Received 15 Aug 2012
Accepted 24 Jan 2013
Running Title: Peanut Resistance Gene Analogs
Co-Editor: Merijn R. Kant

Identification of expressed resistance gene analogs from peanut (*Arachis hypogaea* L.) expressed sequence tags

Zhanji Liu[1,2†], Suping Feng[1,3†], Manish K. Pandey[1,4], Xiaoping Chen[1,5], Albert K. Culbreath[1], Rajeev K. Varshney[4], Baozhu Guo[6]*

[1]University of Georgia, Department of Plant Pathology, Tifton, GA 31793, USA
[2]Shandong Academy of Agricultural Sciences, High-Tech Research Center, Jinan 250100, China
[3]Qiongzhou University, College of Bioscience and Biotechnology, Sanya 572022, China
[4]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, India
[5]Guangdong Academy of Agricultural Sciences, Crops Research Institute, Guangzhou 510640, China
[6]US Department of Agriculture, Agricultural Research Service, Crop Protection and Management Research Unit, Tifton, GA 31793, USA
† These authors contributed equally to this research.
*Corresponding author
Tel: +1 229 387 2334; Fax: +1 229 387 2331; E-mail: Baozhu.Guo@ars.usda.gov

Abstract

Low genetic diversity makes peanut (*Arachis hypogaea* L.) very vulnerable to plant pathogens, causing severe yield loss and reduced seed quality. Several hundred partial genomic DNA sequences as nucleotide-binding-site leucine-rich repeat (NBS-LRR) resistance genes (*R*) have been identified, but a small portion with expressed transcripts has been found. We aimed to identify resistance gene analogs (RGAs) from peanut expressed sequence tags (ESTs) and to develop polymorphic markers. The protein sequences of 54 known *R* genes were used to identify homologs from peanut ESTs from public databases. A total of 1,053 ESTs corresponding to six different classes of known *R* genes were recovered, and assembled 156 contigs and 229 singletons as peanut-expressed RGAs. There were 69 that encoded for NBS-LRR proteins, 191 that encoded for protein kinases, 82 that encoded for LRR-PK/transmembrane proteins, 28 that encoded for Toxin reductases, 11 that encoded for LRR-domain containing proteins and 4 that encoded for TM-domain containing proteins. Twenty-eight simple sequence repeats (SSRs) were identified from 25 peanut expressed RGAs. One SSR polymorphic marker (RGA121) was identified. Two PCR-based markers (*Ahsw-1* and *Ahsw-2*) developed from RGA013 were homologous to the Tomato Spotted Wilt Virus (TSWV) resistance gene. All three markers were mapped on the same linkage group AhIV. These expressed RGAs are the source for RGA-tagged marker development and identification of peanut resistance genes.

Keywords: *Arachis hypogaea*; expressed sequence tags; resistance gene analogs; TSWV.

Introduction

Cultivated peanut, or groundnut (*Arachis hypogaea* L.), is an allotetraploid (AABB; 2n = 4x = 40) and is one of the most important oilseed crops grown. Peanut is a major source of edible oil and digestible protein with kernels composed of about 50% oil and 25% protein (Guo et al. 2012). However, very limited genetic variation has been revealed in peanut by using various molecular markers such as restriction fragment length polymorphisms (RFLP) (Kochert et al. 1991; Paik-Ro et al. 1992), random amplified polymorphic DNA (RAPD) (Subramanian et al. 2000), and simple sequence repeats (SSR) (Liang et al. 2009; Varshney et al. 2009; Qin et al. 2012; Wang et al. 2012). Scarcity of genetic diversity among cultivated peanut accessions is likely derived from the single hybridization event between two ancient diploid species, likely *Arachis duranensis* (A genome) and *Arachis ipaensis* (B genome) (Seijo et al. 2004, 2007; Favero et al. 2006; Burow et al. 2009). The low levels of genetic variation make peanut very vulnerable to many plant pathogens (Ratnaparkhe et al. 2011).

Peanut yield and quality are severely constrained by a wide variety of fungal, bacterial, viral, and nematode pathogens. Among the fungal diseases, early leaf spot (*Cercospora arachidicola*) and late leaf spot (*Cercosporidium personatum*) are most prevalent, and occur throughout peanut growing regions. Late leaf spot and rust (*Puccinia arachidis*) diseases often occur simultaneously and can cause 50%–70% yield loss in India and some African countries (Khedikar et al. 2010). Spotted wilt disease, caused by Tomato Spotted Wilt Virus (TSWV), has become more prevalent and more severe in the Southeastern United States. For example, TSWV caused approximately 12% peanut yield loss in Georgia in 1997, which represented about a USD $40 million value (Culbreath and Srinivasan 2011). Therefore, the most promising solution for managing peanut diseases is using resistant cultivars. A high yielding cultivar with improved resistance to multiple pathogens would present tremendous advantages for peanut farmers to combat these diseases.

Until now, about 70 *R* genes which confer resistance to various diseases have been cloned from different plant species by either map-based cloning or transposon tagging methods (Johal and Briggs 1992; Dixon et al. 1996; Whitham et al. 1996; Sanseverino et al. 2009). The majority of *R* genes share a few highly conserved domains such as a nucleotide binding site (NBS), leucine-rich repeats (LRR), protein kinases (PK), transmembranes (TM), and Toll and interleukin-1

receptor (TIR) domains (Ali and Yan 2012). These conserved domains facilitate the isolation of *R* genes or resistance gene analogs (RGAs) from the same or from other plant species. Based on the presence of specific conserved domains, plant *R* genes can be grouped into at least four classes (Xiao et al. 2006; Sanseverino et al. 2009). *R* genes containing NBS-LRR domains form the largest class of *R* genes (Meyers et al. 1999, 2003; Xiao et al. 2006). This class can be further divided into two sub-classes, TIR-NBS-LRR and non-TIR-NBS-LRR, based on the presence of a TIR domain on the N-terminus (Meyers et al. 1999, 2003). For example, the *N* gene of tobacco, which confers resistance to Tobacco mosaic virus, possesses an N-terminal NBS and C-terminal LRR domains, plus a TIR domain upstream of NBS (Whitham et al. 1994), while the *Rps2* of *Arabidopsis*, which confers resistance to the bacterium *Pseudomonas syringae*, contains an N-terminal NBS and at least 14 imperfect C-terminal LRR domains with a coiled-coil (CC) domain upstream of NBS (Bent et al. 1994). The second class of *R* genes consists of those with LRR and PK domains, such as *Xa21*, found in rice, with an intracellular serine/threonine kinase domain and 23 extracytoplasmic LRRs. *Xa21* confers resistance to rice leaf blight, caused by the bacterium *Xanthomonas oryzae* pv. *oryzae* (Song et al. 1995). The third class is those which contain a serine/threonine kinase (protein kinase) domain. An example of this class is the tomato *Pto* gene with 27 serine and 13 threonine residues, which confers resistance to bacterial speck caused by *Pseudomonas syringae* pv. *tomato* (Martin et al. 1993). The fourth class is those of the *R* genes containing large extracellular LRR domains, like the tomato *Cf-2* gene. The *Cf-2* gene, which provides resistance to the leaf mold pathogen *Cladosporium fulvum*, contains 37 LRRs (Dixon et al. 1996). The fifth class includes all other *R* genes which confer resistance to pathogens by different mechanisms, such as the *Hm1* gene of maize. *Hm1* confers resistance to the leaf spot fungus *Cochliobolus carbonum*. *Hm1* was the first *R* gene to be cloned, and encodes for a reductase enzyme that detoxifies the *C. carbonum* HC-toxin (Johal and Briggs 1992).

Efforts have been made to isolate RGAs from peanut. Bertioli et al. (2003) identified 78 RGAs from the peanut cultivar ‘Tatu’ and four wild diploid species by using degenerate primers targeting the P-loop, GLPL, and RNBS-D motifs of the NBS domains. Subsequently, they mapped 34 candidate RGAs to a linkage map by using a combination of methods of amplified fragment length polymorphism (AFLP), NBS profiling, RGA-AFLP, and sequence characterized amplified region (SCAR) markers (Leal-Bertioli et al. 2009). Yuksel et al. (2005) isolated 234 RGAs from the peanut cultivar ‘Florunner UF-439-16-1003-2’ using degenerate primers based

on the NBS-LRR and LRR-TM conserved domains. Ratnaparkhe et al. (2011) reported 6 RGAs from a BAC clone library. More recently, Wang et al. (2012) reported 3,784 BAC clones containing RGAs and two RGA-SSR markers which were mapped to a linkage map. In summary, there are currently 355 partial genomic DNA sequences for NBS-LRR encoding RGAs (Bertioli et al. 2003; Yuksel et al. 2005). All these RGAs were derived from genomic DNA sequences and not from expressed sequences. Dilbirligi and Gill (2003) concluded that the majority of genomic *R*-gene-like sequences from six crop plants did not have detectable transcripts, indicating that most of them could be non-functional. The peanut expressed sequence tag (EST) project has made significant progress (Guo et al. 2009; Feng et al. 2012b), and this resource has proven to be valuable for various genome-scale experiments (Guo et al. 2008b; Luo et al. 2010; Pandey et al. 2012b). Currently, there are 252,832 peanut ESTs available in the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/) (August 10, 2012), including 178,490 ESTs from cultivated peanut, 35,291 ESTs from *A. duranensis* (AA genome), 6,264 ESTs from *A. stenosperma* (AA genome), and 32,787 ESTs from *A. ipaensis* (BB genome). Therefore, the objectives of this study were to identify expressed RGAs from these peanut ESTs and to develop potential polymorphic markers for a genetic mapping population.

Results

Identification of RGAs from peanut ESTs

We identified 385 putative RGAs from the publicly available peanut ESTs at NCBI. The tBLASTn algorithm was used to identify expressed peanut RGAs homologous to (parts of) the full length protein sequences of 54 known plant *R* genes (Table 1). As a result, a total of 1,053 *R*-gene-like ESTs homologous to the different classes of known *R* genes were recovered with $E \leq$ e-10. After alignment and assembly, 385 unigenes were identified as peanut expressed RGAs, consisting of 156 contigs and 229 singletons (Table S1). The average length of the RGAs was 736 bp, ranging from 313 bp to 2,647 bp. A BLASTX search against GenBank indicated that the expressed RGAs could be classified as (1) NBS-LRR containing proteins (69 total); (2) protein kinases (PK) (191 total); (3) LRR-PK/TM (82 total); (4) Toxin reductase (28 total); (5) LRR-domain containing proteins (11 total); and (6) TM-domain containing proteins (4 total).

Of the 1,053 ESTs, 649 were from cultivated peanut (11 genotypes) and 404 were from wild peanut (3 genotypes) (Table 2). Among the 14 peanut genotypes, wild diploid peanut ‘DUR25’

contributed the largest number of *R*-gene-like ESTs, followed by cultivated peanut '06-4104' and 'Tifrunner'. The percentage of *R*-gene-like ESTs in the wild peanut 'DUR25' is 0.62% (219/35,291), and was higher than that of others except for the cultivars 850 (1.21%, 9/745) and Yueyou 523 (1.30%, 5/385). The 1,053 *R*-gene-like ESTs were mainly derived from tissue obtained from seeds (468), followed by roots (333), leaves (208), and gynospores (44) (Table 2).

Development of SSR markers

The MISA program (http://pgrc.ipk-gatersleben.de/misa) was used to identify SSRs from peanut RGAs (Thiel et al. 2003). A total of 28 SSRs were identified from 25 peanut RGAs with three RGAs containing two SSRs (Table 3). AAG/CTT (8 RGAs) and AAT/ATT (7 RGAs) were the most abundant motifs. Twenty-six SSR primer pairs were designed and used to screen for polymorphisms between the parental lines of two mapping populations (Additional File 2: Primers for 28 peanut RGA-SSRs and two putative *Ahsw* genes). All the primer pairs yielded amplification products using a DNA template from either one of the four parental lines. RGA121-SSR was polymorphic between the parental lines, GT-C20 and Tifrunner, and SunOleic 97R and NC94022 (Figure S1).

Sequence analysis and mapping of RGA121

RGA121 was 1,021 nucleotides long and had an uninterrupted open reading frame (ORF), which translated into a fragment of 339 amino acid residues (Figure S2). Within this fragment, two domains, a leucine-rich repeat N-terminal domain (A_{35} to D_{69}) and a protein kinase domain (T_{87} to Y_{324}), were revealed by searching the Pfam database (http://pfam.sanger.ac.uk/) (Punta et al. 2012), which indicated that RGA121 belongs to the LRR-PK gene family. The program BLASTP was used to search for RGA121 homologs in GenBank, and the soybean receptor protein kinase TMK1-like showed the highest similarity with $E \leq$ 2e-171.

In *Arabidopsis*, at least four LRR-PKs are involved in pathogen resistance. *FLAGELLIN-SENSITIVE 2* (*FLS2*) contributes to the perception of the bacterial elicitor flagellin in *Arabidopsis* (Gomez-Gomez and Boller 2000); *BIR1* and *SOBIR1* are involved in the regulation of cell death and innate immunity (Gao et al. 2009); *ERECTA*, a putative RLK with an extracellular LRR domain and an intracellular kinase domain, functions in both plant development and pathogen defense responses (Godiard et al. 2003). Sequence comparison

showed RGA121 had 34%, 38%, 41%, and 38% sequence identities with *FLS2*, *BIR1*, *SOBIR1*, and *ERECTA*, respectively.

In order to locate markers linked to RGA121 on the peanut genomic map, SSR analysis was conducted using 165 RILs of the T population and 363 RILs of the S population (Figure S3). In the T population, 88 and 72 RILs contained alleles from the parental line GT-C20 and Tifrunner, respectively. In the S population, the numbers of alleles from the parental line NC94022 and SunOleic 97R were 161 and 157, respectively (Table 3). A Chi-squared (χ^2) test indicated the segregation ratio was found to be the expected 1:1 in both mapping populations ($P < 0.05$). Based on the linkage map developed by the total 363 RILs from the S populations (Pandey et al. 2012a), RGA121 was mapped on the linkage group AhIV (**Figure 1**).

Mapping of the putative *Ahsw* gene

Previously, a cDNA clone (GO324202) from peanut with 37% amino acid identity to tomato *Sw-α*, a gene providing resistance to TSWV (Brommonschenkel and Tanksley 1997), was revealed and named as *Ahsw* (Chen et al. 2008; Feng et al. 2012a). Two putative genes, *Ahsw-1* and *Ahsw-2,* were identified by using 5' and 3' rapid amplification of cDNA ends (RACE) and electronic PCR methods. In the present study, GO324202, JK158518 and JR554938, were integrated into RGA013. The length of RGA013 was 1,758 bp. RGA013 contained an uninterrupted ORF which translated into a fragment of 561 amino acids. BLASTP indicated that RGA013 contained a NB-ARC domain from K172 to Q466 (Figure S4) and was similar to the soybean putative disease resistance protein At1g50180-like (LOC100814688) (79% identity). Three gene-specific markers were developed in the 3' variant region for each *Ahsw* gene (Additional File 2: Primers for 28 peanut RGA-SSRs and two putative *Ahsw* genes). Primers *Ahsw*1-c and *Ahsw*2-b detected a polymorphism between the two parental lines SunOleic 97R and NC94022, and were further used to genotype the S population. *Ahsw-1* and *Ahsw-2* were mapped onto the linkage group AhIV, along with RGA121 (**Figure 1**).

Discussion

Peanut resistance to plant pathogens has been identified for a variety of peanut cultivars and wild relatives, which indicates that peanuts do have *R* genes (Bertioli et al. 2003; Qin et al. 2012). Knowledge of expressed *R* genes or RGAs will significantly enhance our ability to better understand the host-pathogen genetic interaction and will facilitate the development of breeding

new resistance peanut cultivars. Identifying and mapping *R* genes from cultivated peanuts is very difficult due to the large size of the peanut genome (~ 2,800 Mb) and the relatively low marker density of linkage maps (Guo et al. 2008a; Varshney et al. 2009; Hong et al. 2010; Qin et al. 2012; Wang et al. 2012). In contrast, there are 252,832 *Arachis* ESTs available in GenBank, which we show to be useful for RGA discovery. In the present study, 385 putative RGAs were successfully identified from the peanut EST database. These RGAs provide a large set of sequence data for RGA-tagged marker development.

Identification of RGAs by data mining

Generally, there are two methods that are used in RGA identification: the PCR-based method and data mining. The PCR-based method uses degenerate primers based on the conserved domains of known *R* genes, and has been successfully used to identify RGAs from a number of plant species (Kanazin et al. 1996; Gowda et al. 2002; Hunger et al. 2003; Gao et al. 2010). In peanut, a total of 355 genomic RGAs have been isolated using the PCR-based method (Bertioli et al. 2003; Yuksel et al. 2005; Nagy et al. 2010; Ratnaparkhe et al. 2011). We compared these genomic RGAs with *Arachis* ESTs available in GenBank (252, 832 ESTs), and found only 61 genomic RGAs that matched 18 *Arachis* ESTs ($E \leq$ e-10). This indicates that only a small fraction of RGAs from genomic DNA sequences are expressed.

Recently, data mining has been successful in RGA discovery from wheat, sugarcane, maize, and common bean (Dilbirligi and Gill 2003; Rossi et al. 2003; Xiao et al. 2006; Liu et al. 2012). Dilbirligi and Gill (2003) compared four different data-mining approaches, namely domain search, individual and multiple motif searches, consensus sequence search, and individual full-length search, to identify *R*-gene-like sequences from wheat ESTs, and concluded that the individual full-length search was the most efficient method (Dilbirligi and Gill 2003). Xiao et al. (2006) adopted three methods, namely modified RACE, AFLP and data mining (individual full-length search) to identify RGAs from maize, and showed that the data mining method revealed the largest number of RGAs. In this study, we used the individual full-length search method and identified 385 peanut expressed RGAs. Nagy et al. (2010) found three *R*-gene homologs (RGC144b, RGC154 and RGC240) linked to *Rma*, a dominant root-knot nematode resistance gene. In the present study, RGA106 is identical to the reported RGC154, but the RGA106 is much longer in the 3' end than RGC154.

RGA121 and *Ahsw* were mapped on the same linkage group AhIV

Mapping of *R* genes or RGAs is an important step toward understanding the relationship between specific chromosomal regions and disease resistance. In the present study, RGA121, *Ahsw-1*and *Ahsw-2* were mapped onto the peanut linkage group AhIV (**Figure 1**), along with a QTL LLS_TF11E6 in control of late leaf spot resistance, and a QTL TSWV_DW10E1 resistant to TSWV (Pandey et al. 2012a).

Despite the importance of disease resistance in peanut, relatively little is known about the chromosomal regions that carry the genes required for disease resistance. This has hampered the use of markers in the development of resistant cultivars. During the last five years, there have been attempts to map peanut disease resistance. Leal-Bertioli et al. (2009) mapped 34 candidate RGAs and five QTLs associated with enhanced resistance to late leaf spot on the genetic map of the A-genome of *Arachis*. A putative QTL, cp4.1 resistant to late leaf spot, was mapped on the linkage group A4 near the marker TC7G10. The two markers most near to cp4.1, p25M46-2 and As26A, are RGAs (Leal-Bertioli et al. 2009). Linkage group A4 and AhIV shared a common marker, TC7G10. Moreover, TC7G10 is located between *Ahsw-1* and *Ahsw-2* in the present study (**Figure 1**). Sujay et al. (2012) identified 28 QTLs for late leaf spot resistance and 15 QTLs for rust resistance by using two RIL populations. Among them, two QTLs, $QTL_{R4\text{-}LLS}13$ and $QTL_{R5\text{-}LLS}14$ conferring resistance to late leaf spot, and one QTL, $QTL_{R4\text{-}Rust}06$ in control of rust infection, were located on the linkage group AhVII near the markers GM1311 and GM2246. These two markers were also mapped on the upper regions of linkage group AhIV in our study. Furthermore, GM2246 was conditioned within the QTL TSWV_DW10E1. Therefore, RGA clusters may exist in the upper regions of the linkage group AhIV. Numerous RGAs have been shown to co-localize with resistance genes/QTLs (Pflieger et al. 2001; Wisser et al. 2005; Xiao et al. 2007), and such RGAs in peanut would be highly suitable candidates for use as markers for breeding. Therefore, further studies will be needed to determine the function of RGA121 and *Ahsw* or RGA013. Recently, Shirasawa et al. (2012) reported a high-density linkage map for cultivated peanut with 1,114 loci. Linkage group LG04.1 shared two markers, TC7G10 and Seq15C12, with linkage group AhIV in the present study. Such high-saturation peanut linkage will hence be needed for the fine mapping of RGA121 and *Ahsw* and for map-based cloning.

Materials and Methods

Data mining of peanut RGAs

The protein sequences of 54 known *R* genes corresponding to different *R* gene classes were used to search for (tBLASTn) peanut homologues in the GenBank EST database (est_others; Organism: Arachis; http://www.ncbi.nlm.nih.gov; Altschul et al. 1997). The 54 *R* genes that were used for the data mining study are listed in **Table 1** along with their original plant species, putative structures and GenBank accession numbers. Those sequences with $E \leq$ e-10 were clustered to develop unigenes with parameters set at 100 bp for overlap length and 95% for nucleotide identity, and all the unigenes were considered as putative peanut RGAs (Xiao et al. 2006; Liu et al. 2012). The resulting unigenes were in turn used to search the GenBank databases by BLASTX to confirm their putative annotations.

Development and mapping of peanut RGA markers

The MISA program (http://pgrc.ipk-gatersleben.de/misa) was used for identification of SSRs in peanut RGAs (Thiel et al. 2003). SSRs were considered to be those sequences containing motifs with a size of two to six nucleotides and di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide motifs with minimum repeat of 6, 5, 5, 5, and 5, respectively. The BatchPrimer3 v1.0 (http://probes.pw.usda.gov/batchprimer3/) was used to design SSR primers for the RGAs containing SSRs with default parameters (You et al. 2008). Two gene-specific markers, *Ahsw-1* and *Ahsw-2*, were also developed based on the gene sequences (Brommonschenkel and Tanksley 1997; Chen et al. 2008; Feng et al. 2012a). These SSRs and gene-specific markers were used for screening the four parental lines of two recombinant inbred line (RIL) mapping populations, the T population derived from Tifrunner and GT-C20, and the S population developed from SunOleic 97R and NC94022 (Qin et al. 2012). Those with polymorphisms were used to genotype individuals of the T and S populations. Markers were tested for segregation distortion by Chi-squared (χ^2) test for goodness of fit to a 1:1 ratio. Linkage analysis was performed by using JoinMap 4.0 (Van Ooijen, 2011).

DNA extraction and PCR

Genomic DNA was isolated from young leaflets of the four parents, 165 RILs of the T population and 363 RILs of the S population (Qin et al. 2012). A Nano-Drop 1000 spectrophotometer (Nano Drop Technologies, USA) was used to evaluate the quality and concentration of all DNA. All of the DNA samples were diluted to 20 ng/µL, and PCR reactions were carried out in 10 µL as described by Qin et al. (2012). PCR products were analyzed on 6% or 9% non-denaturing polyacrylamide gels (PAGE) and visualized by silver staining as described by Fountain et al. (2011).

Acknowledgements

We thank Billy Wilson and Sara Beth Pelham for technical assistance in the field and the laboratory. This research was partially supported by funds provided by the Georgia Agricultural Commodity Commission for Peanuts, the National Peanut Board and the Peanut Foundation. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture (USDA). USDA is an equal opportunity provider and employer.

References

Ali F, Yan J (2012) Disease resistance in maize and the role of molecular breeding in defending against global threat. *J. Integr. Plant. Biol.* **54,** 134–151.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search program. *Nucleic Acid Res.* **25**, 3389–3402.

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ (1994) *RPS2* of *Arabidopsis thaliana*: A leucine-rich repeat class of plant disease resistance genes. *Science* **265**, 1856–1860.

Bertioli DJ, Leal-Bertioli SC, Lion MB, Santos VL, Pappas G Jr, Cannon SB, Guimaraes PM (2003) A large scale analysis of resistance gene homologues in *Arachis*. *Mol. Genet. Genomics* **270**, 34–45.

Brommonschenkel SH, Tanksley SD (1997) Map-based cloning of the tomato genomic region that spans the *Sw-5* tospovirus resistance gene in tomato. *Mol. Gen. Genet.* **256**, 121–126.

Burrow MD, Simpson CE, Faries MW, Starr JL, Paterson AH (2009) Molecular biogeographic study of recently described B- and A-genome *Arachis* species, also providing new insights into the origins of cultivated peanut. *Genome* **52**, 107–119.

Chen X, Culbreath A, Brenneman T, Holbrook C, Guo B (2008) Identification and cloning of TSWV resistance gene(s) in cultivated peanuts and development of markers for breeding selection. *Phytopathology* **98**, S36.

Culbreath AK, Srinivasan R (2011) Epidemiology of spotted wilt disease of peanut caused by tomato spotted wilt virus in the southeastern U.S. *Virus Res.* **159**, 101–109.

Dilbirligi M, Gill KS (2003) Identification and analysis of expressed resistance gene sequences in wheat. *Plant Mol. Biol.* **53**, 771–787.

Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JD (1996) The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat protein. *Cell* 1996, **84**, 451–459.

Favero AP, Simpson CE, Valls JF, Vello NA (2006) Study of the evolution of cultivated peanut through crossability studies among *Arachis ipaensis*, *A. duranensis*, and *A. hypogaea*. *Crop Sci.* **46**, 1546–1552.

Feng S, Chen X, Liu Z, Holbrook C, Culbreath A, Guo B (2012a) Identification of putative TSWV resistance genes and development of gene-specific marker in peanut. *Phytopathology* 102, S2.3.

Feng S, Wang X, Zhang X, Dang PM, Holbrook CC, Culbreath AK, Wu Y, Guo BZ (2012b) Peanut (*Arachis hypogaea*) expressed sequence tag (EST) project: Progress and application. *Comp. Funct. Genomics* 2012, Article ID 373768, 9 pages. doi:10.1155/2012/373768.

Fountain J, Qin H, Chen C, Dang P, Wang ML, Guo B (2011) A note on development of a low-cost and high-throughput SSR-based genotyping method in peanut (*Arachis hypogaea* L.). *Peanut Sci.* **38**, 122–127.

Gao M, Wang X, Wang D, Xu F, Ding X, Zhang Z, Bi D, Cheng YT, Chen S, Li X, Zhang Y (2009) Regulation of cell death and innate immunity by two receptor-like kinases in *Arabidopsis*. *Cell Host Microbe* **6**, 34–44.

Gao Y, Xu Z, Jiao F, Yu H, Xiao B, Li Y, Lu X (2010) Cloning, structural features, and expression analysis of resistance gene analogs in tobacco. *Mol. Biol. Rep.* **37**, 345–354.

Godiard L, Sauviac L, Torii KU, Grenon O, Mangin B, Grimsley NH, Marco Y (2003) ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant J.* **36**, 353–365.

Gomez-Gomez L, Boller T (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Mol. Cell* **5**, 1003–1011.

Gowda BS, Miller JL, Rubin SS, Sharma DR, Timko MP (2002) Isolation, sequence analysis, and linkage mapping of resistance-gene analogs in cowpea (*Vigna unguiculata* L. Walp.). *Euphytica* **126**, 365–377.

Guo B, Chen X, Dang P, Scully BT, Liang X, Holbrook C, Yu J, Culbreath AK (2008a) Peanut gene expression profiling in developing seeds at different reproduction stages during *Aspergillus parasiticus* infection. *BMC Dev. Biol.* **8**, 1–16.

Guo B, Chen X, Hong Y, Liang X, Dang P, Brenneman T, Holbrook C, Culbreath C (2009) Analysis of gene expression profiles in leaf tissues of cultivated peanuts and development of EST-SSR markers and gene discovery. *Int. J. Plant Genomics* **2009**, 1–14.

Guo BZ, Chen CY, Chu Y, Holbrook CC, Ozias-Akins P, Stalker HT (2012) Advances in genetics and genomics for sustainable peanut production. In: Benkeblia N, ed. *Sustainable Agriculture and New Biotechnologies*. CRC Press, Boca Raton. pp. 341–367.

Guo BZ, Chen ZY, Lee RD, Scully BT (2008b) Drought stress and preharvest aflatoxin contamination in agricultural commodity: Genetics, genomics and proteomics. *J. Integr. Plant Biol.* **50**, 1281–1291.

Hong Y, Chen X, Liang X, Liu H, Zhou G, Li S, Wen S, Holbrook CC, Guo B (2010) A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. *BMC Plant Biol.* **10**, 17.

Huang J, Yan L, Lei Y, Jiang H, Ren X, Liao B (2012) Expressed sequence tags in cultivated peanut (Arachis hypogaea): Discovery of genes in seed development and response to *Ralstonia solanacearum* challenge. *J. Plant Res.* **125**, 755–769.

Hunger S, Di Gaspero G, Mohring S, Bellin D, Schafer-Pregl R, Borchardt DC, Durel CE, Werber M, Weisshaar B, Salamini F, Schneider K (2003) Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (*Beta vulgaris* L.). *Genome* **46**, 70–82.

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. *Proc. Natl. Acad. Sci. USA* **93**, 11746–11750.

Johal GS, Briggs SP (1992) Reductase activity encoded by the *Hm1* disease resistance gene in maize. *Science* **258**, 985–987.

Khedikar YP, Gowda MV, Sarvamangala C, Patgar KV, Upadhyaya HD, Varshney RK (2010) A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* **121**, 971–984.

Kochert G, Halward T, Branch WD, Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theor. Appl. Genet.* **81**, 565–570.

Leal-Bertioli SC, Jose AC, Alves-Freitas DM, Moretzsohn MC, Guimaraes PM, Nielen BS, Perira RW, Favero AP, Parniske M, Varshney RK, Bertioli DJ (2009) Identification of candidate genome regions controlling disease resistance in *Arachis*. *BMC Plant Biol.* **9**, 112.

Liang X, Chen X, Hong Y, Liu H, Zhou G, Li X, Guo B (2009) Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biol.* **9**, 35.

Liu Z, Crampton M, Todd A, Kalavacharla V (2012) Identification of expressed resistance gene-like sequences by data mining in 454-derived transcriptomic sequences of common bean (*Phaseolus vulgaris* L.). *BMC Plant Biol.* **12**, 42.

Luo M, Liu J, Lee RD, Scully BT, Guo BZ (2010) Monitoring the expression of maize genes in developing kernels under drought stress using oligo-microarray. *J. Integr. Plant Biol.* **52**, 1059–1074.

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432–1436.

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

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.

Nagy ED, Chu Y, Guo Y, Khanal S, Tang S, Li Y, Dong WB, Timper P, Taylor C, Ozias-Akins P, Holbrook CC, Beilinson V, Nielsen NC, Stalker HT, Knapp SJ (2010) Recombination is suppressed in an alien introgression in peanut harboring *Rma*, a dominant root-knot nematode resistance gene. *Mol. Breed.* **26**, 357–370.

Paik-Ro OG, Smith RL, Knauft DA (1992) Restriction fragment length polymorphism evaluation of six peanut species within the *Arachis* section. *Theor. Appl. Genet.* **84**, 201–208.

Pandey MK, Feng S, Culbreath A, Varshney RK, Wang ML, Barkley NA, Holbrook CC, Guo BZ (2012a) Saturation of genetic maps for identification of QTLs controlling biotic resistance, morphological descriptors and oil quality in tetraploid peanut (*Arachis hypogaea* L.).

American Peanut Research and Education Society Annual Conference, July 10-12, 2012, Raleigh, NC.

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimarães P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B, Cook DR, Bertioli DJ, Michelmore R, Varshney RK (2012b) Advances in *Arachis* genomics for peanut improvement. *Biotech. Adv.* **30**, 639–651.

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: A review. *Mol. Breed.* **7**, 275–291.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res.* 2012, **40**, D290–D301.

Qin H, Feng S, Chen C, Guo Y, Knapp S, Culbreath A, He G, Wang ML, Zhang X, Holbrook CC, Ozias-Arkins P, Guo B (2012) An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. *Theor. Appl. Genet.* **124**, 653–664.

Ratnaparkhe MB, Wang X, Li J, Compton RO, Rainville LK, Lemke C, Tang H, Paterson AH (2011) Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny. *New Phytol.* **192**, 164–178.

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA, D' Hont A (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. *Mol. Genet. Genomics* **269**, 406–419.

Sanseverino W, Roma G, Simone MD, Faino L, Melito S, Stupka E, Frusciante L, Ercolano MR (2009) PRGdb: A bioinformatics platform for plant resistance gene analysis. *Nucleic Acids Res.* **38**, D815.

Seijo G, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EA (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea* – Leguminosae) and its close relatives revealed by double GISH. *Am. J. Bot.* **94**, 1963–1971.

Seijo JG, Lavia GI, Fernandez A, Krapovickas A, Ducasse D, Moscone EA (2004) Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and

A. ipaensis are the wild diploid progenitors of *A. hypogaea* (Leguminosae). *Am. J. Bot.* **9**, 1294–1303.

Shirasawa K, Koilkonda P, Aoki K, Hirakawa H, Tabata S, Watanabe M, Hasegawa M, Kiyoshima H, Suzuki S, Kuwata C, Naito Y, Kuboyama T, Nakaya A, Sasamoto S, Watanabe A, Kato M, Kawashima K, Kishida Y, Kohara M, Kurabayashi A, Takahashi C, Tsuruoka H, Wada T, Isobe S (2012) *In silico* polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. *BMC Plant Biol.* **12**, 80.

Shiu SH, Bleecker AB (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. USA* **98**, 10763–10768.

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

Staskawicz BJ, Ausubel FM, Baker BJ, Ellis JG, Jones JD (1995) Molecular genetics of plant disease resistance. *Science* **268**, 661–667.

Sujay V, Gowda MV, Pandey MK, Bhat RS, Khedikar YP, Nadaf HL, Gautami B, Sarvamangala C, Lingaraju S, Radhakrishan T, Knapp SJ, Varshney RK (2012) Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.)**.** *Mol. Breed.* **30**, 773–788.

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422.

Van Ooijen JW (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet. Res.* (Camb) **93**, 343–349.

Varshney RK, Bertioli DJ, Moretzsohn MC, Vadez V, Krishnamurthy L, Aruna R, Nigam SN, Moss BJ, Seetha K, Ravi K, He G, Knapp SJ, Hoisington DA (2009) The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* **118**, 729–739.

Wang H, Penmetsa RV, Yuan M, Gong L, Zhao Y, Guo B, Farmer AD, Rosen BD, Gao J, Isobe S, Bertioli DJ, Varshney RK, Cook DR, He G (2012) Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* **12**, 10.

Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene *N*: Similarity to toll and the interleukin-1 receptor. *Cell* **78**, 1101–1115.

Wisser RJ, Sun Q, Hulbert SH, Kresovich S, Nelson RJ (2005) Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. *Genetics* **169**, 2277–2293.

Xiao WK, Xu ML, Zhao JR, Wang FG, Li JS, Dai JR (2006) Genome-wide isolation of resistance gene analogs in maize (*Zea mays* L.). *Theor. Appl. Genet.* **113**, 63–72.

Xiao WK, Zhao J, Fan SC, Li L, Dai JR, Xu ML (2007) Mapping of genome-wide resistance gene analogs (RGAs) in maize (*Zea mays* L.). *Theor. Appl. Genet.* **115**, 501–508.

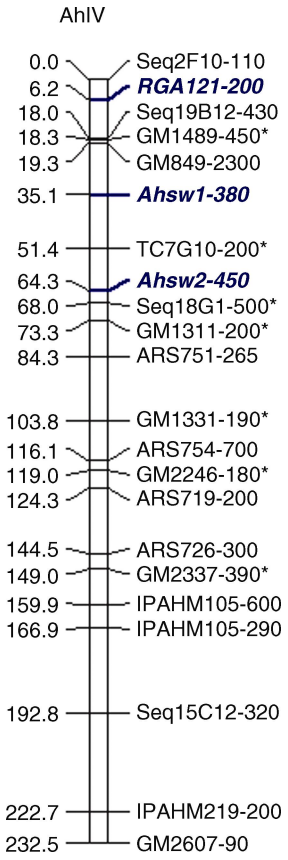
You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**, 253.

Yuksel B, Estill JC, Schulze SR, Paterson AH (2005) Organization and evolution of resistance gene analogs in peanut. *Mol. Genet. Genomics* **274**, 248–263.

Supporting Information

Figure

Figure 1. Linkage mapping of RGA121and *Ahsw* putative genes on peanut map.



Tables

Table 1. Known *R* genes from plants used in this study

Plant	*R* genes	Protein GB ID	Structure	Plant	*R* genes	Protein GB ID	Structure
Apple	*Vf1*	CAC40825	LRR	Potato	*R1*	AAU95638	CC-NBS-LRR
Arabidopsis	*Fls2*	BAB11088	LRR-PK		*Rgc1*	AAF76163	CC-NBS-LRR
	Pbs1	ABR46085	PK		*Rx*	CAB50786	CC-NBS-LRR
	Rpm1	CAA61131	CC-NBS-LRR	Rice	*Pib*	BAA76282	CC-NBS-LRR
	Rpp1	AEE77906	TIR-NBS-LRR		*Pi-ta*	BAF91352	CC-NBS-LRR
	Rpp4	AAE83818	TIR-NBS-LRR		*Rpr1*	BAA75812	CC-NBS-LRR
	Rpp5	AAE83827	TIR-NBS-LRR		*Xa1*	BAA25068	CC-NBS-LRR
	Rpp8	AAC78631	CC-NBS-LRR		*Xa21*	AAC80225	LRR-PK
	Rpp13	AAF42832	CC-NBS-LRR		*Xa26*	ABD36512	LRR-PK
	Rps2	AAM90883	CC-NBS-LRR	Sugarbeet	*Hs1*	AAW03319	LRR-TM
	Rps4	CAB50708	TIR-NBS-LRR	Tobacco	*N*	AAA50763	TIR-NBS-LRR
	Rps5	AAC26126	CC-NBS-LRR	Tomato	*Cf-2*	AAC15779	LRR-TM
	Rpw8.1	ACJ05900	TM		*Cf-4*	CAA05268	LRR-TM
	Rpw8.2	AAP45326	TM		*Cf-5*	AAC78591	LRR-TM
	Ssi4	AAN86124	TIR-NBS-LRR		*Cf-9*	CAA05274	LRR-TM
	Rcy1	BAC67706	CC-NBS-LRR		*Hero*	CAD29728	CC-NBS-LRR
Barley	*Rpg1*	ABK51312	PK		*I2*	AAD27815	CC-NBS-LRR
	Mla1	AAG37354	CC-NBS-LRR		*Mi-1*	AAC97933	CC-NBS-LRR
	Mla6	CAC29242	CC-NBS-LRR		*Prf*	AAC49408	CC-NBS-LRR
	Mlo	CAB06083	TM		*Pto*	AAB47421	PK
Flax	*L6*	AAA91022	TIR-NBS-LRR		*Sw-5*	AAG31013	CC-NBS-LRR
	M	AAB47618	TIR-NBS-LRR		*Cre3*	AAC05834	CC-NBS-LRR
	P2	AAK28806	TIR-NBS-LRR	Wheat	*Vrga1*	AAF19148	CC-NBS-LRR
Maize	*Hm1*	AAC04333	Toxin reductase		*Yr10*	AAG42168	LZ-NBS-LRR
	Hm2	ABY68564	Toxin reductase		*Lr1*	ABS29034	CC-NBS-LRR
	Rp1-d	AAD47197	CC-NBS-LRR		*Lr10*	AAQ01784	CC-NBS-LRR
Lettuce	*Dm3*	AAD03156	CC-NBS-LRR	Pepper	*Bs2*	AAF09256	CC-NBS-LRR

Table 2. The distribution of peanut *R*-gene-like expressed sequence tags (ESTs)

cDNA-ESTs	Leaves	Roots	Seeds	Gynospores	Total
GT-C20	27 (0.32%)*		18 (0.21%)		45 (0.26%)
Tifrunner	65 (0.35%)		129 (0.39%)		194 (0.37%)
Florunner	1 (0.40%)				1 (0.40%)
850	9 (1.21%)				9 (1.21%)
A13	3 (0.19%)		2 (0.40%)		5 (0.24%)
Luhua 14			31 (0.36%)		31 (0.36%)
Minhua 6			4 (0.52%)		4 (0.52%)
Shanyou 523			13 (0.59%)		13 (0.59%)
Yueyou 523			5 (1.30%)		5 (1.30%)
06-4104	33 (0.15%)	95 (0.46%)	72 (0.35%)		200 (0.32%)
Chiba-Handachi	70 (4.60%)	28 (0.42%)		44 (0.77%)	142 (0.52%)
A. ipaensis PI468321		73 (0.46%)	83 (0.49%)		156 (0.48%)
A. stenosperma V10309	0 (0.00%)	29 (0.48%)			29 (0.46%)
A. duranensis DUR25		108 (0.74%)	111 (0.54%)		219 (0.62%)
Total	208 (0.53%)	333 (0.52%)	468 (0.42%)	44 (0.77%)	1053 (0.42%)

*****The percentage of *R*-gene-like ESTs in the total ESTs from corresponding tissues.

Table 3. The genotyping data of RGA121-SSR in T and S populations

RIL population	No. of allele from Parent 1	No. of allele from Parent 2	Heterozygous	Missed band	Total
T population	88	72	5	0	165
S population	161	157	16	29	363