

Double Reduction in Allotetraploid Peanut and the Role of Chromosomal Imbalance in Unexpected Linkage Map Artifacts

Samuele Lamon¹, Peter M. Bourke², Brian L. Abernathy³, João F. dos Santos⁴, Ignácio J. de Godoy⁴, Soraya C.M. Leal-Bertioli^{1,3,5} and David J. Bertioli^{1,3,6}

¹Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA 30602, USA

²Plant Breeding, Wageningen University & Research, Wageningen 6708 PB, The Netherlands

³Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602, USA

⁴Campinas Agronomical Institute, Botafogo, Campinas, SP 13075-630, Brazil

⁵Department of Plant Pathology, University of Georgia, Athens, GA 30602, USA

⁶Department of Crop & Soil Sciences, University of Georgia, Athens, GA 30602, USA

Corresponding author: David J. Bertioli

Running head

Phased Linkage Mapping and Double Reduction in Peanut

Keywords

Arachis hypogaea, polyploidy, double reduction, homoeologous exchange, genetic instability, linkage mapping, segmental allopolyploidy

Abstract

Polyploidization in peanut (*Arachis hypogaea* L.) provided evolutionary advantages by increasing heterosis, the response to selection, and enhancing adaptability. However, it also caused a genetic bottleneck by isolating cultivated peanut from its wild diploid relatives. Mechanisms such as homoeologous exchange can partially restore genetic diversity by generating new allelic combinations. Double reduction is a rare segregation pattern restricted to polyploids, in which a single-dosage locus yields duplex gametes. It requires multivalent formation and crossing over between non-sister chromatids, both of which are associated with homoeologous exchange. Although peanut mainly exhibits disomic pairing, occasional multivalents theoretically allow low-frequency double reduction. To estimate double reduction and examine its relationship with genetic instability, a high-density phased linkage map was constructed using a backcross population from a cross between a neoallotetraploid [*A. magna* K 30097 x *A. stenosperma* V 15076]^{4x} (MagSten) and cultivated peanut. The final map included 9,717 SNP markers with an average spacing of 0.22 centiMorgans. Some progenies showed unbalanced genomic compositions, creating artifacts in linkage analysis. Removing these progenies improved the map and suggested a common origin for artifacts previously observed in other linkage maps, revealing a novel aspect of mapping in allotetraploid peanut. Analysis of the phased map revealed double reduction in 12% of progenies. Notably, one event produced a genomic composition consistent with theoretical predictions, supporting the expectation that double reduction causes unbalanced

genomes in allopolyploids. These results indicate that double reduction is a low but significant frequency genetic phenomenon in the segmental allotetraploid peanut, contributing to the genetic instability and evolutionary dynamics of this and likely other allopolyploid genomes.

Article Summary

This study investigated double reduction, a rare genetic event in segmental allopolyploid peanut, which can create unbalanced genomic compositions and affect genetic diversity. We generated a backcross population using neoallotetraploid and cultivated peanuts, then constructed a high-density phased linkage map. Analysis revealed unbalanced genomic compositions in some progenies caused by homoeologous exchanges, which reduced map quality. Double reduction was estimated to occur in approximately 12% of progenies, aligning with theoretical expectations for genomic imbalance. These results demonstrate that double reduction contributes to genetic instability, inheritance patterns, and genome evolution in allopolyploid organisms such as peanut.

Introduction

Peanut (*Arachis hypogaea* L.) is an allotetraploid species ($2n = 4x = 40$; AABB-type genome) that originated less than 10,000 years ago through a spontaneous hybridization event involving two diploid wild *Arachis* species, *A. duranensis* Krapov. & W.C. Greg. (A-type genome) and *A. ipaënsis* Krapov. & W.C. Greg. (B-type genome). The hybridization of these two wild species was likely facilitated by humans, who introduced *A. ipaënsis* into the growing habitat of *A. duranensis* (Bertioli et al. 2016; Bertioli et al. 2019; de Blas et al. 2025).

Either through unreduced gametes or through hybridization followed by whole genome duplication (Garcia et al. 2020), cultivated peanut became tetraploid and genetically isolated from its wild diploid *Arachis* relatives. This event led to a significant genetic bottleneck in the peanut population and resulted in a drastic reduction of genetic diversity within the peanut gene pool (Kochert et al. 1996; Seijo et al. 2007; Bertoli et al. 2011; Moretzsohn et al. 2013; Bertoli et al. 2019).

However, the merging of two distinct diploid genomes in peanuts triggered a ‘genetic shock,’ resulting in genetic instability, large-scale deletions and inversions (McClintock 1984; Bertoli et al. 2019; Mason and Wendel 2020). As a result, peanuts evolved morphological diversity, eventually becoming a globally recognized crop, aided by an increased response to selection compared to their diploid relatives (Lamon et al. 2025a). Nowadays, peanuts are classified into two subspecies: *hypogaea* and *fastigiata*, and six botanical varieties - *hypogaea* and *hirsuta* (within the *hypogaea* subspecies), and *fastigiata*, *vulgaris*, *aequatoriana*, and *peruviana* (in the *fastigiata* subspecies). Additionally, peanut encompasses four major cultivated market types: Runner, Spanish, Valencia, and Virginia, and thousands of landraces and cultivars, each one with its own distinct characteristics (Bertoli et al. 2011, 2019).

Similar destabilizing mechanisms, including homoeologous exchange and its associated unbalanced chromosomal compositions, such as BBBB, ABBB, AAAB, and AAAA configurations, are still observed in modern cultivated peanut (Lamon et al. 2025b). Understanding the drivers of genetic instability may offer deeper insights into the mechanisms underlying increased genetic and morphological diversity in peanut, as well as its enhanced response to selection.

Double reduction (DR) is that circumstance in polyploid organisms where an individual carrying a locus in a single dosage produces duplex gametes for that same locus during meiosis (Darlington

1929). DR increases homozygosity, diminishes the frequency of deleterious alleles in a population over time through negative selection and has the potential to increase rare advantageous alleles in plant material if combined with marker-assisted selection (Butruille and Boiteux 2000; Bourke et al. 2015). In order for DR to manifest, three main events need to occur: multivalent pairing of chromosomes during prophase I, followed by crossing-over between non-sister chromatids, and a subsequent non-disjunctional chromosomal separation pattern, which implies that first and second meiotic divisions are reductional for the considered loci (Mather 1936; Ronfort et al. 1998; De Silva et al. 2005; Huang et al. 2019). The first two requirements, multivalent formation and crossing-over between non-sister chromatids, are commonly associated with homoeologous exchange (Mason and Wendel 2020).

Polysomic inheritance and multiple chromosomal pairing are typical of autopolyploid organisms since extremely high similarity is shared amongst different sets of chromosomes. In autopolyploids, ploidy increase takes place within a single species, often in a sole individual. Conversely, allopolyploids are formed by the hybridization of two different species and are also generally characterized by disomic inheritance (resulting from chromosome pairing within but not between subgenomes). In 2002, Ramsey and Schemske estimated that on average the percentage occurrence of multivalents in autopolyploid and allopolyploid plants are 28.8% and 8.0%, respectively. Even though multivalents are rarer in allopolyploids than autopolyploids, they arguably have a much higher biological significance in such species, given that recombination has a higher impact if genomes are more divergent (McGrath and Lynch 2012).

Detecting DR is a complex process that often requires a focus on specific markers and cross-combination approaches. Similarly, accurately estimating DR rate can be challenging, frequently relying on computer simulations (Tai 1982; Haynes and Douches 1993; Bourke et al. 2015).

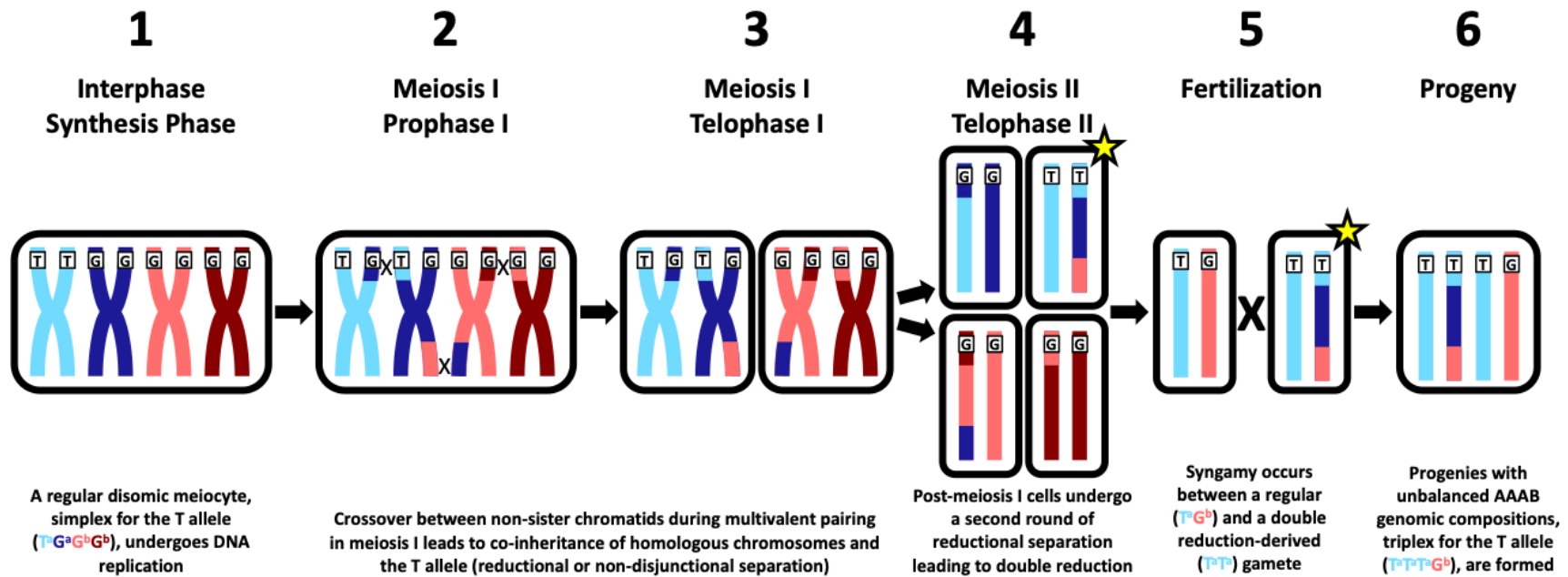
Percentages of DR have been mainly detected and estimated in autopolyploid plants e.g. purple loosestrife (*Lythrum salicaria* L.) (Fisher 1943; Fisher and Mather 1943), potato (*Solanum tuberosum* L.) (Haynes and Douches 1993; Bourke et al. 2015), great yellowcress (*Rorippa amphibia* (L.) Besser) and yellow fieldcress (*R. sylvestris* (L.) Besser) (Stift et al. 2008), purple yam (*Dioscorea alata* L.) (Nemorin et al. 2012), clementine (*Citrus clementina* Hort. ex Tan.) (Aleza et al. 2016), hybrid tea rose (*Rosa hybrida* L.) (Bourke et al. 2017), chrysanthemum (*Chrysanthemum* × *morifolium* (Ramat.) Hemsl.) (Bourke et al. 2021) and much more rarely in animals such as rainbow trout (*Oncorhynchus mykiss* Walbaum) (Allendorf and Danzmann 1997) or Pacific oysters (*Crassostrea gigas* (Thunberg)) (Curole and Hedgecock 2005).

In contrast, the detection and quantification of significant DR frequencies in allopolyploid organisms have been much more challenging. To our knowledge, DR has not been conclusively demonstrated in any confirmed allopolyploid species. Ahmed et al. (2020) estimated DR levels in the ‘Giant Key’ lime, which is likely a doubled diploid derivative of the ‘Mexican’ lime (*C. aurantifolia* (Christm.) Swingle). However, the precise polyploid origin of this species remains uncertain.

Peanut follows a third genetic model, called segmental allopolyploidy. As such, it exhibits mostly disomic pairing, with occasional multivalent formations and homoeologous exchanges (Husted 1933, 1936; Stalker 1981; Leal-Bertioli et al. 2015a). These features make the occurrence of DR theoretically possible in peanut. However, due to the rarity of multivalents and high homozygosity in cultivated peanut, detecting and estimating DR remains challenging. To overcome this limitation, we generated a BC₁ population using the neoallotetraploid [*A. magna* K 30097 × *A. stenosperma* V 15076]^{4x} (MagSten) (Fávero et al. 2015) as the donor parent and ‘IAC OL4’ (*A.*

131 *hypogaea*) (de Godoy et al. 2014) as the recurrent parent (Moretzsohn et al. 2023). This cross
 132 generated sufficient polymorphism to track individual alleles and test for the presence of DR.

133 In this study, we estimated DR occurrence, with direct evidence of an event associated with an
 134 AAAB genomic composition. Both theoretical considerations and our findings indicate that,
 135 because DR is associated with homoeologous exchange, its occurrence in an allopolyploid species
 136 such as peanut frequently results in unbalanced genomic compositions in the progenies (Figure 1).
 137 To enable accurate estimation of DR, we constructed a high-density, phased genetic linkage map,
 138 which facilitated detailed analysis of allele inheritance patterns. Additionally, we identified double
 139 linkage group (LG) artifacts in Marey plots, which resulted from large-scale unbalanced genomic
 140 compositions caused by homoeologous exchange. Resolving these artifacts improved the
 141 resolution of the genetic map and demonstrated how genetic instability can distort linkage analysis
 142 in segmental allopolyploid species such as peanut.



143

144 **Figure 1:** Mechanism of double reduction (DR) in segmental allotetraploid peanut. Letters “G” and “T” represent two different alleles.
 145 Blue (both light and dark) and red (both light and dark) colors, as well as superscripts “a” and “b”, indicate the two peanut subgenomes,
 146 A and B, respectively. 1) During the synthesis phase of interphase, a disomic meiocyte with a simplex genotype for the “T” allele
 147 ($T^a G^a G^b G^b$) undergoes DNA replication, forming sister chromatids. Although this heterozygous configuration is rare in peanut due to
 148 high subgenome homozygosity, the backcross population structure used in this study provides sufficient polymorphism for such
 149 configurations to occur (see Figure 2). 2) In prophase I, due to the segmental allopolyploid nature of peanut, both homologous and,
 150 occasionally, homoeologous chromosomes can pair, forming multivalents. For homologous and homoeologous chromosomes to pair,

151 crossing over needs to occur between non-sister chromatids (homoeologous exchange). 3) Multivalent pairing and homoeologous
152 exchange allow two homologous chromosomes carrying the “T” allele to be inherited together into the same daughter cell (reductional
153 or non-disjunctional separation), which is not possible under strict disomic inheritance. 4) During meiosis II, a second round of chromatid
154 segregation occurs. The two “T” alleles may again segregate into the same gamete (second round of reductional or non-disjunctional
155 separation), leading to the formation of a gamete with two copies of the “T” allele, which in other words is DR (highlighted by yellow
156 star). 5) During fertilization, the DR-derived gamete will very likely pair with a regular gamete resulting from disomic inheritance
157 ($T^aT^aG^bG^b$ generates T^aG^b gametes), given that inheritance is mainly disomic in peanut. 6) Progenies resulting from DR will likely have
158 unbalanced AAAB genomic compositions ($T^aT^aT^aG^b$) at the considered locus. This demonstrates how DR increases homozygosity but
159 also generates unbalanced genomic compositions at the considered locus in segmental allotetraploid species such as peanut.

Materials and Methods

Plant material

A BC₁ population was developed using the neoallotetraploid [*A. magna* K 30097 × *A. stenosperma* V 15076]^{4x} (MagSten) (Fávero et al. 2015) as the donor parent and ‘IAC OL4’ (*A. hypogaea*) (de Godoy et al. 2014) as the recurrent parent (Moretzsohn et al. 2023). A total of 74 BC₁ progenies, numbered from ‘Ind 1’ to ‘Ind 74,’ were included in the study. Additionally, BC₁F₂ progenies were derived by selfing from Ind 27, 40, and 41. These BC₁F₂ progenies were labeled based on single-seed descent identification. For example, Ind 40.2 indicates the second progeny of the BC₁ plant Ind 40.

Genotyping and sequencing

DNA was extracted from young leaves of BC₁ progenies, selected BC₁F₂ progenies, and ‘IAC OL4’ and MagSten parentals. Genotyping was performed using the Axiom *Arachis* 48K SNP array v2 (Affymetrix, Santa Clara, CA, USA) (Pandey et al. 2017; Korani et al. 2019). In addition, four BC₁ progenies (Ind 33, 39, 47, and 62) and two BC₁F₂ progenies (Ind 40.3 and 40.4) underwent whole-genome sequencing (WGS) at 30× coverage using PCR-free Illumina NovaSeq PE150 technology (Novogene Co., Sacramento, CA, USA).

Genotypic data scoring and linkage mapping

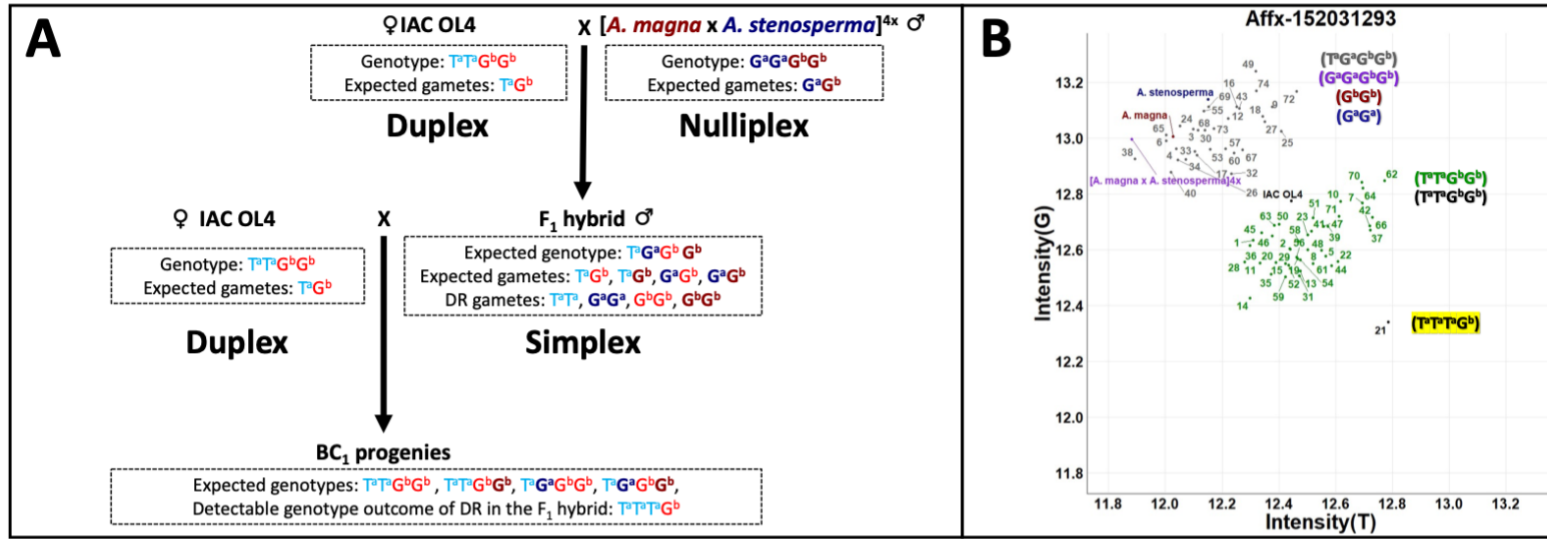
Raw .CEL files from the Axiom *Arachis* 48K SNP array v2 were processed using the Axiom Analysis Suite 2022 (Thermo Fisher Scientific, Waltham, MA, USA) to generate a summarized

signal intensity file, which was then imported into RStudio for further analysis (Posit Team 2025). SNP markers for tetraploid samples were scored on a tetraploid scale (0 to 4) using the ‘fitPoly’ package (Voorrips, Gort and Vosman 2011; Voorrips and Gort 2018), with ‘IAC OL4’ and MagSten given as parents of the BC₁ population, as the F₁ hybrid genotypic data were unavailable (see Supporting Information, Table S1).

Markers filtering and linkage mapping were performed using ‘polymapR’ (Bourke et al. 2018). Initially, markers were filtered for non-missing values in the parents, retaining 25,315 markers (Figure S1A). Subsequently, non-segregating markers were removed and dosages were converted to basic segregating types, resulting in a final set of 11,565 markers (Figure S1B). Next, markers and individuals of the BC₁ population were filtered for missing values, applying a cutoff rate of 0.25 for both. The allele dosage matrix was then filtered for duplicated individuals (cutoff value 0.95), identifying four pairs of duplicated individuals and four individuals resulting from self-pollination in ‘IAC OL4’. After merging these cases, 66 individuals remained for linkage mapping. Finally, the allele dosage matrix was filtered for duplicated markers, retaining 5,225 unique markers (Figure S1C).

Most markers were initially scored by ‘fitPoly’ as simplex x nulliplex and nulliplex x simplex types (Figure S1, B and C). However, because cultivated peanut is characterized by extremely high subgenome homozygosity, simplex x nulliplex markers are expected to be very rare. As MagSten, was entered in place of the F₁ hybrid the pollen parent for genotype calling, it is more accurate to interpret these simplex x nulliplex markers as actually representing duplex x simplex segregation types (Figure 2A). For further mapping purposes, all markers erroneously called as simplex x nulliplex were re-labelled as nulliplex x simplex.

Double Reduction in Phase 1 Linkage Map



Double Reduction in Phase 2 Linkage Map

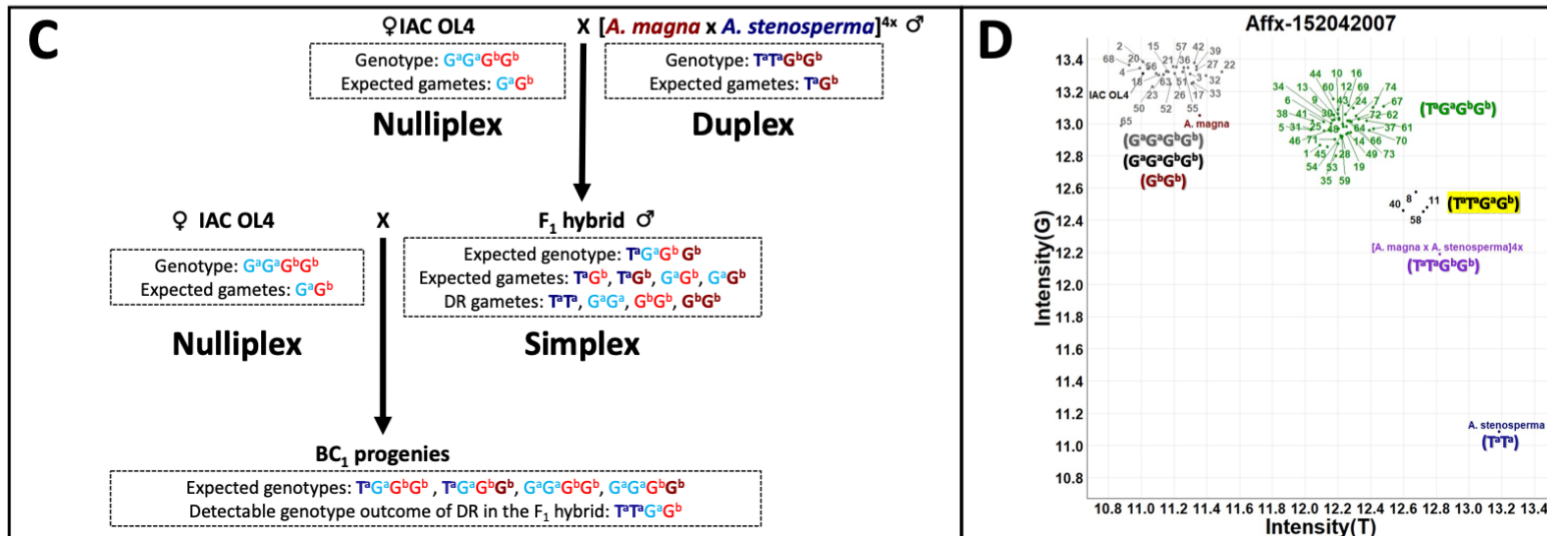


Figure 2: Double reduction (DR) occurrence in segmental allotetraploid peanut. Diagrams show DR events in the ‘IAC OL4’ $\times$ F₁ hybrid BC₁ population. Panels A and B illustrate linkage map phase 1; panels C and D show phase 2. “G” and “T” indicate alleles, with superscripts “a” and “b” for peanut subgenomes. In panels A and C, colors indicate *Arachis* genome-species combinations: ‘IAC OL4’ (light blue/red for a/b), *A. magna* (b; dark red), and *A. stenosperma* (a; dark blue). In panels B and D, numbers label BC₁ progenies, with y-axis showing “G” allele intensity and x-axis “T” allele intensity.

Panel A: If ‘IAC OL4’ is duplex (T^aT^aG^bG^b) and MagSten nulliplex for “T” (G^aG^aG^bG^b), the F₁ hybrid is simplex (T^aG^aG^bG^b), producing T^aG^b or G^aG^b gametes under disomic inheritance, while ‘IAC OL4’ produces only T^aG^b. DR in the F₁ hybrid can generate T^aT^a gametes, resulting in unbalanced AAAB triplex compositions (T^aT^aT^aG^b), distinguishable from simplex/duplex scores.

Panel B: SNP intensities (Affx-152031293) reflect panel A, with color-coded genotype-genome-allele combinations; AAAB DR-derived BC₁ progeny (T^aT^aT^aG^b) is highlighted in yellow.

Panel C: If ‘IAC OL4’ is nulliplex (G^aG^aG^bG^b) and MagSten duplex for “T” (T^aT^aG^bG^b), the F₁ hybrid is simplex (T^aG^aG^bG^b), producing T^aG^b or G^aG^b gametes under disomic inheritance, while ‘IAC OL4’ yields only G^aG^b. DR in the F₁ hybrid can generate T^aT^a gametes, producing unbalanced AAAB duplex compositions (T^aT^aG^aG^b), distinguishable from nulliplex/simplex scores.

Panel D: SNP intensities (Affx-152042007) reflect panel C, with color-coded genotype-genome-allele combinations; AAAB DR-derived BC₁ progenies (T^aT^aG^aG^b) are highlighted in yellow.

Markers were aligned against A and B subgenomes of the ‘Tifrunner’ (*A. hypogaea*) reference genome (version gnm2) to determine their physical positions (Bertioli et al. 2019). The alignment process was conducted in the Unix environment.

Pairwise recombination frequency, LOD scores, and linkage phases were calculated for markers using the ‘polymapR’ linkage() function with ploidy equal 2, with an expectation of 20 LGs. Marker clusters were generated over a range of LOD thresholds, with a minimum cluster size of 11 markers; at a threshold of 11, unique markers were clustered into 20 groups. Concordance of marker positions across the two reference assemblies, along with the application of the phase_SN_diploid() function, was used to phase markers between the A and B subgenomes. Inter-LG linkage signal was then assessed and duplicated markers were reintegrated, yielding a total of 9,717 markers (see Supporting Information, Table S2).

The classification of markers during genotype calling into duplex × simplex (re-labelled as nulliplex × simplex from mapping purposes) and nulliplex × simplex effectively phased markers between the two parental lines, with duplex × simplex corresponding to phase 1 (‘IAC OL4’) and nulliplex × simplex corresponding to phase 2 (F₁ hybrid). A very high concordance was observed between “genotyping calling phasing” and linkage map phasing (Figure S2).

Summary statistics were calculated for the integrated map, and Marey plots were generated to illustrate the relationship between genetic distances and the physical positions on the ‘Tifrunner’ reference genome, with genetic positions reoriented where necessary to match the physical map; pericentromeric regions were highlighted. Out of 9,717 markers, 1,628 were excluded from the plots because their corresponding subgenomes lacked ‘Tifrunner’ physical positions. The plots revealed unexpected double LG artifacts in LGs 4 and 14, with the artifact in LG14 being so pronounced that it resulted in the formation of two distinct LGs.

248

249 **Interspersed genotyping patterns and Marey plots resolution**

250 During linkage map construction, we observed pronounced mapping artifacts in some LGs,
251 including the appearance of doubled lines and even an entire doubled LG in Marey plots. These
252 artifacts suggested that a subset of progenies might violate common assumptions (such as balanced
253 tetraploid inheritance) used for linkage analysis.

254 After initial map refinement by removing markers with missing alleles or lacking physical
255 alignment information, the integrated map was divided into two phases. Examination of the phased
256 maps—particularly phase 2—revealed extensive interspersed genotyping patterns, characterized
257 by rapid switches in inferred allele dosage between consecutive markers ordered by physical
258 position. Such interspersed genotyping patterns are indicative of unbalanced polyploid states (e.g.
259 AAAB or ABBB) that produce allele dosage states incompatible with the genotyping model
260 assumptions and have previously been described by Lamon et al. (2025a). In that study,
261 unbalanced genotypic compositions could not be resolved due to limited dosage discrimination
262 but were partially mitigated by scoring genotypes on a tetraploid scale using ‘fitPoly’ and
263 restricting analyses to markers with clear A- and B-subgenome differentiation. In the present study,
264 however, resolving interspersed genotyping proved more challenging.

265 Although BC₁ progenies were scored on a tetraploid scale, the mapping population comprised four
266 distinct subgenomic sources: A and B subgenomes from the cultivated parent (‘IAC OL4’) and A
267 and B subgenomes from the wild parents (*A. stenosperma* and *A. magna*). In phase 1, A- and B-
268 subgenome alleles were distinguishable in the cultivated parent but identical in the wild parent,
269 whereas the opposite configuration occurred in phase 2 (Figure 2). This phase-dependent

subgenomic symmetry reduced effective dosage resolution and likely contributed to interspersed genotyping patterns, similar to those previously reported, complicating the identification of progenies with unbalanced genomic compositions.

To systematically identify such cases, we developed an algorithm to detect regions of interspersed genotyping. Phased colormaps were first converted into binary matrices indicating whether allele dosage changed between consecutive markers. A sliding window of 11 markers (five upstream, the focal marker, and five downstream) was applied; if more than four dosage changes occurred within the window, the central marker was classified as part of an interspersed genotyping region. These regions were visualized as chromosome-level heatmaps, and the number of interspersed markers per chromosome was summarized using bar plots.

Interspersed genotyping was most pronounced on LG14 of phase 2, which was also the LG exhibiting strong double-LG artifacts in Marey plots. To test whether these artifacts were caused by progenies with abnormal genomic composition, we quantified interspersed genotyping on LG14 (phase 2) for each BC₁ progeny. This analysis identified 18 progenies with markedly elevated interspersed patterns (Figure S3A).

Linkage mapping was then repeated after excluding these progenies. This intervention resulted in the complete disappearance of the double LG artifacts, particularly on LGs 4 and 14, and restored coherent Marey plot structure. These results demonstrate that large-scale unbalanced genomic compositions—manifesting as interspersed genotyping patterns—were responsible for the observed double-LG artifacts, rather than errors in marker ordering or map construction.

Estimation of DR

The integrated linkage map was refined by removing markers with missing alleles or lacking physical alignment information and subsequently divided into the two phases. Regions showing DR-type segregation were identified. While similar patterns could also result from tetrasomic recombination in the female parent during the backcross in phase 1 or from two consecutive rounds of tetrasomic recombination in the male parent in phase 2, DR provides the most parsimonious explanation for the observed patterns, especially in phase 2 (Figure 2).

To minimize the impact of genotyping errors, only regions where multiple consecutive markers displayed DR scores were considered as true DR events. This filtering step was performed manually through visual inspection of phased colormaps. Final DR inferences for both phases were visualized using heatmaps using ‘ggplot2’, providing a chromosome-level overview of DR event distribution.

DR estimation using a hidden Markov model

To independently validate DR inference, we explored an alternative approach based on estimating posterior genotype probabilities using a multivalent-compatible Hidden Markov Model (Zheng et al. 2016). However, such an approach would require an integrated genetic linkage map, which is impossible to create due to lack of linkage between subgenomes (e.g. LG1 and LG11 are syntenic with chromosome 01 but cannot be combined into a single linkage map). To circumvent this difficulty, we used a physical-assembly guided approach to create a pseudo-integrated linkage map of ten groups via centromere alignment. Firstly, pericentromeric regions were identified visually from Marey maps (genetic vs physical marker positions) using consensus physical positions (Figure S4A). The median genetic position of markers within these centromeres was taken as the

approximate genetic position of the centromeres, and for each chromosome, one of the maps was shifted upwards so that centromeres were genetically aligned (Figure S4B). A pseudo-integrated map was created by merging both maps and re-sorting markers on the new consensus genetic positions (Figure S4C, Figure S5). This was converted to a phased map by concatenating the phase information per chromosome across subgenomes into a “pseudo-autotetraploid” phased map (adding a dummy phase for parent 1, for which no segregating markers were present).

The pseudo-integrated phased map and marker genotypes were used to estimate posterior genotype (IBD) probabilities in ‘polyqtlR’ v.0.1.1 (Bourke et al. 2021) using the `estimate_IBD()` function with settings `ploidy = 4` and `bivalent_decoding = FALSE` (i.e. allowing for the possibility of multivalent formation). These probabilities were subsequently re-estimated using the `maxL_IBD()` function using a range of genotyping error priors ($\epsilon = 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35$). Maximum likelihood estimates for the genotyping error rates per individual per chromosome were visualized in a heatmap.

Sequencing data analysis

To assess subgenome balance and evaluate genomic composition in selected progenies, we first identified a robust set of subgenome-diagnostic polymorphisms. This was done by using independent sources of short sequences aligned to a common reference genome: *in silico*–fragmented assemblies of *Arachis duranensis* (A genome) and *A. ipaënsis* (B genome), and Illumina WGS data from two independent sources. Using multiple sources ensured that subgenome-specific alleles were consistently identifiable and reliably recoverable from Illumina sequencing data in the progenies, following Lamon et al. (2025a). Illumina reads were quality-

filtered using BBMap (v39.01) (Bushnell 2014) and, together with the *in silico*-fragmented assemblies, were aligned to the *A. ipaënsis* gnm1 reference genome using BWA (v0.7.17-r1188) (Li and Durbin 2009), which provided a common coordinate framework. Variant calling was performed using BCFtools (version 1.15.1) (Danecek et al. 2021) to identify positions that consistently differentiated the A and B subgenomes. AB-diagnostic variant positions were retained only if they met stringent criteria, including concordance between sequencing datasets, a minimum of 95% consensus among mapped reads, sufficient read depth, and clear allele differentiation between the A and B subgenomes. This procedure yielded a fixed set of high-confidence subgenome-specific polymorphisms.

Genomic composition in progenies was inferred by calculating normalized counts of subgenome-diagnostic alleles at the predefined AB-diagnostic positions. Illumina WGS reads from selected BC₁ and BC₁F₂ progenies (Ind 33, 39, 47, 62, 40.3, and 40.4) were then aligned to the same reference genome, and allele counts were extracted exclusively at these positions. This targeted interrogation allowed quantitative assessment of subgenome allele representation without *de novo* variant discovery. To normalize allele counts and account for potential mapping and coverage biases, synthetic tetraploid controls were generated by combining HudsonAlpha Illumina WGS reads from *A. duranensis* and *A. ipaënsis* in a 1:1 ratio (taking into account the different genome sizes). These synthetic datasets were processed identically to the progeny samples, and allele counts were calculated. Raw allele counts in the progenies were subsequently normalized using the formulas described in Lamon et al. (2025a), enabling direct comparison across samples. Normalized allele distributions for the selected BC₁ and BC₁F₂ progenies were visualized using density plots generated with the ‘ggplot2’ package in RStudio.

RESULTS

Development of a high-density phased peanut linkage map

A high-density phased linkage map was developed, distinguishing two phases in the BC₁ population (Figure 3). The phase 1 map included 3,194 SNP markers, while the phase 2 map contained 6,523 SNP markers, for a total of 9,717 mapped markers (Table 1; see Supporting Information, Table S2). On average, phase 1 had 160 markers per chromosome, with the highest density observed on LG2 (234 markers). Phase 2 averaged 326 markers per chromosome, peaking at 559 markers on LG2.

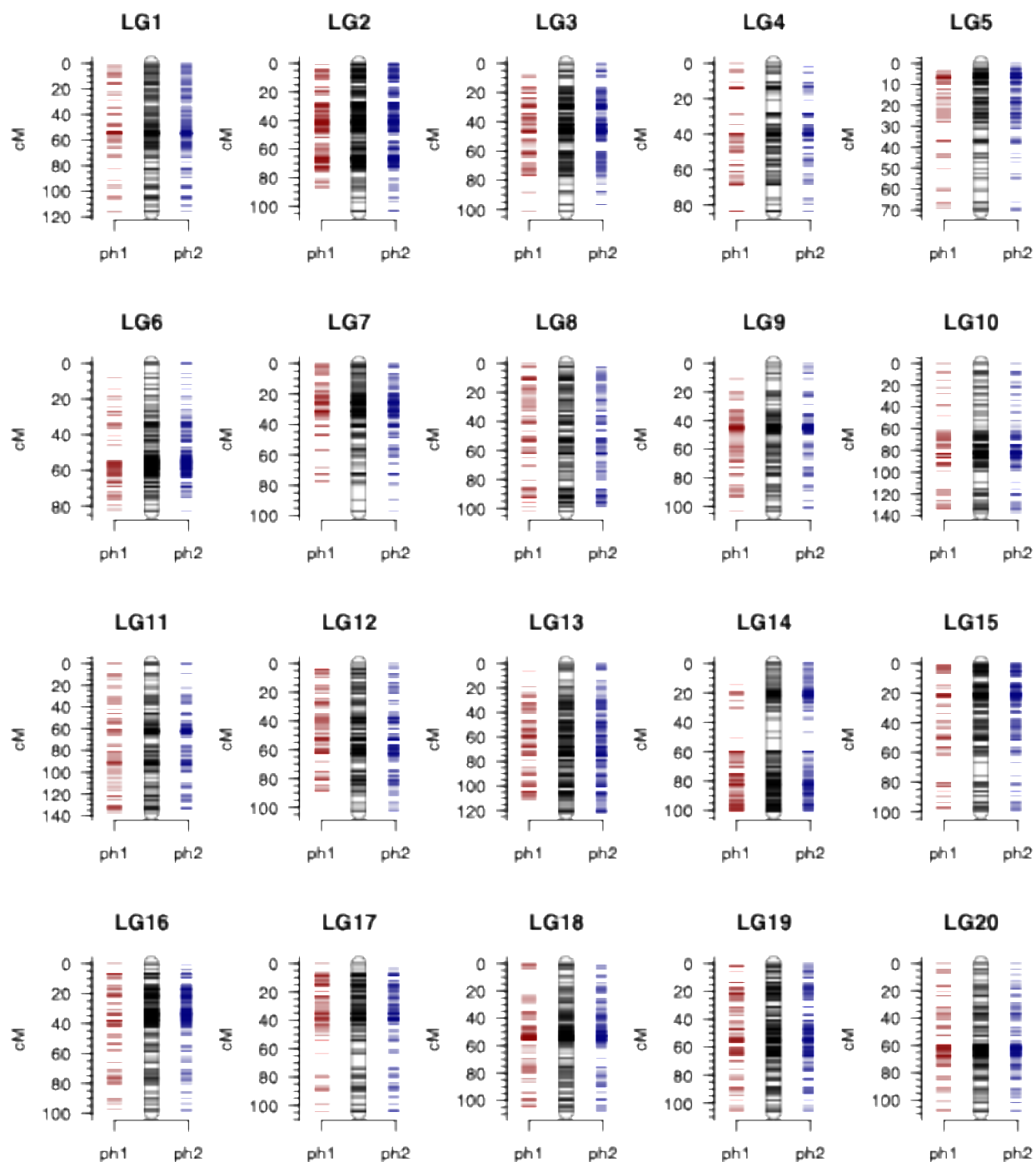


Figure 3: Integrated linkage map displaying 9,717 phased SNP markers (black) across 20 linkage groups (LGs 1–10 represent the A subgenome, and LGs 11–20 represent the B subgenome).

371 Duplex × simplex markers (phase 1) are shown in red, while nulliplex × simplex markers (phase
372 2) are depicted in blue.

373

374 **Table 1:** Summary statistics of the integrated linkage maps, including the number of SNP markers
375 for phase 1 and phase 2, linkage group lengths, average marker distances, and maximum marker
376 gap distances.

Subgenome	Linkage Group	No. of SNPs Phase 1	No. of SNPs Phase 2	Linkage Group Length (cM)	Average Distance (cM)	Maximum Gap Distance (cM)
A-subgenome	LG1	169	505	116.11	0.17	5.65
	LG2	234	559	103.74	0.13	6.25
	LG3	174	345	101.53	0.2	6.6
	LG4	118	145	83.54	0.32	7.55
	LG5	114	214	70.24	0.21	5.08
	LG6	146	337	82.93	0.17	5.54
	LG7	182	343	97.22	0.19	9.2
	LG8	166	305	102.24	0.22	4.13
	LG9	189	304	103.71	0.21	8.52
	LG10	139	295	136.88	0.32	10.43
B-subgenome	LG11	123	349	137.01	0.29	7.44

	LG12	127	244	103.21	0.28	5.56
	LG13	184	345	121.37	0.23	7.54
	LG14	169	277	100.94	0.23	9.04
	LG15	130	304	100.09	0.23	9.12
	LG16	157	298	99.00	0.22	5.77
	LG17	146	326	104.63	0.22	9.39
	LG18	207	361	108.44	0.19	4.41
	LG19	155	326	106.5	0.22	4.12
	LG20	165	341	108.75	0.21	4.78
Across the whole map		3194	6523	2088.08	0.22	10.43

377

378 The integrated map covered a total genetic distance of 2,088.08 cM, with individual LG lengths

379 ranging from 70.24 cM (LG5) to 137.01 cM (LG11). The average marker interval was 0.22 cM,

380 while maximum distances between markers varied from 4.12 cM (LG19) to 10.43 cM (LG10).

381 Strong cross-LG signals were found between LG3 and LG13, LG4 and LG14, and LG9 and LG19

382 (Figure S6).

383 Marey plots displayed distinct S-shapes with smooth transitions (Figure 4A), indicative of a high-

384 quality linkage map and consistent with expected recombination rate variation along the

385 chromosomes. An exception was observed for LG5, which showed near-complete suppression of

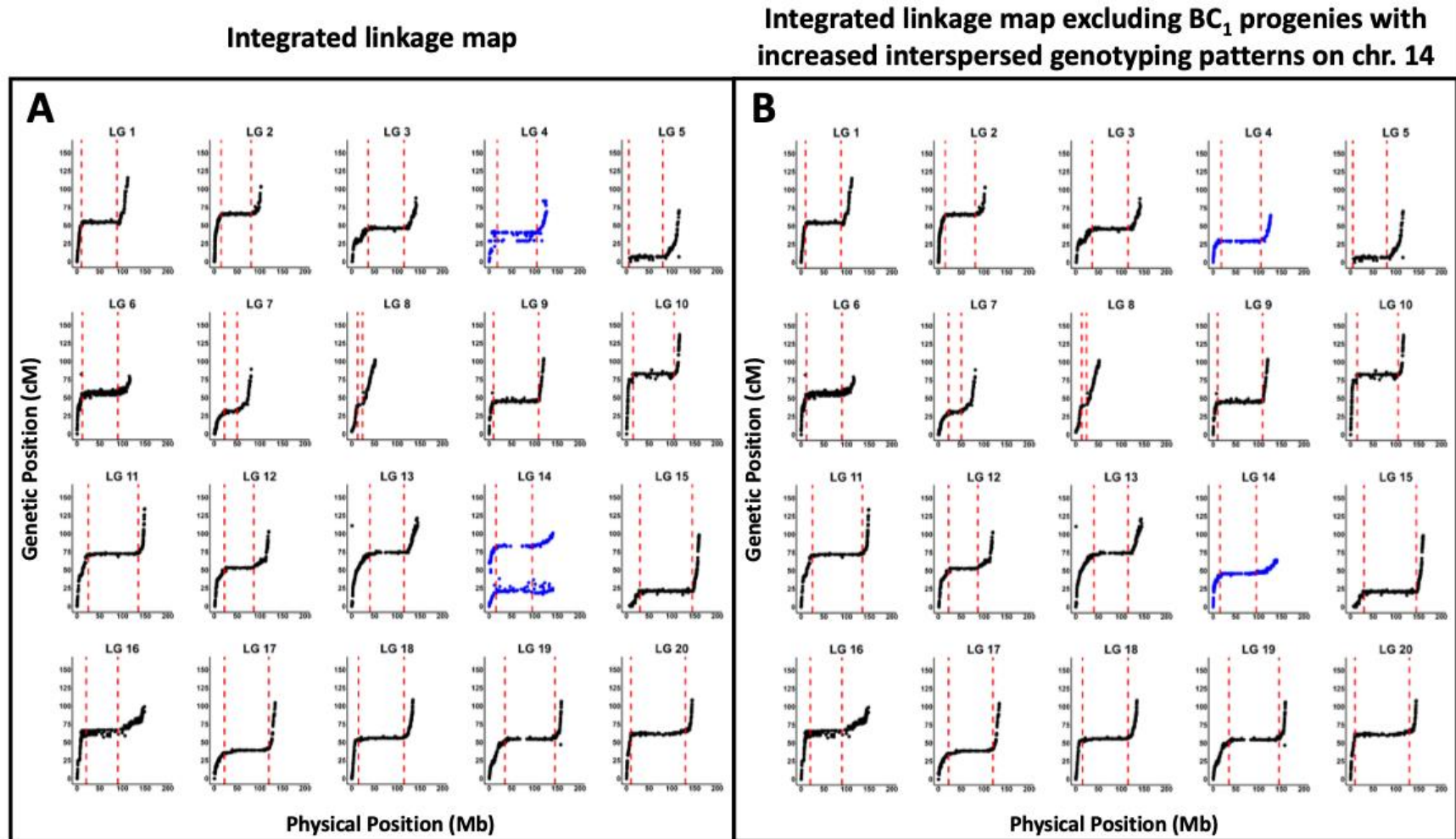
386 recombination along its upper arm. This pattern is consistent with the presence of a large post-

387 polyploidy inversion in this region of *A. hypogaea* (Bertioli et al. 2019), an inversion that likely

388 stabilized the homeologous chromosome pair by disrupting collinearity between the A and B

389 subgenomes and would also be expected to suppress recombination between cultivated and wild

390 chromosome 05 homologs. In addition, LG4 and LG14 showed extreme “doubled-line” artifacts.
391 LG14 was particularly affected, presenting as two complete S-shaped curves.



392

393 **Figure 4:** Marey plots of the integrated linkage map (A) and the map after excluding BC₁ progenies with increased interspersed
 394 genotyping patterns on linkage group (LG) 14 (B) are shown. The y-axis represents genetic positions, and the x-axis corresponds to

395 physical positions in the ‘Tifrunner’ reference genome. In panel A, double LG artifacts are evident, particularly on LGs 4 and 14 (blue).
396 These artifacts disappear in panel B following the exclusion of the affected progenies. Pericentromeric regions are indicated by red
397 dashed lines.

Genotyping and sequencing analyses reveal a correlation between interspersed genotyping patterns and unbalanced genomic compositions, which significantly affect linkage mapping

In this section, we describe how we discovered that two initially unexpected observations—chromosome-scale interspersed SNP genotyping patterns and the appearance of “doubled” LGs—were related, and we identify their underlying cause.

Regions with interspersed genotyping patterns were most frequently observed in chromosome set 4 (LGs 4 and 14), followed by sets 2 (LGs 2 and 12), 6 (LGs 6 and 16), 3 (LGs 3 and 13), 5 (LGs 5 and 15), and 9 (LGs 9 and 19), in descending order of frequency (Figure S3B, Figure S7, A and B). Overall, phase 1 exhibited substantially fewer interspersed genotyping patterns than phase 2, although it still showed a notable level of background noise. Notably, the LGs containing the highest numbers of individuals with interspersed genotyping patterns also showed the most extreme “doubled” LG behavior.

With this in mind, we followed the intuition that these chromosome-scale interspersed genotyping patterns represented some chromosomal entity, or biological circumstance, that did not follow standard assumptions, and that this might underlie the doubled LG patterns. To test this, 18 BC₁ progenies with pronounced interspersed genotyping on LG14 in the phase 2 map were excluded, and linkage mapping was repeated. Satisfyingly, this eliminated the doubled LG patterns (Figure 4B, Figure S3A).

Then, to investigate the nature of the entity or circumstance underlying the interspersed genotyping patterns, we performed WGS on BC₁ progenies with (Ind 47 and 62) and without (Ind 33 and 39) the interspersed genotyping patterns in LG14. In addition, BC₁F₂ progenies Ind 40.3 and 40.4,

419 descendants of BC₁ individual Ind 40, which also exhibited interspersed genotyping on LG14,
420 were sequenced.

421 Sequencing revealed that individuals with interspersed genotyping patterns had unbalanced
422 genomic compositions on the same chromosomes. Specifically, AAAB configurations were
423 detected in chromosome set 4 in Ind 47 and 62 that displayed interspersed genotyping, but not in
424 Ind 33 and 39 that had normal genotyping patterns (Figure 5, A–D). Similarly, BC₁F₂ progenies Ind
425 40.3 and 40.4 showed the same large unbalanced AAAB configuration in chromosome set 4
426 (Figure 5, E and F).

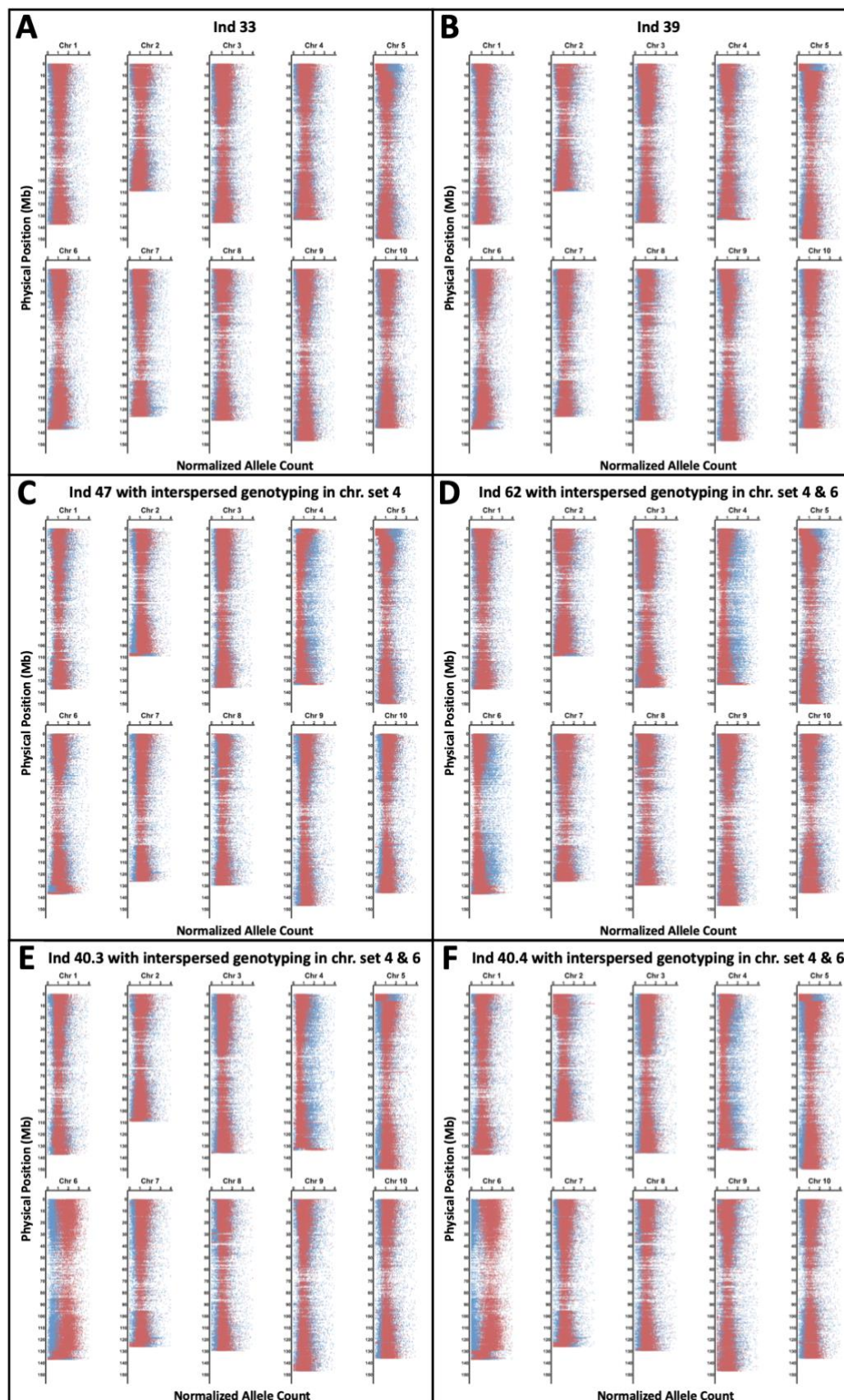


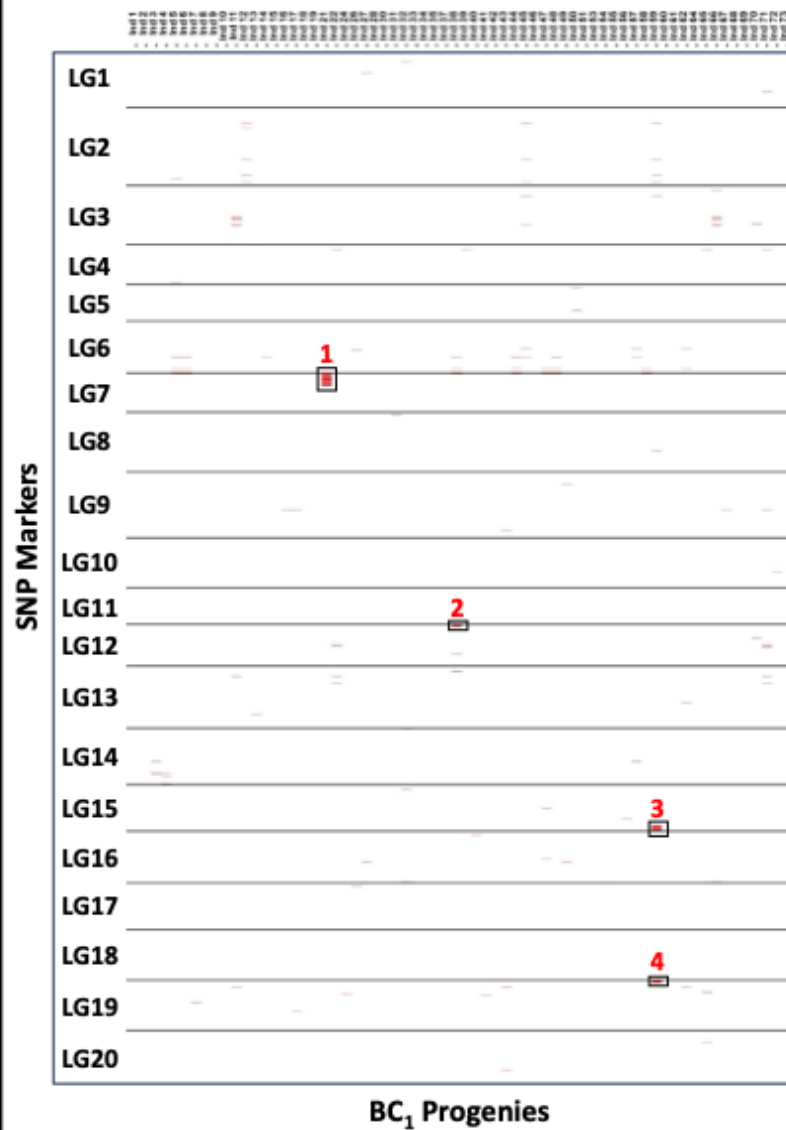
Figure 5: Density plots illustrating read counts by chromosome for sequenced genotypes distinguish between *Arachis duranensis*-like (AA-genome; blue) and *A. ipaënsis*-like (BB-genome; red) subgenomes. Chromosomes with a balanced composition of A and B subgenomes (AABB genome) show read counts of 1 for both A and B reads. In contrast, chromosomes with genome shifts towards either *A. duranensis* or *A. ipaënsis* display varying values, with common scenarios being A = 1.5 and B = 0.5 for the AAAB-genome, and A = 0.5 and B = 1.5 for the ABBB-genome. Panels A and B show Ind 33 and Ind 39, both with regular (balanced) chromosome set 4. Panel C displays Ind 47, which has an unbalanced chromosome set 4, consistent with the presence of interspersed genotyping patterns. Panel D shows Ind 62, which exhibits unbalanced chromosome sets 4 and 6, also reflecting the interspersed genotyping patterns observed in genotyping analysis. Panels E and F present BC₁F₂ progenies Ind 40.3 and 40.4, both derived from Ind 40, each displaying unbalanced chromosome sets 4 and 6 as predicted from genotyping analysis.

Across all individuals, chromosome sets 2–7 exhibited individual-specific genomic composition variation, whereas sets 1, 8, 9, and 10 were largely balanced. Notably, individuals with an AAAB composition at the top of chromosome set 5 had yellow flowers (Figure S8). This is in agreement with a previous report of the genetic loci determining flower color, (Bertioli et al. 2019), and the dominance of the yellow wild allele carried on the B subgenome in this population.

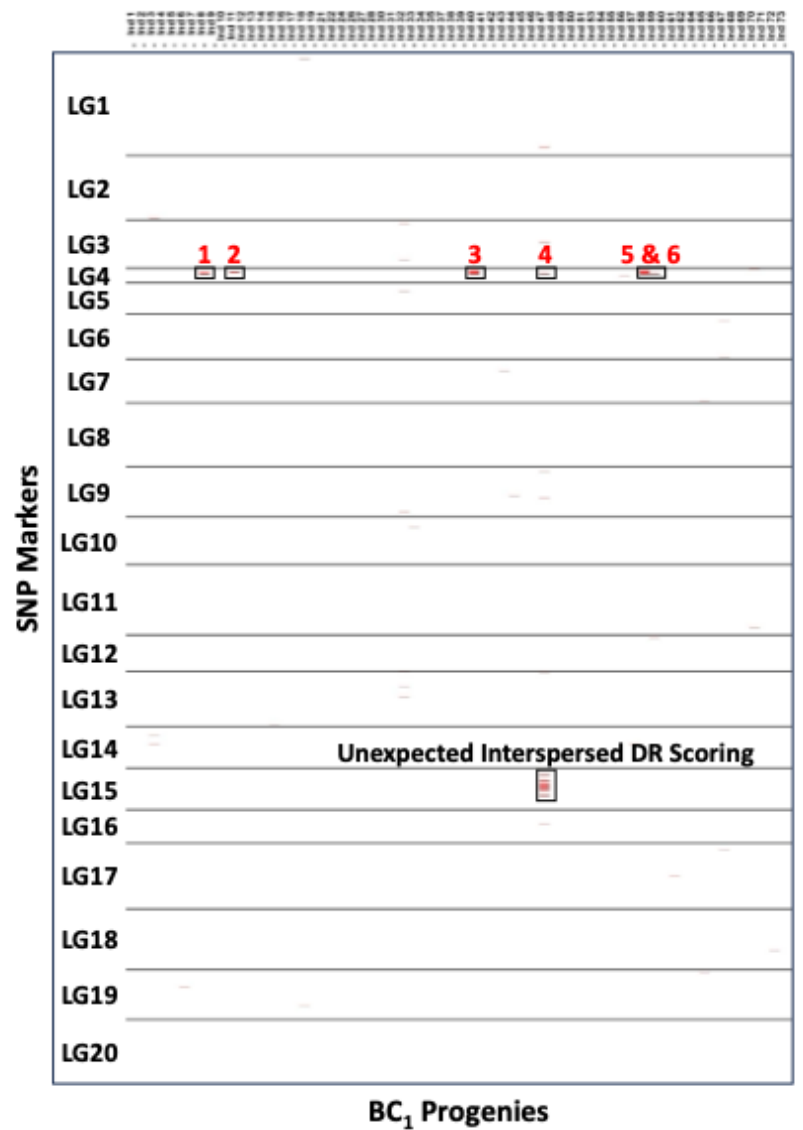
DR estimation in segmental allotetraploid peanut

449 Pericentromeric boundaries for each LG were visually estimated to allow clearer interpretation of
 450 DR event distribution (Figure 4, A and B; see Supporting Information, Table S3). In both phase 1
 451 and phase 2, DR scores were predominantly detected near the telomeric regions, consistent with
 452 previous studies (see Supporting Information, Table S4) (Bourke et al. 2015). Specifically, in
 453 phase 1, DR events occurred at the top of LG7 (0.34–6.54 Mb, Ind 21), the top of LG12 (1.49–
 454 1.77 Mb, Ind 38), the bottom of LG15 (157.87–159.92 Mb, Ind 59), and the top of LG19 (0.18–
 455 0.43 Mb, Ind 59), with all physical positions referenced to the ‘Tifrunner’ genome (Figure 6A). In
 456 phase 2, DR events were consistently observed at the top of LG4 (1.94–3.88 Mb in the ‘Tifrunner’
 457 genome) in six individuals (Ind 8, 11, 40, 47, 58, and 59) (Figure 6B). Thus, DR events were most
 458 frequent on LG4 in phase 2, whereas in phase 1 they were distributed across several LGs but
 459 always confined to telomeric regions.

A Double Reduction in Phase 1 Linkage Map

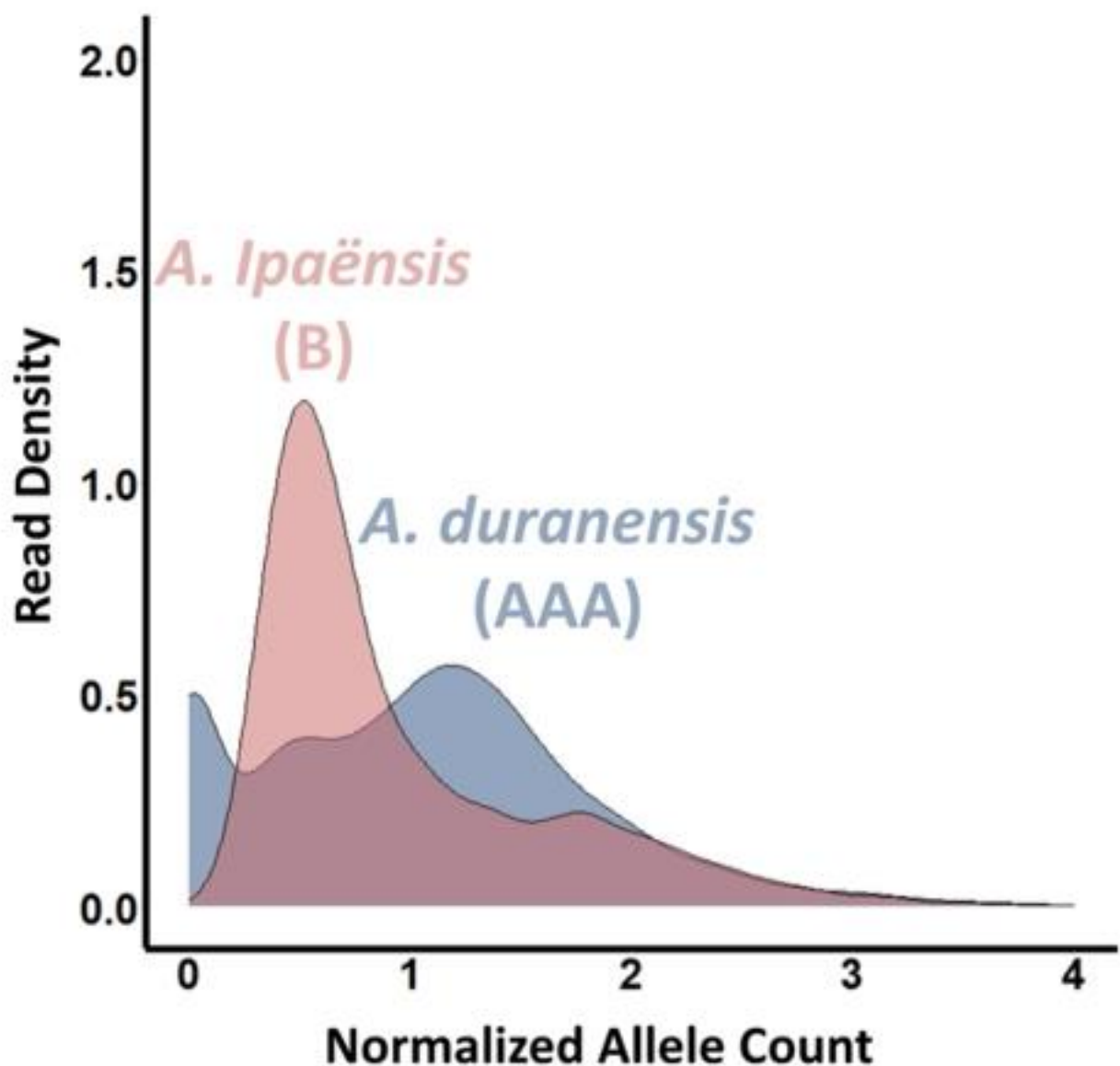


B Double Reduction in Phase 2 Linkage Map



461 **Figure 6:** Heatmaps display double reduction (DR) events in BC₁ progenies based on phase 1 (A) and phase 2 (B) linkage maps, with
462 SNP markers ordered by genetic distance. Red marks indicate SNP scores consistent with DR, while regions with consecutive DR-
463 consistent scores, identified as DR events, are highlighted with black rectangles and numbered in red. In agreement with previous reports,
464 DR events occurred either at the top or bottom of the linkage groups: the top of LG7 (1), top of LG12 (2), bottom of LG15 (3), and top
465 of LG19 (4) in the phase 1 map, and the top of LG4 (1–6) in the phase 2 map. In phase 2, unexpected interspersed DR scoring was
466 observed across LG15 in Ind 47.

467 Among the BC₁ progenies, 12% (8 out of 66) showed evidence of DR in either phase 1 or phase 2
 468 linkage maps (see Supporting Information, Table S4). Because only a subset of marker
 469 combinations was analyzed, this figure likely underestimates the true frequency of DR. In addition,
 470 an AAAB genomic configuration was detected in the sequenced individual Ind 47 within the DR
 471 region (Figure 5C, Figure 7). This result confirms the theoretical expectation that DR events
 472 generate unbalanced genomic composition in progenies (Figure 1).



473

Figure 7: Normalized allele count read density plot for the upper region of LG4 (3.15–3.88 Mb in the ‘Tifrunner’ genome; 4.27–4.97 Mb in *Arachis ipaënsis* gnm1) of individual 47, where double reduction duplex scores were detected. The expected genome composition for this region is $T^aT^aG^aG^b$, where “T” and “G” are arbitrary alleles, with the T^a allele provided by *A. stenosperma*, the G^a allele from the A subgenome of ‘IAC OL4’, and the G^b allele from the B subgenome of ‘IAC OL4’. The plot demonstrates a shift in genome composition toward *A. duranensis* (AAAB), as indicated by a normalized allele count of approximately 1.5 for the A subgenome and 0.5 for the B subgenome. This scenario corresponds to the genomic configuration depicted in Figure 2, C and D.

However, Ind 47 also displayed an unexpected interspersed pattern of DR scores across LG15, including substantial presence in the pericentromeric region, an area typically not associated with DR (Bourke et al. 2015). Notably, Ind 47 did not exhibit the anticipated unbalanced genomic configuration in this region (Figure 5C), suggesting that a different genetic mechanism, rather than DR, may underlie these patterns.

Hidden Markov model revealed homoeologous exchange but could not confirm DR

The Hidden Markov Model using a tetrasomic model allowed for both bivalent and multivalent pairing. The model was run at the default error prior setting (Figure S9, Figure S10). With these setting, two individuals were predicted to have arisen from multivalent pairing in the parental meioses, namely Ind 47 on chromosome 05 and Ind 21 on chromosome 07 (Figure S11). However, when the process was repeated over a range of possible error priors, the maximum likelihood

results no longer predicted multivalent pairing for these (or any other) individuals (Figure S11). However, there were six individuals that were predicted to have arisen from homoeologous exchanges on chromosomes 2, 3, 5 and 7 (Figure S12).

Discussion

In this study, we estimated DR occurred in 12% of the analyzed BC₁ peanut progenies, a frequency that is unexpectedly high for a segmental allotetraploid species. This observation is consistent with previous findings in peanut (Leal-Bertioli et al. 2015a), which reported elevated levels of tetrasomic inheritance, despite the predominance of disomic inheritance. While DR frequencies in autopolyploids are generally higher due to their higher rate of multivalent formation, our results suggest that, in peanut, DR is still a rare but recurrent event and can be reliably estimated when sufficient genomic resolution and appropriate population structures are used.

However, it is important to recognize that several methodological limitations likely resulted in an underestimation of DR rates in our study. For example, the number of informative marker combinations available for estimating DR was limited, and stringent filtering of marker dosages was necessary to ensure accurate segregation scoring. Additionally, the inherent structure of the BC₁ population only permits estimation of certain DR events. Alternative mechanisms, such as tetrasomic inheritance in the female parent in phase 1 or consecutive rounds of tetrasomic inheritance in the male parent in phase 2, could also generate genotypic patterns that resemble DR, potentially leading to overestimation. Nevertheless, our conservative scoring approach, which required multiple consecutive markers with consistent dosage, minimizes the risk of false positives and strengthens confidence in the DR events we report.

Most notably, our results provide direct evidence that DR, which depends on homoeologous exchange, could lead to an increase of unbalanced genomic compositions, such as ABBB and AAAB states, in the progeny of allopolyploid species like peanut (Figure 1, Figure 7). These findings establish a new theoretical framework for estimating DR in allopolyploids by demonstrating that the presence of ABBB and AAAB genome compositions can serve as a possible indicator of DR events.

All identified DR events in our study were localized to telomeric regions, in agreement with established theory and previous empirical observations (Mather 1936; Fisher 1947; Haynes and Douches 1993; Butruille and Boiteux 2000; Stift et al. 2008; Nemorin et al. 2012; Bourke et al. 2015). This localization likely reflects higher crossover frequency and multivalent formation near telomeres compared to the more structurally constrained pericentromeric regions. Interestingly, we observed an unusual interspersed pattern of DR-like scores across the pericentromeric region of LG15 in Ind 47, which was not associated with an unbalanced genomic configuration in the corresponding chromosome set. This finding suggests other genetic phenomena could underlie these observations. Ind 47 was also initially predicted to have arisen from DR using the HMM, although this prediction was subsequently lost at a higher error prior. However, the autotetraploid HMM we employed is reliant on a high-quality integrated linkage map. Our pseudo-integrated map was visibly discordant on certain LGs including chromosome 05 (Figure S5), which may have affected the quality of the results. Therefore, determining presence or absence of unbalanced genomic compositions could become useful for distinguishing true DR events from other genotypic patterns in allopolyploid mapping populations.

The generation and analysis of a BC₁ population was critical for uncovering these findings. Because peanut is highly self-pollinating, establishing a backcross population requires substantial

manual effort. However, this approach was essential for generating the segregation patterns needed to detect the transition from a single dosage to duplex gametes (Moretzsohn et al. 2023). Furthermore, the development of the BC₁ population allowed us to construct a high-density phased linkage map for peanut, using SNP chip technology and the ‘fitPoly’ and ‘polymapR’ packages in RStudio.

This map achieved high marker density compared to previous studies and proved crucial for estimating DR (Zhou et al. 2014; Leal-Bertioli et al. 2015b; Huang et al. 2016; Li et al. 2017; Han et al. 2018; Ballén-Taborda et al. 2019; Liu et al. 2020; de Blas et al. 2021; Sun et al. 2022; Miao et al. 2023; Tsai et al. 2024). High marker density enhances both the sensitivity and accuracy of DR estimation, as well as the ability to resolve complex patterns of recombination and genome composition in allopolyploid backgrounds.

Our mapping and sequencing analyses further revealed that regions with frequent interspersed genotyping patterns, especially in LG14, coincided with unbalanced genomic compositions and double LGs artifacts in Marey plots. In some cases, these mapping artifacts were so extensive that they split a single LG into two distinct segments. Although similar issues have been reported informally in previous studies, they have rarely been addressed in detail in the literature. Commonly, researchers have focused on the primary LG and excluded alternative signals to simplify map interpretation. While this practice improves clarity, it also reduces genetic resolution and may overlook important aspects of genetic instability.

The distortions caused by unbalanced genome composition and DR are likely to remain a recurring challenge in genetic mapping of peanut and other allopolyploid organisms. Unbalanced configurations resulting from homoeologous exchange or DR not only complicate map construction, but also affect the interpretation of genetic diversity and inheritance patterns in

breeding populations (Lamon et al. 2025b). This highlights the need for analytical tools and pipelines that can detect and account for these genomic features during linkage analysis.

To address these challenges, future mapping strategies in allopolyploids should incorporate genome composition data, ideally through sequencing or improved dosage scoring, as early as possible in map construction. As demonstrated in this study, the systematic identification and filtering of hotspots for interspersed genotyping prior to map building, combined with sequencing to confirm the presence and boundaries of unbalanced genomic segments, can help prevent mapping artifacts and improve the accuracy of both map construction and downstream analyses.

Another option is to improve genotyping software so that it can dynamically accommodate the complexity of allopolyploid genome structures. For example, in peanut, in genomic regions where frequent multivalent formation and recombination between divergent subgenomes are expected, standard tetraploid scoring (0–4 scale) may not be sufficient to represent all possible allele dosages. Software that allows a more granular and context-dependent dosage scale, coupled with linkage mapping tools capable of handling such data, would be particularly valuable for peanut and other segmental allopolyploids.

Beyond their technical implications, our findings have evolutionary significance. The estimation of DR as a rare but consistent form of genetic instability in cultivated peanut provides direct evidence for the ongoing generation of new allelic combinations, even in crops with narrow genetic bottlenecks and a high degree of subgenome fixation. Both DR and homoeologous exchange contribute to genome plasticity, offering a partial escape from the genetic constraints imposed by polyploidization bottleneck and domestication. These mechanisms may have played an important role in the morphological diversification and adaptability that have characterized peanut's evolutionary history.

In summary, this study estimated DR, a rare but biologically meaningful outcome of genetic instability in peanut, which depends on homoeologous exchange and multivalent formation. Our findings underscore the value of high-density linkage mapping in detecting and interpreting rare meiotic events and call for the continued development of analytical approaches that can accurately accommodate the genomic complexity of segmental allopolyploids. Addressing these challenges will be essential for improving the precision of genetic studies and breeding strategies not only in peanut, but also in other important allopolyploid crops.

Conclusions

This study presents a high-density phased linkage map for peanut, constructed using 9,717 SNP markers. Unbalanced genomic compositions were identified in several BC₁ progenies, mainly caused by homoeologous exchange, which were shown to reduce map quality and introduce artifacts such as double LGs. Excluding progeny with extensive genomic imbalance improved overall map resolution and accuracy. Evidence of DR was found in 12% of the BC₁ population progenies, across LGs 4, 7, 12, 15, and 19. Of particular interest, a DR event on LG4 corresponded with an AAAB genome configuration, lending empirical support to the hypothesis that DR can result in unbalanced chromosome sets in segmental allopolyploids like peanut. The ability to estimate such rare events has significant implications for understanding the evolution and genetic complexity of allopolyploid organisms.

Data availability statement

WGS data are currently under embargo and will be released upon publication. Sequencing analysis scripts are available at <https://github.com/brianabernathy/ABGD>. All other data generated or analyzed during this study are included in the supporting information.

Acknowledgments

Special thanks are extended to Adriana Regina Custodio, Jennifer Leverett, and Mark Hopkins for their valuable contributions. Samuele Lamon also acknowledges the Institute of Plant Breeding, Genetics, and Genomics at the University of Georgia for the 2023 Peggy Ozias-Akins Leadership Award.

Study Funding

Peter Bourke received funding from the USDA’s NIFA Speciality Crop Research Initiative project ‘Tools for genomics-assisted breeding in polyploids: Development of a community resource’ (2020-51181-32156/SCRI).

Conflict of Interest

The authors declare no competing interests.

Author Contributions

Samuele Lamon conducted the genotypic data analysis, including linkage mapping, identification of interspersed genotyping patterns, and estimation of double reduction. He also performed sequencing data visualization and co-wrote the manuscript. Peter Bourke assisted with genotypic data analysis, linkage mapping, and estimation of double reduction, and contributed to manuscript preparation. Brian Abernathy oversaw the sequencing data analysis. João Francisco dos Santos and Ignácio José de Godoy developed and provided the plant material for the study. Soraya Leal-Bertioli conceptualized the project. David Bertioli conceptualized the project, contributed to sequencing data analysis, and co-wrote the manuscript.

References

- Ahmed D, Curk F, Evrard JC, Froelicher Y, Ollitrault P. 2020. Preferential disomic segregation and *C. micrantha*/*C. medica* interspecific recombination in tetraploid ‘Giant Key’ lime; Outlook for Triploid Lime Breeding. *Front Plant Sci.* 11. doi:10.3389/fpls.2020.00939.
- Aleza P, Cuenca J, Juárez J, Navarro L, Ollitrault P. 2016. Inheritance in doubled-diploid clementine and comparative study with SDR unreduced gametes of diploid clementine. *Plant Cell Rep.* 35(8):1573–1586. doi:10.1007/s00299-016-1972-4.
- Allendorf FW, Danzmann RG. 1997. Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. *Genetics.* 145(4):1083–1092. doi:10.1093/genetics/145.4.1083.
- Ballén-Taborda C, Chu Y, Ozias-Akins P, Timper P, Holbrook CC, Jackson SA, Bertioli DJ, Leal-Bertioli SCM. 2019. A new source of root-knot nematode resistance from *Arachis*

stenosperma incorporated into allotetraploid peanut (*Arachis hypogaea*). Sci Rep. 9.
doi:10.1038/s41598-019-54183-1.

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger JP, Dash S, et al. 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nature Genetics, 48(4):438–446. doi:10.1038/ng.3517.

Bertioli DJ, Jenkins J, Clevenger JP, Dudchenko O, Gao D, Seijo G, Leal-Bertioli SCM, Ren L, Farmer AD, Pandey MK, et al. 2019. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nat Genet. 51(5):877–884. doi:10.1038/s41588-019-0405-z.

Bertioli DJ, Seijo G, Freitas FO, Valls JFM, Leal-Bertioli SCM, Moretzsohn MC. 2011. An overview of peanut and its wild relatives. Plant Genetic Resources: Characterization and Utilization. 9(1):134–149. doi:10.1017/S1479262110000444.

de Blas FJ, Bruno CI, Arias RS, Ballén-Taborda C, Mamani E, Oddino C, Rosso M, Costero BP, Bressano M, Soave JH, et al. 2021. Genetic mapping and QTL analysis for peanut smut resistance. BMC Plant Biol. 21. doi:10.1186/s12870-021-03023-4.

Bourke PM, Arens P, Voorrips RE, Esselink GD, Koning-Boucoiran CFS, van't Westende WPC, Leonardo TS, Wissink P, Zheng C, van Geest G, et al. 2017. Partial preferential chromosome pairing is genotype dependent in tetraploid rose. Plant Journal. 90(2):330–343. doi:10.1111/tpj.13496.

Bourke PM, Van Geest G, Voorrips RE, Jansen J, Kranenburg T, Shahin A, Visser RGF, Arens P, Smulders MJM, Maliepaard C. 2018. polymapR—linkage analysis and genetic map

669 construction from F1 populations of outcrossing polyploids. *Bioinformatics*. 34(20):3496–
670 3502. doi:10.1093/bioinformatics/bty371.

671 Bourke PM, Voorrips RE, Hackett CA, Van Geest G, Willemsen JH, Arens P, Smulders MJM,
672 Visser RGF, Maliepaard C. 2021. Detecting quantitative trait loci and exploring
673 chromosomal pairing in autopolyploids using polyqtlR. *Bioinformatics*. 37(21):3822–
674 3829. doi:10.1093/bioinformatics/btab574.

675 Bourke PM, Voorrips RE, Visser RGF, Maliepaard C. 2015. The double-reduction landscape in
676 tetraploid potato as revealed by a high-density linkage map. *Genetics*. 201(3):853–863.
677 doi:10.1534/genetics.115.181008.

678 Bushnell B. 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner. [accessed 2025 Nov 6].
679 <https://escholarship.org/uc/item/1h3515gn>.

680 Butruille DV, Boiteux LS. 2000. Selection-mutation balance in polysomic tetraploids: Impact of
681 double reduction and gametophytic selection on the frequency and subchromosomal
682 localization of deleterious mutations. *PNAS*. 97(12):6608-6613.
683 doi:10.1073/pnas.100101097.

684 Curole JP, Hedgecock D. 2005. Estimation of preferential pairing rates in second-generation
685 autotetraploid pacific oysters (*Crassostrea gigas*). *Genetics*. 171(2):855–859.
686 doi:10.1534/genetics.105.043042.

687 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
688 McCarthy SA, Davies RM. 2021. Twelve years of SAMtools and BCFtools. *Gigascience*.
689 10(2):1–4. doi:10.1093/gigascience/giab008.

690 Darlington CD. 1929. Chromosome behaviour and structural hybridity in the *Tradescantiae*. J
691 Genet. 21:207–286.

692 Fávero AP, Pádua JG, Costa TS, Gimenes MA, de Godoy IJ, Moretzsohn MC, Michelotto MD.
693 2015. New hybrids from peanut (*Arachis hypogaea* L.) and synthetic amphidiploid crosses
694 show promise in increasing pest and disease tolerance. Genetics and Molecular Research.
695 14(4):16694–16703. doi:10.4238/2015.December.11.17.

696 Fisher RA. 1943. Allowance for double reduction in the calculation of genotype frequencies with
697 polysomic inheritance. Ann Eugen. 12(1):169–171. doi:10.1111/j.1469-
698 1809.1943.tb02320.x.

699 Fisher RA. 1947. The theory of linkage in polysomic inheritance. Philosophical Transactions of
700 the Royal Society of London. Biological Sciences. 233(594):55-87.
701 doi:10.1098/rstb.1947.0006.

702 Fisher RA, Mather K. 1943. The inheritance of style length in *Lythrum salicaria*. Ann Eugen.
703 12(1):1–23. doi:10.1111/j.1469-1809.1943.tb02307.x.

704 García AV, Ortiz AM, Silvestri MC, Custodio AR, Moretzsohn MC, Lavia GI. 2020. Occurrence
705 of 2n microspore production in diploid interspecific hybrids between the wild parental
706 species of peanut (*Arachis hypogaea* L., Leguminosae) and its relevance in the genetic
707 origin of the cultigen. Crop Science. 60(5):2420-2436. doi: 10.1002/csc2.20233.

708 de Godoy IJ, dos Santos JF, de Carvalho CRL, Michelotto MD, Bolonhezi D, de Freitas RS, Kasai
709 FS, Ticelli M, Finoto EL, Martins ALM. 2014. IAC OL 3 and IAC OL 4: new Brazilian
710 peanut cultivars with the high oleic trait. Crop Breeding and Applied Biotechnology.
711 14:200–203. doi:10.1590/1984-70332014v14n3a30.

712 Han S, Yuan M, Clevenger JP, Li C, Hagan A, Zhang X, Chen C, He G. 2018. A snp-based linkage
713 map revealed QTLs for resistance to early and late leaf spot diseases in peanut (*Arachis*
714 *hypogaea* L.). Front Plant Sci. 9(1012):1–9. doi:10.3389/fpls.2018.01012.

715 Haynes KG, Douches DS. 1993. Estimation of the coefficient of double reduction in the cultivated
716 tetraploid potato. Theoret. Appl. Genetics. 85(6):857-862. doi:10.1007/BF00225029.

717 Huang K, Wang T, Dunn DW, Zhang P, Cao X, Liu R, Li B. 2019. Genotypic frequencies at
718 equilibrium for polysomic inheritance under double-reduction. G3: Genes, Genomes,
719 Genetics. 9(5):1693-1706. doi:10.1534/g3.119.400132.

720 Huang L, Ren X, Wu B, Li X, Chen W, Zhou X, Chen Y, Pandey MK, Jiao Y, Luo H, et al. 2016.
721 Development and deployment of a high-density linkage map identified quantitative trait
722 loci for plant height in peanut (*Arachis hypogaea* L.). Sci Rep. 6(1):1–10.
723 doi:10.1038/srep39478.

724 Husted L. 1933. Cytological studies on the peanut, *Arachis*. I. Chromosome number and
725 morphology. Cytologia. 5(1):109–117. doi:10.1508/cytologia.5.109.

726 Husted L. 1936. Cytological studies on the peanut, *Arachis*. II. Chromosome number, morphology
727 and behavior, and their application to the problem of the origin of the cultivated forms.
728 Cytologia. 7(3):396–423. doi:10.1508/cytologia.7.396.

729 Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K. 1996. RFLP and cytogenetic
730 evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis*
731 *hypogaea* (Leguminosae). Am J Bot. 83(10):1282–1291. doi:10.1002/j.1537-
732 2197.1996.tb13912.x.

Korani W, Clevenger JP, Chu Y, Ozias-Akins P. 2019. Machine Learning as an effective method for identifying true single nucleotide polymorphisms in polyploid plants. *Plant Genome*. 12. doi:10.3835/plantgenome2018.05.0023.

Lamon S, Abernathy BL, Leal-Bertioli SCM, Bertioli DJ. 2025a. From bottleneck to boom: polyploidy, genetic instability and response to artificial selection resolve the peanut paradox. *The Plant Journal*. 124(6):1-17. doi:10.1111/tpj.70618.

Lamon S, Abernathy BL, Leal-Bertioli SCM, Bertioli DJ. 2025b. Surprisingly frequent chromosomal instability in cultivated peanut. *The Plant Journal*. 124(6):1-13. doi:10.1111/tpj.70617.

Leal-Bertioli SCM, Shirasawa K, Abernathy BL, Moretzsohn M, Chavarro C, Clevenger JP, Ozias-Akins P, Jackson SA, Bertioli DJ. 2015a. Tetrasomic recombination is surprisingly frequent in allotetraploid *Arachis*. *Genetics*. 199(4):1093–1105. doi:10.1534/genetics.115.174607.

Leal-Bertioli SCM, Cavalcante U, Gouvea EG, Ballén-Taborda C, Shirasawa K, Guimarães PM, Jackson SA, Bertioli DJ, Moretzsohn MC. 2015b. Identification of QTLs for rust resistance in the peanut wild species *Arachis magna* and the development of KASP markers for marker-assisted selection. *G3: Genes, Genomes, Genetics*. 5(7):1403–1413. doi:10.1534/g3.115.018796.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14):1754–1760. doi:10.1093/bioinformatics/btp324.

753 Li Y, Li L, Zhang X, Zhang K, Ma D, Liu J, Wang X, Liu F, Wan Y. 2017. QTL mapping and
754 marker analysis of main stem height and the first lateral branch length in peanut (*Arachis*
755 *hypogaea* L.). Euphytica. 213. doi:10.1007/s10681-017-1847-y.

756 Liu H, Sun Z, Zhang X, Qin L, Qi F, Wang Z, Du P, Xu J, Zhang Z, Han S, et al. 2020. QTL
757 mapping of web blotch resistance in peanut by high-throughput genome-wide sequencing.
758 BMC Plant Biol. 20. doi:10.1186/s12870-020-02455-8.

759 Mason AS, Wendel JF. 2020. Homoeologous exchanges, segmental allopolyploidy, and polyploid
760 genome evolution. Front Genet. 11(1014):1–10. doi:10.3389/fgene.2020.01014.

761 Mather K. 1936. Segregation and linkage in autotetraploids. J Genet. 32(2):287–314.

762 McClintock B. 1984. The significance of responses of the genome to challenge. Science.
763 226(4676):792–801. doi:10.1126/science.15739260.

764 McGrath CL, Lynch M. 2012. Evolutionary significance of whole-genome duplication. In: Soltis
765 PS, Soltis DE, editors. Polyploidy and genome evolution. Berlin: Springer.
766 doi:10.1007/978-3-642-31442-1_1.

767 Miao P, Meng X, Li Z, Sun S, Chen CY, Yang X. 2023. Mapping quantitative trait loci (QTLs)
768 for hundred-pod and hundred-seed weight under seven environments in a recombinant
769 inbred line population of cultivated peanut (*Arachis hypogaea* L.). Genes. 14(9).
770 doi:10.3390/genes14091792.

771 Moretzsohn MC, dos Santos JF, Moraes ARA, Custódio AR, Michelotto MD, Mahrajan N, Leal-
772 Bertioli SCM, de Godoy IJ, Bertioli DJ. 2023. Marker-assisted introgression of wild

773 chromosome segments conferring resistance to fungal foliar diseases into peanut (*Arachis*
774 *hypogaea* L.). Front Plant Sci. 14. doi:10.3389/fpls.2023.1139361.

775 Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SCM, Valls JFM, Bertioli DJ. 2013. A
776 study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely
777 related wild species using intron sequences and microsatellite markers. Ann Bot.
778 111(1):113–126. doi:10.1093/aob/mcs237.

779 Nemorin A, Abraham K, David J, Arnau G. 2012. Inheritance pattern of tetraploid *Dioscorea alata*
780 and evidence of double reduction using microsatellite marker segregation analysis.
781 Molecular Breeding. 30(4):1657–1667. doi:10.1007/s11032-012-9749-0.

782 Pandey MK, Agarwal G, Kale SM, Clevenger JP, Nayak SN, Sriswathi M, Chitikineni A, Chavarro
783 C, Chen X, Upadhyaya HD, et al. 2017. Development and evaluation of a high density
784 genotyping “Axiom-*Arachis*” array with 58 K SNPs for accelerating genetics and breeding
785 in groundnut. Sci Rep. 7. doi:10.1038/srep40577.

786 Posit Team. 2025. RStudio: integrated development environment for R. Boston (MA): Posit
787 Software, PBC. [cited 2025 Nov 3]. Available from: <http://www.posit.co/>.

788 Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. Annu Rev Ecol Syst. 33:589–
789 639. doi:10.1146/annurev.ecolsys.33.010802.150437.

790 Ronfort J, Jenczewski E, Bataillon T, Rousset F. 1998. Analysis of population structure in
791 autotetraploid species. Genetics. 150(2):921-930. doi:10.1093/genetics/150.2.921.

792 Seijo G, Lavia GI, Fernández A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EA. 2007.
793 Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae)

794 and its close relatives revealed by double GISH. *Am J Bot.* 94(12):1963–1971.
795 doi:10.3732/ajb.94.12.1963.

796 De Silva HN, Hall AJ, Rikkerink E, McNeilage MA, Fraser LG. 2005. Estimation of allele
797 frequencies in polyploids under certain patterns of inheritance. *Heredity.* 95(4):327–334.
798 doi:10.1038/sj.hdy.6800728.

799 Stalker HT. 1981. Hybrids in the genus *Arachis* between sections *Erectoides* and *Arachis*. *Crop*
800 *Sci.* 21(3):359–362. doi:10.2135/cropsci1981.0011183x002100030002x.

801 Stift M, Berenos C, Kuperus P, Van Tienderen PH. 2008. Segregation models for disomic,
802 tetrasomic and intermediate inheritance in tetraploids: A general procedure applied to
803 *Rorippa* (yellow cress) microsatellite data. *Genetics.* 179(4):2113–2123.
804 doi:10.1534/genetics.107.085027.

805 Sun Z, Qi F, Liu H, Qin L, Xu J, Shi L, Zhang Z, Miao L, Huang B, Dong W, et al. 2022. QTL
806 mapping of quality traits in peanut using whole-genome resequencing. *The Crop Journal.*
807 10(1):177–184. doi:10.1016/j.cj.2021.04.008.

808 Tai GCC. 1982. Estimation of double reduction and genetic parameters of autotetraploids.
809 *Heredity.* 49(1):63–70. doi:10.1038/hdy.1982.65.

810 Tsai YC, Brenneman TB, Gao D, Chu Y, Lamon S, Bertoli DJ, Leal-Bertoli SCM. 2024. The
811 Identification of the peanut wild relative *Arachis stenosperma* as a source of resistance to
812 stem rot and analyses of genomic regions conferring disease resistance through QTL
813 mapping. *Agronomy.* 14(7). doi:10.3390/agronomy14071442.

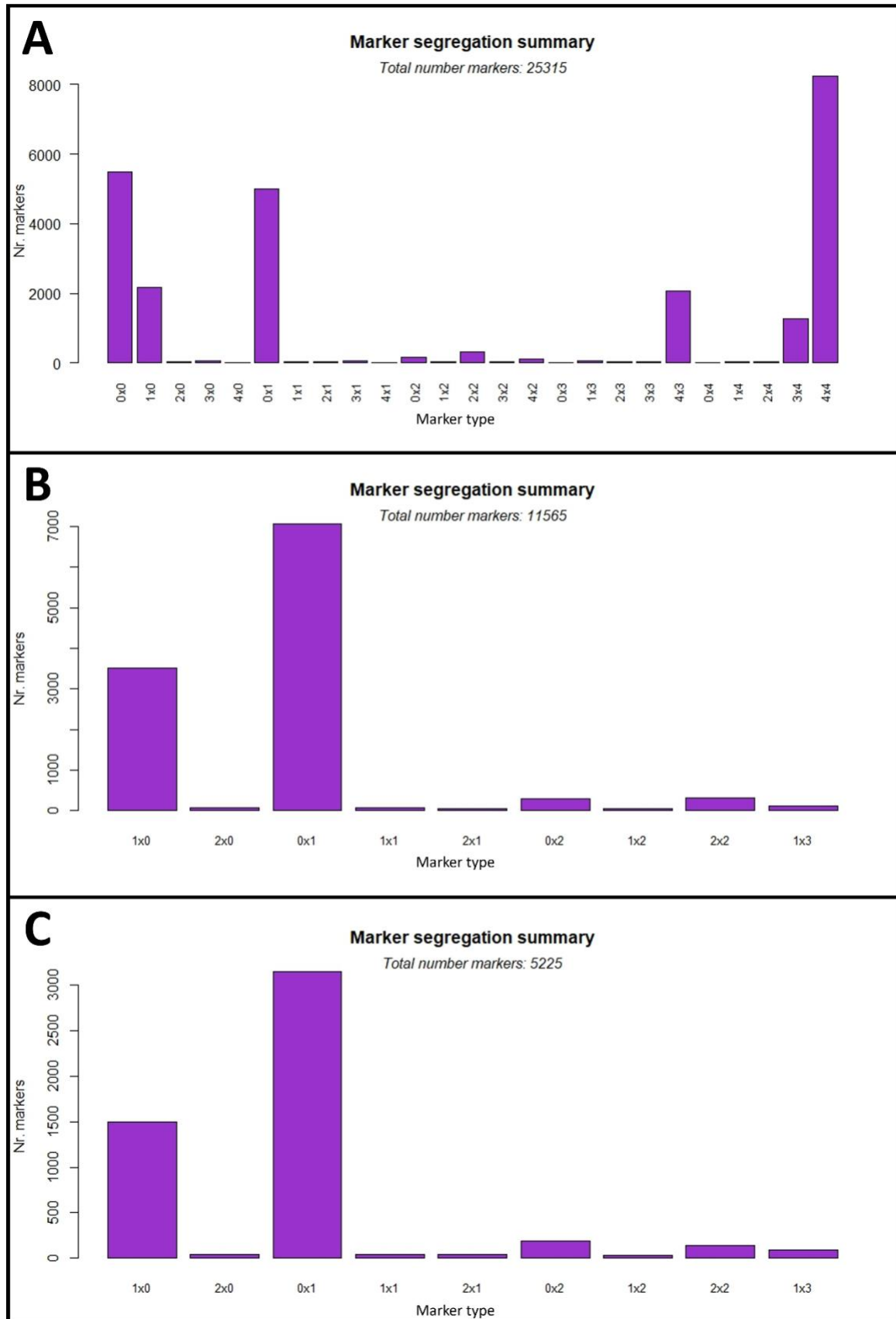
814 Villanueva RAM, Chen ZJ. 2019. Ggplot2: elegant graphics for data analysis. Measurement:
815 Interdisciplinary Research and Perspectives. 17(3):160–167.
816 doi:10.1080/15366367.2019.1565254.

817 Voorrips RE, Gort G. 2018. fitPoly: genotype calling for bi-allelic marker assays. R package
818 version 3.0.

819 Voorrips RE, Gort G, Vosman B. 2011. Genotype calling in tetraploid species from bi-allelic
820 marker data using mixture models. BMC Bioinformatics. 12(1):172. doi:10.1186/1471-
821 2105-12-172.

822 Zheng C, Voorrips RE, Jansen J, Hackett CA, Ho J, Bink MCAM. 2016. Probabilistic multilocus
823 haplotype reconstruction in outcrossing tetraploids. Genetics. 203(1):119–131.
824 doi:10.1534/genetics.115.185579.

825 Zhou X, Xia Y, Ren X, Chen Y, Huang L, Huang S, Liao B, Lei Y, Yan L, Jiang H. 2014.
826 Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale
827 marker development using next-generation double-digest restriction-site-associated DNA
828 sequencing (ddRADseq). BMC Genomics. 15(1):1–14. doi:10.1186/1471-2164-15-351.



830 **Figure S1:** Panel A shows the segregation of markers after filtering for non-missing values in both
 831 ‘IAC OL4’ and MagSten, retaining 25,315 markers from the BC₁ population, with MagSten as the
 832 donor and ‘IAC OL4’ as the recurrent parent. Panel B displays the segregation after converting
 833 marker combinations to basic types and removing non-segregating markers, resulting in 11,565
 834 retained markers. In panel C, the segregation is shown after removing non-segregating markers
 835 and screening for duplicates, yielding 5,225 retained markers. The x-axis represents marker
 836 combinations: nulliplex (0), simplex (1), duplex (2), triplex (3), and tetraplex (4).

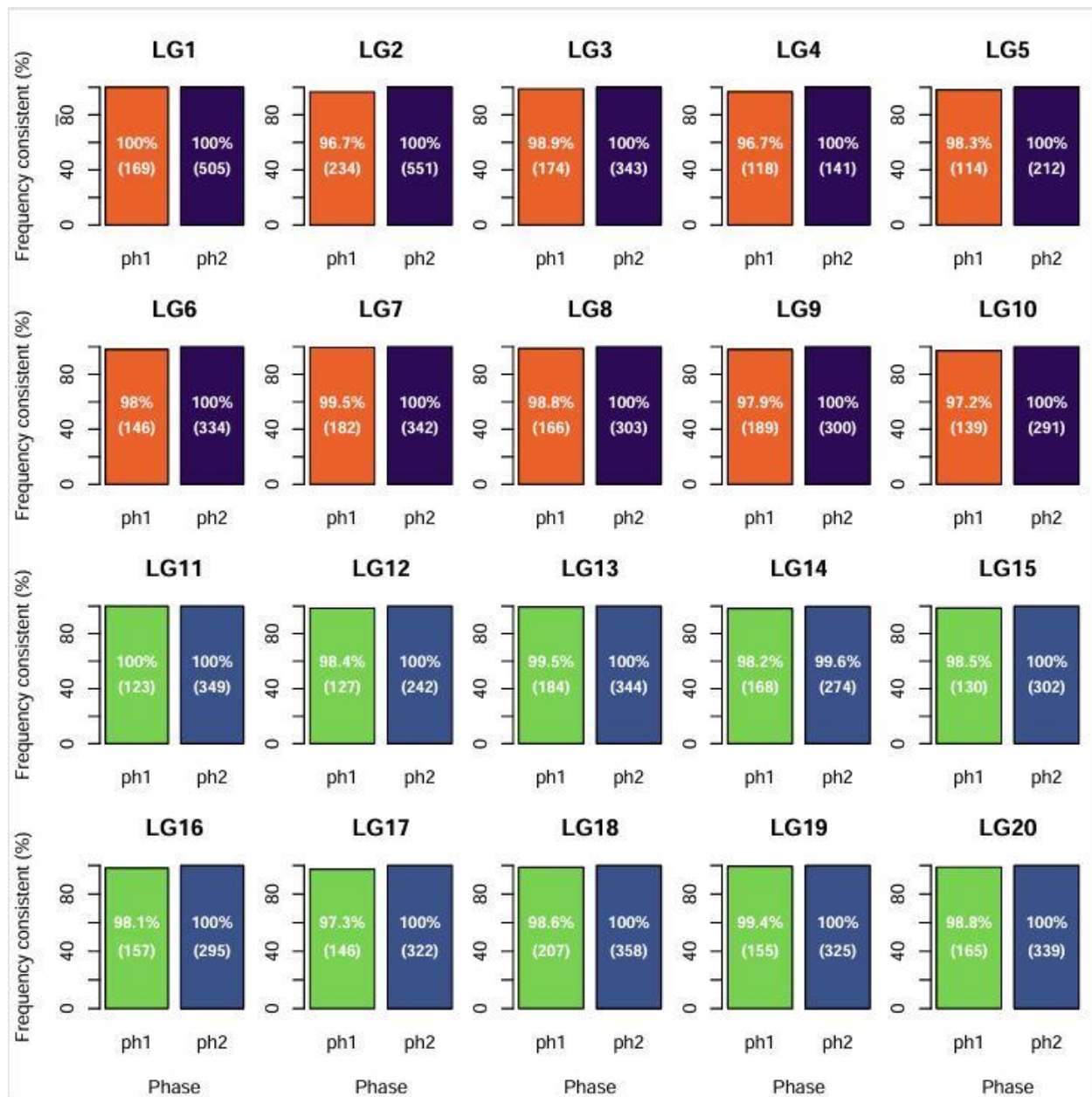
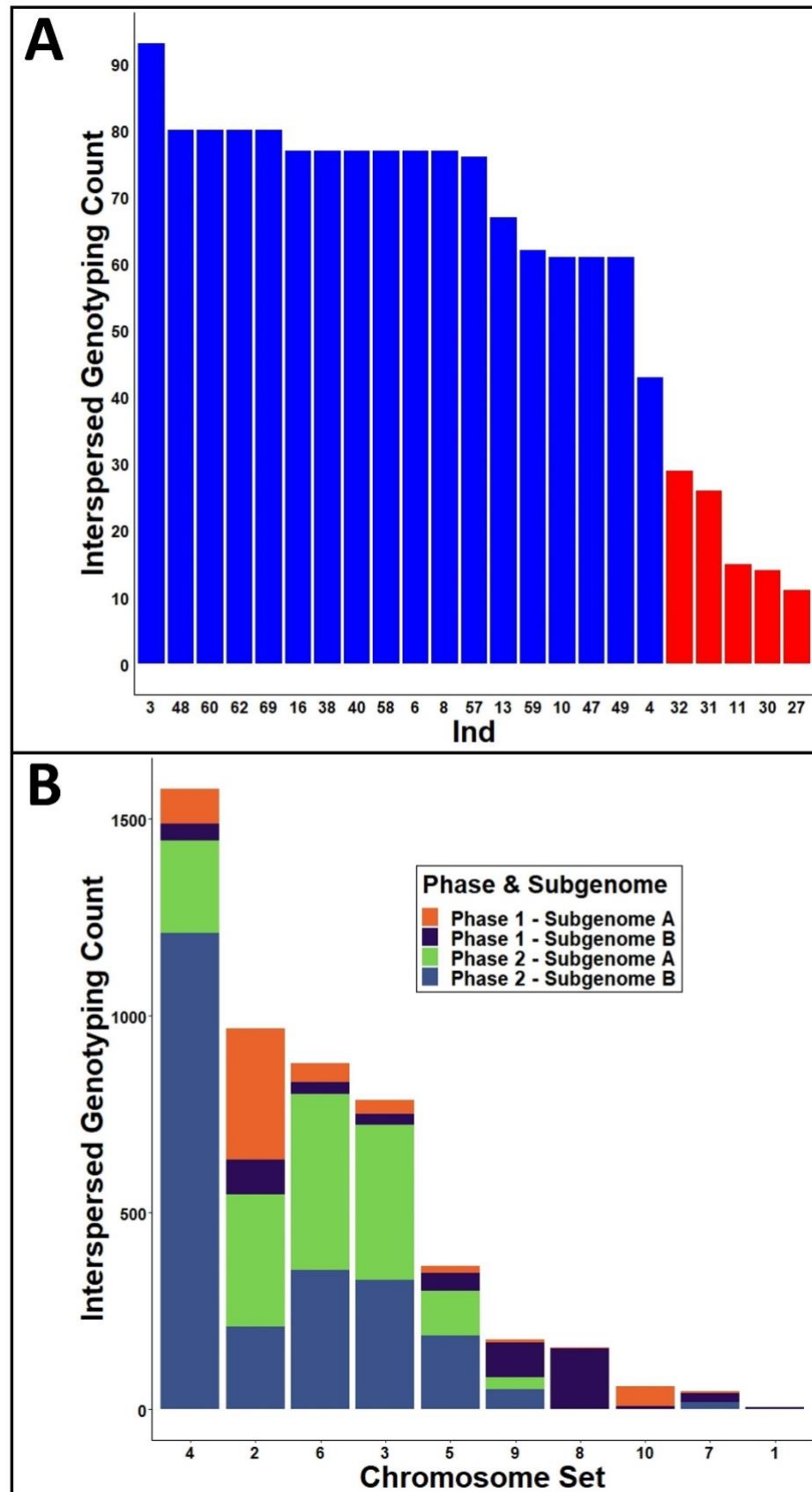


Figure S2: Bar plots of phasing consistency by linkage group (LG), where LGs 1–10 correspond to the A subgenome and LGs 11–20 correspond to the B subgenome. Phase 1 is shown in orange (A subgenome) or green (B subgenome), and phase 2 is shown in dark blue (A subgenome) or light blue (B subgenome). The y-axis indicates the percentage of markers whose phasing from genotype calling matches the linkage mapping phasing, with the total number of markers given in parentheses.



845 **Figure S3:** Distribution and genomic context of interspersed genotyping events detected in BC₁
846 progenies. In panel A, bar plot displaying 18 BC₁ individuals (Ind) with interspersed genotyping
847 events specifically on LG14 (corresponding to chromosome set 4 subgenome B) of phase 2.
848 Individuals with increased counts are shown in blue, while those with reduced counts are shown
849 in red. Progenies with no interspersed events are excluded. In panel B, stacked bar plot
850 summarizing total interspersed genotyping counts across all chromosome sets. Each bar is
851 subdivided by phase (1 or 2) and subgenome (A or B), as indicated in the legend.

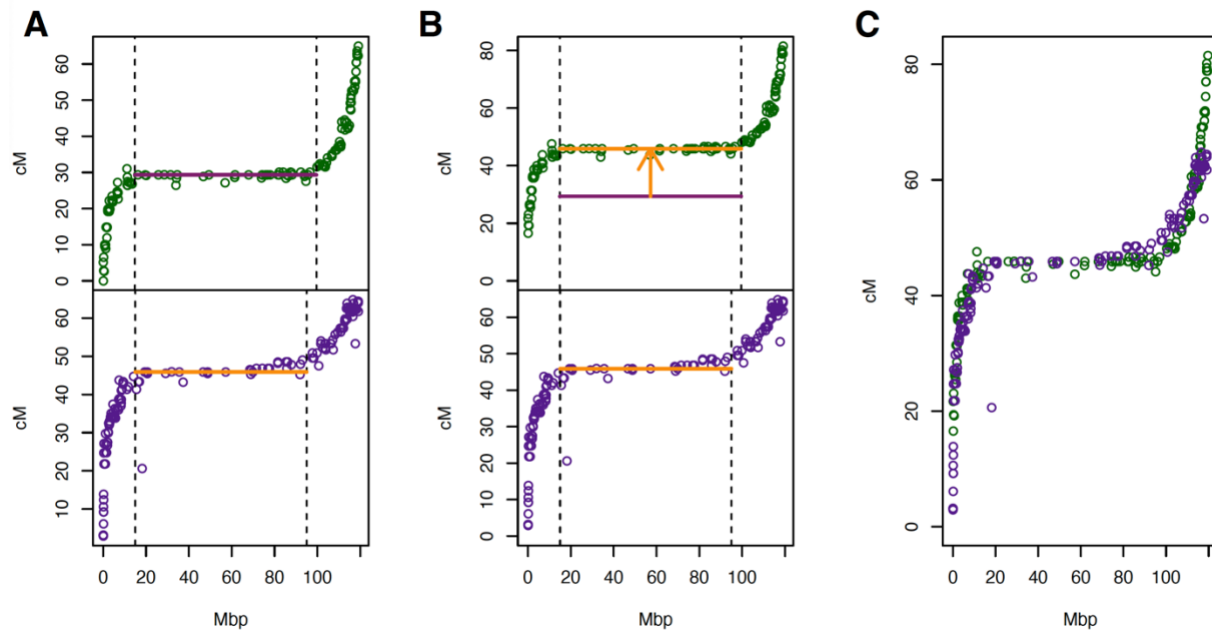


Figure S4: Steps in the generation of pseudo-integrated linkage map of chromosome 04. In panel A, centromeric bounds were visually identified (vertical dashed lines) and the genetic position of the centromeres were estimated as the median cM position of markers within this interval. Upper plot shows linkage group (LG) 4 (green points) while the lower plot shows LG14 (purple points). In panel B, markers on LG4 were shifted upwards by 16.5 cM, the distance between the two centromeric lines (denoted by arrow). Panel C shows the pseudo-integrated map with centromeres aligned (green points for LG4 markers and purple for LG14).

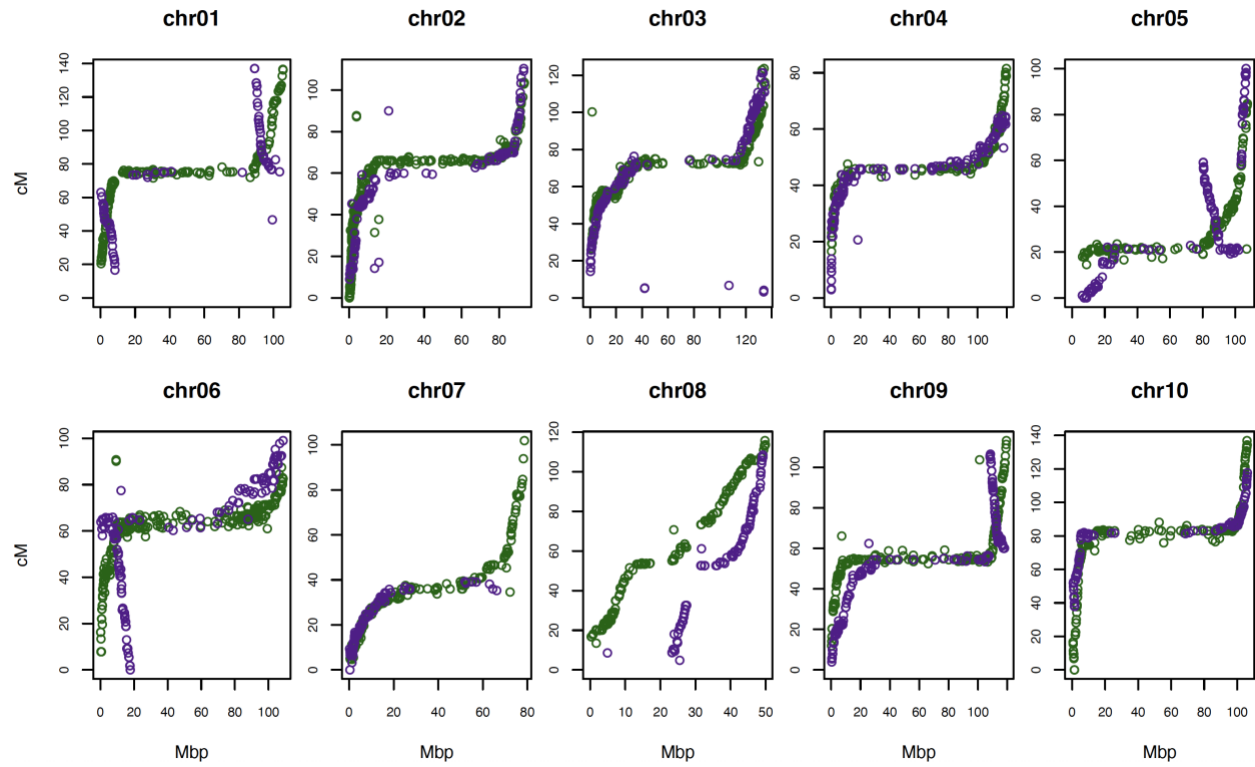
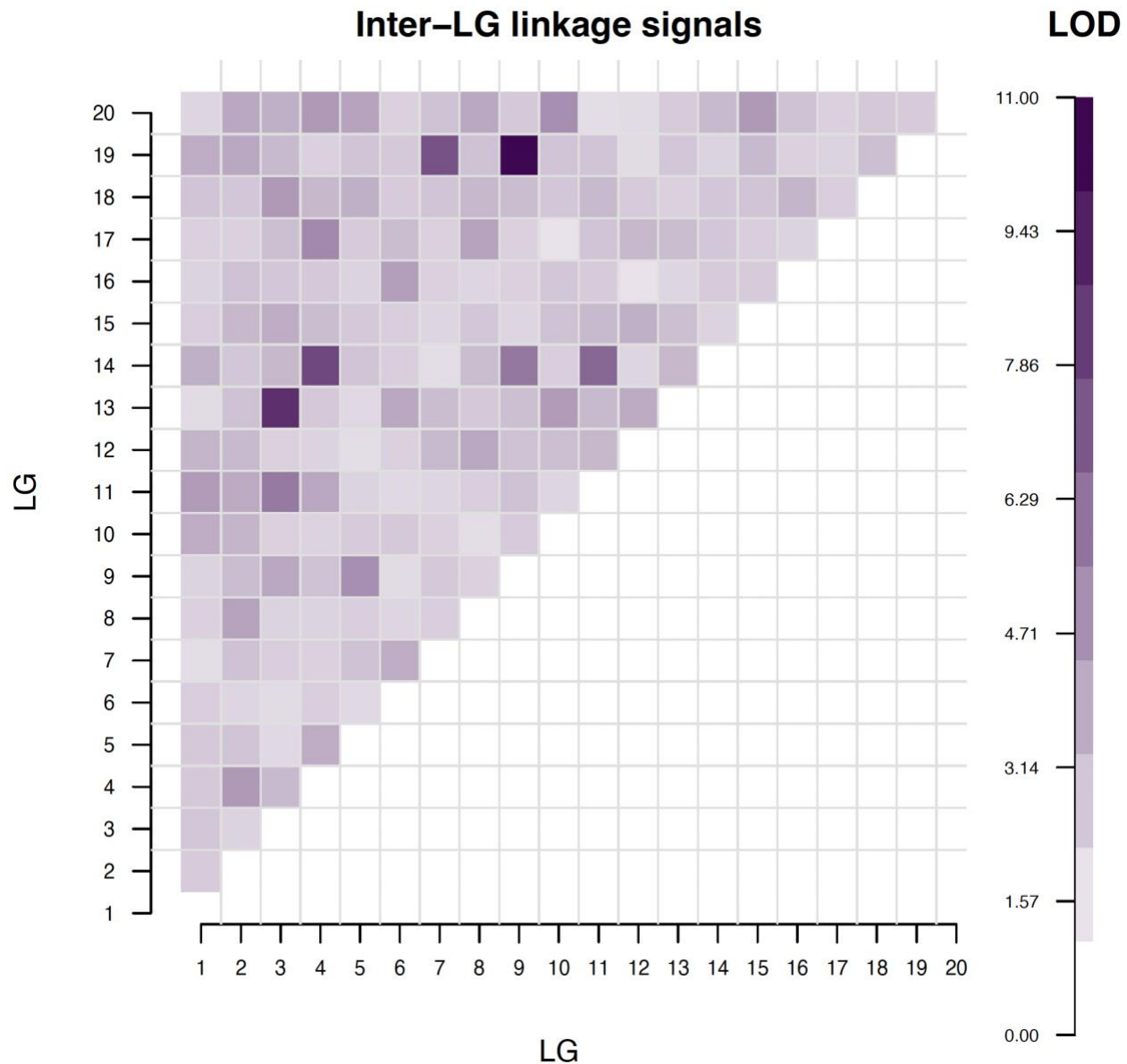


Figure S5: Pseudo-integrated linkage maps. Linkage group (LG) 1 – LG10 are shown in green, while LG11 – LG20 are shown in purple.

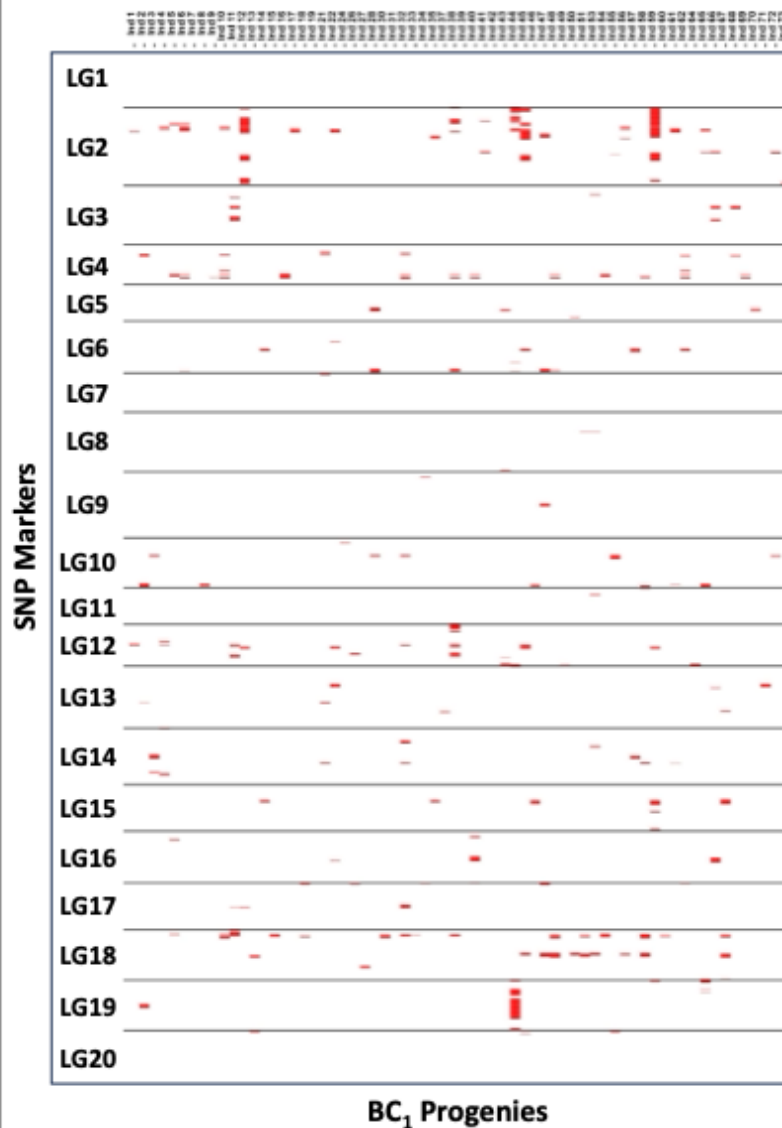


863

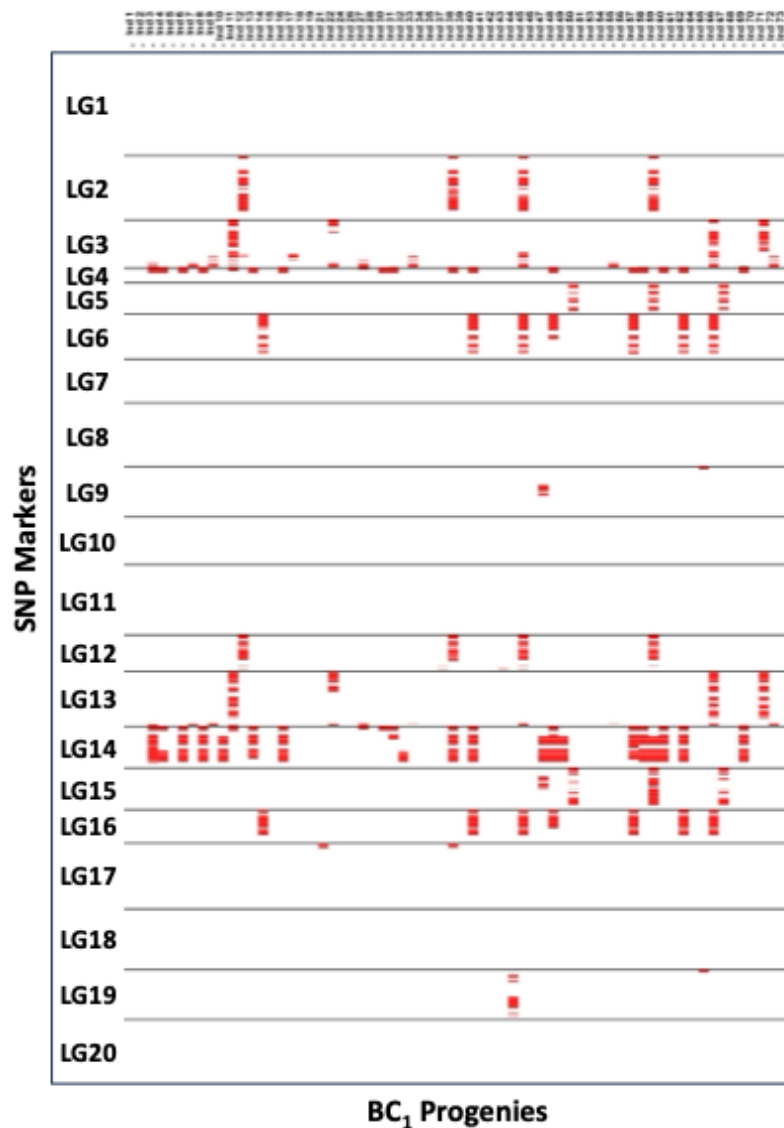
864 **Figure S6:** Heatmap of the maximum LOD for linkage detected between pairs of markers located
865 on different linkage groups (LGs). The strongest cross-LG signals were found between LG3 –
866 LG13 (chromosome 03), LG4 – LG14 (chromosome 04), and LG9 – LG19 (chromosome 09)
867 shown in darker shades of purple. The pairs of markers involved in generating these maximum
868 LOD scores were “well-linked” within their respective LGs, suggesting that the linkage signals

869 were not due to mis-assigned markers. In these three groups, multiple other cross-LG linkages at
870 elevated LOD scores were also detected.

A Interspersed Pattern in Phase 1 Linkage Map



B Interspersed Pattern in Phase 2 Linkage Map



872 **Figure S7:** Heatmaps displaying interspersed genotyping patterns (in red) in BC₁ progenies based on phase 1 (A) and phase 2 (B)
873 linkage maps. SNP markers are ordered by ‘Tifrunner’ physical distance. In panel A, phase 1 shows generally low levels of interspersed
874 patterns, with scattered noise and localized concentrations in linkage group (LG) 2, LG12, and LG19. In panel B, phase 2 exhibits a
875 markedly higher prevalence of interspersed genotyping, with pronounced patterns observed across homologous chromosome sets: LG2
876 and LG12, LG3 and LG13, LG4 and LG14, LG5 and LG15, LG6 and LG16, and LG9 and LG19.



877

878 **Figure S8:** BC₁ and BC₁F₂ derived from a cross between the neoallotetraploid [*Arachis magna* K
879 30097 x *A. stenosperma* V 15076]4x (MagSten) and 'IAC OL4' exhibited yellow flower
880 phenotypes.

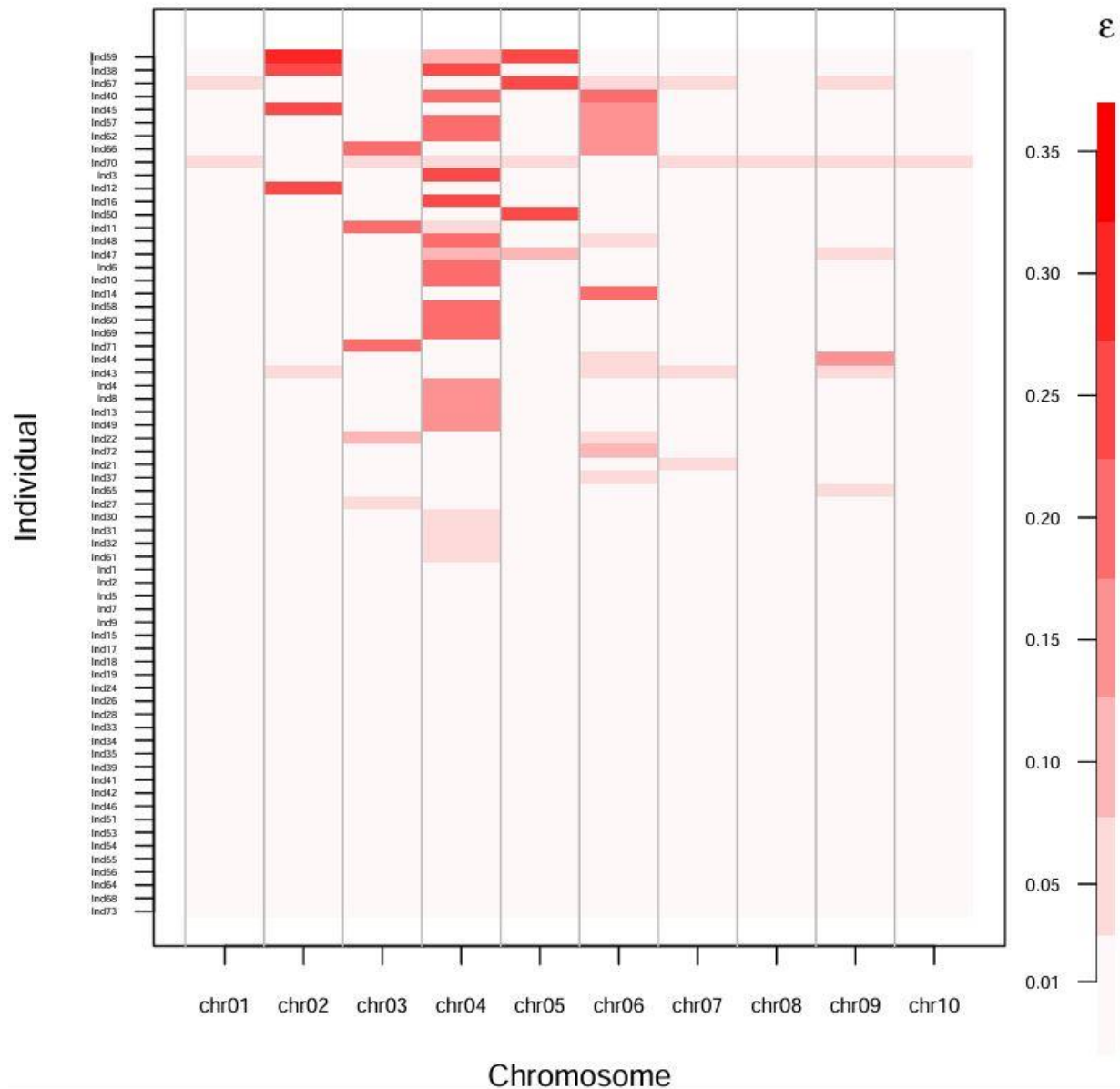
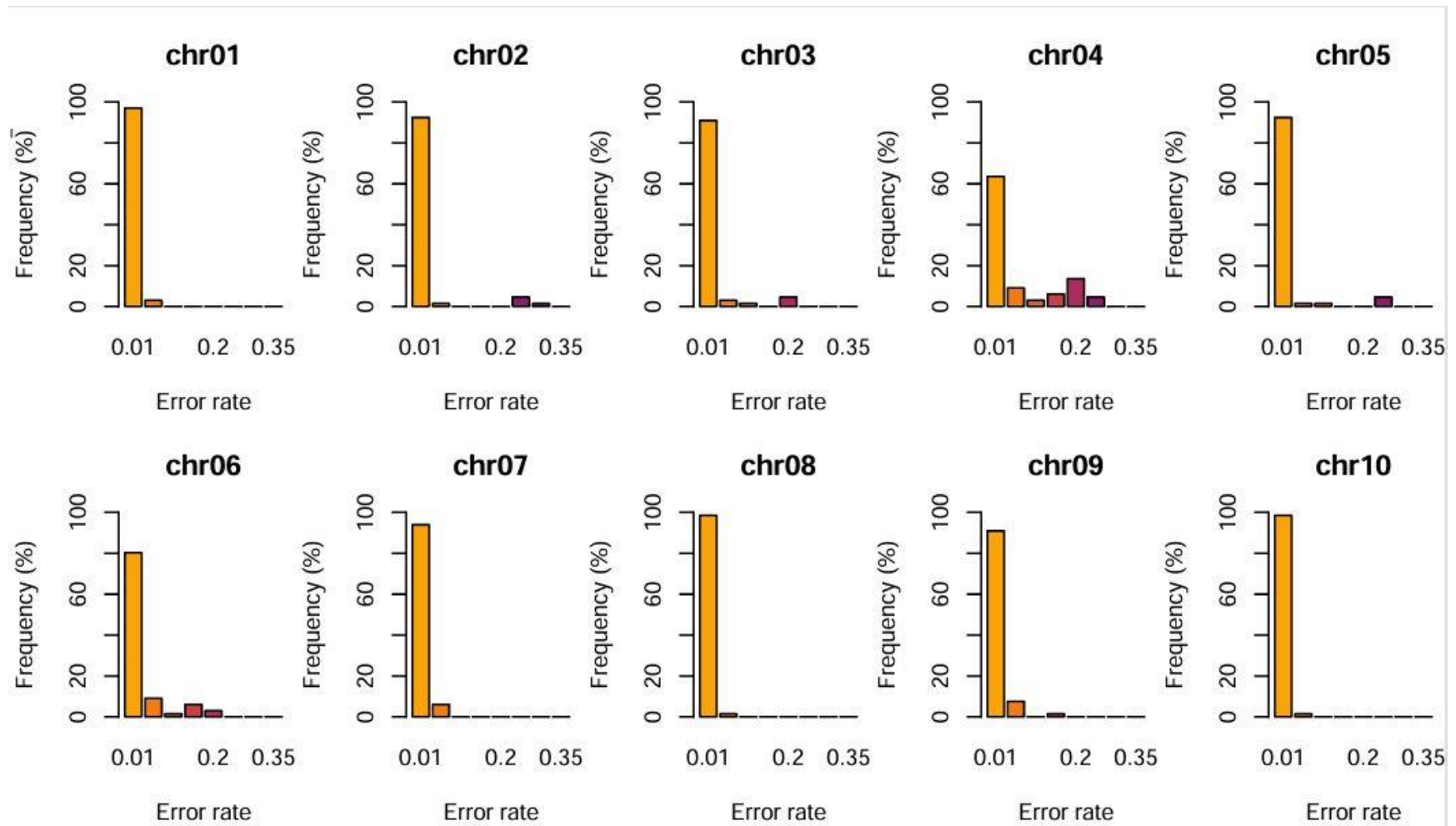


Figure S9: Heatmap showing the predicted genotyping error rates for each individual (y-axis) and chromosome (x-axis) combination, using a tetraploid HMM run over a range of error priors (ranging from 0.01 – 0.35 as shown in the color-bar on the right).



885

886 **Figure S10:** Distribution (histogram) of the predicted genotyping error rates per chromosome using a tetraploid HMM run over a range
887 of error priors (ranging from 0.01 – 0.35). The underlying data is the same as that used to generate Figure S9.

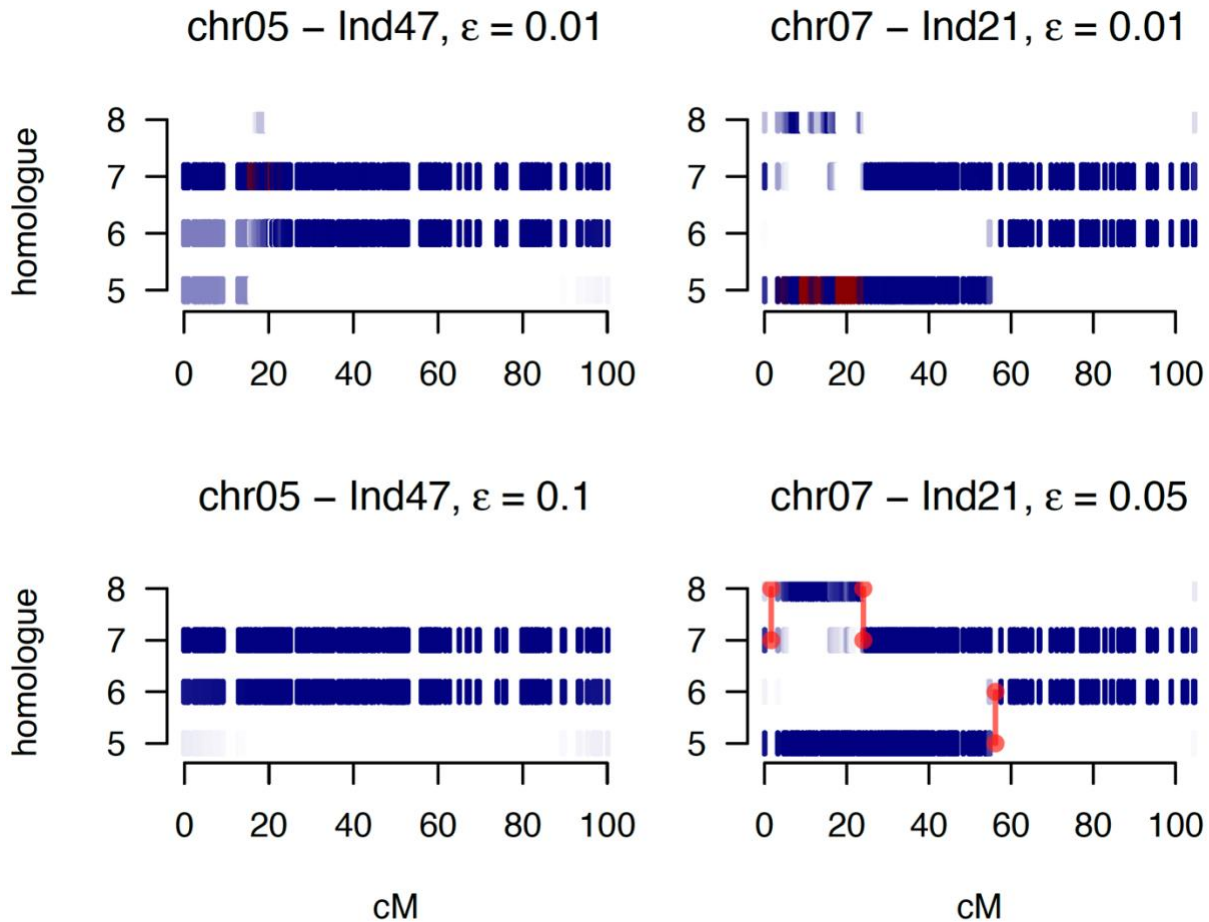
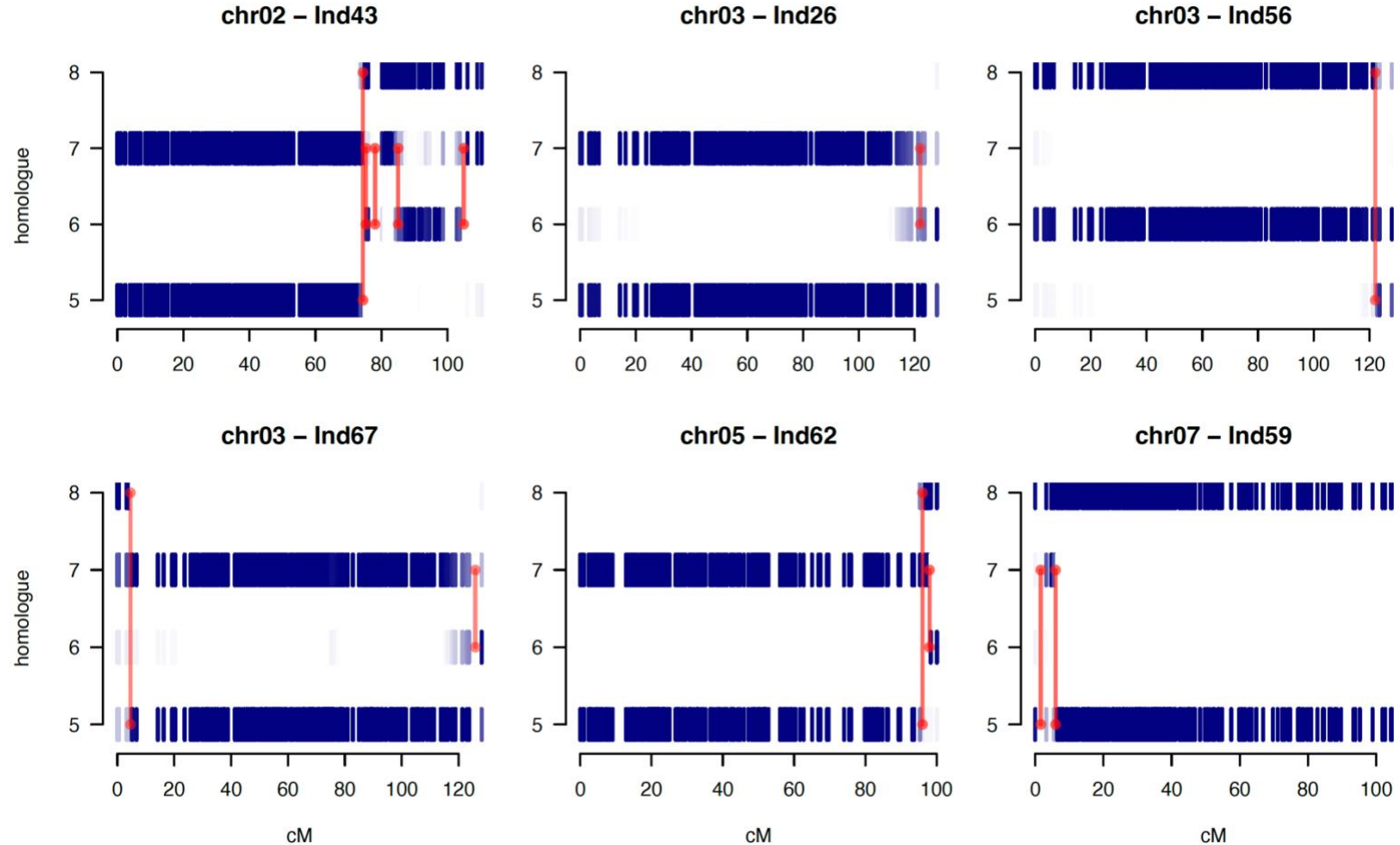


Figure S11: Visualization of posterior genotype (IBD) probabilities from a tetraploid Hidden Markov Model. Two individuals initially predicted to carry double reduction products (colored dark red segments, upper plots) at the default genotyping error prior of 0.01 were subsequently predicted to have arisen from bivalent pairing (within subgenomes) at error priors (0.1 and 0.05) that maximized the likelihood function. Homologue numbering 5 & 6 refers to phase 1 and 2 of LG5 (and LG7) while homologue 7 & 8 denote phase 1 and 2 of linkage group (LG) 15 (and LG17, respectively). Predicted recombination break-points are shown with red dots and lines in the bivalent-pairing case only.



897

898 **Figure S12:** Visualization of posterior genotype (IBD) probabilities for individuals with predicted pairing and recombination across
 899 subgenomes (*i.e.* homoeologous exchanges). Allotetraploid pairing is only expected between homologues 5 & 6 or 7 & 8 (see legend of
 900 previous Figure S11 for further explanatory details).