References

1 Cheung, V.G. *et al.* (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.* 33, 422–425
2 Oleksiak, M.F. *et al.* (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* 32, 261–266
3 Brem, R.B. *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755
4 Enard, W. *et al.* (2002) Intra- and interspecific variation in primate gene expression patterns. *Science* 296, 340–343
5 Cowles, C.R. *et al.* (2002) Detection of regulatory variation in mouse genes. *Nat. Genet.* 32, 432–437
6 Yan, H. *et al.* (2002) Allelic variation in human gene expression. *Science* 297, 1143
7 Lo, H.S. *et al.* (2003) Allelic variation in gene expression is common in the human genome. *Genome Res.* 13, 1855–1862
8 Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* 3, 662–673
9 Plath, K. *et al.* (2002) *Xist* RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* 36, 233–278
10 Tremblay, K.D. *et al.* (1995) A paternal-specific methylation imprint marks the alleles of the mouse *H19* gene. *Nat. Genet.* 9, 407–413
11 Thorvaldsen, J.L. *et al.* (1998) Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev.* 12, 3693–3702
12 Li, E. *et al.* (1993) Role for DNA methylation in genomic imprinting. *Nature* 366, 362–365
13 Sakatani, T. *et al.* (2001) Epigenetic heterogeneity at imprinted loci in normal populations. *Biochem. Biophys. Res. Commun.* 283, 1124–1130
14 Bray, N.J. *et al.* (2003) *Cis*-acting variation in the expression of a high proportion of genes in the human brain. *Hum. Genet.* 113, 149–153
15 Yvert, G. *et al.* (2003) *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* 35, 57–64
16 Schadt, E.E. *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302
17 Hoogendoorn, B. *et al.* (2003) Functional analysis of human promoter polymorphisms. *Hum. Mol. Genet.* 12, 2249–2254
18 Singer-Sam, J. *et al.* (1992) Parental imprinting studied by allele-specific primer extension after PCR: paternal X chromosome-linked genes are transcribed prior to preferential paternal X chromosome inactivation. *Proc. Natl. Acad. Sci. U. S. A.* 89, 10469–10473
19 Singer-Sam, J. *et al.* (1992) A sensitive, quantitative assay for measurement of allele-specific transcripts differing by a single nucleotide. *PCR Methods Appl.* 1, 160–163
20 Knight, J.C. *et al.* (2003) *In vivo* characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat. Genet.* 33, 469–475
21 Norton, N. *et al.* (2002) Universal, robust, highly quantitative SNP allele frequency measurements in DNA pools. *Hum. Genet.* 110, 471–478
22 King, M.C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116

0168-9525/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tig.2004.01.001

Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes

Dario Leister

Abteilung für Pflanzenzüchtung und Ertragsphysiologie, Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné Weg 10, D-50829 Köln, Germany

NBS-LRR genes are the major class of disease resistance genes in flowering plants, and are arranged as single genes and as clustered loci. The evolution of these genes has been investigated in *Arabidopsis thaliana* by combining data on their genomic organisation and position in phylogenetic trees. Tandem and segmental duplications distribute and separate NBS-LRR genes in the genome. It is, however, unclear by which mechanism(s) NBS-LRR genes from different clades are sampled into heterogeneous clusters. Once physically removed from their closest relatives, the NBS-LRR genes might adopt and preserve new specificities because they are less prone to sequence homogenization.

Plant resistance (*R*) genes mediate phenotypic resistance against pests and pathogens expressing avirulence genes (a situation known as 'gene-for-gene' interaction). Genes that encode proteins containing a nucleotide-binding site (NBS) and C-terminal leucine-rich repeats (LRRs) represent the largest class of *R* genes in flowering plants [1]; NBS-LRR genes also exist in gymnosperms, non-vascular plants and mammals [2–4]. Based on their N-termini, two subclasses of NBS-LRR resistance proteins are known: the first is characterized by the TIR-domain homologous to the *Drosophila* Toll and mammalian Interleukin-1 receptors, and the second is characterized by a coiled-coil (CC) structure. Truncated versions of NBS-LRR genes exist encoding proteins that lack either a domain close to the N-terminal of the NBS, or the LRR region, or consist only of a TIR-domain. In grass species, TIR-NBS-LRR genes have not yet been identified, but the CC-type is very common [5].

In different plants, NBS-LRR loci are found both as isolated genes (singletons) and as tightly linked arrays of related genes (gene clusters) [6]. In some cases, gene clusters contain copies of NBS-LRR genes from different phylogenetic clades [HETEROGENEOUS CLUSTERS (see Glossary)] [7]. Before the complete sequences of plant genomes became available, analyses of NBS-LRR gene evolution in diverse species were based on relatively few loci. Recently, with the complete sequence of the genome of *Arabidopsis thaliana* available, several groups have carried out genome-wide analyses of the organisation and evolution of NBS-LRR genes [8–12]. The distribution of NBS-LRR genes in the genome has been, in general,

Corresponding author: Dario Leister (leister@mpiz-koeln.mpg.de).

Glossary

Heterogeneous cluster: gene cluster that contains genes from different gene sub-families, indicating an origin that is different from tandem duplication.
Birth–death model: interprets the expansion or contraction of gene clusters as the result of unequal crossover and the evolution of individual genes as the result of diversifying selection.
Diversifying selection: type of selection that leads to an increase in genetic diversity.
Segmental duplication: copying of entire blocks of genes from one chromosome to another. This event leads to duplication of genes to unlinked sites, even when a segment is translocated to the same chromosome. Segmental duplications can follow whole-genome duplication events.
Ectopic duplication: duplication of individual or small groups of genes to an unlinked locus; a small-scale event compared with segmental duplications. Candidate mechanisms for ectopic duplication are recombination events between homologous sequences at unlinked sites, or the effects of the action of transposable elements.
Concerted evolution and/or sequence homogenization: process by which a series of nucleotide sequences or different members of a gene family remain similar or identical through time.
Tandem duplication: type of duplication where the duplicated segment is contiguous with the original duplication.
Clade: group of species, organisms, genes or proteins comprising a common ancestor and all of its descendants.
Phylogeography: a branch of evolutionary biology, also called biogeography that compares species phylogenies to the movement of continents. The order and relationship of species from the different continents in the phylogenetic trees reflects the effect of time passed after a continent splits. Analogously, in genome evolution, duplications and rearrangements can move homologous gene sequences to different regions, similar to how continental drift affects species. In this view, genome segments can be treated as geographic populations and analysed using phylogeographic approaches.
Synteny: the occurrence of genomic co-linearity between homologous genes in different organisms.
Non-synonymous nucleotide substitution: results in changes at the amino acid level, whereas synonymous substitution does not. During diversifying selection the rate of non-synonymous substitution is higher than the rate of synonymous substitution, leading to amino acid diversity.

explained by tandem gene duplications and by duplication of individual or small groups of genes to unlinked loci (ectopic duplication) [8,9]. In addition, recombination and homogenization have played major roles in the distribution of NBS-LRR genes [7]. BIRTH–DEATH MODELS have also been proposed, emphasizing the importance of inter-allelic sequence exchange and DIVERSIFYING SELECTION [13] (Box 1). Recently, Baumgarten *et al.* [10] have suggested that most of the genomic dispersion of NBS-LRR genes originates from duplication and translocation of entire chromosomal segments (SEGMENTAL DUPLICATION), rather than from small-scale ECTOPIC DUPLICATION events. The authors also suggested that, when physically separated from their closest relatives, NBS-LRR genes might adopt and preserve new functions by escaping SEQUENCE HOMOGENIZATION occurring as a result of recombination. Because other gene families in *Arabidopsis* show analogous genomic organisations, for example, those encoding cytochrome P450 proteins [14], UDPG-glycosyltransferases [15], receptor-like kinases [16] and also the mammalian fibroblast growth factor (FGF) gene family [17], mechanisms of tandem and segmental duplication, in combination with recombinational isolation, are supposed to function, in general, in the diversification of gene families.

Genome-wide organisation of NBS-LRR genes

The genome of *A. thaliana* contains >150 NBS-LRR genes [8–10] (Figure 1), the precise number depending on the model of gene annotation. Richly *et al.* have listed a total of 166 NBS-LRR sequences, including 33 truncated sequences, arranged as 51 singletons and 40 clusters [8]. Using extensive manual re-annotation of the genomic sequence of the same species, Meyers *et al.* have detected 149 NBS-LRR genes and 58 truncated genes; the 149 non-truncated genes are distributed as 40 singletons and 43 clusters [9]. In *A. thaliana*, TIR-NBS-LRR genes outnumber CC-NBS-LRR genes, indicating either a recent amplification of the former family or loss of the latter family of genes [8,9,18]. Similar to the situation in other plants, *Arabidopsis* NBS-LRR gene loci are not evenly distributed in the genome: superclusters exist on

Box 1. Chromosomal organisation of *R* genes and mechanisms contributing to their evolution

R loci, consisting of only a single gene, can have multiple and functionally distinct alleles, recognizing different varieties of a pathogen (e.g. the flax *L* locus and the *Rpp13* locus of *Arabidopsis*), or only one resistant allele (e.g. *Rpm1* and *Rps2* of *Arabidopsis thalina*). Other *R* genes belong to families of tightly linked genes (clusters), some of which include copies of *R* genes from different clades.

Genetic recombination between alleles or between related sequences contributes to genetic variation (Figure Ia–c). Duplication can increase the number of genes and disperse them in the genome (Figure Id–f). Selection of mutation-induced diversity can favour the generation of novel specificities or promote the linkage of *R* genes acting in the same pathway (Figure Ig–i). The figure illustrates how intragenic crossover can generate novel alleles with different specificities [23] (Figure Ia). Unequal crossover (symbolized by a distorted cross) has the potential to change the number of family members in *R* gene clusters and rearrange them into new combinations [24] (Figure Ib). Such crossovers can be intragenic or intergenic. In addition, the repeated action of equal and unequal recombination within a clustered gene family can homogenize them (a phenomenon known as CONCERTED EVOLUTION) [25,26] (Figure Ic). In some *R* gene clusters, unequal recombination occurs frequently (e.g. in the *Rp1* and *Rp3* gene clusters of maize), whereas in others it is rare (e.g. in *Dm3* of lettuce and *Pto* of tomato) [13]. As a consequence, at loci similar to *Dm3* and *Pto*, orthologous genes from two different lines are more similar to each other than they are to paralogous genes within the same cluster. Duplication of genes can occur as transfer of the duplicated segment to a site contiguous to the original one (tandem duplication; Figure Id), or it can involve the duplication of large stretches of DNA containing many genes (segmental duplication; Figure Ie). During ectopic duplication (Figure If), individual or a small groups of genes are duplicated to unlinked sites by uncharacterized mechanisms. It has been suggested that heterogeneous clusters might be derived from ectopic recombination between different *R* gene loci [27], thereby uncoupling NBS-LRR genes from syntenic relationships. The concept of ectopic duplication is supported by comparative mapping of NBS-LRR genes in cereals [28] but not by the chromosomal locations of *R* gene loci in different *Solanaceae* species [29].

In addition to genetic recombination and duplication events, selective forces play a major role in *R* gene evolution. Selection should favour the linkage of *R* genes participating in the recognition of the same pathogen. This could be the case for the *Pto* locus of tomato, which contains copies of two different classes of *R* genes (*Pto* and *Prf*), operating in the same resistance pathway [30] (Figure Ig). A different type of selection generates novel or diverse recognition capabilities of *R* genes when non-synonymous substitutions are favoured over synonymous exchanges as a result of diversifying selection (Figure Ih). Diversifying selection can act on LRRs [21,24,31,32] and on other domains [33,34] of NBS-LRR genes. In addition, the absence of *R* gene alleles in certain susceptible cultivars might result from a reduced fitness associated with the expression of *R* proteins in the absence of pathogens (e.g. *Rpm1* [35]) (Figure Ii).

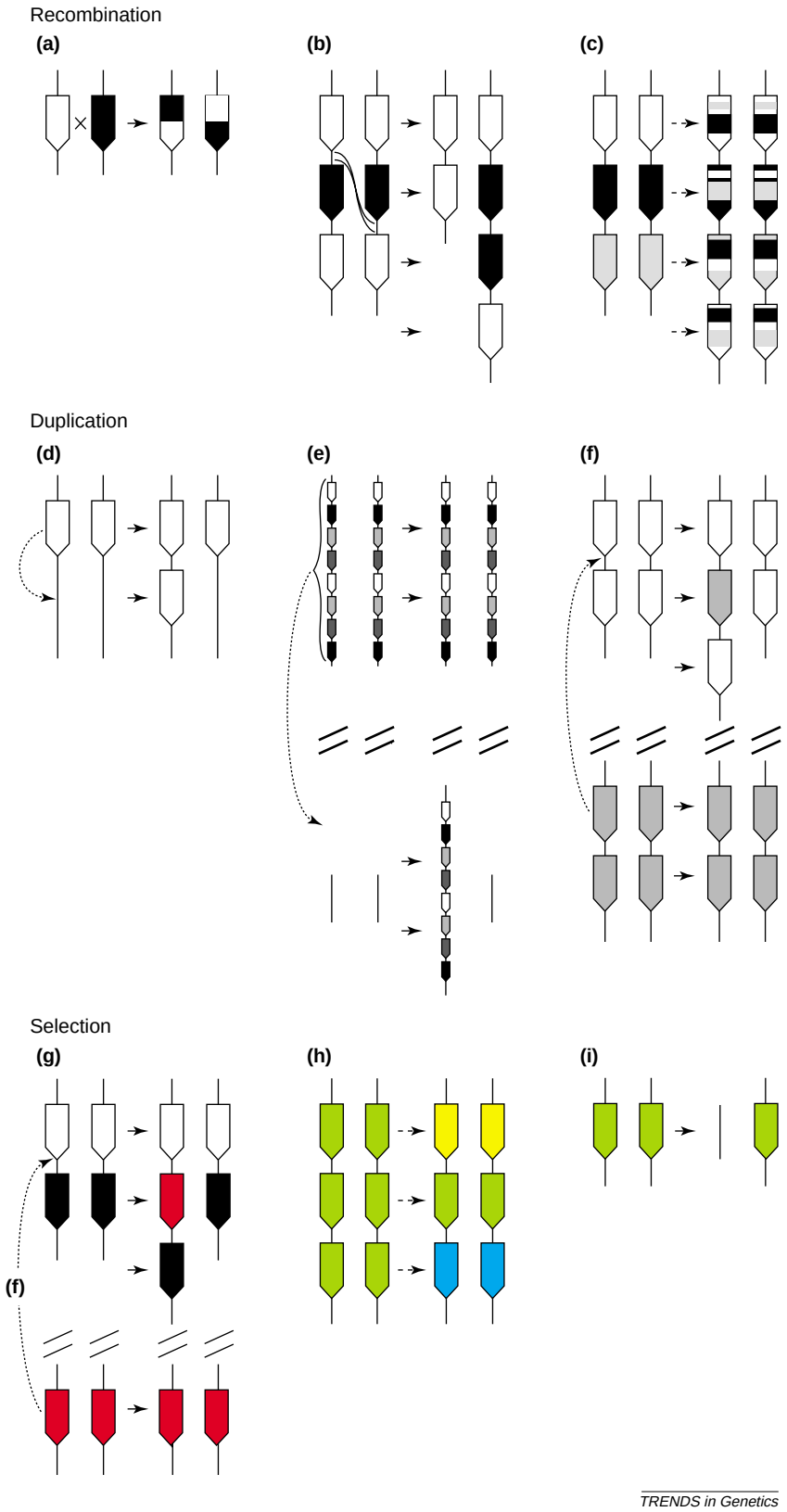


Figure I. The mechanisms that contribute to diversity of *R* genes and their loci. **(a)** Intragenic (equal) crossing over leads to domain swaps in the protein. **(b)** Unequal crossover changes the number of genes in a cluster. **(c)** Sequence homogenization (concerted evolution), a multi-step process (indicated by the broken arrow), results from the repeated action of (a) and (b). **(d)** Tandem duplication in which the copy is contiguous to the original copy. **(e)** Segmental duplication involves the duplication of entire chromosomal regions. **(f)** Ectopic duplication transfers individual or small sets of genes to unlinked sites. **(g)** Positive selection for linkage of genes participating in the same defence pathway. This should result in the linkage of different types of *R* genes and can be interpreted as the result of ectopic duplication (f) followed by gene loss at the original locus (not shown; see i). **(h)** Diversifying selection, also a multi-step process (indicated by the broken arrow), which increases the genetic diversity and antagonizes concerted evolution (c). **(i)** Gene loss because of negative selection in the absence of pathogens as a result of a reduced fitness associated with the expression of *R* proteins.

chromosome 1 and 5 (Figure 1), whereas chromosomes 2 and 3 are relatively deficient in NBS-LRR genes [8,9].

Phylogenetics meets genomics

The phylogenetic analysis of Richly *et al*. and the combined analysis of phylogenetic trees, protein motifs and intron positions by Meyers *et al*. have distinguished nine (seven TIR and two CC) [8] and twelve (eight TIR and four CC) [9] clearly distinguishable clades of NBS-LRR genes, respectively (Figure 1). When the phylogenies were projected on the *Arabidopsis* physical map, most of the ~40 clusters consisted of genes of the same phylogenetic lineage, suggesting that TANDEM DUPLICATION of these DNA sequences had occurred. However, ten [8] and eleven [9] clusters, respectively, contained genes from different clades. Seven heterogeneous clusters were interpreted as the amplification products of an ancient two-gene module [8,9] (Figure 1). For the remaining four clusters, two origins were suggested: (i) association of NBS-LRR genes from different clades by chance (if a large numbers of related genes exists in a genome, some should be found in close vicinity) [8,9]; or (ii) a mechanism that samples NBS-LRR genes of different clades into clusters [9]. The latter would involve recombination events between either conserved stretches of otherwise dissimilar NBS-LRR genes or their flanking regions, if they exhibit a certain degree of homology. It is, however, somewhat unrealistic to consider heterogeneous clusters containing NBS-LRR genes from more than two different clades (for example Figure 1c) as deriving solely from chance association.

Duplication and recombination of NBS-LRR genes in *Arabidopsis*

The genomic analyses in *A. thaliana* present a picture of NBS-LRR gene evolution in which the organisation of NBS-LRR genes in arrays of members of the same CLADE is mainly a result of tandem duplications [8,9]. However, duplication events to distant positions have also been observed [8,9]. Some of these rearrangements could be associated with segmental duplications of entire chromosomal regions, when large segments containing NBS-LRR genes were moved to new positions, even on different chromosomes. The movement of singletons or clusters to unlinked sites via ectopic duplication [8,9], which, in contrast to segmental duplication, is a small-scale event translocating only few genes, and deletion of genes after segmental duplications [9,19,20] appear to be more frequent than segmental events.

The contention that duplication of NBS-LRR genes to unlinked sites might occur independently of segmental duplication was recently challenged [10]. Baumgarten *et al*. employed a PHYLOGEOGRAPHIC approach, treating chromosomal regions as geographic populations, to determine the evolutionary mechanisms by which NBS-LRR genes were dispersed across the *Arabidopsis* genome. Of the 89 duplication events resolved in their analyses, 71 were classified as 'local', given that the duplicated genes were located <2-Mb apart. Of the remaining 18 'non-local' events, 15 fell within segmentally duplicated regions and only three duplications were candidates for an ectopic origin. In addition, using an association test, Baumgarten *et al*. have shown that colocation of sequence-divergent NBS-LRR genes across different genomic regions indicates a duplication of chromosomal segments rather than of individual genes [10]. Future analyses have to clarify the genetic mechanism(s) besides tandem duplication that distribute NBS-LRR genes within the 'local' 2-Mb intervals that were considered by Baumgarten and coworkers. Because in *Arabidopsis* a 2-Mbp stretch of DNA contains ~400 genes, such 'local' events also include the duplication of single or small groups of genes to non-contiguous sites (Figure 1c), resembling or corresponding to the mechanism defined by Richly *et al*. [8] and Myers *et al*. [9] as 'ectopic' duplication. A realistic possibility is that similar genetic mechanisms might have driven 'local' (but non-tandem) and 'non-local' duplication of NBS-LRR genes: certain heterogeneous clusters (indicated by the light-green and dark-green arrowheads in Figure 1c) have been duplicated locally and to unlinked chromosomal regions.

Regardless of the uncharacterized genetic mechanism(s) responsible for these local non-tandem duplications, the data of Baumgarten *et al*. suggest that frequent ectopic translocation of NBS-LRR genes among different chromosomes does not occur. Therefore, the syntenic relationship between orthologous NBS-LRR genes in different *Arabidopsis* accessions (or ecotypes), including closely related species, should not have been markedly affected by the amplification of NBS-LRR genes. When a loss of SYNTENY is observed, it might have originated from gene deletion rather than from ectopic events.

Diversification of NBS-LRR gene sequences

Depending on the plant species and the NBS-LRR gene clade, several mechanisms have been suggested as being responsible for the diversification of NBS-LRR gene sequences: (i) intralocus crossovers; (ii) sequence exchange between unlinked loci (ectopic recombination); and (iii) mechanisms of selection (Box 1). Baumgarten *et al*. have also investigated the frequency of recombination between NBS-LRR genes of the same clade and between those of different clades; no significant sequence-exchange events were detected among members of different NBS-LRR clades [10]. However, many more recombination events were detected between sequences of the same clade when they were located in the same 2-Mb region. Sequence exchanges were also evident among NBS-LRR genes of the same clade that were located in different genomic regions, but they were less frequent. Most of these recombination events were restricted to genes found in duplicated regions of the genome, suggesting that the responsible sequence exchange occurred before the segmental duplication of the chromosomal regions concerned [10].

Towards a model for NBS-LRR evolution

Clusters of closely related genes originate from tandem gene duplications. But how are heterogeneous clusters, which are common in *Arabidopsis* and in other species, generated? In *Arabidopsis*, almost all heterogeneous clusters contain sequences belonging to the major gene clades. Therefore, it seems unlikely that heterogeneous clusters are derived from diversification within homogenous clusters as a result of diversifying selection. According to Baumgarten *et al*. [10],

ectopic duplication of individual NBS-LRR sequences, a mechanism that breaks synteny, appears to be rare in *Arabidopsis*, if not absent. This notion favours the idea that heterogeneous clusters might originate from chance associations of distantly related genes, rather than from ectopic recombination events between conserved stretches of otherwise dissimilar NBS-LRR genes. Moreover, if sequence exchange among NBS-LRR genes of different clades does not occur (as claimed in [9] and [10]), heterogeneous clusters should, once generated, persist because the dissimilar genes that they contain are not threatened by sequence homogenization. It is also possible that linkage of NBS-LRR genes from different clades, which mediate resistance against the same or even different pathogens, is under positive selection because their co-segregation increases the total fitness of the plant [7].

In most cases where NBS-LRR genes are closely related and physically linked, intergenic exchange is frequent and should lead to sequence homogenization. Nonetheless, the ratio of synonymous versus NON-SYNONYMOUS NUCLEOTIDE SUBSTITUTION found in NBS-LRR gene family members [12] points to the effects of a strong diversifying selection, which would make sequences less similar and so would tend to antagonize sequence homogenization. A novel mechanism to escape the maelstrom of sequence homogenization has been presented by Baumgarten *et al*. [10]: the duplication of NBS-LRR sequences to unlinked chromosomal regions where they are out of reach of intergenic-sequence exchange. In this context, segmental duplications might play an important, and previously unrecognized, role in the generation of novel gene functions, promoting the generation of new *R* gene specificities. In this 'recombinational isolation' scenario, physically isolated single-gene loci play a major role in the evolution of *R* specificities. This is why future studies will have to test the recombination frequencies and the generation of novel *R* specificities at such loci. Because the genomic organisation of other multigene families in *Arabidopsis* and in other species has not been found to be dramatically different from that of NBS-LRR genes, such mechanisms of recombinational isolation could function, more universally, in the diversification of large gene families.

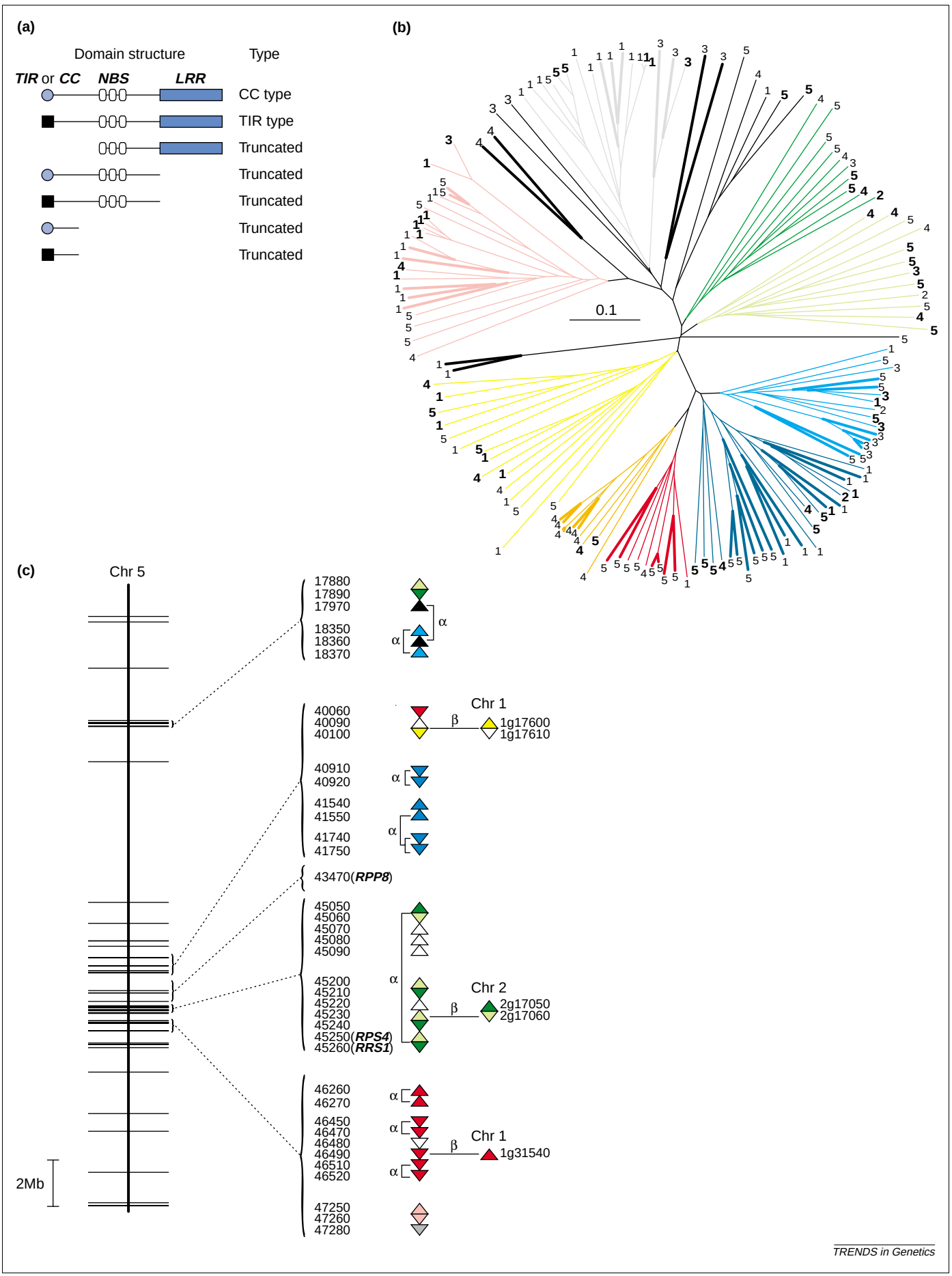
Concluding remarks

Because the function and mode of action remain to be discovered for most *Arabidopsis* NBS-LRR genes, it will therefore be important in the future to attempt the integration of function analyses, population dynamics and genome evolution. With respect to the dissection of the phylogeny of NBS-LRR genes, the genomic analysis of the Col-0 accession of *Arabidopsis* provides significant evolutionary information. The limitation of sequence data from only a single accession of *Arabidopsis* is that it does not enable the quantification of exchanges between alleles or loci or the reconstruction of deletions. Systematic intergenomic comparisons, using the DNA sequences from other accessions of *Arabidopsis* [9,21], will facilitate measurements of the relative frequency with which rearrangements, duplications and deletions occur. In addition, the analysis of NBS-LRR gene organisation in other plant genomes will help to determine whether the model developed in *Arabidopsis* is unique or can be used as a paradigm for NBS-LRR evolution. Obvious candidates for such comparisons are other dicotyledonous plants such as *Medicago*, and the monocot rice, which contains a large and diverse set of NBS-LRR genes [22].

Acknowledgements

I thank Francesco Salamini, Christos Noutsos, Andrew Baumgarten, Georgiana May and Jeff Dangl for stimulating discussions and comments on the manuscript. This work was supported by the Max-Planck-Gesellschaft (MPG) and a Heisenberg fellowship provided by the Deutsche Forschungsgemeinschaft (DFG).

Figure 1. Phylogeny and genomic organisation of nucleotide binding site C-terminal leucine-rich repeat (NBS-LRR) genes in the genome of *Arabidopsis thaliana*. **(a)** Domain structures of the coiled-coil- (CC) and Toll and Interleukin 1 receptor- (TIR) type NBS-LRR genes and of their truncated versions. The NBS domain is defined by a tripartite motif. **(b)** Phylogenetic analysis of NBS-LRR protein sequences of *A. thaliana*. Analysis of all NBS-LRR protein sequences listed at the Plant Gene Family Evolution Page (http://www.tc.umn.edu/~cann0010/genefamilyevolution/index.html) was based on distance-matrix (neighbour-joining) analyses [MEGA version 2.1 (http://www.megasoftware.net/), bootstrap of 1000] after alignment by ClustalX [36]. The numbers (1–5) refer to the chromosome on which the corresponding gene is located and the branch lengths reflect the estimated number of substitutions per 100 sites. The branch colours refer to those used by Richly *et al*. [8] to highlight different clades. Most closely related genes can be derived from tandem duplications and are positioned in the same chromosomal region (indicated by thick lines) or they can be located in different regions of the same, or even a different, chromosome (indicated by bold numbers) as a result of segmental or ectopic duplication events. **(c)** Examples of NBS-LRR gene clusters and gene phylogenies on chromosome 5 of *A. thaliana*. The colours indicate membership of the corresponding clades as shown in (b), those clusters depicted by white arrowheads refer to truncated NBS-LRR genes. On the left hand side, the locations of NBS-LRR genes on chromosome 5 of *A. thaliana* (accession Col-0) are indicated with horizontal lines. At the right hand side, the accession numbers of NBS-LRR genes in some selected clusters of chromosome 5, their orientation and their clade membership are shown. The 5-digit number refers to the MIPS database protein entry code (http://mips.gsf.de/proj/thal/db/index.html), omitting the first four original digits (i.e. 17880 refers to MIPS protein entry At5 g17880-*Arabidopis thaliana* Chromosome 5 17880). NBS-LRR genes are depicted as arrowheads directed towards the 3′-end of genes, and clusters as contiguous arrowheads, whereby linked NBS-LRR genes are grouped into clusters when they were not interrupted by more than 8 other open reading frames encoding non-NBS-LRR proteins. Note that many of the clusters are real tandem arrays and not interspersed with non-NBS-LRR genes (adjacent genes usually differ in their accession numbers by '10'). Heterogeneous clusters contain NBS-LRR genes from at least two different clades: in the case of *At5 g17880–At5 g17970*, genes from three different clades were found at the same locus. The *R* genes with known resistance to particular pathogens such as *RPS4* [37], *RRS1* [38] and *RPP8* [39] are also indicated. Recent duplications found by Richly *et al*. [8] and Meyers *et al*. [9] are indicated by lines joining the genes concerned. According to Baumgarten *et al*., 'α' indicates local duplications and 'β' duplications to unlinked sites, which are associated with segmental duplications [10]. Interestingly, *At5 g46490* and *At1 g31540* fall in blocks resulting from a relatively ancient duplication event [19], whereas their relatively low sequence divergence suggests either a recent duplication event, or recent sequence homogenization. Note that in the case of *At5 g45050–At5 g45060* and *At5 g45250*(*RPS4*)*–At5 g45260*(*RRS1*), and of *At5 g40090–At5 g40100* and *At1 g17600–At1 g17610*, pairs of genes (modules) and not individual NBS-LRR genes have been duplicated. The physical distance of NBS-LRR genes should have profound effects on the degree of sequence exchange among them. According to Baumgarten *et al*. [10], recombination between *At5 g17880*, which is located on chromosome 5, and *At2 g17060*, which is located on chromosome 2, should be low because they are located on different chromosomes. Recombination will also be low between At5 *g17880* and the other genes of the same cluster (i.e At5 g17890 and At5 g17970) because they belong to different clades. By contrast, frequent sequence exchange should occur between genes *At5 g40910–At5 g41750* (indicated in blue) and *At5 g46260–At5 g46520* (indicated in red), because these genes are located in the same chromosomal region and belong to the same clade.


(a)
Domain structure
Type
TIR or CC
NBS
LRR
CC type
TIR type
Truncated
Truncated
Truncated
Truncated
Truncated
(b)
0.1
(c)
Chr 5
17880
17890
17970
18350
18360
18370
α
40060
40090
40100
β
Chr 1
1g17600
1g17610
40910
40920
41540
41550
41740
41750
43470(RPP8)
45050
45060
45070
45080
45090
45200
45210
45220
45230
45240
45250(RPS4)
45260(RRS1)
Chr 2
2g17050
2g17060
46260
46270
46450
46470
46480
46490
46510
46520
Chr 1
1g31540
47250
47260
47280
2Mb
TRENDS in Genetics

References

1 Dangl, J.L. and Jones, J.D. (2001) Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833
2 Meyers, B.C. *et al.* (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332
3 Akita, M. and Valkonen, J.P. (2002) A novel gene family in moss (*Physcomitrella patens*) shows sequence homology and a phylogenetic relationship with the TIR-NBS class of plant disease resistance genes. *J. Mol. Evol.* 55, 595–605
4 Bertin, J. *et al.* (1999) Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-κB. *J. Biol. Chem.* 274, 12955–12958
5 Pan, Q. *et al.* (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* 50, 203–213
6 Holub, E.B. (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2, 516–527
7 Hulbert, S.H. *et al.* (2001) Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* 39, 285–312
8 Richly, E. *et al.* (2002) Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Mol. Biol. Evol.* 19, 76–84
9 Meyers, B.C. *et al.* (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15, 809–834
10 Baumgarten, A. *et al.* (2003) Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics* 165, 309–319
11 The Arabidopsis Genome Initiative, (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
12 Mondragon-Palomino, M. *et al.* (2002) Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res.* 12, 1305–1315
13 Michelmore, R.W. and Meyers, B.C. (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 8, 1113–1130
14 Schuler, M.A. and Werck-Reichhart, D. (2003) Functional genomics of P450s. *Annu. Rev. Plant Biol.* 54, 629–667
15 Paquette, S. *et al.* (2003) On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62, 399–413
16 Shiu, S.H. and Bleecker, A.B. (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol.* 132, 530–543
17 Ornitz, D.M. and Itoh, N. (2001) Fibroblast growth factors. *Genome Biol.* 2 REVIEWS30051
18 Cannon, S.B. *et al.* (2002) Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *J. Mol. Evol.* 54, 548–562
19 Blanc, G. *et al.* (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis genome*. *Genome Res.* 13, 137–144
20 Wolfe, K.H. (2001) Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* 2, 333–341
21 Noel, L. *et al.* (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. *Plant Cell* 11, 2099–2112
22 Bai, J. *et al.* (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res.* 12, 1871–1884
23 Ellis, J.G. *et al.* (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* 11, 495–506
24 Parniske, M. *et al.* (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. *Cell* 91, 821–832
25 Hickey, D.A. *et al.* (1991) Concerted evolution of duplicated protein-coding genes in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 88, 1611–1615
26 Walsh, J.B. (1987) Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? *Genetics* 117, 543–557
27 Parniske, M. and Jones, J.D. (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5850–5855
28 Leister, D. *et al.* (1998) Rapid reorganization of resistance gene homologues in cereal genomes. *Proc. Natl. Acad. Sci. U. S. A.* 95, 370–375
29 Pan, Q. *et al.* (2000) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and *Arabidopsis*. *Genetics* 155, 309–322
30 Salmeron, J.M. *et al.* (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell* 86, 123–133
31 Wang, G.L. *et al.* (1998) *Xa21D* encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. *Plant Cell* 10, 765–779
32 Meyers, B.C. *et al.* (1998) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell* 10, 1833–1846
33 Luck, J.E. *et al.* (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell* 12, 1367–1377
34 Dodds, P.N. *et al.* (2001) Six amino acid changes confined to the leucine-rich repeat β-strand/β-turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. *Plant Cell* 13, 163–178
35 Tian, D. *et al.* (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423, 74–77
36 Thompson, J.D. *et al.* (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882
37 Gassmann, W. *et al.* (1999) The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant J.* 20, 265–277
38 Deslandes, L. *et al.* (2002) Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive *RRS1-R* gene, a member of a novel family of resistance genes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2404–2409
39 McDowell, J.M. *et al.* (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. *Plant Cell* 10, 1861–1874

0168-9525/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tig.2004.01.007

Genome Analysis

Fixation biases affecting human SNPs

Matthew T. Webster and Nick G.C. Smith

Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

Under neutrality all classes of mutation have an equal probability of becoming fixed in a population. In this article, we describe our analysis of the frequency distributions of >5000 human SNPs and provide evidence of biases in the process of fixation of certain classes of point mutation that are most likely to be attributable to biased gene conversion. The results indicate an increased fixation probability of mutations that result in the incorporation of a GC base pair. Furthermore, in transcribed regions this process exhibits strand

Corresponding author: Matthew T. Webster (matthew.webster@ebc.uu.se).