Genetics: Early Online, published on February 20, 2015 as 10.1534/genetics.115.174607

Tetrasomic recombination is surprisingly frequent in allotetraploid *Arachis*

Soraya Leal-Bertioli*,†, Kenta Shirasawa§, Brian Abernathy†, Marcio Moretzsohn*, Carolina Chavarro†, Josh Clevenger‡, Peggy Ozias-Akins‡, Scott Jackson†, and David Bertioli,†**

*Embrapa Genetic Resources and Biotechnology, PqEB W5 Norte Final, Brasília, DF, 70770-917, Brazil

§Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba, 292-0818, Japan

†Center for Applied Genetic Technologies, University of Georgia, Athens, GA, 30602-6810, U.S.A.

‡Department of Horticulture, University of Georgia, Tifton, GA, 31973, U.S.A.

**University of Brasília, Institute of Biological Sciences, Campus Darcy Ribeiro. 70910-900. Brasília, DF, Brazil

KEY WORDS

Arachis, peanut, groundnut, induced allotetraploid, tetrasomic genetics, segmental genetics

Corresponding author:

David Bertioli

Center for Applied Genetic Technologies

111 Riverbend Road

Athens, GA 30602-6810

Phone: 706/542-4021

Email address: bertioli@uga.edu

Running title: *Arachis* segmental genetics

ABSTRACT

Arachis hypogaea L. (cultivated peanut) is an allotetraploid (2n = 4x = 40) with an AABB genome type. Based on cytogenetic studies it has been assumed that peanut and wild-derived induced AABB allotetraploids have classic allotetraploid genetic behavior with diploid-like disomic recombination only between homologous chromosomes, at the exclusion of recombination between homeologous chromosomes. Using this assumption, numerous linkage map and Quantitative Trait Loci studies have been carried out. Here, with a systematic analysis of genotyping and gene expression data, we show that this assumption is not entirely valid. In fact, autotetraploid-like tetrasomic recombination is surprisingly frequent in recombinant inbred lines generated from a cross of cultivated peanut and an induced allotetraploid derived from peanut's most probable ancestral species. We suggest that a better, more predictive genetic model for peanut is that of a "segmental allotetraploid" with partly disomic, partly tetrasomic genetic behavior.

This intermediate genetic behavior has probably had a previously overseen, but significant impact, on the genome and genetics of cultivated peanut.

INTRODUCTION

Whole genome duplications have occurred in virtually all angiosperm lineages at some time, and are believed to be a major driving force in their evolution (Cui *et al.* 2006). They generally result in polyploids that are new biological species with increased heterosis and robustness and whose genomes have new possibilities for gene and genome evolution (eg. Adams and Wendel, 2005). There is believed to be a correlation between polyploidy and the ability to adapt to new environments (te Beest *et al.* 2012) and many crop plants are polyploids, perhaps reflecting an increased adaptability to, and increased yield under, cultivation (Renny-Byfield and Wendel 2014).

Perhaps the most frequent class of polyploidy are tetraploids, the result of the duplication of a diploid genome. These are categorized as autotetraploids that result from chromosome doubling of a non-hybrid genome, or allotetraploids that result from hybridization between two species and subsequent chromosome doubling. In autotetraploids, all chromosome sets can pair in meiosis leading to tetrasomic inheritance. Allotetraploids generally undergo disomic recombination because only homologous chromosomes pair at meiosis, but homeologous chromosomes do not. However, some allotetraploids have meiotic behavior that is intermediate between these two extremes: whilst most chromosomes pairings in meiosis are between homologous chromosomes, some pairing and genetic recombination also occurs between homeologous chromosome pairs (Wu, *et al*. 2001).

Arachis hypogaea L. (cultivated peanut), an allotetraploid (2n = 4x = 40) with an AABB genome type, is believed to be derived from the wild diploid species *A. duranensis* Krapov. & W.C.Greg. and *A. ipaënsis* Krapov. & W.C.Greg. that contributed the A and B genomes respectively (Kochert, *et al*. 1991; Seijo, *et al*. 2007; Moretzsohn, *et al*. 2013). These two component genomes diverged an estimated 2.9 Mya (million years ago; Morezsohn *et al*. 2013). Although relatively recent, this divergence should be put in the biological context of the unusual reproductive biology of *Arachis* species. Their high rates of self pollination, the peculiar habit of underground pod and seed development, short distance seed dispersal, and consequent frequent population bottle-necks are likely linked to the accumulation of Bateson-Dobzhansky-Muller incompatibilities and the high estimated rates of speciation (0.95 speciation events per million years, compared to 0.15 for legumes in general; (Magallon and Sanderson 2001; Moretzsohn, *et al*. 2013). Accordingly, diploid hybrids derived from crosses between A-genome and B-genome species are highly infertile (Krapovickas and Gregory 1994; Mallikarjuna, *et al*. 2011).

Since most wild species of *Arachis* are diploid, synthetic allotetraploids have been developed to enable their use in peanut breeding programs (Simpson, *et al*. 1993; Fávero, *et al*. 2006; Mallikarjuna, *et al*. 2011; Leal-Bertioli *et a*l., 2014). These plants are usually produced by crossing diploid species with A- and B-genomes, followed by a treatment of the sterile F_1 hybrid with colchicine to induce chromosome doubling. The resultant tetraploid plant can be crossed with cultivated peanut. Based on cytogenetics and molecular mapping, *A. hypogaea* and induced allotetraploids derived from wild diploid species have been considered to have classic allotetraploid genetic behavior. Using this assumption, numerous linkage map construction and QTL identification studies have been successfully carried out (Burow, *et al*. 2001; Varshney, *et al*. 2009;

Hong, *et al*. 2010; Fonceka, *et al*. 2012; Gautami, *et al*. 2012; Qin, *et al*. 2012; Shirasawa, *et al*. 2012; Sujay, *et al*. 2012; Wang, *et al*. 2012; Shirasawa, *et al*. 2013; Zhou, *et al*. 2014).

However, in three datasets we observed patterns among unexpected datapoints. The expression of numerous *A. ipaënsis* transcribed regions "disappeared" in a derived artificially-induced allotetraploid and, unexpectedly, these regions were genetically clustered. Likewise, in recombinant inbred lines (RILs) derived from a cross of cultivated peanut with an induced allotetraploid, "missing" PCR marker data showed unexpected genetic clustering. Finally, rare unexpected single nucleotide polymorphism (SNP) genotypes, unexplainable with disomic recombination, were observed in both the induced allotetraploid itself and RILs (recombinant inbred lines). We realized that these data could be explained by a degree of tetrasomic genetic recombination. Here we report a systematic examination of the datasets to test this hypothesis.

MATERIALS AND METHODS

Plant material

Plants were maintained in pollinator-free greenhouses and seeds saved from year to year. Accessions of the diploid species *A. ipaënsis* KGBSPSc30076 (hereafter referred to in the abbreviated form K30076), *A. duranensis* V14167, and their derived colchicine-induced allotetraploid [*A. ipaënsis* x *A. duranensis*]^{4x} (termed an amphidiploid in the original publication; Fávero *et al*. 2006) were obtained from the Active Germplasm Bank of Embrapa Genetic Resources and Biotechnology (Brasília, Brazil). The tetraploid cultivated peanut, *A. hypogaea* cv. Runner IAC-886 was from the Instituto Agronômico de Campinas (Campinas, Brazil).

Recombinant inbred lines in F_6 generation were derived from a cross between *A. hypogaea* cv. Runner IAC-886 and the colchicine-induced allotetraploid. These RIL lines are the same population described in Shirasawa *et al*. (2013) and Bertioli, *et al*. (2014). Lines were produced from a single F_1 plant, cloned by cuttings in order to produce enough F_2 seeds, and single seed descent to the F_6 generation.

The colchicine-induced allotetraploid was derived from chromosome doubling, and so is completely homozygous. Our assumption, shown here to be false, was that it would be genetically stable over generations (see Results). Therefore, we will here indicate the number of generations the induced allotetraploid plants were removed from the original polyploidy event for each experimental procedure. For the production of the RIL population, the parental colchicine-induced allotetraploid was only one or two generations removed from the original polyploidy event. For the DNA control for genotyping, the induced allotetraploid was about seven generations removed, and for expression analysis the induced allotetraploid was about nine generations removed.

Comparison of gene expression of the two diploid species and their derived allotetraploid

The colchicine-induced allotetraploid [*A. ipaënsis* K30076 x *A. duranensis* V14167]^{4x} (Fávero *et al.* 2006) and its diploid parents were grown in the greenhouse at the same time, under the same conditions. The induced allotetraploid was an approximately 9th generation progeny derived by single seed descent from the original polyploidy event. Total RNA was extracted from the first expanded leaf of the main axis using the Qiagen Plant RNeasy kit (Qiagen, USA) with on-column DNAse treatment. cDNA libraries were constructed using equal amounts of RNA from five

individuals of each genotype using the TruSeq v2 library construction kit (Illumina, USA). To obtain long reads to improve transcriptome assemblies, size-selected libraries were sequenced using MiSEQ v3.0. For gene expression analyses, libraries were sequenced using HiSEQ 2000 2x150 paired end sequencing (Illumina, USA).

Adapter and quality trimming was performed using Trim_galore! v0.3.5. (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Transcripts were assembled using Trinity with default parameters (Haas, *et al*. 2013). Assembled transcripts were filtered to include only the longest isoform from each read cluster. The longest isoforms were then aligned to each other using NCBI blastn v2.2.29. Alignments with 100% sequence identity and $\geq$ 90% sequence length were considered redundant and removed from the final assembly.

Mapping of reads from *A. duranensis*, *A. ipaënsis*, and the 9th generation induced allotetraploid [*A. ipaënsis* x *A. duranensis*]^{4x} onto a joint database of assembled *A. ipaënsis* and *A. duranensis* transcripts was performed using bowtie2 v2.2.2, tophat v2.0.11 and cufflinks v2.2.1 (Trapnell, *et al*. 2012). The densities of mapping (FPKM) were used as measures of expression of the diploid species transcribed regions in the diploid and tetraploid contexts. The use of a joint database helped to avoid erroneous cross comparison of A and B genome transcripts, which have high similarity, because each read from the induced allotetraploid could only map once onto either an *A. ipaënsis* or an *A. duranensis* transcript. Furthermore, we only considered counts for *A. duranensis* assembled transcripts to which no *A. ipaënsis* reads mapped, and *A. ipaënsis* transcripts to which no *A. duranensis* reads mapped (“cleanly mapping reads”).

Transcripts that were expressed in the diploid species, but not in the induced allotetraploid were identified by processing data in Excel, with regard to statistical significance of the comparisons.

Transcripts corresponding to sequences that had been previously genetically mapped onto maps from two different crosses, both having *A. duranensis* as one of the parents (Nagy, *et al*. 2012; Shirasawa *et al.* 2013) were identified using NCBI blastn v2.2.29.

Genetic markers and maps

Simple sequence repeat (SSR) and transposable element PCR-based markers genotyping and map data was from Shirasawa *et al*. (2013), SNP marker data was from a 1536 Illumina GoldenGate single nucleotide polymorphism assay (Nagy *et al*. 2012; Bertioli *et al*. 2014). The SNPs for this assay were called between two A-genome accessions: *A. duranensis* PI 475887 (KSSc 36036) and *A. duranensis* Grif 15036 (PI 666084; WiSVg 1510-B).

In the search for tetrasomic recombinants we used both standard polymorphic SNP markers (here referred to as “intragenomic SNPs”, see Figure 1) and SNP markers that were not polymorphic for this population, but which harbored different bases on the A- and B-genomes (here referred to as “intergenomic SNPs”, see Figure 2). Intergenomic SNPs, provided unexpected key information for the study. Being non polymorphic they could not be genetically mapped on the tetraploid mapping population used here. Therefore the diploid SNP A-genome map generated by Nagy *et al*. (2012) using an F_2 population derived from two *A. duranensis* accessions was used to position markers. Linkage groups were reoriented where necessary to ensure forward compatibility with pseudomolecules produced by the Peanut Genome Sequencing Consortium

(www.peanutbase.org).

Analysis of genotyping data

Genotypes resultant from tetrasomic genetic behavior were identified by structured inspection of the results of genotyping assays (Bertioli, *et al*. 2014; Shirasawa, *et al*. 2013). This inspection was different from marker scoring for creating standard genotyping data for linkage mapping: it was focused on data points that did not fit the expected disomic model of recombination and assays that were monomorphic between tetraploid parents. Tetrasomic and null alleles were identified in genomic regions from SNP markers and from PCR-based SSR and transposable element markers (the full rationale for the scoring is explained in the Results section).

Single marker association and Quantitative trait loci analyses

To investigate if tetrasomic recombination was influenced by genetic factors, we carried out quantitative trait loci analyses using, as a phenotype, the number of null PCR marker data points for each RIL from the PCR marker linkage map of Shirasawa *et al.* (2013) with only co-dominant markers (Supplementary file 1 – worksheet "PCR and SMA", line 1205).

Single-marker analysis was done by comparing the association between a marker genotype and the trait value using simple linear regression. Composite interval mapping (CIM; Zeng 1993; Zeng 1994) was executed in WinQTL Cartographer, version 2.5 (Wang *et al*. 2006). CIM analysis was performed using the Standard Model (Model 6), scanning the genetic map and estimating the likelihood of a QTL and its corresponding effects at every 1 cM, while using 8 significant marker

cofactors to adjust the phenotypic effects associated with other positions in the genetic map. A window size of 10 cM was used, and therefore cofactors within 10 cM on either side of the QTL test site were not included in the QTL model. Threshold was determined by permutation tests (Churchill and Doerge 1994; Doerge and Churchill 1996), using 1,000 permutations and a significance level of 0.05.

RESULTS

Phenotypic observation of parentals and RILs

Over almost a decade of observation of the diploid and tetraploid parental accessions and the induced allotetraploid lines seemed phenotypically stable.

Through the advancement of over six generations, the RILs derived from cultivated x induced allotetraploid were vigorous and fertile. This observation is consistent with the current consensus that *A. ipaënsis* and *A. duranensis* are the genome donors to *A. hypogaea*, and so an induced allotetraploid derived from a hybrid between them effectively "re-creates" the species *A. hypogaea* (or more strictly speaking, *A. monticola* Krapov. & Rigoni which is the wild form of *A. hypogaea*; Grabiele *et al.* 2012).

Comparison of gene expression of the two diploid species and their derived allotetraploid

HiSeq sequencing of cDNA libraries produced 20,843,737 reads of *A. ipaensis*, 18,393,012 reads

of *A. duranensis* and 48,644,287 reads of the induced allotetraploid. The software Trinity produced 79,828 assembled transcripts with a contig N50 of 1,330 bp for *A. ipaënsis* and 82,089 assembled transcripts with a contig N50 of 1,625 bp for *A. duranensis*.

The software pipeline of bowtie2, tophat and cufflinks mapped 90.7%, 91.4%, and 89.9% of the reads from *A. ipaënsis*, *A. duranensis* and the induced allotetraploid respectively to the combined database of *A. ipaënsis* and *A. duranensis* assembled transcripts. 15,099 assembled *A. ipaënsis* transcripts were cleanly mapping, that is, only *A. ipaënsis* reads, and no *A. duranensis* reads mapped to them. 12,685 assembled *A. duranensis* transcripts were cleanly mapping, that is, only *A. duranensis* reads and no *A. ipaënsis* reads mapped to them.

Of these, 909 *A. ipaënsis* and 35 *A. duranensis* transcripts, were expressed in the diploid state but not in the induced allotetraploid (null expression transcripts). Searches against genetically mapped sequences (Nagy, *et al*. 2012; Shirasawa, *et al*. 2012; Shirasawa, *et al*. 2013) showed that the null expression *A. ipaënsis* transcripts were disproportionally located and broadly distributed on Linkage Group 04 (LG04; Table 1; Figure 3). (Here when referring to both A- and B-genome homeologous linkage groups we will use the generic “LG”. When referring specifically to A- or B-genomes we will use “LGA” or “LGB”.) Furthermore, on LG04, there were disproportionally few *A. ipaënsis* transcripts for which there was robust support for expression in the tetraploid (FPKM >10) (Table 1, Figure 3). This genetic grouping of B-genome null expression transcripts suggests a loss of alleles rather than silencing that would tend to be scattered over the genome. Furthermore, inspection of SNP and PCR-based marker data also indicated loss of B-genome alleles on LG04 (see below).

Analysis of genotyping data

Identifying tetrasomic/nullisomic genomic regions from SNP markers:

For SNP assays of the type used for genetic mapping, the contrasting bases are harbored on homologous bases of the same component A- or B-genome; in other words, the SNP assay is "intragenomic". For this type of assay, data points generally clustered into two main groups on GoldenGate signal intensity plots in accordance with the expected genotypes for an allotetraploid RIL population. These correspond to the genotypes of the parentals: eg. $A^aA^aC^bC^b$ and $C^aC^aC^bC^b$ (here bases A and C are used as exemplars of SNP polymorphism any other combination of two different bases could have been used, "a" and "b" superscript letters indicate the component genome of origin, we shall use "x" where it cannot be assigned; see Figures 1a and b and 2). Here, the base C is the ancestral base and A is the variant base. In this case, the expected genotypes considering disomic recombination are $A^aA^aC^bC^b$, $A^aC^aC^bC^b$, $C^aC^aC^bC^b$. Genotypes that are tetraplex and triplex for the base A are not expected, and can only arise through tetrasomic recombination ($A^aA^aA^aC^b$, $A^aA^aA^aA^a$, see Figure 1a and b). Note that genotypes that are duplex, monoplex or nulliplex for the base A may also be the result of tetrasomic recombination (eg. $A^aA^aC^aC^a$, $A^aC^aC^aC^a$, $C^aC^aC^aC^a$, $C^bC^bC^bC^b$), but they are indistinguishable from genotypes derived from disomic recombination. Therefore, this assay can only distinguish two of the products of tetrasomic recombination, those that are tetraplex and triplex for the variant base. Since SNPs for the GoldenGate assay were identified using two *A. duranensis* (A-genome) accessions; almost all the intragenomic SNP assays can distinguish genotypes that are tetraplex for *A. duranensis* (A-genome), but not genotypes that are tetraplex for *A. ipaënsis* (B-genome) or *A. hypogaea* SNPs. (Clearly, this type of assay cannot distinguish $C^bC^bC^bC^b$ from $C^aC^aC^bC^b$). Signal intensity plots for SNP assays were inspected and SNP genotypes indicative of tetrasomic recombination were

recorded manually.

An initially unexpected source of data was also collected from SNP assays that are not polymorphic for classic allotetraploid mapping, “intergenomic SNPs” (Figure 2). These are assays where both parental genotypes are the same and are of the type $G^aG^aC^bC^b$, that is, the contrasting bases reside on homeologous genomes (here G and C are used as exemplars of DNA bases, any other combination of two different bases could have been used). With strictly disomic recombination, genotypes of all progeny RILs should be the same as the parents and recombinants would not be found. However, for many assays, some genotypes of the type $G^aG^aG^aG^a$, $G^aG^aG^aC^b$, $C^bC^bC^bG^a$, $C^bC^bC^bC^b$ were evident in a proportion of the progeny (Figure 2). These genotypes arise through tetrasomic recombination and datapoints corresponding to them were recorded for all intergenomic SNP assays. Data collection was performed using GenomeStudio’s software and subsequent data transformations were done in Excel. Note that, for quantifying the products of tetrasomic recombination, the intergenomic SNP is more sensitive than the intragenomic SNP, because, with the former, all products of tetrasomic recombination can be distinguished, but with the latter only two products can.

All 1,536 SNP genotyping assays were inspected of which 222 were informative for tetrasomic recombination. Of these, 97 were intragenomic polymorphic SNP markers (almost all informative only for tetraplex and triplex alleles of *A. duranensis*), and 125 were intergenomic (A- vs. B-genome) SNPs (Supplementary file 1 – worksheets “IntERgenomic SNPs” and “IntrAgenomic SNPs”). Ninety one RIL lines were analyzed, but two of these were eliminated because they showed contaminant or very aberrant genotypes. This makes 19,758 datapoints in total, which, after removing 529 with low quality, gives 19,229 effective datapoints. A total of 604 datapoints

(3.1%) were informative of tetrasomic recombination. Of these 403 were informative of A-genome tetrasomic recombinants, 193 of B-genome recombinants, and 8 tetrasomic events could not be assigned.

The 89 RIL lines had an average of 3% of marker calls indicative of tetrasomic recombination. Of these, only nine lines had zero tetraplex/nulliplex calls, 32 had 0-2% tetraplex/nulliplex calls and 48 had more than 2% of tetraplex/nulliplex calls. The line most affected by tetrasomic recombination (T141) had 14.5% tetraplex/nulliplex calls (See Figure 4 for a frequency distribution).

When the lines and tetrasomic/nullisomic genotyping calls were ordered using the Nagy *et al.* (2013) linkage map, it was obvious that the tetraplex/nullplex genotype calls formed genetically meaningful blocks (Figure 5; Supplementary file 1, worksheets "IntERgenomic SNPs", "IntrAgenomic SNPs" and "Both SNP types"). The spatial distribution of the tetraplex/nullplex regions was highly structured. Most lines harbor only one or two well-supported tetra/nullisomic segments, and these regions cover an entire or a substantial part of a linkage group. Several points of tetrasomic recombination were shared between different lines. This is particularly noticeable on LG03 where twenty lines have large tetra/nullisomic regions (Figures 5 and 6). Overall, although the total number of A-genome tetraplex calls is larger than B-genome tetraplex calls, this mostly reflects the greater density of A-genome informative markers, and the genetic distances covered by A- and B-genome tetraplex regions are similar.

The degree of tetrasomic behavior of the linkage groups varies considerably. LG03 displays the most autotetraploid behavior (6.4% of calls are tetra/nullisomic and 20 RIL lines have large tetra/nullisomic regions). In contrast we found no robustly supported autotetraploid behavior in

LG01 (0.2% of calls tetra/nullisomic, all from isolated points; no lines with well supported large tetra/nullisomic regions) (Figure 6)

Overall, although the percentage of tetra/nulliplex calls is low, the influence on the genotypes of the RILs is very considerable. Out of 89 lines, at least 35 have at least one large genome segment derived from tetrasomic recombination. The RIL line most affected by tetrasomic recombination (T141) has a substantial lower region of LG03 tetraplex for B-genome, and a substantial lower region of LG09 tetraplex for the A-genome (Far right column Figure 5). For this line, these regions cover about 82 cM out of a total of 1077 cM. One line (T38) is nulliplex for the B-genome at the top of LG04 and nulliplex for the A-genome at the bottom (Supplementary File 1 – worksheet "Both SNP types").

Identifying tetrasomic/null alleles using PCR-based markers:

Whilst inspecting a "colormap" generated from the PCR-based markers for the construction of the tetraploid linkage maps published elsewhere (Supplementary file 1 worksheet "PCR and SMA") we noted blocks of "missing" SSR and transposable element PCR-based marker data. Initially, these data were considered failed assays. However, under closer inspection, and because of the block structure, we suspected that these datapoints may not be missing, but in fact null genotypes derived from tetrasomic recombination. For the RILs with the most prominent blocks of "missing" data, the original genotyping data in the form of gel images and fluorescent traces generated by GeneMapper (Applied Biosystems, USA), was inspected to check if the marker assays had, in fact, failed, or if the alleles were actually null. This can be done because many PCR-based markers amplify two pairs of loci from the allotetraploid genomes, one pair from the A-genome and one pair from the B (Figure 7). Usually only the A or the B alleles are polymorphic and the non-polymorphic

loci are ignored. However, in this context, the non-polymorphic loci provided positive controls for the functioning of the genotyping assay. This close observation of PCR markers was also repeated over the whole of the homeologous linkage groups of the RILs in question (For selected traces, see Supplementary Files 2 and 3).

Observations of a colormap of the PCR based marker map (Supplementary file 1 worksheet "PCR and SMA") indicated that the RILs T138, T101, T125 and T141 displayed prominent blocks of "missing" PCR marker data in LGA02, LGB06, LGB09 and LGB09 respectively. These were further investigated by manual inspection of PCR-based marker traces. All markers in the linkage groups of interest were observed, although some markers were more informative than others. Some markers were dominant and some co-dominant, some amplified A- and B-genome alleles and some did not, furthermore some marker electrophoretic profiles gave more distinct and well separated peaks than others. In addition, lines identified by SNP markers to harbor nulli/tetraplex regions on LG03 and LG04 were chosen for inspection of the PCR marker data. A selection of marker gels and traces are available as a Supplementary file 2. Results of marker inspections are described below. All genetic distances in this section are with respect to Shirasawa *et al.* (2013).

LGA02 and line T138 (for annotated traces see Supplementary File 2)

Genotyping for 32 PCR-based markers on this linkage group was inspected for the line T138, and, as controls, the parentals and lines T101, T125 and T141. Eight microsatellite markers (AHGS1677, AHS1351, AHG2733, AHGS2026, AHS0946, AHGS3738, AHS1949, AHS1008) were particularly suitable for detection of null alleles, they are all co-dominant, amplify a product from both the A- and B-genomes, and have easy to score, well separated electrophoretic peaks. All lines presented A- and B-genome alleles except for T138 which presented only B-genome alleles.

This is consistent with SNP data that scores LG02 in T138 to be tetrasomic for the B-genome and nullisomic for the A-genome. (See Figure 7 for marker profiles of AHGS1677, for more marker traces see Supplementary File 2).

LG03 and lines T040, T074, T095 (for annotated traces see Supplementary File 2)

Of the 13 particularly suitable markers between 76.5 and 101.9 cM, seven and eight showed null B alleles in T040 and T074, respectively. All these 13 markers between were null for the A-genome in T095. This is consistent will the SNP data except that the genetic distances involved are shorter than those indicated by the SNP markers which are positioned according to the Nagy map. This difference could easily be accounted for by different recombination characteristics of the different populations used to calculate the genetic distances.

LG04 and the induced allotetraploid and line T038 (for annotated traces see Supplementary File 2)

SNP and expression analysis indicated that the seventh/ninth generation induced allotetraploid was nulli/tetraplex over much of the genetic space of LG04 (Figure 3). SNP data indicated that T038 is nulliplex for B alleles at the top of LG04 and nulliplex for A alleles at the bottom. Inspection of 19 markers that are particularly suitable for the detection of null alleles confirmed the expected genotypes. The induced allotetraploid was nulliplex for B alleles at all marker positions except for marker AHGS2048 that mapped at 56.7 cM on A04 and AHGS2390 which mapped at 71.1 cM on A04. Line T038 is nulliplex for B-genome alleles from 0-31 cM, diplex from 56.4 - 67.45 cM and nulliplex for A alleles from 71.1- 91.0 cM. This shows that line T038 had at least two tetrasomic recombination events involving the same linkage groups.

LGB06 and line T101 (for annotated traces see Supplementary File 3)

Data for 25 PCR-based markers distributed along the linkage group B06 was inspected for the line T101. Parentals and lines T125, T138 and T141 were compared as controls. Six markers (AHGS1943, AHGS1464, AHS1510, AHGS1756, AHGS2243, AHGS2195) were particularly suitable for detection of null alleles. These markers detected A- and B-genome alleles in all lines except for T101, which presented only A-genome alleles for all the linkage group except three markers at 25.1, 19.9 and 15.6cM (AHGS2058, AHGS2324, AHGS2317; traces not shown in Supplementary File 3). This is consistent with SNP data that detects A-genome tetrasomics / B-genome nulls on LG06.

LGB09 and lines T125 and T141 (for annotated traces see Supplementary File 3)

Data for 14 PCR-based markers distributed along the linkage group was inspected for lines T125 and T141 and controls. Five markers (AHGS1500, AHG2205, AHGS2324, AHGS2696, AHS1950) were particularly suitable for detection of null alleles. B-genome alleles were detected at the top marker only (AhTE0222, 0cM; trace not included in Supplementary File 3), no other marker displayed any B alleles. This matches with the SNP data that scores LG09 in T125 and T141 to be disomic at the top of the linkage group and tetrasomic for the A-genome and nullisomic for the B-genome at the bottom.

When the dominant markers are removed from the map of Shirasawa *et al*. (2013) and it is incorporated into a colormap coded to enhance missing or null data (Supplementary file 1 worksheet “PCR and SMA”) then null regions are apparent as blocks within genotypes and linkage groups. The concordance with the SNP calls of nulli/tetraplex genomic regions is good. In summary, the inspection of PCR-based marker data independently identified and/or confirmed the tetraplex/nulliplex regions of these RILs that were identified using SNPs.

Single marker association and Quantitative trait loci analyses:

Single marker analysis for tendency to undergo tetrasomic recombination (number of null PCR-markers per genotype) indicated significant marker associations on LGA02 and LGA07. For LGA02, marker Seq16C03 positioned at 39.6 cM was significantly associated ($P \leq 0.1$: R^2 14.2%). For LGA07, almost all markers between 35.7 and 54.6 cM ($P \leq 0.1$), were significantly associated with the most significant markers positioned at 36.1, 52.2, 52.6 and 54.6 cM ($P \leq 0.01$: R^2 4.7%, 15.1%, 15.1% and 11.2% respectively) (Supplementary file 1 worksheet "PCR and SMA", columns CP to CV)

Composite interval mapping identified three QTLs with peaks: LGA02 at 37.5 cM, LGA07 at 51.3 cM and LGB03 at 86 cM (Table 2). Supplementary file file 1 worksheet "CIM QTLs").

Therefore both CIM and single-marker analysis indicated LGA02 at around 38 cM, and LGA07 at around 51 cM (Table 2). Although mapping of genes that influence genotype must be undertaken with caution because of the possible difficulties of distinguishing cause from effect, neither of these genetic positions is much affected by tetrasomic recombination (having two and zero null datapoints respectively). Therefore the QTLs are very likely valid. Note that the additive effects of the QTLs are opposite, indicating that both wild and cultivated alleles favor tetrasomic recombination.

DISCUSSION

The morphology of *Arachis* chromosomes and their mechanisms of recombination have long been of interest to peanut researchers. Pioneering work showed that *A. hypogaea* is an allotetraploid species (2n = 4x = 40) with an AABB type genome (Husted 1936, Smartt *et al*. 1978). Meiotic chromosomes were shown to consist of 20 chromosome bivalents in 88–98% of cells, exceptions being rare univalents, trivalents, and quadrivalents, suggesting limited homeologous pairing between A- and B-genomes (Singh and Moss 1982; Wynne and Halward 1989). Molecular mapping in populations derived from intraspecific crosses of *A. hypogaea* was for some time impossible because of very low DNA polymorphism. However, as the wild diploids related to the component genomes of *A. hypogaea* became better defined, it was possible to make inferences using diploids in hybridization schemes. Hybrids between A- and B-genome species are typically very infertile, and bivalent formation is significantly reduced compared to hybrids between species of the same genome type (Krapovickas & Gregory 1994, Robledo & Seijo, 2010). The use of wild species also allowed the introduction of more allelic diversity into the cultivated species facilitating molecular mapping. The first molecular study of alien introgression into cultivated peanut (Garcia *et al.* 1995) used a map previously constructed from a diploid cross to order markers (Halward *et al.*1993). A small but significant amount of recombination between the A-genome alien species (*A. cardenasii* Krapov. & W.C. Gregory) and the B-genome of cultivated peanut was detected. However, since the study used the complex hexaploid route for introgression, its relevance to normal allotetraploid *Arachis* genetics was uncertain. The first map construction in allotetraploid *Arachis* used a BC_1 population derived from a cross of *A. hypogaea* with an artificially-induced allotetraploid [*A. batizocoi* x (*A. cardenasii* x *A. diogoi*)]^{4x} (termed an amphidiploid in the original paper; Burow *et al.* 2001). In spite of the complexity of the cross, and the distinctness of *A. batizocoi* Krapov. & W.C.Greg. from the B-genome of *A. hypogaea* (Robledo and Seijo 2010; Moretzsohn, *et al*. 2013; Leal-Bertioli, *et al*. 2014), the material of this study more closely mimicked

the natural genetics of *A. hypogaea*. Although there were some anomalies, as expected, inheritance was reported to be disomic. Since then, to our knowledge, all the molecular mapping studies in allotetraploid *Arachis* have assumed disomic inheritance for map construction and QTL identification (Varshney, *et al*. 2009; Hong, *et al*. 2010; Robledo and Seijo 2010; Fonceka *et al*. 2012; Gautami, *et al*. 2012; Qin *et al*. 2012; Shirasawa *et al*. 2012; Sujay *et al*. 2012; Wang *et al*. 2012; Shirasawa *et al*. 2013; Zhou, *et al*. 2014).

To investigate the genetic behavior and genome structure of peanut we have over some years built up datasets centered around the two most probable diploid ancestral species of cultivated peanut, an induced allotetraploid derived from them, cultivated peanut and RILs derived from a cross between them (Shirasawa *et al.* 2013; Bertioli *et al.* 2014 and unpublished data). As our datasets became more populated, infrequent datapoints that could easily have been dismissed as noise began to form regularities that seemed to imply the involvement of some biological phenomena that we had not accounted for. A systematic review of this data prompted us to interpret it in a fresh light.

Indications of tetrasomic genetic behavior came together from several lines of evidence. Gene expression and genotyping data showed that a line of induced allotetraploid studied at nine and seven generations removed from the original induced polyploidy had become nulliplex for B-genome alleles over a very large proportion of the genetic space on LG04. Genotyping data showed us that most RILs harbor the expected B-genome alleles on LGB04. Therefore, this genetic recombination must have happened after the second generation when the allotetraploid was used as male parent to generate the F_1 from which the RILs were derived.

This in itself may have been dismissed as a rare unexpected genetic event. However, both SNP and PCR-based marker data indicated that many of the RILs had also undergone tetrasomic recombination (we recognize that some of null marker calls may be due to deletions, but overall the patterns that are observed are clearly best explained by tetrasomic recombination). Although the number of datapoints indicating tetrasomic recombination was only about 3% of the whole dataset, the impact of this recombination on the genomic structure of the RILs was surprisingly large. Most lines show some evidence for tetraplex/nulliplex genomic regions and many of these regions cover substantial proportions or even whole linkage group arms (Figure 5). Their frequency was different for different linkage groups, being most frequent in LGs 02, 03 and 06, of moderate frequency in LGs 04, 09 and 10, infrequent in LGs 05, 07 and 08, and possibly absent from LG01 (Figure 6).

Because the unexpected datapoints were infrequent, for some time we simply did not notice them, dismissed them as noise, or assumed they were failed marker assays. Because of their relative infrequency, map construction and QTL identification were possible not taking them into account (Shirasawa *et al.* 2013; Bertioli *et al.* 2014, Table 2 and unpublished). Furthermore, we believe that the linkage maps and QTLs identified are broadly valid. However, without taking tetrasomic recombination into account, the inferred genotypes of many lines would be substantially incorrect.

We recognize that the induced allotetraploid is of recent origin and so may have an enhanced instability (Soltis and Soltis 1995). We also recognize that different genetic backgrounds are likely to influence the tendency to tetrasomic recombination. However, because of the very high similarity of A- and B-genomes in genic regions (Bertioli, *et al*. 2013; Moretzsohn *et al.* 2013) the fact that the induced allotetraploid in this study was derived from the most likely ancestral species to cultivated peanut, the indication that both wild and cultivated alleles positively influence tetrasomic

recombination (Table 2), and the previous studies that indicate a small amount of tetravalent pairing in the meiosis of cultivated peanut (Singh and Moss 1982; Wynne and Halward 1989), we think it probable that tetrasomic recombination has gone unnoticed in previous studies. The Burow, *et al*. (2001) study used a backcross population, where null alleles would not be apparent. Furthermore, the linkages created by low frequency tetrasomic recombination would be too weak to significantly influence the linkage maps. Interestingly, one linkage group consisting of markers originating from different component genomes was reported and the authors do mention the possibility that this may represent a genome region with tetrasomic genetic behaviour. The mapping studies based on intraspecific crosses of *A. hypogaea* have mostly generated lower density data where the patterns in datapoints derived from tetrasomic recombination would be difficult to detect. The higher density maps produced by Shirasawa, *et al*. (2012) are based on F_2 populations, so again null data would be much more difficult to detect.

Occam's razor, a central principle of scientific enquiry, proposes that among competing hypotheses, the one with the fewest assumptions should be preferred. Frequently, the most parsimonious hypotheses can be made when aberrant data is discarded or ignored. Sydney Brenner (with tongue firmly in cheek) dubbed this methodology “Occam’s broom”, because inconvenient facts are “swept under the carpet”. Plant genetics has a long and distinguished history of using Occam’s broom (see the “process of sophistication” probably used in the collection of Mendel’s data (Fisher 1936) and it is routinely used in mapping algorithms even today. The working assumption of disomic inheritance is parsimonious, and has served peanut genetics well. However, to build better, more predictive genetic models, for peanut, and perhaps other plants where disomic inheritance has been assumed, “in due course, the edge of the carpet must be lifted and the untidy reality confronted” (Robertson 2009). Artificially-induced *Arachis* allotetraploids, and

the spontaneous allotetraploid *A. hypogaea* are probably "segmental allotetraploids" (*sensu* Stebbins 1947, 1950) with predominantly disomic, but partly tetrasomic genetic behavior. As shown here, even infrequent tetrasomic recombination could have had a significant impact on the genome of *A. hypogaea* and have an unexpected importance in genetics and breeding programs today.

ACKNOWLEDGEMENTS

This work was performed with the support of the Georgia Genomics Facility at the University of Georgia, funding from Mars Inc., EMBRAPA, the Peanut Foundation and the USDA National Institute of Food and the Agriculture National Research Initiative Competitive Grants Program (#2006-35604-17242).

REFERENCES

Adams K. L. and J. F. Wendel. 2005 Polyploidy and genome evolution in plants. Current Opinion in Plant Biology, 8:135–141.

Bertioli, D. J., P. Ozias-Akins, Y. Chu, K. M. Dantas, S. P. Santos *et al.* 2014 The Use of SNP Markers for Linkage Mapping in Diploid and Tetraploid Peanuts. G3: Genes|Genomes|Genetics 4**:** 89-96.

Bertioli, D. J., B. Vidigal, S. Nielen, M. B. Ratnaparkhe, T. H. Lee *et al.* 2013 The repetitive component of the A genome of peanut (Arachis hypogaea) and its role in remodelling intergenic sequence space since its evolutionary divergence from the B genome. Ann Bot 112**:** 545-559.

Burow, M., C. Simpson, J. Starr and A. Paterson, 2001a 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.

Burow, M. D., C. E. Simpson, J. L. Starr and A. H. Paterson, 2001b 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.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics 138**:** 963 - 971.

Cui, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis, *et al.* 2006 Widespread genome duplications throughout the history of flowering plants. Genome Res. 16: 738-749.

Doerge, R. W., and G. A. Churchill, 1996 Permutation tests for multiple loci affecting a quantitative character. Genetics 142.

Fávero, A. P., C. E. Simpson, J. F. M. Valls and N. A. Vello, 2006 Study of the evolution of cultivated peanut through crossability studies among Arachis ipaensis, A. duranensis, and A. hypogaea. Crop Sci 46**:** 1546 - 1552.

Fisher, R. A., 1936 Has Mendel's work been rediscovered? . Annals of Science 1.

Fonceka, D., H.-A. Tossim, R. Rivallan, H. Vignes, I. Faye *et al.* 2012 Fostered and left behind alleles in peanut: interspecific QTL mapping reveals footprints of domestication and useful natural variation for breeding. BMC Plant Biology 12**:** 26.

Gautami, B., D. Foncéka, M. K. Pandey, M. r. C. Moretzsohn, V. Sujay *et al.* 2012 An International Reference Consensus Genetic Map with 897 Marker Loci Based on 11 Mapping Populations for Tetraploid Groundnut (*Arachis hypogaea* L.). PLoS ONE 7**:** e41213.

Grabiele M., L. Chalup, G. Robledo, G. Seijo. 2012 Genetic and geographic origin of domesticated peanut as evidenced by 5S rDNA and chloroplast DNA sequences. Plant Syst Evol 298: 1151–1165.

Haas, B. J., P. A., M. Yassour, M. Grabherr, P. D. Blood *et al.* 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8**:** 1494-1512.

Hong, Y. B., X. P. Chen, X. Q. Liang, H. Y. Liu, G. Y. Zhou *et al.* 2010 A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. BMC Plant Biology 10**:** 1-13.

Kochert, G., T. Halward, W. D. Branch and C. E. Simpson, 1991 RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theoretical and Applied Genetics 81**:** 565 - 570.

Krapovickas, A., and W. Gregory, 1994 Taxonomia del genero *Arachis* (Leguminosae). Bonplandia 8**:** 1 - 186.

Leal-Bertioli, S. C. M., S. P. Santos, K. M. Dantas, P. W. Inglis, S. Nielen *et al.* 2014 *Arachis batizocoi*: a study of its relationship to cultivated peanut (*A. hypogaea* L.) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. Annals of Botany in press.

Magallon, S., and M. Sanderson, 2001 Absolute diversification rates in angiosperm clades. Evolution 55**:** 1762-1780.

Mallikarjuna, N., S. Senthilvel and D. Hoisington, 2011 Development of new sources of tetraploid Arachis to broaden the genetic base of cultivated groundnut (Arachis hypogaea L.). Genetic Resources and Crop Evolution 58**:** 889-907.

Moretzsohn, M. C., E. G. Gouvea, P. W. Inglis, S. C. M. Leal-Bertioli, J. F. M. Valls *et al.* 2013 A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Annals of Botany 111**:** 113-126.

Nagy, E. D., Y. Guo, S. Tang, J. E. Bowers, R. A. Okashah *et al.* 2012 A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genomics 13**:** 469.

Qin, H., S. Feng, C. Chen, Y. Guo, S. Knapp *et al.* 2012 An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. TAG Theoretical and Applied Genetics 124**:** 653-664.

Renny-Byfield S., J. F. Wendel 2014. Doubling down on genomes: polyploidy and crop plants. Am J Bot. 101:1711-25.

Robertson, M., 2009 Ockham's broom: A new series. Journal of Biology 8**:** 79.

Robledo, G., and G. Seijo, 2010 Species relationships among the wild B genome of *Arachis* species (section *Arachis*) based on FISH mapping of rDNA loci and heterochromatin detection: a new proposal for genome arrangement. Theoretical and Applied Genetics 121**:** 1033-1046.

Seijo, G., G. I. Lavia, A. Fernandez, A. Krapovickas, D. A. Ducasse *et al.* 2007 Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. American Journal of Botany 94**:** 1963-1971.

Shirasawa, K., D. J. Bertioli, M. C. Moretzsohn, S. C. M. Leal-Bertioli, R. K. Varshney *et al.* 2013 An integrated consensus map of cultivated peanut and wild relatives revealsed structures of the A and B genomes of *Arachis* and divergences with other of the legume genomes. Plos One Submitted.

Shirasawa, K., P. Koilkonda, K. Aoki, H. Hirakawa, S. Tabata *et al.* 2012 In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biology 12**:** 80.

Simpson, C. E., S. C. Nelson, L. J. Starr, K. E. Woodard and O. D. Smith, 1993 Registration of TxAG-6 and TxAG-7 peanut germplasm lines. Crop science 33**:** 1418.

Singh, A. K., and J. P. Moss, 1982 Utilization of wild relatives in genetic improvement of *Arachis hypogaea* L. Part 2. Chromosome complements of species of section *Arachis*. Theoretical and Applied Genetics 61**:** 305–314.

Soltis, D. E., and P. S. Soltis, 1995 The dynamic nature of polyploid genomes.

Stebbins, G. L., 1947 Types of polyploids: their classification and significance. Advances in Genetics 1**:** 403-429.

Stebbins, G. L., 1950 *Variation and Evolution in Plants*. Columbia University Press, New York.

Sujay, V., M. V. C. Gowda, M. K. Pandey, R. S. Bhat, Y. P. Khedikar *et al.* 2012 Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (Arachis hypogaea L.). Molecular Breeding 30**:** 773-788.

te Beest M., J. J. Le Roux, D. M. Richardson, A. K. Brysting, J. Suda, *et al.* 2012 The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot. 109: 19-45.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim *et al.* 2012 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7**:** 562-578.

Varshney, R. K., D. J. Bertioli, M. C. Moretzsohn, V. Vadez, L. Krishnamurty *et al.* 2009 The first SSR based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). Theoretical and Applied Genetics 118**:** 729 - 739.

Wang, H., R. V. Penmetsa, M. Yuan, L. Gong, Y. Zhao *et al.* 2012 Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biology 12**:** 10.

Wang, S., C. J. Basten and Z. B. Zeng, 2006 *Windows QTL Cartographer 2.5*, Raleigh.

Wu, R., M. Gallo-Meagher, R. C. Littell and Z. B. Zeng, 2001 A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. Genetics 159**:** 869–882.

Wynne, J. C., and T. Halward, 1989 Genetics and cytogenetics of *Arachis*. The Critical Review in Plant Sciences 8**:** 189-220.

Zeng, Z. B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90**:** 10972 - 10976.

Zeng, Z. B., 1994 Precision mapping of quantitative trait loci. Genetics 136**:** 1457-1468.

Zhou, X., Y. Xia, X. Ren, Y. Chen, L. Huang *et al.* 2014 Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics 15**:** 351.

Table 1 Numbers of *A. ipaënsis* transcripts positioned on Nagy *et al.* (2012) and Shirasawa *et al.* (2013) maps that are expressed in diploid leaf and have null expression or positive expression in the 9th generation induced allotetraploid leaf. Transcripts have "disappeared" disproportionally from LG04 (in bold).

	Expression of *A. ipaënsis* transcripts in the 9th generation induced allotetraploid	
mapped sequences	**Null**	**Positive (> 10 FPKM)**
LG01	6	49
LG02	0	24
LG03	1	62
LG04	**46**	**6**
LG05	0	26
LG06	2	35
LG07	1	16
LG08	1	33
LG09	0	32
LG10	4	25

Table 2 Quantitative Trait Loci for tetrasomic recombination identified by Composite Interval Mapping (CIM) concurring with single marker association (a QTL was also identified on LGB03 by CIM, but not by single marker association and is not shown here). The table presents the linkage groups (LG), map position in accordance with the Shirasawa *et al.*2013 PCR-marker linkage map, with the markers linked to the QTL, maximum LOD score obtained for the QTL, additive effects and the phenotypic variance explained by each QTL. Note that the additive effects of the QTLs are opposite, indicating that both wild and cultivated alleles favor tetrasomic recombination.

LG	**Position (cM)**	**Nearest markers**	**LOD**	**Additive effect**	**R^2 (%)**
LGA02	37,5	AHS0519 / Seq16C03	4.6	-8,9027	9.8
LGA07	51,3	AHGS1357_t1 / AHS2066	4.7	9,3873	10.9

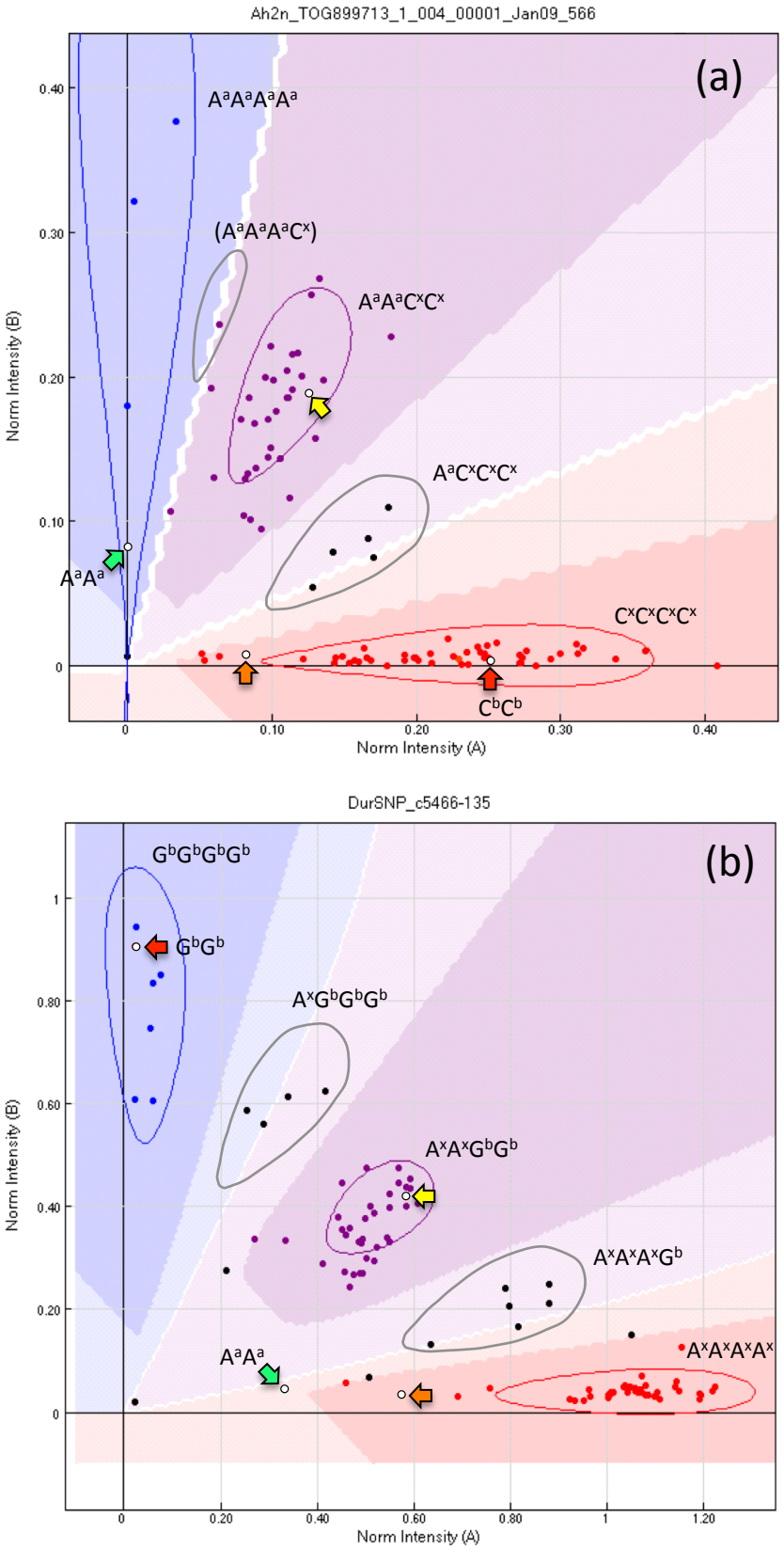


Figure 1
Examples of plots of signal intensities generated using GenomeStudio from "intragenomic" SNP assays on the RIL population derived from *A. hypogaea* and [*A. ipaensis* x *A. duranensis*]^{4x} (These parentals indicated by orange and yellow arrows respectively). Diploid controls, *A. duranensis* and *A. ipaensis* are indicated with green and red arrows respectively. For (a), tetraploid parental genotypes are $C^aC^aC^bC^b$ and $A^aA^aC^bC^b$ (A and C refer to the DNA bases, the superscript letter refers to the subcomponent genomes; superscript "x" is used when it is not possible to assign the subgenome). For the classic allotetraploid model, progeny RILs are expected to form two frequent disomally homozygous clusters ($C^aC^aC^bC^b$ and $A^aA^aC^bC^b$) with an additional cluster with some residual heterozygotes ($A^aC^aC^bC^b$). For the autotetraploid model five clusters are expected. Note that this example broadly conforms to the expectations of the classic allotetraploid model, but three progenies are tetraplex for the base A, and one progeny may be triplex for A (in parenthesis). These genotypes can only be formed by tetrasomic recombination. Note that genotypes that are duplex, monoplex or nulliplex for the base A may also be the result of tetrasomic recombination, but they are indistinguishable from genotypes derived from disomic recombination. **Therefore, this assay can only reliably score one ($A^aA^aA^aA^a$) of the four most frequent genotypes resultant from tetrasomic recombination** (the others being $A^aA^aC^aC^a$, $C^aC^aC^aC^a$, $C^bC^bC^bC^b$). Example (b) only conforms to the classic allotetraploid model approximately. Although the two expected clusters are the most populated, five clusters of genotypes are clearly visible. This SNP assay is an extreme example in this respect.

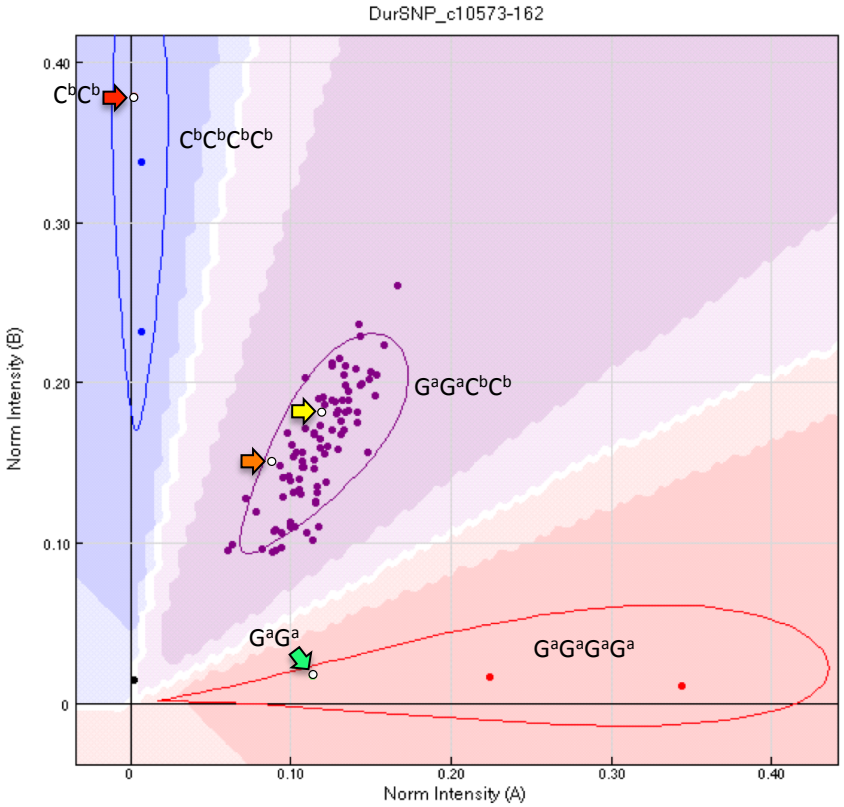


Figure 2

Example of a plot of signal intensities generated using GenomeStudio from an "intergenomic" SNP assay on the RIL population derived from *A. hypogaea* x [*A. ipaënsis* x *A. duranensis*]^{4x}. These parentals indicated by orange and yellow arrows respectively. Diploid controls for *A. duranensis* and *A. ipaënsis* are shown as green and red arrows respectively. For the classic allotetraploid genetic model no recombinants are expected, but here, four progenies have tetraplex/nulliplex genotypes. These genotypes can only arise from tetrasomic recombination. **These intergenomic assays can score all genotypes resultant from tetrasomic recombination**

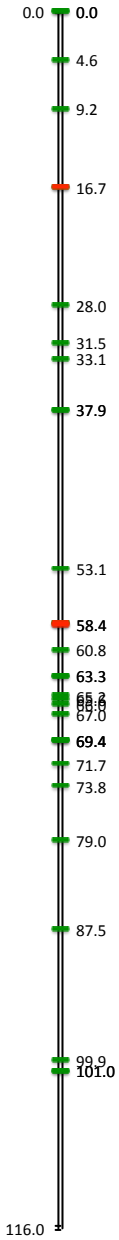


Figure 3
Distribution of *A. ipaënsis* transcripts positioned on LG04 of Nagy et al 2012 map that are expressed in diploid leaf and have null expression (in green) or positive expression (FPKM >10; in red) in the induced allotetraploid [*A. ipaënsis* x *A. duranensis*]^{4x} leaf. Distances are in centiMorgans. Other LGs have very few null expressed transcripts and many more expressed transcripts (see Table 1). The *A. ipaënsis* null expressing transcripts are widely distributed on LG04. This disappearance of *A. ipaënsis* transcripts is consistent with marker data that shows that alleles on LG04 have become nulliplex for B alleles and tetraplex for A alleles (see first column in Figure 5).

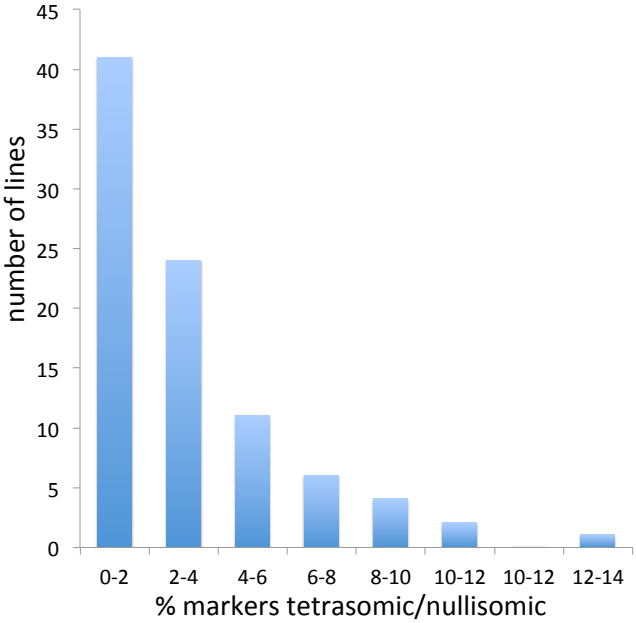


Figure 4

Frequency distribution of number of recombinant inbred lines (RILs) derived from *A. hypogaea* x [*A. ipaensis* x *A. duranensis*]^{4x} *vs.* percentage of markers with tetra/nullisomic genotyping calls.

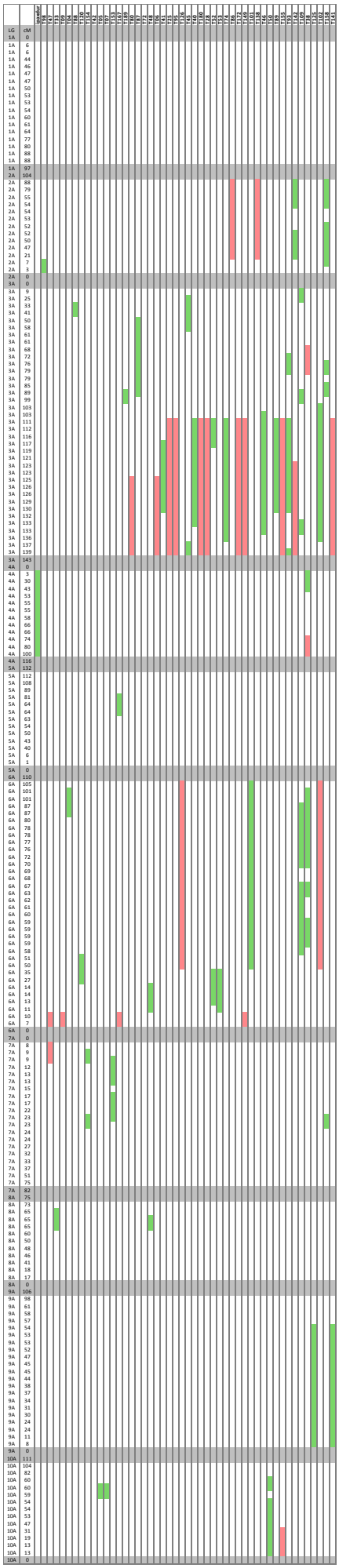


Figure 5

A colour-map representation of the 47 RILs derived from *A. hypogaea* x [*A. ipaënsis* x *A. duranensis*]^{4x} with more than 2% of SNP markers indicating genotypes derived from tetrasomic recombination plus one control colchicine-induced allotetraploid ([*A. ipaënsis* x *A. duranensis*]^{4x}) that was *c.* seven generations removed the original polyploidy event (far left genotype). The genotypes, inferred from inter- and intragenomic SNPs (Figures 1 and 2), are represented in columns, and markers on the horizontal (full data is in Supplementary file 1). Markers are ordered with respect to the A genome linkage map of Nagy *et al* (2012). Linkage groups are delimited by horizontal grey lines. Only tetrasomic regions supported by at least two adjacent informative markers are shown. Green regions are tetrasomic for A genome alleles and nullisomic for B genome alleles. Red regions are nullisomic for A genome alleles and tetrasomic for B genome alleles. Tetrasomic recombination events often cover large portions of linkage groups. Tetrasomic/nullisomic calls were almost absent from LG01 (none shown here). In LG03 they are most frequent: 20 lines represented here have large tetrasomic/nullisomic regions of this linkage group. Note that the allotetraploid [*A. ipaënsis* x *A. duranensis*]^{4x} has undergone tetrasomic recombination that covers all informative markers on LG04, this is consistent with PCR marker (Supplementary file 2) and expression data that is also presented in this manuscript (see Figure 3).

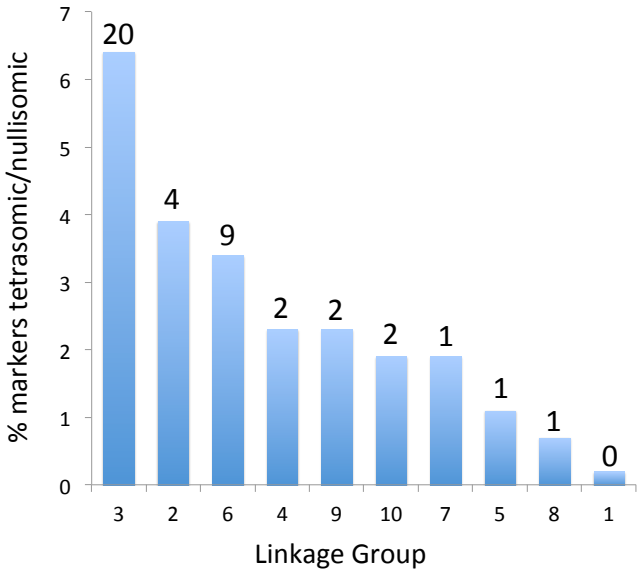
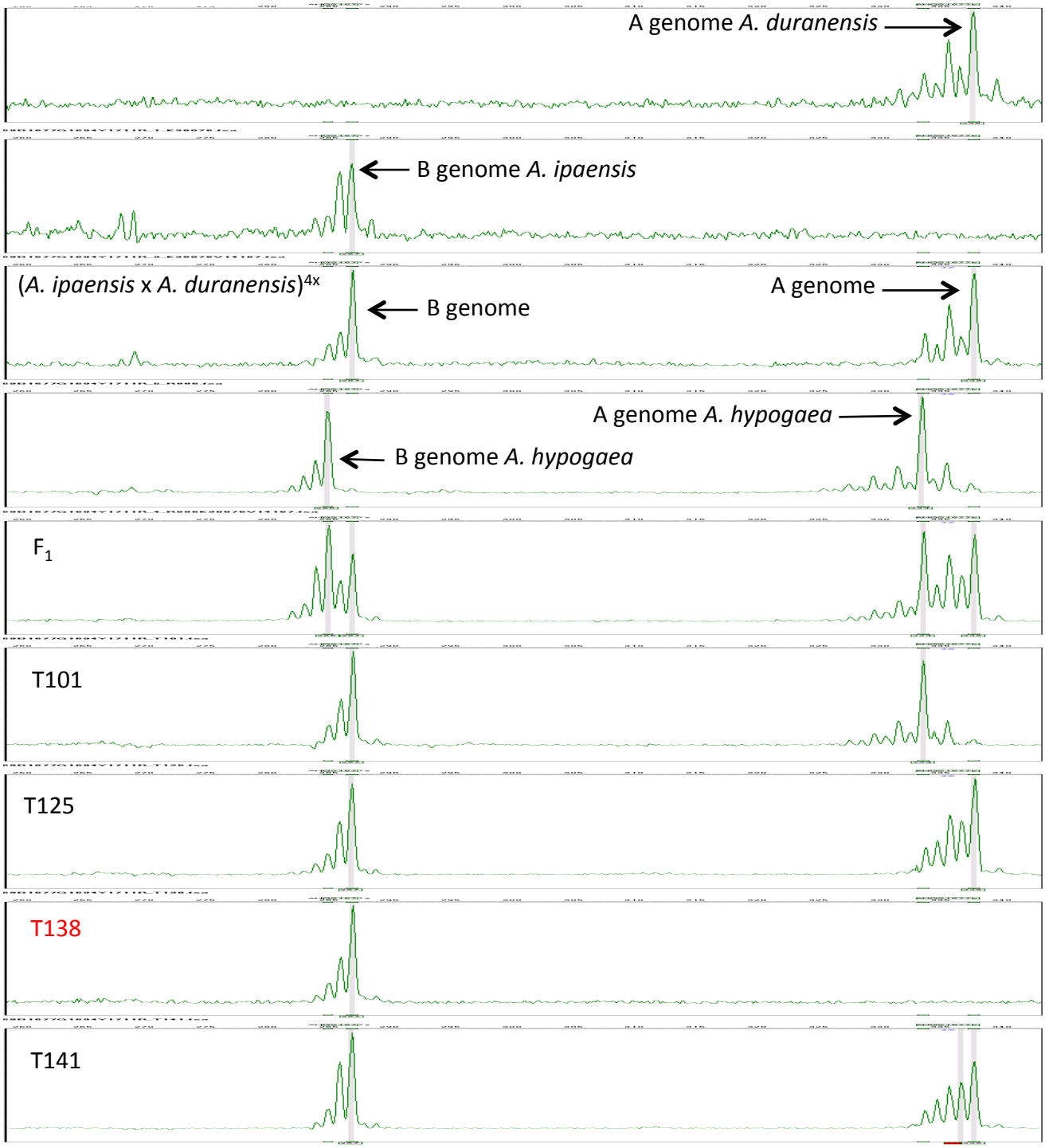


Figure 6
Distribution of percentage of markers with tetra/nullisomic genotyping calls for the linkage groups of recombinant inbred lines (RILs) derived from *A. hypogaea* x [*A. ipaensis* x *A. duranensis*]^{4x} and the induced allotetraploid parent, as inferred from SNP genotyping (Figures 1 and 2). Numbers above columns are the number of plants with tetra/nulliplex segments supported by at least three contiguous informative markers on the respective linkage group.



AHGS1677_A – LGA02 9.8 cM

Figure 7

An example of a trace from capillary electrophoresis (ABI 3730 XL) of a microsatellite marker (AHGS1677), generated by the GeneMapper software, revealing an A genome null allele of a recombinant inbred line (RIL); T138. The marker is polymorphic in both A and B subgenomes. Therefore, assuming allotetraploid genetics, all lines should show peaks in both A and B genomes, but T138 only has a B genome allele. At this locus the line is tetraplex for B alleles (in this case of *A. ipaënsis*) and nulliplex for A genome alleles.