EVOLUTIONARY, GENETIC AND PHYSICAL ANALYSIS OF RESISTANCE GENE-LIKE SEQUENCES IN THE PEANUT GENOME AND EXPLORATORY WORK TOWARD INTEGRATION OF PEANUT GENETIC AND PHYSICAL MAPS

by

BAYRAM YÜKSEL

(Under the Direction of Andrew H. Paterson)

ABSTRACT

The scarcity of genetic diversity in the domesticated peanut, *Arachis hypogaea*, gene pool makes the future of peanut cultivation especially vulnerable. We constructed and characterized a peanut *Hin*dIII BAC library containing 182,784 clones with an average insert size of 104 kb with the hope of providing better means for analysis of peanut genome organization and evolution. Although no disease resistance genes have been cloned from peanut itself, the conserved motifs in cloned resistance genes from other plants provided a means to isolate and analyze similar genes from peanut. To survey the number, diversity, evolutionary history, and genomic organization of resistance gene-like sequences in *Arachis hypogaea*, we isolated 234 resistance gene analogs (RGA) from the peanut genome by using primers designed from conserved regions of different classes of resistance genes including NBS-LRR, and LRR-TM classes. Phylogenetic and sequence analyses of the RGA sequences were done to explore evolutionary relationships. Overgos designed from the RGA sequences on the basis of their phyletic association were applied to the peanut BAC library; 736 uniquely detected BAC clones were fingerprinted and contigs were formed in order to gain insights into the genomic organization of these genes. As a

result, we identified 250 putative resistance gene loci. As a secondary part of this project, we investigated the practicality of physical mapping in *A. hypogaea*, by integrating 117 previously genetically mapped probes into the physical map. A total of 576 *Arabidopsis* derived overgos, which were designed from the most conserved regions of orthologous genes, were also applied to the library with about a 61.5% success rate. The study has shown that *Arabidopsis* genome sequence is a valuable stepping stone towards opening new avenues in the analysis of complex genomes like peanut.

INDEX WORDS: Peanut, *Arachis hypogaea*, BAC Library, Resistance Gene Analogs, Physical Map, *Arabidopsis*, Overgo

EVOLUTIONARY, GENETIC AND PHYSICAL ANALYSIS OF RESISTANCE GENE-LIKE SEQUENCES IN THE PEANUT GENOME AND EXPLORATORY WORK TOWARD INTEGRATION OF PEANUT GENETIC AND PHYSICAL MAPS

by

BAYRAM YÜKSEL

B.S., Boğaziçi University, TURKEY, 1996

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

EVOLUTIONARY, GENETIC AND PHYSICAL ANALYSIS OF RESISTANCE GENE-LIKE SEQUENCES IN THE PEANUT GENOME AND EXPLORATORY WORK TOWARD INTEGRATION OF PEANUT GENETIC AND PHYSICAL MAPS

by

BAYRAM YÜKSEL

Major Professor: Andrew H. Paterson

Committee: Roger H.Boerma
Wayne A. Parrott
Richard S. Hussey
Joe H. Bouton

Electronic version approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2004

DEDICATION

I dedicate this thesis to my family, especially my sister, Emine Yüksel, for their unwavering support and patience they showed during my long years of graduate school adventure.

ACKNOWLEDGEMENTS

I am deeply grateful to all of the Plant Genome Mapping Lab personnel for their generous support and friendship.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW:

Arachis hypogaea L. is one of the world's most important food and oil-seed crops with 37,057,652 Gg of world-wide production (Food and Agriculture Organization 2003). Peanuts are grown worldwide; leading producers are China and India. Although *Arachis* spp. are widely distributed throughout most of South America, their center of origin is usually believed to be in the region between northern Argentina and southern Bolivia (Stalker and Simpson 1995). Section Arachis of the genus *Arachis* contains about 69 species, which are mostly A and B genome diploids with the exception of the D genome diploid *A. glandulifera* (Stalker 1991; Stalker et al. 1991). The cultivated peanut, *Arachis hypogaea* L., and its wild relative *Arachis monticola* (2n=4x=40) are the only known AABB tetraploids (Krapovickas and Gregory 1994). Most estimates of the genome size of cultivated peanut and its wild relative fall between 2800 to 3000 Mb (Arumuganathan and Earle 1991; Singh et al. 1996; Temsch and Greihulber 2000). The diploid ancestry of the allotetraploid peanut remains somewhat unclear. Molecular markers developed by Kochert et al. (1996) suggested *A. duranensis* Krapovickas as A and *A. ipaensis* as B genome donors. Additionally, Jung et al (2003) argued for *A. duranensis* and *A. ipaensis* as the possible ancestors of peanut on the basis of the evolution of stearoyl-ACP desaturase and oleoyl-PC desaturase genes. However, cytogenetical and molecular marker studies done by Raina et al. (1999; 2001) suggested *A. villosa* as A and *A. ipaensis* as B genome donors. Better understanding of genome structure of *A. hypogaea* may help to clarify questions related to its ancestry.

The low level of genetic diversity both among peanut cultivars and between the two tetraploid species has been very problematic for peanut breeders and has made peanut cultivation vulnerable to many diseases. This lack of genetic diversity caused breeders to look for agronomically significant gene sources among the wild diploid relatives. Although the diploid species have been successfully exploited for the introgression of a few important genes such as the root-knot nematode (*Meloidogyne arenaria*) resistance gene (*Ma*) (Abdel-Momen and Starr 1998; Starr and Simpson 1990), the process is very time consuming, costly, and the effectiveness of this type of monogenic resistance under constant selective pressure is uncertain. This fact necessitates the investigation of alternative approaches.

Genetic and molecular studies in peanut have been very slow relative to other legume species such as soybean and *Medicago*. The first genetic map of *Arachis spp.* was constructed by crossing two related diploids *A. stenosperma* and *A. cardenasii* (Halward et al. 1993), and consists of 117 loci. Only one tetraploid map, containing 367 loci on 23 linkage groups has been developed by crossing *A. hypogaea* and a synthetic amphidiploid TxAG6 (Burow et al. 2001). Six polymorphic Simple Sequence Repeat (SSRs) markers were described several years ago (Hopkins et al. 1999), and another 110 have been recently reported (Ferguson et al. 2004). Ribosomal DNA repeats have been used for assessment of the *Arachis* evolutionary lineage (Singh et al. 2001).

Technology for the construction of large insert libraries has added new dimensions to genome analysis of a wide range of plants and animals. Yeast Artificial Chromosome (YAC) libraries were the method of choice initially (Burke et al. 1987). YAC libraries of some important crops have been successfully employed for map-based cloning of important genes such as *pto* from tomato, and *RPS2* and *RPM1* from *Arabidopsis* (Bent et al. 1994; Grant et al. 1995;

Martin et al. 1993). Problems inherent to YACs resulted in reduced use of these libraries (Anderson 1993). The most fundamental issue is difficulties associated with manipulation of yeast clones. Instabilities and rearrangements, especially in clones consisting of highly repetitive DNA, have also been reported (Dunford et al. 1993; Neil et al. 1990). The abundance of chimeric clones, which interferes with map-based cloning, has been another major glitch (Larionov et al. 1994; Schmidt et al. 1996).

Bacterial Artificial Chromosome (BAC) libraries have emerged as a method of preference for the construction of large insert clones. Overall, bacteria grow much faster than yeast and bacterial clones are much easier to work with. The first F-factor based vector in *E. coli* for the propagation of large inserts was described by O'Connor et al. (1989). pBAC, which allowed cloning and maintenance of DNA fragments up to 300 kb, was constructed by Shizuya et al. (1992). F-factor based genes *parA* and *parB* ensure the single copy maintenance of the plasmid, preventing co-transformation and recombinant clones, common problems in YACs (Larionov et al. 1994; Willets and Skurray 1987). Addition of the lacZ gene to the vector system allowed color-based positive screening (Kim et al. 1996). Few chimeric clones have been detected by Fluorescence *In Situ* Hybridization (FISH) of BACs with large inserts (Kim et al. 1996; Shizuya et al. 1992; Woo et al. 1994). Serial dilutions and longer cycles of growth have supported the secure and stable propagation of BAC clones (Shizuya et al. 1992). Modified BAC vectors made possible the direct transfer of large DNA fragments into plant cells (Hamilton et al. 1996). Although BAC clones with insert sizes up to 300 kb have been reported, the average insert size of BAC libraries is usually between 100 kb and 150 kb. Therefore, it could be still necessary to use YACs, if clones with insert size larger than 300 kb are required.

BAC libraries for many important crops from diverse taxonomic families of plants already exist; *S. bicolor* (Woo et al. 1994), *Arabidopsis* (Choi et al. 1995), papaya (*Carica papaya*-Ming et al. 2001), sunflower (*Helianthus annuus*-Gentzbittel et al. 2002), lettuce (*Lactuca sativa*-Frijters et al. 1997), peach (*Prunus persica*-Wang et al. 2001a), rice (*Oryza sativa*-Wang et al. 1995), wheat (*Triticum aestivum*-Cenci et al. 2003), banana (*Musa acuminata*-Vilarinhos et al. 2003), and cauliflower (*Brassica oleracea var. Botrytis*-Li et al. 2003). In the Leguminosae, multiple libraries have been made for soybean (*Glycine max*) and *Medicago* (*Medicago sativa*-Danesh et al. 1998; Meksem et al. 2000; Nam et al. 1999; Salimath and Bhattacharyya 1999; Tomkins et al. 1999). Common bean (*Phaseolus vulgaris*) is another crop from this family for which a BAC library has previously been constructed (Vanhouten and MacKenzie 1999). No BAC library has previously been constructed for any member of the *Arachis* genus.

BAC libraries have been utilized for many purposes including map-based cloning of genes such as ebisu dwarf (*d2*) from rice (Hong et al. 2003) and rust resistance (*Rpg1*) from barley (Brueggeman et al. 2002). Similarly, BAC libraries are often used for seeking particular regions of genomes, for instance; for identification of putative transmembrane receptor gene loci throughout the sunflower genome (Gentzbittel et al. 2002) and for studying the distribution of resistance gene candidates in the citrus genome (Deng et al. 2001a). In the plants for which dense genetic maps exist, chromosome landing for a gene of interest is possible (Wang et al. 2001b). BAC-based physical maps are extremely valuable in whole-genome sequence assembly, and have been a cornerstone in the sequencing of many complex genomes (Venter et al. 1998). BAC-based physical maps assist in pinpointing gene-rich regions in complex genomes, in exploring colinearity among genomes at the local level (Chen et al. 1997), and in formation of cytogenetic maps integrating genetic distances with physical distances (Kulikova et al. 2001).

In this work, we describe the first BAC library for peanut, its application to localizing resistance gene analogs, and progress toward its utilization as a physical framework for a wide range of applications in peanut genomics.

I. CONSTRUCTION AND CHARACTERIZATION OF BAC LIBRARIES AND EXPLORATION OF THEIR APPLICATIONS IN PEANUT GENOMICS

A. Basic Steps For BAC Library Construction

a) Isolation of HMW DNA

Isolation of intact, digestible high molecular weight (HMW) DNA is pivotal for BAC library construction. This step is more complicated in plants than animals because of the presence of a cell wall. Initial methods of choice for isolation of HMW DNA relied on the production of protoplasts by dissolution of cell-walls with cellulase. This method had very limited applicability due to lack of standard procedures for protoplast production for many plant species. Another major problem has been high ratios of chloroplast and mitochondrial DNA contamination. A better and more universal procedure was developed later by Woo et al. (1995). Isolation of intact and clean nuclei is the most crucial step for the success of this procedure. Since the development of general DNA extraction protocols based on the isolation of nuclei, some adjustments in these protocols has been enough for isolation of intact nuclei in many species. The isolated nuclei are lysed in agarose plugs providing a supportive matrix for fragile DNA fragments. This method has produced satisfactory results for many plant species. However, an abundance of starch granules, which have similar buoyancy as nuclei, and phenolic compounds, have been major obstacles in some species such as banana (Vilarinhos et al. 2003).

b)Vector isolation and preparation:

The low copy number of the BAC vector requires large amounts of culture for isolation of enough plasmid. This problem has been solved by cloning pINDIGO536 into the high-copy vector pGEM-4Z (pCUGIBAC1) (Luo et al. 2001). A more important issue is the false positives produced by damaged cohesive ends; up to 17 % false positives have been reported (Wang et al. 2001a). Freeze-thaw cycles exacerbate this problem resulting in high numbers of clones with no insert.

c) Digestion and fractionation of HMW DNA:

Optimization of fractionation conditions to maximize the size of DNA fragments that are suitable for ligation, is essential. The major obstacle in this step is dilution of large size ligatable fragments either because of inefficient resolution of small fragments or degradation of DNA during digestion, producing large numbers of fragments with damaged ends. For good quality DNA fractionation, one size selection is usually not sufficient; application of a second size selection is necessary (Osoegawa et al. 1998).

d) Elution of DNA from gel:

The digestion of agarose with β-agarase is one way of elution, but this often damages DNA fragments causing low ligation efficiency. Electroelution has produced the best results (Osoegawa et al. 1998; Strong et al. 1997).

e) Ligation

The adjustment of vector/insert ratio is vital for efficiency and minimizing the number of chimeric clones. This ratio can vary from 1:4 to 1:8 depending on the species or more likely on the ratio of ligatable DNA fragments in the eluted sample after size selection. Low insert

concentration may improve the average insert size, but usually diminishes efficiency. On the other hand, high insert concentration may produce too many chimeras.

II. EVOLUTIONARY, GENETIC, AND PHYSICAL ANALYSIS OF RESISTANCE GENE ANALOGS IN PEANUT

A. Peanut and Disease Resistance

In monophyletic polyploid species such as cultivated peanut, *Arachis hypogaea*, genetic variation is very limited. The low level of genetic polymorphism makes peanut vulnerable to many pathogens. For instance, the peanut cultivars cultivated in the southern USA are highly susceptible to the nematode *Meloidogyne arenaria*, and damage in highly infested fields can be devastating. Yield losses are estimated to be about 30% on average (Nelson et al. 1989). A soil-borne pathogen *Sclerotium rolfsii* Sacc causes stem rot with devastating results (Branch and Brenneman 1999). *Cylindrocladium* black rot and tomato spotted wilt virus are also very harmful diseases of peanut requiring a solution (Branch and Brenneman 2003). Cultivated peanut is also vulnerable to other pathogens such as bud necrosis caused by tospovirus (Pensuk et al. 2004), Sclerotinia blight, collar rot caused by *Lasiodiplodia theobromae* (Phipps and Porter 1998) and many others. Until now the main strategy of defense against these diseases has been either screening of different cultivars for resistance or transfer of resistance genes from primary or secondary gene pools of *Arachis* spp. However, it is rare that the first approach works, because most peanut cultivars are very similar genetically. The second approach is very cumbersome and time-consuming. Toward the development of better approaches, it is a valuable starting point to analyze the spectrum of peanut genes that resemble disease resistance genes from other plants, their physical distribution in the peanut genome, and their relationships. Such information

provides a valuable starting point for rapidly searching for specific gene(s) conferring resistance to existing or new pathogens.

B. Resistance Genes:

Plants must defend themselves against a wide range of pathogens including bacteria, viruses, fungi, insects, and nematodes. Essential to a successful defense strategy is the recognition of the intruder; plant resistance genes play an early role in this perception by acting as sentinels according to the "guard hypothesis" (Dangl and Holub 1997; Fluhr 2001; Hulbert et al. 2001). Contact between a resistance gene and its corresponding avirulence product of pathogens, or possibly virulence factors, which can occur intra- or extracellularly depending on the location of the gene product (Baker et al. 1997), usually triggers a hypersensitive response (HR-Dangl et al. 1996; Morel and Dangl 1997) and activates downstream defense-related signal pathways. Such direct physical interactions have been experimentally confirmed either by the yeast-two-hybrid system as in *avrPto* and *Pto* conferring resistance to *Pseudomonas syringae* (Tang et al. 1996) or by *in vitro* binding assays as in *avrPita* and *Pi-ta* rendering resistance to the fungal pathogen *Magnaporthe grisea* (Jia et al. 2000). Moreover, *in vivo* experiments done by Leister et al. (2000) have further supported the ligand-receptor interaction by demonstrating direct contact between *RPS2* and *RPM1* products and their corresponding pathogen *avrPt2* and *avrB* products respectively, in mesophyll protoplasts of the *Arabidopsis* leaf. After initial perception, regardless of the type of pathogens, pathways leading to final response may converge (Staskawicz et al. 1995). Although most of the characterized resistance genes fit into this gene-for-gene concept (Flor 1956; Flor 1971), there are some exceptions such as *Hm1* in maize, which provides resistance against fungal pathogen *Cochliobolus carbonum race 1* by neutralizing HC toxin (Johal and Briggs 1992).

Most of the 46 plant resistance genes cloned to date can be categorized into the following classes on the basis of their cellular location and functionally conserved domains:

a) Class I:

The majority of these genes belong to the NBS-LRR class, all of which share about a 400 amino acid long NBS (Nucleotide-Binding Site), which is found in all ATP or GTP binding proteins, and functions by creating a pocket for specific binding of these nucleotides (Saraste et al. 1990; Traut 1994). The same domain is also found in mammalian apoptosis regulatory proteins such as *Apf-1* in humans (van der Biezen 1998). Specific interactions between plant receptor and pathogenic ligand are likely to initiate signal transduction by phosphorylation of downstream mediators ultimately resulting in induction of defense related responses or leading to autophosphorylation of the R gene as is the case in *Xa21* genes of rice (Liu et al. 2002; Tameling et al. 2002). The NBS region can be further subdivided into P-loop or kinase-1a domain functions by binding of ATP or GTP (Saraste et al. 1990). The kinase-2a domain acts by binding metallic ion Mg-ATP and kinase-3a domain coordinates purine base binding (Traut 1994).

The N-termini of these genes are composed of two different kinds of motif; TIR, which is homologous to human interleukin-1 and Drosophila toll-like receptor regions, and coiled-coil (CC) leucine zips. This difference creates two distinct subcategories; TIR-NBS-LRR and CC-NBS-LRR. The C-terminal region of each of these two categories is composed of varying numbers of XXLXLXX motifs. Typical examples for the first subclass are the tobacco mosaic virus resistance gene *N* (Whitham et al. 1996), and flax rust resistance genes *L6* and *M* (Anderson et al. 1997; Lawrence et al. 1995). The tomato root-knot nematode resistance gene *Mi* (Milligan et al. 1998), *Psuedomonas syringae* resistance genes *RPS2* and *RPM1* in *Arabidopsis* (Bent et al. 1994; Grant et al. 1995), and *Fusarium oxysporum f sp lycopersici* resistance gene I2

(Simons et al. 1998) are some samples for the other subcategory, CC-NBS-LRR. Resistance genes with other combinations of these conserved domains, such as TIR-X, TIR-NBS, or NBS-LRR are also found in plant genomes (Meyers et al. 2003; Meyers et al. 2002).

To shed light on the issue of how plants recognize pathogens, detailed analysis of binding specificity between an avR gene and its corresponding resistance gene is a prerequisite. The high level of polymorphism in LRR components of allelic and non-allelic resistance genes have suggested that these repetitive motifs could play a role in recognition specificity (Bergelson et al. 2001). For example, most of the 13 flax *L* resistance gene alleles differed in their LRR region (Ellis et al. 1999); in the *Cf5* locus of tomato, the majority of allelic variation occurs in the same region (Dixon et al. 1998). Additionally, the observation of diversifying selection in genome-wide analysis of NBS-LRR genes further emphasized the potential role of these repeats in generation of the specificity (Kajava 1998; Mondragon-Palomino et al. 2002). Moreover, a single amino acid change in the same region of *Pi-ta* prevented interaction with *avRPita* (Jia et al. 2000). However, swapping LRR domains of functional *Mi1.2* and its nonfunctional copy *Mi1.1* has suggested that not just LRR, but also N-terminal structures may be involved in the recognition process (Hwang et al. 2000; Hwang and Williamson 2003). Likewise, exchanging *L2* TIR region with the corresponding region of *L6* supported the same possibility (Luck et al. 2000). Homolog structure to the N-terminal LZ motif in the light-response pathway gene *COP1* plays an essential role in regulation of light perception during seedling development (Torii et al. 1998). It is quite possible that hetero- and homo-dimeric structures may also take place in creation of avR-R interaction specificity. Overall, it is quite possible that for different R-avR pairs, specificity may lie in different regions or combinational interactions of these regions.

The distribution of these two classes of genes across plant species is quite different; both TIR-NBS-LRR and NonTIR-NBS-LRR genes are found in dicots, while only Non-TIR class genes have so far been observed in monocots (Cannon et al. 2002; Meyers et al. 2003; Pan et al. 2000b). After thorough analysis of the rice genome sequence, Bai et al. (2002) could not find any TIR homologs. Since some TIR homologs have been cloned from gymnosperms (Liu and Ekramoddoullah 2003), it is likely that both of these classes existed in a common ancestor of angiosperms and were eliminated from monocots. However, the absence of the TIR class in the *Beta vulgaris* genome (a dicot) may suggest that the fate of these two classes may be different in different plant lineages (Tian et al. 2004).

Arabidopsis CC-NBS-LRR (CNL) and TIR-NBS-LRR (TNL) classes differ in their signaling pathway components; while most TNL genes act through *EDS1*, CNL genes act through *NDR1*(Aarts et al. 1998b). This difference may suggest that different sets of defense responses may be activated. However, it should be emphasized that there is not much known about how these two different classes function in plant species other than *Arabidopsis*. Moreover, some genes in these two categories may not depend on either of these signaling components, or may use both of them suggesting that there could be functional overlaps (Aarts et al. 1998b).

Many resistance genes are located in close proximity to one another in plant genomes. For example, about 2/3 of 150 NBS-LRR class genes in the *Arabidopsis* genome are organized in 43 different clusters (Cannon et al. 2002; Meyers et al. 2003; The Arabidopsis Genome Initiative 2000). The longest array consisting only of R genes in *Arabidopsis* is the RPP4/RPP5 complex, which is composed of seven TIR-NBS-LRR genes extending over 90 kb (Meyers et al. 1999; Noel et al. 1999). Megaclusters (i.e. clusters, which are interrupted with other genes) in the

Arabidopsis genome extend 2.53 Mb in chromosome 1 and 4.46 Mb in chromosome 5 (Richly et al. 2002)

The clustering characteristics of resistance genes have also been observed in other plant species. For instance, in lettuce a region containing 10 *Dm* genes, which confer resistance to oomycete downy mildew fungus *Bremia lactucae* with other very divergent *RGC2* NBS-LRR gene complexes, spans more than 3.5 Mb (Meyers et al. 1998). Seven *I2* gene homologs on tomato chromosome 11 extend over a 90 kb distance with 8 to 10 kb long intergenic space (Simons et al. 1998). The rust resistance locus, *M*, in flax includes 15 other related genes (Anderson et al. 1997), and the *L* locus consists of six paralogs of rust resistance genes. The complex rust resistance locus *Rp1* in maize is composed of *Rp1-D* and eight of its paralogs (Sun et al. 2001). The *Xa21* locus in rice is composed of A1, A2, B, C, D, E, and F families and in some of them sequence similarity is as high as 98.0 % (Song et al. 1997). The clustered nature of many resistance gene families is likely to be a result of the evolutionary history of this family of genes.

The same genomic organization of resistance genes has been detected in legumes. A BAC sequence from soybean contained 16 different R-genes with 86% sequence homology in the exons (Graham et al. 2000). Soybean mosaic virus resistance gene *Rsv1* and other important resistance genes mapped to the same cluster on linkage group F (Hayes and Saghai Maroof 2000).

The phyletic makeup of resistance gene clusters is important for understanding the evolutionary history of this class. TNL and CNL classes never coexist in the same cluster in *Arabidopsis* (Richly et al. 2002). Although the majority of clusters in *Medicago* were homogenous, occasional mixed clusters containing both TNL and CNL were observed (Zhu et al.

2002). A total of 25 out of 43 clusters in *Arabidopsis* are composed of genes from monophyletic clades (Meyers et al. 2003). The presence of some heterogeneous clusters could be explained by intragenomic rearrangements at least in the *Arabidopsis* genome (Baumgarten et al. 2003; Cannon et al. 2002; Meyers 2003; Pan et al. 2000b). Phyletically heterogeneous clusters were also detected in the *Medicago truncatula* genome (Zhu et al. 2002). The monophyletic origin of most clusters supports the idea of derivation of paralogs from a single ancestral gene, possibly through unequal crossing-over and duplication processes.

For the maximal exploitation of model plant genome sequences such as *Arabidopsis*, rice, or *Medicago*, it is important to know the degree of microcolinearity between orthologous loci. There have been conflicting reports about the degree of rearrangement from different plant families. Leister et al. (1998) observed a high level of disparity in copy numbers of paralogs intraspecifically and the orthologous loci among foxtail millet, barley, and rice were highly rearranged. Comparative analysis of the *Rp1* orthologs of maize and sorghum revealed numerous nonsyntenic characteristics despite some shared features (Ramakrishna et al. 2002b). These observations could be interpreted as suggestive that synteny is not well maintained because of the dynamically evolving nature of resistance gene clusters, at least for grasses. However, *Capsicum* homologues of cloned R genes, *Sw-5*, *N*, *Pto*, *Prf*, and *I2* were found in syntenic positions in the *Lycopersicon* and *Solanum* genomes (Grube et al. 2000). A similar degree of conservation is observed among corresponding R gene loci among *Solanum* genomes, but these syntenies were lost between *Arabidopsis* and tomato (Pan et al. 2000a). In contrast, some incongruities were observed in synteny between orthologous R gene loci of these genomes, for example, *Globodera spp*., potato virus Y, and tobacco mosaic virus resistance genes either did not exist in some genomes or mapped to entirely different locations (Grube et al. 2000). It may

be concluded that the speed of micro-scale rearrangements in R gene clusters may have been much slower in the *Solanaceae* than the *Poaceae* family. Although markers flanking the *Rpm1* locus existed in both the *Arabidopsis* and *Brassica* genomes, *Rpm1* itself was missing in *Brassica* (Grant et al. 1998). On the other hand, some resistance gene clusters were syntenic between *M. truncatula* and soybean, and between *M. truncatula* and pea (Zhu et al. 2002). Overall, it would be prudent to approach very cautiously any attempt in using comparative map-based approaches for resistance gene cloning.

Plants have to keep up with constantly evolving pathogen populations; therefore, the generation of novel specificities is vital. One way of generating novel specificities is modification of a gene for defense against multiple pathogens. These kinds of genes with multiple functions have been found in some plants; for example; the root-knot nematode resistance gene *Mi* confers resistance against very distantly related pathogens such as the aphid, *Macrosiphum euphorbiae* and the whitefly, *Bemisia tabaci* (Milligan et al. 1998; Nombela et al. 2003; Rossi et al. 1998). The other strategy is the creation of a new function from duplicates of a resistance gene. Since redundancy of extra copies of a gene may be free of initial negative selectional constraints, duplicate copies may evolve to encounter challenges from different strains of the same pathogen. One could also argue that the genes would become irrelevant and eventually die with the decline of a particular strain of pathogen against which they function. The presence of multiple pseudogenes in most resistance gene clusters has led to the idea of birth-and-death cycles in plant-pathogen coevolution (Michelmore and Meyers 1998). On the other hand, there have been some counterarguments against this hypothesis; the coexistence of susceptible and resistance alleles at the *RPM1* locus of *Arabidopsis* have suggested concerted action of diversifying and purifying selections (Stahl et al. 1999). Likewise, phenotypic variation

at *RPS2* loci further supported the same idea (Mauricio et al. 2003). Better understanding of plant-pathogen coevolution would likely play a critical role in development of future strategies in the fight against plant pathogens.

The main mechanisms for generation of novel specificities are unequal crossing-over, (this could be either intergenic, resulting in chimeric copies, or intragenic, resulting in an increase in internal repeat numbers), gene conversion, point mutations, and transposon insertions (Ellis et al. 2000; Holub 2001; Hulbert et al. 2001; Richly et al. 2002; Ronald 1998). Deng et al. (2001a) have studied the Dm locus in lettuce and detected 16 spontaneous mutations; 11 of 12 mutations at the *Dm3* locus caused chromosomal deletions and increase in copy number, but none of them produced chimeras. Only one of these 12 mutations was a result of intragenic recombination causing gene conversion and generating a novel chimeric gene. The maize rust resistance gene locus *Rp1* has been extensively studied and numerous illegitimate recombination and gene conversion events have been observed (Collins et al. 1999; Hu and Hulbert 1994; Ramakrishna et al. 2002a; Richter et al. 1995; Sudupak et al. 1993; Sun et al. 2001). Varying numbers of LRR repeats among *Cf-5* and *L* gene homologs has further affirmed the significance of intragenic recombination for the creation of novel specificities (Dixon et al. 1998; Ellis et al. 1999). Likewise, at another tomato fungal resistance locus, *Cf-9*, interlocus exchange among homologs has also been observed (Parniske and Jones 1999). Transposon insertion is likely to be a secondary driving force behind the rapid diversification of resistance genes, since there are many cases where resistance genes are closely associated with retrotransposons, such as *RPP5*, *RPS5* and *Mla* loci in *Arabidopsis* and barley, respectively (Henk et al. 1999; Noel et al. 1999; Wei et al. 2002). Intergenic recombination and gene conversion may not be among the major factors determining the fate of resistance genes, since in most cases it has been observed that paralogs

are much more diverse than orthologs (Michelmore and Meyers 1998). Even in some cases such as the *Mla* locus in barley, recombination is suppressed (Wei et al. 2002). In conclusion, all of these evolutionary forces: unequal crossing-over, gene conversion, point mutations, and transposon insertion may have played roles in resistance gene evolution though with varying degrees of significance.

b) Class II:

Genes in this category are mainly composed of two major domains, extracytoplasmic LRR repeats and a short transmembrane (TM) domain. The first gene from this class, *Cf9* providing race-specific resistance against *Cladosporium fulvum*, was cloned in tomato by transposon-tagging (Jones et al. 1994). Homologs of this gene, *Cf2*, *Cf4*, and *Cf5* conferring resistance to different strains of the same fungus have been cloned separately and investigated in detail (Dixon et al. 1998; Dixon et al. 1996b; Thomas et al. 1997). *Cf* homologs mainly differ in their LRR domains; *Cf9* lacks two LRR repeat of *Cf4*, confirming the possible role of this domain in creation of novel recognition specificities (Van der Hoorn et al. 2001). The critical role of the same repeats is further pinpointed by Wulff et al. (2001), strongly suggesting that recognition specificity of these genes lie in these repeats.

The molecular mechanism of how these genes function has not yet been fully unraveled. The similarity between *Cf* and CLV, meristem regulatory proteins has led to speculation about the possibility of transduction of signal via protein complexes formed on the cell surface, however, direct contact between *Cf9* and *Avr9* has not been observed and none of the Rho-GTPase related proteins have been found in the complex (Rivas et al. 2002). This raises the possibility that these genes may induce resistance via entirely different pathways.

Genes similar to tomato "*Cf*" have been cloned and characterized from two other plant species. One of them is the cyst nematode, (*Heterodera schachtii Schmidt*) resistance gene *Hs1pro-1* from sugar beet. *Hs1pro-1* is composed of imperfect LRR and transmembrane domains (Cai et al. 1997). Four paralogs *Vfa1*, *Vfa2*, *Vfa3*, and *Vfa4* with similar structure to *Cf*-like genes; conferring resistance against fungal pathogen, *Venturia inaequalis* at the *Vf* locus of apple, have been identified and characterized (Xu and Korban 2002). Recently, a similar gene, *HcrVf2* from a wild apple species, has been cloned (Belfanti et al. 2004). That most fungal resistance genes have been in this category may hint at a common defense strategy of plants against ectoparasites.

c) Class III:

The structure of these genes is constituted from a single serine/threonine protein kinase domain. Cytoplasmically located *Pto*, conferring race-specific resistance against *Pseudomonas syringae pv. Tomato* is the first resistance gene ever cloned fitting the gene-for-gene model (Martin et al. 1993). *Pto*, like most later-characterized resistance genes, is located in a cluster consisting of about six paralogs in *L. pimpinellifolium*; *Pto* is missing in the syntenic locus of susceptible *L. esculentum*. (Riely and Martin 2001). Initial contact of *avrPto* and *Pto*, together with downstream mediators and eventual induction of cell death and oxidative burst responses have been extensively analyzed. The direct physical interaction between *Pto* and *avrPto* has been depicted *in vivo* by the yeast-two hybrid system (Tang et al. 1996). However, this interaction is not sufficient for activation of downstream mediators; *Prf*, which is a LZ-NBS-LRR gene, is also required (Salmeron et al. 1996). *Prf* is located in the same cluster with *Pto* and *Fen*, the Fenthion sensitivity gene. Autophosphorylation of two amino acid residues, thr38 and ser198 is necessary for physical contact between *AvrPto* and for elicitation of defense related responses (Sessa et al.

2000). Signal from initial contact is transduced to downstream mediators, such as *Pti-1*, *Pti-4,* and *Pti-6* by phosphorylation (Bogdanove and Martin 2000; Bryan et al. 2000; Gu et al. 2002; Zhou et al. 1995)

d) Class IV

Genes in this group are receptor-like protein kinases, which contain extracytoplasmic LRR repeats, followed by a transmembrane (TM) and intracellular serine/threonine kinase domains. *Xa2*1, providing race-specific resistance against rice pathogen *Xanthomonas oryzae pv. oryzae race 6*, is the first gene to be cloned from this group (Song et al. 1995). The presence of both putative extra-and-intracellular domains suggested that specific recognition of the pathogen could occur outside the cell and signal could be transduced through the intracellular kinase domain. Autophosphorylation of a candidate ser/thr kinase domain in maltose-binding and glutathionine S-transferase fusion proteins further affirmed the possibility of cross membrane signal transduction (Liu et al. 2002). *Xa21* is part of a multigene family consisting of seven other paralogs in the same stretch of DNA on the rice genome; five of these paralogs are truncated. Only one encodes a full receptor-like kinase (Song et al. 1997). *Xa21* exemplifies an ongoing evolutionary process in resistance gene clusters, with large duplicated sequences possibly due to intergenic illegitimate recombination and truncated open-reading frames caused by retrotransposon insertions (Song et al. 1997). In fact, one of these interrupted paralogs, *Xa21D*, functions in a similar manner to *Xa21*, but only renders partial resistance, demonstrating creation of possible novel function (Wang et al. 1998). Inter- and intragenic unequal crossing-over, point mutations, and retrotransposon insertions as in other resistance gene clusters are likely to be main factors shaping the evolutionary fate of this group of genes.

e) CLASS V

Genes in this category do not encode receptor-like proteins. The *Hm1* gene in maize, which encodes for HC toxin reductase (HCRT), confers resistance against the fungus *Cochliobolus carbonum* race 1 in a race-specific manner (Johal and Briggs 1992). The *Mlo* gene of barley is a membrane anchored protein with at least six membrane spanning helices; it provides broad spectrum resistance to the fungal pathogen *Erysiphe graminis f. sp. Hordei* (Buschges et al. 1997). Genes in this group do not fit the classical 'sentinel' role of pathogen perception like most other resistance genes. In contrast, they encode defense proteins that are directly effective against pathogens.

C. **Plant Nematode Resistance Genes:**

The root-knot nematodes, *Meloidogyne spp*. are one of the most important plant pathogens affecting a wide-range of crop plants; more than 2000 plant species worldwide have been reported (Sasser 1980). This is likely to be an underestimate, since many yield losses caused by nematodes are often attributed to other factors such as drought. The cultivated peanut does not have significant genetic resistance to this devastating pathogen. Holbrook et al. (1992) could find only moderate levels of resistance after screening more than 1000 *A. hypogaea* landraces. A recent similar study on Florunner breeding lines, UF81206-Q4, UF81206, and UF93111, has also shown only a moderate level of resistance (Holbrook et al. 1998). On the other hand, some diploid wild relatives of peanut have a high level of resistance to *Meloidogyne* spp. including *A. batizocoi* (Krapov. and W.C. Gregory), *A. cardenasii* (Krapov. and W.C. Gregory), and *A. diogoi (Hehne)*. One resistance gene to *Meloidogyne arenaria* has been successfully introduced by backcrossing of a recipient parent, *A. hypogaea* to a synthetic tetraploid TxAG6 (*A. batizocoi*× (*A. cardenasii*× *A. diogoi*-Simpson et al. 1991). This gene mapped on linkage group 1 of peanut

between the RFLP markers R2430E and R2545E (Burow et al. 1996). These two markers were successfully utilized for identification of individuals homozygous for nematode resistance in BC_7F_2 populations (Church et al. 2000). Later analyses in advanced backcross breeding populations showed that a single gene was responsible for the resistance in some BC5F2 populations, TP259-3, TP262-3, and TP271-2. On the other hand, two genes were likely to confer resistance in other populations, TP259-2, TP263-2, and TP268-3 (Choi et al. 1999). This study has shown that additional nematode resistance genes may have been transferred from wild relatives of peanut, which could be exploited in the future. Nematode resistant peanut breeding lines "COAN" and "NemaTAM" from BC_5 and BC_7 respectively of the same population have been introduced to market, one of the first cultivars developed by marker-assisted selection (Simpson et al. 2001b, 2003). Another two dominant root-knot nematode resistance genes, *Mae* and *Mag*, which act by restricting egg numbers and galling, have been identified in a segregating population derived from the cross of *Arachis hypogaea* ×*Arachis cardenasii* (Garcia et al. 1996). The exact locations of these genes are not clear yet. Even though genetic resistance against nematodes in cultivated peanut is very scarce, its wild relatives provide a potential genetic resource. However, in order to speed up the process of introgression, it is important to analyze the physical location of single and clustered resistance genes in peanut. Because of often-tight association between resistance-gene like sequences throughout genomes, these islands might be used for tracking the potential resistance loci in segregating populations.

Many plant nematode resistance genes have been mapped and several have been cloned (Davis et al. 2000; Williamson 1999). The root-knot nematode resistance gene *Mi* from tomato is homologous to the widely present NBS-LRR resistance gene class (Milligan et al. 1998). A nematode resistance gene, *Hero*, conferring broad spectrum resistance to potato cyst nematode

G. rostochiensis and partial resistance against *G. pallida* shares the same structural features as well (Ernst et al. 2002). Other cyst nematode resistance genes such as *Gpa2* from potato and *Cre3* from wheat share the similar motifs (Lagudah et al. 1997; Van der Vossen et al. 2000). However, one nematode resistance gene *Hs1pro-1* has very distinctive structure with LRR like repeats at N-termini, and little in common with *Cf* gene LRRs, followed by a transmembrane domain (Cai et al. 1997). In any amplified RGA pool, a certain proportion could be from genes functioning in nematode resistance.

Although race-specific resistance has generally received more attention, some nematode resistance characters are polygenic. Mapped QTLs conditioning resistance against nematodes is helpful for plant breeders in creation of durable resistance (Young 1996). Especially in two crop plants, soybean and potato, numerous QTLs conferring nematode resistance have been mapped and characterized. For example, Tamulonis et al. (1997a; 1997b) mapped four QTLs, in soybean, *Glycine max*; two of which condition resistance to root-knot nematode *M. arenaria* and the other two providing resistance against *M. incognita*. Many QTLs conditioning field resistance in soybean to cyst nematode, (*Heterodera glycines* I.) have been localized and characterized (Concibido et al. 2004). Likewise, introgressed segments of hybrids between cultivated potatoes and wild diploids, which contain QTLs against the cyst nematode *Globodera pallida*, have been analyzed (Bryan et al. 2002). *Mae* and *Mag* are two dominant complex trait loci in peanut, which confer resistance to *M. arenaria* by restricting egg numbers and galling respectively (Garcia et al. 1996). Similarly, two genes were responsible for conditioning resistance in some BC5 populations formed by crosses between *A. hypogaea* and a synthetic tetraploid TxAG6 (Choi et al. 1999). Besides conventional race specific resistance genes, QTLs are also likely to be important for breeding nematode resistant cultivars. It should be also elucidated that in some

cases resistance QTLs are often closely associated with resistance gene clusters. Therefore, physical characterization of resistance gene rich islands in the peanut genome would also facilitate the identification of possible QTLs conferring resistance to nematodes and a wide range of other pathogens.

The mode of action of nematode resistance genes is not understood as well as bacterial or viral resistance genes. The *Mi-1.2* gene triggers a hypersensitive response in tomato as a result of interaction with nematode. Conversely, the resistance mechanism of *Ma1*, a dominant peanut root-knot nematode resistance gene, does not involve necrotic or hypersensitive response (Starr et al. 1990). Furthermore, Bendezu et al. (2003) speculated that resistance in COAN, a cultivar containing *Ma1*, could be due to constitutive factors in roots. Also, another nematode resistance gene *Hs1pro-1* conferring resistance to sugar beet cyst-nematode does not trigger necrosis (Cai et al. 1997). Later studies on the same gene have suggested the direct regulation of the expression of *Hs1pro-1* in sugar beet and *Arabidopsis* roots by nematode induced factors at feeding sites (Thurau et al. 2003). Like initial recognition specificity, downstream mediators and induced-defense-related genes in plant-nematode interactions have not been well described.

The root-knot nematode resistance gene *Mi-1.2* is the most extensively studied nematode resistance gene. *Mi-1.2* shares strong homology with *prf*, a resistance gene that mediates *pto* resistance to *Pseudomonas syringae* (Milligan et al. 1998). *Mi-1.2* conditions resistance not only against root-knot nematodes, but also against potato aphid, *Macrosiphum euphorbiae* and white fly, *Bemisia tabaci* (Nombela et al. 2003; Rossi et al. 1998; Vos et al. 1998).This multifunctional aspect of the *Mi-1.2* gene may engender the presence of possible common virulence mechanisms of these pathogens. A comparative study of the interactions between *Mi-1.2* and these pathogens could assist in shedding light on the complex issue of nematode-plant recognition specificity.

The structural aspect of recognition-specificity has been studied at length. A domain swapping experiment between *Mi.1-2* and its nonfunctional homolog *Mi.1-1* supported a possible role of N-terminal LRR repeats in conditioning resistance. However, the LRR domain of the functional homolog alone was not sufficient to complement the nonfunctional homolog; in fact, the resulting phenotype was lethal (Hwang et al. 2000). This experiment confirms possible involvement of other domains such as the C-terminal domain (Hwang and Williamson 2003). Comprehensive analysis of nematode-plant interactions at the molecular level may require cloning and characterization of genes from other plant families.

Many concepts about pathotype specific resistance in nematode-plant interactions, unlike bacterial, viral, and fungal, are still vague and need to be clarified. First of all, a bacterial avr equivalent has never been cloned or characterized from any nematode species. The question of how interaction specificity is established remains a mystery (Davis et al. 2000). Additionally, nematode parasitism genes have been another area of research, but putative roles of only a few genes such as endoglucanases have been described. Another major subject in nematode-plant interactions has been study of establishment and maintenance of highly specialized nematode feeding sites (Gheysen and Fenoll 2002). The downstream components of plant defense systems against nematodes remain to be investigated (Williamson 1999; Williamson and Gleason 2003). A study conducted by de Ilarduya et al. (2003) analyzes the temporal and spatial induction of possible defense-related transcripts as a result of both compatible and incompatible interactions after infection with potato and peach green leaf aphids. It will be quite interesting to see how much the *Mi*-mediated responses to aphids and nematodes overlap and how the plant distinguishes these two different pathogens.

The wheat *Cre* locus consists of multiple copies of related NBS-LRR genes providing resistance to the root endoparasitic nematode *Heterodera avenae*. One of the homologs, *Cre3* in this locus has been cloned and characterized (Lagudah et al. 1997). Several genes, *Cre1*, *Cre6*, and *CreF*, conferring varying degrees of resistance to different pathotypes of the same nematode have been cloned (Ogbonnaya et al. 2001). The presence of multiple homologs of *Cre3* gene presented an opportunity for amplification of other genes from the same family. Wheat *Cre3* sequence was used as a template for designing primers to amplify homologous genes from the wheat genome (Seah et al. 1998).

Nematode resistance genes are also clustered with other pathogen resistance genes, for example, *Mi* maps about 1 cM from fungal resistance genes *Cf4* and *Cf5* in tomato (Dickinson et al. 1993; Dixon et al. 1996a). Similarly, a region delimited by RFLP markers K644 and B212 on soybean linkage group F contains genes conferring resistance to bacterial blight, *Phytophthora* root rot, and various viral diseases, plus QTLs conditioning resistance to root knot nematode and corn earworm (Jeong et al. 2001). These observations may suggest that the genes conferring resistance to nematodes and other pathogens result from the same evolutionary process. If so, it may suggest the formation of gene clusters around the root-knot nematode resistance gene in peanut. Therefore, physical characterization of resistance gene rich regions of the peanut genome could help to limit the regions containing possible quantitative or qualitative nematode resistance gene loci. Furthermore, the identification of resistance gene clusters as focal points in the large genome of peanut would simplify dealing with the complexity and size of the genome. Contigs of large insert DNA clones around the root-knot nematode resistance gene or other resistance genes would be a good resource for fine-scale mapping and cloning.

D. Resistance Gene Analogs (RGA)

Resistance genes have very little sequence similarity outside a few well conserved domains such as the Nucleotide Binding Site (NBS) that is the most commonly shared among all classes of R genes. The commonality of functionally significant subdomains such as the P-loop, kinase-2, kinase-3a, GLPL, and RNBSD in the NBS region of cloned resistance genes has inspired the design of degenerate primers for amplification of these families of genes from other plant genomes. In particular, P-loop motif GXGXXGXV and downstream hydrophobic region GLPL are extensively utilized as anchor points for degenerate primers (Leister et al. 1996). Large numbers of RGAs from soybean have been amplified with degenerate primers developed from NBS (Kanazin et al. 1996; Yu et al. 1996). Similar primers have been used for RGA amplification from many other plants such as maize (Collins et al. 1998), wheat and barley (Lagudah et al. 1998), tomato (Ohmori et al. 1998), lettuce (Shen et al. 1998), Arabidopsis (Speulman et al. 1998), rice (Ilag et al. 2000), *Brassica* spp. (Fourmann et al. 2001), coffee (Noir et al. 2001), chickpea (Huettel et al. 2002), *Arachis* spp. (Bertioli et al. 2003), and cassava (Lopez et al. 2003). This approach has proved to be very useful for analysis of resistance gene analogs even from genomes that have not been particularly well studied such as cassava and coffee.

a) Targeted Amplification of the NonTIR Group

Although conserved sequence blocks in NBS domain are usually common to both TIR and nonTIR subgroups of NBS-LRR genes, some blocks contain group-specific motifs. For example, a 5 amino acid long motif in the RNBSA domain of two classes is clearly distinct between the two groups, while the FXXXF is common in CLN, FXXXXW is common in the TNL group. Likewise, the kinase 2 domain contains a distinctive pattern too; the common LDDV pattern is

usually followed by an aspartate in CNL, but the same residue is usually substituted by a W in the other class. Similar distinctive patterns in downstream conserved blocks such as RNBSC and RNBSD are also observed (Cannon et al. 2002; Pan et al. 2000b). These distinguishing amino acids in conserved blocks of the NBS region allowed the design of degenerate primers for targeted amplification of particular groups. Penuela et al. (2002) were able to isolate 50 RGAs with nonTIR specific primers from the soybean genome, which were homologous to cloned nonTIR genes. The targeted isolation of the nonTIR class in dicots is significant, since nonspecific primers are more likely to amplify the TIR class because of its prevalence. For instance, in *Arabidopsis* more than two thirds of NBS-LRR genes belong to the TNL class.

b) Amplification of Receptor-Like Ser/Thr Kinases

All receptor-like kinase genes such as *Xa21*, *Pto*, *FLS2*, and *SRF2* are homologous in their kinase domains. This homology allows the design of degenerate primers to amplify similar genes in other genomes. However, it should be acknowledged that a significant proportion of amplicons from such experiments, may not be functionally related to resistance genes due to the existence of large numbers of unrelated genes that share similar domains. Nevertheless, Vallad et al. (2001) were able to amplify five different families of Ser/Thr Kinase (STK) like genes from common bean. Similarly, 26 kinase analogs have been amplified from the wheat genome (Maleki et al. 2003), and 53 amplicons have been generated from the grape genome that resemble both cytoplasmic kinases such as *Pto* and STKs with extracellular LRR such as *Xa21* (Di Gaspero and Cipriani 2003). Therefore, kinase domains of cloned genes from this family could be exploited as an anchor point for isolation of similar genes from other plant genomes.

c) Amplification of Cf Like Genes with Primers Designed From the LRR Region

The Cf gene family confers resistance to the fungal pathogen *Cladosporium fulvum* in tomato. Main structural components of genes in this family are an extracellular domain mainly composed of varying numbers of LRR repeats, followed by a transmembrane domain (TM). The presence of sufficient similarity between homologs, such as *Cf2*, *Cf5*, and *Cf9* allowed the design of degenerate primers from the LRR region to amplify similar genes. The LRR domains of *Cf2* and *Cf9* genes have been used for amplification of similar genes in nearly isogenic tomato lines (Ohmori et al. 1998). However, it should be noted that the hypervariable nature of the LRR domains complicates the amplification process.

d) Amplification of RGAs From *Arachis* spp.

The cultivated peanut, *Arachis hypogaea*, has likely been formed by a recent single hybridization of A and B genome ancestral diploids. Because of this recent polyploidization, genetic diversity among all peanut cultivars is scarce, making peanut more vulnerable to a variety of pathogens. The most comprehensive characterization of disease resistance gene like sequences in *Arachis spp*. to date has been conducted by Bertioli et al. (2003). They generated 78 NBS sequences from *A. duranensis*, *A. cardenasii*, *A. stenosperma*, and *A. simpsonii*. The number of NonTIR sequences was much less than TIR sequences. About 12 NBS region sequences were amplified from *A. hypogaea* with primers based on homology to root-knot nematode resistance gene *Mi* (Dumbala and Sivaramakrishnan, unpublished). These studies are very unlikely to represent the total resistance gene family diversity in the *Arachis* genome. For instance, different sets of primers in soybean produced distinct groups of resistance gene-like sequences (Kanazin et al. 1996; Penuela et al. 2002; Yu et al. 1996). Therefore, maximal

representation of resistance gene family diversity in *Arachis* may require use of different sets of degenerate primers.

e) Alternative Approaches For Analysis of Resistance Gene Like Sequences

Other methods have been also used for analysis of resistance gene families, albeit not as often. Expressed resistance genes have been isolated and characterized in soybean by screening a cDNA library with R gene specific probes (Graham et al. 2000). He et al. (2003) was able identify a putative resistance gene (*KR1*) from soybean by employing a similar strategy. In wheat, 184 putative expressed resistance genes have been isolated with a modified RNA-fingerprinting method; primer sets were designed from conserved R gene domains and PolyT plus selective bases (Dilbirligi et al. 2004). However, it should be noted that the expression of most resistance genes is very brief, therefore, it is quite possible that most resistance genes would not be represented in a RNA population.

f) Genomic Distribution of RGAs in Plant Genomes

As it has been discussed above, resistance genes are usually located in tamdem gene compexes. Likewise, most RGAs have either mapped to the vicinity of known R genes or in very close proximity to each other. This feature of RGAs makes them very useful for pinpointing resistance genes in map-based cloning approaches. For instance, three NBS-LRR homologs were closely linked to an *Anthracnose* resistance gene in *Phaseolus vulgaris* (Creusot et al. 1999). Similarly, RGAs were located on the same linkage groups with known resistance genes in *Arabidopsis*, melon, grape, and wheat (Aarts et al. 1998a; Brotman et al. 2002; Donald et al. 2002; Yan et al. 2003). NBS-LRR gene homologs played a significant role in cloning and characterization of wheat rust resistance gene paralogs (Feuillet et al. 1997; Feuillet et al. 2003). Finally, RGAs have been cleverly employed in demarcating the barley *Mla* gene family to a 240

kb BAC contig near telomeric region of barley chromosome 5S (Wei et al. 1999). Some QTLs for resistance against a variety of pathogens have been also closely associated with RGAs; for instance, an R-gene homolog very tightly linked to a cucumber mosaic virus resistance QTL in the pepper genome (Pflieger et al. 1999). QTLs defining resistance to corn earworm, root knot nematode, and white mold mapped to the same linkage group with some NBS-LRR homologs in soybean (Jeong et al. 2001). QTLs conditioning resistance to *C. lindemuthianum* mapped in the same location in the common bean genome as some expressed NBS-LRR homologs (Ferrier-Cana et al. 2003) Overall, RGAs could be efficiently used for focusing on R gene islands and also for studying the comparative evolutionary history of R gene families.

RGAs have been frequently used for either identification of possible candidates from BAC contigs containing R-gene loci, or during the formation of these contigs by screening BAC libraries as a initial step for cloning. For example, Wang et al. (1996) used RGAs for identification of BACs containing possible resistance genes. RGAs have been also employed for identification of BACs associated with citrus tristeza virus resistance and other resistance genes in citrus (Deng et al. 2001a; Deng et al. 2001b). Possible candidates in a BAC contig spanning the *Lr1* region of wheat have been identified with RGAs (Ling et al. 2003). In conclusion, RGAs could be beneficial for both contig formation and identification of possible candidates.

III. INTEGRATION OF MAPPED PEANUT PROBES INTO A PHYSICAL MAP AND APPLICATION OF *ARABIDOPSIS* DERIVED OVERGOS ONTO A PEANUT BAC LIBRARY FOR COMPARATIVE ANALYSIS

A. Peanut

a) Introduction:

The bunch and runner types of domesticated peanut are among the most widely-grown crops both in the USA and worldwide. According to FAO estimates, peanut production worldwide stood at 37,057,652 (Mt) in 2003 (FAOSTAT data, 2004). Peanut was originally domesticated by South American natives; after arrival of Europeans in the New World, peanuts were disseminated and became widely cultivated in temperate regions of world. Peanut seeds, with 25% protein and high monounsaturated fat content are an essential part of the daily diet in many parts of the world. Peanut seeds are also important for industrial oil production. Despite its economic and agronomic significance, genomic studies in peanut lag behind those in other crops. This study will advance peanut genomics with potential practical applications in peanut agriculture.

b) Phylogeny of the Genus *Arachis*

The cultivated peanut, *Arachis hypogaea L*., belongs to the genus *Arachis*, which contains about 70 species, most of which are either unnamed or poorly described. These species are usually grouped into 8 to 9 different sections, *Caulorrhizae, Arachis, Erectoides, Extranervosae, Heteranthae, Procumbentes, Rhizomatosae, Trierectoides, and Triseminatae*, on the basis of morphological and cross-compatibility differences (Fernandez and Krapovickas 1994; Resslar 1980). Section *Arachis* has been the focal point of research, because it contains cultivated peanut, *Arachis hypogaea* and its possible diploid progenitors. The 30 species in *Arachis* are

discerned from other sections in terms of their tap roots, lack of adventitious roots or rhizomes, and vertical pegs, which do not grow horizontally after soil penetration (Resslar 1980). Most of these species are geographically distributed in an area encompassing parts of four different countries; Paraguay, southern Brazil, eastern Bolivia, and northern Argentina (Jarvis et al. 2003). Unfortunately, *Arachis spp*. habitat is dwindling with an imminent risk of losing *in situ* stocks of valuable members of the secondary and tertiary gene pool. This fact calls for a more complete picture of the peanut genome to facilitate peanut breeding and classification.

Almost all species in section *Arachis* of genus *Arachis* are diploids (2n=20) with only two exceptions, *Arachis hypogaea L.*, and *Arachis monticola*, which are allotetraploids (2n=4x=40) (Stalker and Dalmacio 1986; Stebbins 1957). Most diploid species are classified as having A or B genomes. Eleven species are recognized as A genomes; some of the well defined ones are *A. cardenasii*, *A. duranensis*, *A. stenosperma*, and *A. villosa* (Lavia 1996). Initially only *A. batizocoi* was identified as B genome, but later more diploids were included in this genomic class such as *A. ipaensis*, and *A. magna* (Gimenes et al. 2000; Krapovickas and Gregory 1994). The only known exception to these genomic categories is *A. glandulifera*, classified as D genome (Stalker 1991). It needs to be noted that most of these genomic assignments have been done on the basis of morphological and cytogenetic data; thus, it is possible that the level of genetic diversity could vary both within and among groups.

Arachis hypogaea has been divided into two subspecies based on criteria such as plant growth habit (e.g. erect or spreading), fruit traits such as number of seeds per pod, and several other distinguishing morphological features (Gregory et al. 1980). Even though these two subspecies, *A. hypogaea spp. hypogaea* and *A. hypogaea spp. fastigiata* are quite distinctive in their appearances, RFLP and other molecular markers have shown that landraces and cultivars of

these subspecies are genetically very similar (i.e. very little polymorphism). Further, a low level of polymorphism has been also observed between *A. hypogaea* and its weedy relative *A. monticola* (Kochert et al. 1991; Kochert et al. 1996). This paucity of genetic polymorphism led to exploration of diploid relatives as genetic resources, making introgression of desirable traits much more laborious.

c) Progenitors of *Arachis hypogaea*

The time of the polyploidization event that generated allotetraploid *Arachis hypogaea* is not clear, but its low level of genetic diversity may suggest either a recent single amphidiploidization event, or a possible recent severe genetic bottleneck. The only other tetraploid in section *Arachis*, *A. monticola*, has been suggested as a direct ancestor or a close wild-relative of *A. hypogaea* (Kirti et al. 1982; Raina and Mukai 1999; Simpson et al. 2001a). Even though the nature of the phylogenetic relationship between these two species is still unclear, there is more evidence disproving *A. monticola* as a direct wild precursor of *A. hypogaea* (Jung et al. 2003). It is clear that *A. monticola* is cross-compatible with the domesticated peanut and very little genetic polymorphism exists between the two species. Thus, *A. monticola* does not have high value as a source of genetic variation.

The identities of the A and B genome progenitors of peanut are still uncertain. The majority of data about this subject comes from three sources: a) molecular data such as markers and gene sequences; b) cytological data (e.g. genomic *in situ* hybridization); and c) morphological data. Since the first known B genome diploid was *A. batizocoi*, it was propounded as a presumed ancestor (Singh and Smartt 1998). However, recent evidence suggests otherwise. RFLP data discounted *A. batizocoi* as a progenitor, instead suggesting *A. ipaensis* as B and *A. duranensis* as A genome donors (Kochert et al. 1991; Kochert et al. 1996). Fluorescently labeled *A. villosa*

genomic DNA mixed with unlabelled *A. ipaensis* genomic DNA uniformly probed only 20 chromosomes in tetraploid species, hence, suggesting them as ancestral donors (Raina and Mukai 1999). Assessment of genetic variation among *Arachis spp.* with RAPD and ISSR markers further supported the same candidate donors (Raina et al. 2001). Lastly, comparison of stearoyl-ACP desaturase and oleoyl-PC desaturase gene sequences denoted that sequences from *A. hypogaea* shared higher homology with diploid species, *A. ipaensis* and *A. duranensis* than *A. monticola*, therefore, supporting the same diploids as genome donors and discounting *A. monticola* as progenitor (Jung et al. 2003). Accumulation of more genomic data from *Arachis spp.* will eventually settle this debate.

There have been conflicting reports about the genome sizes of *Arachis* species. For instance, *Arachis hypogaea* genome size is variously reported to be 1.78 pg/C (Bennett and Smith 1976), 2.91 pg/C (Arumuganathan and Earle 1991), 5.13 pg/C (Singh et al. 1996), and 2.95 pg/C (Temsch and Greihulber 2000). These contradictory reports are likely to be a result of using different techniques, and also the possible influence of secondary plant metabolites such as flavanols on Feulgen densitometry readings. Similar paradoxical results have been also reported about the flexibility of nuclear DNA content in *Arachis* species. Some speculations were made about diminishing nuclear DNA content of *A. duranensis* with altitude and latitude (Temsch and Greihulber 2001). However, one needs to be cautious about any postulations by relying on these genome size reports. On the other hand, genome sizes of putative diploid progenitors, *A. villosa*, *A. duranensis*, and *A. ipaensis*, have been usually considered to be as large as amphidiploids or much larger (Singh et al. 1996). On the basis of this fact, it may be concluded that either loss of large amounts of nuclear DNA has occurred in allopolyploids or there has been an expansion in genome sizes of diploids after amphidiploidization.

d) Peanut Genetic Maps

Attempts to form genetic maps from peanut amphidiploid species were hindered by the lack of sufficient polymorphism both at interspecific and intraspecific levels. In contrast, an abundance of polymorphic sites have been detected among wild diploids. The first genetic map from *Arachis* spp. was constructed by crossing two wild diploids (*A. stenosperma* and *A. duranensis*). The map is composed of 117 RFLP markers on 11 linkage groups with a total length of 1603 cm. Several methods have been tried to overcome the genetic diversity problem for creation of a tetraploid map. One of these was creation of diploid-tetraploid hybrid lines; for which a diploid was crossed with *A. hypogaea* resulting in an unstable triploid, followed by doubling chromosome number to create a hexaploid, with subsequent reduction to tetraploidy either by selfing or backcrossing with a tetraploid. Garcia et al., (1995) created 46 introgression lines by crossing *A. hypogaea L.* with *A. cardenasii Krapov. & W.C. Gregory*. They detected introgressed regions on 10 of 11 linkage groups formed by mapping 34 RFLP and 45 RAPD probes. Successful introduction of desired traits from diploid to tetraploid species has also been accomplished through initial creation of a synthetic tetraploid from diploids, and subsequent crossing of it with the cultivated tetraploid. One such synthetic amphidiploid, TxAG-6, was formed by crossing [*A. batizocoi* K9484× (*A. cardenasii* GKP10017× *A. diogoi* GKP10602)]$^{4\times}$ (Simpson et al. 1993; Simpson et al. 1991). TxAG-6 was used as the donor parent in formation of a BC_1 mapping population with 78 progeny. The map contains 370 RFLP loci on 23 linkage groups, extending over 2210 cM (Burow et al. 2001). Even though this map has been very beneficial for shedding light on peanut genome organization and practical applications, the marker density on the map is not yet sufficient for future attempts for anchoring of the physical map to the genetic map.

B. Physical Maps

a) Large Insert Clone Libraries:

Construction of large insert clone libraries is a primary step for any physical mapping project. Yeast artificial chromosomes were the first vectors, which allowed cloning of DNA fragments up to 1 Mb in size. However, due to problems such as recombination, instability, and development of better cloning systems, [specifically, Bacterial Artificial Chromosomes; BACs], YAC's are now rarely utilized for physical mapping. The quality of BAC libraries, in terms of insert size and lack of chimerism, is vital for a successful mapping project. For many important crops and model plants such as soybean and *Medicago*, several BAC libraries have been constructed and these libraries employed either in large scale physical analysis or in targeted analysis of particular regions of genomes (Danesh et al. 1998; Marek and Shoemaker 1997; Nam et al. 1999; Salimath and Bhattacharyya 1999; Tomkins et al. 1999). However, a BAC or YAC library had not previously been generated for any *Arachis spp*. The BAC library, constructed in our lab, with 182,784 clones and 104 kb average insert size is a valuable genomic resource for peanut (Chapter 2). The characteristics of the library are discussed at length in the chapter.

b) Screening Techniques:

Two main approaches have been commonly used for screening large-insert arrayed libraries. One approach is based on pooling of DNA from multiple clones, and subsequent screening of these pools with site-specific primers. For instance, Green et al. (1990) were able to pull out the specific YACs from a pool with DNA from 1920 different clones. New pooling strategies have been developed to maximize experimental efficiency and for large scale applications (Bruno et al. 1995). Semi-arbitrary PCR primers accommodate some general applications such as overall physical mapping. For instance, AP-PCR (Arbitrary Primer-PCR)

has been used in identification of 245 BAC clones from a rice BAC library (Xu et al. 1998). Likewise, in a larger scale experiment, Klein et al. (2000) were able to pick out ~2400 BACs from BAC pools of a sorghum library by using AFLP primers. Another approach is the application of radioactively labeled probes onto high-density BAC filters. The availability of robots that can array large numbers of clones onto membranes in a short time reduced the need for the pooling approaches for large scale mapping projects. High-density BAC filters allow screening of large numbers of BAC clones for either multiple or single probes. Oligonucleotide based probes called "overgos", improved screening due to efficient radiolabeling and low cost. Overgos are two 24 bp oligonucleotides with 8 bp overlap at the 3′ end; thus, allowing synthesis of complementary strands with radioactive nucleotides (Ross et al. 1999). Multiplexing of overgos enabled the hybridization of large number of probes in a single experiment. Cai et al. (1998) showed feasibility of a multiplexing strategy in two dimensions by applying 320 overgos to a mouse BAC library. With a similar experimental setting, Han et al. (2000) were able to identify 5,187 positive BAC clones with 230 overgo probes from a human library. Finally, 10,642 overgos designed from ESTs were applied to 165, 888 maize BACs in a 24×24 multiplex with 88 % success rate (Gardiner et al. 2004). The multiplexing and the practicality of overgos have made it the method of preference for screening large-insert arrayed libraries

c) Plant Physical Maps:

Hybridization data alone is not sufficient for the formation of a complete physical map; a combination of hybridization data, BAC fingerprinting data, and BAC-end sequence data is required to minimize gaps between contigs. Fingerprinting is commonly based on digestion of BAC clones with a six-cutter restriction enzyme and subsequent resolution of the digestion products (Marra et al. 1997). Afterwards, gel images are analyzed (Fuhrmann et al. 2003; Sulston

et al. 1989). Finally contigs are formed on the basis of shared banding patterns with software such as FPC V4.7 (Soderlund et al. 2000). The additional information may contribute significantly to closure of gaps between contigs. Hence, integration of several data types is a must for both generation of robust contigs and anchoring them to a genetic map.

A physical map of the human genome provided a template for choosing clones to sequence, and guided accurate assembly of finished sequences (The International Human Genome Mapping Consortium 2001). Physical maps with varying coverage have been constructed for a few plant species. *Arabidopsis* is the first flowering plant for which the whole genome has been sequenced and characterized (The Arabidopsis Genome Initiative 2000). A clone-by-clone strategy has been employed in sequencing *Arabidopsis* (Mozo et al. 1999). Several rice genome sequences from two different subspecies, *Oryza sativa L. spp. japonica* (Goff et al. 2002) and *Oryza sativa L. spp. indica* (Yu et al. 2002) have been produced with random-fragment shotgun sequencing. In contrast, a public rice genome sequencing consortium has employed a clone-by-clone approach. Although clone-by-clone sequencing is more laborious compared to the shotgun approach, the higher rate of assembly accuracy and modularity of the assembled sequence may justify this approach. Besides, several rice physical maps have been developed with more than 90 % coverage rates (Chen et al. 2002; Tao et al. 2001). In addition, shotgun sequencing may be less feasible for complex plant genomes, which contain large numbers of recently amplified highly repetitive sequences. Hence, physical maps for large plant genomes such as maize, peanut, and wheat, may provide templates for targeting regions of interest. Besides rice and *Arabidopsis*, physical maps have been constructed for a few other plants too. Wu et al. (2004) generated 2905 BAC/BIBAC contigs, covering 1408 Mb of the soybean genome. For several other plants, preliminary work on physical maps has been accomplished, including sorghum

(Draye et al. 2001), maize (Gardiner et al. 2004; Yim et al. 2002), and two legumes, *Lotus japonicus*, and *Medicago truncatula* (VandenBosch and Stacey 2003; Young et al. 2003). These preliminary studies have shown the feasibility of formation of physical maps with sufficient coverage to be used for multiple purposes such as map-based cloning. However, no preliminary studies have been conducted regarding feasibility and applicability of physical mapping for any *Arachis* species. Exploratory work on this subject would open new avenues for peanut genomics in terms of both comparative evolutionary analysis, and practicality of map-based cloning methods.

Since resistance gene families are usually organized in clusters, it is possible to tag the location of these clusters in the genome by screening BAC libraries. The phyletic origin of sequences is often correlated with their physical proximity in the genome. In other words, resistance gene islands are often monophyletic. Therefore, identifying resistance gene rich regions of a genome would create focal points for the detailed analysis of genes in particular families.

There are at least two ways that resistance genes could be physically located onto genomes; RGA sequences can be used as probes for screening a BAC (or other large-insert) library, or 40 bp long overgos (as described above) could be designed from sequences and applied to the library.

Three hundred and twenty-two BAC clones have been isolated from a 7× citrus BAC library with the application of 13 RGA as probes (Deng et al. 2001b). In the same study, one to four fragments in each BAC hybridized to single probes showing possible multiple copy presence (i.e., clusters). In a similar study, the same group was able to characterize the distribution and organization of RGAs from kinase domains of *pto* and *Xa21* by applying them

as probes to a citrus library (Deng and Gmitter 2003). Likewise, 42 BAC clones were isolated from a cassava library by screening with RGA probes (Lopez et al. 2003). These studies have proven the usefulness of RGAs as a tool in identification and characterization of resistance gene rich regions of genomes.

REFERENCES:

Aarts MG, Hekkert A, Baste L, Holub EB, Beynon JL, Stiekema WJ, Pereira A (1998a) Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Mol Plant Microbe Inter 11: 251-258

Aarts N, Metz M, Holub E, Staskawicz BJ, Daniels MJ, Parker JE (1998b) Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in *Arabidopsis*. Proc. Natl. Acad. Sci 95: 10306-10311

Abdel-Momen SM, Starr JL (1998) *Meloidogyne javanica Rhizoctonia solani* disease complex of peanut. Fund Appl Nemat 21: 611-616

Anderson C (1993) Genome shortcut leads to problems. Science 259: 1684-1687

Anderson PA, Lawrence GJ, Morrish BC, Ayliffe MA, Finnegan EJ, Ellis JG (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell 9: 641-651

Arumuganathan K, Earle D (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208-218

Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res. 12: 1871-1884

Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar SP (1997) Signaling in plant-microbe interactions. Science 276: 726-733

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

Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, Vinatzer BA, Gianfranceschi L, Gessler C, Sansavini S (2004) The *HcrVf2* gene from a wild apple confers scab resistance to a transgenic cultivated variety. Proc. Natl. Acad. Sci 101: 886-890

Bendezu IF, Starr JL (2003) Mechanism of resistance to *Meloidogyne arenaria* in the peanut cultivar COAN. J Nematol 35: 115-118

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philosophical Transactions of the Royal Society, London

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.

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. Science 292: 2281-2285

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

Bogdanove AJ, Martin GB (2000) *AvrPto*-dependent *Pto*-interacting proteins and *AvrPto*-interacting proteins in tomato. Proc. Natl. Acad. Sci 97: 8836-8840

Branch WD, Brenneman TB (1999) Stem rot disease evaluation of mass-selected peanut populations. Crop Protection 18: 127-130

Branch WD, Brenneman TB (2003) Field resistance to cylindrocladium black rot and tomato spotted wilt virus among advanced runner-type peanut breeding lines. Crop Protection 22: 729-734

Brotman Y, Silberstein L, Kovalski I, Perin C, Dogimont C, Pitrat M, Klingler J, Thompson A, Perl-Treves R (2002) Resistance gene homologues in melon are linked to genetic loci conferring disease and pest resistance. Theor App Genet 104: 1055-1063

Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J, Druka A, Steffenson B, Kleinhofs A (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc. Natl. Acad. Sci 99: 9328-9333

Bruno WJ, Knill E, Balding DJ, Bruce DC, Doggett NA, Sawhill WW, Stallings RL, Whittaker CC, Torney DC (1995) Efficient pooling designs for library screening. Genomics 26: 21-30

Bryan GJ, McLean K, Bradshaw JE, De Jong WS, Phillips M, Castelli L, Waugh R (2002) Mapping QTLs for resistance to the cyst nematode *Globodera pallida* derived from the wild potato species *Solanum vernei*. Theor App Genet 105: 68-77

Bryan GT, Wu K-S, Farrall L, Jia Y, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R, Valent B (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 12: 2033-2046

Burke D, Carle G, Olson M (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-812

Burow MD, Simpson CE, Paterson AH, Starr JL (1996) Identification of peanut (*Arachis hypogaea L*) RAPD markers diagnostic of root-knot nematode (*Meloidogyne arenaria* (Neal) Chitwood) resistance. Mol Breeding 2: 369-376

Burow 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

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J (1997) The Barley Mlo Gene: A novel control element of plant pathogen resistance. Cell 88: 695-705

Cai D, Kleine M, Kifle S, Harloff HJ, Sandal NN, Marcker KA, Klein-Lankhorst RM, Salentijn EM, Lange W, Stiekema WJ (1997) Positional cloning of a gene for nematode resistance in sugar beet. Science 275: 832-834

Cai W-W, Reneker J, Chow C-W, Vaishnav M, Bradley A (1998) An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. Genomics 54: 387-397

Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2002) diversity, distribution, and ancient taxonomic relationships within the TIR and Non-TIR NBS-LRR resistance gene subfamilies. J Mol Evol 54: 548-562

Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003) Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. durum). Theor App Genet 107: 931-939

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P,

Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14: 537-545

Chen M, SanMiguel P, de Oliveira AC, Woo S-S, Zhang H, Wing RA, Bennetzen JL (1997) Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc. Natl. Acad. Sci USA 94: 3431-3435

Choi K, Burow MD, Church G, Burow G, Paterson AH, Simpson CE, Starr JL (1999) Genetics and mechanism of resistance to *Meloidogyne arenaria* in peanut germplasm. J Nematol 31: 283-290

Choi S, Creelman RA, Mullet JE, Wing RA (1995) Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis-thaliana*. Plant Mol Biol Rep 13: 124-128

Church GT, Simpson CE, Burow MD, Paterson AH, Starr JL (2000) Use of RFLP markers for identification of individuals homozygous for resistance to *Meloidogyne arenaria* in peanut. Nematology 2: 575-580

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T (1999) Molecular Characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell 11: 1365-1376

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol Plant Microbe Inter 11: 968-978

Concibido VC, Diers BW, Arelli PR (2004) A decade of QTL mapping for cyst nematode resistance in soybean. Crop Sci 44: 1121-1131

Creusot F, Macadre C, Cana EF, Riou C, Geffroy M, Sevignac M, Dron M, Langin T (1999) Cloning and molecular characterization of three members of the NBS-LRR subfamily located in the vicinity of the *Co-2* locus for anthracnose resistance in *Phaseolus vulgaris*. Genome 42: 254-264

Danesh D, Penuela S, Mudge J, Denny RL, Nordstrom H, Martinez JP, Young ND (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor App Genet 96: 196-202

Dangl J, Holub E (1997) La Dolce Vita: A Molecular feast in plant–pathogen interactions. Cell 91: 17-24

Dangl JL, Dietrich RA, Richberg MH (1996) Death don't have no mercy: Cell death programs in plant-microbe interactions. Plant Cell 8: 1793-1807

Davis EL, Hussey RS, Baum TJ, Bakker J, Schots A, Rosso M-N, Abad P (2000) Nematode parasitism genes. Annual Review of Phytopathology 38: 365-396

de Ilarduya OM, Xie QG, Kaloshian I (2003) Aphid-induced defense responses in *Mi-1*-mediated compatible and incompatible tomato interactions. Mol Plant Microbe Inter 16: 699-708

Deng Z, Gmitter FG (2003) Cloning and characterization of receptor kinase class disease resistance gene candidates in citrus. Theor App Genet 108: 53-61

Deng Z, Huang S, Ling P, Yu C, Tao Q, Chen C, Wendell MK, Zhang HB, Gmitter FG (2001a) Fine genetic mapping and BAC contig development for the citrus tristeza virus resistance gene locus in *Poncirus trifoliata* (Raf.). Mol Gen Genet 265: 739-747

Deng Z, Tao Q, Chang YL, Huang S, Ling P, Yu C, Chen C, Gmitter FGJ, Zhang H-B (2001b) Construction of a bacterial artificial chromosome (BAC) library for citrus and

identification of BAC contigs containing resistance gene candidates. Theor App Genet 102: 1177-1184

Di Gaspero G, Cipriani K (2003) Nucleotide binding site/leucine-rich repeats, *Pto*-like and receptor-like kinases related to diseases resistance in grapevine. Mol Gen Genomics 269: 612-623

Dickinson M, Jones DA, Jones JDG (1993) Close linkage between the *Cf-2/Cf-5* and *Mi* resistance loci in tomato. Mol Plant-Micr Inter 6: 341-347

Dilbirligi M, Erayman M, Sandhu D, Sidhu D, Gill KS (2004) Identification of wheat chromosomal regions containing expressed resistance genes. Genetics 166: 461-481

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JDG (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell 10: 1915-1926

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

Donald TM, Pellerone F, Adam-Blondon AF, Bouquet A, Thomas MR, Dry IB (2002) Identification of resistance gene analogs linked to a powdery mildew resistance locus in grape. Theor App Genet 104: 610-618

Draye X, Lin Y-R, Qian X-Y, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren S-x, Wing RA, Paterson AH (2001) Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. Plant Physiol. 125: 1325-1341

Dunford R, Vilageliu L, Moore G (1993) Stabilization of a yeast artificial chromosome containing Plant DNA using recombination-deficient host. Plant Mol Biol 21: 1187-1189

Ellis J, Dodds P, Pryor T (2000) The generation of plant disease resistance gene specificities. Trends in Plant Science 5: 373-379

Ellis JG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. Plant Cell 11: 495-506

Ernst K, Kumar A, Kriseleit D, Kloos D-U, Phillips MS, Ganal MW (2002) The broad-spectrum potato cyst nematode resistance gene (*Hero*) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. Plant J 31: 127-136

Ferguson ME, Burow MD, Schulze SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S (2004) Microsatellite identification and characterization in peanut (*A. hypogaea L.*). Theor App Genet 108: 1064-1070

Fernandez A, Krapovickas A (1994) Cromasomas y evolucion en *Arachis* (Leguminosae). Bonplandia 8: 187-220

Ferrier-Cana E, Deffroy V, Macadre C, Creusot F, Imbert-Bollore P, Sevignac M, Langin T (2003) Characterization of expressed NBS-LRR resistance gene candidates from common bean. Theor App Genet 106: 251-261

Feuillet C, Schachermayr G, Keller B (1997) Molecular cloning of a new receptor-like kinase gene encoded at the *Lr10* disease resistance locus of wheat. Plant J 11: 45-52

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc. Natl. Acad. Sci 100: 15253-15258

Flor HH (1956) The complementary genic systems in flax and flax rust. Adv Genet 8: 29-54

Flor HH (1971) Current status of the gene-for-gene concept. Annual Review of Phytopathology 9: 275-296

Fluhr R (2001) Sentinels of disease. Plant resistance genes. Plant Physiol. 127: 1367-1374

Food and Agriculture Organization (2003) FAOSTAT Agriculture. Available at: http://apps.fao.org/page/collections?subset=agriculture; verified 10 February 2004.

Fourmann M, Charlot F, Froger N, Delourme R, Brunel D (2001) Expression, mapping, and genetic variability of *Brassica napus* disease resistance gene analogues. Genome 44: 1083-1099

Frijters ACJ, Zhang Z, van Damme M, Wang GL, Ronald PC, Michelmore RW (1997) Construction of a bacterial artificial chromosome library containing large *Eco*RI and *Hin*dIII genomic fragments of lettuce. Theor App Genet 94: 390-399

Fuhrmann DR, Krzywinski MI, Chiu R, Saeedi P, Schein JE, Bosdet IE, Chinwalla A, Hillier LW, Waterston RH, McPherson JD, Jones SJM, Marra MA (2003) Software for automated analysis of DNA fingerprinting gels. Genome Res. 13: 940-953

Garcia GM, Stalker HT, Kochert G (1995) Introgression analysis of an interspecific hybrid population in peanuts (*Arachis hypogaea L.*) using RFLP and RAPD markers. Genome 38: 166-176

Garcia GM, Stalker HT, Shroeder E, Kochert G (1996) Identification of RAPD, SCAR, and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* into *Arachis hypogaea*. Genome 39: 836-845

Gardiner J, Schroeder S, Polacco ML, Sanchez-Villeda H, Fang Z, Morgante M, Landewe T, Fengler K, Useche F, Hanafey M, Tingey S, Chou H, Wing R, Soderlund C, Coe EH, Jr. (2004) Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol. 134: 1317-1326

Gentzbittel L, Abbott A, Galaud JP, Georgi L, Fabre F, Liboz T, Alibert G (2002) A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. Mol Gen Genet 266: 979-987

Gheysen G, Fenoll C (2002) Gene expression in nematode feeding sites. Annual Review of Phytopathology 40: 191-219

Gimenes MA, Lopes CR, Galgaro ML, Valls JEM, Kochert G (2000) Genetic variation and phylogenetic relationships based on RAPD analysis in section Caulorrhizae, genus *Arachis* (Leguminosae). Euphytica 116: 187-195

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H (2002) Draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 296: 92-100

Graham MM, Marek LF, Lohnes D, Cregan P, Shoemaker RC (2000) Expression and genome organization of resistance analogs in soybean. Genome 43

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dang JL (1995) Structure of the *Arabidopsis RPM1* Gene enabling dual specific disease resistance. Science 269: 843-846

Grant MR, McDowell JM, Sharpe AG, de Torres Zabala M, Lydiate DJ, Dangl JL (1998) Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. Proc. Natl. Acad. Sci 95: 15843-15848

Green ED, Olson MV (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci 87: 1213-1217

Gregory WC, Krapovickas A, Gregory MP (1980) Advances in legume science. Royal Botanic Gardens, Kew

Grube RC, Radwanski ER, Jahn M (2000) Comparative genetics of disease resistance within the Solanaceae. Genetics 155: 873-887

Gu Y-Q, Wildermuth MC, Chakravarthy S, Loh Y-T, Yang C, He X, Han Y, Martin GB (2002) Tomato transcription factors *Pti4*, *Pti5*, and *Pti6* activate defense responses when expressed in *Arabidopsis*. Plant Cell 14: 817-831

Halward T, Stalker HT, Kochert G (1993) Development of an RFLP linkage map in diploid peanut species. Theor App Genet 87: 379-384

Hamilton GM, Frary A, Lewis C, Tanksley SD (1996) Stable transfer of intact high-molecular-weight DNA into plant chromosomes. Proc. Natl. Acad. Sci 93: 9975-9979

Han CS, Sutherland RD, Jewett PB, Campbell ML, Meincke LJ, Tesmer JG, Mundt MO, Fawcett JJ, Kim U-J, Deaven LL, Doggett NA (2000) Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. Genome Res. 10: 714-721

Hayes AJ, Saghai Maroof MA (2000) Targeted resistance gene mapping in soybean using modified AFLPs. Theor App Genet 100: 1279-1283

He C-Y, Tian A-G, Zhang J-S, Zhang Z-Y, Gai J-Y, Chen S-Y (2003) Isolation and characterization of a full-length resistance gene homolog from soybean. Theor App Genet 106: 786-793

Henk AD, Warren RF, Innes RW (1999) A New Ac-like transposon of *Arabidopsis* is associated with a deletion of the *RPS5* disease resistance gene. Genetics 151: 1581-1589

Holbrook CC, Noe JP (1992) Resistance to peanut root-knot nematode (*Meloidogyne arenaria*) in *Arachis hypogaea*. Peanut Science 19: 35-37

Holbrook CC, Noe JP, Gorbet DW, Stephenson MG (1998) Evaluation of peanut breeding lines with resistance to the peanut root-knot nematode. Crop Sci 38: 260-262

Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. Nature Rev. Genetics 2: 516-527

Hong Z, Ueguchi-Tanaka M, Umemura K, Uozu S, Fujoka S, Takatsuto S, Yoshida S, Ashikari M, Kitano H, Matsuoka M (2003) A rice brassinostreoid-deficient mutant, *ebisu dwarf (d2)*, is caused loss of function of a new member of cytochrome P450. Plant Cell 15: 2900-2910

Hopkins MS, Casa AM, Wang T, Mithcell SE, Dean RE, Kochert GD, Kresovich S (1999) Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Sci. 39: 1243-1247

Hu GS, Hulbert SH (1994) Evidence for the involvement of gene conversion in meiotic instability of the *Rp1* rust resistance genes of maize. Genome 37: 742-746

Huettel B, Santra D, Muehlbauer F, Kahl G (2002) Resistance gene analogs of chickpea (*Cicer arietinum* L.): isolation, genetic mapping and association with a *Fusarium* resistance gene cluster. Theor App Genet 105: 479-490

Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: Evolution and utilization. Annu Rev Phytopathol 39: 295-312

Hwang C-F, Bhakta AV, Truesdell GM, Pudlo WM, Williamson VM (2000) Evidence for a role of the N terminus and leucine-rich repeat region of the *Mi* gene product in regulation of localized cell death. Plant Cell 12: 1319-1330

Hwang C-F, Williamson VM (2003) Leucine-rich repeat-mediated intramolecular interactions in nematode recognition and cell death signalling by the tomato resistance protein *Mi*. Plant J 34: 585-593

Ilag LL, Yadav RC, Huang N, Ronald PC, Ausubel FM (2000) Isolation and characterization of disease resistance gene homologues from rice cultivar IR64. Gene 255: 245-255

Jarvis A, Ferguson ME, Williams DE, Guarino L, Jones PG, Stalker HT, Valls JFM, Pittman RN, Simpson CE, Bramel P (2003) Biogeography of wild *Arachis* assessing conservation status and setting future projects. 43: 1100-1108

Jeong SC, Hayes AJ, Biyashev RM, Maroof MAS (2001) Diversity and evolution of a non-TIR-NBS sequence family that clusters to a chromosomal "hotspot" for disease resistance genes in soybean. Theor App Genet 103: 406-414

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 J. 19: 4004-4014

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

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JDG (1994) Isolation of tomato *Cf9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266: 789-793

Jung S, Tate R, Horn G, Kochert G, Moore K, Abbott AG (2003) The phylogenetic relationship of possible progenitors of the cultivated peanut. J Hered 94: 334-340

Kajava AV (1998) Structural diversity of leucine-rich repeat proteins. JMB Online (Journal of Molecular Biology) 277: 519-527

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

Kim UJ, Birren BW, Slepak T, Mancino V, Boysen G, Kang HL, Simon MI, Shizuya H (1996) Construction and characterization of a human bacterial artificial chromosome library. Genomics 34: 213-218

Kirti P, Murty U, Bharati M, Rao N (1982) Chromosome-pairing in F1-hybrid *Arachis hypogaea* L. X *A. monticola* Krap Et Rig. Theor App Genet 62: 139-144

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A High-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res. 10: 789-807

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

Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Amer. Jour. Bot. 83: 1282-1291

Krapovickas A, Gregory WC (1994) Taxonomía del género *Arachis* (Leguminosae). Bonplandia 8: 1-186

Kulikova O, Gualtieri G, Geurts R, Kim D-J, Cook DR, Huguet T, de Jong H, Fransz PF, Bisseling T (2001) Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. Plant J. 27: 49-58

Lagudah ES, Moullet O, Appels R (1997) Map-based cloning of a gene sequence encoding a nucleotide-binding domain and a lecuine-rich region at the *Cre3* nematode resistance locus of wheat. Genome 40: 659-665

Lagudah ES, Seah S, Sivasithamparam K, Karakousis A (1998) Cloning and characterization of a family of disease resistance gene analogs from wheat and barley. Theor App Genet 97: 937-945

Larionov V, Kouprina N, Nikolaishvili N, Resnick M (1994) Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YACs). Nucl. Acids. Res. 22: 4154-4162

Lavia G (1996) Estudios cromosomicos en *Arachis* (Leguminosae). Bonplandia 9: 111-120

Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. Plant Cell 7: 1195-1206

Leister D, Ballvora A, Salamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genetics 14: 421-429

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proc. Natl. Acad. Sci 95: 370-375

Leister RT, Katagiri F (2000) A resistance gene product of the nucleotide binding site-leucine rich repeats class can form a complex with bacterial avirulence proteins *in vivo*. Plant J 22: 345-354

Li L, Lu S, O'Halloran DM, Garvin DF, Vrebalov J (2003) High-resolution genetics and physical mapping of the cauliflower high-beta-carotene gene *Or* (*Orange*). Mol Gen Genet 270: 132-138

Ling HQ, Zhu Y, Keller B (2003) High-resolution mapping of the leaf rust disease resistance gene *Lr1* in wheat and characterization of BAC clones from the *Lr1* locus. Theor App Genet 106: 875-882

Liu G-Z, Pi L-Y, Walker JC, Ronald PC, Song W-Y (2002) Biochemical characterization of the kinase domain of the rice disease resistance receptor-like kinase *Xa21*. J. Biol. Chem. 277: 20264-20269

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

Lopez CE, Zuluaga AP, Cooke R, Delseny M, Tohme J, Verdier V (2003) Isolation of resistance gene candidates (RGCs) and characterization of an RGC cluster in cassava. Mol Genet Genomics 269: 658-671

Luck JE, Lawrence GJ, Dodds PN, Shepherd KW, Ellis JG (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. Plant Cell 12: 1367-1378

Luo MZ, Wang YH, Frisch D, Joobeur T, Wing RA, Dean RA (2001) Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium wilt* (*Fom-2*). Genome 44: 154-162

Maleki L, Faris JD, Bowden RL, Gill BS, Fellers JP (2003) Physical and genetic mapping of wheat kinase analogs and NBS-LRR resistance gene analogs. Crop Sci 43: 660-670

Marek LF, Shoemaker RC (1997) BAC contig development by fingerprint analysis in soybean. Genome 40: 420-427

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH (1997) High throughput fingerprint analysis of large-insert clones. Genome Res. 7: 1072-1084

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

Mauricio R, Stahl EA, Korves T, Tian DC, Kreitman M, Bergelson J (2003) Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. Genetics 163: 735-746

Meksem K, Pantazopoulos P, Njiti VN, Hyten LD, Arelli PR, Lightfoot DA (2001) 'Forrest' resistance to the soybean *cyst* nematode is bigenic: saturation mapping of the *Rhg1* and *Rhg4* loci. Theor App Genet 103: 710-717

Meksem K, Zobrist K, Ruben E, Hyten D, Quanzhou T, Zhang H-B, Lightfoot D (2000) Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome wide physical mapping. Theor App Genet 101: 747-755

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10: 1817-1832

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: 317-322

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

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8: 1113-1130

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM (1998) The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10: 1307-1319

Ming R, Moore PH, Zee F, Abbey CA, Ma H, Paterson AH (2001) Construction and characterization of a papaya BAC library as a foundation for molecular for molecular dissection of a tree-fruit genome. Theor App Genet 102: 892-899

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

Morel JB, Dangl JL (1997) The hypersensitive response and the induction of cell death in plants. Cell Death Differ. 4: 671-683

Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martienssen R, Meier-Ewert S, Altmann T (1999) A complete BAC-based physical map of the *Arabidopsis thaliana* genome. Nature Genetics 22: 271-275

Nam YW, Penmetsa RV, Endre G, Uribe P, Kim D, Cook DR (1999) Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. Theor App Genet 98: 638-646

Neil DL, Villasante A, Fisher RB, Vetrie C, Cox B, Tyler-Smith C (1990) Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. Nucl. Acids. Res. 18: 1421-1428

Nelson SC, Simpson CE, Starr JL (1989) Resistance to *Meloidogyne arenaria* in *Arachis* spp. germplasm. J Nematol Suppl 21: 654-660

Noel L, Moores TL, van der Biezen EA, Parniske M, Daniels MJ, Parker JE, Jones JDG (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. Plant Cell 11: 2099-2112

Noir S, Combes M-C, Anthony F, Lashermes P (2001) Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea L.*). Mol Gen Genet 265: 654-662

Nombela G, Williamson VM, Muniz M (2003) The root-knot nematode resistance gene *Mi-1.2* of tomato is responsible for resistance against the whitefly *Bemisia tabaci*. Mol Plant Microbe Inter 16: 645-649

O'Connor M, Peifer M, Bender W (1989) Construction of large DNA segmets in *Escherichia coli*. Science 244: 1307-1312

Ogbonnaya FC, Seah S, Delibes A, Jahier J, Lopez-Brana I, Eastwood RF, Lagudah ES (2001) Molecular-genetic characterization of a new nematode resistance gene in wheat. Theor App Genet 102: 623-629

Ohmori T, Murata M, Motoyoshi F (1998) Characterization of disease resistance gene-like sequences in near-isogenic lines of tomato. Theor App Genet 96: 331-338

Osoegawa K, Woon PY, Zhao B, Eirik F, Tateno M, Catanese JJ, de Jong PJ (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52: 37994

Pan Q, Liu Y-S, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D, Fluhr R (2000a) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and *Arabidopsis*. Genetics 155: 309-322

Pan Q, Wendel J, Fluhr R (2000b) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50: 203-213

Parniske M, Jones JDG (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc. Natl. Acad. Sci 96: 5850-5855

Pensuk V, Jogloy S, 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

Penuela SD, Danesh D, Young ND (2002) Targeted isolation, sequence analysis, and physical mapping on nonTIR NBS-LRR genes in soybean. Theor App Genet 104: 261-272

Pflieger S, Lefebvre V, Caranta C, Blattes A, Goffinet B, Palloix A (1999) Disease resistance gene analogs as candidates for QTLs involved in pepper-pathogen interactions. Genome 42: 1100-1110

Phipps PM, Porter DM (1998) Collar rot of peanut caused by *Lasiodiplodia theobromae*. Plant Dis 82: 1205-1209

Raina SN, Mukai Y (1999) Genomic in situ hybridization in *Arachis* (Fabaceae) identifies the diploid wild progenitors of cultivated (*A-hypogaea*) and related wild (*A-monticola*) peanut species. Plant Systematics and Evol. 214: 251-262

Raina SN, Rani V, Kojima T, Ogihara Y, Singh KP, Devarumath RM (2001) RAPD and ISSR fingerprintings as useful genetic markers for analysis of genetic diversity , varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. Genome 44: 763-772

Ramakrishna W, Emberton J, Ogden M, SanMiguel P, Bennetzen JL (2002a) Structural analysis of the maize *Rp1* complex reveals numerous sites and unexpected mechanisms of local rearrangement. Plant Cell 14: 3213-3223

Ramakrishna W, Emberton J, SanMiguel P, Ogden M, Llaca V, Messing J, Bennetzen JL (2002b) Comparative sequence analysis of the sorghum *Rph* region and the maize *Rp1* resistance gene complex. Plant Physiol. 130: 1728-1738

Resslar PM (1980) A review of the nomenclature of the genus *Arachis* L. Euphytica 29: 815-819

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

Richter TE, Pryor TJ, Bennetzen JL, Hulbert SH (1995) New rust resistance specificities associated with recombination in the *Rp1* complex in Maize. Genetics 141: 373-381

Riely BK, Martin GB (2001) Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene *Pto*. Proc. Natl. Acad. Sci 98: 2059-2064

Rivas S, Romeis T, Jones JDG (2002) The *Cf-9* Disease resistance protein is present in an ~420-kilodalton heteromultimeric membrane-associated complex at one molecule per complex. Plant Cell 14: 689-702

Ronald PC (1998) Resistance gene evolution. Current Opinion in Plant Biology 1: 294-298

Ross MT, LaBrie T, McPherson J, Stanton VM (1999) Screening large-insert libraries by hybridization. Wiley, New York, NY

Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM (1998) The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. Proc. Natl. Acad. Sci 95: 9750-9754

Salimath SS, Bhattacharyya MK (1999) Generation of a soybean BAC library, and identification of DNA sequences tightly linked to the *Rps-1* disease resistance gene. Theor App Genet 98: 712-720

Salmeron JM, Oldroyd GE, Rommens CM, Scofield SR, Kim HS, Lavelle DT, Dahlbeck D, Staskawicz BJ (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86: 123-133

Saraste M, Sibbald PR, Wittinghofer A (1990) The P-Loop - a Common motif in ATP-binding and GTP-binding proteins. Trends Biochem Sci 15: 430-444

Sasser JN (1980) Root knot nematodes : a global menace to crop production. Plant Dis 64: 36-41

Schmidt R, Putterill J, West J, Cnops G, Robson F, Coupland G, Dean C (1996) Detailed description of four YAC contigs representing 17 Mb of chromosome 4 of *Arabidopsis thaliana* ecotype Columbia. Plant J. 9: 755-760

Seah S, Sivasithamparam K, Karakousis A, Lagudah ES (1998) Cloning and characterization of a family of disease resistance gene analogs from wheat and barley. Theor App Genet 97: 937-945

Sessa G, D'Ascenzo M, Martin GB (2000) Thr38 and Ser198 are *Pto* autophosphorylation sites required for the *AvrPto-Pto*-mediated hypersensitive response. EMBO J. 19: 2257-2269

Shen KA, Meyers BC, Islam-Faridi MN, Chin D, Stelly DM, Michelmore RW (1998) Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Mol Plant-Microbe Int 8: 815-823

Shizuya H, Birren B, Kim U-J, Mancino V, Slepak T, Tachiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA 89: 8794-8797

Simons G, Groenendijk J, Wijbrandi J, Reijans M, Groenen J, Diergaarde P, Van der Lee T, Bleeker M, Onstenk J, de Both M, Haring M, Mes J, Cornelissen B, Zabeau M, Vos P (1998) Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10: 1055-1068

Simpson CE, Krapovickas G, Valls JFM (2001a) History of Arachis including evidence of *A. hypogaea* L. progenitors. Peanut Science 28: 78-79

Simpson CE, Nelson SC, Starr JL, Woodard KE, Smith OD (1993) Registration of TxAG-6 and TxAG-7 peanut germplasm lines. Crop Sci 33: 1418

Simpson CE, Smith OD, Grichar WJ, Melouk HA (1991) Pathways for introgression of pest resistance into *Arachis hypogea* L. Peanut Science 18: 22-26

Simpson CE, Starr JL, Church GT, Burow MD, Paterson AH (2001b) Registration of 'COAN' peanut. Crop Sci 41: 918-918

Simpson CE, Starr JL, Church GT, Burow MD, Paterson AH (2003) Registration of 'NemaTAM' peanut. Crop Sci 43: 1561-1561

Singh AK, Smartt J (1998) The genome donors of the groundnut/peanut (*Arachis hypogaea* L.) revisited. Genetics Res and Crop Evol 45: 113-118

Singh KP, Raina SN, Singh AK (1996) Variation in chromosomal DNA associated with the evolution of *Arachis* species. Genome 39: 890-897

Singh KP, Singh A, Raina SN, Singh AK, Ogihara Y (2001) Ribosomal DNA associated with polymorphism and heritability in peanut (*Arachis hypogaea* L.) accessions and related wild species. Euphytica 123: 221-230

Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. 10: 1772-1787

Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald PC (1997) Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9: 1279-1287

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

Speulman E, Bouchez D, Holub EB, Beynon JL (1998) Disease resistance gene homologs correlate with disease resistance loci of *Arabidopsis thaliana*. Plant J 14: 467-474

Stahl E, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400: 667-671

Stalker H (1991) A new species in section *Arachis* of peanuts with a D genome. Am Journal Bot 78: 630-637

Stalker H, Dalmacio R (1986) Karyotype analysis and relationships among varieties of *Arachis hypogaea* L. Cytologia 51: 617-629

Stalker H, Dhesi J, Parry D, Hahn J (1991) Cytological and Interfertility relationships of *Arachis* Section *Arachis*. Am Journal Bot 78: 238-246

Stalker HT, Simpson CE (1995) Germplasm resources in *Arachis*. Advances in peanut science. American Peanut Research and Education Society, Inc., Stillwater, OK

Starr JL, Simpson CE (1990) Segregation of resistance to *Meloidogyne arenaria* in progeny of interspecific *Arachis* spp Hybrid. Peanut Science 17

Staskawicz BJ, Ausubel FM, Baker BJ, Ellis JG, Jones JD (1995) Molecular genetics of plant disease resistance. Science 268: 66-667

Stebbins GL (1957) Genetics, evolution, and plant breeding. Indian J. Genet. Plant Breed 17: 129-141

Strong S, Ohta Y, Litman G, Amemiya C (1997) Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. Nucl. Acids. Res. 25: 3959-3961

Sudupak MA, Bennetzen JL, Hulbert SH (1993) Unequal exchange and meiotic instability of disease-resistance genes in the *Rp1* region of maize. Genetics 133: 119-125

Sulston J, Mallett F, Durvin R, Horsnell T (1989) Image analysis of restriction enzyme fingerprint audioradiograms. Comput. Appl. Biosci. 5: 101-106

Sun Q, Collins NC, Ayliffe M, Smith SM, Drake J, Pryor T, Hulbert SH (2001) Recombination between paralogues at the *rp1* rust resistance locus in maize. Genetics 158: 423-438

Tameling WIL, Elzinga SDJ, Darmin PS, Vossen JH, Takken FLW, Haring MA, Cornelissen BJC (2002) The tomato R gene products *I-2* and *Mi-1* are functional ATP binding proteins with ATPase activity. Plant Cell 14: 2929-2939

Tamulonis JP, Luzzi BM, Hussey RS, Parrott WA, Boerma HR (1997a) DNA marker analysis of loci conferring resistance to peanut root-knot nematode in soybean. Theor App Genet 95: 664-670

Tamulonis JP, Luzzi BM, Hussey RS, Parrott WA, Boerma HR (1997b) RFLP mapping of resistance to southern root-knot nematode in soybean. Crop Sci 37: 1903-1909

Tang XY, Frederick RD, Zhou JM, Halterman DA, Jia YL, Martin GB (1996) Initiation of plant disease resistance by Physical Interaction of *AvrPto* and *Pto* Kinase. Science 274: 2060-2063

Tao Q, Chang Y-L, Wang J, Chen H, Islam-Faridi MN, Scheuring C, Wang B, Stelly DM, Zhang H-B (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158: 1711-1724

Temsch EM, Greihulber J (2000) Genome size variation in *Arachis hypogaea* and *A-monticola* re-evaluated. Genome 43: 449-451

Temsch EM, Greihulber J (2001) Genome size in *Arachis duranensis*: a critical study. Genome 44: 826-830

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815

The International Human Genome Mapping Consortium (2001) A physical map of the human genome. Nature 409: 934-941

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, Jones J (1997) Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. Plant Cell 9: 2209-2224

Thurau T, Kifle S, Jung C, Cai DG (2003) The promoter of the nematode resistance gene *Hs1(pro-1)* activates a nematode-responsive and feeding site-specific gene expression in sugar beet (*Beta vulgaris L.*) and *Arabidopsis thaliana*. Plant Mol Biol 52: 643-660

Tian Y, Fan L, Thurau T, Jung C, Cai D (2004) The absence of TIR-type resistance gene analogues in the sugar beet (*Beta vulgaris* L.) genome. J Mol Evol 58: 40-53

Tomkins JP, Mahalingam R, Smith H, Goicoechea JL, Knap HT, Wing RA (1999) A bacterial artificial chromosome library for soybean PI 437654 and identification of clones associated with *cyst* nematode resistance. Plant Mol Biol 41: 25-32

Torii KU, McNellis TW, Deng XW (1998) Fuctional dissection of *Arabidopsis COP1* reveals specific roles of its three structural modules in light control of seedling development. EMBO J. 17: 5577-5587

Traut TW (1994) The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. European J Biochem 222: 9-19

Vallad G, Rivkin M, C V, McClean P (2001) Cloning and homology modelling of a *Pto*-like protein kinase family of common bean (*Phaseolus vulgaris* L.). Theor App Genet 103: 1046-1058

van der Biezen EA JJ (1998) The NB-ARC domain: A novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr Biology 8: 226-227

Van der Hoorn RAL, Roth R, De Wit PJGM (2001) Identification of distinct specificity determinants in resistance protein *Cf-4* allows construction of a *Cf-9* mutant that confers recognition of avirulence protein *AVR4*. Plant Cell 13: 273-285

Van der Vossen EAG, Rouppe van der Voort JNAM, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ, Klein-Lankhorst RM (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. Plant J 23: 567-576

VandenBosch KA, Stacey G (2003) Summaries of legume genomics projects from around the globe. Community resources for crops and models. Plant Physiol. 131: 840-865

Vanhouten W, MacKenzie S (1999) Construction and characterization of a common bean bacterial artificial chromosome library. Plant Mol Biol 40: 977-983

Venter JC, Adams MD, Granger GS, Anthony RK, Smith HO, Hunkapiller M (1998) Shotgun sequencing of the human genome. Science 280: 1540-1542

Vilarinhos AD, Priffanelli P, Lagoda P, Thibivilliers S, Sabau X, Carreel F, D'Hont AD (2003) Construction and characterization of a bacterial artificial chromosome library of banana (*Musa acuminata* Colla). Theor App Genet 106: 1102-1106

Vos P, Simons G, Jesse T, Wijbrandi J, Heinen L, Hogers R, Frijters A, Groenendijk J, Diergaarde P, Reijans M, Fierens-Onstenk J, de Both M, Peleman J, Liharska T, Hontelez J, Zabeau M (1998) The tomato *Mi-1* gene confers resistance to both root-knot nematodes and potato aphids. Nat Biotechno 16: 1365-1369

Wang GL, Holsten TE, Song WY, Wang H-P, Ronald PC (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance gene locus. Plant J 7: 525-533

Wang G-L, Ruan D-L, Song W-Y, Sideris S, Chen L, Pi L-Y, Zhang S, Zhang Z, Fauquet C, Gaut BS, Whalen MC, Ronald PC (1998) *Xa21D* encodes a receptor-like molecule with a leucine-rich repeat comain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10: 765-780

Wang GL, Warren R, Innes R, Osborn B, Baker B, Ronald PC (1996) Construction of an arabidopsis BAC library and isolation of clones hybridizing with disease resistance gene-like sequences. Plant Mol Biol Rep 14 107-114

Wang Q, Zhang K, Qu X, Jia J, Shi J, Jin D (2001a) Construction and characterization of a bacterial artificial chromosome library of peach. Theor App Genet 103: 1174-1179

Wang W, Zhai W, Luo M, Jiang G, Chen X, Li X, Wing RA, Zhu L (2001b) Chromosome landing at the bacterial blight resistance gene *Xa4* locus using a deep coverage rice BAC library. Mol Gen Genet 265: 118-125

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP (1999) The *Mla* (Powdery Mildew) Resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 153: 1929-1948

Wei F, Wing RA, Wise RP (2002) Genome dynamics and evolution of the *Mla* (Powdery Mildew) resistance locus in barley. Plant Cell 14: 1903-1917

Whitham S, McCormick S, Baker B (1996) The *N* gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. Proc. Natl. Acad. Sci 93: 8776-8781

Willets N, Skurray R (1987) Structure and function of F-factor and mechanism of conjugation. In:. In: FC N (ed) *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol 2. Am Soc Microbiol, Washington D.C, pp 110-1137

Williamson VM (1999) Plant nematode resistance genes. Current Opinion in Plant Biology 2: 327-331

Williamson VM, Gleason CA (2003) Plant-nematode interactions. Current Opinion in Plant Biology 6: 327-333

Woo S-S, Jiang J, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucl. Acids. Res. 22: 4922-4931

Woo SS, Rastogi VK, Zhang HB, Paterson AH, Schertz KF, Wing RA (1995) Isolation of megabase-size DNA from sorghum and applications for physical mapping and bacterial and yeast artificial chromosome library construction. Plant Mol Biol Rep 13: 82-94

Wu C, Sun S, Nimmakayala P, Santos FA, Meksem K, Springman R, Ding K, Lightfoot DA, Zhang HB (2004) A BAC and BIBAC-based physical map of the soybean genome. Genome Res 14: 319-316

Wulff BBH, Thomas CM, Smoker M, Grant M, Jones JDG (2001) Domain swapping and gene shuffling Identify sequences required for induction of an avr-dependent hypersensitive response by the tomato *Cf-4* and *Cf-9* proteins. Plant Cell 13: 255-272

Xu J, Daichang Y, Domingo J, Ni J, Huang N (1998) Screening for overlapping bacterial artificial chromosome clones by PCR analysis with an arbitrary primer. Proc. Natl. Acad. Sci 95: 5661-5666

Xu M, Korban SS (2002) A cluster of four receptor-like genes resides in the *Vf* locus that confers resistance to apple scab disease. Genetics 162: 1995-2006

Yan GP, Chen XM, Line RF, Wellings CR (2003) Resistance gene-analog polymorphism markers co-segregating with the *YR5* gene for resistance to wheat stripe rust. Theor App Genet 106: 636-643

Yim Y-S, Davis GL, Duru NA, Musket TA, Linton EW, Messing JW, McMullen MD, Soderlund CA, Polacco ML, Gardiner JM, Coe EH, Jr. (2002) Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. Plant Physiol. 130: 1686-1696

Young ND (1996) QTL Mapping and quantitative cisease resistance in plants. Annual Review of Phytopathology 34: 479-501

Young ND, Mudge J, Ellis TN (2003) Legume genomes: more than peas in a pod. Current Opinion in Plant Biology 6: 199-204

Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H,

Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A Draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296: 79-92

Yu Y, Buss G, Maroof M (1996) Isolation of a superfamily of candidate resistance genes in soybean based on a conserved nucleotide binding site. Proc. Natl. Acad. Sci 93: 11751-11756

Zhou J, Loh YT, Bressan RA, Martin GB (1995) The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by *Pto* and is involved in the hypersensitive response. Cell 83: 925-935

Zhu HY, Cannon SB, Young ND, Cook DR (2002) Phylogeny and genomic organization of the TIR and non-TIR NBS-LRR resistance gene family in *Medicago truncatula*. Mol Plant Microbe Inter 15: 529-539

CHAPTER 2

CONSTRUCTION AND CHARACTERIZATION OF A PEANUT BAC LIBRARY[1]

[1] Yuksel B. and A.H. Paterson. To be submitted to *Theoretical and Applied Genetics*.

ABSTRACT:

BAC libraries have been an essential tool for physical analyses of genomes of many crops. We constructed and characterized the first large-insert DNA library for *Arachis hypogaea L*. The library contains 182, 784 clones; only 5484 (3%) had no inserts; and the average insert size is 104.05 kb. Chloroplast DNA contamination was very low, only nine clones. r-DNA content was 1208, 0.66% of clones. The depth of coverage is estimated to be 6.5 genome-equivalents, allowing the isolation of virtually any single-copy locus. We also applied 24 overgos derived from the genetically mapped RFLP probes to confirm the rate of coverage. A total of 655 hits were obtained. After exclusion of the four overgos that hybridized to repetitive sequences, the number of hits, 305, to the remaining 20 overgos was consistent with the estimated depth. The identification of multiple loci by most probes in polyploids complicates anchoring of physical and genetic maps. We explored the practicality of a hybridization-based approach for determination of map locations of BAC clones in peanut by analyzing 94 clones detected by seven different overgos. The banding patterns on Southern blots were good predictors of contig composition. This BAC library has great potential to advance the future research about the peanut genome.

Key words: *Arachis hypogaea*, large-insert DNA library, overgo, rDNA, physical mapping

INTRODUCTION:

Arachis hypogaea L is one of the most important food and oil-seed crops with 37,057,652 Gg of world-wide production (Food and Agriculture Organization 2003). Peanut is grown worldwide; leading producers are China and India. The section Arachis of genus *Arachis* contains about 69 species, which are mostly A and B genome diploids with the exception of the D genome diploid *A. glandulifera* (Stalker 1991; Stalker et al. 1991). The cultivated peanut *Arachis hypogaea* L. and its wild relative *Arachis monticola* (2n=4x=40) are the only known AABB tetraploids (Fernandez and Krapovickas 1994).

Like other monophyletic amphidiploid species, the genetic variation within and between tetraploid peanut species is very limited, hindering genome analysis by genetic mapping. A large-insert DNA clone library will be valuable for detailed analysis of the peanut genome, by permitting the use of physical mapping approaches that are not dependent upon genetic polymorphism.

Bacterial artificial chromosome libraries (BAC) have emerged as the preferred method for the construction of large insert size clone libraries, supplanting YACs (Burke et al. 1987) due to ease of manipulation (Anderson 1993), lower levels of instabilities and rearrangements in clones consisting of highly repetitive DNA (Dunford et al. 1993; Neil et al. 1990), and the abundance of chimeric clones in YACs (Larionov et al. 1994; Schmidt et al. 1996). The first F-factor based vector in *E. coli* for the propagation of large inserts has been described by O'Connor et al. (1989). F-factor-based genes parA and parB ensure the single copy maintenance of the plasmid, preventing co-transformation and recombinant clones, common in YACs (Larionov et al. 1994; Willets and Skurray 1987). Overall, bacteria grow much faster than yeast and bacterial clones are

much easier to work with. Although BAC clones with insert sizes up to 300 kb have been reported, the average size of BAC libraries is usually between 100 kb and 150 kb.

BAC libraries for many important crops from very diverse plant families already exist; *Sorghum bicolor* (sorghum-Woo et al. 1994), *Arabidopsis thaliana* (Choi et al. 1995), *S. propinquum* (Lin et al. 1999), *Carica papaya* (papaya-Ming et al. 2001), *Helianthus annuus* (sunflower-Gentzbittel et al. 2002), *Lactuca sativa* (lettuce-Frijters et al. 1997), *Prunus persica* (peach-Wang et al. 2001a), *Triticum aestivum* (wheat-Cenci et al. 2003), *Musa acuminata* (banana-Vilarinhos et al. 2003), and *Brassica oleracea var. botrytis* (broccoli-Li et al. 2003). In Leguminosae, multiple libraries have been made for *Glycine max* (soybean), *Medicago truncatula* (Danesh et al. 1998; Meksem et al. 2000; Nam et al. 1999; Salimath and Bhattacharyya 1999; Tomkins et al. 1999), and *Phaseolus vulgaris* (common bean-Vanhouten and MacKenzie 1999). However, no BAC libraries have been previously constructed for *Arachis spp*.

The primary objective of this study was the construction and characterization of an *A. hypogaea* BAC library with sufficient genome coverage for genomic analyses.

METHODS AND MATERIALS:

a) Plant Material

Arachis hypogaea L. cv. Florunner UF-439-16-1003-2 was grown in the greenhouse. Leaf material was collected at the seedling stage. The collected leaf tissue was washed in ddH_2O and frozen in liquid nitrogen in 3 to 4 g packages then stored at -80 ºC until needed.

b) Preparation of HMW DNA

The protocol for isolation of peanut nuclei was created by modifying a DNA extraction method developed by Burow et al. (2001). The frozen plant material was ground in liquid

nitrogen with a mortar and pestle until very fine powder was obtained , then the powdered sample was dissolved in extraction buffer: [0.005 M citric acid, 0.5 M glucose, 0.01 M Na_2EDTA, 2.0% (w/v) polyvinylpyrrolidone-40 (PVP-40), 5% (v/v) Triton X-100, 0.25% (w/v) spermidine, titrated to pH 6.5 with NaOH] to which fresh antioxidants [0.1% (w/v) ascorbic acid, 0.2% (v/v) 2-mercaptoethanol, 0.1% (w/v) disodium diethylthiocarbamate (Na_2Et_2dtc), and 0.4% (w/v) $NaHSO_3$] were added. After dissolving on ice for about 10 min, the samples were filtered twice through 8 layers of cheese cloth and 2 layers of miracloth. The filtrates were centrifuged for 20 min at 160 g to precipitate the nuclei. The precipitated nuclei were dissolved in extraction buffer and centrifuged at the same speed, for another 3-4 times until the nuclei looked clean. The last wash of nuclei was done in the same extraction buffer without Triton X-100. The nuclei were pre-warmed at 42 ºC for 5 min before mixing with an equal volume of 1% LMP agarose, and then plugs were formed. Plugs were incubated at 50 ºC for 24 hr in lysis buffer [0.005 M citric acid, 0.14 M NaCl, 0.05 M Na_2EDTA, 2% (w/v) PVP-40, 1% (w/v) SDS, 1% Sodium Lauryl Sarcosine titrated to pH 6.5 with NaOH, and autoclaved] and the same antioxidants at similar proportions as in extraction buffer and 0.06 mg/ml Proteinase K were freshly added. The buffer was replaced and plugs were incubated at the same temperature for another 24 hrs. The plugs were incubated at room temperature for at least 4 hrs in 70% ethanol then stored at -20 ºC until use.

We increased pH of extraction and lysis buffers from 5.0 to 6.5, after noticing that the low pH damaged DNA and resulted in low ligation efficiency. HMW DNA with size of larger than 1000 kb is necessary (Fig 2.2)

c) Quality Check

The intactness and concentration of plugs were checked by electrophoresis on 1%, 0.5 × TBE (45 mM TRIS, 45 mM TRIS-Borate, 1 mM EDTA) agarose for 18 hrs at 6 V/cm, with 1 sec to 40 sec switch times, linear ramp, in a contour clamped homogenous electric field apparatus (CHEF) DRIII (BioRad). The gel was stained in ethidium bromide and imaged.

d) Test Digestion

Plugs were incubated in 0.05 M EDTA on ice until ethanol effused. After that, plugs were incubated for 1 hr at 50 ºC in 0.5 M EDTA, pH 9.3, then another 1 hr on ice in 0.05 M EDTA. Following that, the plugs were incubated on ice in $T_{10}E_{10}$ for 30 min, and then incubated in 0.1 µM PMSF, $T_{10}E_{10}$ solution for another 1 hr. Next, the plugs were washed with $T_{10}E_{10}$ for 2 hr on ice. The washed plugs were minced and incubated in a 1 ml buffer cocktail (1×NEB buffer 2, 1mM BSA, 1 mM DTT, and 4 mM spermidine) for 30 min, and then the buffer refreshed and incubated another 30 min on ice. Finally, buffer cocktail was removed again, and 250 µl fresh buffer cocktail was added to each sample. Serial dilutions of *Hin*dIII (NEB) of 0, 0.1, 0.15, 0.2, 0.3, 0.5, 0.75, 1 and 2 units per mg of plug were added and incubated at 4 ºC for 4 hrs. Partial digestion was done at 37 ºC for 7 min, and 0.5 M EDTA added to the tubes to stop the reactions. The partially digested samples were resolved in 1% agarose run in 0.5 × TBE buffer by pulsed field gel electrophoresis (PFGE) with at 6 V/cm, with 1 sec to 40 sec switch times, and linear ramp, for 18 hr. Optimum enzyme concentration was determined by visualizing maximum fragment concentration in the 150 kb to 350 kb range.

e) Size Selection and Elution

The plugs were treated the same as described for test digestions, partially digested in predetermined enzyme concentrations and DNA electrophoresced under the same conditions.

Markers on the gel were stained, and the gel fraction 125-350 kb was excised. The gel fraction was divided into three different equal sub-fractions, X, Y, and Z. A second round of size selection was applied to X, Y, and Z fractions in order to augment the desired size DNA fragments and purge small fragments stuck behind large fragments after the first size selection. Electrophoresis parameters were set to 3 s to 5 s switch times, 6 V/cm, linear ramp, and 18 hrs. In the second size selection gel, the size fraction larger than 125 kb was excised and gel fragments were saturated in 70% ethanol at room temperature for 4 hrs before storing at -20 ºC until use. Excised gel pieces were washed in 1 × TAE buffer on ice until ethanol diffused out of them. The DNA was eluted with an Electro-Eluter (Bio Rad) by applying a constant 10 mA of current for each cuvette for 2 hrs. The concentration of the eluted DNA was checked and concentrated with 0.025 μm nitrocellulose membranes (MilliPore) on 10% PEG, if needed.

f) BAC Vector

Three types of BAC vector were used in the construction of the library including, pBeloBAC (Home-made), pCUGIBAC1 (kindly provided by Dr. J. Tomkins, Clemson University Genomics Institute, SC) and pIndigoBAC-5 (Epicentre Technologies). The vector used in each ligation and number of colonies picked are denoted in Table 2.1. The pBeloBAC: Plasmid was isolated by an alkaline-lysis method (Sambrook and Russell, 2001) from 10 L of LB CM culture. Isolated plasmid was dissolved in TE overnight at 4º and closed-circle plasmid was purified by applying to a CsCl ethidium bromide gradient twice (Sambrook and Russell, 2001. The purified plasmid samples were dialyzed twice in TE for 24 hrs at 4º, and plasmid was precipitated with 95% ethanol, 1/10 (v/v) of 7.5 M $NH_4CH_3CO_2$, and washed once in 70% ethanol. Ten μg of the plasmid was digested with *Hin*dIII and dephosphorylated with HK phosphatase (Epicentre Technologies). The dephosphorylated vector was self-ligated and run on

0.8%, 1 × TAE gel and the open circle plasmid band was excised and eluted. λHIII fragments were ligated into the vector to test the efficiency of dephosphorylation. pCUGIBAC1**:** The same method as pBeloBAC was used for vector isolation except much smaller amount of culture (1 L) was required because of the high-copy nature of the plasmid. CsCl-ethidium bromide gradient was not applied, but vector was digested, dephosphorylated and tested in the same way. The pIndigoBAC-5: This vector was purchased from Epicentre Technologies, dephosphorylated, which is ready-to-clone.

Ligation, Transformation, Insert-Size Characterization and Picking: For ligation, a constant 15 ng of vector was used and varying amounts of insert ranging from 60 ng to 120 ng were tested for each size selection. The vector/insert ratio, which gave the best efficiency and average insert size, was chosen. Ligation reactions were performed in 60 µl volume and incubated at 16º for 10 hrs. After desalting, 2-3 µl of reactions were transformed into *E. coli* DH10B competent cells (DH10B ElectroMAX^{TM} T1 Phage resistant and ElectroMAX DH10B competent cells, Invitrogen) by electroporation (BioRad Gene Pulser II). For electroporation, 2.5 kV and 200 Ω resistance was used. The electroporated cells were immediately mixed with 1ml of SOC media (Sambrook and Russell 2001), and grown for 1 hr at 37 ºC before separation on selective medium (LB, Luria-Bertani medium) with 12.5 µg chloroamphenicol, 0.55 mM IPTG, and 80 µg/ml X-gal. After 18 hr of incubation at 37 ºC, a sampling of 10-20 colonies were picked and tested.

Randomly selected white colonies were inoculated into 1 ml LB CM liquid medium and grown for 16 hrs at 37º. The liquid cultures were miniprepped by an alkaline lysis protocol and the DNA was digested with 10 units of *Not*I (NEB) for 4 hrs at 37º. The digested samples were resolved on 1%, 0.5×TBE agarose by pulsed field gel electrophoresis with parameters of 3 to 20

s linear ramp, 6 V/cm for 16 hr. The ligation reactions with average insert size of 100 kb or more were mass-transformed, plated, and directly picked with a Q BOT (Genetix) into 384-well plates. The clones were stored in FM medium [LB + 36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM Sodium Citrate, 0.4 mM $MgSO_4$, 6.8 mM $(NH_4)_2SO_4$ and 4.4% glycerol] at -80 ºC.

g) Gridding Peanut BAC Library and Preparing High-Density Filters for Hybridization:

A total of 182, 784 *Arachis hypogaea* BAC clones was gridded on 22.5 cm^2 Hybond N+ membranes (Amersham Life Sciences, Arlington Heights, IL, USA) with a Q BOT (Genetix). Each clone was double spotted in 4×4 arrays allowing representation of 18, 432 different clones per filter. Thus, the whole peanut library fit in total onto 10 filters. The filters were incubated on 1% LB agar containing 12.5 µg/µl CM, at 37 ºC for 12 to 18 hrs until optimal colony growth-size was obtained. The high-density BAC filters were processed according to a standard alkaline-lysis method (Sambrook and Russell 2001). The filters were dried overnight and stored at -4 ºC.

h) Fingerprinting, Blotting, and Labeling

BAC clones were inoculated into 1.5 ml 2×YT medium (Sambrook and Russell 2001) containing 12.5 µg/µl CM, and grown at 37 ºC for 18 hrs. Minipreps were performed according to a standard alkaline-lysis method. Afterwards, the samples were digested with 40 U of *Hin*dIII (NEB) at 37 ºC for 4.5 hrs. Subsequently, the digestion products were resolved on 1% agarose gel in 1×TAE buffer for 16 hrs at 95 V. The gel was stained with SYBR Green and imaged. After analysis of the gel image with IMAGE (Sulston et al. 1989), the contigs were built with FPC V 4.7 (Soderlund et al. 2000).

The fingerprinting gel was blotted onto Hybond N+ (Amersham-Pharmacia, Piscataway, NJ) in 0.4 N NaOH solutions.

Overgo labeling reactions were set in a total volume of 15 µl containing 0.0067 nM forward and reverse oligonucleotide primers that were denatured at 94 ºC for 5 min and cooled on ice, 1 µg BSA, 2.5 U Taq polymerase, 1 µl α^{32}-dATP (6000 Ci/mmol) (MP Biomedicals, CA), 1 µl α^{32}-dCTP(6000 Ci/mmol) (MP Biomedicals, CA), and 3 µl OLB (Oligo labeling buffer without dATP, dCTP, and random hexamers) (Ross et al. 1999). The reaction mixture was incubated at 37 ºC for 2 hours (Fig 2.1). The labeled overgos were filtered through Sephadex columns to remove any unincorporated radioactive nucleotides, and heat denatured at 94 ºC for 5 minutes before adding to the hybridization bottle.

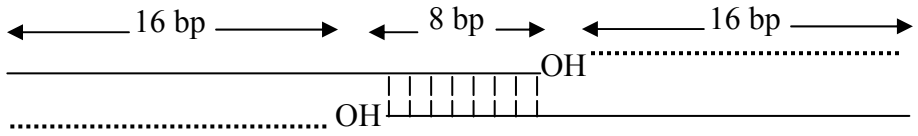


Fig 2.1 Schematic representation of an overgo structure and labeling; the dashed lines indicate the ^{32}P labeled portion of the probe

A 100 ng of four cotton chloroplast sequences (300-400 bp) were labeled by random hexamer protocol, using 25 µCi of α^{32}-dCTP(6000 Ci/mmol) and Klenow fragment for primer extension at 37 ºC for 2 hours (Chittenden et al. 1994). The unincorporated radioactivity was removed by filtering through a Sephadex column.

The membranes separated by a nylon mesh were placed into the rotisserie bottles with 15 ml (Southern blot) and 85 ml (BAC filters) of hybridization buffer [0.5 M sodium phosphate, pH 7.2, 7% (w/v) SDS, 1 mM EDTA, and 0.01% (w/v) BSA] and incubated at 58 ºC for Southern

and 55 ºC for BAC filters before adding the denatured radioactive overgo probes. Hybridizations were performed at 58 ºC and 55 ºC (for BAC filters) with rotation speed of 4.5 rpm for 18 to 36 hrs in a rotisserie oven. Following the hybridization process, the membranes were washed at 55 ºC for 30 min each with constant shaking in a tray; first at low stringency in wash buffer II [1×SSPE (0.15 M NaCl, 10 mM $NaH_2PO_4.H_2O$, and 1 mM Na_2EDTA, pH 7.4), 1% (w/v) SDS)], then at high stringency in wash buffer III (Same as wash buffer II except (0.5×)), and one last wash in wash buffer II in order to remove nonspecific bindings to reduce background without effecting the specific probe hybridization. Afterwards, the membranes were blotted dry and wrapped in a sheet protector, and autoradiographed with two intensifying screens (10" × 12" L-Plus, Optonix, NJ) and X-ray film (Blue, Medical X-ray film, SourceOne, CA) for 10 days at 80 ºC before developing.

For hybridization of the chloroplast clones, BAC membranes were prepared the same way as overgo and Southern hybridizations. Before addition of radioactively labeled probes, the membranes were pre-hybridized within 85 ml of hybridization buffer [0.5 M Na_2HPO_4, 1 mM EDTA, and 7% SDS], for four hours at 62 ºC, and the hybridization was performed at the same temperature for another 20 hours. After hybridization, the membranes were washed twice at 62 ºC with washing solution [0.5×SSPE (0.15 M NaCl, 10 mM $NaH_2PO_4.H_2O$, and 1 mM Na_2EDTA, pH 7.4), 0.5% SDS] for 15 minutes each. Subsequently, the membranes were autoradiographed as described above.

RESULTS:

A. Construction and Characterization of the Peanut BAC Library

a) HMW DNA isolation

HMW DNA quality is a primary prerequisite for a successful ligation. The DNA needed to be larger than 1000 kb without too much drag. A typical HMW DNA with varying concentrations is denoted in Fig 2.2.

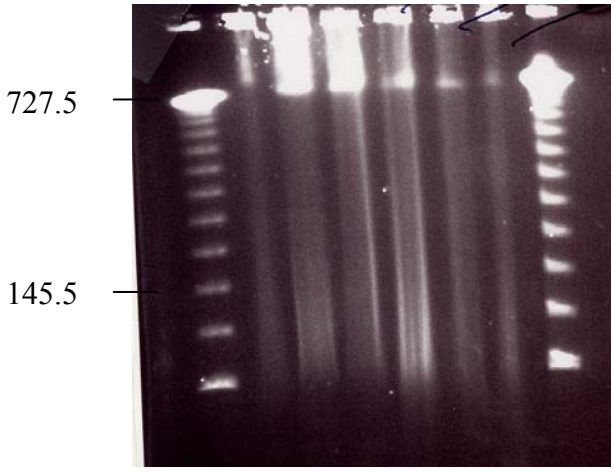


Figure 2.2 The first and last lanes are Lambda ladder PFG markers. The other lanes contain peanut HMW DNA with diminishing concentrations.

b) Size Selections:

The number of plugs used for the size selection of suitable fragments for ligations was adjusted according to the quality and the DNA concentration. Even though three different equal fractions from the first size selection gel (Fig 2.3A), in the range of 125 kb to 350 kb, were subjected to the second size selection for further purging of the smaller fragments, the best results were always obtained from the middle fraction, which corresponds to the fragments

between ~160-200 kb. The lowest fraction usually produced insert size that was too small, while the top fraction yielded too much variation in insert size (Fig 2.3B).

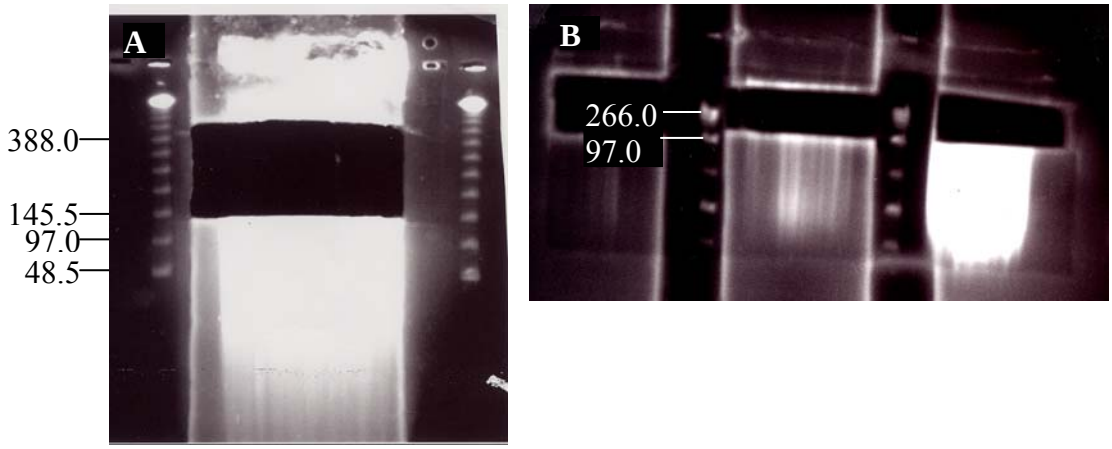


Figure 2.3 Size range of DNA in the excised gel fragment after the first size selection (**A**) and second size selection (**B**). The size markers (kb) are indicated

c) Ligation:

For most of the ligation reactions ¼ V/I ratio produced the best average insert size and the best ligation efficiency, but we adjusted this ratio for each size selection.

d) BAC Library of *Arachis hypogaea*:

We used 16 ligation reactions for construction of the peanut BAC library. Efficiency was variable from ligation to ligation (Table 2.1), due to the intactness and cleanliness of HMW DNA used. The proportion of colonies without insert was also significantly different between ligations, ranging from 0 to 14%. Overall, we estimated that 5484 clones (3%) of 182,784 total clones picked, had no inserts. This percentage is typical of lacZ-gene based cloning systems (Li et al. 2003)

Three kinds of vector were used, pBeloBAC, pCUGIBAC1 and pIndigoBAC-5 (Table 2.1).The host cells of 97, 152 clones were T1 phage resistant. The best efficiency was obtained from pIndigoBAC-5 in terms of blue/white ratio and number of colonies.

Assuming that the genome size of *Arachis hypogaea* is about 2813 Mb (Arumuganathan et al. 1991, Singh et al. 1996 and Temsch et al. 2000), the library provides about 6.5 genome equivalent coverage resulting in a 99.88% probability of recovering any single-copy sequence from the genome.

DNA from 206 randomly sampled colonies were digested with NotI and resolved by PFGE; a sample of 28 of these clones are shown (Fig 2.4). Only 8 out of 206 colonies contained less than 20 kb insert or had no insert (Fig 2.5). The insert sizes of about 50% of colonies were between 100 and 110, and less then 2% of colonies contained inserts of 160 kb or larger (Fig. 2.5).

e) Assessment of the depth of genome coverage by hybridization:

A sample of 24 overgos (40 bp oligonucleotides) derived from genetically mapped peanut RFLP probe sequences (Burow et al. 2001) were applied to the peanut BAC library to assess the depth of the genome coverage. Overgos that are likely to be present as a single locus in the tetraploid peanut genome were chosen on the basis of number of *Hin*dIII bands (i.e. 2 or 3 bands for tetraploid genome) on survey blots (Burow et al. 2001). A total of 655 hits was identified by the 24 overgos (Table 2.2). The number of hits obtained from overgos POVS1131 (41 hits), POVS1161 (42 hits), POVS1129 (106 hits), and POVS1265 (161 hits) was significantly different than the expectation at the presumed depths of genome coverage (p value=0.001) suggesting possible multiple locus recognition by these overgos. Thus, these overgos were not considered for further calculations. The average number of hits from the remaining 20 overgos was 15.25

(±7.49), which suggests 7.6 genome equivalent coverage of the peanut genome. The range of number of hits observed, from 5-28, is consistent with a range of library coverage from 6 to 9.5 genome equivalence based on a Poisson distribution, and likelihood threshold of 95%.

Table 2.1 Specificities of peanut BAC library derived from randomly picked colonies

Ligation number	Vector	Total no. of colonies picked	No of 0 (kb)[a] inserts/tested (%)	Average insert size[b]	Total Mb cloned
1	pBeloBAC	3,072	0/26 (0%)	95.20	292.45
2	pBeloBAC	4,224	0/8 (0%)	106.25	448.80
3	pBeloBAC	3,456	1/20 (5%)	94.20	309.28
4	pBeloBAC	2,688	0/13 (0%)	90.23	242.54
5	pBeloBAC	8,832	0/13 (0%)	105.30	930.01
6	pCUGIBAC	2,688	3/21 (14%)	96.00	221.18
7	pIndigoBAC-5	18,048	1/42 (2%)	108.46	1910.96
8	pIndigoBAC-5	14,592	0/9 (0%)	106.35	1551.86
9	pCUGIBAC	8,832	0/10 (0%)	100.00	883.20
10	pIndigoBAC-5	14,208	0/14 (0%)	100.05	1421.51
11	pIndigoBAC-5	4,992	0/10 (0%)	99.78	498.10
12[e]	pIndigoBAC-5	36,864	0/21 (0%)	111.90	4125.08
13[e]	pIndigoBAC-5	16,128	0/8 (0%)	114.37	1844.56
14[e]	pIndigoBAC-5	13,056	0/7 (0%)	90.33	1179.35
15[e]	pIndigoBAC-5	17,664	1/7 (14%)	90.33	1367.65
16[e]	pIndigoBAC-5	13,440	1/7 (14%)	106.10	1222.30
Total		182,784	5484.07[c]	104.05[d]	18448.83

[a] False positives (i.e. white colonies without inserts)
[b] Colonies without insert were included in calculation of the average insert size
[c] The total number of colonies without insert = $\Sigma_{i=1 \text{ to } 16}$ (((%) of 0 (kb) inserts × (number of colonies picked)), where i=ligation number
[d] Grand average=($\sum_{i=1}$ to 16 (average insert size × (total no. of colonies picked)$_i$))/total number of colonies)
[e] T1 phage resistant host cells were used for these ligations

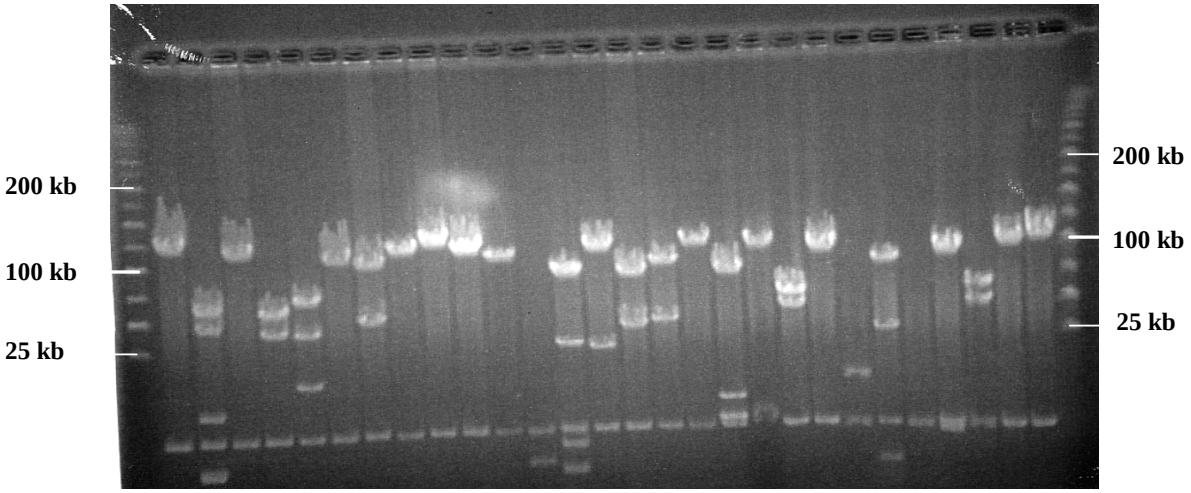


Fig. 2.4 Characterization of the peanut BAC library. DNA was extracted from randomly selected clones and digested with *NotI*. The first and last lanes are MidRange II PFG marker (NEB).

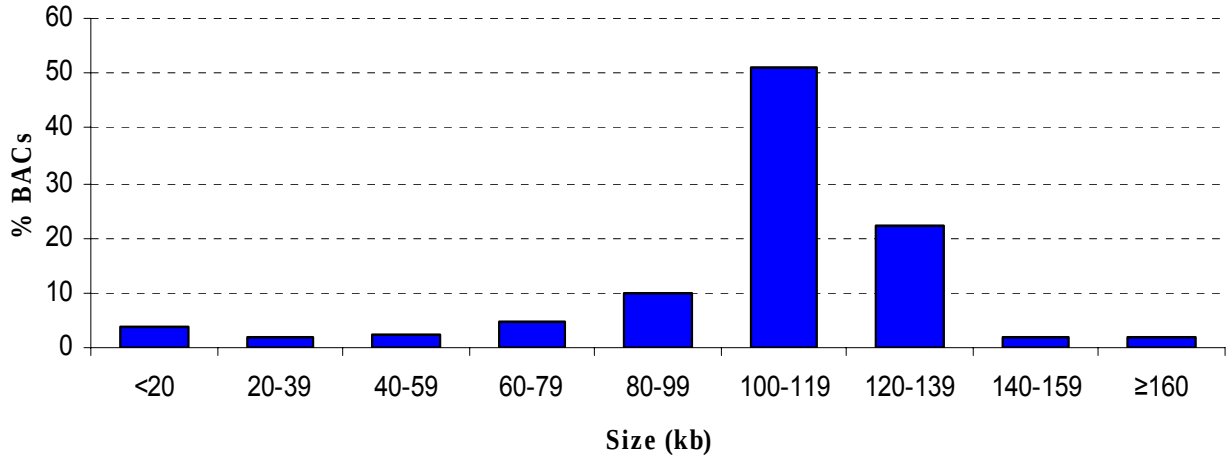


Fig 2.5 Randomly selected BAC clones from each ligation were digested and the percentage of colonies in different size ranges was calculated.

Table 2.2 The results of the peanut BAC library screening with different types of probes

Overgo Id	No. of Hits	No. of *Hin*dIII Bands[a]
POVS1170	2	3
POVS1279	5	2
POVS1091	7	3
POVS1098	9	3
POVS1114	9	2
POVS1130	10	3
POVS1090	11	2
POVS1186	11	3
POVS1066	13	2
POVS1159	13	2
POVS1178	14	2
POVS1105	18	3
POVS1108	18	3
POVS1071	19	2
POVS1034	21	3
POVS1027	22	2
POVS1136	23	2
POVS1002	25	3
POVS1053	27	2
POVS1067	28	3
POVS1131	41	3
POVS1161	42	2
POVS1129	106	3
POVS1265	161	3
SOG6231[b]	9	-
SOG2761[b]	0	-
Garb6VRC12[c]	0	-
Garb11VRA06[c]	0	-
Garb19VRC09[c]	0	-
Garb29VRG08[c]	0	-
SOG6768[d]	866	-
SOG6770[d]	1123	-

[a] The bands were obtained from surveying blots, where *Hin*dIII digested peanut genomic DNA were hybridized to the probes that were used for derivation of the overgos (Burow et al. 2001)
[b] These overgos are specific to sorghum chloroplast.
[c] These probes derived from *G. arboreum* chloroplast sequence
[d] These overgos are specific to sorghum rDNA (45S)

Four clones (300-400 bp long sequences) were included from different segments of *Gossypium arboreum* chloroplast (Garb6VRC12, Garb11VRA06, Garb19VRC09, Garb29VRG08) with high homology to *Fabaceae* chloroplast sequences to ensure the completeness of the screening (Table 2.2). In addition, two overgos specific to the sorghum chloroplast (SOG6231, and SOG2761) were used to screen the peanut high-density BAC filters to elucidate the level of chloroplast DNA contamination in the library. Hybridization of neither cotton chloroplast based probes nor the sorghum chloroplast overgo, SOG2761, yielded any hits, but one overgo, SOG6231 had nine hits. This could mean that the percentage of chloroplast clones is very low. To rule out possible problems during the labeling stage of the cotton probes, we counted the total incorporated $^{32}\alpha$-P, and it was on average 2.3×10^{-6} dpm, which is more than enough to produce signal on BAC filters based on prior experience.

Two sorghum overgos with good homology to *Fabaceae* 45S rDNA sequences were employed to assess the copy number and organization of rDNA clusters in peanut. One of the overgos, SOG6768, identified 866 BAC clones and the other one, SOG6770, had 1123 hits (Table2.2). Since both of them targeted the same cluster, the hits overlapped, and the total number of unique hits was 1208. SOG6768 detected fewer loci possibly due to its homology to ITS (Internal Transcribed Sequence), which is comparatively less conserved than transcribed rDNA sequences (Nickrent and Doyle 1995). Nevertheless, about 0.66% of BAC clones contain rDNA clusters, which means that at least 0.1% of the peanut genome is composed of rDNA clusters assuming 6.5 genome equivalence of the peanut BAC library.

B. Determination of Subgenomic Specificity of Individual BACs:

To test the practicality of a hybridization based approach for peanut physical mapping, we selected 94 BAC clones, which were identified by seven different probes. These BAC clones

were first subjected to fingerprinting (Fig 2.7), and contig analyses were performed with FPC V4.7. The results of the contig analysis are shown in Table 2.3. However, grouping of BACs into contigs does not anchor contigs to genetic maps. To resolve this, the fingerprinting gel was blotted and labeled with the same overgos. The expectation was that the BAC clones belonging to different map locations would produce different banding patterns when labeled with respective overgos, which can be directly compared with RFLP patterns from the genetic map. This would assist in integration of physical and genetic maps making subgenomic chromosome walking relatively plausible. The results of the experiment are shown in Fig 2.6. For example, positive BAC clones for POVR2609 produced two distinctive banding patterns suggesting different subgenomic localization, and the contig data from the fingerprinting gel confirmed the dual characteristics of the clones (Table 2.3). The contig data for overgo POVR2080, was almost 100 percent in agreement with the observed banding pattern; the only exceptional clone had no fingerprint data. Likewise, another overgo, POVR2100, produced very similar results. However, some of the clones in POVR2100 showed more than one band on the film (at least 4) suggesting a possible partial digestion. For overgo POVR2029 only 50% of the banding patterns agreed with the contig information; however; it needs to be noted that the resolution of the film may not be sufficient to discern very closely spaced bands. Some of the other disagreements might have also been caused by an insufficient resolution on the fingerprinting gel (Fig 2.7). Overall, the banding pattern on the Southern blot was usually concordant with the contig association of the clones (71% overall average). We believe that with a better resolution of samples, the technique could be further improved.

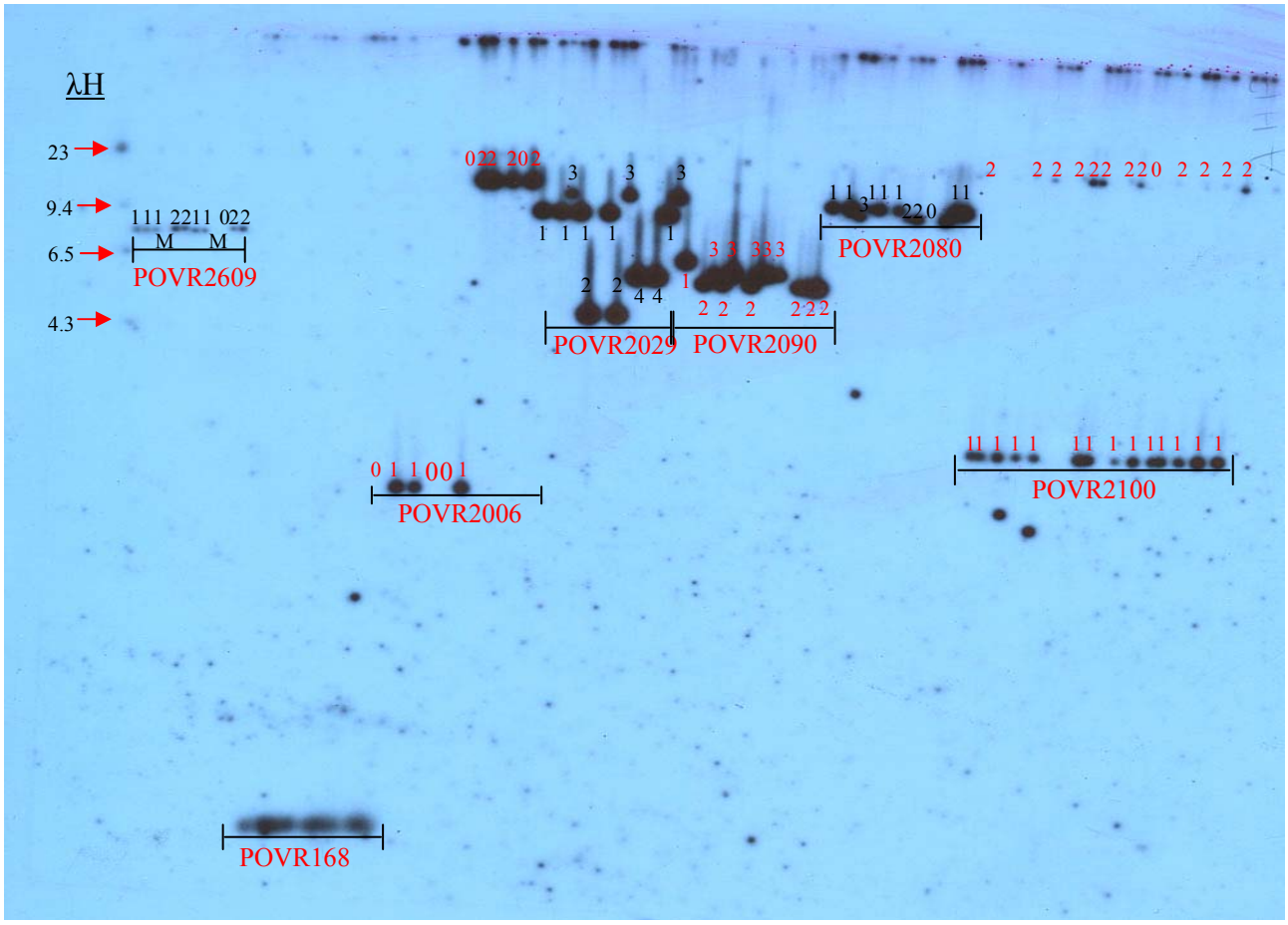


Figure 2.6 BAC-RF (Lin et al. 2000) on single BAC clones hybridizing to low-copy probes. Starting from the first lane every fifth lane is markers (M). The second lane is λHIII, where the sizes of the fragments are indicated (kb) The number of clones for the overgos, POVR2609, POVR168, POVR2006, POVR2029, POVR2090, POVR2100 and POVR2100 are 10, 12, 12, 12, 12, 12, and 24, respectively. The gel was hybridized with corresponding overgo primers. The polymorphic bands on the autoradiograph depict possible contigs belonging to different loci. The approximate positions of positive clones for each probe are indicated. The numbers (i.e. 1, 2, 3, and 4) denote possible polymorphic bands between contigs within the same probe and "0" indicates the clones that were recognized as positives at screening the BAC library, but failed to

produce hybridization in the southern blot. The numbers for the clones specific to the consecutive overgos were depicted in alternating black and red colors.

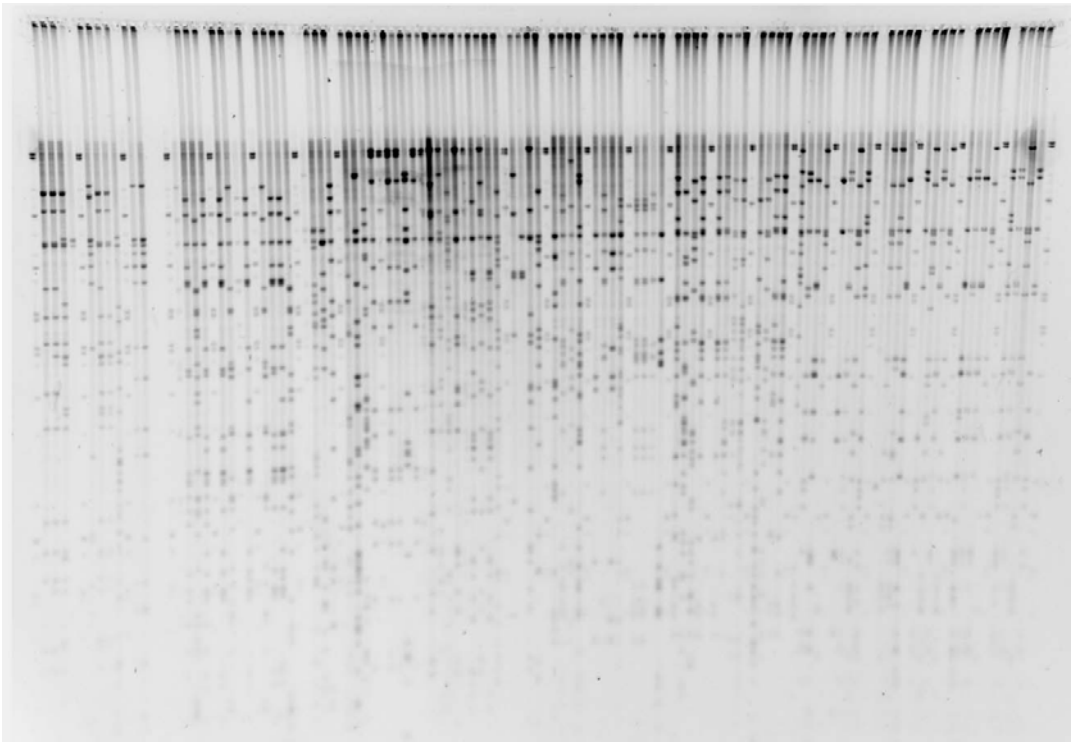


Figure 2.7 *Hin*dIII digested BACs from peanut. Starting from the first lane, every fifth lane is markers. The first 10 samples are for overgo POVR2609, and next 60 samples are respectively for POVR168, POVR2006, POVR2029, POVR2090, and POVR2080 overgos, where each probe has 12 positives. The last 24 sample lanes are positive clones for POVR2100.

It can be concluded that this technique can be employed for large scale applications. Perhaps the biggest advantage of this approach is that it allows the hybridization of multiple probes in a single hybridization experiment. The subgenomic identity of contigs could be determined with this method rather efficiently.

Table 2.3 The listing of the contigs per overgo and their relationship with the polymorphic bands on the autoradiograph (Fig 2.6)

Probe	Fragment[a] size	No. of clones	No of clones that form a contig	False positives[b]	Agreement(%)[c]
POVR2609	1	5	5	1	89
	2	4	3		
POVR2100	1	14	14	1	88
	2	[d]9	9		
POVR2006	1	3	0	5	57
	2	4	4		
POVR2029	1	5	2	0	50
	2	2	2		
	3	3	2		
	4	2	0		
POVR2090	1	1	0	0	50
	2	6	4		
	3	5	2		
POVR2080	1	3	2	0	92
	2	7	7		
	3	2	2		
POVR168	-	-		-	-

[a] The polymorphic fragments within the probe are shown on **Fig 2.6**
[b] Normal clones that were recognized as positives in BAC library screening, but failed to show hybridization in the southern blot.
[c] The overall percentage of the clones that were contigged within the overgo
[d] Some fragments were not clearly discernible, perhaps due to partial digestion
- Not able to discern the different band sizes, where three different contigs were formed.

DISCUSSION:

We describe the construction and characterization of the first BAC library for any *Arachis spp*. This BAC library will considerably advance genomic studies in peanut. The amount of coverage, 6.5 times, will make possible the isolation of clones containing virtually any gene of interest. In the near future, map-based cloning may become a reality in peanut. Because of its low level of diversity at the DNA level, the identification and characterization of economically significant genes is urgent for the future of peanut cultivation. This BAC library will also be a very useful tool for comparative evolutionary analysis among different legume species.

Optimization of the enzyme concentration and digestion time was pivotal for the maximization of the numbers of DNA fragments that contain cohesive ends in the desired size range. The cleanliness of the embedded DNA in the plugs was one of the major factors determining the success of a partial digestion reaction. In terms of the efficiency of digestion, we usually obtained more promising results by increasing the amount of enzyme rather than the duration of digestion

Adjusting the DNA concentration during size fractionation was very critical. Too much DNA caused wide variation in insert size, probably due to poor resolution during gel electrophoresis. Therefore; the higher the DNA concentration, the higher the number of small fragments 'stuck' in the desired size range. Even a second round of size selection is not enough to completely purge the small fragments, since the small fragments tend to ligate better. At high DNA concentration, uniformity and average size of inserts dropped dramatically. On the other hand, low DNA concentration usually makes it impossible to obtain a sufficient amount of DNA for ligation. Therefore, it is necessary to optimize the amount of DNA used in size selection. Another major problem we observed was that if the initial DNA is too degraded, size selections usually failed to produce DNA fragments of sufficient quality to ligate. We could not get any

successful ligation from inserts isolated by β-agarase method, perhaps because large DNA fragments may be damaged during the process (Strong et al. 1997, Osoegawa et al. 1998).

Although the DNA fragments used for ligations were selected to be in the range of 150 to 190 kb, the final average insert size was much lower, 104 kb. Similar results were observed in the construction of other BAC libraries (Danesh et al. 1998; Meksem et al. 2000). This result is consistent with the suggestions made by Frijters et al.(1997); that is, the resolution of the agarose gels is not sufficient to remove all the 'trapped' smaller fragments resulting in much smaller average insert size than expected. A more effective way of purging the trapped fragments might significantly improve the average insert size.

The efficiency of the ligation reactions was highly variable depending on the quality of the source DNA and efficiency of the size selections. More than 10 hr of ligation at 16 ºC usually produced more false positives (clones with no insert and rearranged vector). The proportion of false positives in BAC libraries has been a problem, with up to 17% reported (Wang et al. 2001b). In our library, only 3% of the clones were false positives. The vector/insert ratio had to be adjusted for each size selection, suggesting that the quality (i.e. the ratio of fragments with the damaged ends) was variable among the insert DNA samples used for library construction.

The depth of coverage was confirmed by screening of the library with 24 probes (Table2.2). The average number of hits was 15.25, which is somewhat more than the expected number of hits per probe, for estimated 6.5 genome equivalence. The inflated number of hits could mean that some of these overgos might have targeted multiple loci. This is not unusual; especially for overgos derived from evolutionarily conserved domains. Nevertheless, the hybridization results approximately confirm the presumed coverage depth of the library.

Typically around 1% of most BAC libraries are chloroplast clones (Ming et al. 2001; Wu et al. 2004). The level of chloroplast DNA contamination in our library was very low. The low contamination percentage might be a result of the technique utilized for the isolation of HMW DNA; where nuclei were precipitated at very low centrifugal force, 160 g, possibly preventing precipitation of any organelle.

rDNA sequences are very well conserved across distantly related plant taxa, even though the length and composition of the clusters vary greatly (Shi et al. 1996). By relying on this assumption, we used two sorghum 45S rDNA sequence derived overgos for screening of the peanut BAC library. The total number of BAC clones recognized by the two overgos was 1208. Raina et al.(1999) has detected two major clusters that are associated with NORs of the chromosomes, and six inactive condensed rDNA sites in the *A. hypogaea* genome. Besides the number of ribosomal RNA cluster sites, the number of *Hin*dIII sites within the repeat would affect the ratio of BAC clones carrying rDNA repeats. A segment of the 45S rDNA repeat unit in peanut has been cloned and sequenced (Bhagwat et al. 2001); there was a single *Hin*dIII restriction site within the 12 kb rDNA unit; thus, the frequency of *Hin*dIII sites is less than the expected average of one per 4 kb for a six-cutter. In conclusion, the length of the clusters in peanut could be an explanation for the anomaly in the observed number of hits. Alternatively, one of these overgos, SOG 6770, could have detected other loci besides the rDNA clusters. Nevertheless, the number of hits roughly verifies previous assessments about the depth of library coverage.

Like many agriculturally important plants, cultivated peanut is polyploid. A complication in physical mapping of polyploids is ambiguities associated with anchoring contigs to genetic loci, due to presence of multiple loci that hybridize to most probes. Several strategies have been

devised to deal with this issue; however, approaches based on hybridization have been the most promising for high throughput applications. We investigated possible high-throughput techniques for the assignment of subgenomic specificity of the probes with multiple hits in the peanut genome. Although very little genetic variation exist among the *Arachis hypogaea* cultivars, the A and B subgenomes usually different with regard to most loci, increasing the chance of success for subgenomic assignment on the basis of polymorphism. The technique tested in this paper was rather easy and would allow testing of multiple contigs in a single experiment. The overall success rate, 71%, was acceptable, and this rate could be improved with better resolution. Hence, the integration of physical and genetic maps for polyploid species such as *A. hypogaea* becomes more manageable.

Acknowledgments: Requests for the BAC library (or subsets) should be directed to Dr. Paterson (paterson@uga.edu).

REFERENCES:

Anderson C (1993) Genome shortcut leads to problems. Science 259: 1684-1687

Bhagwat AS, Krishna TG, Jawali N, Mitra RM (2001) Cloning and characterization of a ribosomal RNA gene repeat unit from groundnut. Plant Cell Rep. 20: 193-197

Burke D, Carle G, Olson M (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-812

Burow 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

Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003) Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. durum). Theor App Genet 107: 931-939

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) RFLP mapping of a cross between *Sorghum bicolor* and *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of Sorghum chromosomes. Theor App Genet 87: 925-933

Choi S, Creelman RA, Mullet JE, Wing RA (1995) Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis-thaliana*. Plant Mol Biol Rep 13: 124-128

Danesh D, Penuela S, Mudge J, Denny RL, Nordstrom H, Martinez JP, Young ND (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor App Genet 96: 196-202

Dunford R, Vilageliu L, Moore G (1993) Stabilization of a yeast artificial chromosome containing plant DNA using recombination-deficient host. Plant Mol Biol 21: 1187-1189

Fernandez A, Krapovickas A (1994) Cromasomas y evolucion en *Arachis* (Leguminosae). Bonplandia 8: 187-220

Food and Agriculture Organization (2003) FAOSTAT Agriculture. Available at: http://apps.fao.org/page/collections?subset=agriculture; verified 10 February 2004.

Frijters ACJ, Zhang Z, van Damme M, Wang GL, Ronald PC, Michelmore RW (1997) Construction of a bacterial artificial chromosome library containing large Eco RI and *Hin*dIII genomic fragments of lettuce. Theor App Genet 94: 390-399

Gentzbittel L, Abbott A, Galaud JP, Georgi L, Fabre F, Liboz T, Alibert G (2002) A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. Mol Gen Genet 266: 979-987

Larionov V, Kouprina N, Nikolaishvili N, Resnick M (1994) Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YACs). Nucl. Acids. Res. 22: 4154-4162

Li L, Lu S, O'Halloran DM, Garvin DF, Vrebalov J (2003) High-resolution genetics and physical mapping of the cauliflower high-beta-carotene gene *Or* (*Orange*). Mol Gen Genet 270: 132-138

Lin Y, Draye X, Qian X, Ren S, Zhu L, Tomkins J, Wing R, Li Z, Paterson A (2000) Locus-specific contig assembly in highly-duplicated genomes using the BAC-RF method. Nucl. Acids. Res. 28: e23

Lin Y-R, Zhu L, Ren S, Yang J, Schertz KF, Paterson AH (1999) A Sorghum propinquum BAC library, suitable for cloning genes associated with loss-of-function mutations during crop domestication. Molecular Breeding 5: 511-520

Meksem K, Zobrist K, Ruben E, Hyten D, Quanzhou T, Zhang H-B, Lightfoot D (2000) Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome wide physical mapping. Theor App Genet 101: 747-755

Ming R, Moore PH, Zee F, Abbey CA, Ma H, Paterson AH (2001) Construction and characterization of a papaya BAC library as a foundation for molecular for molecular dissection of a tree-fruit genome. Theor App Genet 102: 892-899

Nam YW, Penmetsa RV, Endre G, Uribe P, Kim D, Cook DR (1999) Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. Theor App Genet 98: 638-646

Neil DL, Villasante A, Fisher RB, Vetrie C, Cox B, Tyler-Smith C (1990) Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. Nucl. Acids. Res. 18: 1421-1428

Nickrent DL, Doyle JJ (1995) A molecular phylogeny of diploid *Glycine* (Fabaceae) based upon nuclear ribosomal ITS sequences. Am Journal Bot 82: 153

O'Connor M, Peifer M, Bender W (1989) Construction of large DNA segmets in *Escherichia coli*. Science 244: 1307-1312

Raina SN, Mukai Y (1999) Detection of a variable number of 18S-5.8S-26S and 5S ribosomal DNA loci by fluorescent in situ hybridization in diploid and tetraploid Arachis species. Gemome 42: 52-59

Ross MT, LaBrie T, McPherson J, Stanton VM (1999) Screening large-insert libraries by hybridization. Wiley, New York, NY

Salimath SS, Bhattacharyya MK (1999) Generation of a soybean BAC library, and identification of DNA sequences tightly linked to the *Rps-1* disease resistance gene. Theor App Genet 98: 712-720

Sambrook J, Russell DW (2001) Molecular Cloning: A laboratory manual. vol 1-3. Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York

Schmidt R, Putterill J, West J, Cnops G, Robson F, Coupland G, Dean C (1996) Detailed description of four YAC contigs representing 17 Mb of chromosome 4 of *Arabidopsis thaliana* ecotype Columbia. Plant J. 9: 755-760

Shi L, Zhu T, Keim P (1996) Ribosomal RNA genes in soybean and common bean: chromosomal organization, expression, and evolution. Theor App Genet 93: 136-141

Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. 10: 1772-1787

Stalker H (1991) A New Species in section *Arachis* of peanuts with a D genome. Am Journal Bot 78: 630-637

Stalker H, Dhesi J, Parry D, Hahn J (1991) Cytological and interfertility relationships of *Arachis* Section *Arachis*. Am Journal Bot 78: 238-246

Sulston J, Mallett F, Durvin R, Horsnell T (1989) Image analysis of restriction enzyme fingerprint audioradiograms. Comput. Appl. Biosci. 5: 101-106

Tomkins JP, Mahalingam R, Smith H, Goicoechea JL, Knap HT, Wing RA (1999) A bacterial artificial chromosome library for soybean PI 437654 and identification of clones associated with *cyst* nematode resistance. Plant Mol Biol 41: 25-32

Vanhouten W, MacKenzie S (1999) Construction and characterization of a common bean bacterial artificial chromosome library. Plant Mol Biol 40: 977-983

Vilarinhos AD, Priffanelli P, Lagoda P, Thibivilliers S, Sabau X, Carreel F, D'Hont AD (2003) Construction and characterization of a bacterial artificial chromosome library of banana (*Musa acuminata* Colla). Theor App Genet 106: 1102-1106

Wang Q, Zhang K, Qu X, Jia J, Shi J, Jin D (2001a) Construction and characterization of a bacterial artificial chromosome library of peach. Theor App Genet 103: 1174-1179

Wang W, Zhai W, Luo M, Jiang G, Chen X, Li X, Wing RA, Zhu L (2001b) Chromosome landing at the bacterial blight resistance gene *Xa4* locus using a deep coverage rice BAC library. Mol Gen Genet 265: 118-125

Willets N, Skurray R (1987) Structure and function of F-factor and mechanism of conjugation. In:. In: FC N (ed) *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol 2. Am Soc Microbiol, Washington D.C, pp 110-1137

Woo S-S, Jiang J, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucl. Acids. Res. 22: 4922-4931

Wu CC, Nimmakayala P, Santos FA, Springman R, Scheuring C, Meksem K, Lightfoot DA, Zhang HB (2004) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor App Genet online

CHAPTER 3

ORGANIZATION AND EVOLUTION OF RESISTANCE GENE ANALOGS IN PEANUT[1]

[1] Yuksel, B. and A.H. Paterson. To be submitted to *Molecular Genetics and Genomics*.

ABSTRACT:

The scarcity of genetic polymorphism in *Arachis hypogaea* (peanut), like other monophyletic polyploid species, makes it especially vulnerable to nematode, bacterial, fungal, and viral pathogens. Although no disease resistance genes have been cloned from peanut itself, the conserved motifs in cloned resistance genes from other plants provide a means to isolate and analyze similar genes from peanut. To survey the number, diversity, evolutionary history, and genomic organization of resistance gene-like sequences in *Arachis hypogaea*, we isolated 234 resistance gene analogs from the peanut genome by using primers designed from conserved regions of different classes of resistance genes including NBS-LRR, and LRR-TM classes. Phylogenetic and sequence analyses of the RGA sequences were done to explore evolutionary relationships both among peanut RGAs and with orthologous genes from other plant taxa. Fifty-six overgos designed from the RGA sequences on the basis of their phyletic association were applied to the peanut BAC library; 736 hybridizing BAC clones were fingerprinted and contigs were formed in order to gain insights into the genomic organization of these genes. All the fingerprinting gels were blotted and screened with the respective overgos in order to verify the authenticity of the hits from initial screens and to explore the physical organization in terms of both copy number and spatial distribution. As a result, we identified 250 putative resistance gene loci. We found correlation between the phyletic positions of the sequences and their physical locations. The BACs we isolated will serve as a valuable resource for potential future applications such as map-based cloning and better understanding evolution and organization of these genes in the peanut genome.

Keywords: *Arachis hypogaea*, NBS-LRR, physical mapping, TM-LRR, RGA, BAC, Disease, Cf-like

INTRODUCTION:

In monophyletic allopolyploid species such as cultivated peanut, *Arachis hypogaea*, genetic variation is often very limited. Scarcity of genetic polymorphism makes peanut especially vulnerable to a wide variety of pathogens. For instance, the peanut cultivars grown in the Southern USA are highly susceptible to *Meloidogyne arenaria*, and damage in highly infested fields can be devastating. Yield losses are estimated on average to be about 30% (Nelson et al. 1989). Peanut is also vulnerable to several viral, fungal, and bacterial pathogens (Branch and Brenneman 1999; Pensuk et al. 2004; Phipps and Porter 1998). Gaining insight into the evolution and genomic organization of peanut resistance-gene-related sequences will be a start toward lessening the threats of these pathogens for peanut cultivation.

The majority of plant resistance genes belong to the NBS-LRR class, all of which share about a 400 aa long characteristic NBS (Nucleotide-Binding Site) found in ATP or GTP binding proteins, that function by creating a pocket for specific binding of these nucleotides (Saraste et al. 1990; Traut 1994). The C-terminus of these genes is composed of varying numbers of imperfect LRRs (leucine-rich-repeats). The NBS-LRR containing genes are further subdivided into two categories according to conserved domains at the N-terminus: 1) Genes containing a leucine-zip like conserved pattern (NonTIR), for example; the tomato root-knot nematode resistance gene *Mi* (Milligan et al. 1998), *Pseudomonas syringae* resistance genes *RPS2* and *RPM1* in *Arabidopsis* (Bent et al. 1994; Grant et al. 1995), and *Fusarium oxysporum f sp lycopersici* resistance gene *I2* (Simons et al. 1998) and 2) Genes containing a homologous structure to human interleukin-1 and Drosophila toll-like receptor regions (TIR), for instance; tobacco mosaic virus resistance gene *N* (Whitham et al. 1996) and flax rust resistance genes *L6* and *M* (Anderson et al. 1997; Lawrence et al. 1995). Besides NBS-LRR genes, other resistance

gene classes also exist. Genes like *Cf2*, *Cf4*, *Cf5*, and *Cf9* rendering resistance to fungal pathogen *Cladosporium fulvum* are mainly composed of extracytoplasmic LRRs and a short TM (Transmembrane) domain. Furthermore, *pto*, the first cloned resistance gene, contains only a cytoplasmic Ser/Thr kinase domain (Martin et al. 1993). Finally, resistance genes such as *Xa21* from rice are receptor-like kinases with their extracellular LRRs, transmembrane domains, and cytoplasmic kinase domains (Song et al. 1995).

The commonality of functionally significant subdomains such as the P-loop, kinase-2, kinase3a, GLPL, and RNBS-D in the NBS region of cloned resistance genes has inspired the design of degenerate primers for amplification of the same domain of similar genes (resistance gene analogs: RGAs) from other plants. For example, large numbers of RGAs from soybean have been amplified with degenerate primers developed from the NBS (Kanazin et al. 1996; Yu et al. 1996). Similarly, RGAs from maize (Collins et al. 1998), lettuce (Meyers and Michelmore 1998), chickpea (Huettel et al. 2002), and *Arachis spp*. (Bertioli et al. 2003) are also developed. The distinguishing amino acid residues in conserved blocks of NBS regions of NonTIR and TIR subgroups allowed a targeted amplification of the NonTIR class (Penuela et al. 2002). For example, a 5 aa long motif in the RNBS-A domain of two classes is clearly distinct between the two groups. While the FXXXF is common in CLN, FXXXXW is common in the TNL group. Likewise, a kinase-2 domain contains a distinctive pattern too; the common LDDV pattern is usually followed by a D (aspartate) in CNL, but the same residue is usually substituted by a W (tryptophan) in the other class. A similar distinctive pattern in downstream conserved blocks such as RNBS-C and RNBS-D are also observed (Cannon et al. 2002; Pan et al. 2000). By exploiting these subgroup-specific amino acid residues, Penuela et al. (2002) were able to isolate 50 RGAs with NonTIR specific primers from the soybean genome, which were homologous

cloned NonTIR genes. Targeted isolation of the NonTIR class is required due to their rarity in dicot genomes, for instance; less than one-third of NBS-LRR genes belonged to the NonTIR class.

Degenerate primers have also been employed in amplification of other classes of resistance gene-like sequences such as ser/thr kinase-like genes (Di Gaspero and Cipriani 2003; Vallad et al. 2001), LRR domains of *Cf2* and *Cf9* genes (Ohmori et al. 1998), and the wheat *Cre3* sequence (Seah et al. 1998). For a comprehensive analysis of resistance gene-related sequences in a genome, employment of several of these primers would be required.

The main objectives of this study were to explore the full diversity of resistance gene-related sequences in the peanut genome, also assessing the evolutionary relationships both among the peanut sequences and with similar sequences from other taxa. In addition, to investigate the genomic organization of resistance-gene-related sequences, we isolated and characterized BACs containing these sequences by using overgo primers specific to different sets of sequences from distinct cladistic groups of RGAs.

MATERIALS AND METHODS:

a) Plant material:

Arachis hypogaea L. cv. Florunner UF-439-16-1003-2 seeds were grown on germination mix in the greenhouse. The collected leaf tissue from seedlings was washed in ddH_2O, and frozen in liquid N_2 in 3-4 g packages, then stored at -80 ºC until use.

b) DNA extraction:

For extraction, a standard protocol developed by Burow et al. (2001) was followed. Briefly, 3-4 g of the frozen plant material was ground in liquid nitrogen with a mortar and pestle until fine powder was obtained , and the powdered sample was dissolved in 20 ml of extraction buffer:

[0.005 M citric acid, 0.5 M glucose, 0.01 M Na2EDTA, 2.0% (w/v) polyvinylpyrrolidone-40 (PVP-40), 5% (v/v) Triton X-100, 0,25% (w/v) spermidine, titrated to pH 6.5 with NaOH] to which fresh antioxidants [0.1% (w/v) ascorbic acid, 0.2% (v/v) 2-mercaptoethanol, 0.1% (w/v) disodium diethylthiocarbamate (Na2Et2dtc) and 0.4% (w/v) NaHSO3] were added. After dissolving on ice for about 10 minutes, the samples were filtered through 4 layers of cheese cloth, and the filtrates were centrifuged for 20 min at 3000 × g to sediment nuclei. Subsequently, the pellets were homogenized with a motorized tissue homogenizer at 11,500 rpm for 30 sec in 20 ml of clearing solution [absolute ethanol consisting 0 .5% (w/v) Sodium Dodecyl Sulfate (SDS), 1% (w/v) PVP-40, 0.167 M acetic acid titrated with 4 N NaOH to pH 5.5 and the same antioxidants and in the same proportions as in the extraction buffer with the exception of NaHSO3 freshly added just before use], and were incubated in a 65 ºC water bath for 15 min to break down unbroken cell masses. Following incubation, the nuclei were re-precipitated by centrifuging for 20 min at 3,000 × g, and the precipitates were suspended in 10 ml of lysis buffer [0.050 M citric acid, 0.14 M NaCl, 0.05 M Na2EDTA, 2% (w/v) PVP-40, 2% (v/w) SDS, titrated with 4 N NaOH to PH 5.5, autoclaved, and the antioxidants added freshly] by vortexing before another incubation at 65 ºC for 15 min to lyse nuclei. Following centrifugation at 3000 × g for 15 min, 3.5 ml of 5 M KOAc pH 5.2 was added into the transferred supernatants and the solutions were incubated on ice for 30 min; then, DNA was precipitated with about 7 ml of isopropanol. Finally, DNA was washed with 10 ml of 70% Ethanol and the pellets were air-dried before suspension in TE.

c) Primer sets:

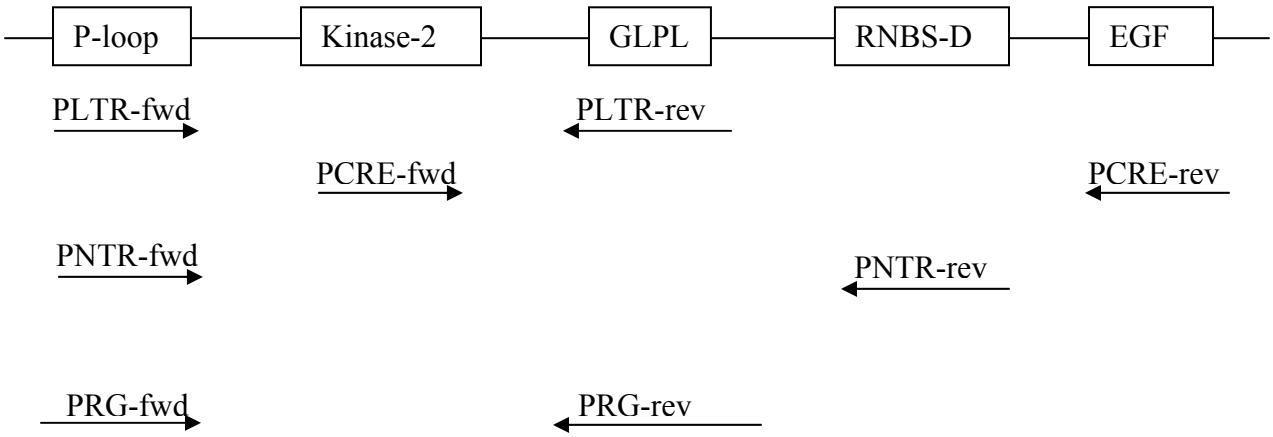


Fig 3.1 Conserved motifs at the NBS (Nucleotide Binding Site) used as the anchorage points for the design of forward and reverse RGA primer sets: PNTR, PLTR, PRG, and PCRE. The drawing is not to scale.

Six different primer sets were used in this experiment; 4 were degenerate and 2 were specific primers (Table 3.1). The priming sites for PNTR, PLTR, PRG and PCRE were based on the conserved motifs of NBS (Nucleotide Binding Site) of resistance gene-like sequences (Fig 3.1). For the amplification of *pto*-like and *Cf*-like genes, priming sites were chosen on the basis of the most conserved sequence motifs in the functional domains of sequenced resistance genes; the kinase domain for *pto* and the LRR domain for *Cf*-like genes (Table 3.1)

Table 3.1 The primer sequences used for the amplification of peanut resistance gene analogs (RGA)

Primer Id	Motif	Motif sequence	Primer sequence (5'-3')
PLTR-fwd[a]	P-loop	GMGGVGKTT	GGNATGGGNGTNGGNAARACNACN
PLTR-rev[a]	GLPL	GLPLALKVLG	NCANCARAANGGNTGNGGNGGGTANGG
PNTR-fwd[b]	P-loop	GGVGKTT	GGNGGNGTNGGNAANACNAC
PNTR-rev[b]	RNBS-D	CFLYCALFP	CGRAANARNSHRCARTANVNRAARC
PCRE-fwd[c]	Kinase-2	LILDDVW	TGATACTGGATGATGTCTGG
PCRE-rev[c]	EGF	EGFIRNT	GTGCTTCTTATGAACCCTTC
PPTO-fwd[d]	*Pto* kinase 1		GCATTGGAACAAGGTGAA
PPTO-rev[d]	*Pto* kinase 2		AGGGGGACCACCACGTAG
PCf-fwd[e]	LRR	SNKLHGPI	WSNAAYAARYTNCAYGGNCCNAT
PCf-rev[e]	LRR	GEIPQQLA	GCNARYTGTCKNGGNATYTCNCC
PRGA-fwd[f]	P-loop	GVGKTT	GGNGGNGTNGGNAANACNAC
PRGA-rev[f]	GLPL	GLPLAL	ARNGCTARNGGNARNCC

[a, b, c, d, e, f] Originally designed by (Aarts et al. 1998), (Penuela et al. 2002), (Lagudah et al. 1998), (Leung et al. 1998), (Ohmori et al. 1998), and (Kanazin et al. 1996) respectively
Degenerate code: N=A, G, C or T, R=A or G, H=A, C or T, S=C or G, V=A, C or G W=A or T, Y=C or T, K=G or T

d) PCR Amplifications:

Fifty ng of *A. hypogaea* genomic DNA were used as template in 50 µl total reaction volume consisting of 3 mM MgCl2, 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.2 mM dTTP, 1µM of each forward and reverse primers (PLTR, PNTR, PPTO, PCf , and PCRE), 1 × PCR buffer and 2.5 U Taq polymerase for all the PCR reactions. The annealing temperatures for each reaction were optimized by using a Gradient Cycler (MJ Research PTC-200) in such a way that the highest annealing temperature with a sufficient amount of product at the expected size range was chosen. The following thermal profiles were used for amplification of the indicated primer sets: PLTR: 94 ºC for 3 min; 35 cycles of 94 ºC for 1 min, 45 ºC for 30 s and 72 ºC for 1.30 min; PCRE: 94 ºC for 3 min; 40 cycles of 94 ºC for 1 min, 55 ºC for 30 sec; PPTO: 94 ºC for 3 min; 38 cycles of 94 ºC for 1 min, 45 ºC for 30 sec, and 72 ºC for 1:50 min; PCf: 94 ºC for 1 min; 35 cycles of 94 ºC for 1 min, 45 ºC for 30 sec, and 72 ºC for 1:45 sec; and PNTR: 94 ºC for 3 min; 35 cycles of 94 ºC for 1 min, 45 ºC for 30 sec, and 72 ºC for 1:50 min. In all cases a 10 minutes extension at 72 ºC was added at the end.

e) Cloning of PCR products:

PCR products were resolved on a 1% agarose gel in 1×TAE buffer for size and quality check. The appropriate size fragments (i.e. ~600 bp for PNTR, ~400 bp for PPTO, ~600 bp for ~PCRE, ~450 bp for PCf, and ~600 bp for PLTR), were excised from the gels; subsequently, the DNA from the excised gel slices was eluted with a GENECLEAN III KIT (BIO 101, Inc.). The quality, quantity, and size of the eluants were checked by running on a gel before cloning into the pGEM-T Easy Vector (Promega, Madison, Wis.). The ligation reactions were performed according to the manufacturer's instructions, and the ligation products were desalted before being transformed into *E. coli* strain, DH10B Electro-Competent cells, by using a GenePulserII (Bio-Rad) with parameters of 5 kV and 200 Ω resistance. After that, the transformed cells were

grown with constant shaking at 250 rpm in 1 ml of SOC (Sambrook and Russell 2001) medium without any antibiotic and subsequently, the cells were plated on LB Xgal/IPGT/Amp solid media and incubated at 37 ºC for 16-18 hours. A total of 192 clones were picked from each ligation and grown in FM (Freezing Medium), which consists of LB, 4.4% (v/v) glycerol and additional salts [36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM Sodium Citrate, 0.4 mM $MgSO_4$ 6.8 mM $(NH4)_2SO_4$]. The clones were sequenced with an Applied Biosystems 3700 by following the ABI BigDye Terminator protocol and the sequences were processed with Phred/Phrap software. Sequences were translated in all six reading-frames with transeq (EMBOSS package) and those found to be in frame (based on the known motifs) were considered for further detailed analysis.

f) Sequence and phylogenetic analysis:

The complete sequences were compared against GenBank sequences with blastx and blastn algorithms (Altschul et al. 1997). Initially, all sequences were aligned within a class using CLUSTAL_X (Thompson et al. 1997). Following that, all sequences containing the same stretch of conserved domains were re-aligned. The alignments were edited with Jalview (http://www.ebi.ac.uk/~michele/jalview/download.html) for redundancy. For the construction of phylogenetic trees, MEGA V2.1 was used (Kumar et al. 2001). The initial trees were generated with the Neighbor-Joining algorithm. The reliability of tree topologies was tested by bootstrapping 1000 times.

g) Physical Distribution Analysis of Peanut Resistance Gene Analogs:

1) Gridding Peanut BAC Library:

High-density membranes were generated from the BAC clones by double spotting 18,432 clones onto 500 cm^2 Hybond N+ membranes (Amersham Life Sciences, Arlington Heights,

USA) with a Q Bot (Genetix). The membranes were grown overnight at 37 ºC on LB CM agar plates until optimal growth was obtained, then fixed according to an alkaline-lysis protocol.

2) Design of PRGOs (Peanut Resistance Gene Overgos):

PRGOVs (Peanut Resistance Gene Overgos) were generated from multiply-aligned PLTR and PNTR sequences. Initially, the sequences were divided into 19 different groups on the basis of their phyletic association. The sequences within each group were multiply–aligned both with each other and with the sequences from the two flanking groups by using CLUSTAL_X (Thompson et al. 1997). Stretches of 40 bp long DNA sequence were chosen from regions containing less than 2 bp difference within the group, but 6 or more bp differences with other groups. The overgo primers, which are composed of 24 bp long forward and reverse primers with an 8 bp overlapping region at the center, were designed from these selected sequences. All the overgos were blasted first against each other to eliminate redundancy, and then blasted against the peanut resistance gene analogs to determine the degree of homology between the overgos and the resistance gene analog sequences. Initially, 54 overgos meeting all the previous criteria were generated, and 47 were chosen for hybridization to high-density *Arachis hypogaea* BAC library filters. Two of these overgos were made from the sequences amplified with the PCRE primer set, using a similar approach.

In addition to these 47 overgos that were designed based on their phyletic relationships, we designed nine more overgos from the maximally conserved regions such as P-loop, kinase2a and kinase3 regions after the multiply aligning of all sequences. Overall, a total of 56 PRGOVs were employed in this experiment.

3) Labeling:

Overgo labeling reactions were set in a total volume of 15 µl containing 0.0067 nM forward and reverse oligonucleotide primers that were denatured at 94 ºC for 5 min and cooled on ice, 1 µg BSA, 2.5 U Taq polymerase, 1 µl α^{32}-dATP (6000 Ci/mmol) (MP Biomedicals, CA), 1 µl α^{32}-dCTP(6000 Ci/mmol) (MP Biomedicals, CA), and 3 µl OLB [oligo labeling buffer without dATP, dCTP, and random hexamers (Ross et al. 1999)]. The reaction mixtures were incubated at 37 ºC for 2 hours (Fig 3.2). The labeled overgos were filtered through Sephadex columns to remove any unincorporated radioactive nucleotides, and heat denatured at 94 ºC for 5 minutes before adding to the proper bottles.

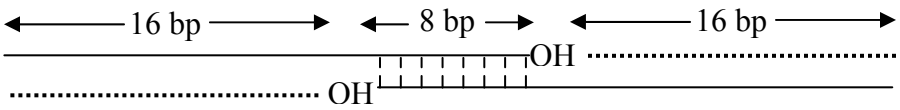


Fig 3.2 Schematic representation of an overgo structure and labeling; the dashed lines indicate the ^{32}P labeled portion of the probe.

4) Hybridization and Washing

Membranes were separated with nylon mesh, placed into rotisserie bottles with 85 ml of hybridization buffer [0.5 M Sodium Phosphate, pH 7.2, 7% (w/v) SDS, 1 mM EDTA, and 0.01% (w/v) BSA] and incubated at 55 ºC for 2 hours before addition of the denatured radioactive overgo probes. Hybridizations were performed at 55 ºC with a rotation speed of 4.5 rpm for 18 to 36 hrs in rotisserie ovens. Following the hybridization process, the membranes were washed at 55 ºC for 30 min each with constant shaking in trays; first at low stringency in wash buffer II

[1×SSPE (0.15 M NaCl, 10 mM $NaH_2PO_4.H_2O$, and 1 mM Na_2EDTA, pH 7.4), 1% (w/v) SDS)], then at high stringency in wash buffer III (same as wash buffer II except 0.5×SSPE), and one last wash in wash buffer II in order to remove nonspecific binding to reduce background with a minimal effect on the site-specific hybridizations. Afterwards, the membranes were blot-dried, wrapped in sheet protectors, and autoradiographed with two intensifying screens (10" × 12" L-Plus, Optonix, NJ) and X-ray film (Blue, Medical X-ray film, SourceOne, CA) for two weeks at -80 ºC before developing.

5) Data Entry and Analysis

The hits on the films were scored manually onto transparent templates, scanned, and read by ABBYY FineReader 5.0 with manual checking and correction when the optical character recognition software was uncertain. The BAC hit scores were converted to the individual BAC clone addresses with BACEater (a Microsoft Basic script, James Estill, UGA).

6) Fingerprinting and Physical Mapping:

A total of 736 BAC clones identified by hybridization with 56 different PRGOVs was re-arrayed in 96-well plates. All the BAC clones were sorted on the well plates according to their overgo specificities. For fingerprinting, the BAC clones were inoculated into 1.5 ml of 2×YT, CM 12.5 µg/ml medium and grown at 37 ºC for 16 hours. For plasmid DNA isolation, a standard alkaline-lysis protocol was used. The DNA samples were digested with 40 U of *Hin*dIII for 4.5 hours. Afterwards, the digestion products were loaded onto a 1% agarose gel and resolved for 16 hours at 95 V. The fingerprint gel images were analyzed with IMAGE (Sulston et al. 1989), and contigs were formed with FPC V 4.7 with a cutoff value of 1×e-10 (Soderlund et al. 2000).

Fingerprinting gels were blotted onto Hybond N+ (Amersham-Pharmacia, Piscataway, NJ) in 0.4 N NaOH. Membranes were pre-hybridized for 4 hours in 45 ml of the same hybridization

buffer as PRGOVs, and the overgos for the corresponding gels labeled as described above. Afterwards, the labeled overgos were added to their respective bottles and hybridizations were performed at 60 ºC for 16 hours. The same washing procedure was applied as explained above, but the washing temperature was 60ºC. The radioactively labeled membranes were exposed to X-ray films in double-screened cassettes for 10 days at -80 ºC.

RESULTS:

A. Amplification and Sequencing of RGAs from *Arachis hypogaea* Genomic DNA:

A total of 1028 amplicons were generated from *Arachis hypogaea* genomic DNA with six different primer sets. The PLTR primer, which is designed from NB-ARC conserved patterns of P-loop and GLPL, was employed in generation of 179 PCR products. Among these 138 (77%) had significant homology at 1×e-7 level to a known resistance gene; only 95 had the correct reading frame. The remaining 43 contained one or more stop codons in their reading frames, and are likely to be pseudogenes or PCR artifacts. Since sequences were only single-pass, it is not possible to distinguish between these two possibilities. Thirty-five of the PLTR sequences did not share any homology with any known sequence or were partial. Six sequences from this primer set showed good match (i.e. E-value>$1\times e^{-10}$) with non resistance genes, one of which (PLTRP2G10) encoded a *gag-pol* protein precursor. The rest of the five non resistance gene coding sequences were reverse transcriptase or transposon-related sequences.

The next category of sequences, PRGA, is also produced with a primer set based on the same conserved patterns as PLTR, but with slightly modified base composition (Table 3.1). Seventy-seven of 171 clones sequenced from this group have yielded a contiguous open-reading-frame. Unlike PLTR sequences, all of the DNA sequences with right reading-frames showed strong homology to known resistance gene analogs with at a threshold of 1 × e-7 level after DNA

sequence comparison. Hence, this primer set generated a less diverse group of resistance gene analogs from peanut than did the PLTR set. A large group of sequences, 52, had some similarity to expected NBS region domains, but their reading frames were interrupted by one or more stop codons. Overall, 129 of 171 sequences (75%) showed significant homology to known resistance gene-like sequences at the threshold of 1×e-7 (Table 3.2). Ten other sequences showed some homology to transposon related sequences, such as reverse transcriptase, transposase, *gag-pol* etc. Another group with a total of 32 sequences did not produce any significant similarity to any known sequence.

A specific primer has been employed for targeted enrichment of Non-TIR like sequences in the final amplicon pool by exploiting polymorphic amino acids in a conserved subdomain of the NBS region, RNBS-D (Penuela SD, 2002). Ninety-one of the 189 sequences did not show significant similarity at 1 × e-7 to any known sequence in GenBank; some contained too many sequencing errors. Another 80 sequences shared significant homology to a known resistance gene (blastx result, e-value=1×e-7). Only 37 of these yielded an uninterrupted open reading frame, with the remaining 43 sequences interrupted by one or more stop codons suggesting that they may be pseudogenes. Only about 41% of the amplicons obtained with this primer set showed significant sequence similarity to a known resistance gene related sequence (Table 3.2), a much lower percentage than the other primer sets. The remaining 18 sequences out of 189 shared sequence similarity with transposases (mostly) or chloroplast genes.

A pair of Cre3 based primers was used for amplifying similar genes from the peanut genome (Lagudah, 1998). The forward primer was based on the kinase-2 and the second primer was a downstream conserved pattern, EGF (Table 3.1). A total of 169 of the cloned amplicons were sequenced (Table 3.2). From these sequences, a total of 78 (46%) genes had significant

similarity with a resistance gene sequence, 20 of which had uninterrupted open reading frames, and the remaining 58 containing one or more stop codons in their reading frames (i.e. putative pseudogenes). Most of the pseudogenes are likely to be real since the stop codons exist at similar positions in all of them. A total of 73 sequences did not match any sequence in GenBank. Only a small proportion of the sequences, 18, matched non-resistance genes, mostly photosystem II and retroelements.

One of the primer pairs, PCf, with limited success in amplifying similar sequences in the peanut genome was based on conserved motifs of extracellular LRR (Leucine-Rich-Repeats) domains of *Cf* genes rendering resistance to *Cladosporium fulvum* in *Lycopersicon esculentum* (Table 3.1: Ohmori et al. 1998). A significant proportion, 107(83%) out of 129, of the sequences had strong homology with chloroplast genes, *NADH dehydrogenase* or *NAD(P)H-quinone oxidoreductase* (Table 3.2). Only six (4%) sequences showed some homology with leucine-rich disease resistance sequences such as *Cf2* and *Cf5* (Table 3.2), and only one of them yielded a contiguous reading frame. A total 16 of the remaining sequences did not match anything in the database or matched to unknown proteins.

The only primer sequence that did not produce any resistance gene-related sequence was based on conserved kinase structures of the *pto* resistance gene of tomato (Table 3.1). Most of the sequences, 107, did not show any similarity with any sequence from GenBank. The remaining 80 sequences were homologous to retrotransposon related genes, such as reverse transcriptase, and *gag-pol* (Table 3.2). This result may suggest an abundance of retroelements in the peanut genome. The lack of kinase-related sequences may be a result of amplification of non-target sequences. In fact, similar sequences have been amplified in common bean (Vallad et al. 2001).

Table 3.2 *Arachis hypogaea* amplicons amplified with degenerate primers from conserved domains of several classes of cloned resistance genes

Sequence Id[a]	No of clones sequenced	No of RGA[c] matches	No of RGAS with ORF	Non-RGA	No match
PLTR	179	138 (77%)	95	6	35
PNTR	189	80 (45%)	37	18	91
PRGA[b]	171	129 (75%)	77	10	32
PCf	129	6 (5%)	1	107	16
PCRE	169	78 (46%)	20	18	73
PPTO	191	0 (0%)	0	80	107
Total	**1028**		**234**		

[a] The sequence Ids are same as the primer Ids used in table 3.1
[b] J. Ballester and A. H. Paterson, unpublished data
[c] The minimum E-value was 1×10^{-7} in blasting against the plant protein database

B. Multiple Alignments and Phylogenetic Analysis of Peanut Resistance Gene Analogs:

1) Peanut Resistance Gene Analogs Generated from Wheat Nematode Resistance Gene Based Primers (PCRE)

The sequences extending from kinase-2 to a downstream conserved motif EGF were subjected to multiple alignments by CLUSTAL_X. To get a more complete picture of the PCRE genes, each of two major classes of NBS-LRR genes, LZ-NBS-LRR (*Prf,* and *Mi1.2*) and TIR-NBS-LRR genes (*L6*, *M*, and *N*) are also included in the alignment (Fig 3.3). The expected conserved motifs of kinase-2, kinase3a, GLPL, RNBS-D, and EGF were observed in all of the sequences with the exception of kinase-2, which was absent in a few due to truncated sequences.

All of the peanut Cre3-like gene sequences (PCREs) are located in the same cladistic group of LZ-NBS-LRR genes (Fig 3.4). Most PCREs were very closely related to each other, with the exception of PCREP2G05, which was phyletically distant and much closer to the wheat

nematode resistance gene *Cre3* and tomato root-knot nematode resistance gene *Mi*. In particular, the sequences, PCREP1H03, PCREP1E02, PCREP2C06, PCREP1D09, and a few other sequences not included in this analysis shared homology of more than 90%. These may represent recently duplicated resistance gene-related sequences in *Arachis hypogaea*.

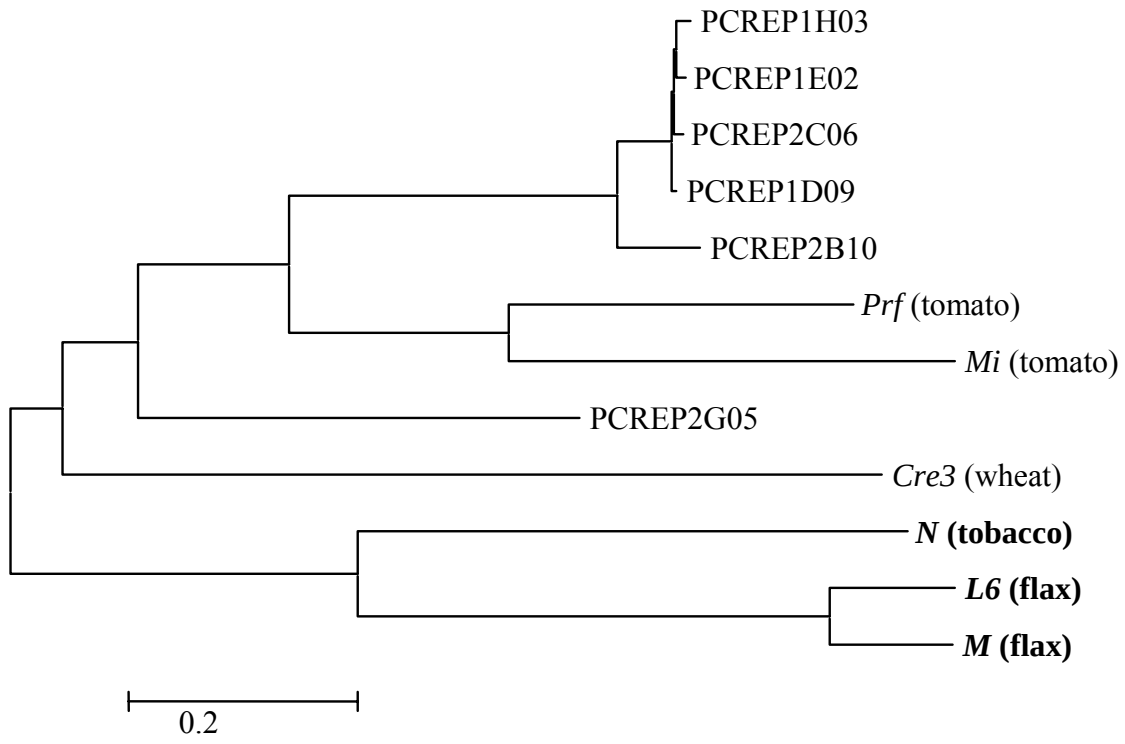


Fig 3.3 Phyletic relationships between PCRE (Peanut *Cre3* like gene sequences) and other resistance genes. The tree is based on CLUSTAL_X alignment. The confidence of the nodes was tested by bootstrapping of 1000 replicates; bootstrap test results varied between 65% and 100%. The branch lengths are proportional to the average number of amino acid substitutions per site as indicated by the scale. TIR class genes are indicated in bold characters, and the species origin for each sequence is denoted in parenthesis.

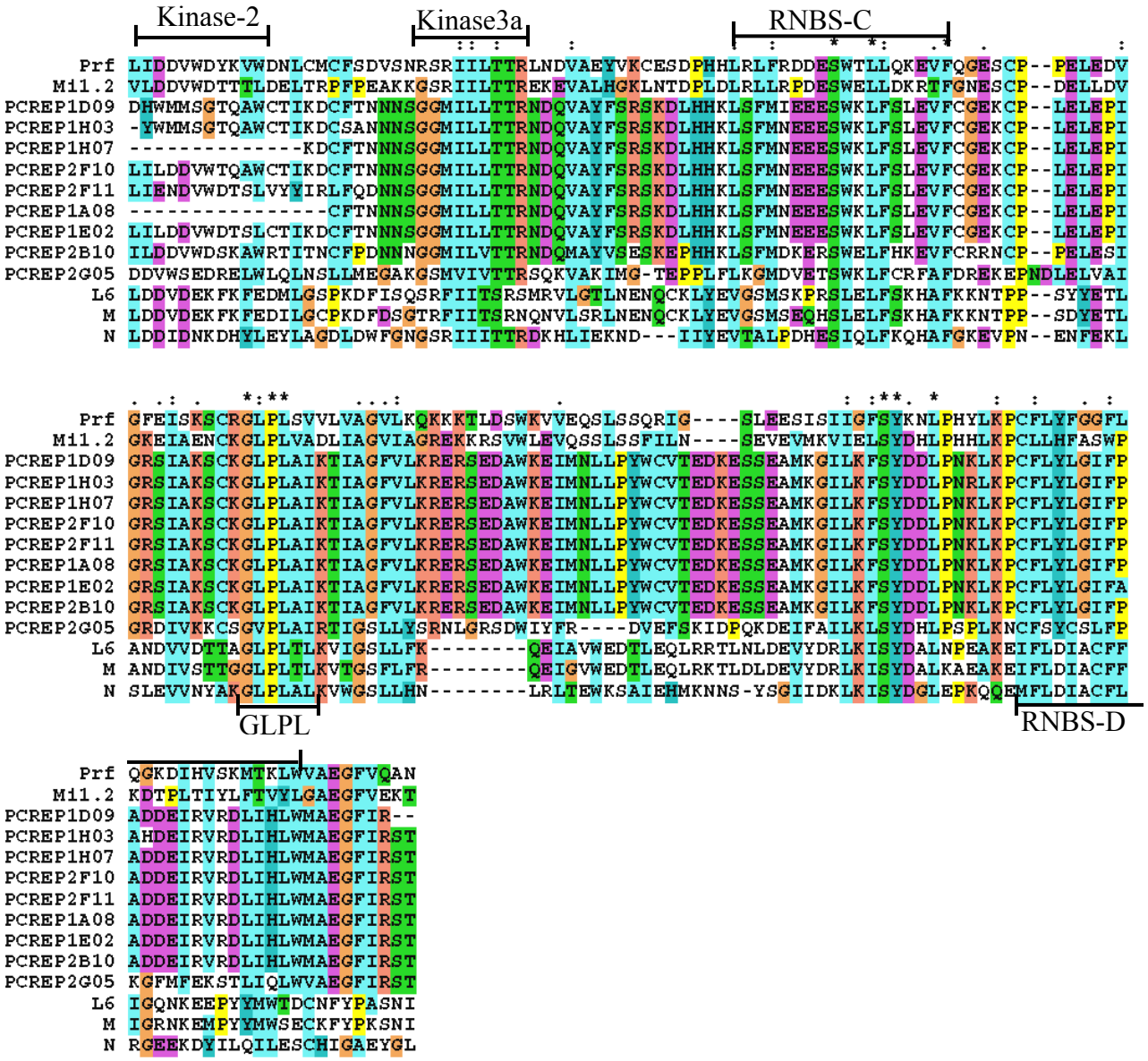


Fig 3.4 Multiple alignment of peanut resistance gene analogs (PCREs) generated with a primer set based on conserved motifs from the wheat nematode resistance gene, *Cre3*. The sequences of *Prf* and *Mi1.2* are LZ-NBS-LRR genes from tomato. *L6*, *M* and *N* are TIR-NBS-LRR resistance genes from tomato (first two) and tobacco. The sequences cover the NBS region from kinase-2 to a downstream conserved region of EGF. The major conserved domains, kinase-2, kinase3a, GLPL, and RNSBD are shown.

2) Analysis of Peanut Leucine-Rich RGAs (Resistance Gene Analogs):

Amplification of peanut leucine-rich disease resistance gene-like sequences has been attempted with a degenerate primer set based on the conserved motifs of two disease resistance genes, *Cf9* and *Cf2* from tomato. Even though six sequences have shown homology (E-value≥1×10^{-7}) (Table 3.2), only one of them, PCfP2F02, had a contiguous open reading frame allowing a multiple alignment comparison with similar sequences from other plant species (Fig 3.5). However, the sequence did not extend the full length of the region of interest.

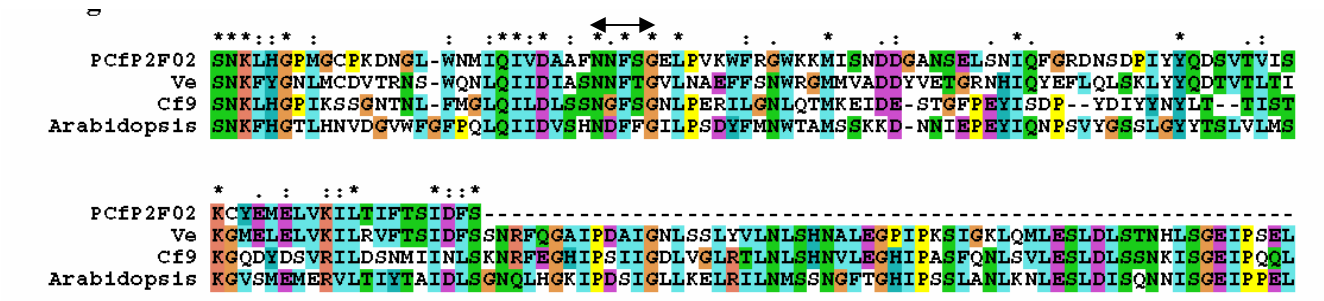


Fig 3.5 A peanut leucine-rich sequence (PCfP2F02) aligned with similar sequences, *Cf9*, *Ve*, from tomato, and a similar gene from *Arabidopsis* by using CLUSTAL_X. Double arrow denotes the possible N-glycosylation consensus sequence NX(S/T).

3) Analysis of Peanut NonTIR and TIR Resistance Gene Analogs:

A total of 132 peptide sequences encompassing a highly conserved region from the P-loop to GLPL of motifs of the NBS domain was multiply aligned with the corresponding regions of orthologous genes from other plant species, and a sample of the alignment is denoted (Fig 3.6). All of the expected motifs, including the P-loop, RNBS-A, kinase-2, kinase3a, RNBS-C and GLPL are observed in all sequences with the exception of ends of a few sequences that were missing GLPL (Fig 3.6).

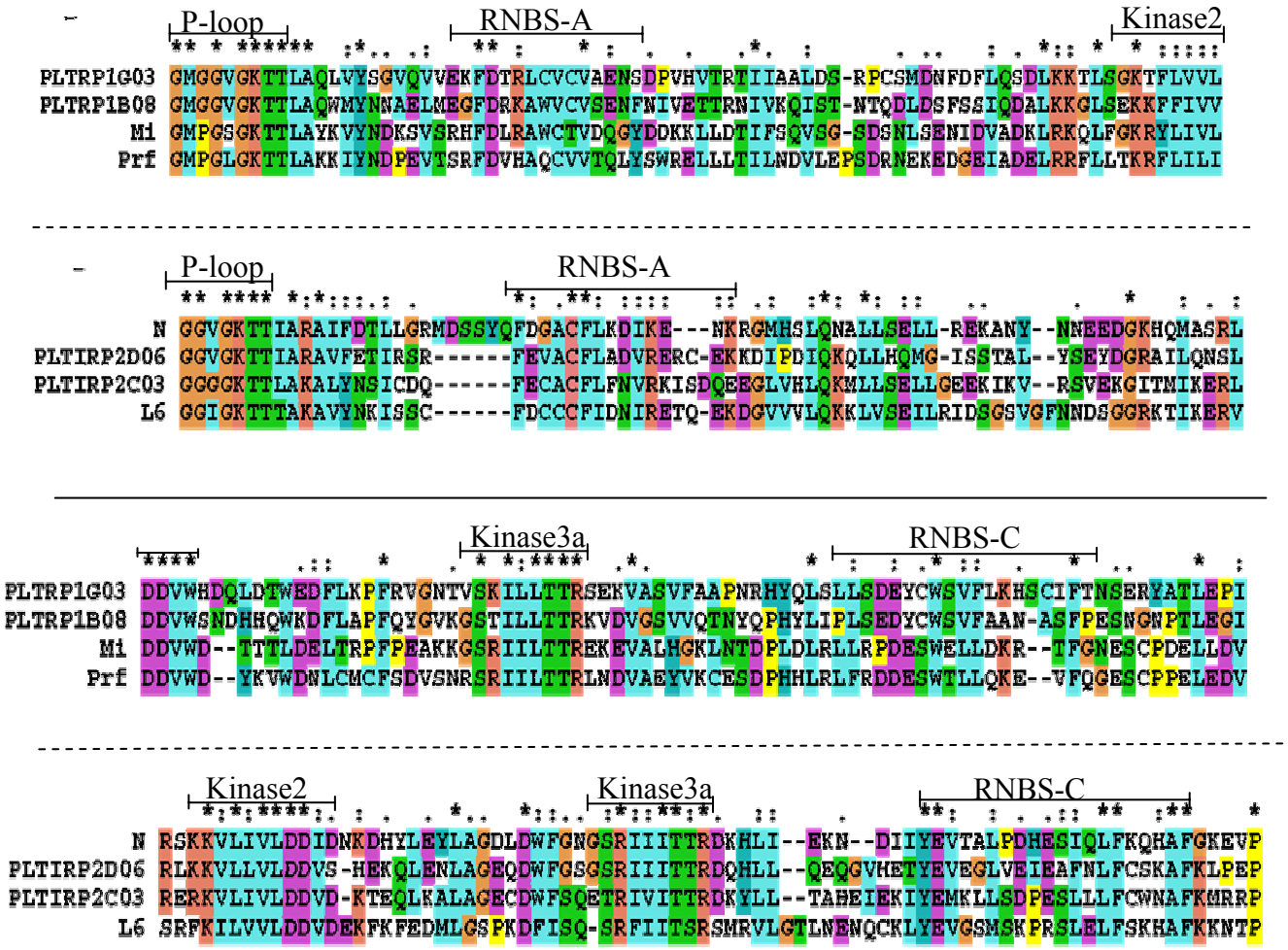


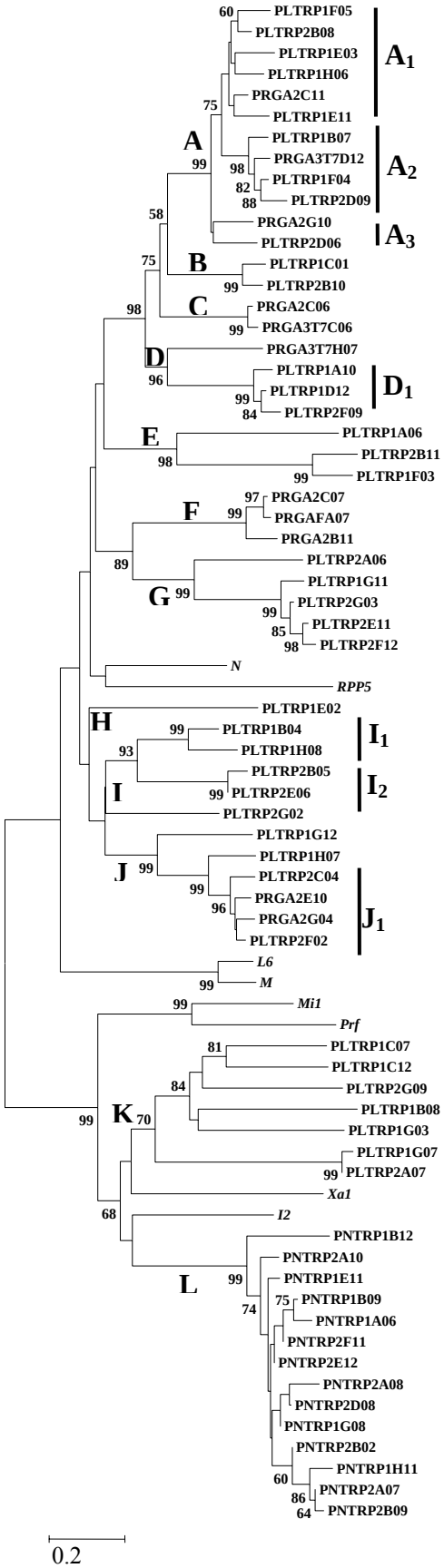
Fig 3.6 Multiple alignments of a sample of peanut resistance gene analogs from groups of TIR and NonTIR sequences with the corresponding domains of other resistance genes, *N* (tobacco), *L6* (flax: TIR), and *Mi* and *Prf* (tomato: NonTIR). The sequences above the dashed line are from the NonTIR class (PLTRP1G03, PLTRP1B08, *Mi*, and *Prf*), and below are from the TIR class. The major conserved motifs, P-loop, RNBS-A, kinase-2, kinase3a, RNBS-C are denoted. The distinctive W (Tryptophan) residue between these two classes in both RNBS-A and Kinase-2 motifs can be seen.

Fig 3.7 Multiple sequence alignment of peanut PNTRs (sequences obtained by targeted amplification of NonTIR sequences). The similar sequences, *Mi* and *I2* (tomato) *Xa1* (rice),

RPS2 (*Arabidopsis*), and AANs (*Arachis hypogaea*, obtained from GenBank) are also included. The bold letters denote the nearly consensus F (phenyl-alanine) and W (tryptophan).

The primers of PLTR and PRG were not designed for a targeted amplification of either the NonTIR or TIR subclass of NBS-LRR genes. A similar alignment of peanut sequences generated with the PNTR primer set, which is designed for the targeted isolation of the NonTIR class, was done. The sequence homology among PNTR family members was very high and these sequences did not contain a distinguishable RNBS-C motif, which existed in all other genes (Fig 3.7). However, the P-loop, RNBS-A, kinase-2, and kinase3a motifs were present

For the phylogenetic analysis of the NBS domain sequences, a neighbor-joining tree was constructed, after the multiple-alignment of a total of 132 sequences containing the expected domains from Ploop to GLPL. The sequences that are more than 95% similar are not depicted on the tree due to space limitations (Fig 3.8). Phylogenetic analysis of PLTR and PNTR sequences has predictably revealed two distinct clades, NonTIR and TIR (Fig 3.8). Seven PLTR sequences, PLTRP1C07, PLTRP1C12, PLTRP2G09, PLTRP1B08, PLTRP1G03, PLTRP1G07, and PLTRP2A07, which shared common features with the NonTIR class (Fig 3.8), plus all PNTR sequences and known NonTIR sequences, *Mi*, *Prf*, *Xa1*, *I2* formed a group; thus, supporting their classification as putative NonTIR sequences (Fig 3.8).



Overgos	Clade	Shared	Unique(%)	
PRGOV44[1]	A1	64	15	(19)
PRGOV52	A1&A2	48	0	(0)
PRGOV45	A2	27	64	(70)
PRGOV49	A3	20	1	(5)
PRGOV01[2]	B	14	29	(67)
PRGOV07[3]	PRGA3T7H07	53	14	(21)
PRGOV04[4]	D1	28	9	(24)
PRGOV12	PLTRP1A06	0	10	(100)
PRGOV14[5]	E1	15	56	(79)
PRGOV36[6]	F	58	103	(64)
PRGOV34[7]	G	7	8	(53)
PRGOV16[8]	PLTRP1E02	5	35	(88)
PRGOV19[9]	I1	15	76	(84)
PRGOV22[10]	I2	0	18	(100)
PRGOV50	PLTRP2G02	3	0	(0)
PRGOV25[11]	PLTRP1G12	11	12	(52)
PRGOV28[12]	PLTRP1H07	25	10	(29)
PRGOV31[13]	J1	24	28	(54)
PRGOV40	PLTRP1C07	1	13	(93)
PRGOV39	PLTRP2G09	1	35	(97)
PRGOV53	PLTRP1G07	0	6	(100)
PRGOV41[14]	L	8	11	(58)

TIR

NonTIR

Fig 3.8 The region from the conserved motif P-loop to GLPL of NBS-LRR class resistance genes from peanut and other plant species are phyletically analyzed by construction of a Neighbor Joining tree using MEGA.2.1(Kumar et al. 2001) after multiple alignment of peptide sequences with CLUSTAL_X (Thompson et al. 1997). The sequences which are more than 95% similar are not depicted on the tree. The genes that are not from peanut are denoted in bold italicized characters. Different cladistic and subphyletic groups are designated with bold capital letters. The confidence levels of nodes were tested by bootstrapping of 1000 replications, and the bootstrap values that are more than 60% are indicated on the branches. The branch lengths are determined by the proportion of average number of amino acid substitutions per site as indicated by the scale. Overgos designed on the basis of homology for the each cladistic group, which are depicted under the "clade" column, are shown on the side of the tree (See the methods and materials for details of overgo design). **Note:** The following overgos are also part of the list; **[1]** PRGOV46, PRGOV47, PRGOV48, PRGOV51; **[2]** PRGOV03; **[3]** PRGOV08, PRGOV09; **[4]** PRGOV05, PRGOV06; **[5]** PRGOV15; **[6]** PRGOV37, PRGOV38; **[7]** PRGOV35; **[8]** PRGOV17, PRGOV18; **[9]** PRGOV20, PRGOV21; **[10]** PRGOV23; **[11]** PRGOV27, **[12]** PRGOV29, PRGOV30; **[13]** PRGOV32, PRGOV33; and **[14]** PRGOV43. The **"**Shared" column is the number of BACS identified by more than one probe, only 33% of shared hits were because of multiple probes within the clades; for clades with multiple probes the total number of hits is indicated, and the "Unique" column is the number of BACs identified by only one probe, where the percentage of the unique hits were indicated in parenthesis. The upper part of the tree contains TIR genes, while the lower section is composed of NonTIR genes.

C. Physical Mapping and Genomic Distribution of Peanut Resistance Gene Analogs

1) Screening of a Peanut BAC Library with Overgos Based on Peanut Resistance Gene Analog Sequences:

Initial screening of the library has been done with the overgos, which were designed with maximum sequence homology in mind. The motifs used as templates for the Ploop1 and Ploop2 was P-loop; for kin21 and TIRKIN2 was kinase-2; for kinase3a1 was kinase3a; and for RNBSA-1, RNBSA-2, and RNBSA-3 was RNBS-A. From nine overgos, a total of 228 hits were obtained (Fig 3.9). Among these, 117 were unique hits; the rest co-occurred with other probes hinting at the presence of possible multiple copies on the same BAC. This result would be consistent with the physical distribution of resistance genes in many plant genomes.

To attempt to maximize the number of resistance gene loci identified in peanut, and to explore possible correlations between phyletic origin and genomic location, we designed overgo primers specific to each clade (Fig 3.8). Each one of the overgo primers, PRGs, had sequence homology at the level of at least 38/40 bp with all sequences within one clade. To minimize the likelihood of overlapping hits due to sequence similarity, overgo primer sequences were designed in such a way that they did not share more than 30 bp with any sequence outside their clade. Much prior experience across many taxa has shown that overgo primers with more than six bp mismatch would identify different loci at the stringency level used in the experiment (see the methods and materials section). Hence, this design strategy should permit us to explore the resistance gene locus diversity in peanut and gain further insight into the evolution of these resistance genes by analyzing the physical distributions of sequences belonging to different clades. A total of 47 overgos designed as described above was used in screening the peanut BAC

library. A total of 1028 hits was obtained, 701 of which were unique. The number of BACs hybridized per overgo ranged from as few as 3 to as many as 98 (Fig 3.9).

Overall, from application of a total of 56 overgo probes designed from peanut resistance gene analog sequences to the peanut BAC library, a total of 1256 hits was obtained (Fig 3.9). Of these hits, 736 were to unique BAC clones, the rest overlapping between overgos.

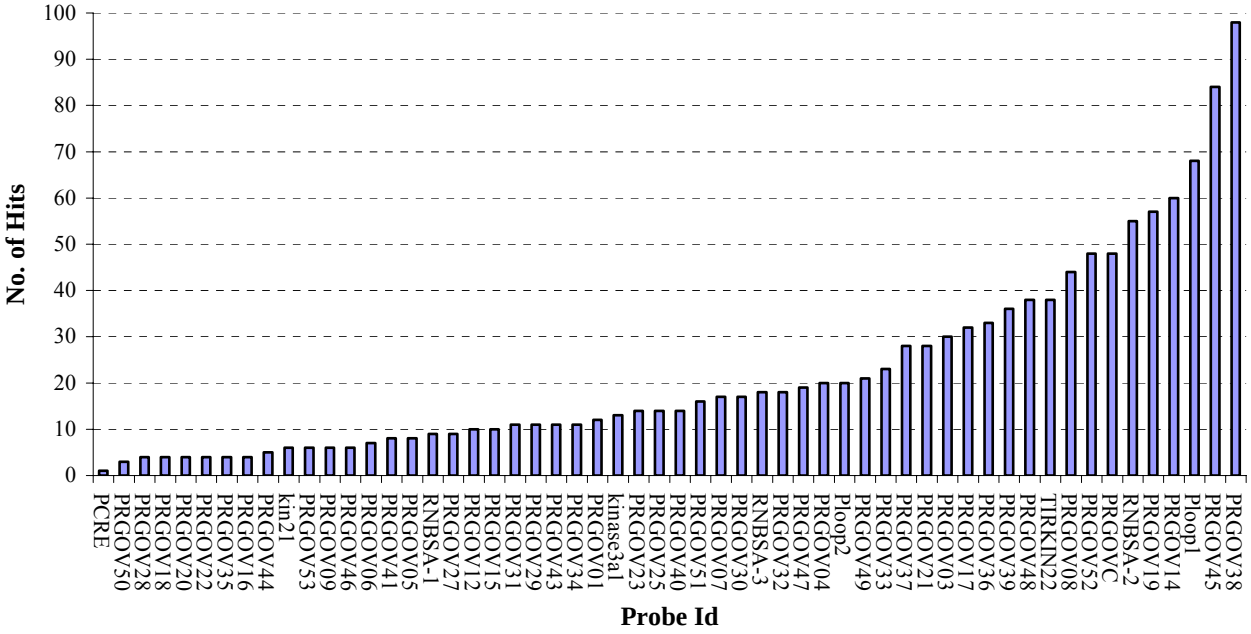
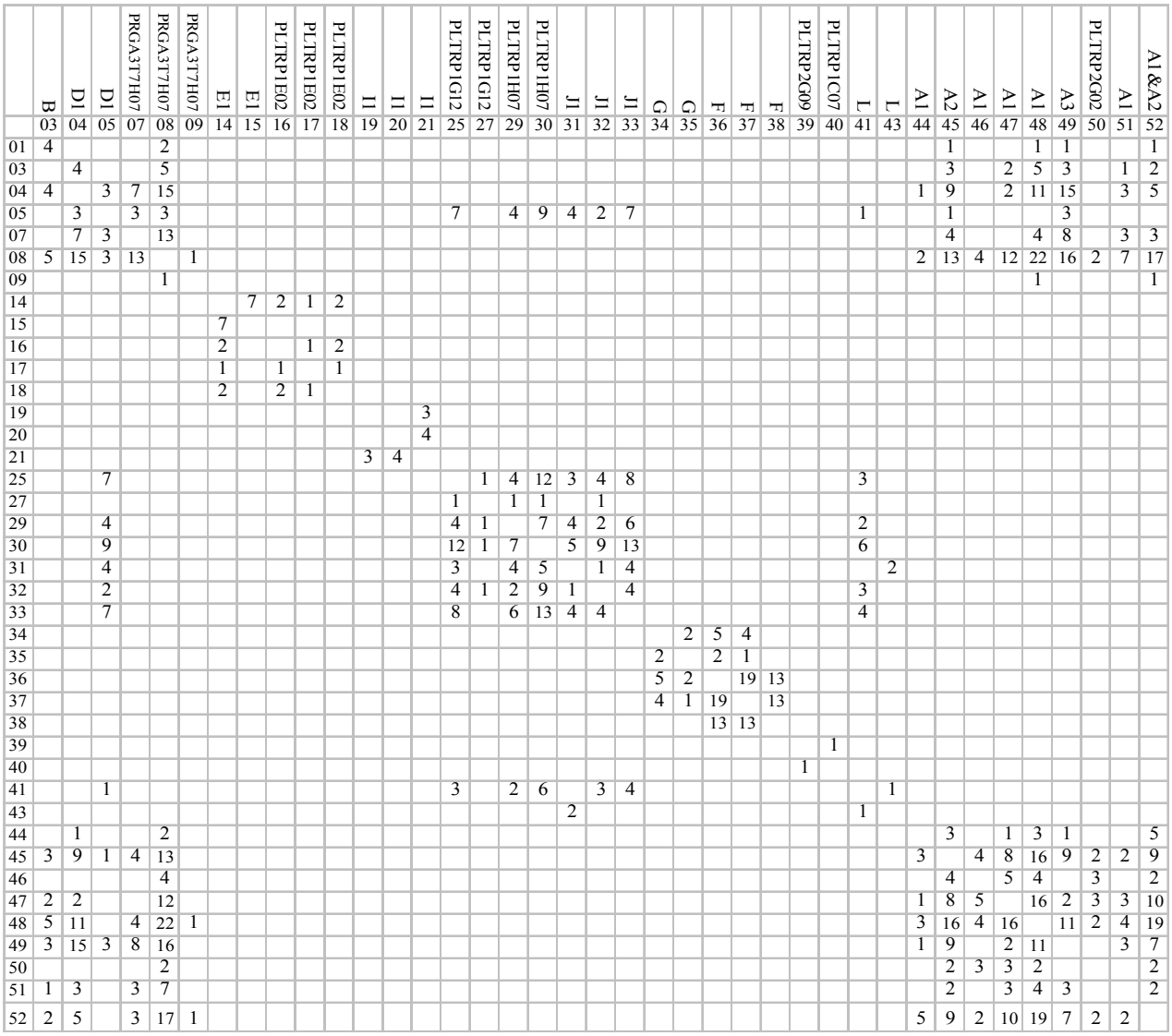


Fig 3.9 Number of BAC hits for each overgo probe designed from peanut resistance gene analogs on the basis of sequence homology. The "PRGOV" prefix represents the overgos designed specifically for a particular clade of resistance gene analog (For more detailed explanation, see Fig 3.8, and Methods and Materials). Other probes were designed from conserved motifs such as kinase-2, P-loop, and kinase3a.



	B	D1	D1	PRGA3T7H07	PRGA3T7H07	PRGA3T7H07	E1	E1	PLTRP1E02	PLTRP1E02	PLTRP1E02	I1	I1	I1	PLTRP1G12	PLTRP1G12	PLTRP1H07	PLTRP1H07	J1	J1	J1	G	G	F	F	F	PLTRP2G09	PLTRP1C07	L	L	A1	A2	A1	A1	A1	A3	PLTRP2G02	A1	A1&A2
	03	04	05	07	08	09	14	15	16	17	18	19	20	21	25	27	29	30	31	32	33	34	35	36	37	38	39	40	41	43	44	45	46	47	48	49	50	51	52
01	4				2																											1			1	1			1
03		4			5																											3		2	5	3		1	2
04	4		3	7	15																									1	9		2	11	15		3	5
05		3		3	3										7		4	9	4	2	7								1		1				3			
07		7	3		13																										4			4	8		3	3
08	5	15	3	13		1																								2	13	4	12	22	16	2	7	17
09					1																													1				1
14								7	2	1	2																												
15							7																																
16							2			1	2																												
17							1		1		1																												
18							2		2	1																													
19														3																									
20														4																									
21												3	4																										
25			7													1	4	12	3	4	8								3										
27															1		1	1		1																			
29			4												4	1		7	4	2	6								2										
30			9												12	1	7		5	9	13								6										
31			4												3		4	5		1	4									2									
32			2												4	1	2	9	1		4								3										
33			7												8		6	13	4	4									4										
34																							2	5	4													
35																						2		2	1													
36																						5	2		19	13												
37																						4	1	19		13												
38																								13	13													
39																												1										
40																											1											
41			1												3		2	6		3	4									1								
43																			2										1									
44		1			2																										3		1	3	1			5
45	3	9	1	4	13																									3		4	8	16	9	2	2	9
46					4																										4		5	4		3		2
47	2	2			12																									1	8	5		16	2	3	3	10
48	5	11		4	22	1																								3	16	4	16		11	2	4	19
49	3	15	3	8	16																									1	9		2	11			3	7
50					2																										2	3	3	2				2
51	1	3		3	7																										2		3	4	3			2
52	2	5		3	17	1																								5	9	2	10	19	7	2	2	

Fig 3.10 The number of shared hits between the overgos is depicted. The first column and second row denote the overgo names (PRGOV and number), which are same as for Fig 3.8. The overgos, which did not share any hits, PRGOV12, PRGOV22, and PRGOV53 are not indicated in the figure. The clade specificities of the overgos depicted in first row are same as in Fig 3.8.

Table 3.3 Genomic organization of the peanut resistance gene-like sequences

Overgo	First screen[a]	Second screen[b]	No. of singletons[c]	No. of contigged clone[d]	No of loci[e]
PRGOV01	13	4	3	1	4
PRGOV03	30	18	8	7	15
PRGOV04	20	14	1	7	8
PRGOV05	10	6	2	3	5
PRGOV06	7	2	0	1	1
PRGOV07	17	10	1	4	5
PRGOV08	44	28	9	9	18
PRGOV09	6	6	5	1	6
PRGOV12	10	0	0	2	2
PRGOV14	61	14	11	3	14
PRGOV15	10	6	3	2	5
PRGOV16	4	3	3	0	3
PRGOV17	32	19	12	4	16
PRGOV18	4	2	2	0	2
PRGOV19	57	34	13	13	26
PRGOV20	4	4	3	1	4
PRGOV21	30	15	10	2	12
PRGOV22	4	2	2	0	2
PRGOV23	14	7	7	1	8
PRGOV25	14	4	3	2	5
PRGOV27	9	0	0	0	0
PRGOV28	4	1	1	0	1
PRGOV29	11	4	1	2	3
PRGOV30	20	8	6	4	10
PRGOV31	11	7	2	2	4
PRGOV32	18	2	2	3	5
PRGOV33	23	6	4	3	7
PRGOV34	11	3	1	3	4
PRGOV35	4	1	1	1	2
PRGOV36	33	19	8	6	14
PRGOV37	28	16	6	5	11
PRGOV38	100	54	26	16	42
PRGOV39	36	3	1	6	7
PRGOV40	14	9	7	2	9
PRGOV41	8	3	3	3	6
PRGOV43	11	7	3	1	4
PRGOV44	5	3		4	4
PRGOV45	86	47	21	19	40

Overgo	First screen[a]	Second screen[b]	No. of singletons[c]	No. of contigged clone[d]	No of loci[e]
PRGOV46	6	5	2	3	5
PRGOV47	19	13	3	8	11
PRGOV48	38	26	8	14	22
PRGOV49	21	15	3	6	9
PRGOV50	3	3	1	2	3
PRGOV51	16	7	0	4	4
PRGOV52	48	23	9	10	19
PRGOV53	6	0	0	0	0
PRGOVCRE54	48	15	8	8	16
kin21	6	4	2	2	4
kinase3a1	13	9	5	2	7
PCRE	1	1	1	0	1
Ploop1	68	46	11	17	28
Ploop2	20	17	5	5	10
RNBSA-1	9	7	3	2	5
RNBSA-2	55	37	12	14	26
RNBSA-3	18	8	2	6	8
TIRKIN22	38	24	8	12	20
Total			**166[f]**	**84[g]**	**250[h]**

[a] The number of BAC clones identified by the application of overgos to the peanut high-density BAC filters

[b] The number of BAC clones that produced positive results after re-hybridization with the same overgos on the southern blots

[c] The number of clones that were positives in both screens, but did not contig with any other clones

[d] The number of contigs formed, including the clones which were not positives at the second screen

[e] The number of loci recognized by a particular overgo is equal to the number of singletons plus the number of contigs hybridized by the overgo; this is a rough estimate, because some of the singletons may have been caused by insufficient number of restriction fragments for the cutoff-value of 1×10^{-10}, or some clones might have been excluded due to occasional fingerprinting failures.

[f] The total number of BAC clones that are confirmed by both screens, but did not contig (note: The sum of the numbers in the column is larger than 80 because a single BAC often hybridized to more than one probe)

[g] The total number of contigs

[h] The estimated number of resistance gene loci in peanut genome= the number of contigs + the number of singletons that are confirmed by both screens

Supplementary Table 3.1 The listing of the contigs and their overgo association

Ctg[a]	overgos
Ctg26	Ploop1, PRGOV01, PRGOV03, PRGOV04, PRGOV49, RNBSA-2
Ctg06	Ploop1, PRGOV03, PRGOV04, PRGOV05, PRGOV07, PRGOV08, PRGOV45, PRGOV48, PRGOV49, PRGOV51, PRGOV52, RNBSA-2, TIRKIN22
Ctg36	Ploop1, PRGOV03, PRGOV04, PRGOV08, PRGOV47, PRGOV48, PRGOV49, RNBSA-2
Ctg72	Ploop1, PRGOV03, PRGOV04, PRGOV45, PRGOV48, RNBSA-2, TIRKIN22
Ctg49	Ploop1, PRGOV03, PRGOV08, PRGOV09, PRGOV48, PRGOV52, PRGOVCRE54, RNBSA-2, TIRKIN22
Ctg12	Ploop1, PRGOV03, PRGOV08, PRGOV45, PRGOV47, PRGOV48, PRGOV52, PRGOVCRE54, RNBSA-2, TIRKIN22
Ctg01	Ploop1, PRGOV04, PRGOV05, PRGOV07, PRGOV08, PRGOV48, PRGOV49, PRGOV51, PRGOV52, RNBSA-2, TIRKIN22
Ctg04	Ploop1, PRGOV04, PRGOV07, PRGOV08, PRGOV44, PRGOV45, PRGOV46, PRGOV47, PRGOV48, PRGOV49, PRGOV50, PRGOV52, PRGOVCRE54, RNBSA-2, TIRKIN22
Ctg18	Ploop1, PRGOV04, PRGOV08, PRGOV45, PRGOV46, PRGOV47, PRGOV48, PRGOV49, PRGOV51, PRGOV52, PRGOVCRE54, RNBSA-2, TIRKIN22
Ctg11	Ploop1, PRGOV07, PRGOV08, PRGOV44, PRGOV47, PRGOV48, PRGOV51, PRGOV52, PRGOVCRE54, RNBSA-2, TIRKIN22
Ctg15	Ploop1, PRGOV08, PRGOV44, PRGOV45, PRGOV47, PRGOV48, PRGOV52, PRGOVCRE54, RNBSA-2, TIRKIN22
Ctg62	Ploop1, PRGOV38
Ctg10	Ploop1, PRGOV44, PRGOV45, PRGOV48, PRGOV52, PRGOVCRE54, RNBSA-2, TIRKIN22
Ctg44	Ploop1, PRGOV45, PRGOV46, PRGOV47, PRGOV48, PRGOV50, RNBSA-2
Ctg74	Ploop1, PRGOV45, PRGOV47, PRGOV48, PRGOV52, PRGOVCRE54, RNBSA-2, TIRKIN22
Ctg64	Ploop1, PRGOV48, TIRKIN22
Ctg38	Ploop1
Ctg81	Ploop2
Ctg55	Ploop2
Ctg19	Ploop2
Ctg47	Ploop2, kin21, kinase3a1, RNBSA-1
Ctg02	kin21, kinase3a1, Ploop2, PRGOV29, PRGOV30, PRGOV31, PRGOV32, PRGOV33, PRGOV41, PRGOV43, RNBSA-1

Ctg[a]	overgos
Ctg61	PRGOV03
Ctg13	PRGOV05, PRGOV25, PRGOV29, PRGOV30, PRGOV31, PRGOV32, PRGOV33, PRGOV41
Ctg54	PRGOV06
Ctg76	PRGOV12
Ctg07	PRGOV12, PRGOV19
Ctg39	PRGOV14
Ctg28	PRGOV14
Ctg21	PRGOV14, PRGOV15
Ctg70	PRGOV15
Ctg34	PRGOV17
Ctg60	PRGOV17
Ctg80	PRGOV17
Ctg29	PRGOV17
Ctg56	PRGOV19
Ctg16	PRGOV19
Ctg27	PRGOV19
Ctg32	PRGOV19
Ctg43	PRGOV19
Ctg66	PRGOV19
Ctg67	PRGOV19
Ctg78	PRGOV19
Ctg79	PRGOV19
Ctg33	PRGOV19
Ctg23	PRGOV19, PRGOV21
Ctg09	PRGOV19, PRGOV39
Ctg03	PRGOV20, PRGOV21
Ctg63	PRGOV23
Ctg37	PRGOV25
Ctg58	PRGOV30, PRGOV32
Ctg68	PRGOV30, PRGOV41
Ctg22	PRGOV33
Ctg14	PRGOV34, PRGOV35, PRGOV36, PRGOV37, PRGOV38, RNBSA-3
Ctg65	PRGOV34, PRGOV36, PRGOV37
Ctg08	PRGOV34, PRGOV36, PRGOV37, PRGOV38, RNBSA-3
Ctg35	PRGOV36, PRGOV37, PRGOV38, RNBSA-3

Ctg[a]	overgos
Ctg69	PRGOV36, PRGOV37, PRGOV38, RNBSA-3
Ctg50	PRGOV36, PRGOV38, RNBSA-3
Ctg17	PRGOV38
Ctg57	PRGOV38
Ctg48	PRGOV38
Ctg83	PRGOV38
Ctg46	PRGOV38
Ctg31	PRGOV38
Ctg05	PRGOV38
Ctg45	PRGOV38
Ctg42	PRGOV38
Ctg24	PRGOV38, RNBSA-3
Ctg82	PRGOV39
Ctg59	PRGOV39
Ctg75	PRGOV39
Ctg20	PRGOV39, PRGOV40
Ctg51	PRGOV39, PRGOV40
Ctg77	PRGOV45
Ctg73	PRGOV45
Ctg71	PRGOV45
Ctg53	PRGOV45
Ctg52	PRGOV45
Ctg41	PRGOV45
Ctg40	PRGOV45
Ctg25	PRGOV45
Ctg30	PRGOV45
Ctg84	PRGOV45

[a] As a cut-off value, 1×10^{-10} was used.

2) Fingerprinting of BAC Clones Containing Resistance Gene-Like Sequences and Construction of a Physical Map:

In order to gain better insight into the genomic organization of resistance gene-like sequences in the peanut genome, we fingerprinted the BAC clones identified by resistance gene analog specific overgos. Of 736 BAC clones, 254 formed 84 different contigs (Table 3.3 and

3.4). The longest contig was 42 CB Units (number of restriction fragments). By assuming one restriction site for every 4 kb on average for a six-cutter, we can estimate the approximate size of this contig as about 168 kb. Thirty-two of the contigs are composed of more than two BAC clones; the other 52 were formed by only two BACs. The fingerprint gels were blotted and re-hybridized with the corresponding overgos to assess copy number and for verification of the presumed BAC clones that passed the first screen. A total of 333 out of 736 clones was verified. A total of 166 clones did not contig, although they passed both screens suggesting that some of them may represent singleton resistance gene loci (Table 3.3).

The band numbers on the Southern blot gels were as high as 18. The remaining 453 clones did not produce any visible band on the films. There could be several plausible reasons for this including failures at the plasmid preparation stage, false positives from the first screen, or failures at the blotting and labeling stages.

The number of loci recognized by each overgo is estimated by adding up the number of different contigs identified by the overgo and number of singletons (i.e. the clones detected by the overgo at the BAC library screening stage, and were successful in Southern blot re-hybridization step, but failed to form contigs). All clones that are contigged were considered in the analysis, even some that failed in the second screen (the clones fingerprinted and re-hybridized with the corresponding overgos) with the assumption that such associations were unlikely to occur by chance and that there would be many steps in this stage causing failures such as blotting and labeling. For the clones that were not associated with any contigs, only the ones which passed both testing steps were considered in analysis (Table 3.3). At the end of this analysis, we estimated the total number of loci characterized at 250. The highest numbers of loci detected by a single overgo (PRGOV38) was 42, and on average 9.5 loci were identified by a

single overgo. For individual loci, the highest number of overgos was 15, and on average 2.11 overgos detected a single locus.

None of the 15 clones detected by PRGOV27 and PRGOV53 are contigged or positively identified by the second screen (i.e. Southern hybridization) (Table 3.3). Forty-nine out of 84 contigs are only composed of clones detected by a single overgo, and the highest number of probes hybridized to a single contig, ctg04, was 15 (Supplementary table 3.1). The longest contigs formed by the clones that belong to Clade A (Fig3.8) suggesting that most of these sequences are physically located in close proximity to each other in the peanut genome, possibly forming a long cluster such as are common in other plant taxa (Meyers et al. 1998; Michelmore 2000; Michelmore and Meyers 1998).

DISCUSSION:

The amplification of several classes of resistance gene-like sequences from peanut with the employment of five different primer sets designed from the conserved motifs of known resistance gene sequences including the NBS domain of NBS-LRR genes, the kinase domain of receptor-like kinase genes, and LRR domains of LRR-TM (such as *Cf*) has been attempted. The success rate and amplicon diversity differed greatly between the primer sets; possible factors contributing to this difference are discussed. The evaluation of phylogenetic similarities and differences between the peanut RGAs and the corresponding sequences from other plant taxa is shown. The possible correlation between phyletic association of the sequences and their genomic distribution is explored. We also explored the overall physical distribution and the evolutionary history of RGAs in *Arachis hypogaea*.

A. Amplification of Peanut Resistance Gene Analogs:

Although five of the sequences, PLTIRP1G07, PLTIRP2E11, PLTIRP2F12, PLTIRP2G03, and PLTIRP2G07 did not have any match at the DNA level, the translation products from these sequences corresponded to resistance gene sequences from more distantly related species such as *Arabidopsis RPP13* or nematode resistance gene-like sequences from *Solanum tuberosum*. This could mean that the variation at DNA level is very high and only some of these sequences have retained sufficient homology to allow recognition through DNA sequence homology searches. It could also mean that the full diversity of resistance gene sequence populations from species more closely related to peanut such as *Glycine max* and *Medicago truncatula*, has not yet been fully revealed.

Even though both PLTR and PRG primer sets were designed from the same conserved motifs (the lengths and degeneracy levels of the primers were different; see the methods and materials for details), the sequences generated from these primers rarely coexisted in the same phyletic group suggesting that the full exploration of the resistance gene-like sequence diversity may require multiple sets of primers with varying level of degeneracy. These two primer sets were also different with regard to the group of sequences in their amplicon pools; that is, none of the sequences from the PRG primer had characteristics of NonTIR sequences, whereas seven of the sequences from the PLTR primer set were NonTIR sequences. This discrepancy might have been caused by the length of the template motifs. The motifs for PRG primers were much shorter, GGVGKTTL and GLPLAL, than for PLTR primers, GMGGVGKTT, and GLPLAALKVLG; hence, the PRG primers were more likely to amplify abundant sequences than the PLTR primers since a shorter motif might be shared by a large number of sequences. In other words, abundant sequences might have competed more efficiently for PRG primers than

more diverse sequences, restricting the diversity of amplicons. In fact the sequences from the PRG primers were much less diverse than the PLTR sequences (Fig 3.8). Slight differences in the amino acid composition in the P-loop and GLPL motifs of these sequences were observed (Bertioli et al. 2003); for instance, methionine in the P-loop was found more frequently in the NonTIR group. Hence, this could be an explanation for the presence of NonTIR sequences only in the PLTR amplicon pool. In brief, the diversity of the amplified sequences increases with the longer primers, for which less abundant sequences can compete in PCR reactions.

The amplification of the NonTIR sequences were less successful, although a specific primer set, PNTR, was employed for it. This result fully corroborates previous observations and evaluations (Bertioli et al. 2003; Penuela et al. 2002). The fact is that legumes, and dicots in general, contain fewer NonTIR resistance genes than TIR genes; for instance, only about 36% of 149 NBS-LRR genes were NonTIR in *Arabidopsis*. A comprehensive sampling of these sequences in *Arachis* may require multiple sets of primers and sequencing of a larger number of amplicons.

The success rate of the primer set based on the extracellular LRR motif of *Cf*-like genes was very limited. One reason behind this poor success rate could be a higher polymorphism rate in this region. In fact, varying copy numbers, intra-or-inter repeat exchanges, and diversifying selective pressure have been very commonly observed in the LRR domains of *Cf* genes (Dixon et al. 1998; Ellis et al. 1999; Parniske and Jones 1999). Extracellular location of the repeats and this higher rate of modifications have led to speculation about putative roles of these repeats in recognition of avR (avirulence gene). Briefly, homology in the priming sites between tomato and peanut might have been completely lost since the divergence of these two species because of a higher evolutionary pace of these repeats; hence, hindering any homology-based amplifications.

The success rate for the other two primer sets, PPTO (*pto*-based), and PCRE (*Cre3*-based) was also limited. This result was expected because these two primers are not degenerate, limiting the number of target sequences in the peanut genome. Nonetheless, the amplicons from the PCRE contained some (20) resistance gene-like sequences with ORFs. This was not the case for the PPTO. This may imply that the kinase domains of receptor-like kinase genes evolve much faster than the NBS domains; thus, preventing the recognition of the target sequences by tomato based primers in the peanut genome. This problem could be solved by using the multiple kinase domain sequences from a variety of plant taxa to choose the best conserved sequence stretch as template for designing degenerate primers. In fact, the degeneracy of the primers allowed the amplification of similar sequences in common bean (Vallad et al. 2001).

B. Phylogenetic and Sequence Analysis of Peanut RGAs

There was a single nearly invariant amino acid residue among all sequences, which is "F" (phenylalanine) at the RNBS-A conserved motif. The differentiating amino acid, "W" in RNBS-A, was present in seven peanut (the PLTRs of NonTIR subgroup) sequences, but not all of the PNTR sequences (Fig 3.6, and Fig 3.7) agreeing with previous findings in the NonTIR class (Cannon et al. 2002; Meyers et al. 1999). Another NonTIR specific amino acid residue was W at the end of the kinase-2 motif, which was also nearly unanimous in all of the NonTIR sequences (Fig 3.6 and 3.7). Moreover, another tryptophan residue at RNBS-C was also unique to the NonTIR class.

Although PNTR sequences contained most of the NBS domain motifs such as the P-loop, kinase-2 and kinase3a, they contained a deletion in the RNBS-C domains. These sequences also contained a unique motif "GHLLPQ", which could be equivalent of the GLPL motif present in other sequences (Fig 3.7). Thus these may represent a new class of sequences or a group of

homologous pseudogenes, albeit most of them did not contain a stop codon in the reading frame. Further research is required to gain insight into the true nature of these sequences.

The peanut resistance gene analogs from the primer set based on the *Cre3* (wheat cyst nematode resistance gene showed the major characteristics of the NonTIR class. PCRE (Peanut *Cre3* like sequences) sequences had W (tryptophan) at the LXVLDDVW motif, which is consistent with the exclusive presence of W in other NonTIR class genes such as *Prf*, *I2,* and *Mi* (Cannon et al. 2003). Additionally, another downstream conserved motif of RNBS-D is showing strong homology with the nearly consensus sequence pattern of the NonTIR class, which is CFLYGFPAD (Meyers et al. 1999). Thus, both blastx results that denoted the close homology of PCREs with *Prf* and these conserved motifs have provided sufficient information for including them in the LZ-NBS-LRR class.

The peanut leucine-rich sequence, PCfP2F02 shared the highest sequence similarity with a tomato fungal pathogen, *Verticillium dahliae*, resistance gene (E-value=1×10^{-18}) (Kawchuk et al. 2001). A high number of G (Glycine) residues, which is common in all extracytoplasmic receptor-like genes (Jones et al. 1994; Song et al. 1995) and presence of one consensus N-glycosylation motif, NX(S/T) further suggest that PCfP2F02 could be a leucine-rich receptor-like sequence from *Arachis hypogaea*. However, in order to be able to reach a more firm conclusion about the nature of this type of receptor-like sequence in peanut, it is essential to obtain more sequences. Nonetheless, this experiment provides evidence for the presence of leucine-rich receptor-like genes in peanut.

In general, the TIR part of the tree had longer internal and shorter terminal branches, corroborating the observations of Cannon et al.(2002) about the tree topology of NBS-LRR sequences (2002). The majority of TIR clades that are indicated with bold capital letters on the

tree (Fig 3.8) contained more genes than are shown. Sequences with more than 95% similarity were not used for the reconstruction of the tree due to space limitations. The tree topology of the TIR class supports the possibility of a recent burst in their copy numbers in peanut, or that convergent or purifying selection is keeping sequence homology high within the clades (Cannon et al. 2002; Meyers et al. 1999; Richly et al. 2002) The sequences within some clades had homology up to 98%. Segmental or whole genome duplications and subsequent genomic rearrangements (Blanc and Wolfe 2004a, b; Simillion et al. 2002) have been suggested as an important mechanism for the movement of resistance genes and creation of new clusters within *Arabidopsis* (Richly et al. 2002). The recent polyploidization event in the peanut genome might have resulted in creation of new clusters, which have not yet been diversified. However, there were occasional exceptions such as clade E that did not contain large numbers of sequences. On the other hand, the lower number of sequences in that phyletic group might have been caused by insufficient sampling.

In contrast to TIR, the terminal branches of the K clade of the NonTIR class are longer, possibly implying much less selective pressure on these sequences in peanut. This was the case for the *Arabidopsis* genome, where the Ka/Ks ratio was 78% higher for TIR than Non-TIR sequences (Cannon et al. 2002). However, clade L in the NonTIR class seems not to fit to this scenario. Clade L contained sequences which were almost 98% identical. At least some of the sequences must have functional constraints hindering their divergence. In fact, NBS domains in all expressed disease resistance genes contain highly conserved functionally essential motifs (Hammond-Kosack and Jones 1997). It needs to be noted that we may not have a large enough sample to shed light on the evolutionary history of the entire NonTIR gene class in peanut.

The tree formed by PCRE (Peanut *Cre3*-like genes) has also shared some features with the NonTIR part of the tree in Fig 3.8. Like PNTR sequences in the NonTIR part of the tree (Fig 3.8), most PCRE derived sequences had very close sequence homology (more than 90%), even though this result might have been caused by the fact that this primer set was not degenerate, limiting the number of target sequences in the peanut genome. This result still further supports the possibility that some of the NonTIR sequences in the peanut genome could be under purifying selection (keeping the homology high), or be a result of amplification.

C. Analysis of Physical Distribution of Peanut Resistance Gene Analogs

From the overgos for maximum homology with a maximal number of sequences, the highest number of hits was obtained from the probes based on the P-loop (Fig 3.9) suggesting that DNA level variation at the P-loop was lowest. Thus, overgos based on this motif were able to identify more resistance gene loci in peanut. In contrast, the probes designed from the kinase-2 domain yielded the lowest number of hits. This could be explained by the fact that the degeneracy of the code at the kinase-2 motif is much higher than in the P-loop because of a high number of leucines (Fig 3.6) causing a much lower level of homology at the DNA level.

The most striking result from the hybridization of the probes designed from the sequences with phyletically isolated groups (Fig 3.8) was that they identified unique BACs (i.e. the BACs, which are identified by a single probe). For instance, overgos based on the sequences of the NonTIR part of the tree, PLTRP2A07, PLTRP1G07, PLTRP1C07, PLTRP2G09 largely identified BACs which were unique to the respective overgos. There was only a single co-hybridization to the same BAC , which was common between probes for PLTRP1C07 and PLTRP2G09.The uniqueness of hits from these overgo probes corroborates the phylogeny results, which show all of these sequences to be highly diverged, possibly obscuring

identification of any other resistance gene locations (Fig 3.8). Another interesting finding was that none of the overgos from the clade K sequences had hits on the same BAC with TIR group sequences, meaning that NonTIR and TIR sequences were not found in any of the same clusters (Fig 3.10). This is in agreement with the genomic organization of these two classes in *Arabidopsis* (Richly et al. 2002). However, overgo probes based on the PNTR sequences, clade L, (Fig 3.8, Fig 3.10), had some common hits with the TIR sequence based overgos although they grouped with NonTIR sequences (Fig 3.8). These results could imply either that these sequences do not actually belong to the NonTIR class even though they share some characteristics (Fig 3.7), or that NonTIR and TIR class genes may occasionally exist in the same cluster in peanut, unlike *Arabidopsis*. In fact, mixed clusters containing both subclasses were observed in another legume, *Medicago truncatula* (Zhu et al. 2002). Further research is required in order to clarify this issue.

The divergence level of sequences (i.e. sequences from the cladistic groups with few sequences and longer internal branches (Fig 3.8) was loosely correlated with the number of shared hits in the TIR part of the tree. For example, none of the 10 BACs recognized by the PLTIRP1A06 based probe hybridized with any other probe. The result is in agreement with the position of this sequence on the tree, which is by itself in a single lineage (Fig 3.8). In other words, there were no sequences sharing a high level of homology with that sequence, unlike other clades that contained many very closely related sequences. Sequences such as PLTRP1A06 may represent singleton peanut resistance loci in the peanut genome. Similarly, two probes specific to clade I2 identified single hit BACs conforming to the abovementioned assertions. In contrast, 16% of the hits for the adjacent clade I1 based overgos were common with other probes. There were more sequences on the I1 branch or clade I; hence some multiple hits would

be expected (supplementary Table 3.1). The highest number of hits was obtained from the probes of clade F, which contained more than 25 sequences with more than 96% sequence homology. Hence, these sequences could reflect a very recent amplification of resistance genes. Moreover, the results from the rest of the overgos were correlated with their divergence (i.e. the more divergent sequences had more unique hits).

Probes designed from different regions of the same group of sequences usually identified different BAC clones. Only 134 (33%) among the total of 406 shared hits were a result of the use of multiple overgo probes within a clade. The result could mean that the amount of DNA level sequence variation varies significantly over the region spanning NBS domains. Hence, it was necessary to design multiple overgos anchored to different parts of these sequences in order to get a complete picture of the genomic organization of the resistance genes. The number of hits from overgos designed from different regions of the same gene were not significantly different (Single factor ANOVA analysis P-value= 0.1913).

Based on *Hin*dIII digestion of the BACs, the number of bands hybridizing to overgo probes varied from 1 to as many as 18 bands (about 100 BAC clones contained more than 2 bands) suggesting that some of the BAC clones contain multiple copies of resistance gene-like sequences, consistent with observations in many other plant taxa (Meyers 2003; Michelmore and Meyers 1998). On the other hand, some BAC clones produced a single band hinting at the presence of singletons. The average number of bands per BAC clone varied with respect to their overgo specificity. The BAC clones detected by more than one overgo (higher percentage of shared hits) usually had on average more bands. For instance, BAC clones specific to the overgos, PRGOV04, PRGOV05, PRGOV06 (21% unique hits), PRGOV07, PRGOV08, and PRGOV09 (24% unique hits) had 4-5 bands on average, whereas the clones from the overgos,

PRGOV01, and PRGOV03 (67% unique hits) had about 2 bands (Fig 3.8 and 3.10). Similar results were also observed for the BACs identified by overgos specific to the clades I1 and I2; the average number of bands were 3 and 1.6 respectively corroborating the number and percentage of the unique hits, (84, and 100%). However, not all the observations agreed with these assessments; for example, clades A1, A2, and A3 on average had one band, even though BAC clones in those categories were detected by multiple overgos; possibly implying much larger intergenic space or possible cross-hybridization.

Most of the overgos yielded the predictable results in terms of their contig identities and number of singletons. For instance, the clones identified by overgos such as Ploop1, which is designed for maximum sequence homology for all peanut resistance gene analogs, formed contigs with many clones (Supplementary table 3.1). Similar overgos such as TIRKIN22, RNBSA-2, RNBSA-3, and kinase3a1 are associated with many contigs supporting the evaluations made on the basis of unique and shared hits (Fig 3.8 and Fig 3.10).

The contig association and phyletic origin of overgos were related. In other words, the clones identified by the overgos designed from phyletically close sequences were more likely to form contigs than the distant ones. For example, clones specific to PRGOV44, PRGOV45, PRGOV47, PRGOV48, PRGOV51, PRGOV49, PRGOV01, and PRGOV03 usually contigged, agreeing with phyletic closeness of these sequences on the tree (Fig 3.8 and Supplementary table 3.1). This result could mean that these overgos identified resistance gene loci that were in very close physical proximity and also were similar in DNA sequence. A similar conclusion could be reached for other overgos such as PRGOV36, PRGOV37, and PRGOV38. This result was expected from the number of shared hits (Fig 3.8 and Fig 3.10). However, not all clones from these overgos contigged together. There were also singletons and overgo-specific contigs such as

ctg05, which is composed of only PRGOV38 specific clones. This may suggest that the overgos were successful in identifying a diverse group of resistance gene loci in the peanut genome. In contrast, the clones from PRGOV39, and PRGOV40, with one exception, ctg09 (Supplementary table 3.1), did not contig with any clone from the TIR part of the tree agreeing with earlier evaluations about the mostly non-overlapping distributions of the two groups. It also suggests that most of the clones identified by these overgos were locus specific. In contrast, most clones from PRGOV41 and PRGOV43, which are based on PNTR sequences (NonTIR part of the tree, Fig 3.8), usually formed contigs containing TIR clones meaning that these sequences may be closely located with the TIR sequence even though they are phylogenetically more closely associated with the NonTIR sequences (Fig 3.8 and Supplementary table 3.1). This would corroborate the observations of Zhu et al (2002) in *Medicago*, where occasional mixed clusters of TIR and NonTIR were present, unlike *Arabidopsis* (Richly et al. 2002). The main factor shaping genomic distribution of the resistance gene sequences in *Arabidopsis* is claimed to be segmental or whole genome duplication followed by extensive rearrangements, which may result in movement of resistance gene to new locations and dismantling of the clusters (Blanc and Wolfe 2004a, b; Ku et al. 2000; Richly et al. 2002; Simillion et al. 2002). Thus, this conclusion may suggest that the heterogeneous nature (containing both TIR and NonTIR classes) of some clusters in legumes could be because of chance rather than functional constraints.

In summary, we were able to identify about 250 loci consisting of resistance gene-like sequences (Table 3.3). However, it should be noted that this estimate is rough; that is because some clones might have not contigged due to an insufficient number of restriction fragments, or failure of fingerprinting. We suggest a positive correlation between phyletic location of resistance like genes and their physical distribution in the peanut genome. The clusters

containing large numbers of highly homologous sequences such as clade A (Fig 3.8) were also closely associated with each other in the peanut genome. This could suggest a sudden burst in copy numbers of these sequences resulting in very closely related proximal groups of sequences.

This is hitherto the most comprehensive study about the physical distribution of resistance gene-like sequences in peanut. The BAC clones containing RGAs will provide initial steps for the future cloning of localized resistance genes in the peanut genome. The sequencing of BAC clones carrying clusters of resistance gene-like sequences will contribute to better understanding of their organization and evolution. Spatial and temporal expression analysis of the isolated full-length sequences can be done, possibly leading to the discovery of important resistance genes. The BACs we isolated will serve as a valuable resource for future research about the resistance gene family in the peanut genome in terms of practical applications such as map-based cloning and for gaining more detailed insights into their evolution and organization.

REFERENCES:

Aarts MG, Hekkert A, Baste L, Holub EB, Beynon JL, Stiekema WJ, Pereira A (1998) Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Mol Plant Microbe Inter 11: 251-258

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids. Res. 25: 3389-3402

Anderson PA, Lawrence GJ, Morrish BC, Ayliffe MA, Finnegan EJ, Ellis JG (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell 9: 641-651

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 SCM, Lion MB, Santos VL, Pappas G, Cannon SB, Guimaraes PM (2003) A large scale analysis of resistance gene homologues in *Arachis*. Mol Gen Genomics 270: 34-45

Blanc G, Wolfe KH (2004a) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16: 1679-1691

Blanc G, Wolfe KH (2004b) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16: 1667-1678

Branch WD, Brenneman TB (1999) Stem rot disease evaluation of mass-selected peanut populations. Crop Protection 18: 127-130

Burow 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, McCombie WR, Sato S, Tabata S, Denny R, Palmer L, Katari M, Young ND, Stacey G (2003) Evolution and microsynteny of the *apyrase* gene family in three legume genomes. Mol Gen Genet 270: 347-361

Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2002) Diversity, distribution, and ancient taxonomic relationships within the TIR and Non-TIR NBS-LRR resistance gene subfamilies. J Mol Evol 54: 548-562

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol Plant Microbe Inter 11: 968-978

Di Gaspero G, Cipriani K (2003) Nucleotide binding site/leucine-rich repeats, *Pto*-like and receptor-like kinases related to diseases resistance in grapevine. Mol Gen Genomics 269: 612-623

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JDG (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell 10: 1915-1926

Ellis JG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. Plant Cell 11: 495-506

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dang JL (1995) Structure of the *Arabidopsis RPM1* gene enabling dual specific disease resistance. Science 269: 843-846

Hammond-Kosack K, Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant Physiol Plant Mol Biol 48

Huettel B, Santra D, Muehlbauer J, Kahl G (2002) Resistance gene analogues of chickpea (*Cicer arietinum* L.): Isolation, genetic mapping and association with a *Fusarium* resistance gene cluster. Theor App Genet 105: 479-490

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JDG (1994) Isolation of tomato *Cf9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266: 789-793

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

Kawchuk LM, Hachey J, Lynch DR, Kulcsar F, van Rooijen G, Waterer DR, Robertson A, Kokko E, Byers R, Howard RJ, Fischer R, Prufer D (2001) Tomato *Ve* disease resistance genes encode cell surface-like receptors. Proc. Natl. Acad. Sci 98: 6511-6515

Ku H-M, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. Proc. Natl. Acad. Sci 97: 9121-9126

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics 17: 1244-1245

Lagudah ES, Seah S, Sivasithamparam K, Karakousis A (1998) Cloning and characterization of a family of disease resistance gene analogs from wheat and barley. Theor App Genet 97: 937-945

Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. Plant Cell 7: 1195-1206

Leung H, Line RF, Chen M (1998) Genome scanning for resistance-gene analogs in rice, barley, and wheat by high resolution electrophoresis. Theor App Genet 97: 345-355

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: 317-322

Meyers BC, Michelmore RW (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and death process. Genome Res 8: 1113-1130

Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW (1998) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell 10: 1833-1846

Meyers BKA, Griego A, Kuang H, and Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15: 809-834

Michelmore R (2000) Genomic approaches to plant disease resistance. Current Opinion in Plant Biology 3: 125-131

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM (1998) The Root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10: 1307-1319

Nelson SC, Simpson CE, Starr JL (1989) Resistance to *Meloidogyne arenaria* in *Arachis* spp. germplasm. J Nematol Suppl 21: 654-660

Ohmori T, Murata M, Motoyoshi F (1998) Characterization of disease resistance gene-like sequences in near-isogenic lines of tomato. Theor App Genet 96: 331-338

Pan Q, Wendel J, Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50: 203-213

Parniske M, Jones JDG (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc. Natl. Acad. Sci 96: 5850-5855

Pensuk V, Jogloy S, 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

Penuela SD, Danesh D, Young ND (2002) Targeted isolation, sequence analysis, and physical mapping on nonTIR-NBS-LRR genes in soybean. Theor App Genet 104: 261-272

Phipps PM, Porter DM (1998) Collar rot of peanut caused by *Lasiodiplodia theobromae*. Plant Dis 82: 1205-1209

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

Ross MT, LaBrie T, McPherson J, Stanton VM (1999) Screening large-insert libraries by hybridization. Wiley, New York, NY

Sambrook J, Russell DW (2001) Molecular Cloning: a laboratory manual. vol 1-3. Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York

Saraste M, Sibbald PR, Wittinghofer A (1990) The P-Loop - a common motif in ATP-binding and GTP-binding proteins. trends Biochem Sci 15: 430-444

Seah S, Sivasithamparam K, Karakousis A, Lagudah ES (1998) Cloning and characterization of a family of disease resistance gene analogs from wheat and barley. Theor App Genet 97: 937-945

Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci 99: 13627-13632

Simons G, Groenendijk J, Wijbrandi J, Reijans M, Groenen J, Diergaarde P, Van der Lee T, Bleeker M, Onstenk J, de Both M, Haring M, Mes J, Cornelissen B, Zabeau M, Vos P (1998) Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10: 1055-1068

Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. 10: 1772-1787

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

Sulston J, Mallett F, Durvin R, Horsnell T (1989) Image analysis of restriction enzyme fingerprint audioradiograms. Comput. Appl. Biosci. 5: 101-106

Thompson J, Gibson T, Plewniak F, Jeanmougin F, Higgins D (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids. Res. 25: 4876-4882

Traut TW (1994) The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. European J Biochem 222: 9-19

Vallad G, Rivkin M, C V, McClean P (2001) Cloning and homology modelling of a *Pto*-like protein kinase family of common bean (*Phaseolus vulgaris* L.). Theor App Genet 103: 1046-1058

Whitham S, McCormick S, Baker B (1996) The *N* gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. Proc. Natl. Acad. Sci 93: 8776-8781

Yu Y, Buss G, Maroof M (1996) Isolation of a superfamily of candidate resistance genes in soybean based on a conserved nucleotide binding site. Proc. Natl. Acad. Sci 93: 11751-11756

Zhu HY, Cannon SB, Young ND, Cook DR (2002) Phylogeny and genomic organization of the TIR and non-TIR NBS-LRR resistance gene family in *Medicago truncatula*. Mol Plant Microbe Inter 15: 529-539

Supplementary Table 3.2

Seq Id	GI	Subphyl	clades	Overgos(PRGOV)	Shared	Unique	Total	Unique(%)
PLTRP1E03	AY747341	1	A1	44, 46, 47, 48, 51	64	15	79	19
PLTRP1E11	AY747409	1	A1	44, 46, 47, 48, 51				19
PLTRP1H06	AY747355	1	A1	44, 46, 47, 48, 51				19
PLTRP1H11	AY747357	1	A1	44, 46, 47, 48, 51				19
PLTRP2B08	AY747364	1	A1	44, 46, 47, 48, 51				19
PRGA2E11	AY747521	1	A1	44, 46, 47, 48, 51				19
PRGA2G08	AY747526	1	A1	44, 46, 47, 48, 51				19
PRGA3T7B09	AY747578	1	A1	44, 46, 47, 48, 51				19
PRGA3T7C11	AY747541	1	A1	44, 46, 47, 48, 51				19
PRGA3T7G10	AY747581	1	A1	44, 46, 47, 48, 51				19
PLTRP1F05	AY747344	2	A1	44, 46, 47, 48, 51				19
PLTRP1F07	AY747345	2	A1	44, 46, 47, 48, 51				19
PLTRP1G10	AY747427	2	A1	44, 46, 47, 48, 51				19
PLTRP1H02	AY747353	2	A1	44, 46, 47, 48, 51				19
PLTRP2B06	AY747363	2	A1	44, 46, 47, 48, 51				19
PLTRP2C05	AY747368	2	A1	44, 46, 47, 48, 51				19
PLTRP2D10	AY747371	2	A1	44, 46, 47, 48, 51				19
PRGA2C05	AY747507	2	A1	44, 46, 47, 48, 51				19
PRGA2E05	AY747516	2	A1	44, 46, 47, 48, 51				19
PRGA2G12	AY747573	2	A1	44, 46, 47, 48, 51				19
PRGA3T7A09	AY747531	2	A1	44, 46, 47, 48, 51				19
PRGA3T7A12	AY747533	2	A1	44, 46, 47, 48, 51				19
PRGA3T7B12	AY747536	2	A1	44, 46, 47, 48, 51				19
PRGA3T7C08	AY747539	2	A1	44, 46, 47, 48, 51				19
PRGA3T7D12	AY747544	2	A1	44, 46, 47, 48, 51				19
PRGA3T7E08	AY747545	2	A1	44, 46, 47, 48, 51				19
PRGA3T7F09	AY747547	2	A1	44, 46, 47, 48, 51				19
PRGA3T7G05	AY747550	2	A1	44, 46, 47, 48, 51				19
PRGAFA09	AY747559	2	A1	44, 46, 47, 48, 51				19
PLTRP1B07	AY747333	2	A2	52	48	0	48	0
PLTRP1D04	AY747337	2	A2	52				0
PLTRP1F04	AY747425	2	A2	52				0
PLTRP2D09	AY747370	2	A2	52				0
PLTRP2A10	AY747401	3	A3	49	20	1	21	5
PLTRP2D06	AY747415	3	A3	49				5
PRGA2C12	AY747494	3	A3	49				5
PRGA3T7C12	AY747497	3	A3	49				5
PLTRP1A11	AY747421	4	B	01, 03	14	29	43	67
PLTRP1B06	AY747332	4	B	01, 03				67
PLTRP1C01	AY747334	4	B	01, 03				67
PLTRP2B03	AY747412	4	B	01, 03				67
PLTRP2B10	AY747365	4	B	01, 03				67
PLTRP2F08	AY747391	4	B	01, 03				67
PLTRP1F09	AY747426	5	C	-				
PRGA2C06	AY747508	5	C	-				
PRGA3T7C06	AY747537	5	C	-				
PRGA3T7H07	AY747555	6	D	07, 08, 09	53	14	67	21

PLTRP1A10	AY747330	6	D1	04, 05, 06	28	9	37	24
PLTRP1C10	AY747402	6	D1	04, 05, 06				24
PLTRP1D12	AY747339	6	D1	04, 05, 06				24
PLTRP2B01	AY747432	6	D1	04, 05, 06				24
PLTRP2C11	AY747429	6	D1	04, 05, 06				24
PLTRP2F09	AY747378	6	D1	04, 05, 06				24
PLTRP2H03	AY747388	6	D1	04, 05, 06				24
PLTRP1A06	AY747405	7	E	12	0	10	10	0
PLTRP1B03	AY747406	7	E1	14, 15	15	56	71	79
PLTRP1C09	AY747408	7	E1	14, 15				79
PLTRP2A01	AY747411	7	E1	14, 15				79
PLTRP2D05	AY747414	7	E1	14, 15				79
PLTRP2E10	AY747416	7	E1	14, 15				79
PLTRP1A01	AY747420	8	E1	14, 15				79
PLTRP1B10	AY747423	8	E1	14, 15				79
PLTRP1C04	AY747395	8	E1	14, 15				79
PLTRP1F03	AY747343	8	E1	14, 15				79
PLTRP1G04	AY747348	8	E1	14, 15				79
PLTRP2B11	AY747413	8	E1	14, 15				79
PLTRP2G12	AY747386	8	E1	14, 15				79
PLTRP2H02	AY747387	8	E1	14, 15				79
PRGA2A07	AY747493	9	F	36, 37, 38	58	103	161	64
PRGA2B11	AY747506	9	F	36, 37, 38				64
PRGA2C07	AY747509	9	F	36, 37, 38				64
PRGA2E03	AY747568	9	F	36, 37, 38				64
PRGA2F11	AY747571	9	F	36, 37, 38				64
PRGA2H01	AY747574	9	F	36, 37, 38				64
PRGA3T7A02	AY747529	9	F	36, 37, 38				64
PRGA3T7A07	AY747530	9	F	36, 37, 38				64
PRGA3T7A10	AY747532	9	F	36, 37, 38				64
PRGA3T7A11	AY747577	9	F	36, 37, 38				64
PRGA3T7B11	AY747535	9	F	36, 37, 38				64
PRGA3T7C07	AY747538	9	F	36, 37, 38				64
PRGA3T7D07	AY747542	9	F	36, 37, 38				64
PRGA3T7D11	AY747579	9	F	36, 37, 38				64
PRGA3T7F10	AY747548	9	F	36, 37, 38				64
PRGA3T7G11	AY747553	9	F	36, 37, 38				64
PRGA3T7H12	AY747585	9	F	36, 37, 38				64
PRGAFA02	AY747556	9	F	36, 37, 38				64
PRGAFA07	AY747558	9	F	36, 37, 38				64
PRGAFA10	AY747560	9	F	36, 37, 38				64
PRGAFA11	AY747586	9	F	36, 37, 38				64
PRGAFB06	AY747563	9	F	36, 37, 38				64
PRGAFD01	AY747565	9	F	36, 37, 38				64
PRGAFH11	AY747588	9	F	36, 37, 38				64
PRGAFH12	AY747589	9	F	36, 37, 38				64
PLTRP1A02	AY747431	10	G	34, 35	7	8	15	53
PLTRP1G11	AY747351	10	G	34, 35				53
PLTRP2A06	AY747360	10	G	34, 35				53
PLTRP2E11	AY747375	10	G	34, 35				53
PLTRP2F12	AY747379	10	G	34, 35				53

PLTRP2G03	AY747381	10	G	34, 35				53
PLTRP2G06	AY747383	10	G	34, 35				53
PLTRP2G11	AY747385	10	G	34, 35				53
PLTRP1C08	AY747396	11	H	16, 17, 18	5	35	40	88
PLTRP1E02	AY747340	11	H	16, 17, 18				88
PLTRP1E05	AY747342	11	H	16, 17, 18				88
PLTRP2G02	AY747380	21	I	5	3	0	3	0
PLTRP1B04	AY747331	12	I1	19, 20, 21	15	76	91	84
PLTRP1B05	AY747400	12	I1	19, 20, 21				84
PLTRP1D09	AY747338	12	I1	19, 20, 21				84
PLTRP1H08	AY747410	12	I1	19, 20, 21				84
PLTRP1H12	AY747358	12	I1	19, 20, 21				84
PLTRP2B05	AY747362	12	I1	19, 20, 21				84
PLTRP2E09	AY747374	12	I1	19, 20, 21				84
PLTRP1H03	AY747428	13	I2	22, 23	0	18	18	0
PLTRP2B09	AY747397	13	I2	22, 23				0
PLTRP2E06	AY747373	13	I2	22, 23				0
PLTRP1C06	AY747424	14	J	25,27				52
PLTRP1G12	AY747352	14	J	25,27	11	12	23	52
PLTRP1H04	AY747354	14	J	25,27				52
PLTRP2B12	AY747399	14	J	25,27				52
PLTRP2C09	AY747422	14	J	25,27				52
PLTRP2D03	AY747369	14	J	25,27				52
PLTRP2E03	AY747372	14	J	25,27				52
PLTRP2E12	AY747417	14	J	25,27				52
PLTRP2H07	AY747389	14	J	28,29,30	25	10	35	29
PLTRP1A08	AY747398	15	J1	31,32,33	24	28	52	54
PLTRP1E12	AY747419	15	J1	31,32,33				54
PLTRP1F12	AY747346	15	J1	31,32,33				54
PLTRP1G08	AY747350	15	J1	31,32,33				54
PLTRP1H07	AY747356	15	J1	31,32,33				54
PLTRP2A04	AY747359	15	J1	31,32,33				54
PLTRP2A09	AY747403	15	J1	31,32,33				54
PLTRP2C03	AY747366	15	J1	31,32,33				54
PLTRP2C12	AY747404	15	J1	31,32,33				54
PLTRP2F05	AY747377	15	J1	31,32,33				54
PLTRP2H05	AY747430	15	J1	31,32,33				54
PRGAFG08	AY747501	15	J1	31,32,33				54
PLTRP2C04	AY747367	16	J1	31,32,33				54
PLTRP2F02	AY747376	16	J1	31,32,33				54
PRGA2B03	AY747505	16	J1	31,32,33				54
PRGA2C10	AY747511	16	J1	31,32,33				54
PRGA2E10	AY747520	16	J1	31,32,33				54
PRGA2G04	AY747525	16	J1	31,32,33				54
PRGA2G09	AY747527	16	J1	31,32,33				54
PRGA3T7C10	AY747540	16	J1	31,32,33				54
PRGA3T7E10	AY747546	16	J1	31,32,33				54
PRGA3T7G04	AY747562	16	J1	31,32,33				54
PRGA3T7G06	AY747551	16	J1	31,32,33				54
PRGA3T7G09	AY747552	16	J1	31,32,33				54
PRGAFG09	AY747587	16	J1	31,32,33				54

PRGA2A09	AY747503	17	J1	31,32,33				54
PRGA3T7B04	AY747567	17	J1	31,32,33				54
PRGAFB04	AY747561	17	J1	31,32,33				54
PRGAFB12	AY747564	17	J1	31,32,33				54
PRGAFG12	AY747502	17	J1	31,32,33				54
PLTRP1B08	AY747407	18	K					
PLTRP1C07	AY747335	18	K	40	1	13	14	93
PLTRP1C12	AY747336	18	K					
PLTRP1G03	AY747347	18	K					
PLTRP1G07	AY747349	18	K	53	0	6	6	100
PLTRP2A07	AY747361	18	K	53				100
PLTRP2G07	AY747384	18	K					
PLTRP2G09	AY747418	18	K	39	1	35	36	97
PRGA2A03	AY747492	18	K					
PRGA3T7A01	AY747575	18	K					
PRGA3T7A03	AY747576	18	K					
PRGA3T7A05	AY747496	18	K					
PRGA3T7F05	AY747499	18	K					
PRGAFA03	AY747557	18	K					
PNTRP1A01	AY747464	19	L	41, 43	8	11	19	58
PNTRP1A03	AY747465	19	L	41, 43				58
PNTRP1A04	AY747466	19	L	41, 43				58
PNTRP1A06	AY747467	19	L	41, 43				58
PNTRP1A08	AY747468	19	L	41, 43				58
PNTRP1A10	AY747469	19	L	41, 43				58
PNTRP1A11	AY747470	19	L	41, 43				58
PNTRP1B01	AY747433	19	L	41, 43				58
PNTRP1B04	AY747434	19	L	41, 43				58
PNTRP1B06	AY747488	19	L	41, 43				58
PNTRP1B07	AY747471	19	L	41, 43				58
PNTRP1B08	AY747435	19	L	41, 43				58
PNTRP1B09	AY747489	19	L	41, 43				58
PNTRP1B12	AY747472	19	L	41, 43				58
PNTRP1C08	AY747436	19	L	41, 43				58
PNTRP1D01	AY747437	19	L	41, 43				58
PNTRP1D07	AY747473	19	L	41, 43				58
PNTRP1E06	AY747474	19	L	41, 43				58
PNTRP1E07	AY747463	19	L	41, 43				58
PNTRP1E10	AY747475	19	L	41, 43				58
PNTRP1E11	AY747438	19	L	41, 43				58
PNTRP1F01	AY747439	19	L	41, 43				58
PNTRP1F10	AY747476	19	L	41, 43				58
PNTRP1G01	AY747440	19	L	41, 43				58
PNTRP1G03	AY747441	19	L	41, 43				58
PNTRP1G04	AY747442	19	L	41, 43				58
PNTRP1G08	AY747443	19	L	41, 43				58
PNTRP1H04	AY747477	19	L	41, 43				58
PNTRP1H11	AY747444	19	L	41, 43				58
PNTRP2A02	AY747478	19	L	41, 43				58
PNTRP2A07	AY747445	19	L	41, 43				58
PNTRP2A08	AY747446	19	L	41, 43				58

PNTRP2A09	AY747479	19	L	41, 43	58
PNTRP2A10	AY747447	19	L	41, 43	58
PNTRP2A11	AY747480	19	L	41, 43	58
PNTRP2B01	AY747448	19	L	41, 43	58
PNTRP2B02	AY747449	19	L	41, 43	58
PNTRP2B03	AY747481	19	L	41, 43	58
PNTRP2B04	AY747450	19	L	41, 43	58
PNTRP2B05	AY747451	19	L	41, 43	58
PNTRP2B09	AY747452	19	L	41, 43	58
PNTRP2C04	AY747453	19	L	41, 43	58
PNTRP2C08	AY747482	19	L	41, 43	58
PNTRP2C09	AY747483	19	L	41, 43	58
PNTRP2C11	AY747484	19	L	41, 43	58
PNTRP2D07	AY747490	19	L	41, 43	58
PNTRP2D08	AY747454	19	L	41, 43	58
PNTRP2D09	AY747485	19	L	41, 43	58
PNTRP2D10	AY747455	19	L	41, 43	58
PNTRP2E08	AY747486	19	L	41, 43	58
PNTRP2E12	AY747491	19	L	41, 43	58
PNTRP2F03	AY747456	19	L	41, 43	58
PNTRP2F06	AY747457	19	L	41, 43	58
PNTRP2F09	AY747487	19	L	41, 43	58
PNTRP2F11	AY747458	19	L	41, 43	58
PNTRP2G06	AY747459	19	L	41, 43	58
PNTRP2G10	AY747460	19	L	41, 43	58
PNTRP2H05	AY747461	19	L	41, 43	58
PNTRP2H12	AY747462	19	L	41, 43	58
PLTRP1B09	AY747393	20			
PLTRP1F10	AY747392	20			
PLTRP1F11	AY747394	20			
PRGA3T7H11	AY747584	20			
PLTRP1C05	AY747390	22			
PLTRP2G04	AY747382	22			
PRGA3T7B05	AY747534	23			
PRGA3T7F01	AY747566	23			
PRGAFA01	AY747500	23			

Footnote:
GI: GenBank accession number
Phylogeny: Phylogenetic associations of the sequences determined by the construction of neighborjoining tree of the nucleotide sequences with MEGA-2 after the multiple-alignment with CLUSTALX
Clades: Clades are same as denoted at Fig 3.8; they encompass the group of sequences that are little more phyletically distant.
Overgos: Overgos specific to each cladistic group; overgo design is expained in the manuscript
Shared: The number of BAC clones that were detected by more than one overgo
Unique: The number of BAC clones that were recognized by a single overgo
Subphyl: Subphyletic association of the sequences within the clades

CHAPTER 4

TOWARD INTEGRATION OF PEANUT GENETIC AND PHYSICAL MAPS, AND EXPLORATION OF POSSIBLE CONTRIBUTIONS FROM *ARABIDOPSIS*[1]

[1] Yuksel, B. and A. H. Paterson. To be submitted to *Theoretical and Applied Genetics*.

ABSTRACT:

Arachis hypogaea is a widely cultivated crop both as an oilseed and protein source. The genomic analysis of *Arachis* species hitherto has been limited to the construction of genetic maps; the most comprehensive one contains 370 loci over 2210 cM length. However, no attempt has made to analyze the physical structure of the peanut genome. To investigate the practicality of physical mapping in peanut, we applied a total of 117 overgos derived from genetically-mapped RFLP probes onto peanut BAC filters containing 182,784 peanut large-insert DNA clones in a multiplex experimental design; 91.5% of the overgos identified at least one BAC clone. In order to gain insights into the potential value of *Arabidopsis* genome sequence for studies in divergent species with complex genomes such as peanut, we employed 576 *Arabidopsis* derived overgos selected on the basis of maximum homology to orthologous sequences in other plant taxa to screen the peanut BAC library. A total of 353 (61.3%) overgos detected at least one peanut BAC clone. This experiment represents the first steps toward the creation of a physical map in peanut; and also illustrates the potential value of leveraging information from distantly-related species such as *Arabidopsis* for both practical applications such as comparative map-based cloning and shedding light on evolutionary relationships. We also evaluated the possible correlation between functional categories of Arabidopsis overgos and their success rates in hybridization to the peanut BAC library.

Keywords: *Arachis hypogaea*, overgos, interspecific hybridization, Fabaceae, Multiplex-Overgo

INTRODUCTION:

Peanut is a widely-grown crop in many areas of the world. According to FAO estimates, peanut production worldwide stands around 37,057,652 (Mt) in 2003 (Food and Agriculture Organization 2003). Almost all species in section *Arachis* of genus *Arachis* are diploids (2n=20) with only two exceptions, *Arachis hypogaea L*, and *Arachis monticola*, which are allotetraploids (2n=4x=40) (Stalker and Dalmacio 1986; Stebbins 1957). The majority of diploid species are classified as A or B genomes; where A genomes differ from B by presence of an easily discernible pair of small chromosomes (Stalker and Dalmacio 1986). The diploid ancestry of amphidiploids in the genus remains to be clarified, but *A. duranesis*, or *A. villosa* as A and *A. ipaensis* as B genome ancestral donors have been usually proposed (Kochert et al. 1991; Kochert et al. 1996; Raina and Mukai 1999; Raina et al. 2001).

The most detailed tetraploid peanut map contains 370 RFLP loci on 23 linkage groups, extending over 2210 cM (Burow et al. 2001). For the majority of important crops and model plants such as soybean and *Medicago*, several BAC libraries have been constructed and employed either in large scale physical analysis or in targeted analysis of particular genomic regions. However, no prior attempts at physical analysis of *A. hypogaea* have been reported.

Two main approaches have been commonly used for screening arrayed libraries with large numbers of clones. One is based on pooling of DNA from multiple clones, and subsequent screening of these pools with site-specific primers. For instance, AP-PCR (Arbitrary Primer-PCR) has been used in identification of 245 BAC clones from a rice BAC library (Xu et al. 1998a). Likewise, Klein et al. (2000) were able to pick out ~2400 BACs from BAC pools of a sorghum library by using AFLP primers. The second approach is the application of radioactively labeled probes to high-density BAC filters. While individual probes such as cDNA or genomic

clones can be used, oligonucleotide based probes, called "overgo", improve screening due to more efficient radiolabeling and low cost. Overgos are two 24 bp oligonucleotides with an 8 bp overlapping region at the 3′ end; thus, allowing the synthesis of complementary strands with radioactive nucleotides (Ross et al. 1999). Multiplexing of overgos enables the hybridization of large numbers of probes in a single experiment. For example, 10,642 overgos designed from ESTs were applied to 165,888 maize BAC in a 24×24×24 experimental setting with an 88% success rate (Gardiner et al. 2004).

Arabidopsis is the most thoroughly studied genome among all flowering plants (angiosperms). Thus, it is likely to serve as an important model for shedding light on many issues related to plant biological processes. It is critical to learn the scope of conservation of gene order and the level of sequence homology between orthologous regions of *A. hypogaea* and *Arabidopsis* in order to make use of *Arabidopsis* genome sequence for exploration of the peanut genome. *Arachis hypogaea* and *Arabidopsis* are thought to have shared a common ancestor about 100-120 mya (million years ago) (Davies et al. 2004). A study conducted by Lee et al. (2001) has shown that homologous regions among *Glycine max*, *Vigna radiata*, and *Phaseolus vulgaris* are well-conserved with one another, and conserved to a lesser degree in *Arabidopsis*. The BAC-end and subclone sequences derived from BAC contigs on a 10-cM region of *Glycine max* linkage group G revealed that 27 out of 78 sequences had noteworthy homology to *Arabidopsis* (Foster-Hartnett et al. 2002). However, the homologs mapped at six different locations in the *Arabidopsis* genome, and occasional sequence deletions were observed.

The main objectives of this study were to explore the feasibility and the practicality of physical mapping approaches for *Arachis hypogaea*, and to investigate the ability to cross-utilize molecular tools between *Arabidopsis* and peanut by applying *Arabidopsis* based sequences to the

peanut large-insert DNA library. We applied 117 overgos designed from the mapped peanut probe sequences, and 576 *Arabidopsis* overgos to a 6.5 genome equivalent peanut BAC library.

MATERIALS AND METHODS:

a) Designing overgos:

In these experiments, two sets of overgos were used. The first 117 overgos were designed from peanut RFLP probe sequences from the *Arachis hypogaea* genetic map (Burow et al. 2001). Peanut probe sequences used were cDNAs selected from either root-or shoot-derived cDNA libraries (originally provided by Dr. Gary Kochert, University of Georgia). The nucleotide sequences were blasted against GenBank databases after masking known repetitive sequences (Altschul et al. 1997). A Microsoft Visual Basic script (written by Dr. John Bowers, University of Georgia) was used for designing overgos. The program chooses the longest and the best conserved stretch of nucleotide sequence as template for designing forward and reverse overgo primers, which are 24 bp in length and overlap by 8 bp at the 3´ ends.

A total of 576 *Arabidopsis* overgos used in this experiment were designed with the same program. The overgo primers were selected from *Arabidopsis* coding sequences by blasting against GenBank Embryophyta sequences. The primer sites were chosen from regions which share the highest homology with orthologous genes from a maximal number of plant species, to maximize the probability of cross-species hybridization

Both sets of primers were synthesized commercially (MWG Biotech, N.C.) and dissolved in ddH_2O at a final concentration of 0.2 nM.

b) Multiplex Hybridization Setup:

A computer program written in Microsoft Visual Basic supported by Microsoft Access, BACMan (http://www.plantgenome.uga.edu/cotton/CottonDBFrames.htm) was used for

multiplex designs. The program assigned 117 POV (peanut overgos) to the hybridizations in such way that each probe intersects at only a single row×column×diagonal point.

For *Arabidopsis* derived probes, a similar experimental design was generated. The multiplexing factor was 24×24×24 (row/column/diagonal) with a total of 576 overgos allocated to 72 hybridizations.

c) Gridding Peanut BAC Library and Preparing High-Density Filters for Hybridization:

A total of 182,784 *Arachis hypogaea* BAC clones were gridded on 22.5 cm^2 Hybond N+ membranes (Amersham Life Sciences, IL, Arlington Heights, USA) with a Q-BOT (Genetix). Each clone was double spotted in 4×4 arrays allowing representation of 18,432 different clones per filter. Thus, the whole peanut library fit in total onto 10 filters. The filters were incubated on 1% LB agar containing 12.5 µg/µl CM, at 37 ºC for 12-18 hrs until optimal colony growth-size was obtained. The high-density BAC filters were processed according to a standard alkaline-lysis method (Sambrook and Russell 2001). The filters were dried overnight and stored at 4 ºC.

d) Overgo Labeling

Overgo labeling reactions were performed in a total volume of 15 µl containing 0.0067 nM forward and reverse oligonucleotide primers (that were denatured at 94 ºC for 5 min then cooled on ice), 1 µg BSA, 2.5 U Taq polymerase, 0.5 µl α^{32}-dATP (6000 Ci/mmol) (MP Biomedicals, CA), 0.5 µl α^{32}-dCTP(6000 Ci/mmol) (MP Biomedicals, CA), and 3 µl OLB (Oligo labeling buffer without dATP, dCTP, and random hexamers (Ross et al. 1999)). The reaction mixtures were incubated at 37 ºC for 2 hours (Fig 4.1). The labeled overgos were combined in the proper column, row, and diagonal pools, then filtered through Sephadex columns to remove unincorporated nucleotide, and heat denatured at 94 ºC for 5 minutes before adding to the proper bottles

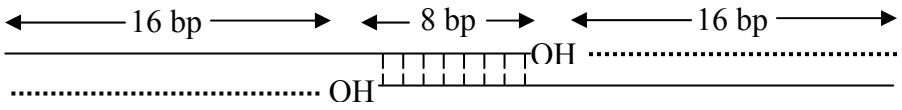


Fig 4.1 Schematic representation of overgo structure and labeling; the dashed lines indicate radiolabeled portion of the probe.

e) Hybridization and Washing

Membranes were separated with nylon mesh, placed into rotisserie bottles with 85 ml of hybridization buffer [0.5 M Sodium Phosphate, pH 7.2, 7% (w/v) SDS, 1 mM EDTA, and 0.01% (w/v) BSA] and incubated at 55 ºC for 4 hours before adding the denatured radioactive overgo probes. Hybridizations were performed at 55 ºC with a constant rotation speed of 4.5 rpm for 18-36 hrs in rotisserie ovens. Following the hybridization process, the membranes were washed at 55 ºC for 30 min each with constant shaking in trays; first at low stringency in wash buffer II [1×SSPE (0.15 M NaCl, 10 mM $NaH_2PO_4.H_2O$, and 1 mM Na_2EDTA, pH 7.4), 1% (w/v) SDS)], then at high stringency in wash buffer III (same as wash buffer II except (0.5×)), and one last wash in wash buffer II in order to remove nonspecific binding to reduce background with minimal effect on the specific probe hybridizations. Afterwards, the membranes were blot dried, wrapped in sheet protectors, and autoradiographed with two intensifying screens (10" × 12" L-Plus, Optonix, NJ) and X-ray film (Blue, Medical X-ray film, SourceOne, CA) for two weeks at 80 ºC before developing.

f) Data Entry and Analysis

The hits on the films were scored manually onto transparent templates, scanned and read by ABBYY FineReader 5.0 with manual checking and correction when this optical character

recognition software was uncertain. The BAC hit scores were converted to the BAC addresses with BACEater (a Microsoft Basic script, James Estill, University of Georgia). Afterwards, the hits were deconvoluted with BACMan assigning each BAC to a specific probe on the basis of common column, row, and diagonal intersections. The hits that matched at least two out of three possible intersections were accepted.

RESULTS:

A. Analysis of Peanut Overgo Hybridization:

From 117 peanut overgos applied to 10 high-density BAC filters containing a total of 182,784 clones, a total of 22,266 data points was scored. The expected number of hits after deconvolution is 7,422; however, the total number of hits that we were able to assign to a specific probe was only 6,663. This discrepancy was mostly due to 'overmatches'; BAC clones that are hit with more than one probe. The experimental design does not allow deconvolution of these hits.

Out of a total number of 6,663 hits successfully ascribed to a specific probe, 2730 were to POVR2012 (i.e. 1.5% of 182,784 peanut BAC clones contained this sequence at least once). In order to find out the reason behind this disproportionately large number of hits to a single probe, we blasted both the original probe and the overgo primer designed from it against various databases including TIGR plant repeat databases. The full R2012 sequence did not have any hits against GenBank. However, POVR2012 had 71% sequence similarity with an *Arabidopsis thaliana* copia-like retrotransposon (gi: 15149815) and even higher similarity with a *cacta*-like transposable element of *Oryza sativa* (78%). Hence, it is very likely that this overgo recognized a repetitive element, which has not been characterized yet, but is present in about 400 copies in the peanut genome. The abundance of this element is not unusual; plants harbor large number of a

diverse group of DNA and RNA elements (Laten et al. 2003; Lee et al. 1990; Manninen and Schulman 1993). To our knowledge, this is the first physically characterized transposable element in *Arachis* spp. These BAC clones could be utilized for further characterization of transposons in the peanut genome assisting in shedding light on peanut genome evolution.

Another overgo with a high number of hits was POVR2545, which had 1277 hits. Both the primer and the full sequence had unique hits to an unknown protein from a *Medicago truncatula* cDNA. Neither of them had similarity to any known repetitive sequence. Thus, if we assume the coverage of the peanut library is about 6.5 genome-equivalents, the copy number of this gene or gene family would be about 193 copies.

Another overgo with a high number of hits, 501, was POVR2022. This sequence had high homology with the 3-*deoxy-D-arabino-heptulosonate 7-phosphate synthase* gene family, which has orthologs in both plants and bacteria and represents a large family of genes (Herrmann and Weaver 1999). This gene encodes the first enzyme in the shikimate pathway, which synthesizes precursors for aromatic secondary metabolites and amino acids Phe, Tyr, and Trp (Entus et al. 2002).The *Arabidopsis* genome contains three copies of this gene. Thus, this gene family may either exist in higher copy numbers in the peanut genome, or the overgo sequence might include a motif which is shared by additional gene families.

POVR2085 identified about 325 BACs. However, neither the overgo nor the full sequence had significant homology to any known sequence or repetitive element. Thus, this cDNA derived from peanut roots could be a peanut specific multiple copy gene (i.e. ~40 copies for the tetraploid genome).

Another overgo that possibly targeted a multigene family, trypsin proteinase inhibitors, was R2403 (179 hits) (Supplementary table 4.1). The trypsin inhibitors, induced as a result of wound

response are a multicopy gene family; for instance, the soybean genome contains 10 copies, mostly in tandem repeats (Jofuku and Goldberg 1989). A similar tandemly-repeated distribution of these genes has been also detected in *Arabidopsis thaliana* (Clauss and Mitchell-Olds 2004).

Three overgos, POVS063, POVR116, and, POVR177, with 72-149 BACs consistent with copy numbers of 10 to 20, did not share sequence similarity with any other known proteins. Five overgos, POVS26, POVR2091, POVR2110, POVR230, POVR190, POVR2405, and POVR2482 had 33-67 hits consistent with copy numbers of 5-10 in the tetraploid genome (Supplementary. table 4.1). Some blast hit results were supportive of the copy number assessment; for example, POVR2482 and POVR230 shared high homology with ribosomal subunit 60S proteins, L10, and S26 respectively, and *Arabidopsis* contains 3-4 copies of each family from these genes (Barakat et al. 2001) (Supplementary table 4.2).

A total of 41 probes had 9 to 28 hits, consistent with copy numbers of 2 to 4 in the tetraploid genome of peanut. The overgos in this category are likely to be specific to a single locus (i.e. two copies in the allopolyploid genome). If we assume that the number of hits to the library will follow a Poisson distribution, 90% of the numbers of hits per overgo would be expected to fall in this range with the assumption of 6.5 genome equivalence of the peanut BAC library, which means 13 hits on average per overgo.

A total of 50 overgo probes had eight or fewer hits, and 10 probes failed to identify any positive BACs. Among these, six had significant overmatch hits (i.e. ~12%). The overmatch hit problem stems from either sequence similarity between different overgos or their recognition of the same or neighboring loci in the peanut genome; for instance, overgos POVR2496 and POVR2497 mapped to the same locus on the genetic map and they had significant overmatch hits (Burow et al. 2001). The remaining probes could be from underrepresented regions of the

peanut genome in our library due to biased distribution of *Hin*dIII restriction sites, or experimental factors such as overgo secondary structures, GC content . It is also plausible that the sequence divergence between some homeolog pairs (i.e. alloalleles) is high or even that one copy of duplicated genes has been deleted, preventing cross-hybridization. Prior experience has shown that more than 3-4 bp difference out in a 40 bp long overgo is sufficient to prevent the identification of the target sequence.

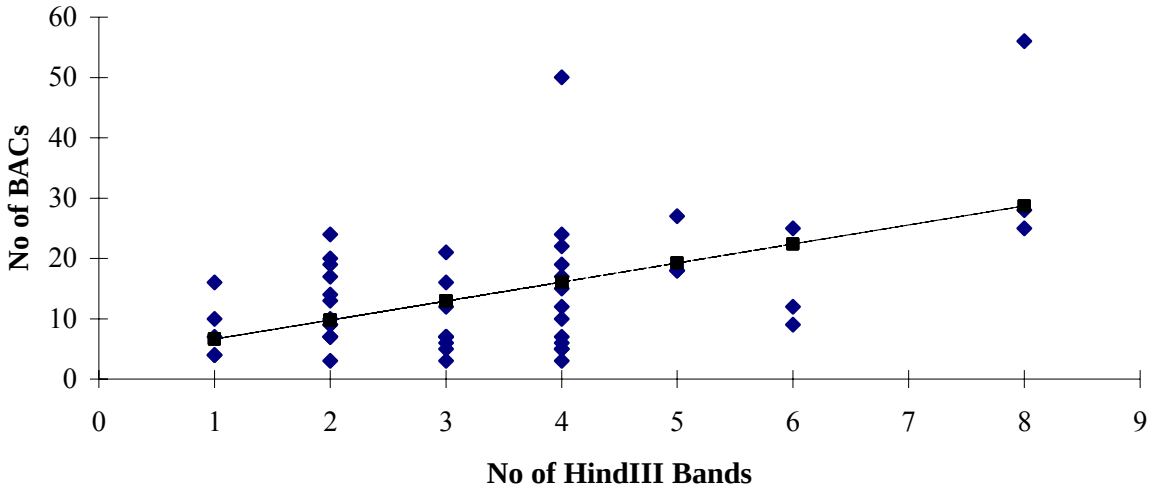


Fig 4.2 Linear regression between number of RFLP bands and number of BAC hits. Probes that had less than three BAC hits are omitted due to high possibility of experimental error stemming from scoring (overmatches). Similarly, overgo probes that have more then 59 hits are also not included.

Fig 4.2 depicts the correlation between the copy number of probes in the peanut genome as estimated by number of *Hin*dIII RFLP bands corresponding to *Arachis hypogaea* on peanut mapping blots, and the number of positive BAC hits generated by overgo primers to corresponding to RFLP probes. The relationship between these two variables (r=0.52) is

statistically significant (p=0.00013). Overgos hybridizing to more than 59 BACs were omitted in this analysis, due to the inability to be confident of resolving 10 or more RFLP bands. As discussed in detail in previous paragraphs of this article, most of the probes which had less then six hits were a result of genetic proximity or sequence similarity. Likewise, some probes that had unexpectedly high numbers of positives were inadvertently designed from highly repetitive domains of cDNA sequences. Overall, the band numbers were good predictors for the number of positive BACs.

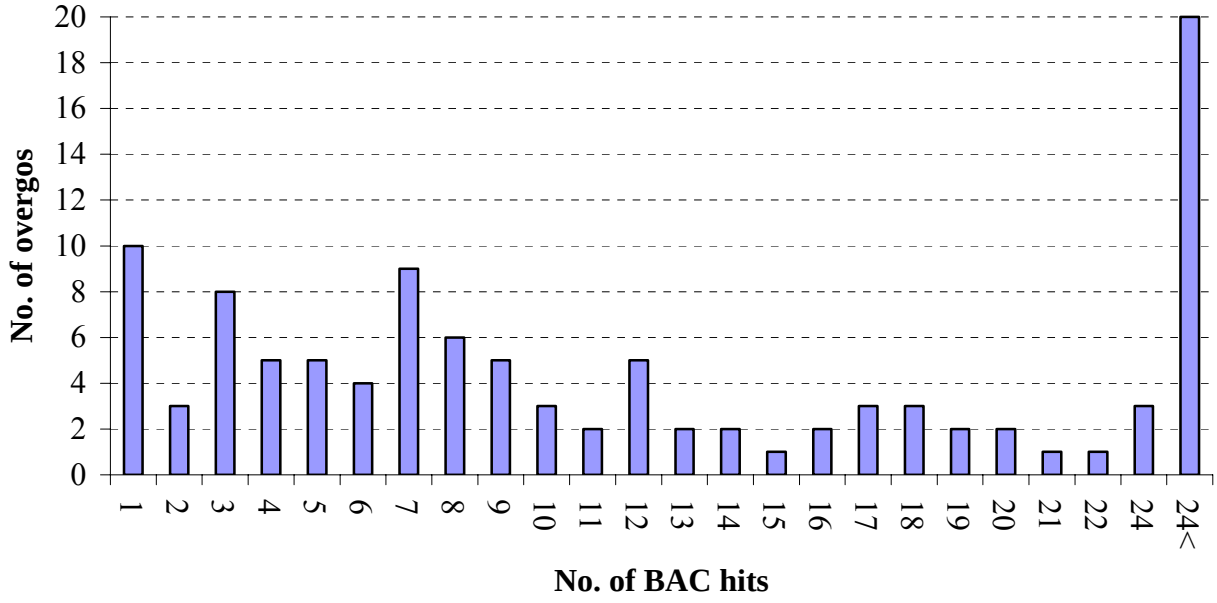


Fig 4.3 The graph denotes the distribution of positive BAC hits for 107 overgos. The Y axis is for number of overgos for particular number of BAC hits, which is demonstrated on the X axis.

Supplementary Table 1. Peanut *Hin*dIII BAC library hybridization results. A total of 10 high density filters were used for hybridization of 117 peanut overgos, allowing screening of 6.5 genome equivalent coverage of the peanut genome

Overgo ID [a]	No. of BAC hits	No of RFLP bands *	Overgo ID [a]	No. of BAC hits	No. of RFLP bands *
POVR32	24	2	POVR255	12	-
POVR72	5	-	POVR256	6	-
POVR74	10	2	POVR258	7	3
POVR78	5	-	POVR2002	7	4
POVR79	8	-	POVR2003	8	-
POVR94	4	-	POVR2005	16	3
POVR111	3	-	POVR2006	12	6
POVR113	3	-	POVR2007	6	4
POVR115	11	-	POVR2008	11	-
POVR116	119	-	POVR2012	2730	-
POVR137	4	-	POVR2013	7	1
POVR140	7	-	POVR2015	5	4
POVR142	1	-	POVR2016	18	5
POVR168	12	-	POVR2022	501	-
POVR177	144	-	POVR2027	9	2
POVR183	9	-	POVR2029	14	-
POVR185	18	-	POVR2031	8	-
POVR190	60	2	POVR2035	20	2
POVR193	1	-	POVR2045	1	4
POVR194	3	-	POVR2046	2	2
POVR199	6	-	POVR2056	1	-
POVR211	10	4	POVR2059	25	8
POVR228	1	3	POVR2060	3	4
POVR230	56	8	POVR2061	9	2
POVR237	1	-	POVR2064	19	4
POVR249	2	4	POVR2067	22	4
POVR251	28	8	POVR2069	14	2
POVR254	4	-	POVR2071	5	4
POVR2072	7	2	POVR2448	1	3
POVR2079	17	4	POVR2452	17	2

Overgo ID [a]	No. of BAC hits	No of RFLP bands *	Overgo ID [a]	No. of BAC hits	No. of RFLP bands *
POVR2080	13	2	POVR2462	27	5
POVR2082	13	-	POVR2463	4	1
POVR2085	325	1	POVR2469	25	6
POVR2087	7	3	POVR2473	3	2
POVR2088	3	1	POVR2482	67	4
POVR2090	12	3	POVR2487	6	3
POVR2091	47	-	POVR2496	1	-
POVR2095	8	-	POVR2497	17	-
POVR2098	2	2	POVR2514	19	2
POVR2099	4	1	POVR2541	9	-
POVR2100	24	4	POVR2545	1277	5
POVR2101	12	4	POVR2579	7	3
POVR2106	60	3	POVR2609	10	1
POVR2110	50	4	POVS8	1	-
POVR2403	179	5	POVS23	20	-
POVR2405	62	3	POVS26	33	-
POVR2408	7	2	POVS29	8	-
POVR2409	3	-	POVS43	24	-
POVR2419	16	1	POVS63	72	-
POVR2421	7	2	POVS65	8	-
POVR2425	5	3			
POVR2434	1	1			
POVR2435	21	3			
POVR2437	15	4			
POVR2439	18	5			
POVR2440	3	3			
POVR2443	9	6			

- Data was not available for these probes

* Data was obtained from mapping survey blots (Burow et al. 2001).

[a] Overgo primers designed from peanut RFLP probe sequences (Burow et al. 2001).

Supplementary Table 2. Blast hits of some peanut RFLP probes, from which overgo primers were designed. The original probes were used in peanut genetic mapping

ProbeId[a]	GI[b]	Homology	E-Value
R032	1208536	Guanine nucleotide regulatory protein (rab2)	3.0e-56
R111	6492269	Nuclear RNA binding protein C (RggC) mRNA	2.0e-21
R113	467564	Lipoxygenase	2.00e-19
R115	18678	Lipoxygenase-3 (EC 1.13.11.12)	8.00e-20
R140	535763	Peroxidase	e-122
R190	30681682	Phosphate translocator-related (At3g11320)	9.0 e-39
R199	951115	Mannose/glucose-binding lectin precursor (lec)	0
R2002	3176097	Annexin	3.0e-10
R2007	13194620	Brassinosteroid biosynthetic protein LKB	e-107
R2008	6006798	Nod factor binding lectin-nucleotide phosphohydrolase	1.0e-59
R2015	2822482	14-3-3 protein homolog	0
R2016	30696258	40S ribosomal protein S2 (RPS2A)	9.0e-76
R2022	37359243	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	1.0e-43
R2027	2558961	Histone H2B1	1.0 e-96
R2034	37359243	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	e-105
R2043	559556	Arabinogalactan-protein (AGP)	6.0e-06
R2046	17736863	Lotus corniculatus var. japonicus genomic DNA	4.0e-61
R2058	15214409	Annexin (AnnGh1)	6.0e-43
R2059	42472383	cDNA of Flowers and buds	2.0e-14
R2061	23197835	Putative protein (At4g27680)	3.0e-16
R2064	457569	Endo-xyloglucan transferase	3.0e-66
R2067	17736884	*Lotus corniculatus* genomic chromosome 6	9.0e-14
R2070	42469547	cDNA sequence of Flowers and buds	1.0e-09
R2071	927382	Elongation factor-1 alpha	3.0e-94
R2079	5031278	Porin (mPOR)	4.0e-19
R2083	18407284	ADP-ribosylation factor 1 (ARF1)	e-129
R2087	395071	Guanine nucleotide regulatory protein	e-110
R2088	41072352	Calmodulin (CaM1)	0
R2090	3129951	Copper amine oxidase	7.0e-48
R2091	34541965	Actin	1.0e-83
R2095	553037	ADP-ribosylation factor	e-117
R2098	16757965	Amygdalin hydrolase isoform AH I precursor (AH1)	8.0e-11

ProbeId[a]	GI[b]	Homology	E-Value
R2099	9621902	Phosphatidic acid phosphatase alpha	9.0e-33
R2100	479144	ATP synthase subunit	1.0e-68
R2101	479089	ATP-sulfate adenyltransferase	7.0e-42
R2110	42491422	*Medicago truncatula* clone mth2-15b16	5.0e-12
R221	17352484	Drought-induced protein RPR-10	2.0e-17
R228	42561754	NOL1/NOP2/sun family protein (At1g06560)	2.0e-15
R230	5706703	Ribosomal protein S26 (RPS26)	6.0e-98
R2403	9367041	Trypsin inhibitor	8.0e-11
R2405	1196896	Acidic ribosomal protein P0	e-177
R2408	19612	Histone H3	1.0e-68
R2425	3818415	Proline-rich cell wall protein	3.0e-13
R2432	1196896	Acidic ribosomal protein P0	0
R2433	1196896	Acidic ribosomal protein P0	e-168
R2434	18138052	Tropinone reductase I (trI gene)	2.0e-09
R2437	849135	Vacuolar H+-ATPase subunit A	2.0e-07
R2445	161369	Histone H2B-2	8.0e-13
R2448	29290711	Vacuolar ATPase subunit E	e-139
R2452	41688363	*Lotus corniculatus* DNA chromosome 1	8.0e-29
R2469	19032303	Succinate dehydrogenase subunit 4	1.0e-31
R2482	5917742	60S ribosomal protein L10 (QM)	2.0e-92
R2487	16416391	Glutathione S-transferase	8.0e-36
R2496	29838628	Alkaline alpha galactosidase I (aga1)	2.0e-45
R2514	6009908	Histone H2A-like protein	3.0e-57
R2541	18146785	CYP1	e-116
R2579	21207472	*Zea mays* PCO081620	1.0e-12
R258	21207747	*Zea mays* PCO092238	2.0e-20
S023	1236950	Nucleoside diphosphate kinase	2.0e-79
S043	20149293	Phenylalanine ammonia-lyase (PAL)	9.0e-19

[a] The probe names same as overgo names, except prefix "POV", which stands for "Peanut Overgo

[b] GenBank Id

B Application of *Arabidopsis* Derived Overgos to Peanut

A total of 576 overgos designed from coding sequences of *Arabidopsis thaliana* that are well-conserved in other plant species have been applied to the peanut BAC library. A total of 19,448 data points were obtained in a multiplex experimental design of 24×24×24 (R×C×D). The built in triple redundancy greatly increases robustness of the data, minimizing errors in ascribing addresses. After deconvolution, 5,434 BAC hits were ascribed to a total of 353 overgos (Table 4.3 and Fig 4.4). The average number of hits per probe was 15.4, slightly higher than the expected 13 hits per probe average for 6.5 genome equivalent library coverage. Thus, the total success rate for this hybridization experiment was about 61.3%. For the remaining 223 overgos that failed to identify any positives in the peanut BAC library, the most likely explanation is that similarity between the A*rabidopsis* sequence and the peanut orthologous region was not sufficient. A few could also have been caused by lack of coverage or experimental errors.

Table 4.3. Hybridization of 576 *Arabidopsis* overgos to a 6.5 × genome-equivalent peanut BAC library

No. of hits	**No. of overgos**	**Percentage (%)**
1-5	126	35.7
6-24	174	49.3
25-60	38	10.8
61-100	9	2.5
>100	6	1.7
Total no. of hits=5,434	Total no of overgos=353	

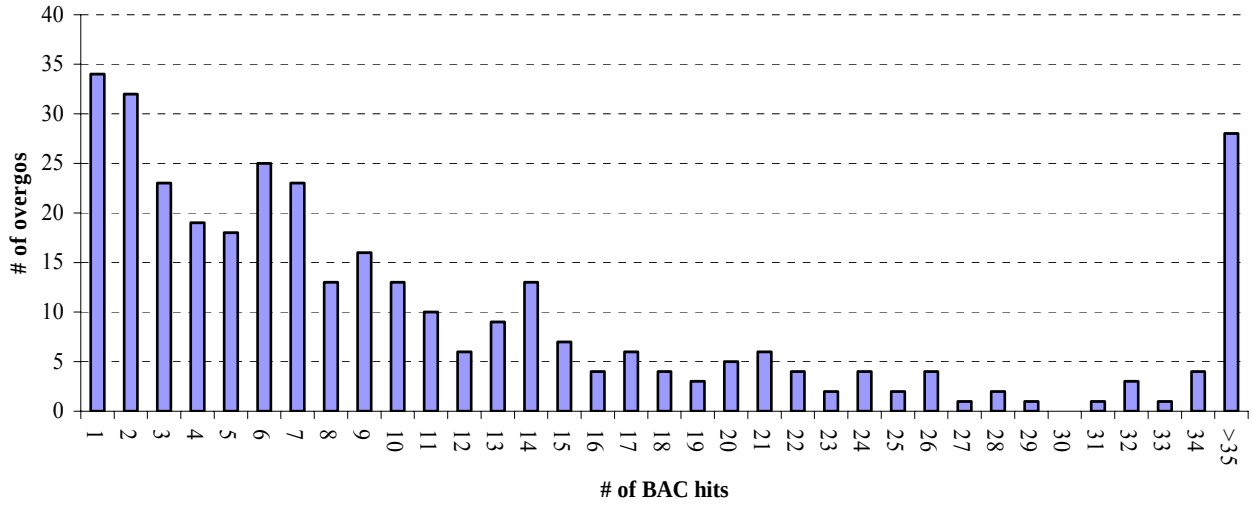


Fig 4.4 The distribution of BAC hits for 353 *Arabidopsis* overgos, which had more than one hit in hybridization. The X axis denotes the number of BAC hits for individual overgo probes, which is shown on the Y axis.

To gain insight into whether there is any effect of the putative functional category of *Arabidopsis* overgo in its recognition of positives in peanut BACs, we blasted the overgos against GenBank and sorted according to their blast hits (Table 4.4). The χ^2 test was applied to check the statistical significance of categorical differences of overgos with respect to their success in identification of positives in a peanut BAC library; only the overgos sharing significant homology with genes involved in energy and translation succeeded more often than expected at p=0.05. However, it should be noted that there were many sequences (151) which did not have any known function. If these sequences were able to be functionally categorized, the statistically significant deviation from the expected values for the categories may have changed.

Table 4.4 Categorization of *Arabidopsis* overgos according to their sequence similarities and BAC hits

Putative functional category [a]	No. of Overgos	No. of Overgos with no hits[b]	% success
Energy	24	4	[c]83.3
Translation	51	12	[c]76.5
Transcription factors	35	9	74.3
Intracellular trafficking	26	8	69.2
Metabolism	80	28	65.0
Cell division and DNA synthesis	19	8	57.9
Other	86	37	57.0
Transport facilitation	18	8	55.6
Unknown	151	67	55.6
Posttranslational modifications	33	16	51.5
total	**576**	**223**	**61.3**

[a] The categorization is based on the blast hits to GenBank. Overgo sequences that matched more than 34 bp (34/40) to any known gene are allocated in the corresponding functional category on the basis of UNIGENE categorization (NCBI)
[b] Overgos that had at least one hit were considered
[c] The observed number of failed and successful overgos was significantly different than the expected values for these categories at $p \leq 0.03$ (χ^2 test)

The 576 overgos employed in this experiment were distributed an average 45 genes apart in the 33 α-duplicated segments (as shown by Bowers et al. 2003) of the *Arabidopsis* genome. We tried to assess the scope of synteny between *Arabidopsis* and peanut by comparing the positions of *Arabidopsis* derived overgos that detected the same BAC clones with both modern and pre-α-duplication gene orders. Out of a total of 54 pairs of overgos that detected the same BAC clone in the peanut library, only 7 pairs (12.9%) were also located in the same α-duplicated segments. There were two pairs whose template sequences were each one gene apart from each other, and the remaining five were separated by an average of 614 genes. The gene density in *Arabidopsis*

is on average one every 5 kb, which could mean one gene for every 50 kb in peanut (the allotetraploid peanut genome is about 20 times larger than *Arabidopsis* genome but contains two copies of every gene). This seems to suggest localized rearrangements of these regions between peanut and *Arabidopsis*.

DISCUSSION:

Of a total of 117 peanut probes applied to the peanut BAC library 107 had one or more hits in the library (Supplementary Table 4.1), a 91.5% success rate. The success rate would have been better if there had not been the unexpected discovery of a transposon, and associated problems with probes that had overmatches. Nonetheless, this percentage is comparable with the results obtained for overgo hybridization data from maize, 88%, (Gardiner et al. 2004), mice ,92%, (Cai et al. 1998), and human , 91% (Han et al. 2000). Thus, the probes that have similar sequences or mapped in close proximity on the genetic map somewhat impeded the experimental results. About 38% of overgos had fewer than 6 hits; the largest proportion of them (10) had only one hit (Fig 4.3). About 48% of probes identified 6 to 24 hits, and only 17% had more than 24 hits (Fig 4.3).The average number of hits was 17.8 after the exclusion of the probes, POVR2085, POVR2022, POVR2545 and POVR2012, which had 325, 501, 1277, and 2730 respectively, supporting a typical organization of 2 to 4 copies per probe in the peanut genome. In conclusion, the number of positives was compatible with the postulated 6.5 times genome coverage for peanut BAC library, perhaps even suggesting that the library is slightly higher coverage.

In this study, a through investigation about the practicality of multiplex-overgo-based experimental approaches for integration of peanut physical and genetic maps has been conducted. The multiplex nature of the experiment permitted the application of multiple probes in a single experiment, significantly diminishing laboriousness of the screening process. The

experimental errors inherent to the design were not significant; only 8.5% of the probes failed. The only drawback of this experimental design was occasional multiple locus hits by overgos with close homology to multiple gene families or repetitive elements. With additional information about the underlying genome, such as sequences representing the complexity of repetitive DNA fractions (Peterson et al. 2002), this setback can be overcome easily by scanning overgos for repetitive sequences. Overall, this approach will be very useful for large scale library screening in a manner efficient of both time and labor.

The overall success rate, 61.3% (353 of 576 overgos) for *Arabidopsis* overgos, is encouraging in terms of developing consensus probes that could work for identification of orthologous regions across distant taxa. This also denotes that these types of probes have great potential for targeting particular regions of interest in large genomes such as peanut for purposes such as chromosome walking or marker enrichment.

The inflated average number of hits per *Arabidopsis* overgo, 17.4, relative to the expected number of hits, 13, could be because of multiple loci recognized by some overgos in peanut. The highest number of hits to a single overgo was 281, and 6 overgos had more than 100 hits (Table 4.3 and Fig 4.4). Therefore, the copy number of these probes per *Arachis hypogaea* genome may range between 15 and 45 copies. It is not rare for overgo primers to identify multiple loci, especially those designed from evolutionarily preserved regions of proteins. About half of the overgo primers, 174 (49%), had 6 to 24 hits, which is equivalent to one to four loci per peanut genome. A total of 126 (21.8%) overgos had less than six hits perhaps corresponding to regions that are underrepresented in the peanut *Hin*dIII BAC library due to bias in restriction sites. Diploidization of loci, or experimental errors such as scoring errors could also contribute to this category.

Almost all overgos which shared strong homology to energy production and conversion genes were successfully hybridized to peanut BACs consistent with conservation of basic cellular machinery between peanut and *Arabidopsis*. On the other hand, *Arabidopsis* overgos that had similarity to genes in signal transduction and posttranslational modifications failed to hybridize to the peanut BAC library especially frequently (Table 4.4). This result was not surprising since genes in this category are more likely to function in providing adaptation to specific environmental conditions. Overgos designed from conserved regions of transcription factor families were more successful than average. Thus, the functional motifs of the transcription factor families such as "myb" genes, "MADs" box etc are preserved due to evolutionary constraints. However, an overgo, AOVG0576, with strong homology to one functional motif, leucine-rich repeats found in *Cf* like disease resistance genes, failed to identify any positives, which may have been caused by hypervariability of these regions (Dixon et al. 1998) On the contrary, an overgo, AOVG0168, with high homology (40bp/40bp match) to tyrosine phosphatase family genes, which are found both in animal and higher plant genomes and play roles in transduction of signals from a wide range of environmental cues including growth factors, cytokines, and hormones (Fordham-Skelton et al. 1999; Xu et al. 1998b), yielded 111 positives. About 35% of overgos in the functional category of metabolism failed, supporting presumed involvement of some of these gene products in synthesis of secondary metabolites, possibly not present in peanut. Even though the majority of overgos similar to enzymes associated with translation machinery (including some ribosomal proteins) produced hits, a certain percentage (23%) did not.

The general conclusion we reached was that the success of any overgo-based interspecific hybridization reaction appears to be related to the functional identity of a sequence and also to

the nature of the domain from which the overgo is designed. In other words, overgos that are based on genes encoding basic cellular machinery or functionally significant common motifs have a higher chance of success, unless these motifs are under diversifying selection such as LRR repeats of some disease resistance genes.

Nevertheless, we were able to demonstrate the preservation of a sufficient level of sequence similarity between a significant number of peanut and *Arabidopsis* orthologs for practical utilization in the peanut genome. This study has also shown that the *Arabidopsis* genome sequence offers significant value for analysis of the peanut genome.

To further assess the value of *Arabidopsis* genome sequence for comparative studies with the peanut genome, we attempted to measure the degree of synteny by comparing the locations of overgos with the hits to the same peanut BAC clones. The comparison was done with the pre-α-duplication gene order (Bowers et al. 2003) of the *Arabidopsis* genome with the hope of observing a higher degree of syntenic association between these two species. However, only 12.9% of the overgos colocalized at the same BAC clone were also located in the same α-duplicated regions. Previous studies between other legume species and *Arabidopsis* have also shown that the scope of colinearity was very limited. Zhu et al. (2003) found only limited microsynteny between *Arabidopsis* and *Medicago* after comparing 82 genetic markers and BAC sequences. A similar observation made by Yan et al. (2003) between soybean and *Arabidopsis* genomes; only 14% of 50 soybean contigs had syntenic relationships, where markers were located less than <100 kb apart. However, much better syntenic relationships between the aforementioned legume species and *Arabidopsis* may have been observed if α-duplication gene order was used. Our result is in concordance with these observations. While the scope of this study is not sufficient for a conclusive assessment of syntenic relationship between *Arabidopsis*

and peanut, an appreciable degree of gene sequence conservation will foster future more definitive studies.

REFERENCES:

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl. Acids. Res. 25: 3389-3402

Barakat A, Szick-Miranda K, Chang I-F, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. Plant Physiol. 127: 398-415

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422: 433-438

Burow 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

Cai W-W, Reneker J, Chow C-W, Vaishnav M, Bradley A (1998) An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. Genomics 54: 387-397

Clauss MJ, Mitchell-Olds T (2004) Functional divergence in tandemly duplicated *Arabidopsis thaliana* trypsin inhibitor genes. Genetics 166: 1419-1436

Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V (2004) Darwin's abominable mystery: Insights from a supertree of the angiosperms. Proc. Natl. Acad. Sci 101: 1904-1909

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JDG (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell 10: 1915-1926

Entus R, Poling M, Herrmann KM (2002) Redox regulation of *Arabidopsis* 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. Plant Physiol. 129: 1866-1871

Food and Agriculture Organization (2003) FAOSTAT Agriculture. Available at: http://apps.fao.org/page/collections?subset=agriculture; verified 10 February 2004.

Fordham-Skelton AP, Skipsey M, Eveans IM, Edwards R, Gatehouse JA (1999) Higher plant tyrosine-specific protein phosphatase (PTPs) contain novel amino-terminal domains: expression during embryogenesis. Plant Mol Biol 39: 593-605

Foster-Hartnett D, Mudge J, Larsen D, Danesh D, Yan H, Denny R, Penuela S, Young N (2002) Comparative genomic analysis of sequences sampled from a small region on soybean (*Glycine max*) molecular linkage group G. Genome 45: 634-645

Gardiner J, Schroeder S, Polacco ML, Sanchez-Villeda H, Fang Z, Morgante M, Landewe T, Fengler K, Useche F, Hanafey M, Tingey S, Chou H, Wing R, Soderlund C, Coe EH, Jr. (2004) Anchoring 9,371 Maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol. 134: 1317-1326

Han CS, Sutherland RD, Jewett PB, Campbell ML, Meincke LJ, Tesmer JG, Mundt MO, Fawcett JJ, Kim U-J, Deaven LL, Doggett NA (2000) Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. Genome Res. 10: 714-721

Herrmann KM, Weaver LM (1999) The shikimate pathway. Annual Review of Plant Physiology and Plant Molecular Biology 50: 473-503

Jofuku KD, Goldberg RB (1989) Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. Plant Cell 1: 1079-1093

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res. 10: 789-807

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

Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Amer. Jour. Bot. 83: 1282-1291

Laten HM, Havecker ER, Farmer LM, Voytas DF (2003) SIRE1, an endogenous retrovirus family from *Glycine max*, is highly homogeneous and evolutionarily young. Mol Biol Evol 20: 1222-1230

Lee DL, Ellis TH, Turner L, Hellens RP, Cleary WG (1990) A copia-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis. Plant Mol Biol 15: 707-722

Lee J, Grant D, Vallejos C, Shoemaker R (2001) Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes. Theor App Genet 103: 765-773

Manninen I, Schulman AH (1993) *BARE-1*, a copia-like retroelement in barley (*Hordeum vulgare L.*). Plant Mol Biol 22: 829-846

Peterson DG, Wessler SR, Paterson AH (2002) Efficient capture of unique sequences from eukaryotic genomes. Trends in Genetics 18: 547-550

Raina SN, Mukai Y (1999) Genomic in situ hybridization in *Arachis* (Fabaceae) identifies the diploid wild progenitors of cultivated (*A-hypogaea*) and related wild (*A-monticola*) peanut species. Plant Systematics and Evol. 214: 251-262

Raina SN, Rani V, Kojima T, Ogihara Y, Singh KP, Devarumath RM (2001) RAPD and ISSR fingerprintings as useful genetic markers for analysis of genetic diversity , varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. Genome 44: 763-772

Ross MT, LaBrie T, McPherson J, Stanton VM (1999) Screening large-insert libraries by hybridization. Wiley, New York, NY

Sambrook J, Russell DW (2001) Molecular Cloning: A laboratory manual. vol 1-3. Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York

Stalker H, Dalmacio R (1986) Karyotype analysis and relationships among varieties of *Arachis hypogaea* L. Cytologia 51: 617-629

Stebbins GL (1957) Genetics, evolution, and plant breeding. Indian J. Genet. Plant Breed 17: 129-141

Xu J, Daichang Y, Domingo J, Ni J, Huang N (1998a) Screening for overlapping bacterial artificial chromosome clones by PCR analysis with an arbitrary primer. Proc. Natl. Acad. Sci 95: 5661-5666

Xu Q, Fu H-H, Gupta R, Luan S (1998b) Molecular Characterization of a tyrosine-specific protein phosphatase encoded by a stress-responsive gene in *Arabidopsis*. Plant Cell 10: 849-858

Yan HH, Mudge J, Kim DJ, Larsen D, Shoemaker RC, Cook DR, Young ND (2003) Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*. Theor App Genet 106: 1256-1265

Zhu H, Kim D-J, Baek J-M, Choi H-K, Ellis LC, Kuester H, McCombie WR, Peng H-M, Cook DR (2003) Syntenic Relationships between *Medicago truncatula* and *Arabidopsis* reveal extensive divergence of genome organization. Plant Physiol. 131: 1018-1026

CHAPTER 5

CONCLUSION:

This study is mainly composed of three sections: 1) Construction and characterization of a large-insert DNA library for *Arachis hypogaea* cv. Florunner; 2) A comprehensive analysis of resistance gene-like sequences including their diversity, evolution and spatial distribution in the peanut genome; and 3) An initiatory work towards physical analysis of the peanut genome. The peanut BAC library formed the basis for the sections 2 and 3.

We constructed and characterized the first *Arachis hypogaea Hin*dIII BAC library to open new avenues towards better understanding physical structure of the peanut genome and to create a new resource that could be utilized for practical purposes such as map-based cloning. Our library contains a total of 182,784 clones with 104 kb average insert size covering peanut genome 6.5 times. The depth of coverage is confirmed to be 6 to 9.5 genome equivalence by application of 24 overgos derived from genetically localized cDNA sequences. To further characterize the library, we applied sorghum rDNA derived overgos, and determined that about 0.66% of the clones contain 45S rDNA clusters. Organelle DNA contamination has been problem in BAC libraries, the percentage of clones carrying chloroplast DNA in our library was insignificant, only nine clones, which was probably result of the technique applied in isolation of HMW DNA (high molecular weight). In conclusion, we believe that this library will serve to the peanut genomic studies in the future.

We explored a hybridization-based approach for determination of the subgenomic identity of the BAC clones recognized by multiple probes. For that purpose, we selected 94 BAC clones

identified by seven different overgos; the BAC clones were fingerprinted and rehybridized with the respective overgos. The contig associations of the clones were in concordance (71% of the time) with the banding pattern observed on the Southern blots. We concluded that this approach could be adapted to high throughput applications and further improved by increasing the gel resolution. In conclusion, the approach tested in this paper will assist in resolving ambiguities stemming from the multiplicity of the probe hits during the integration of physical and genetic maps, especially in polyploids.

The scarcity of genetic polymorphism in monophyletic amphidiploid peanut makes it especially vulnerable to variety of pathogens including bacterial, viral, fungal, and nematode endangering the future of the peanut cultivation. This fact calls for better understanding the diversity, evolution, and spatial distribution of disease resistance gene-like sequences in the peanut genome. The majority of resistance genes contain highly-conserved motifs allowing to design of degenerate primers for amplification of the similar sequences from other plant species. In this study, we obtained 234 resistance gene-like sequences by employing 6 different primer sets (4 degenerate and two site-specific). These RGAs could be utilized for mapping and localization of disease resistance genes in the peanut genome because it is a known fact that most resistance genes form clusters. We multiple aligned our sequences with the orthologs from other plant taxa and constructed phylogenetic trees to gain insights about the evolutionary history. The phylogenetic analysis revealed similar results observed in other plant species. That is, the resistance gene-like sequences are clustered in two main groups, TIR and NonTIR. These two main groups has followed different evolutionary paths resulting in distinctive phyletic makeup, which is more closely associated sequences in the TIR, while much more divergent sequences in the NonTIR part; however, there were some exceptions. To gain insights about the genomic

organization of RGAs, we employed 56 different overgos derived from these sequences on the basis of phyletic association, and identified 736 BAC clones. These BAC clones were fingerprinted, and all the fingerprinting gels were blotted and rehybridized with the corresponding overgos. The clustering characteristics of these sequences were confirmed as a result of these experiments. There was loose correlation between the phyletic association of the sequences and their physical distribution in the peanut genome; that is, the sequences that were phyletically proximal were usually clustered together in the peanut genome. Two main groups of resistance genes, TIR and NonTIR, usually did not colocalize in the same cluster corroborating the observations in other plant species such as *Medicago truncatula*. The previous assessments were supported by both the contig association of the clones carrying these sequences and the hybridization pattern of the overgos derived from these sequences. As a result of these studies, we estimated about 250 putative resistance gene loci including both singletons and clusters. The BAC clones identified in this experiment will assist in better understanding of these sequences in the peanut genome and will serve as an important resource for practical applications such as map-based cloning.

The last part of our study focused on testing multiplex-based library screening technique for the purpose of creating initial steps toward integration peanut physical and genetic maps. We applied 117 overgos derived from genetically-localized cDNA sequences with 91.5% success, which is comparable with the results obtained from other plant species such as maize. This experiment demonstrated the feasibility and practicality of multiplex-overgo-based approach for screening of the peanut BAC library despite the observation of disproportionate number of hits to some overgos due to their repetitive natures. However, these setbacks could be easily surmounted by better screening the overgos.

Arabidopsis is the most thoroughly studied plant genome making it the best candidate for resolving issues related to plant biological processes. Thus, gaining insights about the degree of sequence conservation between *Arabidopsis* and *Arachis hypogaea* is significant for the peanut genome analysis. For that purpose, we designed 576 overgos from *Arabidopsis* coding sequences on the basis of the following criterion: The stretches of 40 bp sequences chosen had best homology with the orthologs from maximum number of plant species, which is determined by blasting against whole embryophyta database. About 61.5% of these overgos were successful in recognizing at least one BAC clone from the peanut BAC library, encouraging in terms of utility of interspecific based approaches for the analysis of complex genomes even between distantly-related plant taxa. We assessed the possible role of the functional category of the template sequence of *Arabidopsis* overgos in recognition of orthologs in the peanut genome. As a result, we observed that the overgos derived from the genes coding basic cellular machinery and universally conserved gene motifs such as "myb" and "MADS" box were more likely to succeed in hybridizing to the peanut BAC library. In contrast, the overgos based on genes that are likely to code for specific adaptive characteristics and on the motifs that are under diversifying selection such as LRRs of some resistance genes failed more often than expected in recognition of the peanut BAC clones. The general conclusion we reached was that the functional identity as well as the functional role of the subdomains from which overgos designed is important for the success of the interspecific hybridization approaches.

In conclusion, the resources we created including the BAC library, resistance gene-like sequences, the BAC clones carrying RGAs will serve to advance the peanut genomic studies. The exploratory work about the peanut physical mapping could be used as a starting point toward creation more detailed physical analysis of the peanut genome.