabidopsis* is very similar, 91 vs. 94, but for most of the subclasses identified in Arabidopsis only very few members could be identified in *Populus*, except for the TNL-D, where about 100 related proteins could be observed.

Using the rice genome data, the *Populus* BN/BNLs best matched with Os01g57870 and Os01g72680, two rice NBS-LRRs with amino acid similarity to RPS2, and not with XA1, a rice *R* protein that contains also a BED finger domain in its N-terminal part (Bai et al. 2002). The two unusual TNs (Poptr1_1:753803 and Poptr1_1:554389) mentioned above, shared high amino acid similarity (58%) with the only three TNs identified in rice so far (Os01g55530, Os08g38970 and Os09g30380) (Zhou et al. 2004).

The NBS sequences of the predicted R proteins were aligned for phylogenetic analyses (Fig. 1, Supplemental Fig. 1a–c). We divided the 402 *Populus* R proteins into three subgroups consisting of 224 non-TIR-NBS-LRR (plus the two TCNL), 91 TIR-NBS-LRR and 85 Non-TIR-NBS and TIR-NBS. In each subdivision, a selection of *Arabidopsis* and rice R proteins representing the proteins with the highest amino acid similarity, as well as some well-characterized *R* proteins from *Arabidopsis* were included for comparison to establish if the recent evolution of R proteins predated speciation in the Salicaceae. From these subdivisions, we generated the 50% majority-rule consensus trees (Supplemental Fig. 1a–c), using a bootstrap heuristic search (Swofford 1999). Parsimony- and distance-matrix-based calculations yielded very similar results (data not shown). Then, we selected 117 representative R proteins from the subdivision trees and generated an overall tree showing the relationships between the different types of *Populus R* genes (Fig. 1). This figure reveals substantial variation in tree shape and depth, indicative of a wide range of evolutionary histories. There was a clear separation between the clades encoding TIR- and non-TIR-type sequences, reflecting the well-known ancient differentiation of *NBS-LRR* genes into two major groups (Meyers et al. 2003). The non-TIR type sequences were clustered in four major subgroups, one composed of BED finger domain containing-sequences (III) and three CNL related subgroups (I, II and IV) (Fig. 1). The *Arabidopsis* R proteins RPM1 and RPP8 (type C and D), as well as another *Arabidopsis* type C R protein (At3g14460) and three rice R proteins (XA1, Os01g71100 and Os05g41290) clustered with the largest subgroup of non-TIR type R proteins (Fig. 1, I). The clade I was further

Table 2 Comparison of *Arabidopsis* and *Populus* subfamilies

Arabidopsis class[a]	Number *Arabidopsis*[a]	Number *Populus* (complete sequence)	Number *Populus* (NBS part)	Ratio *Populus*/ *Arabidopsis*	Putative orthologs (reciprocal blast)
CNL-A	6	5	5	0.8	764584/At5g66900, 710696/At4g33300ADR1/ Os12g39620
CNL-B	26	103	103	**4**	571475/At4g27190, 798001/Atg63020, 417550/ Atg26090RPS2/Os01g57870
CNL-C	10	164	162	**16**	784629/At3g14470, 679204/At3g50950, 755797/ At3g07040RPM1/Os04g46300
CNL-D	16	6	8	0.4	793256/At1g50180, 791093/At1g53350
NL-A	2	1	1	0.5	–
CN-A	1	–	–	–	–
xTNx	2	2	2	1	554389/At4g23440/Os01g55530, 753803/ At5g56220/ Os08g38970
TNL-A	25	1	5	0.04 (0.2)	–
TNL-B	12	8	15	0.7 (1.3)	–
TNL-C	2	3	4	1.5 (2)	806354/At1g27170
TNL-D	9	105	95	**11.7 (10.6)**	581005/At5g17680, 264749/At5g36930
TNL-E	8	1	–	0.13 (–)	–
TNL-F	11	1	–	0.09 (–)	–
TNL-G	18	2	–	0.1 (–)	–
TNL-H	25	–	2	– (0.08)	–

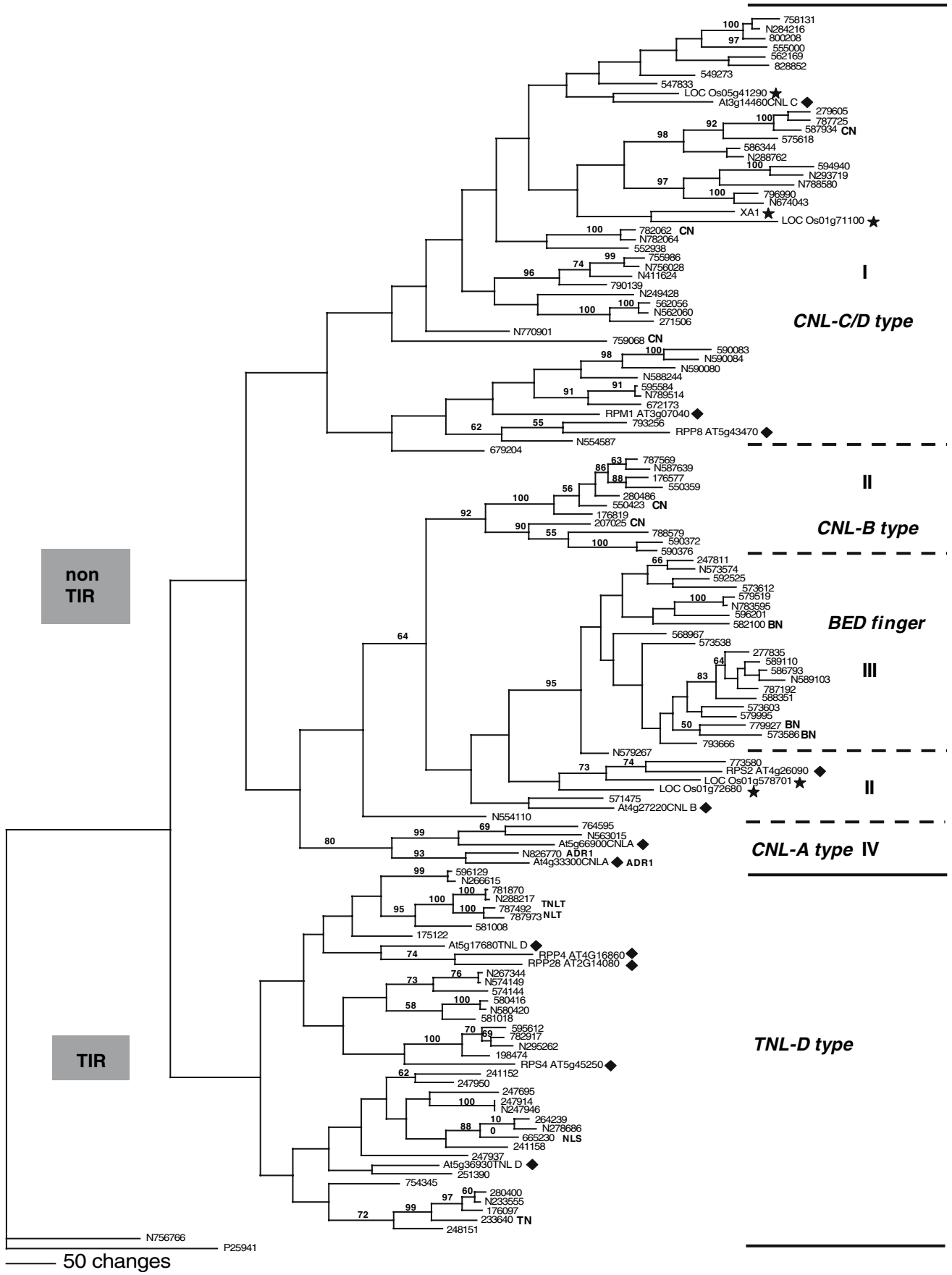
Number of *Populus* members as revealed by BlastP search of the *Arabidopsis* proteome, using either the complete amino acid sequence or only the NBS part

Bold values indicate classes with higher number in *Populus* compared to *Arabidopsis*

[a] Meyers et al. (2003)

divided in two subgroups, one larger, composed of At3g14460-like type C predicted *Populus* R proteins, and a smaller one assembling Poptr1_1:793256, a type D predicted *Populus* R protein, and several RPM1-like type C predicted *Populus* R proteins. RPS2, as well as another type B *Arabidopsis* R protein (At4g27220) and two RPS2-like rice R proteins (Os01g578701 and Os01g72680), formed a cluster with a smaller subgroup of non-TIR type R-proteins (Fig. 1, II, Supplemental Fig. 1a, IV). From this cluster arose the clade containing the BED finger domain containing sequences (III), indicating, as already found for the BLAST searches, a closer relationship of this *Populus* NBS-LRR protein type to RPS2 than to the rice XA1, which also contains a BED finger domain, but was clustered with the subgroup I. The third and smallest subgroup (Fig. 1, IV) contained the type A CNLs, e.g., ADR1 (see also Supplemental Fig. 1a, I). All *Arabidopsis* TNLs formed a cluster with *Populus* TNLs, and as expected, none of the rice R proteins clustered within the TIR subfamily clades. *Arabidopsis* type D TNLs seem to be more closely related to *Populus* TNLs than RPS4 (TNL-B), RPP4 (TNL-E) and RPP28 (TNL-G), which were located on more divergent branches (see also Supplemental Fig. 1b), as already indicated by the Blast search results (Table 2). All five NLS-TNLs were assembled in one clade in the phylogenetic tree (Supplemental Fig. 1b, III). The 33 selected N/NLs were dispersed amongst the different clades, indicating a diverse rather than a monophyletic origin and suggesting that these genes are distinguished either by a loss of a domain or a failure to identify the specific domain due to divergence of the sequence (see also Supplemental Fig. 1a–c). Similar results were obtained when the cladistic analysis was performed using the complete sequences of predicted proteins (data not shown).

Fig. 1 Phylogenetic relationship of the NBS-LRR proteins in ▶ *P. trichocarpa* Nisqually-1. The Maximum Parsimony tree was constructed using the NBS part of the predicted protein sequences of 117 selected *Populus R* proteins, 11 *Arabidopsis* sequences and five rice sequences. From the trees shown in Supplemental Fig. 1, representative predicted protein sequences were aligned. The line indicates the separation between the major non-TIR and TIR-type NBS protein sequences. Sequence homologies to *Arabidopsis* subclasses are indicated in italic. *Arabidopsis* and rice sequences are denoted by diamonds and stars, respectively. N indicates predicted N or NL proteins. Predicted proteins are designated according to their JGI *Populus* Genome Database entry codes, omitting the Poptr1_1: prefix. P25941 from *Streptomyces* rooted the tree as an outgroup. Only bootstrap results >50% were reported


non TIR
TIR
I
CNL-C/D type
II
CNL-B type
BED finger III
II
CNL-A type IV
TNL-D type
50 changes
758131
N284216
800208
555000
562169
828852
549273
547833
LOC Os05g41290
At3g14460CNL C
279605
787725
587934 CN
575618
586344
N288762
594940
N293719
N788580
796990
N674043
XA1
LOC Os01g71100
782062 CN
N782064
552938
755986
N756028
N411624
790139
N249428
562056
N562060
271506
N770901
759068 CN
590083
N590084
N590080
N588244
595584
N789514
672173
RPM1 AT3g07040
793256
RPP8 AT5g43470
N554587
679204
787569
N587639
176577
550359
280486
550423 CN
176819
207025 CN
788579
590372
590376
247811
N573574
592525
573612
579519
N783595
596201
582100 BN
568967
573538
277835
589110
586793
N589103
787192
588351
573603
579995
779927 BN
573586 BN
793666
N579267
773580
RPS2 AT4g26090
LOC Os01g57870
LOC Os01g72680
571475
At4g27220CNL B
N554110
764595
N563015
At5g66900CNLA
N826770 ADR1
At4g33300CNLA ADR1
596129
N266615
781870
N288217 TNLT
787492
787973 NLT
581008
175122
At5g17680TNL D
RPP4 AT4G16860
RPP28 AT2G14080
N267344
N574149
574144
580416
N580420
581018
595612
782917
N295262
198474
RPS4 AT5g45250
241152
247950
247695
247914
N247946
264239
N278686
665230 NLS
241158
247937
At5g36930TNL D
251390
754345
280400
N233555
176097
233640 TN
248151
N756766
P25941

Architectural diversity in *Populus NBS-LRR* genes

To analyze the fine structure of *Populus NBS* genes, we divided them into three parts defined by either their N-terminal, NBS, or LRR & C-terminal domains. These three parts were analyzed separately using the program MEME (Multiple Expectation Maximization for Motif Elicitation) (Bailey and Elkan 1995) (Table 3, Supplemental table 3). For each gene

Table 3 Major MEME motifs in predicted *Populus* CNL, BNL and TNL proteins

Domain	(Sub)group	Motif[a]	Sequence[b]
TIR	TNL	TIR-1	YDVFLSFRGEDTRNNFTDHLYTALCQAGIHTFRDD
	TNL	TIR-2	ELPRGEEISPHLWKAIQESRISIIVFSKDYASPWCLDELVKI
	TNL	TIR-3	CKKxTGQIVLPVFYDVDPSDVRKQTGSFA
	TNL	TIR-4	VQRWRDALTEAANLSGWD
	TNL	NLS	MQKEKRKQSKDEENDSSSRKRRKADLKKP
BED	BNL		SNDPFWNYVEKMDDGSMKCKFCGHLFANGTSISRIKWHLSGHGVA
CC	CNL		See Supplemental table 3
TN linker	TNL		ANGHEAKFIQEIVEDVLYKLNxxxPHVPKHLVGMDSRVEYINSWLQMGTHDVC
BN linker	BNL		LPTSSTKLVGQAFEENKKVIWSLLMDDEV
NBS	TNL	P-loop	VGIYGMGGIGKTTIAKVVYNQIRxQ
	CNL	P-loop	IPIVGMGGVGKTTLAQLVYND
	BNL	P-loop	STIGIYGMGGVGKTTLLQHIHNELLQRPD
	TNL	RNBS-A	VRLQEQLLSDILKER
	CNL	RNBS-A	KIWVCVSQDFDVxRIQKAILE
	BNL	RNBS-A	HVYWVTVSQDFSINRLQNLIAKRLGLDLS
	TNL	Kinase 2	IKDRLCRKKVLIVLDDV
	CNL	Kinase 2	LKGKKFLLVLDDVWN
	BNL	Kinase 2	RKKQKWILILDDLWNN
	TNL	RNBS-B	WFGPGSRIIITTRDK
	CNL	RNBS-B	DGARGSKILVTTRNExVA
	BNL	RNBS-B	VPLKGCKLIMTTRSERVC
	TNL	RNBS-C	YEVEELNDDDALQLFSWHAFKNDQP
	CNL	RNBS-C	VHxLGxLSEEDCWSLFKKxAF
	BNL	RNBS-C	PLSEGEAWTLFMEKLGHDIAL
	TNL	GLPL	VVDYANGLPLALEVLGSSLYGR
	CNL	GLPL	EIAKKCKGLPLAAKTLGGLLR
	BNL	GLPL	RIAKDVARECAGLPLGIITMAGSLRGV
	TNL	RNBS-D	EEKKIFLDIACFFIGMKKDYV
	CNL	RNBS-D	RCFAYCAIFPKDYEIxKEELVRLWMAEGF
	BNL	RNBS-D	QQCLLYCALFPEDHRIEREELIGYLIDEG
	TNL	MHDV	IVSDDKLWMHDLLRDMGREIV
	CNL	MHDV	FKMHDLVHDLAQSIA
	BNL	MHDV	KMHDLIRDMAIQILQENSQGM
NL linker	TNL		EPGKRSRLWTHEDVVLMDNGTEKVEGIFLD
	CNL		See Supplemental table 3
LRR			See Supplemental table 3
C terminus	TNL		EIPGWFNHQGWGSSISFQ
	CNL		GEDWPKIAHIPNIYI
	BNL		EWEHPNAKDVLRPFVEF

[a] Domains and motifs are listed in the order that they occurred in CNL, BNL and TNL proteins, starting with motifs most N terminal in the protein

[b] Consensus amino acid sequence derived from MEME. Related motifs in the NBS and LRR domains of CNL, BNL and TNL proteins are aligned. The complete output is available in the supplemental data online. x indicates a non conserved residue. Underlined residues indicate consensus matches

model, the NBS region was defined from the so-called P-loop to the MHDV-motif using the corresponding conserved motifs from *Arabidopsis* (Meyers et al. 2003).

The N-terminal and NBS regions

The N-terminal region ranged from 150 to 250 amino acids from the start of the coding region to the beginning of the P-loop of the NBS domain. Using the COILS program, the CC motifs, a characteristic feature of CNL *R* proteins, were clearly apparent in 119 predicted proteins (Supplemental table 1) and this CC-containing region shared significant sequence similarity (Supplemental table 3, e.g., CC motifs 3 & 8). The N-terminal region of 34 CNL-related sequences contained an unusual domain, the so-called BED finger or DUF1544 domain (IPR003656/IPR011523). These domains of about 50–60 amino acid residues contain a shared pattern of cysteines and histidines (Supplemental Fig. 2) that is predicted to form a zinc-finger DNA-binding domain (Aravind 2000). This domain is absent from all *Arabidopsis R* genes including type B, which is the closest group of homologs for these poplar genes (see above) (Meyers et al. 2003), but is present in the rice *Xa1* gene and two *XA1-like* genes (Bai et al. 2002). We identified 15 and 56 conserved sequence motifs in the N-terminal region from *TNL* and non-*TNL* genes, respectively (Table 3, Supplemental table 3). Five gene models coding for TNL proteins contained a predicted N-terminal nuclear localization signal (NLS) (Supplemental Fig. 3) suggesting that they are targeted to the nucleus (see below), and possibly inferring similarity of function to the NLS domain identified in the C-terminal region of the *RRS1 R* gene (Deslandes et al. 2003) and the RPS4 *R* gene (Wirthmueller et al. 2007) from *Arabidopsis*.

The NBS region showed lower sequence diversity than the N-terminal region. The eight major motifs (P-loop, RNBS-A, kinase 2, RNBS-C, GLPL, RNSB-D and MHDV) found in the NBS region of *Arabidopsis* CNL and TNL proteins were identified in *Populus* NBS proteins (Table 3, Supplemental table 3). These motifs are divergent between TIR and non-TIR NBS-LRR (Table 3). The MHDV motif in *Populus* was often slightly modified into MHDL (Table 3). As found for *Arabidopsis* and other plants before (Meyers et al. 2003; Chini and Loake 2005), in the *Populus* CNL type A-like subgroup this motif has changed to QHDV (Poptr1_1:764595, Poptr1_1:764584) and QHDL (Poptr1_1:710696, Poptr1_1:826770), respectively. In *Arabidopsis*, all CNs and TNs are lacking the MHDV motif. In *Populus*, only the two unusual TNs already mentioned above (Poptr1_1:753803 and Poptr1_1:554389) lacked the MHDV motifs, all others either contained the motif or the sequence was truncated, probably indicating a different origin of CNs and TNs in *Arabidopsis* and *Populus*.

The LRR and C-terminal regions

The LRR regions were highly variable in sequence and size, but were between 350 and 750 amino acids for most of the predicted proteins. The repeats in most of the genes were imperfect (Supplemental table 3). This high sequence variability is consistent with their role in determining the recognition specificity of pathogen encoded Avr proteins. Several *NBS-LRR* genes have been identified in *Arabidopsis* that code for additional domains after the LRR in their C-terminal region (e.g., NLS domain, (Meyers et al. 2003; Deslandes et al. 2003)). The LRR domain of most analyzed *Populus* NBS-LRR proteins extends to the end of the predicted gene products. However, 15 predicted proteins showed a TIR domain in the C-terminal region (Supplemental table 3).

While mining *NBS* gene sequences from *Populus*, we identified several gene families lacking the LRR region. A total of 13 *TIR-NBS*, 19 *CC-NBS*, and 7 *BED-NBS* gene models without this domain were found. The predicted NBS region of these gene models are scattered throughout the phylogenic trees (Fig. 1) indicating a closer relationship to other members of the TNL, CNL and BNL classes, respectively than to other TN, CN or BN, respectively. The only exceptions are again the two unusual TNs (Poptr1_1:753803 and Poptr1_1:554389), which form a clearly separated clade in the phylogenetic tree (Supplemental Fig. 1c). It is possible that at least some of these genes have been misassembled and additional unique domains have not been identified so far. Finally, 49 predicted proteins consisted only of the NBS region, and these were dispersed in TNL and CNL clades (Fig. 1), suggesting that they may have been derived from *NBS-LRR* genes. At least some of these genes with a different structure are transcribed as *Populus* ESTs (Supplemental table 1).

Conserved intron positions

We analyzed the intron numbers and positions of the different *Populus NBS-LRR* subgroups and compared them with those of *Arabidopsis* (Fig. 2; Supplemental table 1). In general, *BN/BNLs* and *CN/CNLs* contain fewer introns (0–4) than *TN/TNLs* (3–7), as previously found in *Arabidopsis* for *CNLs* and *TNLs* (Meyers et al. 2003). In *TNLs*, the different domains (TIR, NB-ARC, LRR/C-Terminal) are often separated by introns. Figure 2 shows the comparison of the intron/exon structure of *Populus* genes with those of closely related *Arabidopsis* genes. The *Populus*

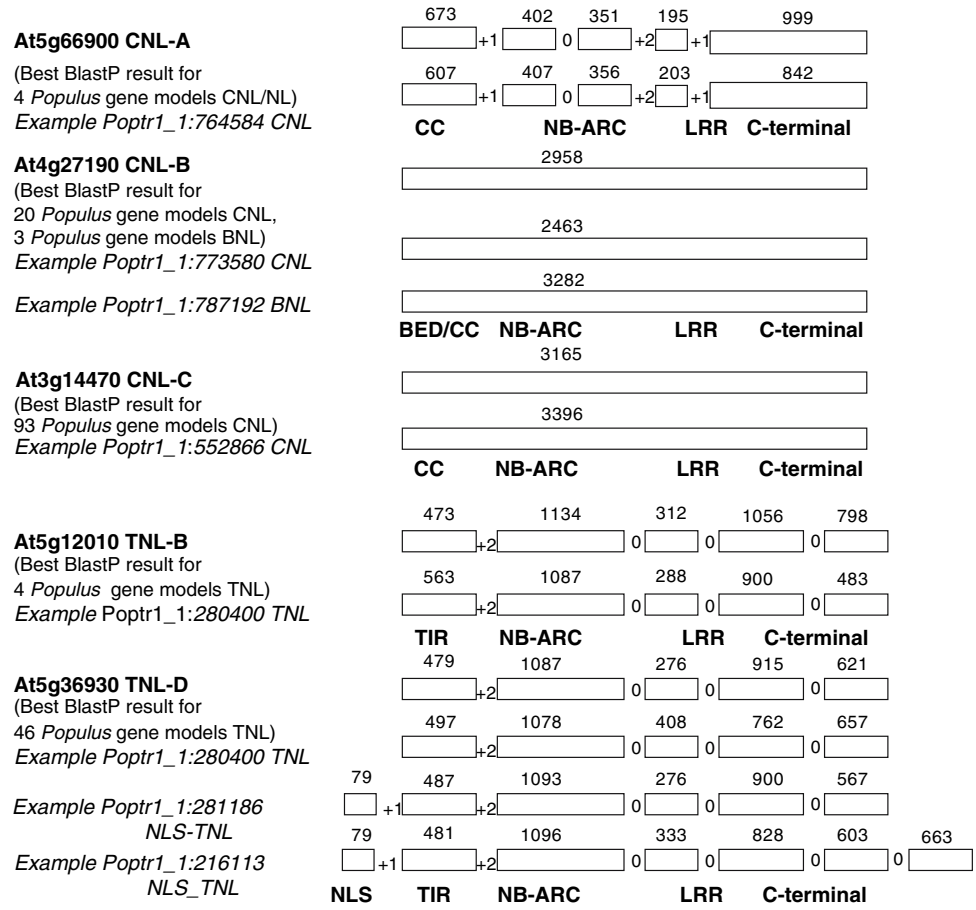


Fig. 2 Comparison of the intron/exon configuration of selected *Populus R* genes with that of related *Arabidopsis R* genes. Numbers above exons indicate the size of the exons (bp). Numbers between exons indicate the phase of the intron. The locations of CC, TIR, BED finger, NB-ARC, LRR and C-terminal domains are indicated in bold

CNL Poptr1_1:764584 exhibited a similar intron/exon distribution as *At5g66900*, a *CNL-A*, and the intron phases were conserved. The same observation could be made for the *Populus TNL Poptr1_1:280400* and a type D *TNL* from *Arabidopsis*, *At5g36930*. Even in the case of the *Populus NLS-TNL Poptr1_1:216113*, where apparently a new N-terminal domain was added during evolution, the intron phases of the rest of the gene were still conserved compared to *Arabidopsis* (Meyers et al. 2003).

Genomic organization and duplication of *NBS* genes

While clustering of *R* genes does occur frequently, complete sequencing of the *Arabidopsis* genome has revealed a large proportion of single-gene loci (Meyers et al. 2003). The physical distribution of *NBS* genes was investigated across the *Populus* genome. We used the same parameters to define a cluster as Meyers et al. (2003) and Richly et al. (2002), i.e., two or more NBS genes that occurred within a maximum of eight non-NBS genes were considered to be clustered. The 402 *Populus NBS* genes were distributed over 228 loci. In the current draft sequence, 170 sequences were located on linkage groups (LG), whereas 232 genes were found on as yet unmapped scaffolds (Table 4). The latter may correspond to different haplotypes (Kelleher et al. 2007). From the 170 *NBS* genes located on LG, 72% consisted of between 2 and 12 tightly linked genes in 37 clusters (Fig. 3; Supplemental table 4). Cluster size varied between 8.5 kbp (cluster LG_XI_5) and 328 kbp (cluster LG_XIX_2). Single-gene loci and clusters were distributed unevenly over the chromosomes, as highlighted by the presence of clusters of clusters (superclusters) on LG II, LG XI and XIX, which cover 1.14, 0.95 and 1.16 Mbp, respectively. LG VIII, X and XVII did not carry any *NBS* genes. Complex *NBS* gene clusters were found on larger regions on LG II (18 *CNL* and *NL* genes on approx. 1 Mb), XI (13 *TNL* and *NL*) and XIX (31 *BNL*, *TNL*, *N*, *BN* and *NB* genes on approx. 1.16 Mb). Truncated *NBS* genes lacking the LRR region or gene models restricted to the LRR region were often located adjacent to complete *NBS-LRR* genes (Figs. 3, 4). At this stage of the *Populus* genome assembly it is possible that some of these genes are truncated due to sequencing errors or that they have been misassembled. As an example, the data obtained from the mapping of *NBS-LRR* genes and from phylogenetic analyses were combined for LG XIX (Fig. 4). The three LG XIX clusters were made up of *NBS* genes from up to three

Table 4 Number of *NBS* genes in LGs and in scaffolds of *P. trichocarpa* genome

Predicted protein domains	Letter code	Total number	No. in LGs	No. in scaffolds
TIR-NBS	TN	13	8	5
TIR-NBS-LRR	TNL	64	37	27
TIR-NBS-LRR-TIR	TNLT	13	1	12
TIR-NBS-LRR-NBS	TNLN	1	0	1
NBS-LRR-TIR	NLT	1	0	1
TIR-CC-NBS-LRR	TCNL	2	1	1
CC-NBS	CN	19	6	13
CC-NBS-LRR	CNL	119	52	67
BED/DUF1544[a]-NBS	BN	6	1	5
NBS-BED/DUF1544[a]	NB	1	1	0
BED/DUF1544[a]-NBS-LRR	BNL	34	13	21
NBS-LRR	NL	80	35	45
NBS	N	49	15	34
Total NBS genes		402	170	232

[a] BED finger and/or DUF1544 domain

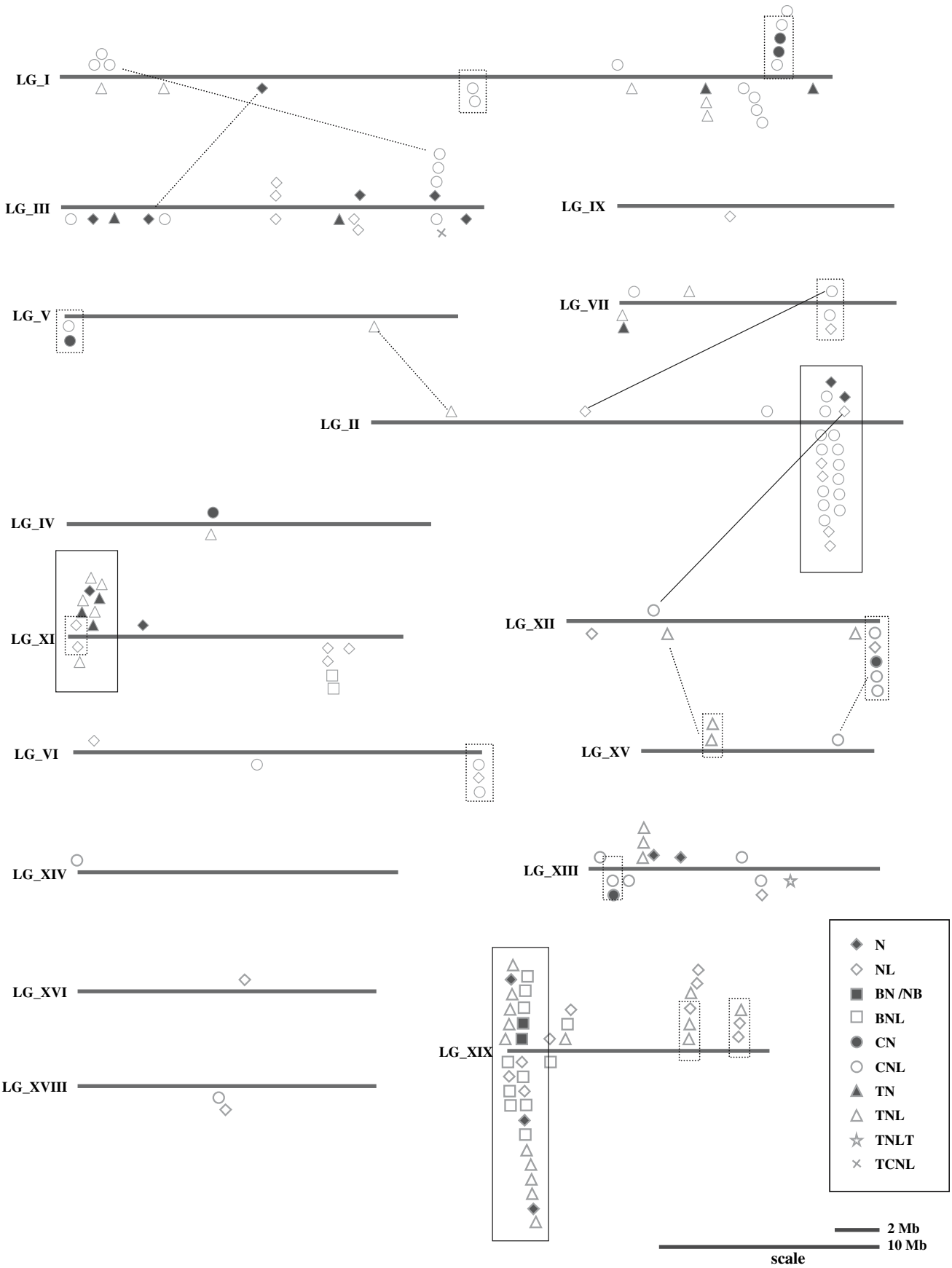
different families, including TNL, BNL, and NL, and from different clades (Supplemental Fig. 1, Fig. 4) suggesting that a mechanism selectively sampling *NBS* genes of different clades/families is operative (Richly et al. 2002). However, clusters XIX_1 and XIX_2 also contained members of the same clade (e.g., *Poptr1_1: 573538/Poptr1_1: 573542/ Poptr1_1: 573543* and *Poptr1_1:665212/Poptr1_1:573612*; Supplemental Fig. 1a). Furthermore, tandems of *NBS* genes with high sequence similarity, i.e., arising from recent duplication events, (e.g., *Poptr1_1:247713/Poptr1_1:247950* and *Poptr1_1: 247732/Poptr1_1:247937*) can be found in head-to-tail orientation in two clusters (XIX_1 and XIX_3) of the supercluster suggesting an amplification by translocation. Dot plot analysis of 20 Kb frames around *NBS-LRR* genes within clusters XIX_1 and XIX_3 with PipMaker (http://pipmaker.bx.psu.edu/pipmaker/) (Schwartz et al. 2000) revealed that 5′ and 3′ non-coding regions of up to 10 Kb around genes *Poptr1_1:247713/Poptr1_1:247950* and *Poptr1_1:247732/ Poptr1_1:247937* were also duplicated (data not shown), suggesting ancient events of large duplication at these loci. The LG XIX supercluster did not contain genes belonging to other *R* gene families (e.g., RLK, *Cf*-like RLP) (Fig. 4). Fragments of retroelements, which may play a role in *R* gene duplication, found in this supercluster, fall into the *Gypsy* and *Pong/Pif* families.

The N-terminal NLS is a functional signal

As indicated above, five gene models coding for TNL proteins contained an amino acid extension rich in arginine and lysine (Supplemental Fig. 3), suggesting a possible bipartite nuclear localization sequence (NLS) that consists of two adjacent basic amino acids (Arg or Lys), a spacer region of ten residues, and more than three basic residues (Arg or Lys) in the five positions after the spacer region. This feature appears to be conserved among *Populus* species as TNL proteins with NLS signal have also been identified in *P. deltoides* (Lescot et al. 2004). To evaluate the functional significance of the putative NLS present in these predicted proteins, the NLS-dependent translocation of a protein fusion made of the *P. tremula* homolog of Poptr1_1:247642 (supercluster LG XIX) and Poptr1_1:216113 (LG VII), and green fluorescent protein (GFP) was investigated in BY2-cells. One fusion was made with, and the other without the NLS region. As shown in Fig. 5, the fusion protein lacking the NLS was localized throughout the cell, whereas the fusion protein containing the bipartite NLS exclusively located in the nucleus, demonstrating the functionality of the amino-terminal peptide in sequestering this *R* protein in the nucleus.

The expression of *NBS* genes

By analyzing dbEST at GenBank, we were able to identify 179 ESTs corresponding to 87 annotated *NBS* genes (Supplemental table 1) among the 376,600 ESTs from various *Populus* species, confirming the low expression of *NBS* genes. Amongst the prominent, an ADR1-like *R* transcript (*Poptr1_1:826770*) was represented by 12 ESTs (from cDNA libraries derived from 9 different tissues), whereas a CNL transcript (*Poptr1_1:747319*) and another ADR1-like *R* gene (*Poptr1_1:710696*) were represented by nine and eight ESTs, respectively (Table 5). Other


LG_I
LG_III
LG_IX
LG_V
LG_VII
LG_II
LG_IV
LG_XI
LG_XII
LG_VI
LG_XV
LG_XIV
LG_XIII
LG_XVI
LG_XIX
LG_XVIII
N
NL
BN /NB
BNL
CN
CNL
TN
TNL
TNLT
TCNL
2 Mb
10 Mb
scale

◀**Fig. 3** Representation of genomic loci of *NBS* genes in *P. trichocarpa* Nisqually-1 chromosomes. The numbered grey bars represent the 19 chromosomes (LGs); scale is indicated at the bottom right. *NBS* genes are represented by symbols according to their classification. The 60 and 12 Myr duplications are linked by black and dotted lines, respectively. Tandem duplications are indicated in dotted boxes and superclusters are boxed

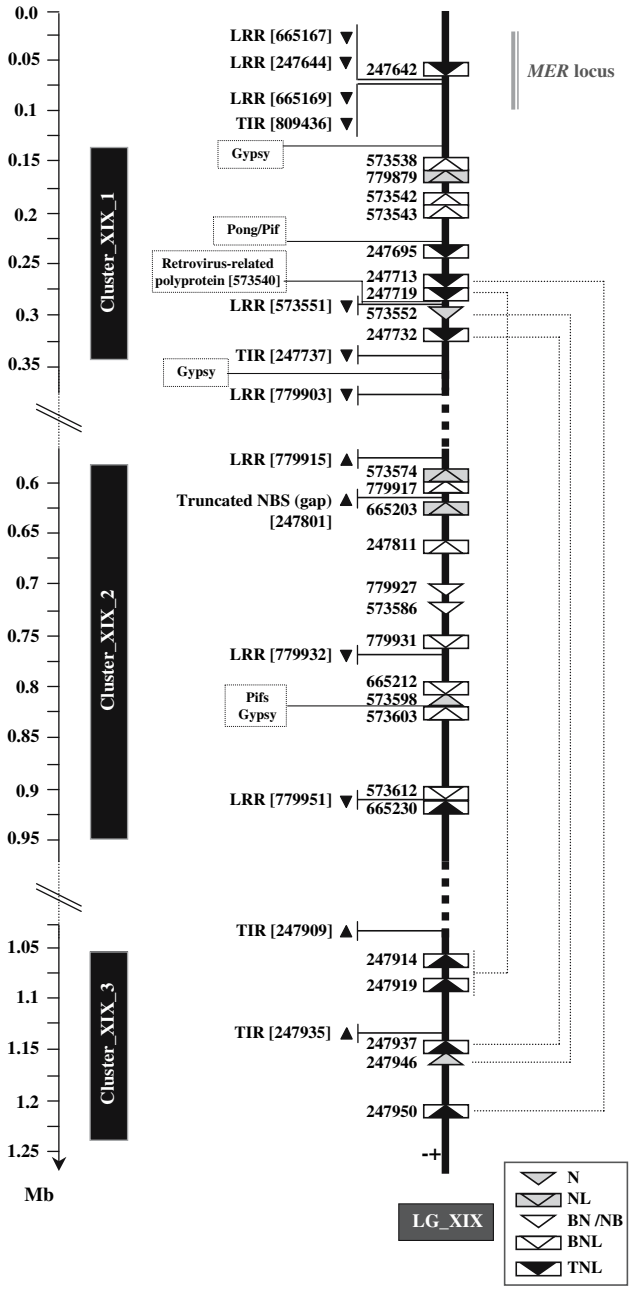


Fig. 4 Combined physical-phylogenetic map of NBS-encoding genes on linkage group XIX of *P. trichocarpa* Nisqually-1. NBS genes are drawn as arrowheads directed toward the 3′ ends of genes and shading indicates membership of the corresponding gene in different NBS protein groups (e.g., TNL, BNL). Duplicated genes are joined by dotted lines. NBS-LRRs are designated according to their six-digit JGI *Populus* Genome Database entry codes, omitting the Poptr1_1: prefix. Positions of repeated elements are designated by dotted boxes

gene models were only represented by a few ESTs (Supplemental table 1). These ESTs originated from 43 different cDNA libraries made from different tissues sampled under different environmental conditions. Interestingly, 19 ESTs (corresponding to 9 *NBS* gene models) were found in a cDNA library constructed from leaves infected by the fungal pathogen *Marssonia* (Huang et al. unpublished). Furthermore, 17 ESTs (corresponding to 10 *NBS* gene models) derived from RNA of terminal vegetative buds from 20 year old trees (Ralph et al. 2006), 13 ESTs (corresponding to 9 *NBS* gene models) were found in a bark cDNA library (Unneberg et al. unpublished) and 11 ESTs (corresponding to 7 *NBS* gene models) derived from a library made from cultured cells grown in media supplemented with either salicylic acid, benzothiadiazole, methyljasmonate, chitosan or an extract from the filamentous fungus *Pollacia radiosa* that causes short tip disease (Ralph et al. 2006) (Supplemental table 1). Amongst ESTs from a subtractive suppressive hybridization cDNA library constructed from RNA of P. trichocarpa × P. deltoides challenged by an incompatible strain of the foliar rust *Melampsora larici-populina*, was only one *TNL* (*Poptr1_1:179317*) (Rinaldi et al. 2007). To learn more about the expression of *Populus NBS-LRR* genes, we performed a gene profiling analysis using the *Populus* NimbleGen whole-genome expression oligo array. This array contains probes for 44,133 annotated gene models including 331 putative *NBS-LRR* genes (Tuskan et al. 2006). We monitored the gene expression in leaves (LPI7-8) of *P. trichocarpa* × *P. deltoides* grown for three months in greenhouse, as well as in rust-infected leaves, in order to assess the possible regulation of *NBS-LRR* genes in this plant-pathogen interaction. In control and rust-infected leaves, only 34 *NBS-LRR* genes were expressed at detectable levels (normalized intensity >200) (Supplemental table 1 and Table 5). None of the expressed *NBS* genes showed a significant alteration in mRNA level 48 h post infection, consistent with most published reports showing that NBS-LRRs are not substantially induced by infection (Tan et al. 2007). The *TNL Poptr1_1:179317* identified in the subtractive suppressive hybridization cDNA library was neither expressed in control leaves nor in leaves 48 h after inoculation, suggesting a possible expression at earlier time-points (12 or 24 h) of the poplar-rust interaction. Most of the *R* genes with a strong EST support were not expressed in the control and rust-infected leaves used for the NimbleGen arrays and vice versa, most of the *R* genes identified as expressed by the NimbleGen arrays had no EST support in the NCBI EST database (Table 5), reflecting probably the low levels of expression or the fact that the *NBS* genes are only expressed under specific conditions in given tissues.

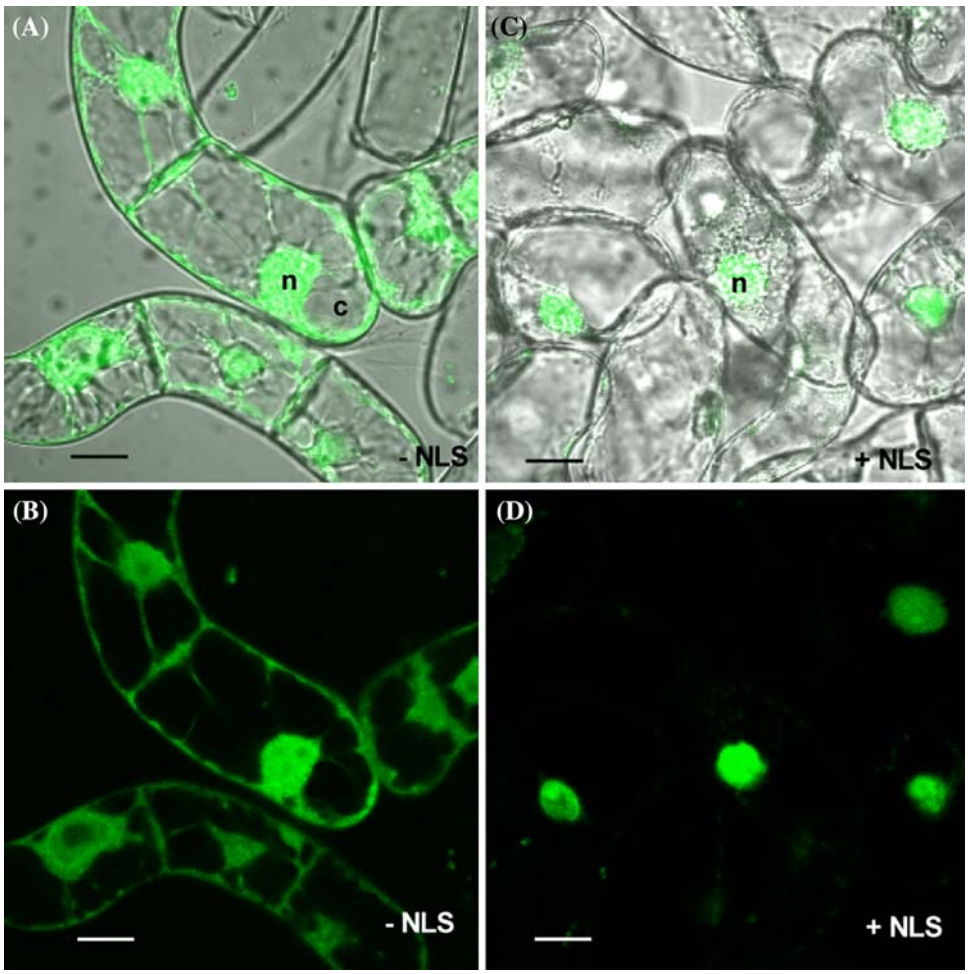


Fig. 5 Intracellular localization of *R*-gene-GFP-fusion proteins, with and without the NLS, in tobacco BY2-cells. (**a**) and (**b**), BY2-cells transformed with *RGONLS-GFP*, i.e., without NLS; (**c**) and (**d**), BY2-cells transformed with *RGNLS-GFP*, i.e., with NLS. (**a**, **c**) Differential interference contrast; (**b**) and (**d**), dark field. The bar represents 50 µm. Images were taken by confocal microscopy. n, nucleus; c, cytoplasm

Discussion

As the largest class of disease resistance genes, the *NBS-LRR* genes play a crucial role in defending plants from a multitude of pathogens and pests. The availability of the complete genomic sequences of *P. trichocarpa*, a perennial species, the related *Arabidopsis* and the distantly related monocot, rice, allows comparative evolutionary analyses of these genes. The dicot *NBS* genes can be divided into two distinct groups, those coding for a TIR domain at their N-termini and those without the TIR domain. The TIR class of *NBS-LRR* genes accounts for the largest proportion of the *NBS-LRR* genes in the *Arabidopsis* genome (Meyers et al. 2003), but this class has not been found in rice and other cereals (Bai et al. 2002; Zhou et al. 2004). In *Populus*, we identified 91 *TIR-NBS-LRR* related genes (94 in *Arabidopsis*), but the largest class is represented by non-*TNL* genes. The numbers of genes in this class has amplified to ~225 or more in *Populus*, compared to ~50 in *Arabidopsis* (Meyers et al. 2003). The higher number in *Populus* is mainly due to the four-fold expansion of the CNL-B class (acquisition of the BED-finger domain) and the CNL-C class (16 fold) compared to *Arabidopsis*. Since we do not know the *R* genes complement of the common ancestor, this could either be explained by an increase of these *R* gene types in *Populus* or a decrease in *Arabidopsis*. But the number of non-*TNL* genes in *Populus* is still much lower than in rice (ca. 480) and probably comparable to the number in *Medicago* (177 CNL in 60% of the genome sequenced so far, Ameline-Torregrosa et al. 2007).

In *Populus*, the ~400 members of the *NBS* family can be divided into multiple subfamilies with distinct domain organizations (Table 1). They represent about 1% of all the predicted proteins in this species, while they represent 1% and 0.43% of the rice and *Arabidopsis* genomes, respectively. The *NBS* family includes 64 *TNL* genes, 224 non-*TNL* genes, and 17 unusual *TIR-NBS*-containing genes (*TNLT, TNLN* or *TCNL*) (Table 1), which are different from the unusual *TNLX* previously identified in *Arabidopsis* (Meyers et al. 2003). On the other hand, *TNL-WRKY*, as well as *TNTNL*, *TNLTX* and *TNLX* identified in *Arabidopsis* were not found in *Populus*. The latter finding indicates that domain acquisition (i.e., biological innovations) has continued to occur through this versatile family

Table 5 Predicted Populus *R* genes with a strong EST support and/or an elevated transcript level in *P. trichocarpa* × *P. deltoides* leaves (as revealed by NimbleGen Poplar expression arrays)

Gene model	Protein ID (Poptr1_1)	Classification	Mean intensity on NimbleGen Array	Number of ESTs in NCBI database
eugene3.01800028	585619	CNL	6332	0
eugene3.01340076	582100	BN	1297	0
grail3.0305000501	676846	CNL	1025	0
fgenesh4_pm.C_LG_I001004	798001	CNL	1014	3
gw1.XIX.1828.1	249428	NL	1014	0
fgenesh4_pg.C_scaffold_64000216	782354	N	977	0
gw1.305.34.1	284216	NL	912	0
fgenesh4_pg.C_LG_VII000940	764595	CNL	813	0
eugene3.00440238	592903	N	809	0
eugene3.01200105	580420	N	650	0
estExt_fgenesh4_pg.C_400229	826770	ADR1-like	314	12
estExt_Genewise1_v1.C_4370003	747319	CNL	Not on array	9
estExt_Genewise1_v1.C_LG_II2154	710696	ADR1-like	<200	8
eugene3.08140001	596330	NL	<200	6
fgenesh4_pg.C_scaffold_134000057	784915	BNL	<200	5
eugene3.00290263	590080	NL	<200	5
fgenesh4_pg.C_LG_XII001317	773542	CNL	<200	4
gw1.VII.1617.1	217312	TNL	<200	4
gw1.XI.1542.1	233562	TNL	<200	4
eugene3.00640255	594667	TNL	<200	4

of proteins in the *Populus* lineage. We also found several classes of truncated *R* genes: 19 *CC-NBS*, 13 *TIR-NBS*, 83 *NBS-LRR* and 46 *NBS* genes. Interestingly, a large class of 34 predicted *R* proteins, so-called *BNL*, carries sequences near its N terminus that are predicted to code for a zinc-finger, DNA-binding domain. At least some of the *BNL* genes are transcribed, as the Zn-finger domain region was found in one *Populus* EST in the NCBI EST database (BU812997; Supplemental table 1) and ESTs of the 3′ end of predicted *BNL* genes exist as well. In addition, two *BNL* genes were found expressed in *Populus* leaves using the NimbleGen arrays. To our knowledge, this DNA-binding domain is absent from *Arabidopsis R* genes (Meyers et al. 2003) and *Medicago* (Ameline-Torregrosa et al. 2007), and was only identified in the rice *Xa1*, a bacterial blight-resistance *NBS-LRR* gene, and two of its homologs (Bai et al. 2002). However, comparison of the *Populus* BNL with R proteins of *Arabidopsis* and rice using Blast search and phylogenetic analyses revealed their closer relationship to *Arabidopsis* and rice CNL type B, e.g., RPS2. Although the family appears to be missing from the *Arabidopsis* genome, it is apparently a very old monophyletic gene family as evidenced by the extensive sequence divergence among members (Supplemental Fig. 1a). A class of 46 genes encoding only the NBS region was observed in the *Populus* genome and this class is also present in rice (Bai et al. 2002; Zhou et al. 2004). It was thought that this family was missing or reduced in dicot genomes (Bai et al. 2002) and the present analysis showed this is not the case. For three of them (*Poptr1_1:580420*, *Poptr1_1:592903*, *Poptr1_1:782354*; Supplemental table 1), we confirmed a transcription in *Populus* leaves by using the *Populus* NimbleGen array. These truncated *R* genes could act as adaptor molecules in the resistance response through the recruitment or interaction with *NBS-LRR* genes (Meyers et al. 2002; Belkhadir et al. 2004).

Within the TNL proteins, a novel N terminal domain was also found. Five predicted proteins showed a NLS (named NLS-TNL) involved in nuclear addressing of the protein (Supplemental Fig. 3). To evaluate the functional significance of this putative NLS domain, the NLS-dependent translocation of TNL protein fusions with green fluorescent protein (GFP) was investigated in tobacco BY2-cells. One fusion contained the *P. tremula* homolog of the Poptr1_1:216113 NBS protein with its NLS, whereas the other protein fusion was constructed without the bipartite NLS domain. The fusion protein lacking the NLS was localized in the cytoplasm throughout the cell, whereas that with the NLS was located exclusively in the nucleus, demonstrating the functionality of the amino-terminal peptide in sequestering the R-protein in the nucleus. The presence of a NLS has been reported before in *Arabidopsis*

at the C-terminal part of the *RRS1 R* disease resistance protein that confers broad spectrum resistance to several strains of *Ralstonia solanacearum*, the causal agent of bacterial wilt (Lahaye 2002). However, no N-terminal NLS has been described for *R* genes in plant species other than *Populus* (Lescot et al. 2004). The *Populus NLS-TNLs* share the highest amino acid similarity with At5g36930 and except for a small N-terminal part containing the NLS, the intron/exon structure between these *NLS-TNLs* and *At5g36930* is well conserved, suggesting a common origin. All five *NLS-TNLs* were assembled in one clade in the phylogenetic tree indicating a more recent evolution. The presence of NLS and Zn-finger DNA-binding domains in several *R* proteins suggests that they may directly control DNA transcription of defense response genes. As shown recently, nuclear and cytoplasmic trafficking plays an important role in *R* gene mediated disease resistance (Shen et al. 2007, Burch-Smith et al. 2007).

Gene duplication is an important mechanism for the evolution of divergent protein functions. In *Arabidopsis*, the ancient duplication of the entire genome and subsequent chromosome rearrangements played a key role in *R* gene amplification (Richly et al. 2002; Meyers et al. 2003). In *Populus*, the ancient and recent segmental duplications had a major impact on several multigene families, including the MYB transcriptional factors and the receptor-like kinases (RLK) (Tuskan et al. 2006). In contrast, ancient segmental duplications and subsequent chromosomal rearrangements only accounted for 5% (20 duplicated genes) of the amplification of *NBS* genes in the progenitors of *Populus*. The combination of phylogenetic analyses and physical mapping of 402 *NBS* genes showed that they are organized in single-gene loci (37%), clusters and superclusters. The ratio of singletons to the total number of NBS genes in the *Populus* genome was slightly higher than that in the rice (24.1%; Zhou et al. 2004) and *Arabidopsis* (26.8%; Meyers et al. 2003) genomes. Within the clusters, genes of the two major clades, i.e., TIR- and non-TIR-types (e.g., BNL), are intermingled, although CC-types are rarely mixed with *TNL* genes (Figs. 3, 4). Most clusters contain genes from different classes (NL, BNL, TNL) and phylogenetic lineages, although several clusters (e.g., cluster XIX_1 and XIX_2) contain highly similar sequences in head-to-tail orientation (Fig. 4) suggesting an intralocus recombination giving rise to translocation of a sequence block. The latter mechanism has the capacity of changing the gene number in clusters and to generate paralogous chromosomal loci with a highly variable number of LRR repeats (Noel et al. 1999). The structural and genetic diversity that exists among NBS proteins in *Populus* is remarkable, and suggests that diversifying selection has played an important role in the evolution of R genes in *Populus*, as has been shown in a number of other plant species. The pattern of R gene and distribution of pair-wise genetic distances suggest a recent expansion of the family, in part due to tandem duplications (Meyers et al. 2003).

Although some resistance loci have been genetically identified in *Populus*, no disease resistance genes have yet been cloned from *Populus* (Cervera et al. 2001; Zhang et al. 2001; Yin et al. 2004; Jorge et al. 2005; Zhang and Li 2005). The linkage established with previously described resistance loci should facilitate the map-based cloning of these resistance loci. For example, the co-localization of *R* genes with known resistance loci uncovered a genetic linkage between the foliar rust, *Melampsora* × *columbiana,* the *MXC3* resistance locus and a region containing several genes having thaumatin-related and serine/threonine protein kinase domains (Yin et al. 2004), including the predicted proteins Poptr1_1:196571 and Poptr1_1:196533. No *NBS* genes are found in this region, located on LG IV (near position 66,000). On the other hand, the *P. trichocarpa* homologs of the *TNL* genes that were mapped to 0.6 centiMorgan from the *M. larici-populina MER* resistance locus in *P. deltoides* (Lescot et al. 2004) are scattered in different locations in the *P. trichocarpa* genome, including the supercluster XIX_1,2,3 (Poptr1_1:247642, Poptr1_1:809436) (Fig. 4; Yin et al. 2004) and several non-assembled scaffolds. The different *Populus* species may have evolved different clusters of *R*-genes through specific intralocus recombination and duplication events in regions such as the supercluster XIX_1,2,3. Further sequence analysis of such *NBS-LRR* rich clusters in different *Populus* genetic background will greatly help in understanding selection mechanisms that occur during long-term interactions with pathogens such as *Melampsora* spp.

By analyzing the *R* gene sequences within the *Populus* draft genome we have begun to make advances in understanding the mechanisms generating novel disease resistance specificities in this perennial tree. Trends emerging from studies in *Populus* and other plant families should allow both the definition and resolution of the processes and patterns that govern the function and evolution of this important class of plant disease resistance genes.

Acknowledgements The *Populus* genome sequence data was produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/). We thank Daniel Rokshar and the other members of the *Populus* sequencing project at the JGI, and Jerry Tuskan (Oak Ridge National Laboratory) for permission to use the draft genome sequence before publication. Thanks to Uffe Hellsten and Stephen DiFazio for the duplication dataset. Annegret Kohler was supported by Postdoctoral Fellowships from the INRA, the Région Lorraine and the EU POPYOMICS FP6 programme. Cécile Rinaldi was supported by a joint doctoral scholarship from the INRA and the Région Lorraine and Sébastien Duplessis was supported by a junior scientist support grant from the Région Lorraine. Marie Baucher is a

Research Associate from the Belgian National Funds for the Scientific Research (FNRS). This work received support from the INRA and the Belgian 'Fonds de la Recherche Fondamentale Collective' (No. 2.4524.02). The EST *ub48dpf09* containing the NLS region was kindly provided by Dr. Stefan Jansson (Umea Plant Genetic Center, Umeå, Sweden).

References

Ameline-Torregrosa C, Wang BB, O'Bleness MS et al (2007) Identification and Characterization of NBS-LRR Genes in the Model Plant *Medicago truncatula*. Plant Physiol Prev. doi: 10.1104/pp.107.104588

Aravind L (2000) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem Sci 25:421–423

Ayliffe MA, Lagudah ES (2004) Molecular genetics of disease resistance in cereals. Ann Bot 94:765–773

Bai J, Pennill LA, Ning J et al (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res 12:1871–1884

Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol 3:21–29

Bailey TL, Gribskov M (1998) Methods and statistics for combining motif match scores. J Comput Biol 5:211–221

Baumgarten A, Cannon S, Spangler R et al (2003) Genome-level evolution of resistance genes in *Arabidopsis thaliana*. Genetics 165:309–319

Belkhadir Y, Subramaniam R, Dangl JL (2004) Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Curr Opin Plant Biol 7(4):391–399

Burch-Smith TM, Schiff M, Caplan JL et al (2007) A novel role for the TIR domain in association with pathogen-cerived elicitors. PloS Biol 13:e68

Cervera MT, Storme V, Ivens B et al (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

Chini A, Grant JJ, Seki M et al (2004) Drought tolerance established by enhanced expression of the *CC-NBS-LRR* gene, *ADR1*, requires salicylic acid, EDS1 and ABI1. Plant J 38:810–822

Chini A, Loake GJ (2005) Motifs specific for the ADR1 NBS-LRR protein family in *Arabidopsis* are conserved among NBS-LRR sequences from both dicotyledonous and monocotyledonous plants. Planta 221:597–601

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833

Deslandes L, Olivier J, Peeters N et al (2003) Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. Proc Natl Acad Sci USA 100:8024–8029

Faigon-Soverna A, Harmon FG, Storani L et al (2006) A constitutive shade-avoidance mutant implicates TIR-NBS-LRR proteins in *Arabidopsis* photomorphogenic development. Plant Cell 18:2919–2928

Geelen DNV, Inzé DG (2001) A bright future for the Bright Yellow-2 cell culture. Plant Physiol 127:1375-1379

Jeanmougin F, Thompson JD, Gouy M et al (1998) Multiple sequence alignment with Clustal X. Trends Biochem Sci 23:403–405

Jorge V, Dowkiw A, Faivre-Rampant P et al (2005) Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. New Phytol 167:113–127

Karimi M, Inzé D, Depicker A (2002) GATEWAY^{TM} vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci 7:193–195

Kelleher CT, Chiu R, Shin H et al (2007) A physical map of the highly heterozygous Populus genome: integration with the genome sequence and genetic map and analysis of haplotype variation. Plant J 50:1063–1078

Lahaye T (2002) The *Arabidopsis RRS1-R* disease resistance gene-uncovering the plant's nucleus as the new battlefield of plant defense? Trends Plant Sci 7:425–427

Lescot M, Rombauts S, Zhang J et al (2004) Annotation of a 95-kb *Populus deltoides* genomic sequence reveals a disease resistance gene cluster and novel class I and II transposable elements. Theor Appl Genet 109:10–22

Liu JJ, Ekramoddoullah AKM (2003) Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl.ex.D.Don). Mol Gen Genomics 270:432–441

Liu JJ, Ekramoddoullah AKM (2007) The CC-NBS-LRR subfamily *in Pinus monticola*: targeted identification, gene EXPRESSION, and genetic linkage with resistance to *Cronartium ribicola*. Phytopathology 97:728–736

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164

McDowell JM, Woffenden BJ (2003) Plant disease resistance genes: recent insights and potential applications. Trends Biotech 21:178–183

Meyers BC, Dikerman AW, Michelmore RW et al (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:31732

Meyers BC, Morgante M, Michelmore RW (2002) TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. Plant J 32:77–92

Meyers BC, Kozik A, Griego A et al (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15:809–834

Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Curr Opin Plant Biol 8:129–134

Mondragón-Palomino M, Meyers BC, Michelmore RW et al (2002) Patterns of positive selection in the complete *NBS-LRR* gene family of *Arabidopsis thaliana*. Genome Res 12:1305–1315

Nagata T, Nemoto Y, Hasezawa S (1992) Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants. Int Rev Cytol 132:1-30

Noel L, Moores TL, van der Biezen EA et al (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. Plant Cell 11:2099–2111

Ralph S, Oddy C, Cooper D et al (2006) Genomics of hybrid poplar (*Populus trichocarpa* × *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences. Mol Ecol 15:1275–1297

Richly E, Kurth J, Leister D (2002) Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. Mol Biol Evol 19:76–84

Rinaldi C, Kohler A, Frey P et al (2007) Transcript profiling of poplar leaves upon infection with compatible and incompatible strains of the foliar rust *Melampsora larici-populina*. Plant Physiol 144:347–366

Schwartz S, Zhang Z, Frazer KA et al (2000) PipMakerA web server for aligning two genomic DNA sequences. Genome Res 10:577–586

Shen QH, Saijo Y, Mauch S et al (2007) Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. Science 315:1098–1103

Swofford DL (1999) PAUP*. Phylogenetic analysis using parsimony (*and other methods) Version 4. Sinauer Associates, Sunderland, Massachusetts

Tan S, Meyers BC, Kozik A et al (2007) Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. BMC Plant Biol 7:56

Tuskan G, DiFazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa*. Science 313:1596–1604

Wirthmueller L, Zhang Y, Jones JDG et al (2007) Nuclear accumulation of the *Arabidopsis* immune receptor RPS4 is necessary fr triggering EDS1-dependant defense. Curr Biol 17:2023–2029

Yin TM, DiFazio SP, Gunter LE et al (2004) Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence. New Phytol 164: 95–105

Zhang Y, Li X (2005) A putative nucleoporin 96 is required for both basal defense and constitutive resistance responses mediated by suppressor of npr1-1, constitutive 1. Plant Cell 17:1306–1316

Zhang J, Steenackers M, Storme V et al (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

Zhou T, Wang JQ, Chen JQ et al (2004) Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics 271:402–415