

Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny

Milind B. Ratnaparkhe^{1*}, Xiyin Wang^{1,2*}, Jingping Li¹, Rosana O. Compton¹, Lisa K. Rainville¹, Cornelia Lemke¹, Changsoo Kim¹, Haibao Tang¹ and Andrew H. Paterson¹

¹Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA; ²Center for Genomics and Computational Biology, School of Life Sciences, School of Sciences, Hebei United University, Tangshan, Hebei 063000, China

Summary

Author for correspondence:

Andrew H. Paterson

Tel: +1 706 583 0162

Email: paterson@uga.edu

Received: 24 March 2011

Accepted: 16 May 2011

New Phytologist (2011) **192**: 164–178

doi: 10.1111/j.1469-8137.2011.03800.x

Key words: disease resistance, evolution, gene conversion, gene synteny, peanut (*Arachis hypogaea*), transposons.

- Plant genomes contain numerous disease resistance genes (R genes) that play roles in defense against pathogens. Scarcity of genetic polymorphism makes peanut (*Arachis hypogaea*) especially vulnerable to a wide variety of pathogens.
- Here, we isolated and characterized peanut bacterial artificial chromosomes (BACs) containing a high density of R genes. Analysis of two genomic regions identified several TIR-NBS-LRR (Toll-interleukin-1 receptor, nucleotide-binding site, leucine-rich repeat) resistance gene analogs or gene fragments. We reconstructed their evolutionary history characterized by tandem duplications, possibly facilitated by transposon activities. We found evidence of both intergenic and intragenic gene conversions and unequal crossing-over, which may be driving forces underlying the functional evolution of resistance.
- Analysis of the sequence mutations, protein secondary structure and three-dimensional structures, all suggest that LRR domains are the primary contributor to the evolution of resistance genes. The central part of LRR regions, assumed to serve as the active core, may play a key role in the resistance function by having higher rates of duplication and DNA conversion than neighboring regions. The assumed active core is characterized by significantly enriched leucine residue composition, accumulation of positively selected sites, and shorter beta sheets.
- Homologous resistance gene analog (RGA)-containing regions in peanut, soybean, *Medicago*, *Arabidopsis* and grape have only limited gene synteny and microcollinearity.

Introduction

Cultivated peanut (*Arachis hypogaea* L.) is grown on 25.5 million ha with a total global production of c. 35 million tons and thus ranks among the top five oilseed crops in the world alongside soybean, cottonseed, rapeseed, and sunflower. Peanut is widely used as a food and cash crop by resource-poor farmers in Africa and Asia to produce edible oil, and for human and animal consumption. Although productivity of peanuts in Asia (1.8 tons ha⁻¹) exceeds the world average, it is still lower than the yields in developed countries (3 tons ha⁻¹). One of the main reasons for low

productivity of this crop in these regions is the exposure of the crop to severe abiotic and biotic stresses.

Cultivated peanut is considered an allotetraploid ($2n = 4x = 40$) originated from a single hybridization event between two wild diploids with A and B genomes. Conservation of genome macrostructure (macrosynteny) has been reported between the respective subgenomes of peanut (Burow *et al.*, 2001; Moretzsohn *et al.*, 2009) and between peanut and other legumes including soybean *Glycine max* (*Gm*), *Medicago truncatula* (*Mt*) and *Lotus japonicus* (*Lj*) (Zhu *et al.*, 2005; Hougaard *et al.*, 2008; Bertoli *et al.*, 2009). A growing number of microsynteny studies describe similarities at the scale of individual bacterial artificial chromosome (BAC) clones or clone contigs between other legume genomes, but microsynteny data for peanut have

*These authors contributed equally to this work.

been lacking. Establishing microcollinearity and microsynteny of peanut with model legumes is critical since its complex genome structure is not well understood and defined.

Scarcity of genetic polymorphism makes peanut especially vulnerable to a wide variety of pathogens. Peanut is vulnerable to viral, fungal, bacterial and nematode pathogens such as groundnut rosette virus, rust (*Puccinia arachidis* Speg.), web blotch (*Phoma arachidicola* Marasas, Pauer & Boerema), early leaf spot (*Cercospora arachidicola* S. Hori), late leaf spot (*Cercosporidium personatum* Berk. & M.A. Curtis), bacterial wilt (*Pseudomonas solanacearum*) and root-knot nematode (*Meloidogyne arenaria* (Neal) Chitwood) (Phipps & Porter, 1998; Branch & Breneman, 1999; Dwivedi *et al.*, 2003; Pensuk *et al.*, 2004).

In the last several years, many different disease resistance genes (R genes) have been cloned from a variety of plant species. Some of the best characterized R-gene proteins comprise an N-terminal coil-coil (CC) or Toll-interleukin-1 receptor (TIR) homology domain, a centrally located nucleotide-binding site (NBS), and C-terminal leucine-rich repeats (LRRs). TIR and CC NBS-LRR genes are generally present in plant genomes as multigene families, often in complex clusters. Clusters of R genes have been reported in several legumes such as soybean, *Lotus*, *Medicago* and *Phaseolus* (Ameline-Torregrosa *et al.*, 2008; Innes *et al.*, 2008; Sato *et al.*, 2008; David *et al.*, 2009).

Early exploration of the host plant defense systems in peanut identified resistance gene analogs (RGAs) using degenerate primers based on the NBS region from *A. hypogaea* var. Tatu and four wild relatives (*A. duranensis*, *A. cardenasii*, *A. stenosperma*, and *A. simpsonii*) (Bertioli *et al.*, 2009). A more comprehensive characterization of disease resistance gene-like sequences in *Arachis* spp. identified RGAs for additional gene families as well as BACs containing R genes (Yuksel *et al.*, 2005), thus shedding early light on the genomic distribution of these elements. Several candidate genes have been associated with quantitative trait loci (QTLs) for late leaf spot disease resistance (Leal-Bertioli *et al.*, 2009). Recently, a dominant root-knot nematode resistance gene introduced into tetraploid peanut was identified and mapped (Nagy *et al.*, 2010). The main objectives of the present study were to explore two genomic regions in peanut suspected from previous study to contain clusters of RGAs; to assess evolutionary relationships both among peanut RGAs and with similar sequences from other taxa; and to explore evolutionary rules influencing functional innovations of domains of host-plant resistance genes.

Materials and Methods

Screen of BAC clones

Overgo screens of BAC libraries were essentially as described by Bowers *et al.* (2005). The BAC hit scores were

converted to the BAC addresses with an in-house script and analyzed with the BACman software (bacman.sourceforge.net) in order to assign each BAC to a specific overgo.

BAC sequencing

A total of 768 subclones (two 384-well plates) from each BAC clone were picked using a QBOT (Genetix, New Milton, UK) and were sequenced. Assembled sequences were visualized and manually edited using Consed (Gordon *et al.*, 1998). Peanut BAC sequences were annotated using gene prediction programs FGENESH (<http://www.softberry.com>). The predicted genes were searched for similarity to known proteins using BLASTP (Altschul *et al.*, 1990) with a threshold of $E < 10^{-10}$ against the National Center for Biotechnology Information (NCBI) nonredundant (nr) protein database. The coding sequences of the predicted genes in each region were used as query for BLASTN and TBLASTX searches against the corresponding homoeologous or orthologous regions with a cutoff of $E < 10^{-05}$ to identify the conserved genes. Synonymous and nonsynonymous substitution rates were estimated using the predicted coding sequences based on the modified Nei-Gojobori method implemented in MEGA 4.0 (Tamura *et al.*, 2007).

Identification and classification of long terminal repeat (LTR) retrotransposons

A combination of structural analyses and sequence homology comparisons were used to identify retrotransposons. The intact LTR elements were identified by using LTR_STRUC, an LTR-retrotransposon mining program (McCarthy & McDonald, 2003), and by homology on the basis of methods previously described (Ma & Bennetzen, 2004; Ma *et al.*, 2004). For all the intact retroelements with two LTRs, the LTR sequences of the same element were aligned by ClustalW (Chenna *et al.*, 2003) using default parameters. The pairwise sequence divergence was calculated with the Kimura 2 parameter model using MEGA 4.0 (Tamura *et al.*, 2007). The time of insertion was calculated using the equation $T = D/2t$, where T is time of insertion; D is divergence; and t is mutation rate per nucleotide site per year. The synonymous substitution rate per site per year of 1.3×10^{-8} was used to estimate the time of retrotransposon insertion (Ma *et al.*, 2004).

Phylogenetic analysis of retrotransposons, resistance gene analogs and other genes

We performed phylogenetic analysis of the domains (NBS, TIR and LRR, if available) of resistance genes and WD40 genes, being proximally located with the resistance genes on the sequenced BACs. Both the peptide and coding sequence (CDS) alignments were used to construct the trees

with various approaches, including maximum likelihood methods implemented in PHYML (Guindon *et al.*, 2005). Bootstrapping tests with 100 sampling repeats were performed on trees. The constructed trees were compared with one another and the best supported trees by those methods were used for interpreting the observations, and were also used as inputs to run PAML (Yang, 1997) in order to estimate the selection pressure along each branch of the trees. Furthermore, we identified positively selected amino acid under model M8 with two Bayesian approaches (naive empirical Bayesian analysis (NEB) and Bayes empirical Bayesian (BEB)) (Yang *et al.*, 2005). We also estimated the evolutionary distance between the genes in each family with the Nei–Gojobori model (Nei & Gojobori, 1986).

Comparative genomic analysis of gene synteny

The sequenced BACs were searched with MCSCAN (Tang *et al.*, 2008) to find conserved gene collinearity with the sequenced eudicot genomes downloaded from public databases (soybean, <http://www.phytozome.net/soybean>) (Schmutz *et al.*, 2010); *Medicago*, *M. truncatula* sequencing resources database (<http://www.medicago.org/>); *Arabidopsis*, TAIR (<ftp://ftp.arabidopsis.org/>); grape, Genoscope (<http://www.genoscope.cns.fr/>)).

Gene conversion inference

Gene conversion between the domains of peanut resistance genes were performed using public GENECONV software as described by Sawyer (1989) with the aligned CDS sequences.

Motif analysis

Profiles of conserved motifs in TIR and NBS domains in *Arabidopsis* TIR-NBS RGAs are adopted from Meyers *et al.* (2003) in MEME format (<http://meme.sdsc.edu/meme/>). Conserved motifs in *Arabidopsis* CC-NBS RGAs are used as control.

LRR structure exploration

To explore the sequence structure of LRR, using the public program DOTTER (Sonnhammer & Durbin, 1995), we produced dotplots between all LRR domains. The dotplots were produced by matched strings from two protein sequences in comparison. A string length of 7 was used to accommodate the reported conserved consensus pattern *xxaxxxx*, in which *a* is an aliphatic residue and *x* is any amino acid. We also predicted three-dimensional structures of LRR domains using a web-based software I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>), which predicts the protein structure and function based on

the sequence-to-structure-to-function paradigm (Roy *et al.*, 2009).

Results

Sequencing, annotation, and analysis of the NBS-LRR sequences

We sequenced two BACs, AHF 303L13 (GenBank accession no. HQ637177) and AHF 205D04 (GenBank accession no. HQ637178), previously identified as showing strong hybridization signals with multiple R-gene probes and thus considered likely to contain clusters of R genes (Yuksel *et al.*, 2005). Sequencing of these two BACs yielded 197 802 bp with an average guanine+cytosine (GC) content of 34%. BAC AHF 303L13 consists of 94 268 bp while AHF 205D04 has 103 534 bp. BAC sequences were annotated using a semi-automated approach to identify both protein-coding genes and repetitive elements. Annotation of these two BACs revealed a plethora of repetitive sequences showing significant similarities to retroviruses such as *Copia*-Ty1 and *Gypsy*-Ty3-like elements, Mutator-like elements (MULEs) and TIR-NBS-LRR genes (Fig. 1). A total of six RGAs were found, with AHF 205D04 containing five genes in a ‘tandem’ (proximal) cluster, while AHF 303L13 has only one. Two of the six RGAs (RGA4 and RGA5) contain the domains TIR-NBS-LRR shared by typical R genes, whereas the other four RGAs lack one or two domains. The TIR domains of three truncated RGAs (RGA2, RGA3, RGA6) were found as separate genes. For phylogenetic analysis, independent TIR domains were considered as part of adjacent RGAs. The RGAs were named RGA1 to RGA6 (Fig. 1).

Synteny and gene content are not conserved in NBS-LRR-rich regions

Functional DNA sequences are likely to be conserved between divergent species and therefore are an important tool in comparative structural genomics. Following this rationale we investigated the level and structure of microcollinearity between peanut and the largely sequenced genomes of *Gm*, *Mt*, *Vitis vinifera* (*Vv*), and *Arabidopsis thaliana* (*At*). For AHF 303L13, identification of collinear regions (Fig. 2) was aided by a highly conserved gene, 303L13_6, which showed 98% sequence similarity with EST-ES756475, a WD40 repeat protein expressed in peanut leaves. A *Medicago* BAC *mth2-77n20*, (GenBank accession no. AC149204) on chromosome 7 was found to share collinearity with peanut AHF 303L13 (Fig. 2). Sequence alignment revealed that peanut contains many more retroelement insertions than *Medicago* and, as a consequence, the WD40 gene and the NBS-LRR gene (303L13_20) have been spread apart. The WD40 gene

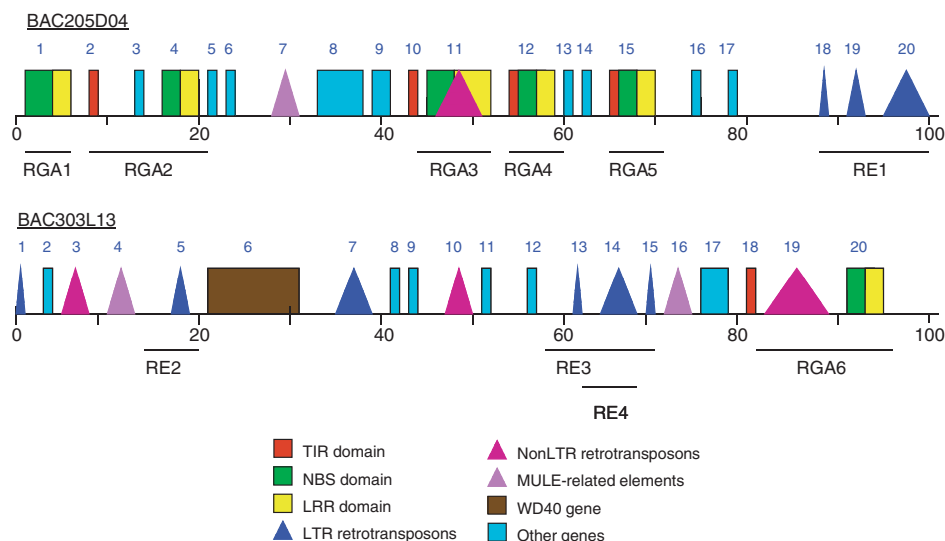


Fig. 1 Position of RGA1–RGA6 on two peanut bacterial artificial chromosomes (BACs) AHF 303L13 and 205D04. RE1–RE4 show the position of the retroelements. Domains of resistance gene analogs (RGAs) and transposable elements are shown in indicated colors. Numbers show the position of genes/open reading frame (ORF) on BACs and the scale indicates the length in kb. LRR, leucine-rich repeat; LTR, long terminal repeats; NBS, nucleotide-binding site; TIR, Toll-interleukin-1 receptor.

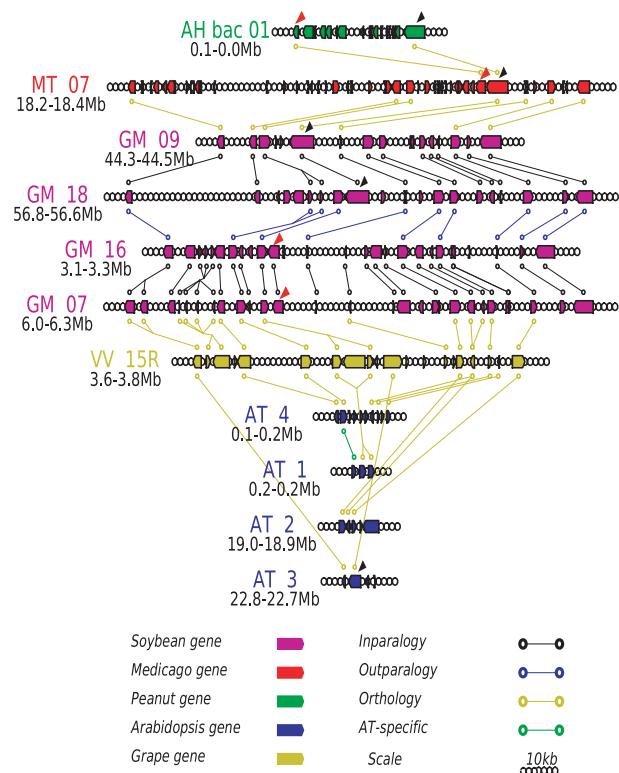


Fig. 2 Alignment of peanut bacterial artificial chromosome (BAC) AHF 303L13 with inferred syntenic regions from *Medicago* (MT), soybean (GM), grape (VV) and *Arabidopsis* (AT). Genes are shown with directional blocks, with specific colors displaying their sources. Gene synteny is shown with lines, and the color scheme displays different homologous relationships. Red arrows, resistance gene analogs; black arrows, WD40 genes.

(Mt7g087880.1) and the NBS-LRR gene (Mt7g087890.1/mt2-77n20_7) are only 872 bp apart in *Medicago*; however, in peanut they are *c.* 58 182 bp apart with several transposons and MULE-related sequences between them.

The relatively gene-rich *Medicago* BAC corresponding to peanut AHF 303L13 permitted us to identify four corresponding homoeologous regions (H1–H4) in the soybean genome (<http://www.phytozome.org/>) on chromosomes 9 (H1), 18 (H2), 16 (H3) and 7 (H4) (Fig. 2). The inference of four corresponding regions is consistent with the soybean genome having experienced at least two whole genome duplication events, estimated to have occurred 13 million yr ago (Mya) and *c.* 59 Mya (Schmutz *et al.*, 2009). Only the WD40 gene on peanut BAC AHF 303L13 had an ortholog in the soybean H1 (Gm9g39270.1) and H2 (Gm18g47050.1) syntenic region in highly conserved size and orientation. This gene was absent from the H3 and H4 regions of soybean, suggesting its loss from one homologue resulting from the *c.* 59 Mya genome duplication. Both soybean H3 and H4 homoeologous regions have one TIR-NBS-LRR gene (Gm7g07390.1 and Gm16g03780.1, respectively), which are each co-orthologs to the six peanut R genes (RGA1–RGA6); however, the TIR-NBS-LRR gene is absent from the H1 and H2 homoeologous regions in soybean. As in *Medicago*, the WD40 repeat and the TIR-NBS-LRR genes appear to have been very close to one another in soybean before the *c.* 59 Mya duplication (indeed, it is inferred to have been in a common ancestor shared with *Medicago*) (Pfeil *et al.*, 2005). While all other genes on the four homoeologous regions in soybean were absent from peanut BAC 303L13, this is not surprising

since the WD40 repeat and TIR-NBS-LRR genes are near the respective termini of the peanut BAC with the intervening region rich in retroelements. Sequence analysis indicated that the WD40 gene is highly conserved among different plant species and shares 90% protein sequence similarity with the soybean WD40 gene on H1 (Gm9g39270.1), 88% with the gene on H2 (Gm18g47050.1), and 68–69% with that in *Arabidopsis* AT3G61480.1 and AT5G28350.1, respectively. The *Arabidopsis* WD40 gene AT3G61480.1 is in the syntenic region, while AT5G28350.1 is located in the nonsyntenic region with the peanut BAC303L13. The peanut WD40 repeat protein also shows high similarity (83%) to a sequence from *Lj* (chr1.LjT10B06.10.nd) and 78% with *Vv* (VV15RG0232). *Lj* BAC Lj10B06 on chromosome 1 is in a region syntenic to the peanut BAC 303L13; however, it has only partial overlap with the peanut BAC. *Lotus* chromosome 1 is reported to be syntenic to linkage group Mt7 in *Medicago* and to linkage group A6 in an *A* genome map of peanut (Leal-Bertioli *et al.*, 2009). Phylogenetic relationships among the WD40 protein families were consistent with the expected relationships among genes from different species (Supporting Information, Fig. S1).

To explore if the syntenic regions are rich in RGAs, we searched for the presence of other closely related TIR-NBS-LRR genes in *Medicago*. Another region on chromosome 7 (BAC mth2-66o11, GenBank accession no. AC169666) was also found to have three RGAs (Mt7g088460.1, Mt7g088470.1, Mt7g099490.1) that shared high homology to the peanut RGAs. These two regions (BAC AC169666 and AC149204) are separated by *c.* 250 kb on *Medicago* chromosome 7. This finding indicates that disease resistance genes were present in this region in the common ancestor of *Medicago* and peanut, and that tandem duplication contributed to the expansion of this family. The *Medicago* sequence could also be used to identify corresponding regions on *Vitis* (grape) chromosome 15, and *Arabidopsis* chromosomes 1, 2, 3 and 4 (Fig. 2). A single corresponding region in *Vitis* is consistent with a lack of genome duplication in *Vitis* since triplication in a common ancestor of peanut and other eudicots (Jaillon *et al.*, 2007), and four regions in *Arabidopsis* is consistent with its two additional genome duplications since the shared triplication. However, microcollinearity could not be discerned directly between peanut and either *Vitis* or *Arabidopsis*, and in particular the TIR-NBS-LRR gene was not nearby in either species. Owing to the nature of genes present (high-copy R genes and transposable elements (TEs)), we could not confirm synteny of peanut BAC 205D04 with the other eudicot genomes; however, sequence data indicate that the RGAs on BACs 205D04 and 303L13 have the same co-orthologs in soybean (Gm16g03780.1, Gm7g07390.1) and *Medicago* (Mt7g087890.1, Mt7g088460.1, Mt7g088470.1, Mt7g088490.1)

Peanut NBS-LRR-rich regions have high densities of transposable elements

In the peanut BACs, we identified four intact retrotransposons, three on AHF 303L13 and one on AHF 205D04; three MULE-related sequences, two on AHF 303L13 and one on 205D04; and four nonLTR retrotransposons between and within NBS-LRR genes (Fig. 1). The retroelement *RE1* is 13 709 bp, is related to the *Gag-pol Copia* polyprotein, and shows high sequence similarity to a *Medicago* retroelement (ABD32582.1). Based on divergence between the respective LTRs, this element is estimated to have been inserted *c.* 0.67 Mya. Retroelement *RE2* is 5948 bp in length and is one of the oldest elements found, inserted *c.* 2.3 Mya. Retroelement *RE3* is the longest found at 16 835 bp, inserted *c.* 0.67 Mya and itself contains another retroelement (*RE4*) of 7007 bp that inserted *c.* 0.33 Mya ago. *RE4* is a *Gypsy* type gag polyprotein and shows high similarity to a retrotransposon gag protein (ABD63142.1) from *Asparagus officinalis*. We found three nonLTR retrotransposons on BAC 303L13 and one on BAC 205D04 (Fig. 1). BLASTN analysis of 303L13_3 indicated that it has 77% similarity to *Gm* LINE gmp1-83i9-re-1 sequence (Wawrzynski *et al.*, 2008). LINE 303L13_19 is inserted between the TIR and NBS domains of RGA6. We also found insertion of one nonLTR retrotransposon in RGA3 on peanut BAC AHF 205D04. The RGA 3 (205D04_11) is 1939 amino acids long with eight exons and a nonLTR retrotransposon is inserted between the NBS and LRR domains encoding a potentially functional protein. This clearly indicates that several nonLTR retrotransposons have been active in this region. In addition to retroelements, we found three sequences related to the MULE superfamily on the two peanut BACs (Fig. 1). Two MULEs were found on BAC 303L13 and one on BAC 205D04 (Fig. 1). These results indicate that TEs might have played a role in restructuring of peanut NBS-LRR-rich regions, as described in the following sections.

A cluster of peanut NBS-LRR genes are derived from one common ancestor and share a high degree of homology to CMR1, a *Phaseolus* TIR-NBS-LRR gene

Nucleotide-binding site, leucine-rich repeat genes on the two peanut BACs are from an ancient R-gene family that is highly conserved in different legumes. BLASTP against the NCBI nonredundant database showed that the NBS domain of RGA5 matches both *Medicago* TIR (Mt7g087890.1, *E*-value 5e-82, GenBank accession no. ABD28703.1) and *Phaseolus* CMR1 (GenBank accession no. ABH07384.1, *E*-value 3e-79) R genes. The TIR domain of RGA2 also has high sequence homology to both *Medicago* and *Phaseolus* sequences. Collectively, these findings indicated that all peanut NBS-LRR genes on BACs 303L13 and 205D04 share a

high degree of similarity to three genes, viz. *Phaseolus* CMR1 (ABH07384.1), *Medicago* TIR (Mt7g087890.1/ABD2870 3.1), and *Lens* (CAD56833.1) disease resistance proteins. The *Phaseolus* R gene CMR1 (ABH07384.1) confers resistance to geminivirus and its expression in transgenic tobacco elicits a hypersensitive response (Seo *et al.*, 2006).

Comparison of CMR1-related sequences with *Medicago* and peanut TIR-NBS-LRR

To better understand the evolution of the NBS-LRR genes in the sequenced region, we performed phylogenetic analyses using the NBS domains with previously published

TIR-NBS-LRR genes from *Medicago* (Ameline-Torregrosa *et al.*, 2008). Fig. 3 shows the phylogenetic analysis of *Medicago* TIR-NBS-LRR genes and their homologs in soybean, *Medicago*, *Phaseolus* and *Lens*. CMR1-related TIR-NBS-LRR sequences from different legumes are grouped together, indicating that these RGAs are highly conserved and were present in the most recent common ancestor of *Phaseolus*, *Glycine*, *Medicago* and peanut. We compared the sequences of all peanut TIR-NBS-LRR genes identified to date (Bertioli *et al.*, 2003; Yuksel *et al.*, 2005) with CMR1-related protein sequences from other plants (Fig. S2). Several important conclusions can be drawn. First, there clearly has been recent expansion of RGA clusters in peanut

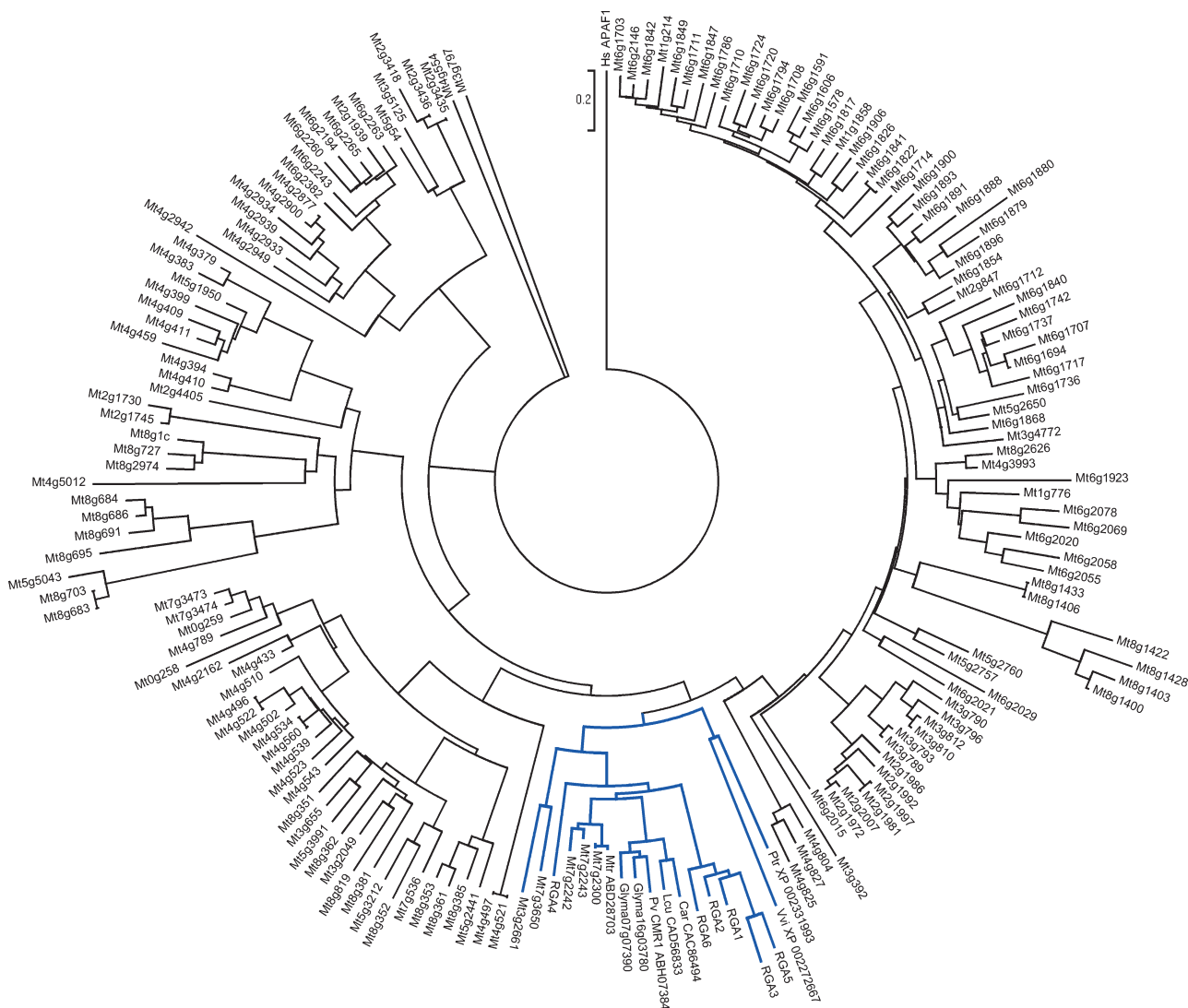


Fig. 3 Phylogenetic analysis of CMR1-related resistance gene analogs (RGAs) with *Medicago* TIR-NBS-LRR (Toll-interleukin-1 receptor, nucleotide-binding site, leucine-rich repeat) sequences. The region extending from the conserved P-loop motif to GLPL of NBS-LRR class resistance genes from peanut, *Medicago*, and other species (chickpea, lentil, *Vitis*, *Populus*) was phylogenetically analyzed by construction of a neighbor-joining tree using MEGA after multiple alignment of polypeptide sequences with CLUSTAL_X. The tree is rooted to the human *Apaf1* gene. CMR1-related RGAs are highlighted in blue. Vvi, *Vitis vinifera*; Glyma, *Glycine max*; Mt, *Medicago truncatula*; Ptr, *Populus trichocarpa*; Car, *Cicer arietinum*; Lcu, *Lens culinaris*; Pv, *Phaseolus vulgaris*.

as evidenced by terminal branches with multiple closely related genes. Peanut RGA1-RGA6 located on BAC 303L13 and 205D04 show high similarity to other peanut paralogs, indicating their expansion by gene duplication events. The phylogenetic analysis also indicates that CMR1-related gene sequences are part of an ancient gene family conserved in related legumes and other eudicots.

Reconstruction of the evolutionary history of the RGA1-6 peanut resistance genes

Phylogenetic analysis of the NBS domain of our sequenced peanut RGAs and their nearest homologs from *Phaseolus*, *Medicago*, and soybean suggested the evolutionary history of these genes (Fig. S3). The six NBS domains from our RGAs form a group on the tree, with RGA4-NBS at the external position. We inferred that each RGA on BAC 205D04 was produced as a complete unit from RGA4 (or an ancestor that closely resembles it) through a series of proximal duplication events. RGA6, on BAC303L13, is grouped with RGA1-3 and RGA5, suggesting it to be phylogenetically related to the RGAs on BAC205D04. The dispersed pattern of derived domains suggests that nonfunctionalization of some may have included TE-mediated gene sequence fragmentation.

Motif analysis indicates high conservation of TIR-NBS domains

Toll-interleukin-1 receptor and NBS functional domains are generally well conserved in plant resistance genes; however, their constituent motifs can differ among RGA subclasses (Meyers *et al.*, 2003). For example, motif profiles of NBS domains in TIR-NBS RGAs and CC-NBS RGAs are quite different in *Arabidopsis*. Comparison of the *Arabidopsis* motifs to peanut RGA1-RGA6 shows high conservation between TIR-NBS RGAs, with most motifs

found in the same order and organization (Fig. 4) in these two species that last shared a common ancestor *c.* 100–150 Mya. Interestingly, however, we found that peanut RGA2 is missing a 43-amino-acid conserved motif₄ (following the nomenclature of Meyers *et al.*, 2003) which was absent in only three of 96 *Arabidopsis* TIR-NBS RGAs (Fig. 4). This motif includes the conserved RNBS-D functional motif previously identified in TIR-NBS genes. Sequence analysis reveals that a remnant of the lost motif₄ can still be discerned at the expected location in RGA2 even though the complete profile is gone. A similar example is the partially missing motif₁₃ in RGA5, which is missing in only five of 96 TIR-NBS RGAs in *Arabidopsis*.

Evolutionary novelty of RGAs evolves mainly in LRR regions

Although some LRR domains studied may be from non-functional genes, to further explore their evolution, we estimated the nucleotide substitution rates (K_a , K_s) on the synonymous and nonsynonymous codon sites and their ratios ($K_a : K_s$) (Table S1). Different domains have similar K_s values, although those of NBS regions are nominally but not significantly smaller (Table 1). LRR domains have significantly larger K_a values than NBS and TIR domains, showing faster protein sequence evolution and suggesting potential action of natural selection. The $K_a : K_s$ ratios for all three domains are < 1 , which is often taken to rule out positive selection. However, even when a ratio for the entirety of a gene or domain is much less than 1, positive selection could act on specific small portions of the gene (Wang *et al.*, 2009a). The LRR domains have significantly larger $K_a : K_s$ ratios than the other domains (Table 1), further implying positive selection (or at least weaker purifying selection). We performed a site-to-site search of the footprints of natural selection on the three domains based on a site-

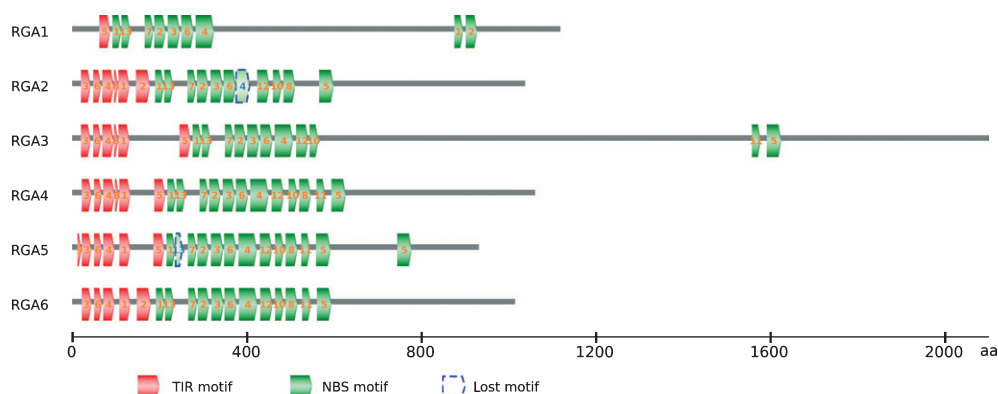


Fig. 4 Motif conservation in peanut and *Arabidopsis* TIR-NBS (Toll-interleukin-1 receptor, nucleotide-binding site) resistance gene analogs (RGAs). *Arabidopsis* motif profiles and numbers are adopted from Meyers *et al.* (2003). Motifs in the single TIR domains are shown in red. Motifs in the single NBS domains are shown in green. Dashed bricks indicate two cases of lost conserved motifs in RGA2 (motif 4) and RGA5 (motif 13), the lengths of which are estimated based on sequence alignment.

Table 1 Average evolutionary rates (K_a , K_s) and their ratios ($K_a : K_s$) in different domains, and P -values between different domains

	K_a	K_s	$K_a : K_s$
NBS	0.17	0.43	0.40
TIR	0.17	0.54	0.31
LRR	0.33	0.51	0.74
<i>P</i> -values			
NBS vs TIR	8.15e-01	2.65e-01	4.35e-03
NBS vs LRR	1.72e-04	4.32e-01	2.60e-07
TIR vs LRR	1.65e-04	7.74e-01	9.97e-07

NBS, nucleotide-binding site; TIR, Toll-interleukin-1 receptor; LRR, leucine-rich repeat.

specific model implemented in PAML. Based on the NEB approach, at a significance level of 0.05, we found evidence of positive selection acting on *c.* 5% of the LRR amino acid sites but no site in the NBS and TIR domains (Table 2). In LRR domains, six sites show extreme significance (0.001), all

Table 2 Positively selected amino acid sites in leucine-rich repeat (LRR) regions

Position	Residue	$P(\omega > 1)$	ω
<i>NEB approach</i>			
3	F	0.999**	1.325
5	K	0.954*	1.285
7	I	0.963*	1.293
38	K	0.993**	1.32
43	S	0.969*	1.298
61	L	0.997**	1.323
68	S	0.970*	1.299
146	K	0.985*	1.313
149	K	0.981*	1.309
168	K	0.999**	1.325
171	S	1.000**	1.325
192	K	0.999**	1.325
216	G	0.955*	1.286
239	K	0.989*	1.316
240	G	0.983*	1.311
248	L	0.967*	1.296
273	G	0.988*	1.315
278	Q	0.968*	1.297
281	K	0.989*	1.316
285	A	0.963*	1.293
286	N	0.969*	1.298
287	T	0.976*	1.305
<i>BEB approach</i>			
168	K	0.964*	1.489 ± 0.172
171	S	0.975*	1.497 ± 0.156
192	K	0.966*	1.49 ± 0.17

The parameter ω , reflecting natural selection pressure, is predicted by using site-specific models implemented in PAML. $P(\omega > 1)$ is the probability that positive selection occurs on a specific residue.

NEB, naive empirical Bayesian analysis; BEB, Bayes empirical Bayesian. * and ** indicate sites that were supported at significance levels of $P > 0.95$ and $P > 0.99$, respectively.

in the initial and middle regions of the LRR domain. Three of the six sites, all from the middle region of the LRR, also show evidence of positive selection by the BEB approach, suggesting that the middle part of the LRR domain plays an essential role in R-gene adaptive evolution.

Appreciable intragenic difference in amino acid composition, sequence structure, and selection pressure in LRR regions

Analysis of the sequences of LRR domains suggested intriguing evolutionary rules preserved across angiosperm species. Comparative dotplot analysis revealed unbalanced evolutionary contributions along the domain sequences (Fig. 5). The middle part (from *c.* 145 to 330 residues) of the LRR domain sequence of RGA5 and the corresponding parts of other peanut LRR domains show an interesting sequence structural pattern compared with the 5'- and 3'-neighboring regions (Fig. 5). Enriched short blocks shown in the dotplot display repetitive sequence accumulation in this region, which is scarce in the other regions. This should reflect the repetitive distribution of the functional units ($xx(a)x(a)xx$, where *a* is any aliphatic residue, and *x* is any amino acid, that form beta strands constituting the potential ligand-binding surfaces of the resistance genes to interact with other proteins. Further characterization of the secondary structure of LRR domains shows that the other regions are also composed of these units encoding beta strands. A likely explanation is that there have been recurring intragenic duplications and/or segmental gene conversion mainly occurring in the middle part of the LRR sequences. That is, the occurrences of duplication and conversion events have been unbalanced along the sequences, with the middle parts being the most affected. Comparative dotplotting shows that the contraction/expansion mainly occurred in the middle and 3'-end of LRR domains, which is a good sign of unequal crossings-over. Comparatively, the structure at the 5'-end of the domains has been relatively preserved, while the 3'-end shows enrichment for breakpoints. This interesting pattern of enriched duplications and conversions in the middle part is shared by LRR domains across the eudicot species studied (Fig. 5). This shows that rapid evolution of the LRR domain, likely the result of ectopic recombination (and associated mutations and breakpoints) and positive selection, has occurred in parallel in corresponding genes in different lineages.

To further explore their evolution, we manually divided the LRR domains into three parts: the 5'-end part, the middle part (active core), and the 3'-end part (Table 3). On average, the three parts have 160, 280 and 213 residues, respectively. We assumed that the middle part could function as an active core of LRR function. Leucine composition differs between the three functional parts in all eudicots studied, with the assumed active core having the

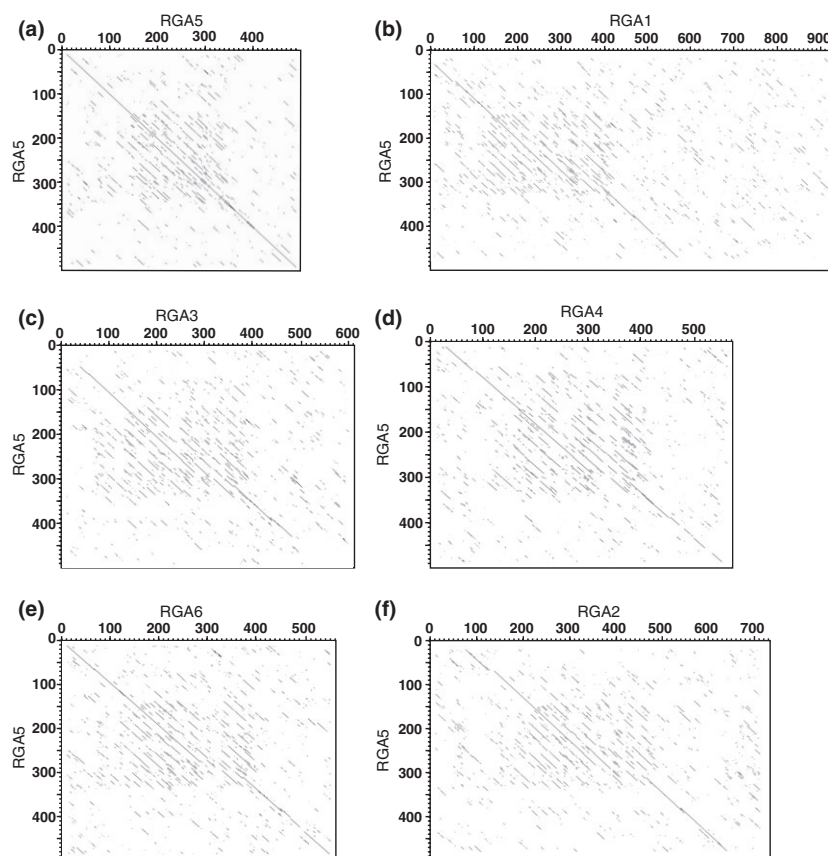


Fig. 5 Dotplot of resistance gene analog 5-leucine-rich repeat (RGA5-LRR) against all other annotated LRR regions, produced by running Dotter. (a) RGA5-LRR itself; (b) RGA1; (c) RGA3; (d) RGA4; (e) RGA6; (f) RGA2. Enriched short blocks display repetitive sequence accumulation in this region.

Table 3 Positions of assumed active core and leucine density of leucine-rich repeat (LRR) domains

ID	Length (LRR)	Active core position		Leucine density			
		Start	End	Global	5'-portion	Active core	3'-portion
RGA5	487	140	350	0.16	0.16	0.21	0.09
RGA4	588	270	420	0.15	0.16	0.21	0.09
RGA3	611	140	400	0.16	0.13	0.21	0.11
RGA2	733	220	480	0.16	0.12	0.23	0.13
RGA1	916	140	420	0.15	0.14	0.22	0.12
RGA6	556	140	400	0.16	0.14	0.21	0.09
GM7g07390.1	423	140	423	0.18	0.17	0.19	0.00
GM16g03780.1	706	140	470	0.15	0.17	0.19	0.09
MT7g088470.1	643	140	470	0.16	0.17	0.20	0.10
MT7g088460.1	631	140	470	0.16	0.17	0.19	0.10
MT7g087890.1	649	140	470	0.16	0.16	0.20	0.10
PV ABH07384.1	651	140	470	0.16	0.16	0.18	0.10
LC CAD56833.1	272	190	272	0.17	0.14	0.23	0.00

highest leucine accumulation (21% of all residues), significantly more than the 5'- and 3'-end regions (15 and 9%, statistically different from the assumed active core with P -values $4e-9$ and $5e-10$, respectively). Leucine residues are the backbone of LRR beta sheets, which are hypothesized to

play a central role in mediating interactions with pathogen molecules (Jia *et al.*, 2000; Jones & Dangl, 2006).

The secondary structures of LRR domains provide further evidence of evolutionary and functional asynchrony along the sequences. The peanut LRR domains contain 22–

Table 4 Beta strands in leucine-rich repeat (LRR) domains

ID	Number of beta strands				Average length of beta strands in amino acids			
	Global	5'-portion	Active core	3'-portion	Global	5'-portion	Active core	3'-portion
RGA5	22	6	10	6	3.55	4.67	2.90	3.50
RGA4	34	13	9	12	3.79	4.38	2.56	4.08
RGA3	33	6	16	11	3.67	4.17	2.81	4.64
RGA2	33	9	11	13	4.45	4.89	5.00	3.69
RGA1	30	5	13	12	4.50	5.40	4.38	4.25
RGA6	29	5	17	7	3.38	4.60	3.06	3.29
GM7g07390.1	17	5	12	0	3.76	5.00	3.25	N.A.
GM16g03780.1	26	5	11	10	3.81	5.00	3.45	3.60
MT7g088470.1	38	5	21	12	3.16	4.60	2.62	3.50
MT7g088460.1	30	6	15	9	4.00	4.33	3.47	4.67
MT7g087890.1	29	6	14	9	4.69	5.00	4.07	5.44
PV ABH07384.1	29	7	15	7	5.07	4.29	4.33	7.43
LC CAD56833.1	13	6	7	0	3.62	4.17	3.14	N.A.

The numbers and lengths of beta strands are shown in full LRR domains, denoted by 'global'. A LRR domain is divided into three parts: 5'- and 3'-portions, and the active core.

34 LRR beta sheets, similar to other eudicot domains except for two severely truncated ones (GM7G07390.1 and LC CAD56833.1) (Table 4). Interestingly, the sheets in assumed active cores are shorter than those in neighboring regions. On average, the sheets have 3.4 residues in the assumed active core, compared with 4.7 and 4.1 in the 5'- and 3'-ends (P -values = $9e-5$ and 0.09). Shorter sheets in the assumed active core may be a sign of gene conversion, which may produce breakages and contribute to fast LRR domain repatterning and evolution of interactions with pathogen molecules.

We predicted the three-dimensional LRR structure of the peanut RGAs, and the best model for each LRR has been shown (Fig. 6). RGA4-LRR has the most compact structure, forming an open circle consisting of parallel beta sheets, with the others often showing a little distortion in structure, especially at the 3'-end. As shown earlier, RGA4 may be the original template that produced the other RGAs on two peanut BACs. A comparative analysis of the secondary structures and three-dimensional structures shows that the predicted neighboring beta sheets in the secondary structure may form a shorter sheets, especially in the active core region. Together with the earlier observations of the difference between the active core regions and the other regions, this provides further evidence of more rapid evolution in the active core, potentially contributing to RGA functional innovation.

Intergenic conversion among peanut resistance genes

Gene conversion may affect gene evolutionary rate and change the topology of phylogenetic trees, consequently affecting further evolutionary analysis (Palomino *et al.*, 2002; Wang *et al.*, 2007, 2009b). Using GENECONV (Sawyer, 1989) we inferred gene conversion between the domains in RGAs1–6 and previously reported peanut

RGAs (96 in total) (Table S2), finding evidence of gene conversion between NBS domains. In our BACs, the NBS domain of RGA2 and RGA6, and those of RGA4 and RGA5 may have experienced gene conversion.

Discussion

Analysis of 197 kb in two peanut contigs has revealed the presence of six genes with similarity to plant disease resistance genes and four retroelements, as well as domains from WD40 repeats, MULE and other transposons, and other genes. Two RGAs appear to be intact, while the remainder are interrupted or truncated. Analysis of these RGA sequences suggests mechanisms contributing to evolution of novel disease resistance specificities in peanut. Phylogenetic and physical analyses revealed that RGAs were already present in this region of the most recent common ancestor of *Phaseolus*, *Glycine*, *Medicago*, and peanut, and appear to have recently expanded in peanut.

Evolution of a local resistance gene cluster in peanut

Nucleotide-binding site, leucine-rich repeat genes are the major class of disease resistance genes in flowering plants. Here, we characterized the evolution of a small cluster of six RGAs from two peanut BACs. Phylogenetic analysis of the NBS domains suggested their likely evolution by a series of proximal duplications. Accumulation of repetitive sequences, including LTR, nonLTR and MULE-like transposons, may have contributed to genome size evolution in the region (the peanut segments being more repeat-rich and gene-poor than their orthologs in other taxa). Some of these transposon insertions appear to be very recent, albeit noting that the accumulation of retrotransposons in the region

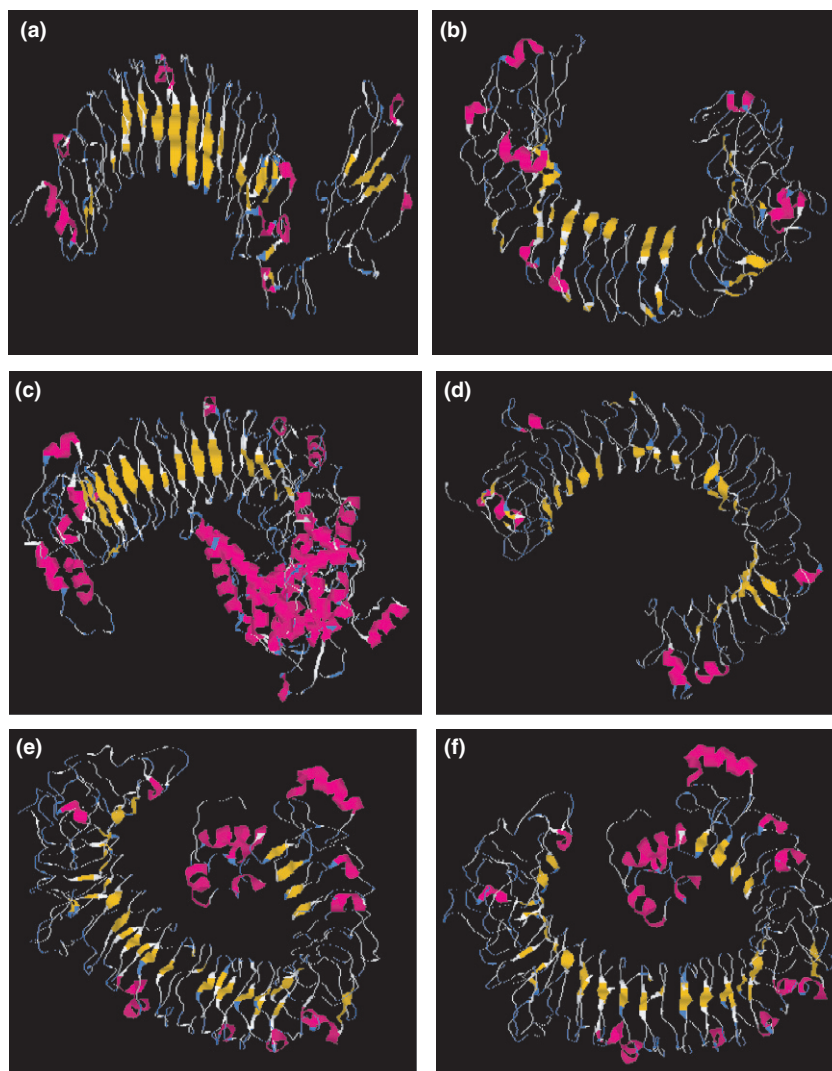


Fig. 6 Three-dimensional structures of leucine-rich repeat (LRR) domains predicted by using I-TASSER. (a) RGA5; (b) RGA4; (c) RGA1; (d) RGA3; (e) RGA6; (f) RGA2. RGA4-LRR has the most compact structure to form an open circle shape consisting of parallel beta sheets, with the others often showing a little distortion in structure, especially at the 3'-end. RGA, resistance gene analog.

could have been ancient with periodic turnover of the specific elements present. TE-mediated genome reorganization and TE-associated methylation may play a role in R-gene evolution (Lee *et al.*, 2009). For example, TE insertion may knock out or knock down R-gene expression (Luck *et al.*, 1998). Transposon-mediated transcriptional activation may also play an important role in the refunctionalization of additional 'sleeping' R genes in the plant genome (Hayashi & Yoshida, 2009).

The LRR domain plays a primary role in resistance function, possibly mainly through the assumed active core

Nucleotide-binding site, leucine-rich repeat RGAs have been proposed previously to encode intracellular receptors,

and at the C-terminal end resistance proteins consist of a series of LRRs, thought to be involved in protein–protein interactions. It has been known for some time that the LRR motif is involved in ligand-binding in porcine ribonuclease inhibitor (PRI) (Kobe & Deisenhofer, 1994, 1995a,b). Comparative analysis of RGAs from different land plants revealed that solvent-exposed residues consisting of LRR motifs are hypervariable and subject to positive selection, which may be a source of new resistance specificities and hence contribute to host–pathogen coevolution (Meyers *et al.*, 1998b; Ellis *et al.*, 2000). Genome-wide characterization in *Arabidopsis* showed that the positively selected residues were disproportionately located in the LRR domain, particularly a nine-amino-acid beta-strand submotif that is likely to be solvent-exposed (Palomino *et al.*, 2002). LRRs serve as the binding domain for the pathogen-

produced elicitor (Baker *et al.*, 1997). In some R genes, the LRR has been shown to interact with pathogen effectors directly (Banerjee *et al.*, 2001; Innes, 2004), and that mutations within the LRRs can affect these interactions (Warren *et al.*, 1998; Bendahmane *et al.*, 2002).

Here, we showed that the LRR domain has a core of functional and evolutionary activities, in that the beta sheets in this core region have more potential for functional innovation than those in other regions. First, the active core retained more mutations that produce changes in the beta sheets and it is also more likely than other regions to be affected by unequal crossing-over and conversion, which may induce mutations. Second, most positively selected residues were located in the active core region. There is significant enrichment of leucine in the active core, increasing the possibility to form an XXLXLXX motif which has been predicted to form a solvent-exposed beta sheet (Jones & Jones, 1997). The conserved leucines (L) project into the hydrophobic core, whereas the other residues (X) form a solvent-exposed surface that is involved in ligand binding (Kobe & Deisenhofer, 1995a). This finding of an active core in LRR domains may contribute to understanding the function and evolution of pathogen resistance held by NBS-LRR genes.

We have to note an important caveat, that the RGAs inferred to have positively selected sites were not proved to be functional in conferring resistance, because of a lack of previously characterized resistance genes in peanut. However in *Arabidopsis*, as previously reported, these RGAs grouped together with functional ones. The worry over this caveat can be, to some extent, mitigated for the following reasons. First, a variable motif like LRR theoretically provides a good place for adaptive evolution to happen. Secondly, the ever-changing environment and pathogens may favor one gene/allele from a reservoir of candidates, as anticipated by the one-gene-one-pathogen model. Allelic variation in many genes may be available that remains neutral until co-occurring with a specific set of conditions that favor a particular pathogen (or strain thereof). When the allele coincides with the right set of conditions, its specific mutations would contribute to positive selection. Finally, most plants are resistant to most pathogens, and it is often only when a resistance gene is overcome by a pathogen that we discover its function. Some of the most durable of resistance mechanisms may remain to be discovered.

Evidence of both intergenic and intragenic illegitimate recombination

Genetic recombination is a driving force of biological evolution, being a major source of genetic novelties by producing new alleles and combinations of alleles (Puchta *et al.*, 1996) that may permit adaptation to environmental changes. Recombination, especially illegitimate recombination

between homeologous sequences, may result in severe chromosomal lesions characterized by various DNA rearrangements which are often deleterious, but may occasionally contribute to elimination of deleterious mutations. Gene conversion can occur as a result of DNA double- or single-strand breaks during meiosis (Chen *et al.*, 2007; Gaeta & Pires, 2009). Illegitimate recombination has been widely invoked to explain the evolution of large gene families, such as histone and rRNA genes (Brown *et al.*, 1972; Ohta, 1984; Wang & Paterson, 2011). The contribution of illegitimate recombination to gene family evolution is multifaceted. On one hand, gene conversion may eliminate mutations by recovering original bases from the other genes, which would contribute to gene conservation. This resulted in the proposition that most multigene families were subject to concerted evolution, in which member genes of a family frequently exchange information and therefore evolve as a unit (Nei & Rooney, 2005). On the other hand, illegitimate recombination may induce mutations since its mechanisms of occurrence may involve single- or double-DNA-stranded breakage. This may increase the evolutionary rate of genes, and consequently some genes may die by accumulating too many mutations. Correspondingly, a birth-and-death model has been proposed to explain the evolution of multigene families (Hughes & Nei, 1988, 1989).

Host-plant resistance genes often form large gene families, frequently containing hundreds of copies distributed throughout genomes (Meyers *et al.*, 2003; Zhou *et al.*, 2004). Unequal crossing-over may have contributed significantly to the formation of such large gene families (Meyers *et al.*, 1998a). Segmental duplications, likely produced by whole-genome duplication events, may also have contributed to the expansion of such families (Baumgarten *et al.*, 2003; Leister, 2004). It has been proposed that resistance gene families are subject to gene conversion. Here, we also found evidence of gene conversion between peanut RGA NBS domains; however, for most cases only ~10% of gene sequences have been affected. Therefore, our results support the previous proposition that gene conversion is not occurring frequently enough to homogenize the resistance gene sequences (Michelmore & Meyers, 1998). Gene conversion, as ectopic recombination, may be parallel with DNA losses. This is supported by the fact of losses and size changes of NBS and LRR motifs, as reported earlier in this paper. There could be intragenic conversions among LRR motifs, especially those forming the active core, as indicated by structural analyses (dotplots). However, because the motifs are quite short, it is beyond any statistical inference to show the significance of inter-LRR motif conversion. We have to note, in view of the rapidly evolving nature of LRR domains, that conversion may have accelerated the evolution of LRR motifs by increasing the mutation rate, thereby providing material for natural selection.

Comparative genomics reveal limited gene synteny in the RGA-clustering regions

The degree to which genome synteny can facilitate cross-species analysis of gene function depends both on the conservation of gene order and content, and on the frequency with which similar traits have a common genetic basis in different species. Our analysis of *c.* 197 kb of genomic sequence from two peanut BAC clones revealed that *Medicago*, soybean and peanut display only limited micro-collinearity in this region. This lack of synteny might be the result of one or more of the following: sampling bias – the representative BAC clones selected contain several RGAs and TEs which directly affect microsynteny; regions in which gene propagations such as RGA clusters are free to evolve tend to be heterochromatic and relatively tolerant of DNA insertions and rearrangements (Bowers *et al.*, 2005). Our results are consistent with the previous observation (Bertioli *et al.*, 2009) that *Medicago* regions with high synteny scores with *Arachis* have low TE density, and vice versa. The *Medicago* genome harbors two ‘super clusters’ of RGAs, one in the upper region of chromosome 3 and one in the lower region of chromosome 6; clusters are also present in the upper regions of chromosomes 4 and 8 (Ameline-Torregrosa *et al.*, 2008). The peanut BAC 303L13 is syntenic to a region on chromosome 7 of *Medicago* and is not part of a super cluster of RGAs. In *Lotus*, clusters of RGAs are present on chromosomes 1, 2, and 3 (Sato *et al.*, 2008). Interestingly, synteny between *Medicago* and *Lotus* appears to be poor in many of the genomic regions that harbor major RGA clusters (Cannon *et al.*, 2006), perhaps supporting the hypothesis that such clusters are more likely to evolve in heterochromatic regions (Bowers *et al.*, 2005).

Conclusion

By exploiting resources for peanut, soybean, *Medicago*, *Arabidopsis*, and *V. vinifera*, in a targeted comparative study centered on *c.* 197 kb spanning peanut R-gene clusters and combining genomic and phylogenetic approaches, we gained insights into both RGA evolution and gene synteny. Analysis of sequence mutations, protein secondary structure and three-dimensional structure all suggest that LRR domains are the primary contributor to the evolution of resistance genes. The central part of LRR regions, assumed to serve as the active core, may play a key role in resistance evolution as a result of and/or reflected by higher rates of duplication and DNA conversion than the neighboring regions. In crops such as peanut where genetic diversity is very limited, targeted searches of gene banks or wild germplasm for novel alleles in such core regions may be a means of identifying promising material warranting empirical study of its value for genetic improvement programs.

Acknowledgements

This work was supported by the CGIAR Generation Challenge Program and the Georgia Peanut Commission. We thank Steven Cannon for providing *Medicago* gene sequences.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- Ameline-Torregrosa C, Wang BB, O’Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB. 2008. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiology* 146: 5–21.
- Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar SP. 1997. Signaling in plant–microbe interactions. *Science* 276: 726–733.
- Banerjee D, Zhang XC, Bent AF. 2001. The leucine-rich repeat domain can determine effective interaction between RPS2 and other host factors in *Arabidopsis* RPS2-mediated disease resistance. *Genetics* 158: 439–450.
- Baumgarten A, Cannon S, Spangler R, May G. 2003. Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics* 165: 309–319.
- Bendahmane A, Farnham G, Moffett P, Baulcombe DC. 2002. Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the Rx locus of potato. *Plant Journal* 32: 195–204.
- Bertioli DJ, Leal-Bertioli SCM, Lion MB, Santos VL, Pappas G, Cannon SB, Guimarães PM. 2003. A large scale analysis of resistance gene homologues in *Arachis*. *Molecular Genetics & Genomics* 270: 34–45.
- Bertioli DJM, Marcio C, Madsen LH, Sandal N, Leal-Bertioli SCM, Guimarães PM, Hougaard BK, Fredslund J, Schauser L, Nielsen AM *et al.* 2009. An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. *BMC Genomics* 10: 45.
- Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC *et al.* 2005. Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. *Proceedings of the National Academy of Sciences, USA* 102: 13206–13211.
- Branch WDB, Brenneman TB. 1999. Field resistance to cylindrocladium black rot and tomato spotted wilt virus among advanced runner-type peanut breeding lines. *Crop Protection* 22: 729–734.
- Brown DD, Wensink PC, Jordan E. 1972. A comparison of the ribosomal DNA’s of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *Journal of Molecular Biology* 63: 57–73.
- Burrow MD, Simpson CE, Starr JL, Paterson AH. 2001. Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploid species. *Genetics* 159: 823–837.
- Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang XH, Mudge J, Vasdewani J, Scheix T *et al.* 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proceedings of the National Academy of Sciences, USA* 103: 14959–14964.
- Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. *Nature Reviews* 8: 762–775.
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* 31: 3497–3500.
- David P, Chen NWG, Pedrosa-Harand A, Thareau V, Sevignac M, Cannon SB, Debouck D, Langin T, Geffroy V. 2009. A nomadic

- subtelomeric disease resistance gene cluster in common bean. *Plant Physiology* 151: 1048–1065.
- Dwivedi SL, Crouch JH, Nigam SN, Ferguson ME, Paterson AH. 2003. Molecular breeding of groundnut for enhanced productivity and food security in the semi-arid tropics: opportunities and challenges. *Advances in Agronomy*, 80: 153–221 San Diego, CA, USA: Academic Press Inc.
- Ellis J, Dodds P, Pryor T. 2000. Structure, function and evolution of plant disease resistance genes. *Current Opinion in Plant Biology* 3: 278–284.
- Gaeta RT, Pires JC. 2009. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytologist* 186: 18–28.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Research* 8: 195–202.
- Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research* 33: W557–W559.
- Hayashi K, Yoshida H. 2009. Refunctionalization of the ancient rice blast disease resistance gene *Pit* by the recruitment of a retrotransposon as a promoter. *Plant Journal* 57: 413–425.
- Hougaard BK, Madsen LH, Sandal N, de Carvalho Moretzsohn M, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S *et al.* 2008. Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. *Genetics* 179: 2299–2312.
- Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167–170.
- Hughes AL, Nei M. 1989. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Molecular Biology & Evolution* 6: 559–579.
- Innes RW. 2004. Guarding the goods: new insights into the central alarm system of plants. *Plant Physiology* 135: 695–701.
- Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NW, Couloux A, Dalwani A, Denny R *et al.* 2008. Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. *Plant Physiology* 148: 1740–1759.
- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choise N, Aubourg S, Vitulo N, Jubin C *et al.* 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.
- Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO Journal* 19: 4004–4014.
- Jones DA, Jones JDG. 1997. The role of leucine-rich repeat proteins in plant defences. *Advances in Botanical Research* 24: 90–167.
- Jones J, Dangl JL. 2006. The plant immune system. *Nature* 444: 323–329.
- Kobe B, Deisenhofer J. 1994. The leucine-rich repeat: a versatile binding motif. *Trends in Biochemical Sciences* 19: 415–421.
- Kobe B, Deisenhofer J. 1995a. Proteins with leucine-rich repeats. *Current Opinion in Structural Biology* 5: 409–416.
- Kobe B, Deisenhofer J. 1995b. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* 374: 183–186.
- Leal-Bertioli SCM, Jose A, Alves-Freitas DMT, Moretzsohn MC, Guimaraes PM, Nielsen S, Vidigal BS, Pereira RW, Pike J, Favero AP *et al.* 2009. Identification of candidate genome regions controlling disease resistance in *Arachis*. *BMC Plant Biology* 9: 112.
- Lee S, Costanzo S, Jia Y, Olsen KM, Caicedo AL. 2009. Evolutionary dynamics of the genomic region around the blast resistance gene *Pi-ta* in AA genome *Oryza* species. *Genetics* 183: 1315–1325.
- Leister D. 2004. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends in Genetics* 20: 116–122.
- Luck JE, Lawrence GJ, Finnegan EJ, Jones DA, Ellis JG. 1998. A flax transposon identified in two spontaneous mutant alleles of the L6 rust resistance gene. *Plant Journal* 16: 365–369.
- Ma JX, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences, USA* 101: 12404–12410.
- Ma JX, Devos KM, Bennetzen JL. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Research* 14: 860–869.
- McCarthy EM, McDonald JF. 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* 19: 362–367.
- Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW. 1998a. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 10: 1817–1832.
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell* 15: 809–834.
- Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW. 1998b. Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell* 10: 1833–1846.
- Michelmore RW, Meyers BC. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research* 8: 1113–1130.
- Moretzsohn MC, Barbosa AVG, Alves-Freitas DMT, Teixeira C, Leal-Bertioli SCM, Guimaraes PM, Pereira RW, Lopes CR, Cavallari MM, Valls JFM *et al.* 2009. A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. *BMC Plant Biology* 9: 40.
- Nagy ED, Chu Y, Guo YF, Khanal S, Tang SX, Li Y, Dong WBB, Timper P, Taylor C, Ozias-Akins P *et al.* 2010. Recombination is suppressed in an alien introgression in peanut harboring *Rma*, a dominant root-knot nematode resistance gene. *Molecular Breeding* 26: 357–370.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology & Evolution* 3: 418–426.
- Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics* 39: 121–152.
- Ohta T. 1984. Some models of gene conversion for treating the evolution of multigene families. *Genetics* 106: 517–528.
- Palomino MM, Meyers BC, Michelmore RW, Gaut BS. 2002. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Research* 12: 1305–1315.
- Pensuk VJ, Wongkaew S, Patanothai A. 2004. Generation means analysis of resistance to peanut bud necrosis caused by Peanut bud necrosis tospovirus in peanut. *Plant Breeding* 123: 90–92.
- Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ. 2005. Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. *Systematic Biologists* 54: 441–454.
- Phipps PM, Porter DM. 1998. Collar rot of peanut caused by *Lasiodiplodia theobromae*. *Plant Disease* 82: 1205–1209.
- Puchta H, Dujon B, Hohn B. 1996. Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proceedings of the National Academy of Sciences, USA* 93: 5055–5060.
- Roy A, Kucukural A, Zhang Y. 2009. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 5: 725–738.
- Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K *et al.* 2008. Genome structure of the legume, *Lotus japonicus*. *DNA Research* 15: 227–239.
- Sawyer S. 1989. Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* 6: 526–538.

- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J *et al.* 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.
- Seo YS, Rojas MR, Lee JY, Lee SW, Jeon JS, Ronald P, Lucas WJ, Gilbertson RL. 2006. A viral resistance gene from common bean functions across plant families and is up-regulated in a non-virus-specific manner. *Proceedings of the National Academy of Sciences, USA* 103: 11856–11861.
- Sonnhammer ELL, Durbin R. 1995. A dot-matrix program with dynamic threshold control suitable for genomic DNA and protein sequence analysis. *Gene* 167: 1–10.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis MEGA software version 4.0. *Molecular Biology & Evolution* 24: 1596–1599.
- Tang HB, Wang XY, Bowers JE, Ming R, Alam M, Paterson AH. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research* 18: 1944–1954.
- Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH. 2009a. Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. *Genome Biology* 10: R68.
- Wang X, Paterson AH. 2011. Gene conversion in angiosperm genomes with an emphasis on genes duplicated by polyploidization. *Genes* 2: 1–20.
- Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH. 2007. Extensive concerted evolution of rice paralogs and the road to regaining independence. *Genetics* 177: 1753–1763.
- Wang X, Tang H, Bowers JE, Paterson AH. 2009b. Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization. *Genome Research* 19: 1026–1032.
- Warren RF, Henk A, Mowery P, Holub E, Innes RW. 1998. A mutation within the leucine-rich repeat domain of the *Arabidopsis* disease resistance gene RPS5 partially suppresses multiple bacterial and downy mildew resistance genes. *Plant Cell*, 10: 1439–1452.
- Wawrzynski A, Ashfield T, Chen NW, Mammadov J, Nguyen A, Podicheti R, Cannon SB, Thareau V, Ameline-Torregrosa C, Cannon E *et al.* 2008. Replication of nonautonomous retroelements in soybean appears to be both recent and common. *Plant Physiology* 148: 1760–1771.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*. 13: 555–556.
- Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology & Evolution* 22: 1107–1118.
- Yuksel B, Estill JC, Schulze SR, Paterson AH. 2005. Organization and evolution of resistance gene analogs in peanut. *Molecular Genetics & Genomics* 274: 248–263.
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. 2004. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics & Genomics* 271: 402–415.
- Zhu H, Choi HK, Cook DR, Shoemaker RC. 2005. Bridging model and crop legumes through comparative genomics. *Plant Physiology* 137: 1189–1196.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Phylogenetic tree of the WD40 gene family.

Fig. S2 Peanut NBS gene tree.

Fig. S3 Reconstructed phylogenetic tree of NBS domains.

Table S1 Synonymous and nonsynonymous nucleotide substitution rates of annotated genes

Table S2 Gene conversion inferred between peanut resistance genes, predicted by running GENECONV

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.