American Journal of Botany

RESEARCH ARTICLE

Segmental allopolyploidy in action: Increasing diversity through polyploid hybridization and homoeologous recombination

Soraya C. M. Leal-Bertioli[1,5], Ignácio J. Godoy[2], João F. Santos[2], Jeff J. Doyle[3], Patrícia M. Guimarães[4], Brian L. Abernathy[1], Scott A. Jackson[1], Márcio C. Moretzsohn[4], and David J. Bertioli[1]

Manuscript received 28 February 2018; revision accepted 18 April 2018.

[1] University of Georgia, Center for Applied Genetic Technologies, 111 Riverbend Road, Athens, GA 30602-6810, USA

[2] Campinas Agronomical Institute, Avenida Barão de Itapura, 1.481, Campinas, SP 13020-902, Brazil

[3] Cornell University, School of Integrative Plant Science, Plant Breeding & Genetics Section, Ithaca, NY 14853, USA

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

[5] Author for correspondence (e-mail: sorayab@uga.edu)

Citation: Leal-Bertioli, S. C. M., I. J. Godoy, J. F. Santos, J. J. Doyle, P. M. Guimarães, B. L. Abernathy, S. A. Jackson, et al. 2018. Segmental allopolyploidy in action: Increasing diversity through polyploid hybridization and homoeologous recombination. *American Journal of Botany* 105(6): 1053–1066.

doi:10.1002/ajb2.1112

PREMISE OF THE STUDY: The genetic bottleneck of polyploid formation can be mitigated by multiple origins, gene flow, and recombination among different lineages. In crop plants with limited origins, efforts to increase genetic diversity have limitations. Here we used lineage recombination to increase genetic diversity in peanut, an allotetraploid likely of single origin, by crossing with a novel allopolyploid genotype and selecting improved lines.

METHODS: Single backcross progeny from cultivated peanut × wild species-derived allotetraploid cross were studied over successive generations. Using genetic assumptions that encompass segmental allotetraploidy, we used single nucleotide polymorphisms and whole-genome sequence data to infer genome structures.

KEY RESULTS: Selected lines, despite a high proportion of wild alleles, are agronomically adapted, productive, and with improved disease resistances. Wild alleles mostly substituted homologous segments of the peanut genome. Regions of dispersed wild alleles, characteristic of gene conversion, also occurred. However, wild chromosome segments sometimes replaced cultivated peanut's homeologous subgenome; *A. ipaënsis* B sometimes replaced *A. hypogaea* A subgenome (~0.6%), and *A. duranensis* replaced *A. hypogaea* B subgenome segments (~2%). Furthermore, some subgenome regions historically lost in cultivated peanut were "recovered" by wild chromosome segments (effectively reversing the "polyploid ratchet"). These processes resulted in lines with new genome structure variations.

CONCLUSIONS: Genetic diversity was introduced by wild allele introgression, and by introducing new genome structure variations. These results highlight the special possibilities of segmental allotetraploidy and of using lineage recombination to increase genetic diversity in peanut, likely mirroring what occurs in natural segmental allopolyploids with multiple origins.

KEY WORDS *Arachis*; Fabaceae; introgression; lineage recombination; peanut; polyploid; pre-breeding; segmental allotetraploid; tetrasomic recombination; wild species.

The 20th century has been marked by an unprecedented growth in food production. A large part of this increase can be attributed to genetic improvement of crops. However, in many cases, increases in yield have been associated with a narrowing of the genetic diversity in elite cultivars that may in fact limit the potential for future crop improvement. Accordingly, several major crops have reached a yield plateau (Tester and Langridge, 2010; www://faostat3.fao.org/home/).

Many major crop plants are polyploid (Renny-Byfield and Wendel, 2014), and as such serve as models for studying the process of whole-genome duplication, which is a major force in plant evolution (Van de Peer et al., 2017). Polyploids in general are also models for studying heterosis (Washburn and Birchler, 2014); allopolyploids such as peanut and wheat combine two or more diploid genomes and are "fixed hybrids". However, genetic diversity is also an issue for allopolyploids, because polyploid formation creates a genetic bottleneck. The polyploidy bottleneck can be mitigated by multiple origins involving different genotypes (Soltis et al., 2004), and gene flow between the polyploids of different origins and lineages

(e.g., Soltis and Soltis, 1993, 1999; Doyle et al., 1999; Welles and Ellstrand, 2016). However, recombination among subgenomes can lead to loss of genetic diversity, even in allopolyploids (Ramsey and Schemske, 2002; Gaeta and Pires, 2010; Wendel, 2015), let alone in autopolyploids, where random pairing among two or more sets of homologous chromosomes is expected; the same problem can occur in plants with intermediate pairing behavior termed segmental allopolyploids (Stebbins, 1971).

Plant species belonging to the genus of peanut (*Arachis hypogaea* L.) are endemic to South America. They are found over a wide area and in diverse environments and comprise 81 described species, grouped in nine taxonomical sections, according to their morphology, geographic distribution, and cross-compatibility (Krapovickas and Gregory, 1994; Valls and Simpson, 2005). Peanut and its approximately 30 most closely related wild species reside in the type section *Arachis*. Their karyotypes have been divided into five groups, named A, B, K, F, and D, with the majority being A or B. Most species are diploid ($2n = 2x = 20$), whilst *A. hypogaea* and its wild equivalent (*A. monticola* Krapov. & Rigoni) are allotetraploids ($2n = 4x = 40$ with AABB type genome). Peanut is derived from a cross of the diploid species *A. duranensis* (Krapov. & W.C.Gregory) and *A. ipaënsis* (Krapov. & W.C.Gregory), that contributed the A and B subgenomes, respectively (Fernández and Krapovickas, 1994; Krapovickas and Gregory, 1994). Unlike several other polyploids that have multiple origins (e.g., *Tragopogon* and *Galax*, Soltis et al., 2012; Servick et al., 2015), a single origin is considered most likely for peanut (Bertioli et al., 2011), which severely restricts the number of alleles within the gene pool in two ways. First, alleles from only two founding individuals contribute to the initial gene pool, and second, any genetic changes that occur in the immediate aftermath of the polyploidy event (Wendel, 2015; Gaeta and Pires, 2010) are likely to become fixed in the whole population.

An important route to increase genetic diversity is the use of wild species in breeding (Hajjar and Hodgkin, 2007; Stalker, 2017). However, because of agronomically unadapted phenotypes, the use of wild species has been rare. For peanut, such use is complicated by barriers in sexual compatibility; almost all wild relatives are diploid. Introgression has been achieved using two main schemes that manipulate ploidy, referred to as the hexaploid and tetraploid routes (Stalker et al., 1979; Simpson, 1991; Simpson et al., 2001). In the hexaploid route, the tetraploid peanut is crossed with a wild diploid species, the resulting sterile triploid hybrid is treated with colchicine to create a hexaploid plant, and after several generations of selfing, plants return to the tetraploid state (Stalker et al., 1979). In the tetraploid route, currently the most utilized strategy, a B genome wild species is crossed with an A genome wild species to create a sterile AB hybrid, which is treated with colchicine to double the chromosome number and restore fertility. However, these complex crossing schemes and multiple generations of selection needed to produce improved lines, combined with difficulties of genetic analysis, have often left poorly defined the actual genetic contribution of the wild species. Well-defined contributions of wild species to commercial peanut cultivars are limited to a large introgression on chromosome A09 from the A genome species *A. cardenasii* Krapov. & W.C.Greg. that confers resistance to root-knot nematode (Chu et al., 2011; Nagy et al., 2012; Clevenger et al., 2017), an introgression at the lower end of A03 from the same species that confers resistance to rust (Kolekar et al., 2016), and three peanut cultivars that harbor distinct chromosome segments from *A. duranensis* Krapov. & W.C.Greg. and *A. ipaënsis* Krapov. & W.C.Greg. (Faye et al., 2016). In these cases, the wild chromosome segments have been reported to substitute the homologous segments from the corresponding subgenome of *A. hypogaea*, in accordance with the common working assumption of classic diploid-like allotetraploid genetics in cultivated peanut and induced allotetraploids (the A subgenome only recombines with the A genome, and the B subgenome only recombines with the B). However, there have been recent reports of limited autotetraploid genetic behavior (A and B subgenomes recombine) in progenies of cultivated peanut and induced allotetraploids (Leal-Bertioli et al., 2015a; Bertioli et al., 2016a; Nguepjop et al., 2016). Thus, it is likely that this genetic behavior has had a previously overlooked impact on the genome and genetics of cultivated peanut and, therefore, of breeding.

Here, we describe the development of peanut lines that incorporate wild alleles. The lines were developed using the tetraploid route of introgression, advancing progeny by single seed descent after a single backcross between *A. hypogaea* cv. Runner IAC-886 and a wild-derived induced allotetraploid, created with accessions of the same species as the peanut progenitors [*A. ipaënsis* × *A. duranensis*]^{4x} (Fávero et al., 2006) ($4x$ indicates ploidy as in $2n = 4x = 40$). Except for the yellow flowers, the general appearance of these peanut lines is of cultivated peanuts of pure pedigree. Using the improved knowledge base and genetic tools that have become available since the sequencing of cultivated peanut's diploid ancestors (*A. duranensis* V14167 and *A. ipaënsis* K30076; Bertioli et al., 2016a; Clevenger et al., 2017; Pandey et al., 2017), we were able to analyze genetic composition in unprecedented detail. Considering their cultivated phenotypes, the lines harbored a surprisingly high percentage of introgressed chromatin. Many introgressions cover large portions of chromosomes. In addition to revealing the introduction of new alleles, the analysis also showed that genetic diversity was introduced by new variations in allotetraploid genome structure. This experimental program is analogous to the process of multiple polyploid origins followed by gene flow and recombination between homeologous chromosomes and affords insights into how diversity can arise in natural polyploid species.

MATERIAL AND METHODS

Initial plant material

The primary cross was performed using an early generation induced allotetraploid [*A. ipaënsis* K30076 × *A. duranensis* V14167]^{4x} (hereafter called IpaDur1), as male (donor) parent, and *A. hypogaea* cv. Runner IAC-886 as female (recurrent parent; for short, here called Runner-886), one of the most widely planted peanut cultivars in Brazil. In the single backcross (BC), Runner-886 was used as the recurrent female parent.

Evaluation and advancement of BC_1 interspecific lines

True BC_1F_1 hybrids were identified using microsatellite markers. BC_1F_2 seeds were harvested from confirmed hybrids. For the first season of selection, 38 BC_1F_2 families were planted in the field at Embrapa Cerrados (latitude 15°35′30″, longitude 47°42′30″) on 30 December 2008. Ten seeds of each family were planted in a randomized block design. Seeds were sown with 1 m intra- and 1.5 m inter-row spacing. To increase disease pressure, *A. hypogea* cv. IAC-Tatu, a Valencia-type cultivar very susceptible to foliar diseases,

was sown in every 11th row. No disease control was applied. Leaves were inspected weekly for the appearance of symptoms of late leaf spot (LLS, *Cercosporidium personatum* syn. *Passalora personata*). Incubation period (time for the appearance of first lesion) was recorded. LLS severity was scored 120 days after sowing (DAS) using a 0–9 visual scale based on defoliation and number of fungal lesions (see Kumari et al., 2014). Also, the area of the leaves affected by LLS was measured by image analysis of 5–10 leaves from the middle third portion of four branches of each plant; leaves were scanned, and percentage diseased leaf area (DLA) was calculated using the program Quant (Vale et al., 2001).

Growth habit was evaluated using a visual scale of 1–10, where 1 indicated a plant with only one or two lateral branches without secondary ramifications (an extreme wild-type phenotype), and 10, a plant with multiple branches with secondary ramifications forming a closed, exceptionally regular runner peanut-type canopy. Plants were harvested between 133 and 137 days after planting. Seeds were manually harvested, shelled, counted and weighed. Twenty-five BC_1F_2 plants were selected based on LLS resistance, growth habit, and seed production.

Progeny from these selected plants, as BC_1F_3 seeds, were taken to a second season of field selection using 20 seeds of each family, planted in a randomized block design. Traits evaluated were: growth habit, yield, and LLS resistance (diseased leaf area, and visual score of severity). Thirteen BC_1F_3 plants were selected and their progeny evaluated the following season.

BC_1F_4 seeds were planted in two rows of 20 plants each for a third season of evaluation of LLS resistance (visual score of severity). Since peanut rust (*Puccinia arachidis*) was not consistently found in the field, detached leaf assays were conducted as described by Leal-Bertioli et al. (2015b) using rust uredospores collected from infested field-grown susceptible plants. Resistance was evaluated as incubation period, total number of lesions, and sporulating lesions per leaf area. Harvested seeds were counted and weighed.

For the final season of evaluation, BC_1F_5 plants, were evaluated for LLS resistance and agronomic and productivity characters.

Phenotypic data were statistically analyzed using the software R (R Core Team, 2010). Data normality was verified using the Shapiro–Wilk test. Average comparisons were done using standard variance analyses, followed by the Tukey test (data with normal distribution) or nonparametric test of Kruskal–Wallis (data with non-normal distribution) at 5% significance levels.

Genotyping and genetic analysis

Genotyping was done using an Affymetrix 58,233 single nucleotide polymorphism (SNP) array (Axiom_Arachis array v01, Clevenger et al., 2017; Pandey et al., 2017). Data were analyzed using Axiom Analysis Suite v.1.1.0.616. Output was filtered and transformed using simple Unix and Perl scripts (Appendix S1, see the Supplemental Data with this article) and data visualized in Excel v.16.12 (Microsoft, Redmond, WA, USA).

SNP assays were done on nine BC_1F_5 lines (BC1 111-10-110, BC1 111-10-121, BC1 111-10-231, BC1 111-4-392, BC1 135-1-107, BC1 135-1-257, BC1 135-1-473, BC1 170-2-56, BC1 37-6-589), their parents, *A. hypogaea* Runner IAC-886 and [*A. ipaënsis* K30076 × *A. duranensis* V14167]^{4x} and several wild diploid *Arachis* species with well-defined phylogenetic relationships: the A genome species *A. duranensis* (accessions V14167, SeSn2848 and K7988), the A genome *A. stenosperma* Krapov. & W.C.Greg. (accessions V10309, V15076, H410); the B genome species sensu lato (K genome sensu stricto; Robledo and Seijo, 2010) *A. batizocoi* Krapov. & W.C.Greg. (accession K9484); the B genome *A. magna* Krapov., W.C.Greg. & C.E.Simpson (accessions K30097 and GKSSc30092) and the B genome *A. ipaënsis* (accession K30076).

Overview of single nucleotide polymorphism analysis—Preliminary tests indicated that the software could not effectively call multiple tetraploid and diploid genotypes together. Therefore, our strategy for processing SNP assay results was divided into different steps. First, informative assays were identified from a SNP calling using predominantly diploid samples. Second, these informative assays were extracted from SNP calling of exclusively tetraploid genotypes. Finally, the different classes of informative SNPs (see below) were processed and visualized in Excel sheets separately.

Identification of introgressions—Our general strategy for detecting wild introgressions in the *A. hypogaea* genome was to identify SNPs that were characteristic (in phylogenetic terminology, autapomorphic) to each of the three-parental species in the pedigree of the selected lines. First, characteristic markers for each species were identified in the SNP calling, which used a panel of diploid genotypes plus a single tetraploid genotype (*A. hypogaea* Runner IAC-886) as follows:

A. duranensis characteristic markers: *A. duranensis* ≠ (*A. hypogaea* = *A. ipaënsis*)

A. ipaënsis characteristic markers: *A. ipaënsis* ≠ (*A. hypogaea* = *A. duranensis*)

A. hypogaea characteristic markers: *A. hypogaea* ≠ (*A. duranensis* = *A. ipaënsis*)

Second, these three sets of characteristic markers were extracted from the SNP calling done using only the tetraploid genotypes of interest.

Overall, our rationale took advantage of the genotypes studied being highly homozygous. If we consider an *A. duranensis* characteristic marker, it could, for instance, be a homozygous G^aG^a, where *A. ipaënsis* is C^bC^b and Runner-886 is $C^aC^aC^bC^b$ (superscript letters "a" and "b" indicate the genome type). Regions of *A. duranensis* introgression would normally be expected to be $G^aG^aC^bC^b$. Thus, because of the equal dosage of G and C bases, these regions are interpreted, incorrectly, by the Axiom software as "heterozygote". However, they are in fact, disomic homozygous introgressions. Filtering and processing were done with simple Unix scripts (Appendix S1). Because of genetic exchange between A and B genomes, this rationale does not always hold, and this phenomenon was accounted for by the steps outlined below.

Detecting recombination between A and B-subgenomes and gene conversion using the Axiom_Arachis array—Most assays in the Axiom_Arachis array detect SNPs in both A and B genomes. Therefore, we could use the rationale described by Leal-Bertioli et al. (2015a) to identify genotypes that can only arise through recombination between subgenomes. Two scenarios, explained below, were studied. We only take account of homozygous states because of the highly selfed genotypes studied here.

In the first scenario, the genotypes of the parents of the original cross are of the type $T^aT^aC^bC^b$ and $C^aC^aC^bC^b$ (superscript letters "a" and "b" indicate the subgenome of origin, bases T and C are used as exemplars of SNP polymorphism, any other combination of bases could be used). Genotypes arising through disomic recombination

would be nulliplex or duplex for the base T ($C^aC^aC^bC^b$, $T^aT^aC^bC^b$). Genotypes tetraplex for the base T ($T^aT^aT^aT^a$) can arise only through recombination between subgenomes. Usually the characteristic alleles of *A. duranensis*, *A. ipaënsis*, or *A. hypogaea* will be expected to be present in the duplex state, and so, be incorrectly called by the Axiom software as "heterozygotes". When tetraplex, they could be expected to be called "homozygote". (These calls were transformed by the Unix scripts to 2DUR, 4DUR; 2IPA, 4IPA; 2HYP, 4HYP indicating the allele dosage, as appropriate; Appendix S1).

The second scenario is where both parental genotypes are the same and 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; any other combination of bases could be used.). With strictly disomic recombination, genotypes of all progeny RILs are the same as the parents and are called by the Axiom software as "heterozygous". (These calls were transformed to AABB, indicating genome composition, by the scripts). However, genotypes of the type $G^aG^aG^aG^a$ and $C^bC^bC^bC^b$ can arise only through recombination between subgenomes. In these cases, the genotypes can be expected to be called by the software as "homozygous" (These calls were transformed by the scripts to AAAA or BBBB, as appropriate). To identify informative assays for this scenario, we aimed to identify SNPs that differentiated the A and B genomes. For that, we used the panel of diploid species accessions assayed for SNPs, identifying the assays that differentiated the A genome accessions (*A. duranensis* accessions and *A. stenosperma*) and the B genome sensu lato species (*A. ipaënsis*, *A. magna*, and *A. batizocoi*). The markers selected for this analysis were defined as follows:

(*A. duranensis* V14167 = *A. duranensis* Se2848 = *A. duranensis* K7988 = *A. stenosperma* V10309) ≠ (*A. batizocoi* K9484 = *A. magna* K30097 = *A. ipaënsis* K30076).

Detecting recombination between A and B subgenomes using low-coverage sequencing—The expected genome composition for allotetraploid peanut is AABB. With the aim of complementing, the SNP analysis and detecting deviations from the expected genome composition, the tetraploid genome structures of selected peanut lines and their parents were analyzed using low-coverage whole-genome sequencing. Short insert Illumina HiSeq2000 pair-end sequencing libraries were constructed and sequenced using the manufacturer's instructions. This analysis generated millions of ~150-bp sequences, most of which could be assigned to the A or B allotetraploid subgenomes by sequence similarity to the combined reference genome sequences of *A. duranensis* and *A. ipaënsis*, donors of the A and B genomes to *A. hypogaea,* respectively (Bertioli et al., 2016a). The sequenced accessions *A. duranensis* V14167 and *A. ipaënsis* K30076 are also the parents of the induced allotetraploid used as the donors of wild alleles in our breeding scheme. The normalized density of sequences assigned to A and B subgenomes, when displayed graphically across chromosomal sequences allowed the tetraploid genome composition to be visualized (see below).

DNAs of selected BC_1F_5 plants, a ninth generation induced allotetraploid [*A. ipaënsis* K30076 × *A. duranensis* V14167]^{4x}, Runner-886, and the diploids *A. ipaënsis* K30076 and *A. duranensis* V14167 were extracted using the Qiagen Plant DNA extraction kit (Qiagen, Germantown, MD, USA). Three micrograms of DNA were used to construct PCR-free DNA libraries using the Kapa-Hyper Prep kit (Illumina Platforms). Libraries were sequenced to an average of 0.83-fold coverage using Illumina NextSeq (300 Cycles) PE150 High Output Flow Cell. Raw reads were trimmed and quality filtered using trim_galore v0.4.2. The filtered reads from each of the genotypes were assigned to A or B chromosomes by "mapping" to the combined chromosomal pseudomolecule sequences of *A. duranensis* V14167 and *A. ipaënsis* K30076 (Bertioli et al., 2016a) using Bowtie2 v2.2.9 (Langmead and Salzberg, 2012) with the -sensitive- local option. Mpileup files were generated using Samtools v1.3 (Li et al., 2009), which were parsed to create average mapping densities for bins of approximately 10 kb covering the entire chromosomes (see below). The average mapping densities (often called "mapping depths") were normalized to 1× coverage by dividing the mapping depths by the trimmed read coverage for each sample. Mapping densities within windows of ~10,000 bp were plotted along the homeologous A and B chromosomes scaled to the same distance along the *x*-axis.

To help visualize the interactions between A and B genomes, windows were defined using synteny blocks detected using DAGchainer (Haas et al., 2004). Taking into account the relative orientation of the blocks and any difference in size, we divided syntenous blocks into corresponding A–B windows of approximately 10 kb (Appendix S2). To display homeologous pairs of chromosomes on the same graph, we used the coordinates of the B genome bins. This method provides a good comparison of syntenous regions of homeologous chromosomes, but does not visualize genomic regions for which DAGchainer could not confidently assign synteny. In this manuscript, we will use the numbering of the chromosomal pseudomolecules of the sequenced genomes of *A. duranensis* and *A. ipaënsis*, which, with one exception, do not have known correspondence to cytogenetic assignments (Bertioli et al., 2016a).

RESULTS

Evaluation and advancement of BC_1 interspecific lines

The induced allotetraploid IpaDur1 has wild-like traits, which are agronomically undesirable: it has trailing branches, produces pods with very long constrictions separating the seeds, and the seeds are smaller than those of cultivated peanut. On the other hand, it has higher levels of resistance to late leaf spot and rust than Runner-886. Successful crosses were obtained between Runner-886 and IpaDur1, and F_1 hybrids were backcrossed with the recurrent parent, Runner-886. Thirty-eight backcrossed genotypes (BC_1), confirmed using SSR markers (data not shown), were then selfed, and the derived families were evaluated in the field and subject to successive rounds of selection. A summary of the selection scheme is described in Fig. 1.

In the first season, the BC_1F_2 lines were evaluated in the field, with no disease control. Disease onset started at around 52 days after planting for most genotypes. IpaDur1 had much lower disease incidence and severity (as indicated by visual score and diseased leaf area). Backcrossed families mostly had intermediate values to those of the parents, with 34 being more resistant than Runner-886 (Appendix S3). Agronomic traits segregated widely between families: all backcrossed families had intermediate values for growth habit; most backcrossed families produced seeds in intermediate numbers between the parents, and four produced more seed than Runner-886, two had higher total seed mass and 100-seed mass (BC1-127 and BC1-173). For selection, each family was scored between 1-5 for each trait. The sum of the scores was ranked, and the top 25 families were selected.

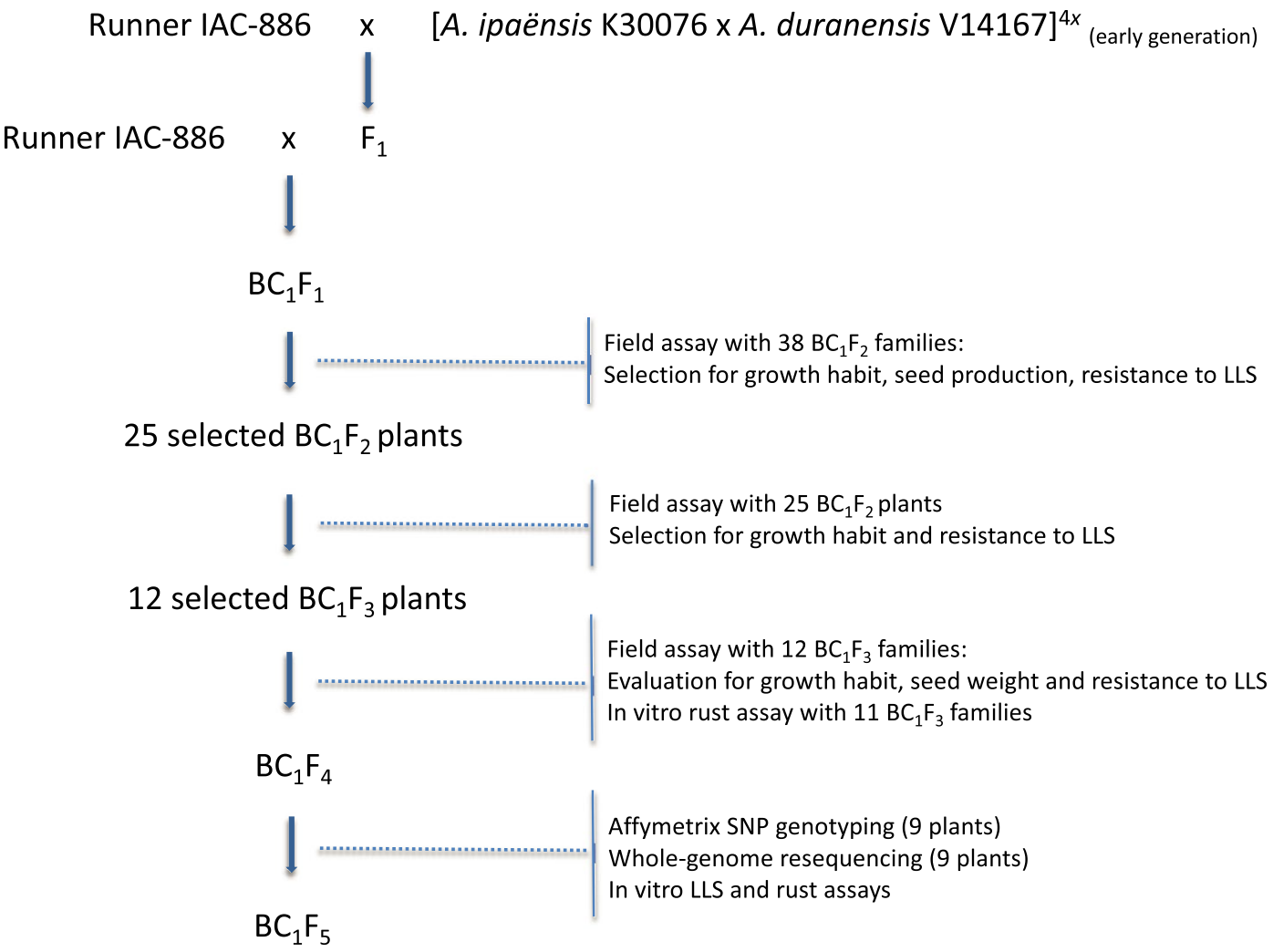

FIGURE 1. Scheme of evaluation and selection of backcrossed lines derived from *A. hypogaea* cv. Runner IAC-886 and the induced allotetraploid IpaDur1.

The 25 BC_1F_3 families were field evaluated in the second season. Growth habit averages shifted to that of the cultivated parent. Based on visual scoring, 21 families were more resistant to LLS than Runner-886 (Appendix S3). The 12 top families were selected.

In the third season, 12 BC_1F_4 families were field evaluated. All families but one (BC1-100-4-249) had lower LLS disease scores than Runner-886. At the end of the growing season, when disease pressure was high, plants of Runner-886 were severely defoliated, while some of the backcrossed families showed improved resistance (Fig. 2). Although the backcrossed lines had not been selected for rust resistance, the in vitro rust assay revealed that two genotypes selected for other traits showed improved resistance to rust (BC1-111-10-121 and BC1-170-2-56) when compared to both parents (Fig. 3A, Table 1). Productivity of 10 of 13 lines was comparable and, in several cases, numerically exceeded that of Runner-886 (Appendix S3).

Ten BC_1F_5 families were evaluated in the fourth season in Pindorama, São Paulo. This region has intense peanut production with higher temperatures and humidity; therefore, disease pressure in this trial was higher than in previous ones. Of 10 lines, seven in field trials and six in laboratory assays had improved resistance to LLS (Fig. 3B, Table 1). All genotypes had domesticated features: compact canopy architecture, large biomass, large seeds (high 100-seed mass), pods with small constriction and large proportion of two-seeded pods (Figs. 3C, D, 4; Table 1). Domestication-related traits were improved at each round of selection (Table 2). Results for all traits for all years are presented in Appendix S3.

Genotyping and genetic analysis

Overview—To analyze the genetic composition of the selected lines, we integrated information from different classes of single nucleotide polymorphisms (SNPs) together with the results of mapping randomly generated whole-genome sequence onto the diploid reference sequences of *A. duranensis* and *A. ipaënsis*. The methodology takes advantage of the diploid genome sequences being from the same genotypes used to make the IpaDur1 allotetraploid, and their having high similarities to the corresponding subgenomes of *A. hypogaea* (Bertioli et al., 2016a). The integration of information of different types was necessary because of the complexity of the genetic structure of the peanut allotetraploids. Whilst there are four genomes segregating (maternal and paternal, A and B genomes), SNP markers are biallelic; they can only detect two

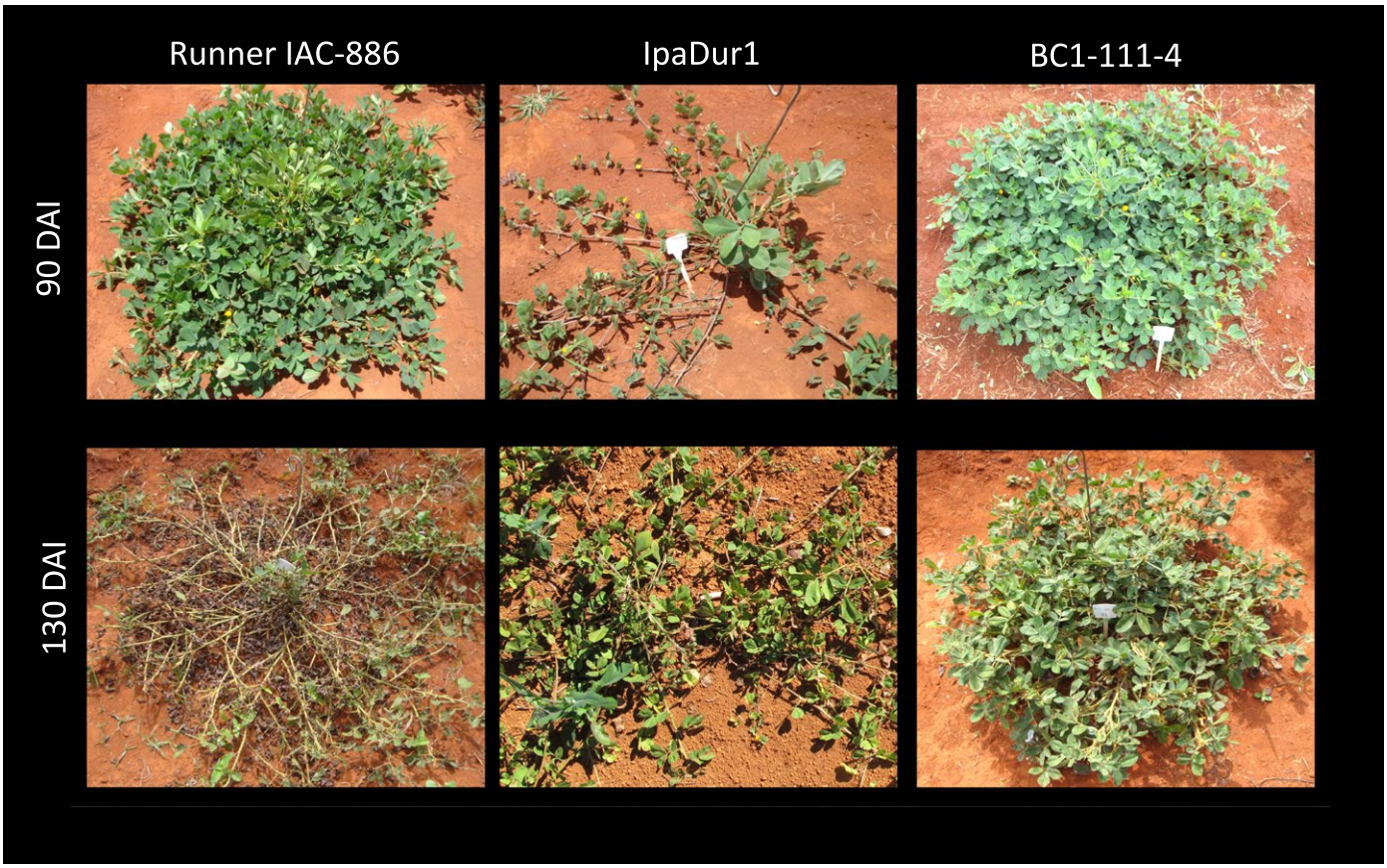

FIGURE 2. Field grown examples of *Arachis* plants at 90 and 130 d after planting: *A. hypogaea* subsp. *hypogaea* cv. Runner IAC-886, IpaDur1 and the backcrossed line BC1-111-4. Note at the end of season, *A. hypogaea* was highly defoliated. The backcrossed line, however, has improved resistance to late leaf spot but retains the growth habit of peanut.

alleles (fortunately, for our data, it was possible to distinguish allelic dosage). Furthermore, genetic studies of peanut have generally assumed that genetic behavior was of a classic allotetraploid (A and B genomes not recombining), to further complicate analyses, recent evidence indicates that there is some genetic exchange between subgenomes of *Arachis* allotetraploids (Bertioli et al., 2016a; Leal-Bertioli et al., 2015a; Nguepjop et al., 2016). These genetic exchanges can change genome composition from the expected AABB to conformations that could be described as AAAA or BBBB. Therefore, we made inferences about introgressions based on the integration of different types of evidence. For instance, the inference that an *A. duranensis* chromosome segment is introgressed into the A-subgenome of *A. hypogaea* was made where the presence of the *A. duranensis* segment was detected together with the absence of the homologous segment from the A subgenome of *A. hypogaea.*

Single nucleotide polymorphism markers (SNPs)—Using the Axiom_Arachis Affymetrix array v01 to assay 58,233 SNPs: 1738 *A. duranensis*-specific markers, 518 *A. ipaënsis*-specific markers, 2575 Runner-886-specific markers, and 2676 markers that distinguish A and B genomes, were identified.

In the publications describing the development of the array (Clevenger et al., 2017; Pandey et al., 2017), SNP markers were given an identification number (ID) with a name beginning "AX-" and also a position relative to the sequenced reference diploid chromosomes. The cited chromosome positions are useful, but, since almost all SNP assays bind to both A and B genome, they don't always indicate the position of the polymorphism, because, in some cases, it actually resides on the homeologous genome. In this case, of 1738 *A. duranensis* markers, 1396 were originally assigned a position on *A. duranensis* chromosomes, whereas 342 were originally assigned a position on *A. ipaënsis* chromosomes; of 518 *A. ipaënsis*-specific markers, 334 were originally assigned a position on *A. ipaënsis* chromosomes, whereas 184 were originally assigned a position on *A. duranensis* chromosomes. For our analyses, markers that were "wrongly" assigned were reassigned positions on the correct genome using the highest sequence similarity determined using the software BLAST (basic local alignment search tool; Altschul et al., 1990). It should be noted that although "wrongly" assigned positions could be corrected for the *A. duranensis* and *A. ipaënsis* characteristic markers, they cannot be identified for *A. hypogaea* markers. For all the *A. hypogaea* markers, the original positions assigned by Clevenger et al. (2017) were used. (Graphic visualizations of genotypes are in Appendix S5.)

Visualizing genotyping calls of *A. duranensis* characteristic markers clearly shows introgressed segments of the wild chromosomes in the selected peanut lines (Appendices S4 and S5). Introgressions are mainly evident as blocks of mostly contiguous duplex calls. Also, some regions are characterized by wild alleles that are not contiguous but scattered; e.g., a region in BC1 111-10-121 covering Aradu.A04 3.7-117 Mbp. The estimated extent of A genome of the selected lines replaced by *A. duranensis* genome varies from 47.8%, in line BC1 111-4-392, to 3.6% in line BC1

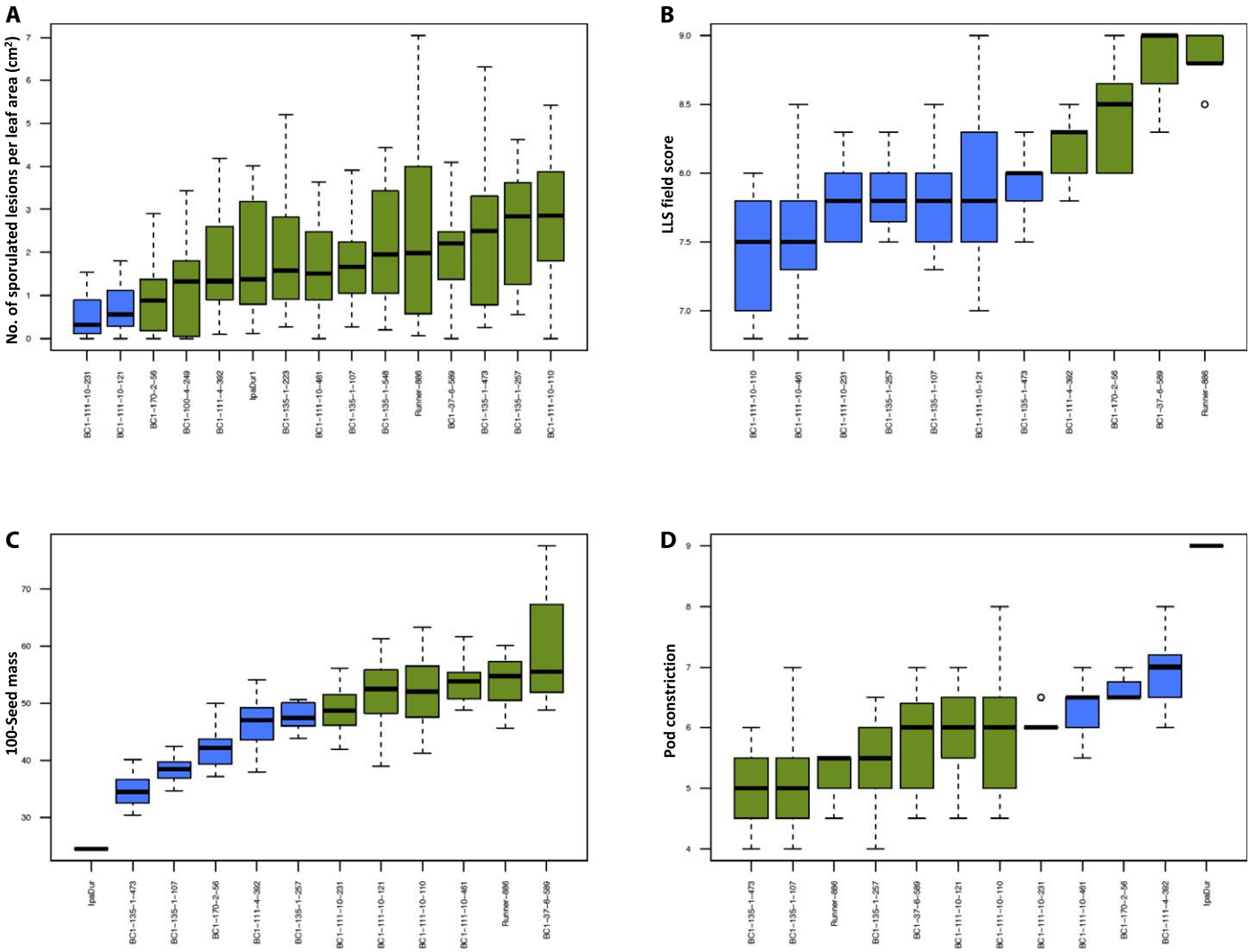

FIGURE 3. Box plot diagrams for variables related to resistance to rust: (A) number of sporulating lesions per leaf area (cm^2), (B) resistance to late leaf spot (visual score on field evaluation), (C) 100-seed mass (g), and (D) pod constriction of selected BC_1F_5 lines derived from the cross between Runner-886 and IpaDur1 (mm). Boxes contain 50% of the data points. Bars across boxes represent the median. The top and bottom ends of the whiskers represent the highest and lowest values observed. Green boxes do not differ significantly from Runner IAC-886 ($P = 0.05$).

37-6-589. Calls indicating *A. duranensis* alleles in the tetraplex dosage are scattered through most of the introgressed segments and also occasionally occur as contiguous regions. Notably in IpaDur1, tetraplex states are indicated for most of chromosomes 04 alleles, consistent with the recombination between A04 and B04 chromosomes that was discovered by independent methods and reported by Leal-Bertioli et al. (2015a). Visualizing the *A. hypogaea* characteristic markers, mostly, but not always, shows that markers assigned to regions of the A subgenome homologous to the *A. duranensis* were absent indicating introgression of the *A. duranensis* segments into the A subgenome of *A. hypogaea*. On average, we estimate that about 98% of the *A. duranensis* chromosome segments were introgressed into the A subgenome (termed *cis* introgression). However, about 2% were introgressed into the B subgenome (*trans* introgression). Genotyping information and summaries of inferred genome structures are provided in Appendices S4 and S5.

Visualizing genotyping calls of *A. ipaënsis* characteristic markers shows similar general patterns, although at much lower resolution and visually more "noisy" (Appendices S4 and S5). The estimated amount of B genome of the selected lines replaced by *A. ipaënsis* genome varies from about 15.3%, in line BC1 111-4-392, to 2.4% in line BC1 135-1-107. On average about 99.65% of the *A. ipaënsis* chromosome segments were introgressed into the B genome (*cis* introgression) and about 0.35% were introgressed into the A genome (*trans* introgression).

Notably, genotyping calls for *A. hypogaea* specific markers show blocks of absence of alleles (Appendix S4). These regions closely correspond to the introgressed segments from the wild species. Most *A. hypogaea* alleles are in duplex with interspersed tetraplex alleles, although, some notable contiguous regions of tetraplex alleles are also apparent. The selected lines show no obvious new regions of tetraplex *A. hypogaea* alleles as compared to Runner-886. However, some regions tetraplex in Runner-886 have been returned to the ancestral duplex state in the selected peanut lines.

Most SNP assays that differentiate the A and B genomes indicate a balanced AABB genome composition. However, significant

TABLE 1. Average and standard deviation of selected BC_1F_4 and BC_1F_5 families for rust and late leaf spot resistance and seed characteristics.

Trait	Rust resistance (BC_1F_4)			LLS resistance (BC_1F_5)		Seed (BC_1F_5)			
Genotype	Susceptibility index	TL/LA	SL/LA	Lab (score 1-5)	Field (score 1-9)	100-Seed mass (g)	Total seed mass (g)	P2 (%)	Pod constriction (1-10)	Seed number
IpaDur1	4.8 ± 3.5 abcd	2.4 ± 1.8 abcd	1.8 ± 1.4 abcde	1.8 ± 0.8 d	2.4 ± 0.1 h	24.8 ± 0.2f	NE	NE	NE	NE
BC1-37-6-589	4.7 ± 2.3 abcd	2.3 ± 1.1 abcd	2.0 ± 1.0 abcde	3.8 ± 0.7 bcd	8.8 ± 0.2 a	58.9 ± 9.3a	79.3 ± 44.5c	83.7 ± 7.8abc	5.83 ± 0.76cde	128.9 ± 67.1e
BC1-111-4-392	4.0 ± 2.8 abcd	1.9 ± 1.1 bcd	1.7 ± 1.2 abcde	3.7 ± 0.5 bcd	8.2 ± 0.3 bcd	46.6 ± 4.4de	90.7 ± 21.1c	74.2 ± 10.2def	6.93 ± 0.67ab	204.4 ± 29.3bcde
BC1-111-10-110	6.3 ± 3.1 ab	3.0 ± 1.4 ab	2.7 ± 1.4 ab	4.1 ± 0.9 abc	7.4 ± 0.4 gh	52.5 ± 5.9ab	81.4 ± 26.8c	78.2 ± 8.8bcde	6.00 ± 1.00cde	167.8 ± 29.3de
BC1-111-10-121	1.6 ± 1.2 cd	0.9 ± 0.6 bcd	0.7 ± 0.5 de	4.8 ± 0.4 a	7.9 ± 0.5 def	51.8 ± 5.0abc	88.5 ± 28.7c	81.9 ± 4.2abcd	5.96 ± 0.59cd	168.8 ± 47.8de
BC1-111-10-231	1.0 ± 1.1 cd	0.3 ± 0.3 d	0.5 ± 0.5 e	4.0 ± 0.3 bcd	7.8 ± 0.2 efg	48.9 ± 3.9bcd	89.7 ± 21.8c	78.6 ± 6.3bcde	6.11 ± 0.21bcd	180.8 ± 29.4de
BC1-111-10-461	3.6 ± 2.5 bcd	1.8 ± 1.2 bcd	1.6 ± 1.1 bcde	3.6 ± 0.4 bcd	7.5 ± 0.4 fgh	53.5 ± 3.5ab	90.6 ± 32.1c	77.7 ± 6.9cde	6.40 ± 0.37abc	166.0 ± 59.1de
BC1-135-1-107	4.0 ± 2.6 abcd	2.0 ± 1.2 bcd	1.8 ± 1.1 abcde	4.3 ± 0.5 ab	7.9 ± 0.3 def	38.3 ± 1.9f	103.6 ± 26.9bc	67.7 ± 7.7fg	5.07 ± 0.74f	279.1 ± 69.5abc
BC1-135-1-257	5.7 ± 2.9 ab	2.7 ± 1.3 abc	2.6 ± 1.3 abc	3.7 ± 0.3 bcd	7.8 ± 0.2 defg	47.6 ± 2.1cde	163.8 ± 71.1ab	73.2 ± 9.2efg	5.47 ± 0.81def	318.9 ± 119.0ab
BC1-135-1-473	5.2 ± 3.8 abc	2.4 ± 1.7 abcd	2.4 ± 1.7 abcd	4.3 ± 0.5 ab	7.9 ± 0.3 cde	34.8 ± 3.2f	99.7 ± 27.0bc	80.4 ± 4.4bcde	4.94 ± 0.57f	288.9 ± 54.5abc
BC1-170-2-56	2.4 ± 1.9 bcd	1.6 ± 1.3 bcd	0.9 ± 0.8 cde	3.9 ± 0.7 de	8.4 ± 0.3 abc	42.0 ± 3.6ef	90.9 ± 35.5bc	64.3 ± 6.4fg	6.63 ± 0.23ab	205.6 ± 82.8cde
Runner-886	6.2 ± 5.5 ab	3.1 ± 2.7 ab	2.5 ± 2.2 abc	5.0 ± 0.1 a	8.8 ± 0.2 ab	53.9 ± 4.4ab	147.9 ± 45.9ab	88.2 ± 0.6a	5.23 ± 0.33ef	286.2 ± 76.4abc
Test results	$F(11,198)=4.72$; $P < 0.000$	$F(11,193)=4.46$; $P < 0.000$	$X^2= 51.11$, df=11, $P= 3.9e^{-07}$	$X^2=81.3$, df=11, $P < 2.4e^{-12}$	$X^2=114.8$, df=11, $P< 2.2e^{-16}$	$F(11,89)= 5.092$; $P < 0.000$	$X^2=107.5$, df=11, $P< 2.2e^{-16}$	$X^2=119.4$, df=11, $P< 2.2e^{-16}$	$X^2=109.1$, df=11, $P < 2.2e^{-16}$	

Notes: TL/LA = total number of lesions/leaf area, SL/LA = number of sporulated lesions/ leaf area, LLS Lab = LLS bioassay using detached leaves; LLS field: score of total plants in the field; 100-Seed mass = mass of 100 grams of seeds; P2 (%) = percentage of pods with two seeds. NE = not evaluated. Cells within each column with the same letter do not differ significantly ($P < 0.05$).

numbers of calls indicate AAAA and BBBB. Of these, conspicuous contiguous blocks frequently confirmed blocks of tetraplex alleles indicated by the species characteristic markers.

Detecting recombination between A and B subgenomes using low-coverage sequencing—We used the low-coverage whole-genome sequences together with the reference genome sequences of *A. duranensis* and *A. ipaënsis* (Bertioli et al., 2016a) to more extensively investigate genome compositions and to complement the SNP analysis. Although the low coverage used was less than would be needed for reliable inferences at base-pair resolution, it was completely adequate for views of genome composition at the ~10,000-bp scale. The methodology takes advantage of the diploid genome sequences being from the same genotypes used to make the IpaDur allotetraploids and their having high similarities to the corresponding subgenomes of *A. hypogaea* (Bertioli et al., 2016a). Random whole-genome DNA sequences were mapped onto the combined diploid genome sequences. Relative mapping depths were normalized and plotted. For better visualization within a unified framework, both A and B homoelogous genes were plotted relative to the chromosome sequences of *A. ipaënsis*. Where genome composition is balanced (the expected AABB), we expect mapping densities to be similar on the A and B chromosomes.

Over most of the chromosomal plots, the expected normalized mapping densities were observed; approximately equal onto A and B genomes (Fig. 5, Appendix S6). However, significant proportions of the genome deviated from equal densities, especially at chromosome ends. Mapping onto one genome decreases to almost zero, and the other doubles. These deviations indicate changes in genome compositions from AABB to what could more accurately be described as AAAA or BBBB and may be derived from meiotic crossovers and/or gene conversion between A and B genomes. Some regions of the genome that were tetraplex AAAA or BBBB in Runner-886 were balanced in one or more of the selected lines; in other words, regions of the A or B subgenome that were absent in Runner-886 have been replaced with their wild homologs (e.g., top of chromosomes 05; Fig. 5; Appendices S4–S6). The resequencing data align closely to the SNPs from the Axiom_Arachis array. In addition to these large deviations in mapping density, we also observed more subtle deviations. For instance, in allotetraploid IpaDur1, in the lower approximately 25% of chromosomes 06, mapping steadily increases on in the A genome, and decreases on the B genome. On chromosomes 05, mapping densities onto A and B genomes form slopes and a "cross-over" (Fig. 5). These subtler deviations in mapping density may indicate regions of strand exchange and gene conversion between homeologous chromosomes.

DISCUSSION

We aimed to study the mechanisms of increasing genetic diversity and introduction of new desirable characters through the use of crop wild relatives in a segmental allotetraploid crop. We used a cross between a popular Brazilian peanut cultivar, Runner-886 and an induced wild-derived allotetraploid, IpaDur1 ([*A. ipaënsis* × *A. duranensis*]^{4x}). For the first stage of selection, we planted progeny from first backcross plants in the field. On average, it is expected that first backcross plants have 25% of donor alleles, we had anticipated

this would result in agronomically unacceptable plants and planned to select for further backcrossing. However, we were surprised by the cultivated-like phenotypes of a proportion of the BC_1F_2 plants and were curious to discover how much breeding progress could be made by simple selection of this material. At each generation, we were encouraged by the results. By selection, it was possible to maintain superior disease resistance whilst successively improving canopy architecture, seed size, pod phenotype and yield.

By the $BC_1F_{4/5}$ generations, all genotypes had domesticated features: compact canopy architecture, large biomass, large seeds, pods with small constriction, and a large proportion of two-seeded pods. Their general appearance is of cultivated peanuts of pure pedigree. The only feature that distinguished some of the lines was yellow flowers, in contrast to the orange flowers borne by cultivated peanuts of pure pedigree. Under very heavy disease pressure from late leaf spot in the hot humid growing season in Pindorama, São Paulo, eight of 10 evaluated lines had significantly improved resistance. Although not actively selected for, two also showed rust resistance.

The induced allotetraploid used here as the donor parent is derived from the cultivated peanut's diploid progenitor species. It was first created with the objective of investigating the origin of cultivated peanut (Favero et al., 2006). Nevertheless, it has proved a useful source of improved traits and new alleles. In Brazil, using the same material, runner-type breeding lines with good agronomic traits, similar yield, but larger seed than the recurrent parent were obtained (Suassuna et al., 2015). In Senegal, the same induced allotetraploid was used to create well-characterized chromosome segment substitution lines within the background of the locally preferred Spanish type variety *A. hypogaea* subsp. *fastigiata*, cultivar Fleur11 (Fonceka et al., 2009, 2012). From these lines, three varieties have been released, one with increased seed size and two with greater haulm production than the recurrent parent Fleur11 (Faye et al., 2016).

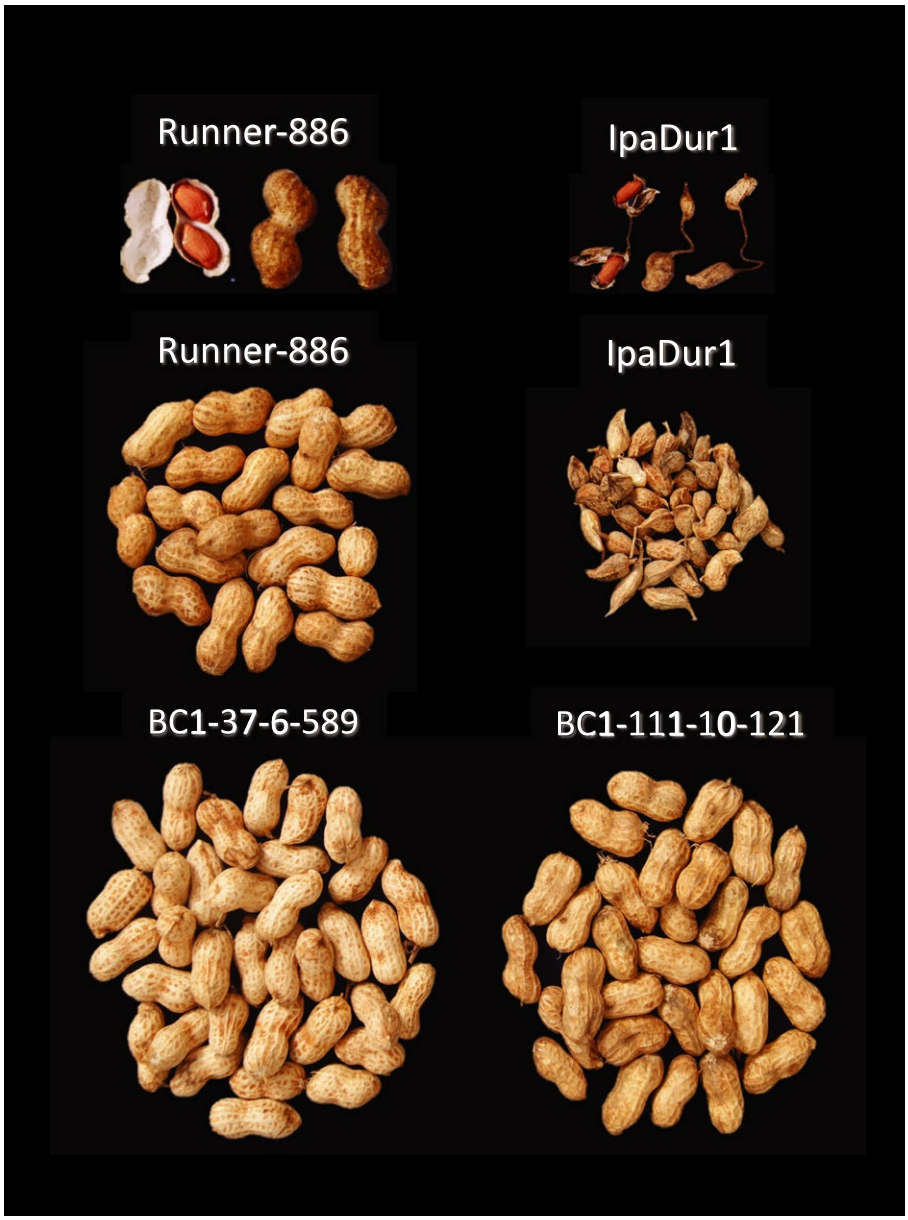

FIGURE 4. Examples of pod and seed shapes of *A. hypogaea* subsp. *hypogaea* cv. Runner IAC-886, IpaDur1, and the backcrossed lines BC1-37-6-589 and BC1-111-10-121. Note contrasting seed size, pod size and pod constriction of the cultivated Runner-886 and IpaDur1. Seeds of wild *Arachis* are separated by a long isthmus, and this character remains on the induced allotetraploid IpaDur1. Through one round of backcrossing and selection, this trait was eliminated from the selected backcrossed lines.

In the work described here, the single backcross and selection gave the potential for multiple recombinations between the cultivated and wild genomes. The results were surprisingly successful, and repeatedly so; progeny of 13% (5/38) of the initial backcrossed plants were represented in the 13 lines finally selected. Using the improved genetic tools and knowledge base that have become available since the sequencing of cultivated peanut's diploid ancestors (*A. duranensis* and *A. ipaënsis*; Bertioli et al., 2016a; Clevenger et al., 2017; Pandey et al., 2017), we were able to investigate the genetic composition in unprecedented detail.

Wild-derived introgressions were identified in all backcrossed lines, covering between an estimated 7.2% (line BC1 135-1-473) and 30.3% (line BC1 111-4-392), with an average of 16%. It was surprising that lines could harbor such high proportions of wild species genomes and have phenotypes essentially indistinguishable from cultivated peanuts of pure pedigree. This result shows the feasibility of greatly increasing the genetic diversity of cultivated peanut and may be an indication that the cultivated phenotype is controlled by relatively few genomic regions. The introgressions varied widely in size, from almost whole chromosomes to much smaller ones. Many chromosomes harbored multiple introgressions. Reflecting

TABLE 2. Parental means, and mean, minimum, maximum and median values of the BC_1 families for growth habit, seed characteristics, and resistance to late leaf spot (LLS).

Trait	Year - Generation	Runner-IAC-886	IpaDur1	Min	Max	Aver	Median
Growth habit							
	2009 - BC_1F_2	10	1	2	10	5.64	5.50
	2010 - BC_1F_3	10	1	3	10	7.48	8.00
	2011 - BC_1F_4	10	1	8.5	9	8.88	8.50
Seed number							
	2009 - BC_1F_2	185.8	10.4	6	490	125.6	107
	2011 - BC_1F_3	127.4	NE	37	428	137.5	118
	2013 - BC_1F_5	286.2	NE	27	475	209.0	200
100-Seed mass (g)							
	2009 - BC_1F_2	52.87	14.30	14.1	71.8	40.95	41.0
	2011 - BC_1F_4	51.41	14.19	21.8	72.3	45.59	46.4
	2013 - BC_1F_5	53.87	NE	30.5	77.5	51.63	51.6
Total seed mass (g)							
	2009 - BC_1F_2	100.38	1.63	0.28	191.7	52.47	46.1
	2011 - BC_1F_4	66.56	NE	8.50	162.8	62.21	57.3
	2013 - BC_1F_5	147.90	NE	17.09	224.7	96.23	92.1
P2 (%)							
	2009 - BC_1F_2	NE	NE	NE	NE	NE	NE
	2011 - BC_1F_4	56.2	0.0	0.0	90.8	52.3	56
	2013 - BC_1F_5	88.2	NE	51.3	100.0	76.0	78
LLS visual score (related to Runner-IAC-886)							
	2009 - BC_1F_2	1.00	0.58	0.3	1.7	1.04	0.33
	2010 - BC_1F_3	1.00	0.39	0.2	1.2	0.79	0.27
	2011 - BC_1F_4	1.00	NE	0.2	1.0	0.55	0.19
	2013 - BC_1F_5	1.00	0.27	0.7	1.0	0.90	0.05

Note: NE = not evaluated/data not available

the patterns of genetic recombination, introgressions in the arms tended to be smaller than those in pericentromeres.

Introgressions seemed to be of two types. Most commonly, wild alleles were detected as mostly continuous blocks, likely formed by meiotic crossovers between the cultivated and wild genomes. Generally, in the regions of the cultivated peanut's genome homologous to the wild alleles, cultivated alleles are not found, indicating replacement by wild alleles from the corresponding diploid progenitor (wild A alleles replaced cultivated A alleles, or wild B alleles replaced cultivated B alleles). However, in a significant number of cases, A alleles replaced B alleles or vice versa. Overall, we estimated that 98% of the introgressed *A. duranensis* genome was incorporated into cultivated peanut's A subgenome, whereas 2% was introgressed on the B subgenome. We named these types of introgressions *cis* and *trans*, respectively. *Trans* introgressions of *A. ipaënsis* segments accounted for about 0.6% of the *A. ipaënsis* alleles. In addition to the continuous blocks of wild alleles, regions of interspersed wild alleles were detected. These regions seem robustly supported and do not seem to be the result of genotyping errors. For instance, in one region (A04: 3.7-117Mb, line BC1 111-10-121), 42% (68/162) of markers indicate the presence of *A. duranensis* alleles. This percentage of wild alleles is much lower than detected in the continuous introgression blocks, which is typically greater than 95%, but at the same time, it seems much too high a proportion of wild alleles to be accounted for by genotyping error. In the same region, *A. hypogaea* A alleles are seen to mostly disappear, but also some *A. ipaënsis* alleles are detected. These regions may represent the action of gene conversion, between *A. duranensis* and *A. ipaënsis* and/or gene conversion between the wild and cultivated genomes. Genetic exchange between *A. duranensis* and *A. ipaënsis* in the induced allotetraploid has been detected previously (Leal-Bertioli et al., 2015a) and was detected in this study. Intriguingly, these results show that introgression from induced allotetraploids will, at some frequency, involve wild-derived genome segments that are not representative of the pure diploid species but contain mixtures of the A and B alleles.

Trans introgression, where wild A genome is introgressed into cultivated B genome or vice versa reflects the occasional recombination between homeologous genomes that has been documented in other studies of peanut (Bertioli et al., 2016a,b; Leal-Bertioli et al., 2015a; Nguepjop et al., 2016). The genetic recombination that occurred appears to fit with a pattern of inheritance intermediate between the disomic, as in diploids, and polysomic as in autopolyploids, consistent with "segmental allopolyploidy" as defined by Stebbins (1947). Other plants that fit this intermediate genetic pattern of inheritance include chrysanthemum (Klie et al., 2014), banana (Jeridi et al., 2012), strawberry (Tennessen et al., 2014), rose (Bourke et al., 2017), and *Tragopogon* (Chester et al., 2012).

Cultivated peanut has itself, in certain genome regions, undergone recombination that has changed the genome formula from the expected AABB to AAAA or BBBB. Once A or B alleles have been lost within a population, they cannot be replaced, a process of genetic degeneration that has been termed the "polyploid ratchet" by Gaeta and Pires (2010). Intriguingly in the selected lines, several genome regions have been restored to the AABB state by the introgression of wild alleles. For instance, a 6-Mb region at the top of A05 that has a genome formula of AAAA in the recurrent cultivated peanut parent, has a genome formula of AABB in several of the selected lines due to B genome introgression. We consider that these *trans* introgressions represent a potential route to effectively reverse the polyploid ratchet and increase heterosis, but they may also be neutral or maladaptive. Also, as a consequence of *trans*

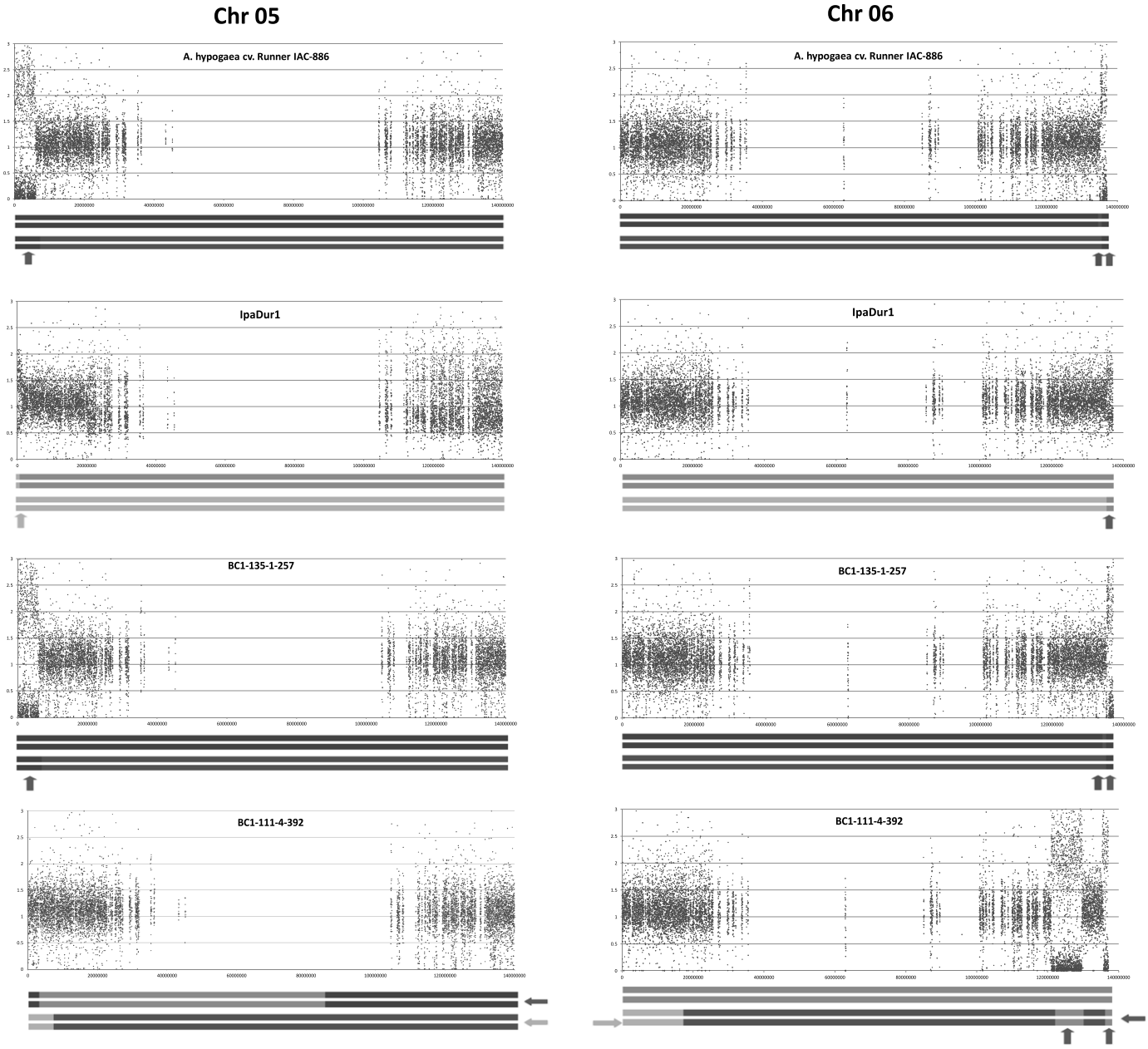

FIGURE 5. Representation of structures of chromosomes 05 and 06 of *Arachis hypogaea* cv. Runner IAC-886, the induced allotetraploid IpaDur1 and two selected lines. The scatterplots are used to infer overall tetraploid genome structure. They show mapping densities of randomly generated Illumina whole-genome sequences from the genotypes onto the chromosome sequences of *A. duranensis* (green dots) and *A. ipaënsis* (red dots), normalized to an expected value of 1 (*y*-axis). Lines below plots represent the chromosomes. Chromosome structures and introgressions were deduced from both the mapping densities and genotyping results from Axiom_Arachis Affymetrix array v01. Dark green and red, A and B subgenomes from cultivated peanut respectively; light green and orange, *A. duranensis* and *A. ipaënsis*, respectively. Horizontal arrows indicate wild introgressions and vertical arrows indicate tetrasomic regions of genome structure. On scatterplots, mapping densities of red and green dots that cluster around 1 indicate the expected genome composition of AABB. Regions where mapping densities deviate represent deviations from the expected genome formula. For instance, at the tops of chromosomes 05 in *A. hypogaea*, genome structure can be described as AAAA. In line BC1-111-4-392, the top of chromosomes 05 has been restored to genome structure of AABB by *A. ipaënsis* introgression (orange segment). At the bottom of chromosomes 06 of BC1-111-4-392, two regions of the genome have structures AAAA that are not present in either of the parents, caused by *A. duranensis* introgression (light green segments). Deviations in mapping densities in IpaDur1 are more subtle and difficult to interpret in terms of genome structure; they may represent the result of extensive gene conversions between the A and B genomes. Therefore, the representation of the genome structure of IpaDur1 is approximate.

introgression, some genome regions in the selected lines have become tetrasomic for the A or B genomes (becoming AAA′A′ or BBB′B′, wild genome being indicated by ′), where, in all cultivated peanuts we have surveyed to date they have the genome formula of AABB (our unpublished results). Again, whether these changes are desirable, neutral, or undesirable, remains to be discovered. Notably, all new tetrasomic regions in the lines involve wild genome segments, no new tetrasomic regions of pure *A. hypogaea* genome were found. Therefore, the use of wild species provides both the possibility to replace subgenome segments and alleles that have been lost in cultivated peanut, and appears to increase genome instability or the tolerance to new genome arrangements. Another view is that peanut breeding using wild species and the tetraploid introgression route effectively adds a new polyploid origin for peanut with each new induced allotetraploid that is used. Creating "new origins" adds possibilities of increasing genetic diversity by what has been termed "lineage recombination" by Doyle et al. (2002), who described allelic recombination within homeologous subgenomes of an allopolyploid but did not discuss recombination between homeologues, which occurs at a low but significant level even in strict allopolyploids (e.g., Flagel et al., 2012). Similar lineage recombination should occur across the partially homeologous chromosome sets of segmental allopolyploids, and analogously, random associations among multiple chromosomes in autopolyploids, though leading to segregational loss of variation, will also create novel genotypes. Given conventional wisdom that autopolyploids, at least, are grossly undercounted (Soltis et al., 2007), lineage recombination is likely to contribute to the genetic diversity of polyploids in nature.

The use of crop wild relatives to improve crop performance is well established with important examples dating back more than 70 years (Hajjar and Hodgkin, 2007). Historically, for peanut, the ploidy differences and incompatibilities between cultivated wild species have hampered the use of wild species in breeding programs (Leal-Bertioli et al., 2017). Improved understanding of the species relationships, together with much improved tools for genetic studies have helped overcome many of the difficulties. The results presented here, together with other studies, reinforce that peanut wild relatives can contribute positive alleles for agronomical traits and much needed diversity that could provide the genetic basis for an expanded range of adaptive traits, including resistance to pests and disease, tolerance to abiotic stresses, and reduced dependence on inputs (Warschefsky et al., 2014). This improved knowledge base and tools for peanut genetics sets the framework for the facilitated use of wild species in breeding that could benefit farmers and families throughout the tropics. The benefits can have wider reach with better sharing of germplasm resources.

In summary, this study produced peanut lines that have improved disease resistance by the incorporation of wild alleles. Genetic analysis showed that a surprisingly high proportion of wild alleles can be introgressed while, at the same time, maintaining the characteristic phenotypes of cultivated peanut and good yields. Also, we showed that, unexpectedly, the breeding scheme introduced genetic diversity not just by introducing new alleles, but also by introducing new allotetraploid genome arrangements. These results highlight the special possibilities of segmental allotetraploidy and of using lineage recombination to increase genetic diversity in peanut, likely mirroring what occurs in natural segmental allopolyploids of multiple origins.

ACKNOWLEDGEMENTS

The authors thank Alessandra Fávero for initial allotetraploid material, Marilda Oliveira for making initial crosses, Leandro Mesquita for greenhouse assistance, Austerclinio Lopes for providing field space at Embrapa Cerrados and The Brazilian National Council for Scientific and Technological Development (CNPq) for MCM Research Productivity Fellowship. This work was funded by MARS Inc., The Peanut Foundation and host institutions. Special thanks are given to the two reviewers for their very useful and thought out comments.

AUTHOR CONTRIBUTIONS

SCMLB, DJB and MCM: experimental design; IG, JFS, MCM, PMG, SCMLB, DJB: field and in vitro phenotyping, BLA and DJB: sequence analyses; SCMLB, DJB, MCM: genotyping analysis; SCMLB, DJB, JJD, SAJ and MCC: wrote the manuscript. All authors read the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Bertioli, D. J., S. B. Cannon, L. Froenicke, G. Huang, A. D. Farmer, E. K. Cannon, X. Liu, et al. 2016a. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nature Genetics* 48: 438–446.

Bertioli, D. J., S. C. M. Leal-Bertioli, and H. T. Stalker. 2016b. The Peanut Genome: the history of the Consortium and the structure of the genome of cultivated peanut and Its diploid ancestors. *In* H. T. Stalker and R. F. Wilson [eds.], Peanuts: genetics, processing, and utilization, 147–161. Elsevier, Atlanta, GA, USA.

Bertioli, D. J., G. Seijo, F. O. Freitas, J. F. M. Valls, S. C. M. Leal-Bertioli, and M. C. Moretzsohn. 2011. An overview of peanut and its wild relatives. *Plant Genetic Resources* 9: 134–149.

Bourke, P. M., P. Arens, R. E. Voorrips, G. D. Esselink, C. F. Koning-Boucoiran, W. P. Van't Westende, T. Santos Leonardo, et al. 2017. Partial preferential chromosome pairing is genotype dependent in tetraploid rose. *Plant Journal* 90: 330–343.

Chester, M., J. P. Gallagher, V.V. Symonds, da A. V. C. Silva, E. V. Mavrodiev, A. R. Leitch, P. S. Soltis, and D. E. Soltis, 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proceedings of the National Academy of Sciences, USA* 109: 1176–1181.

Chu, Y., C. Wu, C. C. Holbrook, B. Tillman, G. Person, and P. Ozias-Akins. 2011. Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. *Plant Genome* 4: 110–117.

Clevenger, J., Y. Chu, C. Chavarro, G. Agarwal, D. J. Bertioli, S. C. M. Leal-Bertioli, M. K. Pandey, et al. 2017. Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. *Molecular Plant* 10: 309–322.

Doyle, J. J., J. L. Doyle, and A. H. D. Brown. 1999. Origins, colonization, and lineage recombination in a widespread perennial soybean polyploid complex. *Proceedings of the National Academy of Sciences, USA* 96: 10741–10745.

Doyle, J. J., J. L. Doyle, A. H. Brown, and R. G. Palmer. 2002. Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: histone H3-D gene sequences. *Evolution* 56: 1388–402.

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

Faye, I., T. Hodo-Abalo, O. Ndoye, and D. Fonceka. 2016. Dossier technique d'homologation de nouvelles variétés d'arachide. *In* Institut Sénegalais de Rescherches Agricoles (ISRA) [ed.], Senegal, 34. ISRA, Dakar, Senegal.

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

Flagel, L. E., J. F. Wendel, and J. A. Udall. 2012. Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. *BMC Genomics* 13: 302.

Fonceka, D., T. Hodo-Abalo, R. Rivallan, I. Faye, M. N. Sall, O. Ndoye, A. P. Favero, et al. 2009. Genetic mapping of wild introgressions into cultivated peanut: a way toward enlarging the genetic basis of a recent allotetraploid. *BMC Plant Biology* 9: 103.

Fonceka, D., T.-A. Tossim, R. Rivallan, H. Vignes, I. Faye, O. Ndoye, M. C. Moretzsohn, 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.

Gaeta, R. T., and J. C. Pires. 2010. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytologist* 186: 18–28.

Haas, B. J., A. L. Delcher, J. R. Wortman, and S. L. Salzberg. 2004. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20: 3643–3646.

Hajjar, R., and T. Hodgkin. 2007. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156: 1–13.

Jeridi, M., X. Perrier, M. Rodier-Goud, A. Ferchichi, A. D'Hont, and F. Bakry. 2012. Cytogenetic evidence of mixed disomic and polysomic inheritance in an allotetraploid (AABB) *Musa* genotype. *Annals of Botany* 110: 1593–1606.

Klie, M., S. Schie, M. Linde, and T. Debener. 2014. The type of ploidy of chrysanthemum is not black or white: a comparison of a molecular approach to published cytological methods. *Frontiers in Plant Science* 5: 479.

Kolekar, R. M., V. Sujay, K. Shirasawa, Y. P. Khedikar, M. V. C. Gowda, M. K. Pandey, R. K. Varshney, and R. S. Bhat. 2016. QTL mapping for late leaf spot and rust resistance using an improved genetic map and extensive phenotypic data on a recombinant inbred line population in peanut (*Arachis hypogaea* L.). *Euphytica* 209: 147–156.

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

Kumari, V., M. V. C. Gowda, V. Tasiwal, M. K. Pandey, R. S. Bhat, N. Mallikarjuna, H. D. Upadhyaya, and R. K. Varshney. 2014. Diversification of primary gene pool through introgression of resistance to foliar diseases from synthetic amphidiploids to cultivated groundnut (*Arachis hypogaea* L.). *Crop Journal* 2: 110–119.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357.

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

Leal-Bertioli, S. C. M., U. Cavalcante, E. G. Gouveia, C. Ballen-Taborda, K. Shirasawa, P. M. Guimaraes, S. A. Jackson, et al. 2015b. Identification of QTLs for rust resistance in the peanut wild species *Arachis magna* and the development of KASP markers for marker assisted selection. *Genes/Genomes/Genetics* 5: 1403–1413.

Leal-Bertioli, S. C. M., M. C. Moretzsohn, S. P. Santos, A. C. M. Brasileiro, P. M. Guimarães, D. J. Bertioli, and A. C. G. Araujo. 2017. Phenotypic effects of allotetraploidization of wild *Arachis* and their implications for peanut domestication. *American Journal of Botany* 104: 379–388.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

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

Nguepjop, J. R., H.-A. Tossim, J. M. Bell, J.-F. Rami, S. Sharma, B. Courtois, N. Mallikarjuna, et al. 2016. Evidence of genomic exchanges between homeologous chromosomes in a cross of peanut with newly synthetized allotetraploid hybrids. *Frontiers in Plant Science* 7: 1635.

Pandey, M. K., G. Agarwal, S. M. Kale, J. Clevenger, S. N. Nayak, M. Sriswathi, A. Chitikineni, et al. 2017. Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Scientific Reports* 7: 40577.

R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramsey, J., and D. W. Schemske. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 33: 589–639.

Renny-Byfield, S., and J. F. Wendel. 2014. Doubling down on genomes: polyploidy and crop plants. *American Journal of Botany* 101: 1711–1725.

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 Applied Genetics* 121: 1033–1046.

Servick, S., C. J. Visger, M. A. Gitzendanner, P. S. Soltis, and D. E. Soltis. 2015. Population genetic variation, geographic structure, and multiple origins of autopolyploidy in *Galax urceolata*. *American Journal of Botany* 102: 973–982.

Simpson, C. E. 1991. Pathways for introgression of pest resistance into *Arachis hypogaea* L. *Peanut Science* 18: 22–26.

Simpson, C. E., A. Krapovickas, and J. F. M. Valls. 2001. History of *Arachis* including evidence of *A. hypogaea* L. progenitors. *Peanut Science* 28: 78–80.

Soltis, D. E., E. V. Mavrodiev, S. C. Meyers, P. M. Severns, L. Zhang, M. A. Gitzendanner, T. Ayers, et al. 2012. Additional origins of Ownbey's *Tragopogon mirus*. *Botanical Journal of the Linnean Society* 169: 297–311.

Soltis, D. E., and P. S. Soltis. 1993. Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Sciences* 12: 243–273.

Soltis, D. E., and P. S. Soltis. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* 14: 348–352.

Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson, B. C. Husband, and W. S. Judd. 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* 56: 13–30.

Soltis, D. E., P. S. Soltis, and J. A. Tate. 2004. Advances in the study of polyploidy since Plant speciation. *New Phytologist* 161: 173–191.

Stalker, H. T. 2017. Utilizing wild species for peanut improvement. *Crop Science* 57: 1102–1120.

Stalker, H. T., J. C. Wynne, and M. Company. 1979. Variation in progenies of an Arachis hypogaea & #x00D7; diploid wild species hybrid. *Euphytica* 28: 675–684.

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

Stebbins, G. L. 1971. Chromosomal evolution in higher plants. Edward Arnold, London, UK.

Suassuna, T. M. F., N. D. Suassuna, M. C. Moretzsohn, S. C. M. Leal-Bertioli, D. J. Bertioli, and E. P. Medeiros. 2015. Yield, market quality, and leaf spots partial resistance of interspecific peanut progenies. *Crop Breeding and Applied Biotechnology* 15: 175–180.

Tennessen, J. A., R. Govindarajulu, T. L. Ashman, and A. Liston. 2014. Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biology and Evolution* 6: 3295–313.

Tester, M., and P. Langridge. 2010. Breeding technologies to Increase crop production in a changing world. *Science* 327: 818–822.

Vale, F. X. R., E. Fernandes Filho, and J. R. Liberato. 2001. Quant—Quantificação de doenças: versão 1.0.1. Software distributed by the authors, Universidade Federal de Viçosa, Viçosa, Brazil.

Valls, J. F. M., and C. E. Simpson. 2005. New species of *Arachis* L. (Leguminosae) from Brazil, Paraguay and Bolivia. *Bonplandia* 14: 35–63.

Van de Peer, Y., E. Mizrachi, and K. Marchal. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* 18: 411–424.

Warschefsky, E., R. V. Penmetsa, D. R. Cook, and E. J. von Wettberg. 2014. Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *American Journal of Botany* 101: 1791–1800.

Washburn, J. D., and J. A. Birchler. 2014. Polyploids as a "model system" for the study of heterosis. *Plant Reproduction* 27: 1–5.

Welles, S. R., and N. C. Ellstrand. 2016. Genetic structure reveals a history of multiple independent origins followed by admixture in the allopolyploid weed *Salsola ryanii*. *Evolutionary Applications* 9: 871–878.

Wendel, J. F. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany* 102: 1753–1756.