Theor Appl Genet (2005) 111: 87–94
DOI 10.1007/s00122-005-1994-8

ORIGINAL PAPER

B. Yüksel · J. E. Bowers · J. Estill · L. Goff
C. Lemke · A. H. Paterson

Exploratory integration of peanut genetic and physical maps and possible contributions from *Arabidopsis*

Received: 15 December 2004 / Accepted: 7 March 2005 / Published online: 5 April 2005

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 2,210 cM in length. However, no attempt has been 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 oligonucleotide-based probes (''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 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

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-005-1994-8

Communicated by I. Paran

B. Yüksel · J. E. Bowers · J. Estill · L. Goff · C. Lemke
A. H. Paterson (✉)
Plant Genome Mapping Laboratory,
The University of Georgia,
111 Riverbend Road,
Athens, GA, 30605, USA
E-mail: paterson@uga.edu
Tel.: +1-706-5830162
Fax: +1-706-5830161

Introduction

Peanut is a widely grown crop in many areas of the world. According to Food and Agriculture Organization (2003) estimates, peanut production worldwide stood around 37,057,652 (Mt) in 2003. Almost all species in section Arachis of genus *Arachis* are diploids ($2n=20$) with only two exceptions, *Arachis hypogaea* L. and *Arachis monticola* (Krapovickas and Gregory 1994), which are allotetraploids ($2n=4\ x=40$, 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 *Arachis duranesis*, or *Arachis villosa* as A- and *Arachis ipaensis* as B-genome ancestral donors have been frequently 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 2,210 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 are 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, 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 ~2,400 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. Whereas individual probes such as cDNA or genomic clones can be used, oligonucleotide-based probes, called "overgos," 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 BACs in a 24×24×24 experimental design with an 88% success rate (Gardiner et al. 2004).

Arabidopsis is the most thoroughly studied genome among 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 the *Arabidopsis* genome sequence for exploration of the peanut genome. *A. hypogaea* and *Arabidopsis* are thought to have shared a common ancestor about 100–120 mya (Davies et al. 2004). A study conducted by Lee et al. (2001) has shown that homologous regions among *Glycine max* L. Merr., *Vigna radiata* L., and *Phaseolus vulgaris*. L. 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 *G. max* linkage group G revealed that 27 of 78 sequences had noteworthy homology to *Arabidopsis* (Foster-Hartnett et al. 2002). However, the homologues mapped at six different locations in the *Arabidopsis* genome, and occasional sequence deletions were observed.

The objectives of this study were to explore the feasibility and the practicality of physical mapping approaches for *A. 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

Designing overgos

In these experiments, two sets of overgos were used. The first 117 overgos were designed from peanut RFLP probe sequences from the *A. hypogaea* genetic map (Burow et al. 2001). Peanut probe sequences used were cDNA selected from either root- or shoot-derived cDNA libraries (originally provided by Dr. Gary Kochert, University of Georgia, Athens, Ga., USA). The nucleotide sequences were BLASTed against GenBank databases after masking known repetitive sequences (Altschul et al. 1997). A Microsoft Visual Basic script (J. Bowers and A.H. Paterson, in preparation) 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*-derived overgos (see Electronic Supplementary Material, Table 4) used in this experiment were designed with the same program. Both sets of primers were synthesized commercially (MWG Biotech, High Point, N.C., USA) and dissolved in ddH_2O at a final concentration of 0.2 nM.

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 peanut overgos (POVs) 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.

Gridding peanut BAC library and preparing high-density filters for hybridization

A total of 182,784 *A. hypogaea* BAC clones were gridded on 22.5-cm^2 Hybond N^+ membranes (Amersham Life Sciences, Arlington Heights, Ill., USA) with a QBot high-precision robot (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 ten filters. The filters were incubated on 1% LB agar containing 12.5 μg/μl CM, at 37°C for 12–18 h 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°C.

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 C for 5 min then cooled on ice), 1 μg BSA, 2.5 U

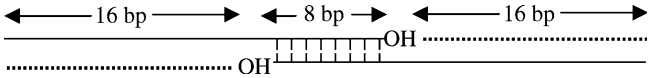

Fig. 1 Schematic representation of overgo structure and labeling; the *dashed lines* indicate radiolabeled portion of the probe

Taq polymerase, 0.5 µl α^{32}-dATP (6000 Ci/mmol, MP Biomedicals, Irvine, Calif., USA), 0.5 µl α^{32}-dCTP (6000 Ci/mmol, MP Biomedicals), and 3 µl oligo-labeling buffer without dATP, dCTP, and random hexamers (Ross et al. 1999). The reaction mixtures were incubated at 37°C for 2 h (Fig. 1). The labeled overgos were combined in the proper column, row, and diagonal pools, and then filtered through Sephadex columns to remove unincorporated nucleotide, and heat-denatured at 94°C for 5 min before adding to the proper bottles.

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°C for 4 h before adding the denatured radioactive overgo probes. Hybridizations were performed at 55°C with a constant rotation speed of 4.5 rpm for 18–36 h in rotisserie ovens. Following the hybridization process, the membranes were washed at 55°C 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, Cedar Knolls, N.J., USA) and X-ray film (Blue medical X-ray film, SourceOne, San Diego, Calif., USA) for 2 weeks at 80°C before developing.

Data entry and analysis

The hits on the films were scored manually onto transparent templates, scanned, and read by an ABBYY FineReader, version 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). Afterward, 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

Analysis of peanut overgo hybridization

From 117 peanut overgos applied to ten high-density BAC filters containing a total of 182,784 clones, a total of 22,266 data points were 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 (see Electronic Supplementary Material, Table 3). This discrepancy was mostly due to "overmatches" (i.e., BAC clones that are hit with more than one probe in the same pool of probes). The experimental design does not allow deconvolution of these hits.

From 6,663 hits successfully ascribed to a specific probe, 2,730 were to POVR2012 (i.e., 1.5% of 182,784 peanut BAC clones contained this sequence at least once). To discover 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 The Institute for Genomic Research (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%, TIGR repeat database). Hence, it is likely that this overgo recognized an uncharacterized repetitive element that 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 to further characterize transposons in the peanut genome and assist in shedding light on peanut genome evolution.

Another overgo with a high number of hits (1,277) was POVR2545. 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 orthologues 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 the amino acids Phe, Tyr, and Trp (Entus et al. 2002). The *Arabidopsis* genome contains three copies of this gene.

Thus, this gene family may exist either 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, see Electronic Supplementary Material, Table 2). The trypsin inhibitors, induced as a result of wound response, are a multicopy gene family; for instance, the soybean genome contains ten copies, mostly in tandem repeats (Jofuku and Goldberg 1989). A similar tandemly repeated distribution of these genes has been also detected in *A. thaliana* (Clauss and Mitchell-Olds 2004).

Three overgos, POVS063, POVR116, and POVR177, with 72–149 BACs consistent with copy numbers of 10–20, did not share sequence similarity with known proteins. Seven overgos, POVS26, POVR2091, POVR2110, POVR230, POVR190, POVR2405, and POVR2482 had 33–67 hits, consistent with copy numbers of 5–10 in the tetraploid genome (see Electronic Supplementary Material, Table 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 of L10 and S26, respectively, and *Arabidopsis* contains three to four copies of each family from these genes (Barakat et al. 2001; see Electronic Supplementary Material, Table 2).

A total of 41 probes had 9–28 hits, consistent with copy numbers of two to four 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, then 90% of the numbers of hits per overgo would be expected to fall in this range, assuming 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 ten 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 in the same overgo pool 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 (Burow et al. 2001) and had significant overmatch hits. The remaining probes could be from underrepresented regions of the peanut genome in our library due to biased distribution of *Hin*dIII restriction sites, experimental factors such as overgo secondary structures, or GC content. It is also plausible that the sequence divergence between some homeologue pairs (i.e., alloalleles) is high, or even that one copy of duplicated genes has been deleted, thus preventing cross-hybridization. Prior experience has shown that more than a 3- to 4-bp difference in a 40-bp long overgo is sufficient to prevent the identification of the target sequence.

Figure 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 *A. hypogaea* on peanut mapping blots, and to 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 our inability to be confident of resolving the expected ten or more RFLP bands. Most of the probes that had less

Table 1 Hybridization of 576 *Arabidopsis* oligonucleotide-based probes (*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	**353**	

Table 2 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	*83.3
Translation	51	12	*76.5
Transcription factors	35	9	74.3
Intracellular trafficking	26	8	69.2
Metabolism	80	28	65.0
Cell division & DNA synthesis	19	8	57.9
Other[c]	86	37	57.0
Transport facilitation	18	8	55.6
Unknown	151	67	55.6
Post translational modifications	33	16	51.5
Cellular com./signal transduction	53	26	50.0
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 (National Center for Biotechnology Information)

[b]Overgos that had at least one hit were considered

[c]This category includes the sequences with known functions other than the ones depicted on the table

*The observed number of failed and successful overgos was significantly different than the expected values for these categories at $P \leq 0.05$ (χ^2 test)

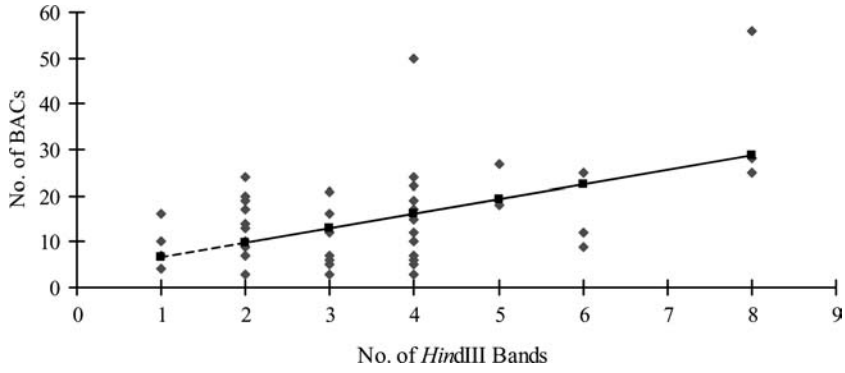

Fig. 2 Linear regression between number of RFLP bands and number of BAC hits. Probes that had less than three BAC hits are omitted due to high possibility of experimental error stemming from scoring (overmatches). Similarly, overgo probes that have more than 59 hits are also not included

than 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.

Application of *Arabidopsis*-derived overgos to peanut

A total of 576 overgos designed from coding sequences of *A. 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 and minimizes errors in ascribing addresses. After deconvolution, 5,434 BAC hits (see Electronic Supplementary Material, Table 3) were ascribed to a total of 353 overgos (Table 1; Fig. 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 positives in the peanut BAC library, the most likely explanation is that similarity between the *Arabidopsis* sequence and the peanut orthologous region was not sufficient. A few could also have been caused by lack of coverage or experimental errors.

To gain insight into whether there is any effect of the putative functional category of *Arabidopsis* overgo on its recognition of positives in peanut BACs, we BLASTed the overgos against GenBank and sorted according to their blast hits (Table 2). 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 $P = 0.05$.

The 576 overgos employed in this experiment were distributed at an average of 45 genes apart along the 34 α-duplicated segments (as shown by Bowers et al. 2003) of the *Arabidopsis* genome. To assess the scope of synteny between *Arabidopsis* and peanut, the positions of *Arabidopsis*-derived overgos that detected the same BAC clones were compared with both modern and pre-α-duplication gene orders. Only seven (12.9%) of a total of 54 pairs of overgos that detected the same BAC clone in the peanut library 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 that there is one gene for every 50 kb in peanut (the allotetraploid peanut genome is about 20× larger than *Arabidopsis* genome but contains two copies of every gene). This suggests 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 (see Electronic Supplementary Material, Table 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 which mapped in close proximity on the genetic map somewhat impeded the interpretation of the experimental results. About 38% of overgos had fewer than six hits, of which the largest proportion (ten) had only one hit (Fig. 3). About 48% of probes identified 6–24 hits, and only 17% had more than 24 hits (Fig. 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. This supports a typical organization of two to four copies per probe in the peanut genome. In conclusion, the number of positives was compatible with the estimated 6.5× genome coverage for peanut BAC li-

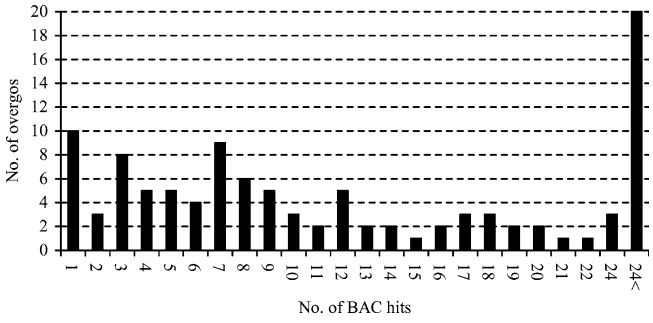

Fig. 3 The graph denotes the distribution of positive BAC hits for 107 overgos. The *y-axis* is for number of overgos for particular number of BAC hits, which is demonstrated on the *x-axis*

brary, perhaps even suggesting that the library has slightly higher coverage.

In this study, a thorough 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 significantly diminished laboriousness of the screening process. The experimental errors inherent to the design were not significant; only 8.5% of the probes failed. The primary 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 the peanut 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 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 six overgos had more than 100 hits (Table 1; Fig. 4). Therefore, the copy number of these probes per *A. 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 6–24 hits, which is equivalent to one to four loci per peanut genome.

Almost all overgos that 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 modification pathways often failed to hybridize to the peanut BAC library (Table 2). This result was not surprising, because 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, the overgo AOVG0576, with strong homology to one functional motif (leucine-rich repeats found in *Cf*-like disease resistance genes) failed to identify any positives. This may have been caused by hypervariability of these regions (Dixon et al. 1998). Conversely, the overgo AOVG0168, with high homology (40 bp/40 bp 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 in identifying any positive clones from the peanut BAC library, supporting presumed involvement of some of these gene products in synthesis of secondary metabolites that may not be present in peanut. Even

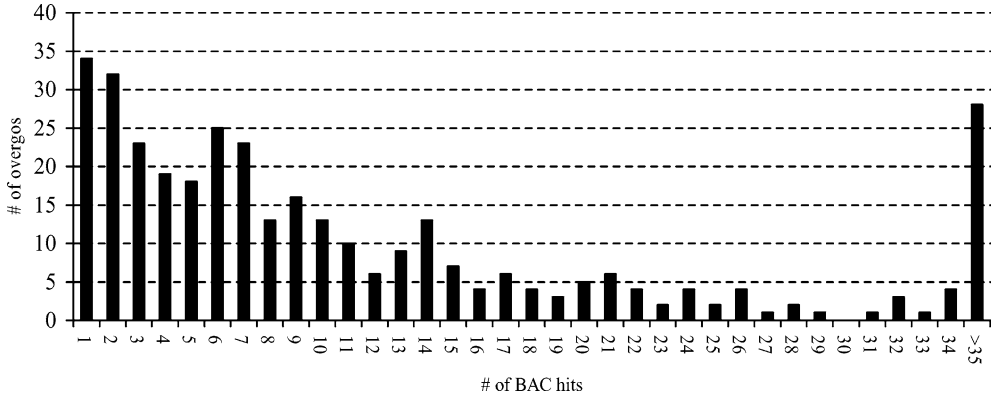

Fig. 4 The distribution of BAC hits for 353 *Arabidopsis* overgos, which had more than one hit in hybridization. The *x-axis* denotes the number of BAC hits for individual overgo probes, which is shown on the *y-axis*

though the majority of overgos similar to enzymes associated with translation machinery (including some ribosomal proteins) produced hits, a certain percentage (23%) did not.

In partial summary, the success of an 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 motifs have a higher chance of success, unless these motifs are under diversifying selection such as leucine-rich repeats of 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* orthologues 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, the degree of synteny was estimated 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* also have 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. Our result is in concordance with these observations. However, much better syntenic relationships between the aforementioned legume species and *Arabidopsis* may have been observed if α-duplication gene order was used. 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, and 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. Nucleic 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 USA 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 via http://apps.fao.org/page/collections?subset=agriculture. Verified 10 February 2004

Fordham-Skelton AP, Skipsey M, Eveans IM, EdwardsR, GatehouseJA (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. Annu Rev Plant Physiol Plant Mol Biol 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 Appl 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). Am J Bot 83:1282–1291

Krapovickas A, Gregory WC (1994) Taxonomía del género *Arachis* (Leguminosae). Bonplandia 8:111–113

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 Appl 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 Syst 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

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor

Stalker HT, 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 USA 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 Appl 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