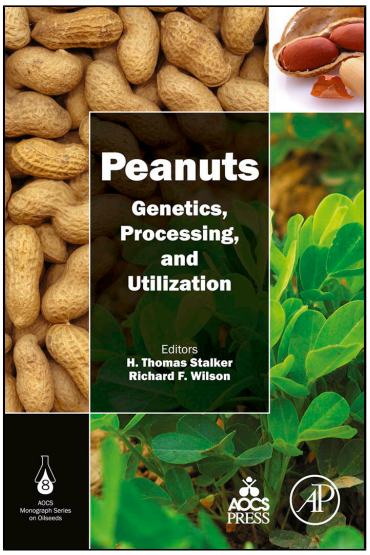

From Bertioli, D.J., Leal-Bertioli, S.C., Stalker, H.T., 2016. The Peanut Genome: The History of the Consortium and the Structure of the Genome of Cultivated Peanut and Its Diploid Ancestors. In: Stalker, H.T., Wilson, R.F. (Eds.), Peanuts: Genetics, Processing, and Utilization. Academic Press and AOCS Press, pp. 147–161.
ISBN: 9781630670382

Academic Press and AOCS Press

Chapter 5

The Peanut Genome: The History of the Consortium and the Structure of the Genome of Cultivated Peanut and Its Diploid Ancestors

David J. Bertioli[1,2], Soraya C. Leal-Bertioli[2,3], H. Thomas Stalker[4]
[1]University of Brasília, Institute of Biological Sciences, Brasília, Brazil; [2]Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA; [3]Embrapa Genetic Resources and Biotechnology, Brasília, Brazil; [4]Department of Crop Science, North Carolina State University, Raleigh, NC, USA

INTRODUCTION

The first plant genome to be sequenced was the model plant *Arabidopsis thaliana* in 2000. Since then, over 140 plant species have had their genomes sequenced (www.ncbi.nlm.nih.gov/genome/browse). Most of the plants chosen fit specific criteria such as being model organisms or economically important, having small genome size, being diploids, access to genetic and physical maps, transcriptome, and other genomic tools, and also, a large research community (Michael and Jackson, 2013). Peanut (*Arachis hypogaea* L.) is an allotetraploid, with a large genome (3.2 Mb), commercially important in the United States and of paramount significance for the livelihoods of people in developing countries. The peanut research community gathered around the sequencing of the peanut genome with the clear objective of aiding peanut breeding, to expedite the production of elite cultivars, and help ensure global food security and safety. In this chapter we first describe the history of the mobilization of the international peanut research community to plan and execute the sequencing of the peanut genome. Later, we describe the biology of the genome of peanut and its wild ancestral species and how their genomes have been essential for the understanding of the peanut genome.

Peanuts. http://dx.doi.org/10.1016/B978-1-63067-038-2.00005-8

Peanuts, First Edition, 2016, 147-161

DISCUSSION

Brief History of Peanut Genomics

To develop a direction for genomics efforts with legume species important in commerce, 26 scientists representing five major crops in the United States (peanut, soybean, dry bean, dry pea, lentils, and model legumes) met at Hunt Valley, Maryland in 2001 along with national commodity leaders to develop a plan for genomic research and ultimately for crop improvement (Boerma et al., 2001). Although advances in science greatly differed among the crops, six areas were identified as being in common across all species, including (1) genome sequencing of strategic legume species, (2) physical map development, (3) functional analysis, (4) development of DNA markers for comparative mapping and breeding, (5) characterization and utilization of biodiversity, and (6) developing resources for analyzing and storing genomics data. A US national workshop was then held in 2003 in Santa Fe, New Mexico as a follow-up to the Hunt Valley workshop. Fifty researchers and several funding agencies worked to refine objectives for genomic research and an action plan to fund genomics research was developed (Gepts et al., 2005). This meeting led to the Legume Genomics Initiative (LGI) that was represented by one scientist, one national commodity representative, and one commercial producer from each of the above-mentioned crops. As a result of the workshops and LGI, nationally competitive grants for legume genomics was funded by the United States Department of Agriculture (USDA), there was an increased awareness by national commodity leaders for the need of additional efforts in plant genomic research. The book Legume Crop Genomics was then published in 2004 (Wilson et al., 2004).

In 2006, the first international peanut conference of aflatoxin and genomics was organized in Guangzhou, China, which brought together international peanut researchers to formulate plans for cooperative research in peanut genomics. The following year, The Peanut Foundation and American Peanut Council sponsored the first International Strategic Planning Workshop held in Atlanta, Georgia where the International Strategic Plan for the Peanut Genome Initiative was developed. As opposed to groups working with other crop species, peanut researchers took a very broad view of genomics research to include all aspects of research leading to crop improvement, including germplasm collection and preservation, trait identification and characterization, molecular marker research, transformation research, genome sequencing and characterization, population development, and plant breeding leading to new cultivars. As a result, peanut populations have been developed and are being characterization in parallel with genomic research. These efforts are expected to significantly shorten the timeline for utilizing genomic tools in peanut and lead to quicker cultivar development. A strategic plan prepared in 2006 (Wilson, 2006b) that originally had six objectives, one of which targeted allergen proteins in the seed,

was revised in 2012 into five areas (Wilson et al., 2012) with specific goals as follows:

1. Allelic Diversity and Germplasm Resources:
   Goal: Characterize genetic diversity and transfer useful genes into new sources of germplasm for crop improvement.
2. Genome Sequencing and Structural Characterization
   Goal: An ordered, anchored, annotated, and accessible genome sequence to facilitate peanut improvement.
3. Genetic Trait Mapping and Gene Discovery
   Goal: Enhancing crop improvement using genetic and genomic tools.
4. Product Quality and Safety
   Goal: Integrated research strategies for major issues that impact global marketing and consumer preferences for peanuts and peanut products.
5. Crop Improvement
   Goal: Ensure that the new genetic information can be used by plant breeders to provide an adequate supply of agronomic and high quality peanut cultivars.

Wilson (2006a) summarized an action plan to accomplish the objectives of the strategic plan and the funding required to accomplish the objectives was presented to the peanut industry by Valentine et al. (2006) in the white paper Biotech Peanut White Paper, Benefits and Issues. Following the 2006 Atlanta workshop, a series of conferences were organized by the international peanut community, such as the Advances in Arachis through Genomic and Biotechnology: An International Strategic Planning Workshop, that have been held on an annual or semiannual basis in India, Mali, Brazil, China, and the USA. Summaries of workshops, strategic and action plans and other activities related to peanut genomics are documented at the website http://www.peanutbioscience.com.

Lastly, The Peanut Genome Consortium (PGC) was initiated in 2010 as an extension of the International Peanut Genome Initiative and is embodied by a coalition of international scientists and stakeholders engaged in the Peanut Genome Project. The goal of this group was to sequence the cultivated peanut genome and its diploid progenitor species. Significant progress has been made to sequence and annotate the peanut genome as will be discussed later in this chapter. The PGC governance is referenced at http://www.peanutbioscience.com.

From quite early, it was clear that the tetraploid peanut genome is complex and difficult to sequence and assemble (Bertioli et al., 2014a). Therefore, the two diploid progenitors, *Arachis duranensis* and *Arachis ipaënsis*, were first sequenced and assembled. Together they provide a good basis and will help guide assembly of the tetraploid genome. In 2014, the sequence of the diploid progenitors was publicly released into the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov), becoming freely available for the scientific community. Even before the publication of the sequencing article in a

scientific journal, the PGC members felt it was important to make available not only the sequence, but also, tools that will help peanut researchers improve cultivars by integrating genetic, genomic, and trait information. With this in mind, a web portal in a searchable format for genomics, genetics, and trait information for peanut was created: PeanutBase (www.peanutbase.org) was started in April 2013 at Iowa State University. PeanutBase develop and deploy tools for harvesting both existing and developing resources, including genetic maps, genetic and molecular markers, quantitative trait locus (QTL) data, reference genomes, gene models, transcriptomes, proteomes, and functional gene model annotations. Data curation plays a prominent role in loading and presenting these data (Cannon et al., 2014). PeanutBase aims to make best use of this data with breeder-centered information such as candidate genes, genetic markers, QTLs, trait information about breeding lines, and allele states for important markers (Dash et al., 2014).

The second part of this chapter will describe the origin of peanut, give an overview of peanut genome and its wild relatives show the genomic similarities between the different peanut subcomponents and those of the wild progenitors, and describe the rationale for sequencing the two wild peanut progenitors, *A. duranensis* and *A. ipaënsis*.

Cultivated Peanut and Its Wild Relatives

Like almost all other crops *A. hypogaea* does not grow wild. The question of the relationships of the wild species to cultivated peanut has long been of interest to researchers. Cytogenetic studies have been particularly informative to determine the biosystematic relationships among *Arachis* species.

The chromosomes of cultivated peanut are of mostly similar size and metacentric. The component A genome is characterized by the presence of strong centromeric heterochromatic bands in the chromosomes and one pair of small chromosomes. In contrast, B chromosomes are all of similar size and have much weaker centromeric bands. The diploid wild *Arachis* species that are most closely related to peanut reside in the botanical section *Arachis* and most have 20 metacentric chromosomes (here we will not deal with species with 18 chromosomes or asymmetric karyotypes). These were initially divided into two genome types in accordance with the component genomes of *A. hypogaea*. Species with the characteristic small pair of chromosomes were assigned to an A genome group. Species with symmetric karyotypes but without the small pair were assigned to a B genome group. Subsequent studies of species biology, cytogenetics, and molecular phylogeny have strongly supported the validity of the A genome grouping where most of the species in the group are perennials, while other species of the section *Arachis* are annuals; almost all chromosomes have AT-rich centromeric bands that strongly stain with DAPI; and they consistently group into molecular phylogenies. However, further cytogenetic studies and molecular phylogenies justified the subdivision of the B group (now known as B

sensu lato) into B (sometimes referred to as B *sensu stricto)*, K, and F genomes (Robledo and Seijo, 2010). Over the years, a number of these diploid species had been considered as possible ancestors of cultivated peanut, including the A genome species *Arachis correntina*, *A. duranensis*, *Arachis cardenasii*, the K genome species *Arachis batizocoi* and the B *sensu stricto* species *A. ipaënsis* (Gregory and Gregory, 1976; Singh and Smartt, 1998; Stalker et al., 1991). However, numerous studies and a large body of evidence, including biogeographic considerations, DNA phylogenies, and detailed cytogenetic analysis support *A. duranensis* and *A. ipaënsis* as being the most closely related to the component genomes of cultivated peanut (Kochert et al., 1996; Moretzsohn et al., 2013; Ramos et al., 2006; Robledo and Seijo, 2010; Seijo et al., 2004, 2007). A comparison of 58 kbp of DNA sequence from *A. duranensis* V 14167 with the A genome of *A. hypogaea* showed 98.66% identity (unpublished data) and a comparison of 61 kbp of DNA sequence from *A. ipaënsis* and the B genome of *A. hypogaea* showed a remarkable 99.99% identity (unpublished data).

The A and B Genomes Are Closely Related

Cultivated peanut (*A. hypogaea*) is an allotetraploid of recent origin with an AB type genome ($2n=4x=40$) of about 2.8 Gb and with a high repetitive DNA content. The very low DNA polymorphism present even between the most distantly related varieties of cultivated peanut indicates that polyploidy was very recent and probably unique (Cuc et al., 2008; Halward et al., 1991, 1992; Kochert et al., 1996; Moretzsohn et al., 2013). This very low level of polymorphism makes the development of informative DNA markers very difficult and has hampered the classical genetic characterization of the crop.

Using sets of orthologous intron sequences from *A. duranensis* and *A. ipaënsis, Lotus*, and both paleopolyploid components of soybean, a molecular clock of DNA sequence divergence was calibrated using the known divergence dates of the two genomic components of soybean (13 million years ago (Mya)), and the Galegoid, Phaseoloid and Dalbergioid clades. Results estimated that the divergence of the *Arachis* A and B genomes was 2.9–3.5 Mya (Moretzsohn et al., 2013; Nielen et al., 2012). Although, in evolutionary terms this is relatively recent, it should be put in the biological context of the unusual reproductive biology of *Arachis* species. Their peculiar habit of underground pod development and consequent limited seed dispersal, combined with infrequent long-range seed dispersal and high rates of self pollination leads to patchy populations that have passed through multiple genetic bottlenecks and have very limited gene flow between them. This is likely linked to the fixation of chromosome rearrangements and the accumulation of Bateson–Dobzhansky–Muller incompatibilities that lead to high rates of speciation (estimated at 0.95 speciation events per million years), compared with 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 (Gregory and Gregory, 1979; Krapovickas and Gregory, 1994, 2007; Mallikarjuna et al., 2011; Stalker et al., 1991).

Comparisons of two pairs of orthologous A and B genome regions showed highly conserved segments (with about 95% identity) punctuated by segments with no significant homology (Bertioli et al., 2013). The similarity of the segments of high identity is compatible with the estimated date of evolutionary divergence and a substitution rate of 1.3×10^{-8} per site per year (Ma et al., 2004). Many of the segments of no significant homology were detectably repetitive, reflecting that 3.5 million years is ample time for very significant transposon activity. Indeed, most easily dated transposons in plant genomes are less than three million years old. Older elements tend to be degraded by mutation or eliminated by unequal crossing-over and illegitimate recombination (Devos et al., 2002; Pereira, 2004; Vicient et al., 1999). A substantial divergence in the repetitive component of the two genome components of peanut is consistent with in situ hybridization experiments where chromosome spreads were probed with *A. duranensis* bacterial artificial chromosome clones (Bertioli et al., 2013), or with genomic in situ hybridization (GISH) using whole genomic DNA of its most probable ancestral diploids *A. duranensis* and *A. ipaënsis* (Seijo et al., 2007). The probes do not hybridize exclusively, but predominantly to the chromosomes of their respective genome components. This shows that the repetitive components of the ancestral species diverged substantially during their separate evolutionary journeys traced since the time of their most recent common progenitor.

This polyploid structure of the peanut genome, with highly similar homeologous segments presents formidable challenges to the direct characterization of the DNA sequence of the whole genome. This is because the high similarity of the homeologous component genome sequences makes it difficult to assign individual sequence reads to the specific component genome from which they are derived. In light of this, the PCG targeted the diploid ancestors, *A. ipaënsis* and *A. duranensis* for sequencing.

How Stable Have the Ancestral Genomes Been since Polyploidy?

At the level of the genome, polyploids have new possibilities for change afforded by their greater genetic redundancy and the interaction of genomes that had previously been isolated (Wendel, 2000). Some of these changes, such as differential gene silencing and point mutations by transposons have limited effects on overall genome structure; but others, such as large-scale deletions and intergenomic translocations will have substantial impact. Below we will examine what impact polyploidy may have had on the component genomes of cultivated peanut.

Generally there is a tendency for polyploids to loose DNA. For cultivated peanut, published estimates of genome size vary considerably. However, the largest variations seem to be due to the methods used by various researchers. For some unidentified reason, estimates made with Feulgen densitometry are roughly double those made with flow cytometry (Guimarães et al., 2008; Samoluk et al., 2014, 2015). Subsequent studies have given more support to the

smaller estimates made with flow cytometry and here we will confine discussion to this methodology. Temsch and Greilhuber (2000, 2001) estimated the haploid genome of cultivated peanut at 2.8 Gbp, a value that closely agrees with the 2.74 Gbp estimate of Samoluk et al. (2014). The later study also estimates the sizes of *A. duranensis* and *A. ipaënsis* as 1.25 and 1.56 Gbp, respectively. The sum of these diploid sizes is very similar to the estimated genome size of the tetraploid indicating that the cultivated peanut genome did not undergo any very large loss (or gain) of DNA during or after the polyploidy event (Samoluk et al., 2014).

Many polyploids undergo large-scale chromosomal rearrangement. For peanut, GISH using whole genomic DNA of *A. duranensis* and *A. ipaënsis* as probes neatly distinguishes the A and B chromosomes (Seijo et al., 2007) and does not reveal any detectable mosaic or chimeric hybridization patterns, this indicates that there have not been very many large-scale translocations between the A and B chromosomes since polyploidization. However, at least five secondary constriction types have been observed in different varieties of *A. hypogaea* which indicate that translocations may have been important for subspecific differentiation in peanut (Stalker and Dalmacio, 1986). Further evidence comes from hybridization and genetic studies. An artificially induced allotetraploid derived from *A. ipaënsis* and *A. duranensis* produced fertile hybrids with varieties of both subspecies of cultivated peanut (Fávero et al., 2006). Furthermore, recombinant inbred lines derived from a hybrid of cultivated peanut and this artificially induced allotetraploid are vigorous and fertile. Interestingly marker segregation distortion for these recombinant inbred lines was 19% (Shirasawa et al., 2013), less distortion has been observed for many cultivated × cultivated crosses: 7.8, 13.5, and 23% (Hong et al., 2010); 3.5–52.3% (Gautami et al., 2012); 32 and 27% (Qin et al., 2012); and 8 and 42% (Shirasawa et al., 2012). These lines of evidence all indicate that the large-scale genomic structure of the recent artificially induced allotetraploid [*A. ipaënsis* K 30076 × *A. duranensis* V 14167]^{4x} is similar to that of the older spontaneous allotetraploid *A. hypogaea.*

Although it seems that the ancestral diploid genomes have remained quite stable since polyploidy, there must have been some changes. At some level, changes in DNA must explain how *A. hypogaea*, which has a very narrow genetic base and extremely low DNA polymorphism, was transformed by domestication into one of the world's most important crops that is completely distinct in plant architecture, seed size, and pod form from its wild ancestors. This transformation is in stark contrast to the much more genetically diverse diploid species, which have been cultivated for at least the same amount of time, but only gave rise to a couple of protodomesticate species cultivated on a very limited scale by indigenous people to this day (Freitas, 2004; J.F.M. Valls, personal communication). Also, cultivated peanut exhibits a remarkable amount of morphological variability and two subspecies are recognized (*hypogaea* Krap. et Rig and *fastigiata* Waldron). These subspecies are morphologically different: subspecies *hypogaea* having a long life cycle, no flowers on the central stem,

and regularly alternating vegetative and reproductive side stems: subspecies *fastigiata* having a shorter life cycle, flowers on the central stem, and reproductive and vegetative stems distributed in a disorganized way. There is even evidence of partial reproductive isolation of these two subspecies (Krapovickas, 1969; Leal-Bertioli et al., 2014). Further, hybrid vigor is observed when cultivated lines in different subspecies are hybridized by not among lines within a subspecies (Wynne and Coffelt, 1982), which indicates genetic differentiation between subspecific groups.

One well-documented change in the diploid genomes following polyploidy is in their ribosomal DNAs. Their distribution in *A. hypogaea* is equivalent to the sum of those present in *A. duranensis* and *A. ipaënsis*. However, while in both diploid species the 18–25S sites bear a thread-like constriction indicating intense transcriptional activity (forming the SAT chromosome (Fernández and Krapovickas, 1994)), in the allotetraploid the constrictions are observed only on the A genome (Seijo et al., 2004). Furthermore, an analysis of 5S loci in the diploid and tetraploid species showed that in *A. hypogaea* there has been an almost complete replacement of the B genome 5S loci with A genome copies (Grabiele et al., 2012). However, this phenomenon is common in polyploids and is called nucleolar dominance (Cermeno et al., 1984; Preuss and Pikaard, 2007) but it is not of major consequence for genome structure.

The genome structure of plants is dominated by transposons and it has been shown that the "genomic shock" (McClintock, 1984) following the formation allopolyploids can lead to transposon activity and consequent genomic changes (Adams and Wendel, 2005). For peanut, as previously mentioned, the overall patterns of GISH indicate that there has not been a very large-scale movement of transposons between the genomes. Studies of the abundant Ty3-gypsy retrotransposon FIDEL, the Ty1-copia retrotransposon Matita, and the LINE element ALI support this by showing that they have not undergone large-scale changes of genomic distribution following the polyploidy event that gave rise to cultivated peanut (Nielen et al., 2010, 2012; Samoluk et al., 2015). This contrasts with evidence regarding one type of small transposon, miniature inverted-repeat transposable elements (MITEs). These transposons are nonautonomous DNA transposons and less than 600 bp in length. In peanut, Patel et al. (2004) reported that following treatment with a chemical mutagen, a MITE insertion caused functional disruption of the fatty-acid desaturase-encoding gene *ahFAD2B*, one of the homeologous genes controlling the very important quality trait of high oleic/linoleic fatty acid ratio in peanut seeds. This MITE did not belong to the most common Tourist or Stowaway families but showed similarities to the Bigfoot family in *Medicago* (Charrier et al., 1999). Later, AhMITE1, a transposon with sequence similarities to the previously reported MITE, was observed to excise from a single locus in spontaneous and artificially induced mutants (Gowda et al., 2010, 2011).

Evidence of activity and a tendency to transpose into genes or their flanking regions (Feschotte et al., 2002) stimulated further interest in MITEs, and

Shirasawa et al. (2012) undertook a large-scale analysis. Using enriched genomic libraries, 504 unique AhMITE1 sequences and their flanking genomic regions were obtained and shown to group into six families. Intriguingly, southern blots showed multiple AhMITE1 copies in the genomes of *Arachis magna* (a wild diploid B genome species very closely related to *A. ipaënsis*) and *A. hypogaea*, but not in the genome of *A. duranensis* (the most probable A genome donor to peanut). This suggests that AhMITE1 elements amplified in the B genome, but not in the A genome after their divergence about 3.5 Mya. Surveying of AhMITE1 insertion sites in cultivated varieties by polymerase chain reaction (PCR) showed 13% polymorphism within a small sample of the Virginia runner type and 30% polymorphism between three Virginia cultivars and a Spanish type. This clearly indicates large-scale activity of AhMITE1 elements since the formation of the cultivated peanut and indicates the possibility that transposition events from the B to the A genome may have occurred in this tetraploid. The distribution of AhMITE1 markers in all the linkage groups of cultivated peanut support that this migration has happened (Shirasawa et al., 2012). This conclusion is compatible with the apparent equivalence in GISH patterns of peanut and synthetic allotetraploid mentioned above because MITEs are small and their movement would not be expected to significantly change genome-wide chromosome hybridization patterns.

The insertion rate of AhMITE1 into BLASTX detectable genes (10.5%) is much more frequent than would be expected by chance, indicates that this family of transposons is likely to have affected the expression of numerous genes since the formation of the tetraploid, and may have had an important role in the generation of present-day morphological diversity of cultivated peanuts (Shirasawa et al., 2012).

We uncovered unexpected evidence that there may have also been genetic exchange driven by meiotic recombination between the A and B genomes following polyploidy. It has long been known that meiotic chromosomes in *A. hypogaea* consist of 20 chromosome bivalents in 88–98% of cells, exceptions being rare univalents, trivalents, and quadrivalents (Husted, 1936; Smartt et al., 1978). This suggests limited homeologous pairing between A and B genomes (Singh and Moss, 1982; Wynne and Halward, 1989). This was supported by molecular data from the first map construction in allotetraploid *Arachis*. The study used a BC_1 population derived from a cross of *A. hypogaea* with an artificially induced allotetraploid [*A. batizocoi* × (*A. cardenasii* × *Arachis diogoi*)]^{4x} (termed an amphidiploid in the original paper by Burow et al. (2001)). Although there were some anomalies, 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 (Foncéka et al., 2009, 2012; Gautami et al., 2012; Hong et al., 2010; Qin et al., 2012; Robledo and Seijo, 2010; Shirasawa et al., 2012, 2013; Sujay et al., 2012; Varshney et al., 2009; Wang et al., 2012). However, in a detailed reanalysis of previously published genotyping data of recombinant inbred lines derived

from cultivated peanut and the induced allotetraploid [*A. ipaënsis* K 30076× *A. duranensis* V 14167]^{4x} (Bertioli et al., 2014b; Leal-Bertioli et al., 2015; Shirasawa et al., 2013), we observed that infrequent unexpected data points formed patterns. This implied the involvement of some biological phenomena for which we cannot account. PCR-based marker data indicated the rare complete disappearance of A or B alleles. These disappearances were genetically grouped. Furthermore, SNP markers indicated that for many markers, five possible allelic states (tetrasomic, trisomic, disomic, monosomic, and nullisomic) were present. Monosomic and nullisomic states were rare, but they too were genetically grouped, indicating that they were not due to errors in marker assays. This type of behavior is expected in autotetraploids like potato, but unexpected in allotetraploids like peanut. Although only about 3% of observed datapoints indicated tetrasomic recombination, the impact of this recombination on the genomic structure of the Recombinant Inbred Lines (RILs) was surprisingly large. Most lines showed some evidence for tetraplex/nulliplex genomic regions and many of these regions covered substantial chromosome proportions or even whole linkage group arms.

Because the study by Burow et al. (2001) was carried out using lines derived from a cultivated peanut crossed with an induced allotetraploid, the applicability of these observations to pure cultivated peanut lines is not completely certain. However, the very high similarity of A and B genomes in genic regions, the fact that the induced allotetraploid in this study was derived from the most likely ancestral species to cultivated peanut, and the cytogenetic studies that show a small amount of tetravalent pairing in meiosis (Singh and Moss, 1982; Wynne and Halward, 1989) suggest that it is most probable that cultivated peanut is a "segmental allotetraploid" (*sensu,* Stebbins, 1947, 1950) with predominantly disomic, but partly tetrasomic genetic behavior. Although almost certainly infrequent, tetrasomic recombination is likely to have had some impact on the genome of *A. hypogaea* and may even have an unexpected importance in genetics and breeding programs today.

CONCLUSIONS

Cultivated peanut (*A. hypogaea* L.) has an estimated genome size of about 3 Gb, similar to the human genome. It is an allotetraploid ($2n=4x=40$) with an AABB genome type. The most probable ancestral species are *A. duranensis* and *A. ipaënsis.* The A and B genomes diverged quite recently in evolutionary terms, about 3 million years ago. This polyploid structure of the genome presents formidable challenges to the direct characterization of the DNA sequence of the whole genome. This is because the homeologous component genome sequences are closely related making it difficult to assign individual sequence reads to the specific component genome from which they are derived. Because of this difficulty, the PCG targeted the diploid ancestors *A. ipaënsis* and *A. duranensis* for sequencing. Various lines of evidence indicate that the changes in the ancestral genomes since polyploidy have been limited and that the sum of the diploid genomes is a

good approximation to the genome of cultivated peanut. The generation of whole genome assemblies for the diploid ancestors of peanut will prove a key landmark in the understanding of the genetics and genome structure of the peanut crop.

REFERENCES

Adams, K.L., Wendel, J.F., 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. 8, 135–141.

Bertioli, D.J., Araujo, A.C.G., Nielen, S., Heslop-Harrison, P., Guimarães, P.M., Schwarzacher, T., Isobe, S., Shirasawa, K., Leal-Bertioli, S.C.M., 2014a. An overview of peanut genome structure. In: Mallikarjuna, N., Varshney, R.K. (Eds.), Genetics, Genomics and Breeding of Peanuts. CRC Press, Boca Raton, FL, USA, 266p.

Bertioli, D.J., Ozias-Akins, P., Chu, Y., Dantas, K.M., Santos, S.P., Gouvea, E., Guimarães, P.M., Leal-Bertioli, S.C.M., Knapp, S.J., Moretzsohn, M.C., 2014b. The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. G3 Genes Genomes Genet. 4, 89–96.

Bertioli, D.J., Vidigal, B., Nielen, S., Ratnaparkhe, M.B., Lee, T.H., Leal-Bertioli, S.C.M., Kim, C., Guimarães, P.M., Seijo, G., Schwarzacher, T., Paterson, A.H., 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.

Boerma, H.R., St John, J., Molen, J.Y., 2001. U.S. Legume Crops Genomics Workshop, July 30-31,2011. White Paper, Hunt Valley, MD. 29 pp.

Burow, M.D., Simpson, C.E., Starr, J.L., Paterson, A.H., 2001. Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.). Broadening the gene pool of a monophyletic polyploid species. Genetics 159, 823–837.

Cannon, E., Dash, S., Kalberer, S., Weeks, N., Huang, W., Farmer, A., Cannon, S., Dickerson, J., July 6–11, 2014. PeanutBase: The genomic data portal for *Arachis*. In: 7th International Conference on Legume Genetics and Genomics, Saskatoon, Canada.

Cermeno, M.C., Orellana, J., Santos, J.L., Lacadena, J.R., 1984. Nucleolar activity and competition (Amphiplasty) in the genus *Aegilops*. Heredity 53, 603–611.

Charrier, B., Foucher, F., Kondorosi, E., D'Aubenton-Carafa, Y., Thermes, C., Kondorosi, A., Ratet, P., 1999. Bigfoot: A new family of MITE elements characterized from the *Medicago* genus. Plant J. 18, 431–441.

Cuc, L.M., Mace, E.S., Crouch, J.H., Quang, V.D., Long, T.D., Varshney, R.K., 2008. Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea* L.). BMC Plant Biol. 8, 55.

Dash, S., Cannon, E.K.S., Kalberer, S., Huang, W., Weeks, N., Dickerson, J., Farmer, A., Cannon, S., November 10–14, 2014. PeanutBase: a community resource to help improve peanut varieties by integrating genetic, genomic, and trait information. In: AAGB 2014: Advances in *Arachis* through Genomics & Biotechnology, Savannah, GA. http://www.peanutbioscience.com (abstract).

Devos, K.M., Brown, J.K.M., Bennetzen, J.L., 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res. 12, 1075–1079.

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

Fernández, A., Krapovickas, A., 1994. Cromosomas y Evolucion en *Arachis* (Leguminosae). Bonplandia 8, 187–200.

Feschotte, C., Jiang, N., Wessler, S.R., 2002. Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. 3, 329–341.

Foncéka, D., Tossim, H.-A., Rivallan, R., Vignes, H., Faye, I., Ndoye, O., Moretzsohn, M.C., Bertioli, D.J., Glaszmann, J.C., Courtois, B., 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 Biol. 12, 26.

Foncéka, D., Hodo-Abalo, T., Rivallan, R., Faye, I., Sall, M., Ndoye, O., Fávero, A., Bertioli, D.J., Glaszmann, J.-C., Courtois, B., 2009. Genetic mapping of wild introgressions into cultivated peanut: a way toward enlarging the genetic basis of a recent allotetraploid. BMC Plant Biol. 9, 103.

Freitas, F.O., 2004. Ampliação da Área de Ocorrência da Espécie. In: Arachis villosulicarpa Embrapa Comunicado Técnico- Embrapa, Brasilia, Brazil.

Gautami, B., Foncéka, D., Pandey, M.K., Moretzsohn, M.R.C., Sujay, V., Qin, H., Hong, Y., Faye, I., Chen, X., BhanuPrakash, A., 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.

Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F., Young, N.D., 2005. Legumes as a model plant family. Genomics for food and feed: report of the cross-legume advances through genomics conference. Plant Physiol. 137, 1228–1235.

Gowda, M., Bhat, R., Sujay, V., Kusuma, P., Varshakumari, V., Bhat, S., Varshney, R., 2011. Characterization of *AhMITE1* transposition and its association with the mutational and evolutionary origin of botanical types in peanut (*Arachis* spp.). Plant Syst. Evol. 291, 153–158.

Gowda, M.V.C., Bhat, R.S., Motagi, B.N., Sujay, V., Kumari, V., Sujatha, B., 2010. Association of high-frequency origin of late leaf spot resistant mutants with *AhMITE*1 transposition in peanut. Plant Breed. 129, 567–569.

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

Gregory, M.P., Gregory, W.C., 1979. Exotic germplasm of *Arachis* L. interspecific hybrids. J. Hered. 70, 185–193.

Gregory, W.C., Gregory, M.P., 1976. Groundnut. In: Simmonds, N. (Ed.), Evolution of Crop Plants. Longman, London, pp. 151–154.

Guimarães, P., Garsmeur, O., Proite, K., Leal-Bertioli, S., Seijo, G., Chaine, C., Bertioli, D., D'Hont, A., 2008. BAC libraries construction from the ancestral diploid genomes of the allotetraploid cultivated peanut. BMC Plant Biol. 8, 14.

Halward, T., Stalker, T., LaRue, E., Kochert, G., 1992. Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.). Plant Mol. Biol. 18, 315–325.

Halward, T.M., Stalker, H.T., LaRue, E.A., Kochert, G., 1991. Genetic variation detectable with molecular markers among unadapted germ-plasm resources of cultivated peanut and related wild species. Genome 34, 1013–1020.

Hong, Y.B., Chen, X.P., Liang, X.Q., Liu, H.Y., Zhou, G.Y., Li, S.X., Wen, S.J., Holbrook, C.C., Guo, B.Z., 2010. A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. BMC Plant Biol. 10, 1–13.

Husted, L., 1936. Cytological studies on the peanut. *Arachis*. II — chromosome number, morphology and behavior, and their application to the problem of the origin of the cultivated forms. Cytologia 7, 396–423.

Kochert, G., Stalker, H.T., Gimenes, M., Galgaro, L., Lopes, C.R., 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., 1969. The origin, variability and spread of the groundnut (*Arachis hypogaea*). In: Ucko, P.J., Dimbleby, G.W. (Eds.), The Domestication and Exploration of Plants and Animals. Duckworth Publishers, London, pp. 427–441.

Krapovickas, A., Gregory, W.C., 1994. Taxonomia del Genero *Arachis* (Leguminosae). Bonplandia 8, 1–186.

Krapovickas, A., Gregory, W.C., 2007. Taxonomy of the genus *Arachis* (Leguminosae). Bonplandia 16 (Suppl.), 1–205.

Leal-Bertioli, S., Shirasawa, K., Abernathy, B., Moretzsohn, M., Chavarro, C., Clevenger, J., Ozias-Akins, P., Jackson, S., Bertioli, D., 2015. Tetrasomic recombination is surprisingly frequent in allotetraploid *Arachis*. Genetics 199, 1093–1105.

Leal-Bertioli, S.C., Santos, S.P., Dantas, K.M., Inglis, P.W., Nielen, S., Araujo, A.C., Silva, J.P., Cavalcante, U., Guimarães, P.M., Brasileiro, A.C., 2014. *Arachis batizocoi*: a study of its relationship to cultivated peanut (*A. hypogaea*) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. Ann. Bot. 115, 237–249.

McClintock, B., 1984. The significance of responses of the genome to challenge. Science 226, 792–801.

Ma, J., Devos, K.M., Bennetzen, J.L., 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. 14, 860–869.

Magallon, S., Sanderson, M., 2001. Absolute diversification rates in angiosperm clades. Evolution 55, 1762–1780.

Mallikarjuna, N., Senthilvel, S., Hoisington, D., 2011. Development of new sources of tetraploid *Arachis* to broaden the genetic base of cultivated groundnut (*Arachis hypogaea* L.). Genet. Res. Crop Evol. 58, 889–907.

Michael, T.P., Jackson, S., 2013. The first 50 plant genomes. Plant Genome 6, 1–7.

Moretzsohn, M.C., Gouvea, E.G., Inglis, P.W., Leal-Bertioli, S.C.M., Valls, J.F.M., Bertioli, D.J., 2013. A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Ann. Bot. 11, 113–126.

Nielen, S., Campos-Fonseca, F., Leal-Bertioli, S., Guimarães, P., Seijo, G., Town, C., Arrial, R., Bertioli, D., 2010. FIDEL — a retrovirus-like retrotransposon and its distinct evolutionary histories in the A- and B-genome components of cultivated peanut. Chromosome Res. 18, 227–246.

Nielen, S., Vidigal, B.S., Leal-Bertioli, S.C., Ratnaparkhe, M., Paterson, A.H., Garsmeur, O., D'Hont, A., Guimaraes, P.M., Bertioli, D.J., 2012. Matita, a new retroelement from peanut: characterization and evolutionary context in the light of the *Arachis* A-B genome divergence. Mol. Genet. Genomics 287, 21–38.

Patel, M., Jung, S., Moore, K., Powell, G., Ainsworth, C., Abbott, A., 2004. High-oleate peanut mutants result from a MITE insertion into the *FAD2 g*ene. Theor. Appl. Genet. 108, 1492–1502.

Pereira, V., 2004. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana g*enome. Genome Biol. 5, R79.

Preuss, S., Pikaard, C.S., 2007. rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. Biochem. Biophys. Acta 1769, 383–392.

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

Ramos, M.L., Fleming, G., Chu, Y., Akiyama, Y., Gallo, M., Ozias-Akins, P., 2006. Chromosomal and phylogenetic context for conglutin genes in *Arachis* based on genomic sequence. Mol. Genet. Genomics 275, 578–592.

Robledo, G., Seijo, G., 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. Theor. Appl. Genet. 121, 1033–1046.

Samoluk, S.S., Robledo, G., Podio, M., Chalup, L., Robledo, G., Ortiz, J.P.A., Pessino, S.C., Seijo, J.G., 2015. First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species. Genetica 143, 113–125.

Samoluk, S.S., Chalup, L., Robledo, G., Seijo, J.G., 2014. Genome sizes in diploid and allopolyploid *Arachis* L. species (section *Arachis*). Genet. Res. Crop Evol. 61, 1–17. http://dx.doi.org/10.1007/s10722-014-0193-3.

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

Seijo, J.G., Lavia, G.I., Fernández, A., Krapovickas, A., Ducasse, D., Moscone, E.A., 2004. Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaënsis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). Am. J. Bot. 91, 1294–1303.

Shirasawa, K., Bertioli, D.J., Varshney, R.K., Moretzsohn, M.C., Leal-Bertioli, S.C., Thudi, M., Pandey, M.K., Rami, J.F., Foncéka, D., Gowda, M.V., et al., 2013. Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes. DNA Res. 20, 173–184.

Shirasawa, K., Hirakawa, H., Tabata, S., Hasegawa, M., Kiyoshima, H., Suzuki, S., Sasamoto, S., Watanabe, A., Fujishiro, T., Isobe, S., 2012. Characterization of active miniature inverted-repeat transposable elements in the peanut genome. Theor. Appl. Genet. 124, 1429–1438.

Singh, A.K., Moss, J.P., 1982. Utilization of wild relatives in genetic improvement of *Arachis hypogaea* L. II. Chromosome complements of species in section *Arachis*. Theor. Appl. Genet. 61, 305–314.

Singh, A.K., Smartt, J., 1998. The genome donors of the groundnut/peanut (*Arachis hypogaea* L.) revisited. Genet. Res. Crop Evol. 45, 113–118.

Smartt, J., Gregory, W., Gregory, M., 1978. The genomes of *Arachis hypogaea*. 1. Cytogenetic studies of putative genome donors. Euphytica 27, 665–675.

Stalker, H.T., Dalmacio, R.D., 1986. Karyotype relations and analysis among varieties of *Arachis hypogaea* L. Cytologia 58, 617–629.

Stalker, H.T., Dhesi, J.S., Parry, D.C., Hahn, J.H., 1991. Cytological and interfertility relationships of *Arachis* section *Arachis*. Am. J. Bot. 78, 238–246.

Stebbins, G.L., 1947. Types of polyploids: their classification and significance. Adv. Genet. 1, 403–429.

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

Sujay, V., Gowda, M.V., Pandey, M.K., Bhat, R.S., Khedikar, Y.P., Nadaf, H.L., Gautami, B., Sarvamangala, C., Lingaraju, S., Radhakrishan, T., et al., 2012. QTL analysis and construction of consensus genetic map for foliar diseases resistance based on two RIL populations in cultivated groundnut (*Arachis hypogaea* L.). Mol. Breed. 32, 773–788.

Temsch, E.M., Greilhuber, J., 2000. Genome size variation in *Arachis hypogaea* and *A. monticola* re-evaluated. Genome 43, 449–451.

Temsch, E.M., Greilhuber, J., 2001. Genome size in *Arachis duranensis*: A critical study. Genome 44, 826–830.

Valentine, H., Grabau, E., Ozias-Akins, P., Nwosu, V., Wilson, R., Gallo, M., Chenault, K., Stalker, H.T., Weissinger, A.W., Donahue, P., et al., 2006. Biotech Peanut White Paper: Benefits and Issues. American Peanut Council, Alexandria, VA, 14 pp.

Varshney, R.K., Bertioli, D.J., Moretzsohn, M.C., Vadez, V., Krishnamurty, L., Aruna, R., Nigam, S.N., Ravi, K., He, G., Knapp, S.J., et al., 2009. The first SSR based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). Theor. Appl. Genet. 118, 729–739.

Vicient, C.M., Suoniemi, A., Anamthawat-Jansson, K., Tanskanen, J., Beharav, A., Nevo, E., Schulman, A.H., 1999. Retrotransposon *BARE*-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell Online 11, 1769–1784.

Wang, H., Penmetsa, R.V., Yuan, M., Gong, L., Zhao, Y., Guo, B., Farmer, A.D., Rosen, B.D., Gao, J., Isobe, S., 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 Biol. 12, 10.

Wendel, J.F., 2000. Genome evolution in polyploids. Plant Mol. Biol. 42, 225–249.

Wilson, R.F., Stalker, H.T., Brummer, E.C., 2004. Legume Crop Genomics. AOCS Press, Champaign, IL.

Wilson, R.F., 2006a. National Strategic Plan for the Peanut Genome Initiative 2004–2008. Version 2.4 USDA, Beltsville, MD. 19 pp. http://www.peanutbioscience.com.

Wilson, R.F., 2006b. National Program Action Plan for the Peanut Genome Initiative: Application of Plant Genomics to Mitigate Peanut Allergy. Version 2.4 USDA, Beltsville, MD, 25 pp. http://www.peanutbioscience.com.

Wilson, R.F., 2012. International Peanut Genomic Research Initiative Strategic Plan for 2012 to 2016 Characterization of the Peanut Genome September 2011. Version 4.3. http://www.peanutbioscience.com.

Wynne, J.C., Halward, T., 1989. Genetics and cytogenetics of *Arachis*. Crit. Rev. Plant Sci. 8, 189–220.

Wynne, J.C., Coffelt, T.A., 1982. Genetics of *Arachis hypogaea* L. In: Pattee, H.E., Young, C.T. (Eds.), Peanut Science and Technology. American Peanut Research and Education Society, Inc., Yoakum, TX, pp. 50–94.