

# Genomic Outcomes of Haploid Induction Crosses in Potato (*Solanum tuberosum* L.)

Kirk R. Amundson,\* Benny Ordoñez,\*† Monica Santayana,† Ek Han Tan,\*‡ Isabelle M. Henry,\*

Elisa Mihovilovich,† Merideth Bonierbale,†,‡ and Luca Comai\*,‡

\*Plant Biology and Genome Center, University of California, Davis, California 95616, †International Potato Center (CIP), Lima 12, Peru, and ‡School of Biology and Ecology, University of Maine, Orono, Maine 04469

ORCID IDs: 0000-0003-3435-626X (K.R.A.); 0000-0003-4554-4354 (B.O.); 0000-0001-8872-7404 (E.H.T.); 0000-0002-6796-1119 (I.M.H.); 0000-0002-8349-9024 (E.M.); 0000-0002-5157-187X (M.B.); 0000-0003-2642-6619 (L.C.)

**ABSTRACT** The challenges of breeding autotetraploid potato (*Solanum tuberosum*) have motivated the development of alternative breeding strategies. A common approach is to obtain uniparental dihaploids from a tetraploid of interest through pollination with *S. tuberosum* Andigenum Group (formerly *S. phureja*) cultivars. The mechanism underlying haploid formation of these crosses is unclear, and questions regarding the frequency of paternal DNA transmission remain. Previous reports have described aneuploid and euploid progeny that, in some cases, displayed genetic markers from the haploid inducer (HI). Here, we surveyed a population of 167 presumed dihaploids for large-scale structural variation that would underlie chromosomal addition from the HI, and for small-scale introgression of genetic markers. In 19 progeny, we detected 10 of the 12 possible trisomies and, in all cases, demonstrated the noninducer parent origin of the additional chromosome. Deep sequencing indicated that occasional, short-tract signals appearing to be of HI origin were better explained as technical artifacts. Leveraging recurring copy number variation patterns, we documented subchromosomal dosage variation indicating segregation of polymorphic maternal haplotypes. Collectively, 52% of the assayed chromosomal loci were classified as dosage variable. Our findings help elucidate the genomic consequences of potato haploid induction and suggest that most potato dihaploids will be free of residual pollinator DNA.

**KEYWORDS** aneuploidy; copy number variation; genome elimination; SNP analysis

**H**IGHLY prized in plant breeding and research, haploid plants can be obtained through culture of immature gametophytes or, more conveniently, through inter- or intra-specific crosses in which the genome of one parent, the haploid inducer (HI), does not appear in the progeny (Forster *et al.* 2007; Ishii *et al.* 2016). Having been documented in 74 crosses between monocotyledonous species and 35 involving dicotyledonous species, this phenomenon is not uncommon [reviewed in Ishii *et al.* (2016)], and is widely exploited for the rapid generation of inbred lines, as well as genetic mapping and germplasm base expansion. However, the genetic

properties that make an HI are largely unknown, with a couple of exceptions. Artificial manipulation of centromeric histone H3 can result in an HI (Ravi and Chan 2010; Karimi-Ashtiyani *et al.* 2015; Kuppu *et al.* 2015; Maheshwari *et al.* 2015; Kelliher *et al.* 2016). Furthermore, natural maize HIs depend on inactivation of the phospholipase encoded by the matrilineal locus (Gilles *et al.* 2017; Kelliher *et al.* 2017; Liu *et al.* 2017).

In potato (*Solanum tuberosum* L.), the world's fourth most important crop in terms of calories consumed per person per day (<http://www.fao.org/faostat/en/#compare>), haploid seed can be routinely obtained via pollination with select HI varieties from the diploid *S. tuberosum* Andigenum Group [formerly *S. tuberosum* Phureja Group or *S. phureja* (Spooner *et al.* 2014)]. Such crosses with tetraploid potato ( $2n = 4x = 48$ ) produce  $2n = 2x = 24$  dihaploids that can be used for genetic mapping (Kotch *et al.* 1992; Pineda *et al.* 1993; Ercolano *et al.* 2004; Velásquez *et al.* 2007; Mihovilovich *et al.* 2014; Bartkiewicz *et al.* 2018). Additionally, these crosses produce hybrids that can be either triploid or tetraploid

Copyright © 2020 by the Genetics Society of America  
doi: <https://doi.org/10.1534/genetics.119.302843>

Manuscript received October 25, 2019; accepted for publication December 9, 2019; published Early Online December 23, 2019.

Supplemental material available at figshare: [https://figshare.com/projects/Genomic\\_Outcomes\\_of\\_Haploid\\_Induction\\_Crosses\\_in\\_Potato\\_Solanum\\_tuberosum\\_L\\_70793](https://figshare.com/projects/Genomic_Outcomes_of_Haploid_Induction_Crosses_in_Potato_Solanum_tuberosum_L_70793).

<sup>1</sup>Present Address: Duquesa Business Centre, P. O. Box 157, Manilva, Malaga, Spain 29692.

<sup>2</sup>Corresponding author: 451 E. Health Sciences Dr., University of California, Davis, CA 95616. E-mail: lcomai@ucdavis.edu

(Wagenvoort and Lange 1975; Hanneman and Ruhde 1978), and can be identified as seed because they express a purple embryo spot, a dominant anthocyanin marker encoded by the HIs that is expected to be absent in the dihaploids (Figure 1) (Hermsen and Verdenius 1973). In embryo spot-negative dihaploid populations, 3.5–11.0% of aneuploids are commonly found, exhibiting  $2n = 2x + 1 = 25$  and, rarely,  $2n = 2x + 2 = 26$  karyotypes (Wagenvoort and Lange 1975).

An ongoing question regarding haploid induction in potato is the cytogenetic mechanism by which it occurs. Two mechanisms have been proposed. The first mechanism is parthenogenesis, in which HI pollen triggers the development of unfertilized egg cells without making a genetic contribution to the embryo. This is supported by three lines of evidence: (i) endosperms from  $4x$  by  $2x$  potato haploid induction crosses are usually hexaploid instead of the expected pentaploid, suggesting abnormal pollen (von Wangenheim *et al.* 1960); (ii) HIs frequently produce 24-chromosome restitution sperm nuclei, thought to be a consequence of failed pollen mitosis II; and (iii) colchicine-treated pollen of noninducer *S. tarjinse* also exhibit restitution sperm nuclei and can induce haploids (Montelongo-Escobedo and Rowe 1969). Based on these observations, it was speculated that a  $2x$  restitution sperm fertilizes the central cell, leaving no sperm to fertilize the egg. The second mechanism is genome elimination, in which HI chromosomes are eliminated from the embryo after fertilization. This alternative hypothesis is supported by the presence of inducer-specific AFLP, RFLP, or isozyme markers in presumably dihaploid progeny. Often, progeny exhibiting genetic markers from the HI are also aneuploid, suggesting inheritance of an entire chromosome from the HI (Clulow *et al.* 1991, 1993; Waugh *et al.* 1992; Wilkinson *et al.* 1995; Allainguillaume *et al.* 1997; Clulow and Roussel-Bourgeois 1997; Straadt and Rasmussen 2003; Ercolano *et al.* 2004). These results are consistent with haploid induction crosses in maize (Riera-Lizarazu *et al.* 1996; Zhao *et al.* 2013), *Arabidopsis* (Kuppu *et al.* 2015; Maheshwari *et al.* 2015; Tan *et al.* 2015), and oat-maize hybrids (Riera-Lizarazu *et al.* 1996) in which one or more HI chromosomes persist in otherwise haploid plants.

Recently, widespread and ubiquitous introgression of very short DNA regions ( $> 100$  bp) from the HI genome into potato dihaploids has been reported by SNP genotyping (Bartkiewicz *et al.* 2018) and by whole-genome sequencing (Pham *et al.* 2019). In the latter case, depending on the progeny, 25,000–300,000 translocation events were inferred, suggesting a massive contribution from the transient HI genome to the maternally contributed genome. Genetic information in short segments of HI DNA could persist through three mechanisms: (i) nonhomologous recombination leading, for example, to insertion; (ii) homologous recombination leading, for example, to gene conversion; and (iii) autonomous replication. To clarify the underlying molecular arrangement, we use the term “addition” to indicate the presence in the dihaploid genome of an additional copy derived from the HI. We use the term “introgression” to indicate

that the DNA from the HI has recombined with the donor genome. If recombination is homologous, this could result in copy-neutral transfer of information. If confirmed, such widespread recombination would require rethinking of both breeding and biotechnology experimental strategies to either avoid or exploit it, depending on the context.

In light of these observations, we resequenced a population of 167 primary dihaploids derived from tetraploid Andigenum Group cultivar Alca Tarma to address three questions: First, does a curated set of phenotypically normal primary dihaploids display aneuploidy? If so, which parent contributes the additional chromosome(s)? Second, are single chromosomes or large chromosome segments from the HI added to otherwise dihaploid progeny? Third, are shorter segments of the HI genome introgressed or added to the dihaploids? If so, on what scale? We considered two hypotheses: first, that occasional failure to eliminate the entire HI chromosome set could result in the persistence of an additional chromosome, whole or fragmentary, in an otherwise dihaploid potato. We have previously demonstrated that events involving entire chromosomes or chromosome segments are readily detectable with low-coverage whole-genome sequencing, and chromosome dosage analyses, in *Arabidopsis* (Henry *et al.* 2010; Kuppu *et al.* 2015; Maheshwari *et al.* 2015; Tan *et al.* 2015) and poplar (Henry *et al.* 2015; Zinkgraf *et al.* 2016). In potato, extensive copy number variation (CNV), which has been described in numerous cytological and genomic surveys of diploid and tetraploid cytotypes, is a potentially confounding factor that should be taken into account (Iovene *et al.* 2013; de Boer *et al.* 2015; Hardigan *et al.* 2016; Hardigan *et al.* 2017; Pham *et al.* 2017). Second, smaller-scale introgression described could be detected by deeper sequencing of selected individuals. Our genomic analysis did reveal frequent whole-chromosome aneuploidy and widespread segmental dosage variation, but these were never attributable to the HI. Notwithstanding the ability of dihaploids to tolerate chromosomal dosage imbalance, we found no evidence that HIs contributed large chromosomal segments to the progeny. Further, by a set of standard criteria, we found no short segmental introgression either.

## Materials and Methods

### Plant material

A population of 167 putative dihaploids described in Velásquez *et al.* (2007) and Mihovilovich *et al.* (2014) was raised from the progeny of tetraploid Andigenum Group landrace cultivar Alca Tarma and one of two HI genotypes: IVP-101 or CIP596131.4. For convenience, we refer to CIP596131.4 as PL4 throughout. Both HIs are homozygous for a dominant embryo spot marker that results in anthocyanin accumulation at the base of the cotyledons visible through the seed coat. Only seeds lacking the embryo spot marker were planted, and seedlings exhibiting more than an average of eight chloroplasts per guard cell (minimum 10 measured

cells) were discarded. Seedlings were germinated on soil and maintained as *in vitro* cuttings thereafter.

#### **Genomic DNA library preparation, sequencing, and preprocessing**

Approximately 750 ng of genomic DNA extracted from leaf tissue as previously described (Ghislain *et al.*, 1999) was sheared to an average size of 300 bp using a Covaris E-220 sonicator in a 50- $\mu$ l reaction volume using the following settings: 175 W peak power, 10% duty factor, 200 cycles per burst, 50-sec treatment time, 4° minimum temperature, and 9° maximum temperature. Genomic libraries were constructed using 375 ng of sheared DNA and a KAPA Hyper Prep Kit (catalog number KK8504) with half-scale reactions, custom 8-bp dual-indexed adapters, and library amplification cycles as specified in Supplemental Material, Table S1. Libraries were sequenced on an Illumina HiSeq 4000 in either 50-nt single-end or 150-nt paired-end mode by the University of California, Davis DNA Technologies Core and Vincent Coates Genome Sequencing Laboratory.

Libraries were demultiplexed using a custom Python script available on our laboratory website (`allprep-12.py`; [http://comailab.genomecenter.ucdavis.edu/index.php/Barcode\\_data\\_preparation\\_tools](http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_data_preparation_tools)).

#### **Variant calling**

For paired-end sequencing, sequence reads were processed with Cutadapt (v.1.15) to remove low-quality (< Q10) bases, adapter sequences, and reads  $\leq$  40 nt after trimming. The DM1-3 v4.04 genome assembly, as well as DM1-3 chloroplast and mitochondrion sequences, were retrieved from ([http://solanaceae.plantbiology.msu.edu/pgsc\\_download.shtml](http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml)), concatenated, and used as the reference sequence for paired-end read alignment with BWA MEM (v.0.7.12-r1039) (Li 2013), with mismatch penalty six and all other parameters left at the program default. PCR duplicates were removed using Picard (v.2.14) MarkDuplicates, and only reads with mates mapping to the same chromosome were retained. For reads with overlapping mates, one of the two reads was soft-clipped in the overlap region using bamUtil::clipOverlap (Jun *et al.* 2015). Variants were called on processed alignment files using freebayes (v.1.1.0) (Garrison and Marth 2012) with minimum read mapping quality 41, minimum base quality 20, population priors not considered, and ploidy specified for each sample as a CNV map. To remove low-quality variants, the following site filters were applied in RStudio (v.3.4.0): NUMALT == 1, CIGAR == 1X, MQM  $\geq$  50, MQMR  $\geq$  50,  $|MQM - MQMR| < 10$ , RPPR  $\leq$  20, RPP  $\leq$  20, EPPR  $\leq$  20, EPP  $\leq$  20, SAP  $\leq$  20, SRP  $\leq$  20, and DP  $\leq$  344. Only sites with called homozygous Alca Tarma genotypes without reads matching HI alleles were retained. Sites that were called homozygous for the Alca Tarma allele in either HI were removed. Several additional quality filters were applied on each sample: depth  $\geq$  10 in all three parents,  $\leq$  5% Alca Tarma allele representation at called homozygous HI loci, and 40–60% Alca Tarma allele

representation at called heterozygous HI loci. After filtering, 798,468 SNPs were retained for analysis. Putative introgression loci were identified via heterozygous genotype calls in any of the three high-coverage dihaploids.

#### **Dosage analysis**

Single-end reads were aligned to the DM1-3 reference genome as described above, and only reads with mapping quality  $\geq$  Q10 were retained. Standardized coverage values were derived by taking the fraction of mapped reads that aligned to a given bin for that sample, dividing it by the corresponding fraction from the same bin for tetraploid Alca Tarma, and doubling the resulting value to indicate the expected diploid state. To mitigate mapping bias due to read type and length, Alca Tarma forward mates were hard trimmed to 50 nt and remapped. When all chromosomes were treated as equivalent, the distribution of per-chromosome standardized coverage values approximated a Gaussian distribution (quantile-quantile (Q-Q) plots not shown), satisfying the assumption of a Z-score analysis. Aneuploidy was then called if chromosomal standardized coverage exceeded the all-chromosome population by  $\geq$  3 SD. We identified local dosage variants by clustering standardized coverage values of nonoverlapping 250-kb bins using the R package MeanShift (CRAN - Package MeanShift). The clustering bandwidth parameter was set to the 50th percentile of interpoint distances in each 250-kb bin.

Parental origin analyses of trisomy and dosage variants were carried out as previously described (Henry *et al.* 2015). Briefly, reads with mapping quality  $\geq$  Q20 and base calls  $\geq$  Q20 were used to compute allele-specific read depth at the subset of 800,384 SNP loci identified above, located on chromosomes 1–12 of the DM1-3 v4.04 assembly. The percentage of reads supporting the HI allele reads among all reads at loci within a nonoverlapping 1-Mb bin was then reported. As a biological positive control was not available for SNP analysis, we empirically evaluated limitations of the SNP dosage assay by comparing dihaploids with simulated triploid hybrids expected to resemble an introgressed HI chromosome segment at all tested genomic loci. To construct triploid hybrids *in silico*, pseudorandom subsets of exactly 2,015,413 and 1,007,706 forward mates were drawn 100 times from raw sequencing reads of Alca Tarma (SRA ID SRR6123032) and IVP101 (SRA ID SRR6123183), respectively. The number of parental reads was chosen such that parental reads would be present in a 2:1 ratio expected for a triploid, and that the number of raw reads for each *in silico* hybrid would match the fifth percentile of raw read counts in the dihaploid sequencing data set. Similarly, 100 *in silico* hybrids were constructed using Alca Tarma and PL4 reads. As a negative control, pseudorandom subsets of 3,023,119 reads were drawn 100 times from Alca Tarma. Raw reads from all simulated hybrids were hard trimmed to 50 nt, and then processed using the SNP dosage pipeline described above with nonoverlapping 1-Mb bins. For each bin, if the ranges of HI allele percentages from either *in silico* hybrid group

overlapped with the corresponding range of the negative control, the bin was withheld from analysis. Using this approach, 608 of 730 bins (83%) were considered in this analysis. Similarly, to determine whether unique dosage variants could be genotyped confidently, we compared the distributions of % HI allele values between simulated hybrid and negative control groups in the affected interval.

To estimate absolute copy number, per-position read was calculated using samtools depth, using only reads with mapping quality  $\geq$  Q20. Median read depth in nonoverlapping 10-kb windows was then determined using custom Python software available from ([https://github.com/kramundson/LOP\\_manuscript](https://github.com/kramundson/LOP_manuscript)). We observed a positive correlation of window median read depth and GC content, suggesting that PCR amplification bias was introduced during sequencing library construction (Benjamini and Speed 2012). For each 10-kb bin, this bias was corrected by dividing median read depth by GC content for that bin. The resulting values were clustered using the MeanShift package in R Ciollaro and Wang (2016). The centroid of the largest cluster was designated as the copy number corresponding to the expected ploidy (four for Alca Tarma and two for dihaploids), and multiples of the centroid were used to designate the remaining copy number states.

### Cytological analysis

Chromosome spreads were prepared from root tips as previously described (Watanabe and Orrillo 1993), with minor modifications. Commercial permethrin was used at a concentration of 75 ppm as a pretreatment to induce chromosome condensation. Roots were kept in an ice-cold water bath for 24 hr before hydrolysis in 1M HCl for 10–15 min, then stained with lacto-propionic acid and squashed.

### Data availability

Sequence data have been deposited in the National Center for Biotechnology Information Sequence Read Archive: Bio-Project identifier PRJNA408137. Analysis code has been deposited at [https://github.com/kramundson/LOP\\_manuscript](https://github.com/kramundson/LOP_manuscript). Supplemental material available at figshare: [https://figshare.com/projects/Genomic\\_Outcomes\\_of\\_Haploid\\_Induction\\_Crosses\\_in\\_Potato\\_Solanum\\_tuberosum\\_L\\_70793](https://figshare.com/projects/Genomic_Outcomes_of_Haploid_Induction_Crosses_in_Potato_Solanum_tuberosum_L_70793).

## Results

### Induction, selection, and sequencing of dihaploids

A population of 167 primary dihaploids was generated from Alca Tarma via pollination with HIs IVP101 or PL4 (Velásquez *et al.* 2007; Mihovilovich *et al.* 2014) over the course of two previous studies (Velásquez *et al.* 2007; Mihovilovich *et al.* 2014). The dihaploids in this population lacked the homozygous dominant embryo spot seed marker present in both HIs, and displayed the expected count of guard cell chloroplast counts and root cell chromosomes (Velásquez *et al.* 2007; Mihovilovich *et al.* 2014) (Figure 1). We carried out genome resequencing to identify dosage variation and aneuploidy among the population. Alca Tarma, IVP101, PL4, and three

selected haploids were sequenced to 40–56 $\times$  coverage. For the remaining dihaploids, we generated an average of 3.88 million reads per individual (Table S1).

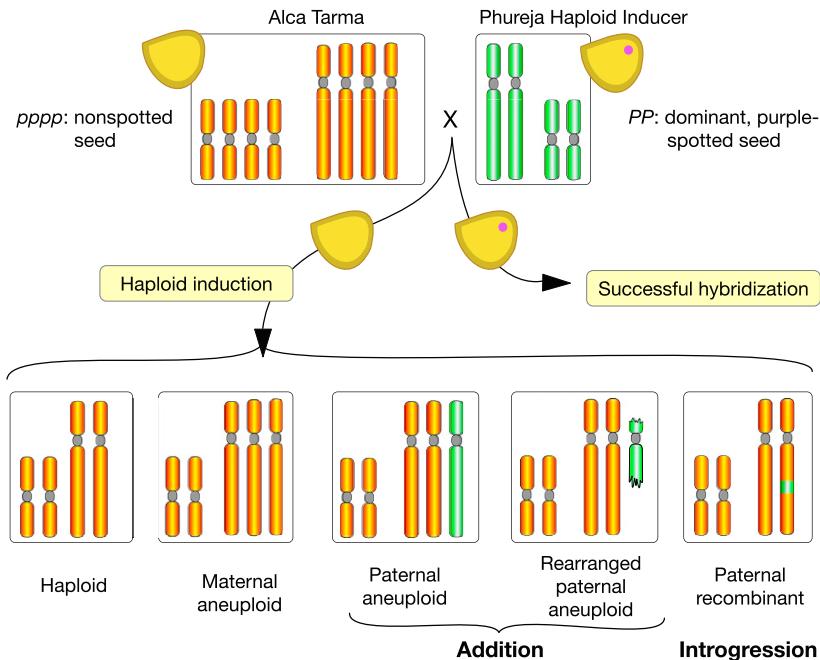
### Maternally derived trisomy

We hypothesized that introgressions into the host genome could derive from at least three types of events, each associated with specific predictions: (i) nonhomologous transposition of HI segments to the host genome, resulting in three copies of the corresponding region with a SNP ratio of 1:2 HI:host; (ii) homologous recombination leading to replacement of a segment, either interstitial or terminal, resulting in no copy number change of the affected region and a SNP ratio of 1:1 HI:host; and (iii) gene conversion from a noncrossover event producing a very short (25–50-bp) conversion tract and resulting mostly in a single SNP with a 1:1 ratio. Formally, homologous recombination could also result in duplication and resemble case (i) (Das *et al.* 1991).

We first screened the population for whole-chromosome aneuploidy. Sequencing reads were aligned to the DM1-3 reference genome and read counts per chromosome were normalized to those of the tetraploid parent, such that values near 2.0 were obtained for chromosomes present in two copies and values deviating from 2.0 indicated aneuploidy. In 19 individuals, standardized coverage was significantly elevated for a single chromosome, suggesting a primary trisomy (Figure 2A). Root tip chromosome spreads were evaluated for 15 of the 19 putative trisomics, confirming the  $2n = 2x + 1 = 25$  karyotype in all cases (Figure 2, B and C and Figure S1). In total, six trisomics of chromosome 2, two of chromosomes 4, 5, 7, and 8, and one of chromosomes 1, 3, 6, 10, and 12 were detected in the population (Figure 2A and Figure S1). To determine the parental origin of each trisomy, 382,967 SNPs homozygous for the same allele in both HIs, but homozygous for an alternate allele in Alca Tarma, were identified from sequencing the parental genomes (Supplemental Data set S1). The fraction of HI-specific allele calls along all chromosomes was then calculated for each trisomic individual. Using this measurement, a trisomy from either HI was expected to exhibit  $\sim$ 33% HI allele across the affected chromosome. To empirically validate this expectation, we evaluated 200 simulated low-coverage hybrids each consisting of 2 million reads from Alca Tarma, and 1 million reads from either IVP101 or PL4 (see *Materials and Methods*). In each trisomic dihaploid, HI alleles were nearly absent from the trisomic chromosome (Figure 2D), indicating inheritance from the noninducer parent.

### Determining which SNP bins are informative

To survey the population for HI chromosome addition or introgression, the SNP dosage analysis described above was repeated using nonoverlapping 1-Mb bins. To account for a low density of homozygous parental SNP markers in some regions, we included an additional set of SNPs that did not fit the optimal criteria of the original set (Supplemental Data set S2). Of the added SNPs, 170,273 were heterozygous in



one HI and homozygous in the other, while 247,144 were heterozygous in both HIs. A consequence of including these additional SNP markers is that an HI allele contribution could be lower than the expected 33%. Therefore, we empirically determined the expected percentages for each bin by comparing the percentages obtained from low-coverage *in silico* triploid hybrids and negative controls. Any bin in which the distributions of observed HI allele percentages of the hybrid and negative control groups exhibited any overlap was withheld from consideration. To call an introgression, we required at least three adjacent bins to exhibit an HI allele percentage that overlapped with the empirical thresholds determined from the *in silico* hybrid analysis. Among 1-Mb bins that were considered in this analysis, no such events were found (Figure S2). Notably, a recombination event that substitutes an Alca Tarma chromosome segment by the corresponding segment from either HI would produce a higher HI allele percentage than an addition event (50 vs. 33%), suggesting that neither addition nor introgression of an HI chromosome segment occurred in the dihaploid population.

#### CNV analysis

To complement the approach described above, we also asked whether rare structural variants existed in the population and, if so, from which parent they were derived. The dosage analysis was repeated as described above for nonoverlapping 250-kb bins of the reference genome. As before, a value of 2 indicates the expected diploid complement and values deviating from 2 indicate structural variation. Inferred karyotypes of a representative  $2n = 2x = 24$  dihaploid and a  $2n = 2x + 1 = 25$  maternal trisomy are shown in Figure 2, C and D, respectively. By overlaying dosage plots for each dihaploid, it became evident that many CNVs are recurrent in the population, with copy number gains and losses of the same locus

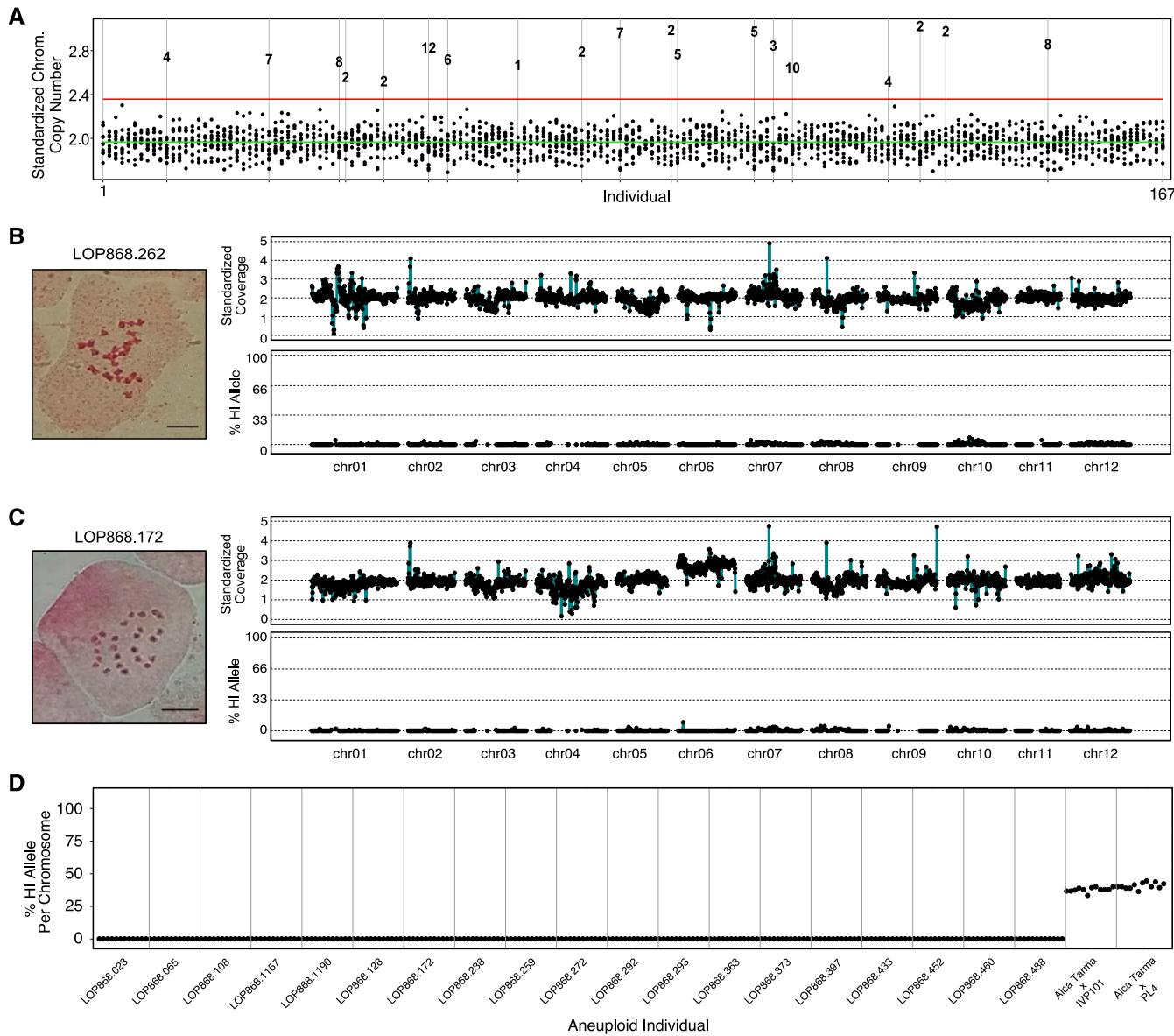
**Figure 1** Production of the dihaploid population and expected types. Haploid inducers IVP101 or PL4, both diploids, were used to pollinate tetraploid cultivar Alca Tarma. For simplicity, the potato genome is represented by two chromosome types. The haploid inducer is homozygous for the dominant seed purple spot marker. Normal fertilization and development resulting in hybrids with spotted seed, defined as successful hybridization, results in biparental triploids or tetraploids, depending on the ploidy of the male gamete. Maternal dihaploids are expected among plants germinated from nonspotted seeds. Plants displaying more than eight stomatic chloroplasts (an indication of increased nuclear content) or unusual phenotypes potentially consistent with aneuploidy were discarded (Velásquez *et al.* 2007; Mihovilovich *et al.* 2014). Genetic haploid inducers can act through either parthenogenesis (Forster *et al.* 2007) (development of an unfertilized egg) or genome elimination (Ishii *et al.* 2016) (rejection of the haploid inducer genome). Addition or introgression of residual haploid inducer DNA indicates the second mode of action.

among the dihaploids (Figure 3, A and B and Figure S3). This behavior is evident in the pericentromeric heterochromatic region of all chromosomes, while the euchromatic, gene-rich arms are more uniform (Figure 3, A–G). Chromosomes 2 and 4 are exceptions in that they display recurring dosage variation in their short arms (Figure 3B and Figure S3).

To define structurally polymorphic loci and alleles at these loci, we clustered relative coverage values separately for each 250-kb bin (Figure S4). Structural variation was widespread, with multiple clusters detected in 48% of 250-kb bins (Figure 3G). From joint consideration of the number of individuals in a cluster and read depth of high-coverage samples, we inferred that most dosage variants represented segregating polymorphism among Alca Tarma haplotypes, as exemplified by a 1-Mb block of chromosome 6 that is present on one chromosome and absent on the other three (Figure 4 and Figure S5). Among 14 duplications that were  $\geq 750$  kb and present in  $< 5\%$  of dihaploids, 5 were clearly derived from Alca Tarma (Table S2). Based on comparison of simulated hybrids, SNP marker density was too low to conclusively resolve the parental origin of the remaining 9, but the fraction of HI SNPs in power analyses classified them as low-probability outliers (Figure S6).

#### SNP loci consistent with addition or introgression were rare, and dispersed

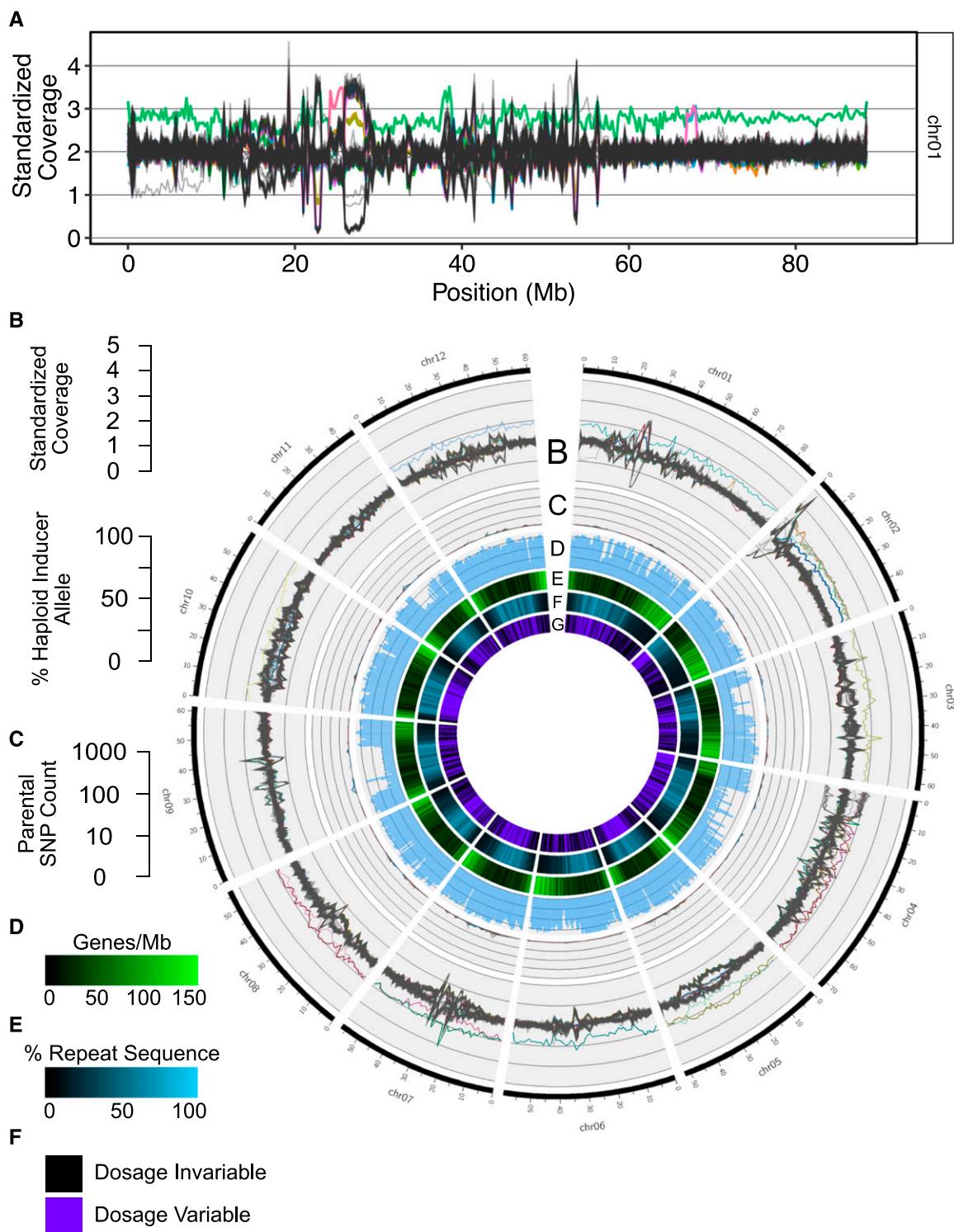
To search for addition and introgression events at higher resolution, three dihaploids were sequenced to 19–30 $\times$  depth with Illumina paired-end sequencing. To minimize spurious genotype calls, SNP loci that passed our quality filtering steps were further filtered to exclude sites with even a single Alca Tarma read that matched an HI allele. In total, 800,384 loci were assayed, with 725,952–745,535 loci assayed in each of the three dihaploids. The fraction of loci



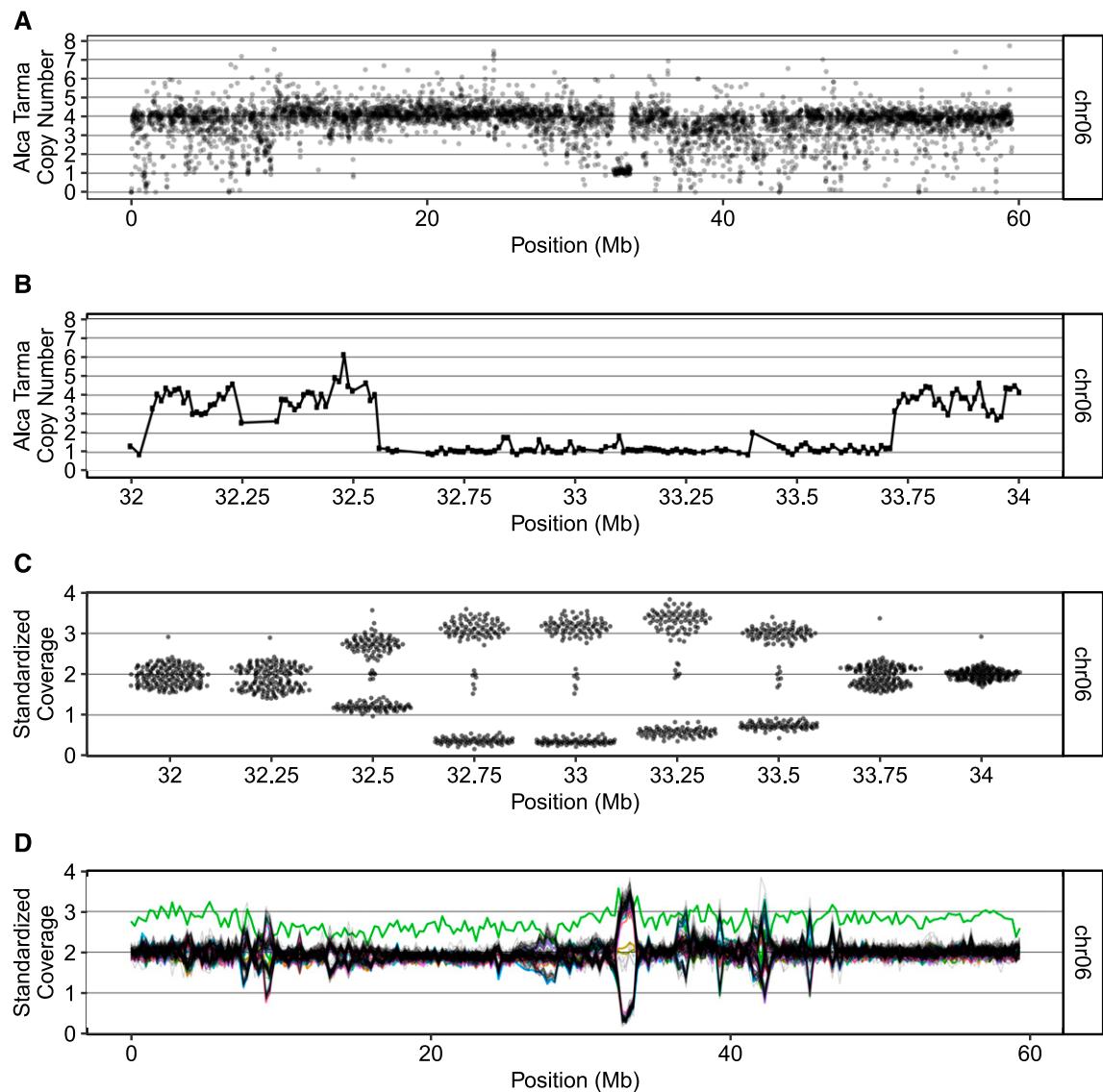
**Figure 2** Aneuploidy detection in 167 primary dihaploids by genome sequencing. (A) Chromosome copy number for each individual in the population. Each individual of the population is displayed along the  $x$ -axis, with the stack of dots at each  $x$  coordinate corresponding to standardized copy number for each chromosome type, with a value of 2.0 representing the expected diploid state. The green line corresponds to the mean chromosome copy number among the population and the red line indicates 3 standard deviations greater than the mean. Outliers in this distribution correspond to the affected chromosome in each trisomic and are numbered according to the chromosome present in excess. (B) Cytogenetic and *in silico* karyotype of a representative euploid dihaploid LOP868.262. Left: Root tip somatic metaphase karyotype. Top right: copy number plot; individual data points represent read depth in nonoverlapping 250-kb bins standardized to tetraploid Alca Tarma counts such that the expected diploid state corresponds to copy number 2.0. Bottom right: haploid inducer SNP allele plot; black points correspond to the fraction of base calls supporting haploid inducer alleles at all informative sites in nonoverlapping 1-Mb bins. (C) Cytogenetic and *in silico* karyotype of chromosome 6 trisomic LOP868.172, illustrating that trisomy of chromosome 6 was not derived from either IVP101 or PL4. See Figure S1 for dosage plots of the remaining trisomics. (D) SNP plot showing the percentage of haploid inducer allele present in the trisomics identified in panel (A). For each individual, the 12 points correspond to the 12 chromosomes displayed in order (*i.e.*, the first dot is chromosome 1, the second chromosome 2, etc.). For each individual, points near 0% for the affected chromosome and all others indicate a maternal trisomy. Observed % haploid inducer allele values from two representative simulated hybrid controls, one Alca Tarma  $\times$  IVP101 and the other Alca Tarma  $\times$  PL4, are shown on the right. Chrom., chromosome; HI, haploid inducer.

with heterozygous genotypes consistent with addition or introgression was very low (0.157–0.195%). Among heterozygous sites in the dihaploids, HI alleles tended to be underrepresented relative to expectations for either addition (~33%) or introgression (50%) of HI DNA (Figure S7), and read depth was lower in

both Alca Tarma and the dihaploid at hand (Figure S8). Upon observing the low read depth and underrepresentation of HI alleles at many putative introgression loci, we applied additional filters (minimum HI allele depth  $\geq 5$  and minimum allele depth representation 10%) to investigate the subset of possible



**Figure 3** Chromosomal distributions of dosage variation and parent-informative markers in the dihaploid population. (A) Read depth standardized to tetraploid Alca Tarma in nonoverlapping 250-kb bins of chromosome 1. Each individual in the dihaploid population is represented by a single line. Dihaploids that are trisomic for any chromosome are represented by colored lines. For example, chromosome 1 trisomic LOP868.238 is displayed as a green line. The plot also displays unique segmental dosage variants, one starting at 24 Mb and involving a trisomic of eight (pink line), a second one at 67 Mb and involving two trisomics of two (overlapping magenta and blue lines), and a terminal deletion of the short arm in an otherwise euploid line (gray line). (B–F) Circos plots for the 12 chromosomes. For each chromosome, 167 lines are shown. Aneuploids of any chromosome are uniquely colored and all other individuals are colored gray. (B) Standardized read-depth plots of all chromosomes smoothed to 1-Mb bins. (C) Percent haploid inducer allele at all parent-informative marker loci in nonoverlapping 1-Mb bins. Refer to Figure S3 for expected haploid inducer allele percentages derived from simulated hybrid analyses. (D) Log<sub>10</sub>-scaled counts of parent-informative markers in nonoverlapping 1-Mb bins. (E) Genes per 1-Mb sliding window and 200-kb step. Chr, chromosome.



**Figure 4** Example of segregating copy number variation among dihaploid population. (A) GC-normalized median read depth of nonoverlapping 10-kb bins for tetraploid *Alca Tarma* (see *Materials and Methods*). Chromosome 6 is shown. Each dot is plotted with partial transparency to emphasize overplotting. (B) GC-normalized median read depth, zoomed in on 1-Mb region of chromosome 6. (C) Swarm plots of standardized coverage values in bins affected by 1-Mb deletion. The population segregates the high- and low-dosage states in an ~1:1 ratio, consistent with random chromosome segregation of a deletion in triplex allele dosage. (D) Population standardized values for chromosome 6. Each dihaploid is displayed as a single contiguous line; green line: chromosome 6 trisomic. The region corresponding to the large deletion shown in panel (A) exhibits segregating dosage variation in the dihaploid population. Chr, chromosome.

introgression events with the best read support in our data set. Applying these filters reduced the number of putative introgression loci from 1217 to 266 in LOP868.004, from 1457 to 358 in LOP868.064, and from 1124 to 198 in LOP868.305 (Supplemental Data set S3).

We further investigated the distribution of all putative introgression loci with respect to the reference genome. As loci that were heterozygous in the two HIs were also included in this analysis and the phase was unknown, introgression of an HI segment may appear discontinuous. Therefore, we estimated a lower bound of the number of introgression events using a seed-and-extend approach: starting at SNP loci with inducer-specific allele introgression events, putative

introgressions were extended in both directions until a parent-homozygous SNP locus with no evidence of HI alleles in the dihaploid was encountered. The total number of SNP markers, as well as markers exhibiting HI alleles, were tallied for each event. Dihaploids LOP868.004, LOP868.064, and LOP868.305 exhibited 1037, 1191, and 806 events, respectively. Approximately one-half (49.6%) of putative introgression events consisted of a singleton SNP, *i.e.*, a single locus exhibiting HI alleles flanked by homozygous parental loci that did not support introgression in the dihaploid (mean distance between loci flanking a singleton = 9110 bp and median distance = 735 bp). Among nonsingleton events, very few markers exhibited HI alleles for each event (Figure S9).

Finally, read alignments were manually inspected for the introgression event with the highest number of parental SNP loci exhibiting HI alleles. This event spans a ~2-Mb region of chromosome 5 in dihaploid LOP868.305, including 5712 parental SNP loci, of which only 43 exhibited HI alleles. Specifically, we looked for phased variants on the same read consistent with introgression of a contiguous HI haplotype. Remarkably, no reads supporting HI introgression at each of these 43 loci matched a local haplotype present in either HI (Figure S10 and Supplemental Data set 4).

Taken together, these analyses revealed rare maternally associated chromosome remodeling among a backdrop of widespread structural heterogeneity (Figure 3A), including several large and novel variants detected in the genome of the tetraploid parent. Robust evidence for chromosomal introgression from the HI was not detected in any case, and a detailed survey of three dihaploid genomes revealed sites that, while superficially consistent with introgression of HI DNA, resemble sequencing or alignment artifacts.

## Discussion

We analyzed the genomes of 167 primary dihaploids produced by pollinating the *S. tuberosum* Andigenum Group cultivar Alca Tarma with HIs IVP101 and PL4. This population is representative of a typical dihaploid progeny set used for breeding in that, during its development, selection has been applied against individuals with DNA content differing from the dihaploid state, against individuals carrying the genetic color marker from the HI (Figure 1), and against severe abnormality. Using low-pass sequencing, we derived karyotypes of each progeny, identifying primary trisomy in 11.4% of individuals. By comparing parental SNP genotypes, we established that whole-chromosome aneuploidy was maternally inherited. Widespread variation in DNA dosage consistent with segregation of maternal structural variation was already evident at 1-Mb resolution (Figure 3). This analysis does not consider small structural events at the genic or transposable element scales. Using randomly downsampled data to simulate a triploid hybrid, we empirically show that our low-pass sequencing approach provided an effective and affordable method to decipher the complexity of populations with highly variable genomic structure, such as is often employed in breeding (Barrell *et al.* 2013; Hirsch *et al.* 2013). Our findings lead to several conclusions.

Genetic contribution from HI was not detected, despite frequent aneuploidy and widespread dosage variation. Genetic HIs act by either stimulating parthenogenesis in the female, or chromosome instability in the embryo resulting in missegregation and loss of one parental chromosome set. In potato, evidence in support of parthenogenesis has been reported (von Wangenheim *et al.* 1960; Montelongo-Escobedo and Rowe 1969; Peloquin *et al.* 1996). At the same time, genome elimination is supported by the detection of genetic markers from the HI in euploids and aneuploids arising from haploid induction crosses (Clulow *et al.* 1991, 1993;

Waugh *et al.* 1992; Wilkinson *et al.* 1995; Allainguillaume *et al.* 1997; Clulow and Roussel-Bourgeois 1997; Straadt and Rasmussen 2003; Ercolano *et al.* 2004). Whole-genome sequencing provides a more informative and reliable method to assess the genetic contribution of the HI. In *Arabidopsis* haploids, DNA from the HI can be identified readily from low-pass sequencing (Tan *et al.* 2015). It consists of whole chromosomes or segmental subsets of a single chromosome, consistent with the incomplete elimination of certain chromosomes, which persist autonomously, whole or rearranged.

We employed a similar approach with the dihaploids derived from Alca Tarma. Chromosomal addition or introgression comparable to the described cases should be evident by the appearance of whole chromosomes or large segments containing an HI centromere (Riera-Lizarazu *et al.* 1996; Zhao *et al.* 2013; Tan *et al.* 2015). Although aneuploidy was common, convincing evidence of chromosomal contribution from the HIs was not observed, consistent with AFLP analysis on this set (Velásquez *et al.* 2007) and analysis of another dihaploid population (Samitsu and Hosaka 2002). The capability of our method to identify long chromosomal segments that diverge in SNP or copy number is validated by *in silico* reconstructions (*Materials and Methods*) and its effectiveness in comparable systems (Tan *et al.* 2015; Henry *et al.* 2015), indicating that the transfer of large segments of HI DNA (Riera-Lizarazu *et al.* 1996; Zhao *et al.* 2013; Kuppu *et al.* 2015; Tan *et al.* 2015) did not take place.

Assessment of small introgressions is more challenging. Recent genotyping (Bartkiewicz *et al.* 2018) or sequencing (Pham *et al.* 2019) of other dihaploid potato populations found that ~1% of SNP loci displayed the presence of HI DNA in very small tracts, and with lower than expected allelic ratios. As in these reports, using high-coverage sequence data from three dihaploids, we detected many, widely dispersed SNPs represented by proportionally fewer aligned reads than expected for the addition of an HI DNA segment, or alternatively, replacement of a non-HI haplotype by homologous recombination. To address these observations, Pham *et al.* have suggested that tens of thousands of recombination events took place in each dihaploid during early growth before the HI genome was eliminated. Further, they have proposed that low allelic frequency could be explained by tissue chimerism. Mechanistically, this type of short introgression could be explained by somatic recombination caused by double-stranded DNA breaks followed by synthesis-dependent strand annealing or dsDNA break repair (Pâques and Haber 1999). Notably, while recombination of ectopic sequences has been demonstrated in plants (Puchta 1999; Čermák *et al.* 2017; Filler Hayut *et al.* 2017), these events are infrequent and require careful interpretation (Puchta and Hohn 2012). In this case, the scale of these changes ranged from ~30,000 to 300,000 per sequenced haploid and affected all examined haploids (Pham *et al.* 2019), implying extremely high efficiency of recombination. The hypothesis of autonomous replication of HI DNA segments, for which there are precedents (Cohen *et al.* 2008; Shibata *et al.* 2012), could

relieve the need for recombination. Nevertheless, it would also require high-efficiency propagation of extrachromosomal elements. These problems suggest a conservative interpretation of our data: these signals are artifactual and could have causes comparable to those identified in other gene conversion studies (Wijnker *et al.* 2013; Qi *et al.* 2014).

We conclude that, for the cross between HIs IVP101 or PL4 and tetraploid Alca Tarma, either the mechanism of haploid induction did not involve egg fertilization or genome elimination resulted in the loss of all HI chromosomes before the plants were evaluated. Events resulting in chromosome addition or introgression may be infrequent. For this reason, it may be premature to rule out genome elimination until more dihaploids derived from different parental combinations are evaluated. If haploid induction acts via genome elimination, both the addition of large DNA segments in the form of chromosomes and the rare introgression of small segments via recombination could be identified, and used for manipulation of the potato genome.

Our findings suggest that after *tuberosum* × *phureja* crosses, plants derived from seeds that did not express the purple spot marker and that display 2x genome content by flow cytometry are likely to be clean dihaploids (free of pollinator genome). That 11.4% of the dihaploid population evidently escaped initial screening against aneuploidy based on chloroplast counts, visual phenotyping, and chromosome counts underscores the difficulty of identifying aneuploids in highly variable dihaploid progeny. Their occurrence is consistent with the high frequency of aneuploid gametes in autotetraploids (Comai 2005) and with the ability of certain genotypes to tolerate imbalance (Rick and Notani 1961; Henry *et al.* 2010). The frequency of chromosome 2 trisomy, which carries the nucleolar organizing region of potato, may be explained if increasing the ribosomal RNA gene copy number offsets the disadvantage of linked dosage imbalance. Alternatively, ribosomal gene transcription (or some other unknown feature) could interfere with segregation (Tomson *et al.* 2006).

## Conclusions

We undertook this investigation with the objective of assessing large-scale structural variation and its causes in dihaploids produced through genetic induction. Using cost-effective low-pass sequencing, we documented extensive, large-scale structural variation affecting > 52% of the genome. We found that 11% of the dihaploids were trisomic, frequently for the chromosome that carries the nucleolar-organizing region. In spite of multiple previous reports of genomic contamination by the HI used as the pollinator, we did not detect any large-scale introgression of the HI genome.

## Acknowledgments

The authors thank Awais Khan for encouraging and coordinating the early phase of this collaboration; Michell Feldmann, Jordan Weibel, Jeanine Montano, and Helen Tsai

for constructive comments on writing and data analysis; and the CGIAR Collaborative Program on Roots Tubers and Bananas for supporting the development of the LOP population used here. This work was supported by the National Science Foundation Plant Genome Integrative Organismal Systems (IOS) Grant 1444612 (Rapid and Targeted Introgression of Traits via Genome Elimination) to L.C.

Author contributions: K.R.A., E.M., E.H.T., M.B., I.M.H., and L.C. designed experiments. M.S. performed cytological analysis. K.R.A., B.O., and E.H.T. performed experiments. K.R.A. and L.C. analyzed data and wrote the manuscript with input from all authors.

## Literature Cited

Allainguillaume, J., M. J. Wilkinson, S. A. Clulow, and S. N. R. Barr, 1997 Evidence that genes from the male parent may influence the morphology of potato dihaploids. *Theor. Appl. Genet.* 94: 241–248. <https://doi.org/10.1007/s001220050406>

Barrell, P. J., S. Meiyalaghan, J. M. E. Jacobs, and A. J. Conner, 2013 Applications of biotechnology and genomics in potato improvement. *Plant Biotechnol. J.* 11: 907–920. <https://doi.org/10.1111/pbi.12099>

Bartkiewicz, A. M., F. Chilla, D. Terefe-Ayana, J. Lübeck, J. Strahwald *et al.*, 2018 Maximization of markers linked in coupling for tetraploid potatoes via monoparental haploids. *Front. Plant Sci.* 9: 620. <https://doi.org/10.3389/fpls.2018.00620>

Benjamini, Y., and T. P. Speed, 2012 Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 40: e72. <https://doi.org/10.1093/nar/gks001>

Čermák, T., S. J. Curtin, J. Gil-Humanes, R. Čegan, T. J. Y. Kono *et al.*, 2017 A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* 29: 1196–1217. <https://doi.org/10.1105/tpc.16.00922>

Ciolaro, M and D. Wang, 2016 Clustering via the Mean Shift Algorithm. R package version 1.1-1. <https://CRAN.R-project.org/package=MeanShift>

Clulow, S. A., and F. Roussel-Bourgeois, 1997 Widespread introgression of *Solatium phureja* DNA in potato (*S. tuberosum*) dihaploids. *Plant Breed.* 116: 347–351. <https://doi.org/10.1111/j.1439-0523.1997.tb01010.x>

Clulow, S. A., M. J. Wilkinson, R. Waugh, E. Baird, M. J. Demaine *et al.*, 1991 Cytological and molecular observations on *Solanum phureja*-induced dihaploid potatoes. *Theor. Appl. Genet.* 82: 545–551. <https://doi.org/10.1007/BF00226789>

Clulow, S. A., M. J. Wilkinson, and L. R. Burch, 1993 *Solanum phureja* genes are expressed in the leaves and tubers of aneuploid potato dihaploids. *Euphytica* 69: 1–6. <https://doi.org/10.1007/BF00021720>

Cohen, S., A. Houben, and D. Segal, 2008 Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. *Plant J.* 53: 1027–1034. <https://doi.org/10.1111/j.1365-313X.2007.03394.x>

Comai, L., 2005 The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6: 836–846. <https://doi.org/10.1038/nrg1711>

Das, O. P., E. Poliak, K. Ward, and J. Messing, 1991 A new allele of the duplicated 27kD zein locus of maize generated by homologous recombination. *Nucleic Acids Res.* 19: 3325–3330. <https://doi.org/10.1093/nar/19.12.3325>

de Boer J. M., E. Datema, X. Tang, T. J. A. Borm, E. H. Bakker, *et al.*, 2015 Homologues of potato chromosome 5 show variable collinearity in the euchromatin but dramatic absence of sequence similarity in the pericentromeric heterochromatin.

BMC Genomics 16: 374. <https://doi.org/10.1186/s12864-015-1578-1>

Ercolano, M. R., D. Carpato, J. Li, L. Monti, A. Barone *et al.*, 2004 Assessment of genetic variability of haploids extracted from tetraploid ( $2n = 4x = 48$ ) *Solanum tuberosum*. Genome 47: 633–638. <https://doi.org/10.1139/g04-020>

Filler Hayut, S., C. Melamed Bessudo, and A. A. Levy, 2017 Targeted recombination between homologous chromosomes for precise breeding in tomato. Nat. Commun. 8: 15605. <https://doi.org/10.1038/ncomms15605>

Forster, B. P., E. Heberle-Bors, K. J. Kasha, and A. Touraev, 2007 The resurgence of haploids in higher plants. Trends Plant Sci. 12: 368–375. <https://doi.org/10.1016/j.tplants.2007.06.007>

Garrison E., and G. Marth, 2012 Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907. Accessed 22 March 2017.

Ghislain M., D. P. Zhang, and M. R. Herrera, 1999 Molecular Biology Laboratory Protocols Plant Genotyping: Training Manual. International Potato Center.

Gilles, L. M., A. Khaled, J. Laffaire, S. Chaignon, G. Gendrot *et al.*, 2017 Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. EMBO J. 36: 707–717. <https://doi.org/10.15252/embj.201796603>

Hanneman, R. E., and R. W. Ruhde, 1978 Haploid extraction in *Solanum tuberosum* group Andigena. Am. J. Potato Res. 55: 259–263. <https://doi.org/10.1007/BF02852134>

Hardigan M. A., E. Crisovan, J. P. Hamilton, J. Kim, P. Laimbeer, *et al.*, 2016 Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated *Solanum tuberosum*. Plant Cell 28: 388–405. <https://doi.org/10.1105/tpc.15.00538>.

Hardigan M. A., F. P. E. Laimbeer, L. Newton, E. Crisovan, J. P. Hamilton, *et al.*, 2017 Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proc. Natl. Acad. Sci. U. S. A. 114: E9999–E10008. <https://doi.org/10.1073/pnas.1714380114>.

Henry, I. M., B. P. Dilkes, E. S. Miller, D. Burkart-Waco, and L. Comai, 2010 Phenotypic consequences of aneuploidy in *Arabidopsis thaliana*. Genetics 186: 1231–1245. <https://doi.org/10.1534/genetics.110.121079>

Henry, I. M., M. S. Zinkgraf, A. T. Groover, and L. Comai, 2015 A system for dosage-based functional genomics in poplar. Plant Cell 27: 2370–2383. <https://doi.org/10.1105/tpc.15.00349>

Hermsen, J. G. T., and J. Verdenius, 1973 Selection from *Solanum tuberosum* group Phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot. Euphytica 22: 244–259. <https://doi.org/10.1007/BF00022632>

Hirsch, C. N., C. D. Hirsch, K. Felcher, J. Coombs, D. Zarka *et al.*, 2013 Retrospective view of North American potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. G3 (Bethesda) 3: 1003–1013. <https://doi.org/10.1534/g3.113.005595>

Iovene M., T. Zhang, Q. Lou, C. R. Buell, and J. Jiang, 2013 Copy number variation in potato - an asexually propagated autotetraploid species. Plant J. 75: 80–89. <https://doi.org/10.1111/tpj.12200>.

Ishii, T., R. Karimi-Ashtiyani, and A. Houben, 2016 Haploidization via chromosome elimination: means and mechanisms. Annu. Rev. Plant Biol. 67: 421–438. <https://doi.org/10.1146/annurev-arplant-043014-114714>

Jun, G., M. K. Wing, G. R. Abecasis, and H. M. Kang, 2015 An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. 25: 918–925. <https://doi.org/10.1101/gr.176552.114>

Karimi-Ashtiyani, R., T. Ishii, M. Niessen, N. Stein, S. Heckmann *et al.*, 2015 Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proc. Natl. Acad. Sci. USA 112: 11211–11216. <https://doi.org/10.1073/pnas.1504333112>

Kelliher, T., D. Starr, W. Wang, J. McCuiston, H. Zhong *et al.*, 2016 Maternal haploids are preferentially induced by CENH3-tailswap transgenic complementation in maize. Front. Plant Sci. 7: 414. <https://doi.org/10.3389/fpls.2016.00414>

Kelliher, T., D. Starr, L. Richbourg, S. Chintamanani, B. Delzer *et al.*, 2017 MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature 542: 105–109. <https://doi.org/10.1038/nature20827>

Kotch, G. P., R. Ortiz, and S. J. Peloquin, 1992 Genetic analysis by use of potato haploid populations. Genome 35: 103–108. <https://doi.org/10.1139/g92-018>

Kuppu, S., E. H. Tan, H. Nguyen, A. Rodgers, L. Comai *et al.*, 2015 Point mutations in centromeric histone induce post-zygotic incompatibility and uniparental inheritance. PLoS Genet. 11: e1005494. <https://doi.org/10.1371/journal.pgen.1005494>

Li, H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997. Accessed 22 March 2017

Liu, C., X. Li, D. Meng, Y. Zhong, C. Chen *et al.*, 2017 A 4-bp insertion at ZmPLA1 encoding a putative phospholipase A generates haploid induction in maize. Mol. Plant 10: 520–522. <https://doi.org/10.1016/j.molp.2017.01.011>

Maheshwari, S., E. H. Tan, A. West, F. C. H. Franklin, L. Comai *et al.*, 2015 Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet. 11: e1004970. <https://doi.org/10.1371/journal.pgen.1004970>

Mihovilovich, E., M. Aponte, H. Lindqvist-Kreuze, and M. Bonierbale, 2014 An RGA-derived SCAR marker linked to PLRV resistance from *Solanum tuberosum* ssp. andigena. Plant Mol. Biol. Report. 32: 117–128. <https://doi.org/10.1007/s11105-013-0629-5>

Montelongo-Escobedo, H., and P. R. Rowe, 1969 Haploid induction in potato: cytological basis for the pollinator effect. Euphytica 18: 116–123.

Pâques, F., and J. E. Haber, 1999 Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. 63: 349–404.

Peloquin, S. J., and A. C. Gabert, R. Ortiz, 1996 Nature of 'pollinator' effect in potato (*Solanum tuberosum* L.) haploid production. Ann. Bot. 77: 539–542. <https://doi.org/10.1006/anbo.1996.0064>

Pham G. M., L. Newton, K. Wiegert-Rininger, B. Vaillancourt, D. S. Douches, *et al.*, 2017 Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. Plant J. 92: 624–637. <https://doi.org/10.1111/tpj.13706>.

Pham, G. M., G. T. Braz, M. Conway, E. Crisovan, J. P. Hamilton *et al.*, 2019 Genome-wide inference of somatic translocation events during potato dihaploid production. Plant Genome. DOI: 10.3835/plantgenome2018.10.0079. <https://doi.org/10.3835/plantgenome2018.10.0079>

Pineda, O., M. W. Bonierbale, R. L. Plaisted, B. B. Brodie, and S. D. Tanksley, 1993 Identification of RFLP markers linked to the H1 gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. Genome 36: 152–156. <https://doi.org/10.1139/g93-019>

Puchta, H., 1999 Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. Genetics 152: 1173–1181.

Puchta, H., and B. Hohn, 2012 In planta somatic homologous recombination assay revisited: a successful and versatile, but delicate tool. Plant Cell 24: 4324–4331. <https://doi.org/10.1105/tpc.112.101824>

Qi J., Y. Chen, G. P. Copenhaver, and H. Ma, 2014 Detection of genomic variations and DNA polymorphisms and impact on analysis of meiotic recombination and genetic mapping. Proc. Natl. Acad. Sci. U. S. A. 111: 10007–10012. <https://doi.org/10.1073/pnas.1321897111>

Ravi, M., and S. W. L. Chan, 2010 Haploid plants produced by centromere-mediated genome elimination. *Nature* 464: 615–618. <https://doi.org/10.1038/nature08842>

Rick, C. M., and N. K. Notani, 1961 The tolerance of extra chromosomes by primitive tomatoes. *Genetics* 46: 1231–1235.

Riera-Lizarazu, O., H. W. Rines, and R. L. Phillips, 1996 Cytological and molecular characterization of oat x maize partial hybrids. *Theor. Appl. Genet.* 93: 123–135. <https://doi.org/10.1007/BF00225737>

Samitsu, Y., and K. Hosaka, 2002 Molecular marker analysis of 24- and 25-chromosome plants obtained from *Solanum tuberosum* L. subsp. *andigena* ( $2n = 4x = 48$ ) pollinated with a *Solanum phureja* haploid inducer. *Genome* 45: 577–583. <https://doi.org/10.1139/g02-019>

Shibata, Y., P. Kumar, R. Layer, S. Willcox, J. R. Gagan *et al.*, 2012 Extrachromosomal microRNAs and chromosomal micro-deletions in normal tissues. *Science* 336: 82–86. <https://doi.org/10.1126/science.1213307>

Spooner, D. M., M. Ghislain, R. Simon, S. H. Jansky, and T. Gavrilenko, 2014 Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Bot. Rev.* 80: 283–383. <https://doi.org/10.1007/s12229-014-9146-y>

Straadt, I. K., and O. S. Rasmussen, 2003 AFLP analysis of *Solanum phureja* DNA introgressed into potato dihaploids. *Plant Breed.* 122: 352–356. <https://doi.org/10.1046/j.1439-0523.2003.00878.x>

Tan, E. H., I. M. Henry, M. Ravi, K. R. Bradnam, T. Mandakova *et al.*, 2015 Catastrophic chromosomal restructuring during genome elimination in plants. *Elife* 4: e06516. <https://doi.org/10.7554/elife.06516>

Tomson, B. N., D. D'Amours, B. S. Adamson, L. Aragon, and A. Amon, 2006 Ribosomal DNA transcription-dependent processes interfere with chromosome segregation. *Mol. Cell. Biol.* 26: 6239–6247. <https://doi.org/10.1128/MCB.00693-06>

Velásquez, A. C., E. Mihovilovich, and M. Bonierbale, 2007 Genetic characterization and mapping of major gene resistance to potato leafroll virus in *Solanum tuberosum* ssp. *andigena*. *Theor. Appl. Genet.* 114: 1051–1058. <https://doi.org/10.1007/s00122-006-0498-5>

von Wangenheim, K., S. J. Peloquin, and R. W. Hougas, 1960 Embryological investigations on the formation of haploids in the potato (*Solanum tuberosum*). *Zh. Vyssh. Nerv. Deiat. Im. I. P. Pavlova*. 91: 391–399. <https://doi.org/10.1007/BF00888395>

Wagenvoort, M., and W. Lange, 1975 The production of aneudi-haploids in *Solanum tuberosum* L. group *Tuberosum* (the common potato). *Euphytica* 24: 731–741. <https://doi.org/10.1007/BF00132912>

Watanabe, K. N., and M. Orrillo, 1993 An alternative pretreatment method for mitotic chromosome observation in potatoes. *Am. Potato J.* 70: 543–548. <https://doi.org/10.1007/BF02846755>

Waugh, R., E. Baird, and W. Powell, 1992 The use of RAPD markers for the detection of gene introgression in potato. *Plant Cell Rep.* 11: 466–469. <https://doi.org/10.1007/BF00232692>

Wijnker, E., G. Velikkakam James, J. Ding, F. Becker, J. R. Klasen *et al.*, 2013 The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife* 2: e01426. <https://doi.org/10.7554/elife.01426>

Wilkinson, M. J., S. T. Bennett, S. A. Clulow, J. Allainguillaume, K. Harding *et al.*, 1995 Evidence for somatic translocation during potato dihaploid induction. *Heredity* 74: 146–151. <https://doi.org/10.1038/hdy.1995.21>

Zhao, X., X. Xu, H. Xie, S. Chen, and W. Jin, 2013 Fertilization and uniparental chromosome elimination during crosses with maize haploid inducers. *Plant Physiol.* 163: 721–731. <https://doi.org/10.1104/pp.113.223982>

Zinkgraf, M., K. Haiby, M. C. Lieberman, L. Comai, I. M. Henry *et al.*, 2016 Creation and genomic analysis of irradiation hybrids in *Populus*. *Curr. Protoc. Plant Biol.* 1: 431–450. <https://doi.org/10.1002/cppb.20025>

Communicating editor: J. Birchler