The Plant Cell, Vol. 27: 2370–2383, September 2015, www.plantcell.org

LARGE-SCALE BIOLOGY ARTICLE

A System for Dosage-Based Functional Genomics in Poplar[OPEN]

Isabelle M. Henry,[a] Matthew S. Zinkgraf,[b] Andrew T. Groover,[b,c] and Luca Comai[a,1]

[a] Department of Plant Biology and Genome Center, University of California, Davis, California 95616
[b] Pacific Southwest Research Station, U.S. Forest Service, Davis, California 95618
[c] Department of Plant Biology, University of California, Davis, California 95616

ORCID IDs: 0000-0002-6796-1119 (I.M.H.); 0000-0002-7229-8016 (M.S.Z.); 0000-0002-6686-5774 (A.T.G.); 0000-0003-2642-6619 (L.C.)

Altering gene dosage through variation in gene copy number is a powerful approach to addressing questions regarding gene regulation, quantitative trait loci, and heterosis, but one that is not easily applied to sexually transmitted species. Elite poplar (*Populus* spp) varieties are created through interspecific hybridization, followed by clonal propagation. Altered gene dosage relationships are believed to contribute to hybrid performance. Clonal propagation allows for replication and maintenance of meiotically unstable ploidy or structural variants and provides an alternative approach to investigating gene dosage effects not possible in sexually propagated species. Here, we built a genome-wide structural variation system for dosage-based functional genomics and breeding of poplar. We pollinated *Populus deltoides* with gamma-irradiated *Populus nigra* pollen to produce >500 F1 seedlings containing dosage lesions in the form of deletions and insertions of chromosomal segments (indel mutations). Using high-precision dosage analysis, we detected indel mutations in ~55% of the progeny. These indels varied in length, position, and number per individual, cumulatively tiling >99% of the genome, with an average of 10 indels per gene. Combined with future phenotype and transcriptome data, this population will provide an excellent resource for creating and characterizing dosage-based variation in poplar, including the contribution of dosage to quantitative traits and heterosis.

INTRODUCTION

Rapid changes in environment, markets, and selective pressures demand the accelerated breeding and deployment of new crop phenotypes. However, plants that have long generation times pose a challenge for traditional functional genomics and breeding approaches. This can be further exacerbated in vegetatively propagated crops, which can be recalcitrant to controlled breeding or have limited genetic variation. These liabilities are counterbalanced by specific advantages: In many clonally propagated species, novel genotypes, such as dosage variants involving aneuploidy (loss or gain of one or more entire chromosome arms) or duplication or deletion of smaller chromosomal segments, can be maintained vegetatively, even if they are meiotically unstable, and these new variants can be directly incorporated in breeding programs if they exhibit the desired characteristics. This property provides unique opportunities for functional analyses as well as a tool to create new variation for breeding.

Variation in gene dosage can have important consequences on phenotypes. Gene dosage is particularly important to variation observed in quantitative traits, including many of the traits of economic or ecological importance, including heterosis. Indeed, these traits are often governed by complex regulatory networks. Disrupting these regulatory interactions is more likely to provide trait variation than allelic or induced variation altering the function of a single protein. For example, the two genomes contributing to an F1 hybrid are slightly different, both in terms of allelic variation and because of the presence of deletions and insertions that encompass one or several genes. The effect of these indel mutations can be drastic to many developmental processes, as illustrated in Figure 1. Variation in gene dosage directly affects the expression of the genes located within the indel region (*cis*-effect) but also affects the regulation of genes located outside of the indel mutations (*trans*-effects). The intensity is dependent on the regulatory context but can be quite drastic on the regulation of certain genes and can result in a cascade of regulatory changes. These changes only affect the function of dosage-sensitive genes, but many of the genes that are important to developmental processes and regulatory cascades, such as transcription factors, are dosage sensitive.

Mechanistically, dosage can affect gene transcription or RNA stability through simple or combined *trans*-regulatory changes affecting transcription factors and small RNAs (Birchler and Veitia, 2010; Carlsbecker et al., 2010; Cruz and Houseley, 2014; Mallory et al., 2004; Veitia et al., 2013). These effects are often directly proportional to dosage (Huettel et al., 2008; Horiguchi et al., 2009), but compensatory or dosage-inverse changes have been observed as well (Birchler et al., 2001, 2007; Huettel et al., 2008; Birchler and Veitia, 2010). Additionally, gene dosage effects can be exerted at the protein level, such as in stoichiometric alterations in the interaction of different subunits of a complex (Veitia et al., 2013). Specifically, the gene-balance hypothesis postulates that

[1] Address correspondence to lcomai@ucdavis.edu.
The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Andrew T. Groover (agroover@fs.fed.us) and Luca Comai (lcomai@ucdavis.edu).
[OPEN]Articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.15.00349

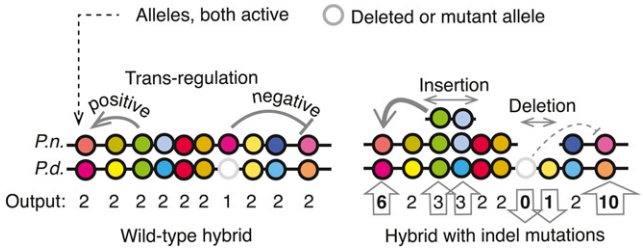


Figure 1. Dosage-Sensitive Gene Regulation.

Chromosomes (straight lines) with genes (circles) represent the two parental genotypes. Positive and negative regulatory interactions are depicted in gray. Left panel: In an F1 hybrid, most genes are balanced with a copy number of 2, except for a few genes that are absent in one of the parental genomes and, therefore, only present in single copy in the F1 interspecific hybrid. Each parent contributes different alleles to the F1 hybrid. The presence of both allelic types as well as the altered expression of the genes in altered copy number both contribute to the phenotypic characteristics of the F1 hybrid, including the heterotic response. Right panel: The presence of insertion and deletion mutations in one of the parental genomes can have drastic consequences on the gene expression patterns. Dosage effects are observed, which correspond to altered expression directly related to the altered gene dosage. Additionally, genes located elsewhere are directly or indirectly regulated by genes with altered copy number. As a result, their output, mRNA or protein, also varies in response to this *trans*-regulation. The direction and intensity of this effect depends on the specific interaction at hand, as exemplified in the up and down arrows located under each gene pair. P.n., *P. nigra*; P.d., *P. deltoides*.

modifying the dosage of proteins that function as part of a multiprotein complex causes a disruption in the stoichiometry of the different subunits of the complex and can partially or fully inhibit the functionality of the entire complex (Papp et al., 2003; Birchler and Veitia, 2010, 2012, 2014). As a result, genes that are part of protein complexes are more likely to be dosage sensitive or haploinsufficient, i.e., losing one genomic copy affects their function, even if the second copy is unaltered (Papp et al., 2003).

Gene dosage effects can be visible even if only one or a few genes are affected. For example, in animals, copy number variation has major effects on developmental traits and disease (Henrichsen et al., 2009; Craddock et al., 2010; Stankiewicz and Lupski, 2010; Alvarez and Akey, 2012). In plants, copy number variation has been identified in multiple species and demonstrated to affect important agronomic traits (Horiguchi et al., 2009; Pearce et al., 2011; Cook et al., 2012; Díaz et al., 2012; Li et al., 2012; Wingen et al., 2012; Ma et al., 2013). On a wider scale, when lesions encompass a larger chromosomal region or even an entire chromosome, gene dosage effects are responsible for aneuploidy syndromes. For example, the traits associated with Down syndrome result from the presence of a third copy of chromosome 21 and the altered dosage of many genes at once. Aneuploid syndromes also occur in plants. Plants carrying extra copies of different chromosome types (trisomics) have been described in many plant species and exhibit phenotypes specific to their chromosomal constitution (Blakeslee et al., 1922; McClintock, 1929; Khush, 1973; Henry et al., 2005, 2010). Analysis of dosage variation can thus provide an efficient approach for connecting phenotypes to regulatory genes.

Dosage variants can be produced rapidly using gamma-irradiation. Ionizing radiations induce DNA breaks, resulting in chromosomal deletions and insertions (indel mutations) of various lengths (Shirley et al., 1992). In plants, dosage mutants can be generated rapidly in a single step, by irradiation of pollen followed by fertilization (Khush and Rick, 1968; Pfahler, 1967; Vizir et al., 1994; Liharska et al., 1997; Farbos et al., 1999; Vizir and Mulligan, 1999; Gao et al., 2006; Kumar et al., 2012). The resulting progeny display male-specific indel mutations that have been used in plant genetic studies in various crops for nearly half a century (Khush and Rick, 1968), mostly in a forward genetics mode. Phenotypic consequences of irradiation vary depending on the size and location of the indel mutations (Vizir and Mulligan, 1999). Importantly, indel mutations provide a range of phenotypes not necessarily available in diploid populations or from other types of induced mutations, although the lack of tools to discover and characterize them has, until recently, prevented genome-scale characterization of dosage effects on traits.

Induction of indel mutations could be particularly well suited for poplar (*Populus* spp) functional genomics and breeding. Interspecific hybridization followed by clonal propagation of hybrids is the most effective method for producing elite commercial poplar clones. The complex interaction between the parental genomes that results in superior hybrid performance presumably includes the effects of insertions, deletions, and sequence divergence at regulatory and protein-coding sequences during speciation (Jiang et al., 2013; Yao et al., 2013). Such structural variation could be enhanced in poplar by underlying triploidy and aneuploidy (Bradshaw and Stettler, 1993). Triploid individuals have been identified in natural populations of poplar (Muntzing, 1936; DeWoody et al., 2008; Mock et al., 2008; Mock et al., 2012), and spontaneously arising triploid cultivars in breeding programs often display superior growth (Yuan et al., 2011). Current sequencing technologies enable genome-wide investigation of dosage effects incorporating both copy number and haplotype information.

To create dosage variation in poplar, we produced interspecific F1 hybrids in a controlled cross, using gamma-irradiated pollen from *Populus nigra* to pollinate *Populus deltoides*. We profiled the genome of 529 F1 hybrid trees by sequencing and, through dosage and haplotyping analysis, detected mutations in the form of large-scale insertions and deletions collectively covering the entire poplar genome multiple times. This population constitutes an attractive resource that can be used for functional genomics and gene discovery as well as a rich source of phenotypic variation available for breeding.

RESULTS

Production of Dosage Variant Using γ-Irradiated Pollen

Poplar breeding is heavily reliant on interspecific hybridization. The *P. deltoides* × *P. nigra* interspecific cross was chosen based on its common use in current breeding programs and high seed yield in controlled crosses (B. Stanton, personal communication). Dried mature pollen from a *P. nigra* individual was subjected to gamma irradiation, with doses varying from 0 to 150 Grays (Gy; Supplemental Table 1), based on previous work (Brewbaker and

Emery, 1962; Rudolph, 1978). Irradiated and nonirradiated control pollen was used the next day to pollinate female *P. deltoides* branches (see Methods). High pollen irradiation doses (125 Gy or higher) resulted in catkin abortion, and no live seed were obtained from those crosses (Supplemental Table 1). Crosses with pollen irradiated with 100 Gy or less produced viable seedlings. A total of 529 individuals, including 14 nonirradiated control individuals, 61 individuals irradiated at 50 Gy, and 454 individuals irradiated at 100 Gy were selected for dosage analysis and ploidy determination using next-generation sequencing approaches (Table 1; see Methods).

Gamma Irradiation Results in High Rates of Indel Mutations

Dosage variation was detected by calculating relative sequencing read coverage in 100-kb consecutive, nonoverlapping bins across the 19 chromosomes of *Populus trichocarpa*. Relative coverage values were normalized in such a way that values close to 2.0 corresponded to diploid copy number, while values closer to 1.0 or 3.0 corresponded to haploid (deletion) or triploid (insertion) copy numbers, respectively (see Methods). A diploid individual is thus expected to produce a line oscillating around 2.0 along the length of all chromosomes. One dosage plot was obtained for each individual, and examples of the types of indel mutations detected can be found in Figure 2. The number and characteristics of the indel mutations obtained are summarized in Table 1 and Supplemental File 1. The current version of the reference sequence also contains scaffolds that have not yet been assigned to specific chromosomes. These scaffolds are generally short (maximum 1 Mb) and contain many repeated sequences, making them particularly unsuitable for the type of analysis performed here. They were removed from our analyses but they cumulatively encode for 1821 genes, which represents 4.4% of the currently annotated genes. Consequently, additional undetected indel mutations might be present in the population.

Of the 515 irradiated seedlings analyzed, ~50% exhibited at least one instance of dosage variation. Surprisingly, while deletions were most common (76%), insertions were observed as well (24%), indicating that dosage variation can be produced in both directions. As expected, this percentage was tightly associated with the dosage of gamma irradiation, with the 100-Gy samples exhibiting much higher incidences of indel mutations than the 50-Gy samples (Table 1). Indels varied in length from whole chromosomes to small fragments. Dosage variants carried between 1 and 10 indels throughout their genome, with a mean number of 2.5 indels per individual. In the 14 control nonirradiated individuals, one individual (IFG_0_09) exhibited a single deletion event spanning ~5.2 Mb on chromosome 7. Deletions and insertions were both similarly partitioned between interstitial indels (~31%), terminal indels (~65%), and whole chromosome aneuploidy (~4%). Finally, indels with unusual patterns were also detected (Supplemental Figure 1). The exact chromosomal architecture of these unusual lesions is difficult to infer and mechanisms underlying their origin remain to be determined.

Increased Indel Incidence Is Associated with Triploidy

In poplar, triploids can arise in interspecific crosses (Bradshaw and Stettler, 1993), although the mechanisms leading to the formation and selection of unreduced (2N) gametes remain unclear. The potential presence of triploid individuals is extremely relevant in our population, since they can provide an additional layer of dosage variation. Moreover, triploids might be expected to be even more tolerant of dosage variation because their higher ploidy background provides a buffer (Khush, 1973; Birchler et al., 2001; Henry et al., 2006). This is even more relevant to the gametic stage: 2N gametes are expected to more easily withstand the loss of one out of two genomic copies compared with the complete loss of the only copy available in haploid gametes.

To survey our population for triploid individuals, all three of the parental genotypes were first sequenced at higher depth and single nucleotide polymorphisms (SNPs) between *P. nigra* and either of the two *P. deltoides* female parents were identified (see Methods; Supplemental Table 2 and Supplemental Files 2 and 3). We then used the parental SNP ratio to detect ploidy variants in the F1 individuals (see Methods; Figure 3). In short, the average percentage of *P. deltoides* allele was calculated for each chromosome and each individual (Figure 3A). Using these measurements, a "regular" diploid hybrid is expected to exhibit a percentage of *P. deltoides* allele close to 50% across the length of all chromosomes. Triploid individuals are easily distinguished from their diploid siblings based on average percentages close to 33% (DNN triploid with two copies of the *P. nigra* genome [N]) or 66% (DDN triploid with two copies of the *P. deltoides* genome [D]). We detected 36 DNN and 3 DDN triploid individuals (Figure 3B). To assess whether the extra chromosomal copies originated from 2N

Table 1. Number and Types of Indel Mutations Detected in the Interspecific Hybrids

Sample Type	*n*	No. with Indel(s) (%)	Total No. of Indels	No. of Indel(s) per Sample with Indel(s)	Indel Type (%)		Indel Location (%)		
					Deletion	Insertion	Interstitial	Terminal	Whole Chrom.
GWR 100 (2×)	298	168 (56)	367	2.2 ± 1.30	73	27	33	63	4
GWR 100 (3×)	36	35 (97)	168	4.8 ± 2.64	78	22	24	70	5
IFG 100 (2×)	117	61 (49)	123	2.0 ± 1.23	77	23	33	63	4
IFG 100 (3×)[a]	3	3 (100)	8	2.7 ± 2.08	100	0	13	75	13
GWR 50 (2×)[a]	61	5 (8)	10	2.0 ± 1.22	100	0	50	50	0
IFG 0[a]	14	1 (7)	1	1.0 ± N.A.[b]	100	0	0	100	0
All irradiated samples	515	272 (53)	676	2.5 ± 1.66	76	24	31	65	4

[a]Percentage statistics for indel type and location are not necessarily representative of the whole population for crosses with low indel numbers.
[b]N.A., not applicable.

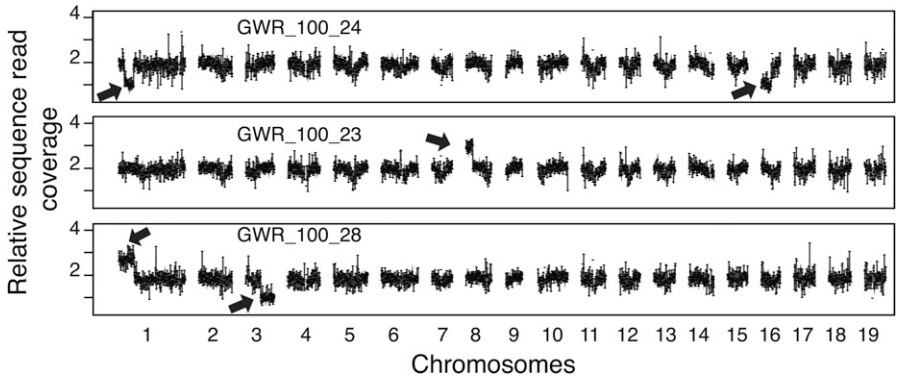


Figure 2. Indel Mutation Detection.

For each sample, sequencing reads were aligned to the reference sequences and sorted into consecutive bins of 100 kb, along the 19 chromosomes present in the current version of the poplar genome (v3.0). Three different samples are represented here, each originating from a cross using pollen that has been irradiated with 100 Gy. Each dot represents the average sequence read coverage for a 100-kb piece of chromosome. For ease of visualization, relative sequence read coverage values are normalized such that fragments present in two copies oscillate around 2.0. Changes up or down in relative sequence read coverage indicate copy number variation and correspond to the presence of an additional copy (insertion) or the deletion of a copy of a particular chromosome fragment, respectively (arrows).

gametes or resulted from high numbers of insertions in a diploid background, we calculated the percentage of the genome present in three copies in all individuals. Individuals fell into two very clear classes containing 0.6% ± 1.1% versus 93.1% ± 4.8% of the genome in triplicated copies, consistent with diploid individuals with a few insertions versus triploid individuals with a few deletions, respectively (Supplemental Figure 2). This bimodal distribution strongly supports the idea that these triploid individuals originated from 2N gametes. Whether these 2N gametes were due to meiotic errors or the result of mitotic error after gamma irradiation remains to be determined. In the case of the two DDN triploids, the latter hypothesis can be discarded since the 2N gamete originated from the nonirradiated female parent. As anticipated, the triploid individuals exhibited higher numbers of indels. Specifically, most of the triploids exhibited at least one indel (38/39 for the triploids versus 229/415 for the diploids, both at 100 Gy) and those with indels carried more than twice as many compared with their diploid siblings (Table 1).

SNP Analysis Also Confirms the Origin of the Indels

We also used the SNP analysis described above to determine the parental origin of each indel (Figure 4). To do so, the percentage of *P. deltoides* allele was calculated across each of the indels and compared with the *P. deltoides* ratio observed for the rest of the genome of that same individual (Figure 4A). Deletions are expected to result in a shift to 100% *P. deltoides* or 100% *P. nigra*, depending on whether the deletion event occurred in the pollen grain (loss of the *P. nigra* alleles) or the ovule (loss of the *P. deltoides* alleles), respectively. Insertion events are expected to result in a shift of *P. deltoides* alleles from 50 to 33% in the case of a gain of a *P. deltoides* fragment or 66% *P. deltoides* when the duplicated fragment originated from the *P. nigra* genome (Figure 4A). In all cases, indel mutations were associated with the expected shift in parental allele (Figure 4A). With the exception of one, all indels originated from the irradiated *P. nigra* genome (Figure 4A). The combination of both types of data allowed us to derive in silico karyotypes for each individual, as exemplified in Figure 4B.

The Indels Cover the Entire Poplar Genome Multiple Times

The ultimate goal of a functional genomics resource such as this one is to associate phenotypes with specific genes. With the 515 irradiated samples analyzed to date (61 at 50 Gy and 454 at 100 Gy; Table 1), 99.5% of the entire genome is affected by at least one indel, most of the time several (Figure 5). Specifically, each gene is affected by an average of 10.7 ± 5.22 indels and only two small pieces (400 kb of chromosome 1 and 700 kb of chromosome 13) are not affected in any of the mutant individuals. These missing regions only encompass a total of 15 genes.

Much like a population in which each individual contains one or a few small segments of one genome introgressed into another genome, this population offers variation of specific chromosomal segments in different individuals. However, instead of providing only sequence variation, our population carries dosage variation as well. The indel mutations that we have detected average 5.8 Mb, although there is tremendous variation in length and gene content. Additionally, for most positions in the genome, several indel mutations overlap each other but are never exactly the same, allowing for the effect of genes located within an indel mutation to be teased apart from the effect of adjacent genes shared by some indel mutations but not all (Figure 5). To get a better grasp of the resolution that our population currently offers, we determined the average number of "covariant" genes, i.e., genes whose functions cannot be teased apart because they vary together in all lines. We found that genes were on average linked to 131 ± 96 other genes. In other words, once we associate a specific phenotype with a genomic region, on average, 131 candidate genes remain. Four hundred more samples will soon be added to this resource, which will lower this number further. Additionally, future transcriptomic analyses will further refine the identity of candidate genes and

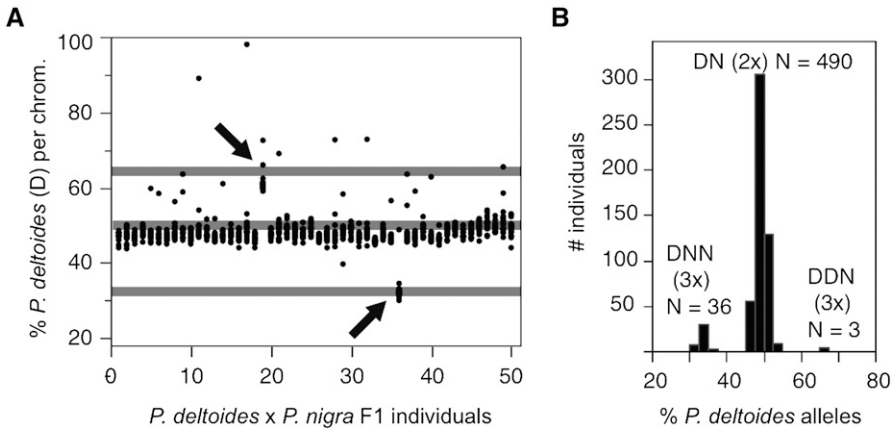


Figure 3. Triploid Detection.

For each irradiated F1 hybrid, the ratio of maternal *P. deltoides* (D) to paternal *P. nigra* (N) alleles was calculated for each chromosome, based on sequence data at polymorphic sites. A percentage of *P. deltoides* alleles of 50% is expected for diploid individuals (DN genotype), while it is expected at 33 or 66% for DNN and DDN triploids, respectively.

(A) For each individual, 19 data points represent the 19 chromosomes of the poplar genomic reference (v3.0). Data points deviating from the expected 50% indicate loss or duplication of that chromosome or a fragment of it (indel mutation or aneuploidy). Consistent deviation of all chromosomes indicates triploidy (indicated by arrows). Data from 48 samples are shown.

(B) Summary of all samples originating from *P. nigra* × *P. deltoides* cross ($n = 529$). For each sample, the mean percentage of *P. deltoides* is calculated across all chromosomes. Triploids are clear outliers on both flanks of the distribution of diploid individuals.

associated transcriptional modules based on the observed variation in gene expression.

The availability of a high number of indels provides additional information about the architecture of the *P. nigra* genome. For example, deletions and insertions are expected to accumulate at the ends of chromosomes, with the inner limit of their coverage delineating the centromeric region. Indeed, deletion of the centromeric region is expected to result in the loss of the entire chromosome, while duplication of the centromere is expected to result in chromosome breakage during mitosis. This is particularly clear for the chromosomes covered with the most indels (see chromosomes 6 and 19 in Figure 5B), when the density of indels is plotted along with gene density. Our data suggest a centromeric location for most of the chromosomes, which overlaps the gene-poor regions of the annotated *P. trichocarpa* genome.

The Dosage Variants Are Devoid of Point Mutations and Small Indels

One concern associated with mutagenized populations is the possible presence of undetected background mutations that could confound downstream analyses aimed at associating phenotypes with specific genomic loci. In the case of our gamma-irradiated population, two types of mutations might have gone unnoticed, in addition to indels located on unsequenced portions of the genome or on the current unassigned scaffolds: point mutations and indels small enough to escape our low-coverage analysis. To address these concerns, we obtained higher coverage of sequence reads from a few individuals and searched their genomes for these smaller scale mutations.

First, 11 samples were selected for medium coverage sequencing (on average 15 million reads per sample as opposed to ~1 million read per sample for the low-coverage analysis). This higher coverage increased our resolution for dosage variation detection 10 times (10-kb bins were used instead of 100-kb bins). Coverage values were examined in detail for those 10 individuals, but no further indel could be detected. Next, even higher coverage was obtained for an additional, unrelated four samples. Similarly, dosage variation analysis was performed using smaller bins (this time 5 kb) and using the nonirradiated sample (sample IFG_0_1) as the control for the other four samples. A few additional indels were detected (Supplemental Figure 3). A 54-kb terminal insertion was newly identified on chromosome 16 for one sample (Supplemental Figure 3A). Low pass coverage had identified unusual dosage patterns for chromosome 1 in another individual (IFG_100_61; Supplemental Figure 3F). This individual was thus chosen for higher coverage analysis as well. The patterns observed at high coverage confirmed the results of low coverage analysis, with multiple dosage transitions within the same chromosome (Supplemental Figure 3E). Three other additional indels were found in that individual as well, on chromosomes 3, 12, and 8 (Supplemental Figures 3B to 3D, respectively). The mechanism leading to such a dosage variation is unclear. Overall, these results suggest that the number of undetected lesions after our low resolution analysis is low but that it may be necessary to perform higher resolution analysis on lines that exhibited aberrant dosage plots in the original analysis.

Data from these five individuals were also input into a previously developed mutation detection pipeline, MAPS (Mutation and Polymorphism Surveyor; see Methods) (Henry et al., 2014). In short, MAPS searches for very short indels (1 to 5 bp) or point mutations that are specific to a single individual and not present in the others. This method searches for mutations within mapped reads, by comparing allele calls at each position in the genome

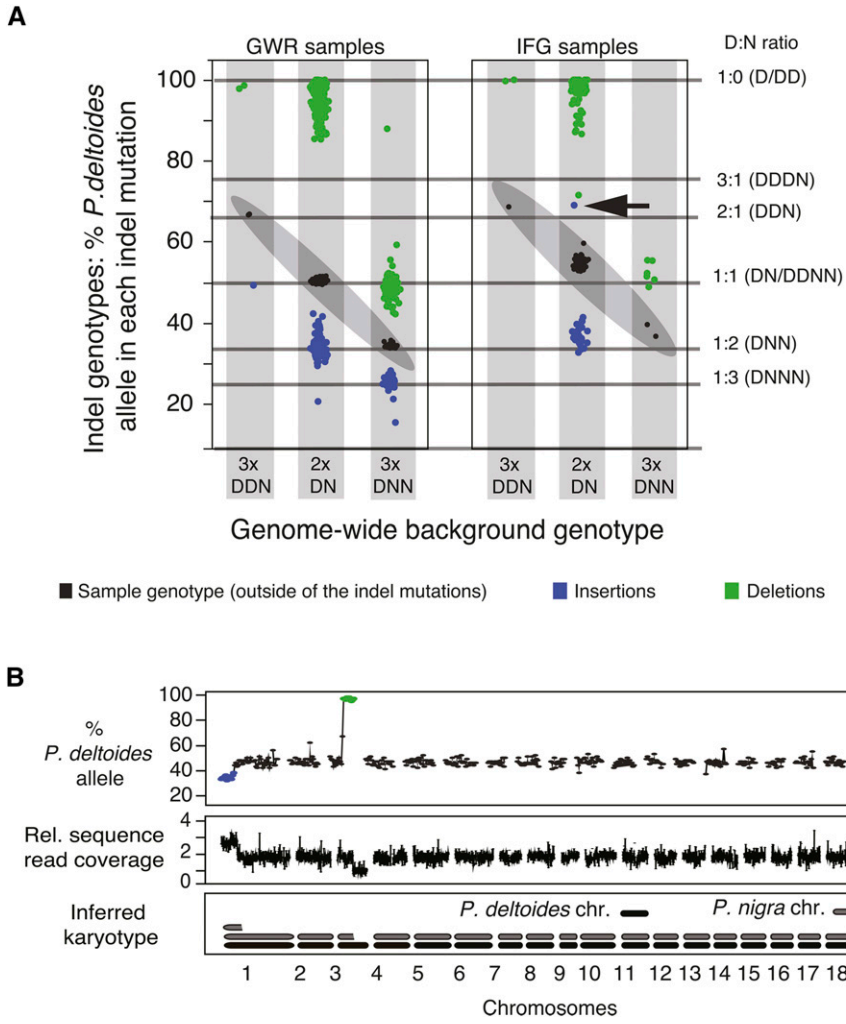


Figure 4. Inferring Karyotype from Dosage and Genotype Information.

(A) Determination of the parental origin of each indel mutation. For each indel mutation identified (Figure 1), the percentage of *P. deltoides* allele is calculated. For each individual contributing an indel mutation, the percentage of *P. deltoides* allele is also calculated for the rest of the genome and depicted in black. Indel mutations are expected to originate from the gamma-irradiated pollen parent (*P. nigra*). Therefore, deletions (green dots) are expected to result in an increased percentage of *P. deltoides* allele, while insertions (blue dots) are expected to result in a decrease in the percentage of *P. deltoides* allele across those indel mutations, relative to the genome-wide values (black dots and gray oval). This is observed in all cases expect for one (arrowhead), for which an insertion of a piece of the *P. deltoides* genome is observed.

(B) Example of in silico karyotyping for one of the dosage variants (100 Gy gamma-irradiated pollen) F1 hybrid individuals. Reads are aligned to the reference genome. Variation in sequence read coverage along a chromosome is used to diagnose copy number variation (aneuploidy or large indels). Ratios of alleles at positions polymorphic between the two parental genotypes inform us about ploidy level and about the origin of each of the chromosomal copies. Top panel: Chromosomes are divided into 5-Mb bins, and, for each bin, the percentage of *P. deltoides* allele is calculated. Values oscillating around 50% indicate equal allelic ratios, as is expected for a diploid individual. Middle panel: Dosage plot based on relative sequence read coverage in consecutive bins of 100 kb along each of the chromosomes. Two indels are visible on chromosomes 1 (insertion) and chromosome 3 (deletion). Bottom panel: Inferred karyotype from data presented in the top and middle panels.

between the different samples. These very short indels cannot be detected using the dosage analysis described above. Conversely, large-scale indels such as those detected by our dosage analysis cannot be detected using MAPS because they are not contained within single reads and cannot be distinguished from missing data. Using this method, we could not detect any homozygous mutations in the samples analyzed out of a total of ~450 assayed Mb. This is consistent with expectation for an F1 generation and indicates that false-positive mutations are uncommon. A total of 15 potential heterozygous mutations were detected out of a cumulative ~150 Mb of assayed space, corresponding to one mutation every 10 Mb. The control individual included in our samples did not exhibit a lower mutation rate, suggesting that these potential mutations are either artifacts or correspond to natural polymorphism, preexisting or new. Based on previous estimates of the frequency of functionally deleterious mutations

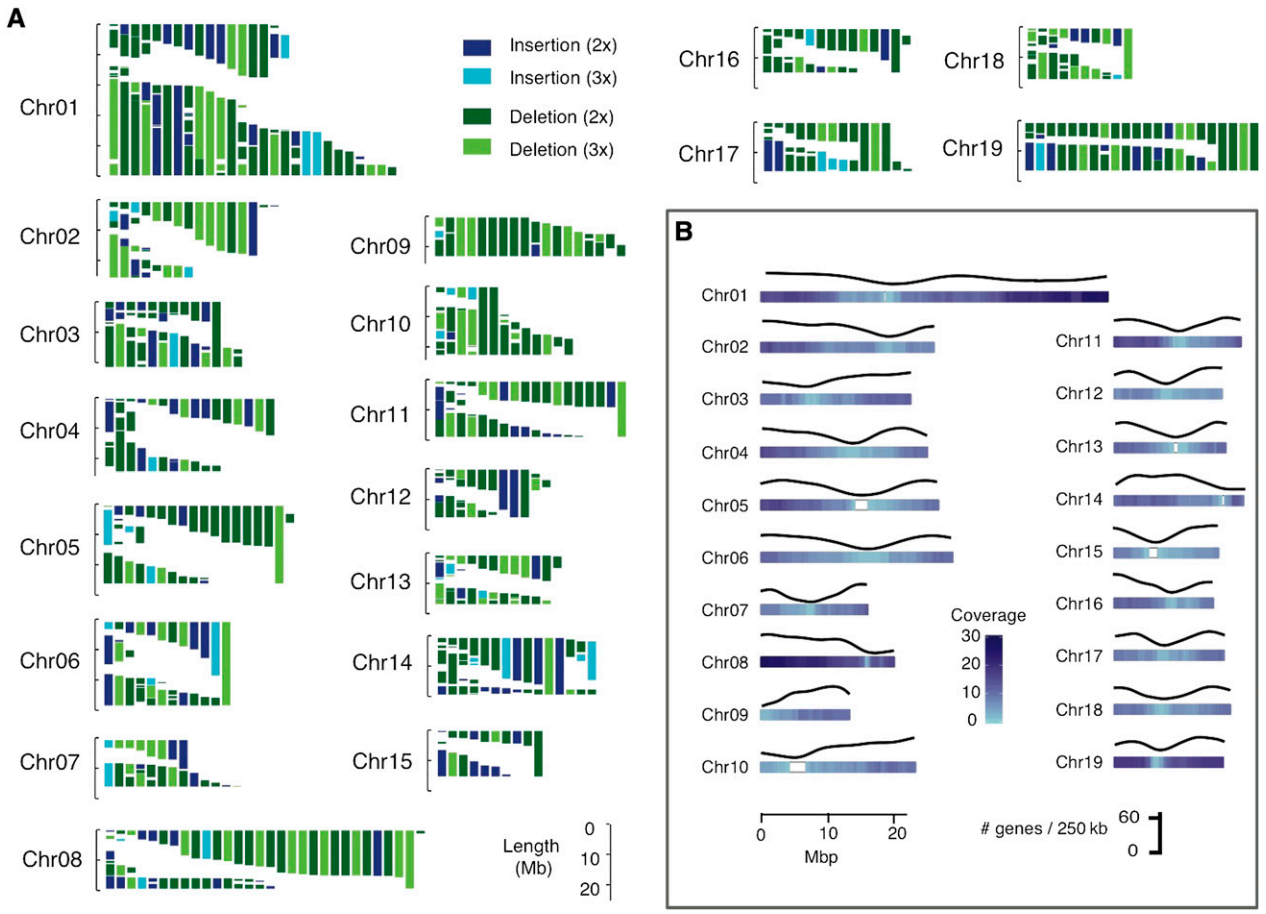


Figure 5. Summary of the Indel Mutations Identified in the *P. deltoides* × Gamma-Irradiated *P. nigra* Population.

(A) For each chromosome, indel mutations are tiled across their length. Indel mutations vary in length (up to a whole chromosome) and a single individual carries up to 10 indels. Deletions were most common (green) but insertions were observed as well (blue). Background ploidy (2× or 3×) is indicated as well.
(B) Density of indel mutations across the poplar genome. All indel mutations were pooled together, except for whole-chromosome aneuploidies. For each chromosome, the number of indel mutations covering every consecutive 100-kb bin was recorded (blue scale). Gene density data were obtained from the genomic annotation (gff3 file), and the number of genes per 250-kb bin was calculated (black curves). Segmental insertions and deletions are not expected to include the centromere. The density of indel mutations thus loosely delineates the centromeric region of each chromosome. This is particularly visible for the most highly covered chromosomes, chromosomes 8 and 19.

(Henry et al., 2014) and the exon space in the annotated poplar genome (Tuskan et al., 2006), we expect that this would result in one to two deleterious mutations per individual. Since these mutations are heterozygous, a functional effect is unlikely. In summary, our results suggest that the irradiated samples do not carry significant numbers of point or small-scale mutations.

The F1 Hybrid Population Exhibits Wide Phenotypic Variation

The F1 hybrid seedlings have been clonally replicated and entered into phenotypic trials. At their current young age, it is not yet possible to evaluate the long-term phenotypic consequences of the indels present in the population but it is clear that many of the dosage mutants display transgressive variation in specific traits (Figure 6). The frequency and intensity of these phenotypes manifestly exceed those found in control crosses. They affect traits such as leaf shape (Figure 6A), phenology (Figure 6C), and apical dominance (Figures 6B and 6D). For example, one of the recurrent phenotypes is a "viney" growth habit (Figure 6B), manifested by procumbent growth and thin stems in contrast to control individuals.

DISCUSSION

A Rich Resource for Functional Genomics

Poplar represents a singular challenge to functional genomic exploitation: It is dioecious (i.e., consists of male and female individuals), it takes several years to flower after germination, it is heterozygous at many loci, and it suffers from inbreeding

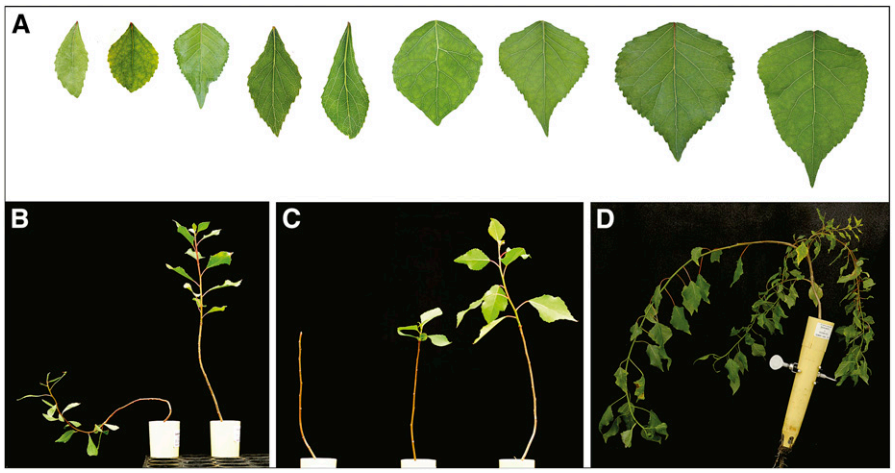


Figure 6. Phenotypic Diversity in the Dosage Variant Population.

(A) Diversity of leaf morphologies. Each of the leaves photographed correspond to plastochron 5 from a single F1 individual. The three leaves on the right come from F1 individuals exhibiting normal leaf morphology, while the other leaves exhibit morphologies that are slightly or strongly altered. Variation in venation patterns is also visible.

(B) Viney mutant (left) and wild-type unirradiated sibling (right) from the F1 hybrid population.

(C) Variation in phenology within the mutagenized population.

(D) Weepy mutant. Robust connections between each phenotype and indel mutations requires ongoing clonal replication.

depression. Approaches tailored to and successful in inbred rapidly cycling plant systems are impracticable in poplar and alternative approaches need to be considered. Dosage is an important aspect of gene regulation and a critical component of the heterotic response that contributes to superior performance in many crops. Altering gene dosage through structural chromosomal changes is a powerful approach to addressing questions regarding gene function, but one that is not easily applied to sexually transmitted species such as maize (*Zea mays*) or *Arabidopsis thaliana*, where structural variation is not usually transmitted to the next generation due to meiotic errors and selection in the haploid gametophyte. In vegetatively propagated crops, on the other hand, meiotic fidelity is not necessary and dosage variation is easily maintained through clonal replication. This allows for robust, replicated analyses examining the relationships between dosage variation and gene function or phenotypic outcomes.

Here, we exploit the advantage of clonal propagation and the known relationship between quantitative gene expression changes and phenotype to provide a resource that probes the effect of indel mutations, which on average allow attribution of each identified quantitative trait locus (QTL) to ~200 genes. Although allelic variation in the parents of our population could contribute to some of the phenotypic variation observed, preliminary comparison (Figure 6) to control populations and the abundantly documented effects of dosage in plants and animals suggest that the latter is the main contributor. Our system exploits the advantages of clonal propagation in species known to exhibit strong interspecific heterosis by saturating the genome with molecularly characterized deletions and insertions. The resource provides an ideal platform for advanced studies that can systematically examine the mechanisms associated with gene dosage, as well as consequences of dosage variation on gene expression, QTLs, transcriptional networks, and complex plant phenotypes.

Our population of poplar dosage variants was produced by performing interspecific crosses between *P. deltoides* females and gamma-irradiated *P. nigra* pollen. Approximately 500 of the resulting F1 hybrids are available (Supplemental Table 2) and have been screened for the presence of indel mutations, using sequencing read depth and composition to evaluate copy number of parental haplotypes (Figures 2 and 4). Approximately half of them carry at least one indel mutation (Table 1; Supplemental File 1). This population constitutes an important resource both for functional genomics and for breeding. First, all together, our population already carries >250 indel mutations that cumulatively tile the entire poplar genome several times (Figure 5). All genes are affected by at least one indel mutation, and most by multiple. Four hundred more individuals are awaiting molecular characterization and should enhance this resource. Next, the dosage variations detected include both deletions and insertions, allowing for the investigation of increased and decreased dosage. At the same time, our population contains ~7% triploid individuals (Figure 3), which provide additional levels of dosage variation. Indeed, the presence of insertions and deletions in two different ploidy backgrounds has the potential to provide a complete dosage series (1×, 2×, 3×, and 4×) for any affected gene (Figure 7). Additionally, we have demonstrated that our population carries negligible frequencies of small indels and single nucleotide polymorphisms (Supplemental Figure 3 and Supplemental Table 3). This supports the idea that the dosage variants detected using our approach are representative of the variation present within the population and can be connected to phenotypes. Finally, observation of our dosage mutant confirms that dosage variation is

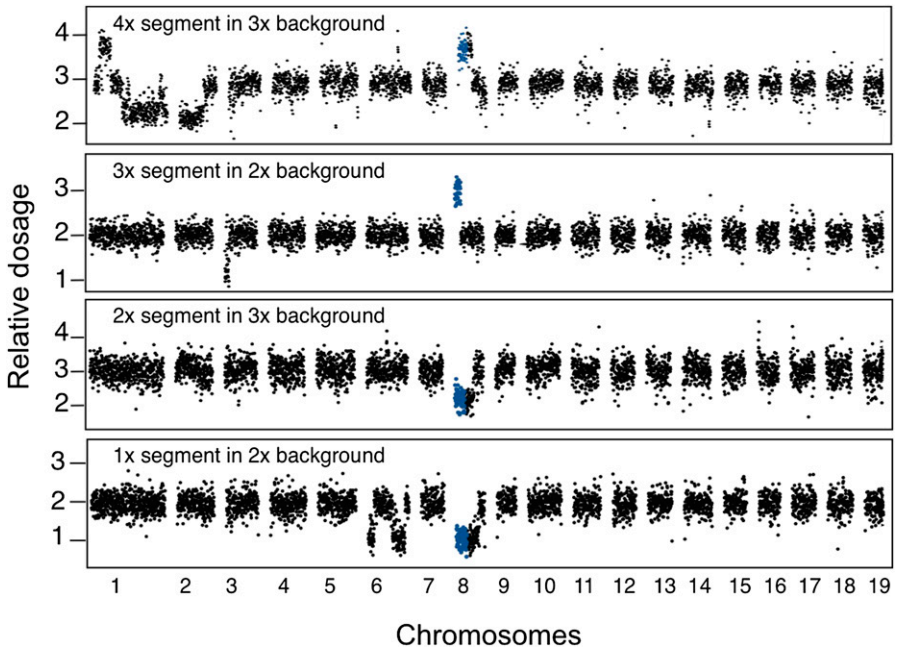


Figure 7. Example of a Dosage Series.

Four of the irradiated individuals share indel mutations on the top arm of chromosome 8 (light blue region). These individuals vary in background ploidy (diploid and triploid) and in the type of dosage variation (insertion and deletion), creating a dosage series including individuals with four different ratios of mutated segment to background genome (4:3, 3:2, 2:3, and 1:2).

tolerated in poplar and that novel phenotypes are present within the population (Figure 6). Detailed phenotypic characterization of the clones described here will complement the resource.

The functional genomic applications of this population are numerous. It represents an excellent opportunity to investigate the phenotypic impact of genome-wide dosage variation in a tree species and to obtain an increased understanding of the mechanisms influencing superior performance of hybrids and ploidy variants. It includes novel genetic and phenotypic variation that can be surveyed in forward and reverse genetic screens. It can be used to assist in identifying genes underlying previously identified QTLs (Dillen et al., 2009; Novaes et al., 2009; Marron et al., 2010; Rohde et al., 2011; Induri et al., 2012). For example, sex determinants have been mapped to chromosome 19 in several different species of poplar, although the exact location seems to vary between species (Tuskan et al., 2012; Kersten et al., 2014; Geraldes et al., 2015) and the exact molecular mechanisms leading to sex determination in poplar remain unclear. Chromosome 19 is particularly well covered by indel mutations (Figure 5), which can be used to further our understanding of the sex determination mechanisms in poplar. A similar approach was successfully applied in another dioecious species, white campion (*Silene latifolia*), in which gamma irradiation of pollen resulted in deletions of fragments of the Y chromosome, resulting in asexual flowers (Farbos et al., 1999). Finally, lesions produced by gamma irradiation also bear information about genomic architecture. In the case of poplar, for which a reference sequence, including chromosome-level assemblies, is available, the position of the segmental insertions and deletions could be compared with known gene density values along each of the chromosomes. One might expect, for example, that indel mutations would be more frequent in gene-poor regions. To the contrary, fewer indel mutations were detected in gene-poor regions. This is most likely explained by strong selection against indel mutations that affect the centromere: A centromere-less chromosome will missegregate, while a duplicated centromere will cause chromosome breakage at anaphase. Consistent with this hypothesis, the potential location of each centromere is evident in the mutation density plot (Figure 5B). The resolution of this analysis will improve further as the population of characterized irradiated trees increases. Exploitation of this resources for functional inference awaits the ongoing establishment of an outdoor plantation in which replication provided by clones of known genotypic constitution will enable robust assignment of trait values to each indel.

A Toolbox for Breeding and Production Purposes

Breeding in poplar is currently limited by the existing genetic variation present in natural populations. This resource can be mined to investigate the role of gene dosage in poplar hybrid performance and to identify chromosomal regions, genes, and even specific alleles that contribute to poplar bioenergy traits, all very important aspects of current poplar breeding programs. Every gene that is dosage sensitive can be surveyed in our germplasm, even if a given locus is fixed (nonvariant) in natural populations or between parental species. Because our dosage variants were produced in a modern plantation hybrid, they can provide usable new variation and it is possible that our germplasm

collection already includes new cultivars with unique and desirable traits. Since this kind of population can be developed in a relatively short time, is not transgenic, and can be vegetatively propagated for commercial use, this approach should be attractive to other systems as well.

Gamma Irradiation on Dry Pollen Results in High Occurrence of Large-Scale Indels but Few Point Mutations

Gamma mutagenesis has been used for decades to create populations of mutants in many crop species. Yet, our understanding of the effect of gamma mutagenesis on the genome remains superficial, mostly because the tools required to systematically detect mutations have only recently become available. Gamma rays are known to induce large-scale rearrangements (deletions, translocations, and sometimes insertions), as well as point mutations (Naito et al., 2005; Sato et al., 2006). The type and frequency of recovered mutations depend both on tissue type and irradiation dose, as well as on the approach used to detect them. For example, in rice (*Oryza sativa*), several populations have been characterized, with variable results (Bruce et al., 2009; Morita et al., 2009).

Our analysis revealed that our samples carried extensive insertions and deletions of variable length, including whole chromosomes. This is in contrast with previous reports indicating that gamma irradiation mostly results in smaller insertions and deletions. The difference could originate from the fact that most other studies focused on mutations that had been transmitted through meiosis to the progeny. This might be only a very small, biased subset of the original mutation load (Stadler and Roman, 1948; Naito et al., 2005). It is important to note that, while we can easily detect each dosage mutation in our population, our data do not currently allow us to derive the exact chromosomal architecture of each sample. This is exemplified in Supplemental Figure 4, which illustrates how dosage plots can be consistent with several genomic architectures, depending on how the original double-stranded breaks resulting from the gamma irradiation were resolved. For example, the deletion of a chromosome piece can also be accompanied by a translocation ("deletion" in Supplemental Figure 4). Similarly, the detection of insertions indicates the presence of additional copies of the same chromosomal fragment but is not informative as to where these fragments are located (ins_I, II, and III in Supplemental Figure 4). Irrespective of the actual chromosomal architecture, these observed large-scale indels and whole-chromosome aneuploidies would probably not be transmitted to the next generation because they would hinder chromosome pairing. This is not a problem in this clonally propagated system though, as each specific variant can easily maintained and replicated clonally.

We searched for point mutations using methods that were successful in ethyl methanesulfonate-mutagenized wheat (*Triticum aestivum*) and rice but found virtually none (Supplemental Table 3). One possible explanation for the discrepancy between our results and those of others is the type of material used for irradiation. Specifically, temperature, moisture content, and oxygen concentration are all thought to affect the effectiveness of gamma irradiation (Da Silva Aquino, 2012), as well as the post-irradiation storage length (IAEA, 1977). We irradiated dry pollen grains and used them for crossing immediately after irradiation (a few hours to a day later). This may have contributed to reducing the amount of postirradiation damage due to oxygen reactive species and resulted in the very low frequency of point mutations observed.

Triploidy Is Common in Specific Poplar Interspecific Crosses

Triploids provide an additional level of variation that shifts the ratio of parental alleles throughout the genome. This can have drastic consequences on development, especially in the endosperm, for which the ratio of maternal to paternal genome is particularly important for proper development (Scott et al., 1998). We found a high percentage of triploids in the progeny of one of the crosses performed in the context of this work. The fact that another cross performed with the same irradiation dose (100 Gy) and using pollen from the same male individual did not produce such a high percentage of triploids (Supplemental Table 2) suggests that this high frequency of 2N gamete is not linked to irradiation but possibly specific to the particular cross at hand. This is consistent with previous findings. Indeed, progeny with altered ploidy levels have arisen spontaneously in breeding programs utilizing interspecific hybridization. For example, the F1 progeny of a controlled cross of *P. trichocarpa* × *P. deltoides* showed an unusually high rate (~30%) of triploidy (Bradshaw and Stettler, 1993). On the other hand, other breeders using similar germplasm in interspecific crosses reported much lower frequencies of triploids, potentially indicating that environmental factors might also influence the rate of 2N gamete formation. Similarly, unreduced male and female gametes have be induced experimentally in Asian poplar species using heat treatments (Lu et al., 2013), and temperature-dependent 2N gamete formation has been reported in other species as well (Pécrix et al., 2011; De Storme et al., 2012).

Regardless of the exact mechanism leading to the production of triploid F1 hybrids, their presence is an unanticipated asset to this resource, as we found that they tend to carry more indels than their diploid counterparts. This is consistent with previous work, showing that higher mutation rates could be obtained when using polyploid material (Slade et al., 2005; Tsai et al., 2013; Henry et al., 2014).

METHODS

Source of Materials

Mature dried pollen grains from male *Populus nigra* trees were collected from individual SO361SL from Greenwood Resources. Female branches were collected from *Populus deltoides* female trees SO546SL and SO598SL, also from Greenwood Resources.

Production of Gamma-Irradiated Hybrids

Crosses were performed using methods described previously (Stanton and Villar, 1996). Briefly, dormant fertile branches were collected from mature trees in January of 2012. Branches from the male *P. nigra* individual were forced in greenhouse conditions and pollen was collected, desiccated, and stored at 4°C prior to being irradiated. The dried pollen was transferred to scintillation vials and subjected to various doses

(Supplemental Table 1) of gamma irradiation using a JL Shepherd Model 143-68 Irradiator with a Cesium-137 source (Center for Comparative Medicine, UC Davis, CA).

Two different genotypes were used as the source of female branches, and crosses were performed at two independent locations (Supplemental Table 1). Branches from the *P. deltoides* female were freshly cut at the base, treated with Dip 'N Grow hormone treatment (containing 1% indole-3-acetic acid and 0.5% 1-naphthaleneacetic acid), and rooted using the false bottom bucket method (Stanton and Villar, 1996) at 4°C with soil warmed to ~20°C using bucket heaters (Grainger). After 6 weeks rooting, females were forced in pollination enclosures in greenhouse conditions. Irradiated and nonirradiated control pollen was used to fertilize female branches the day following gamma irradiation to minimize viability loss after irradiation. Each female branch was pollinated twice at 2-d intervals.

Crosses were kept in separate enclosures to avoid contamination. Mature seeds were collected and counted. Collected seeds were cleaned and sown on SunShine Mix (Sun Grow Horticulture) for germination. Young leaves were sampled from each F1 hybrid progeny and immersed in silica prior to DNA extraction.

Sequencing Library Preparation

Leaf samples were processed for DNA extraction and preparation of sequencing libraries. DNA extraction was performed using the Qiagen DNeasy kit (Qiagen) following the manufacturer's recommendations, except that 4 μL of proteinase K (20 mg proteinase K in 10 mM Tris-HCl, 20 mM $CaCl_2$, and 50% glycerol) was added to the lysis buffer and the mix was incubated at 55°C for 30 min prior to extraction. The resulting DNA was used to produce sequencing libraries in one of two ways.

Lower complexity RESCAN libraries were produced from 44 samples (samples IFG_0_1 through IFG_0_4 and IFG_100_1 through IFG_100_40), as described earlier (Monson-Miller et al., 2012), with the following exceptions: fragments were selected for sizes between 300 and 600 bp (1:0.8 and 1:0.5 ratios of sample to AMPure XP beads volume for the bottom and top size selections, respectively). *Nla*III incubation time was flexible (between 1 h and overnight) and 1 to 4 μL of 0.5 μM adapter A and B mixture were used. Final enrichment was performed using 2.5 μL of 5 μM paired end primer mix and 10 to 12 cycles of amplification. Sequencing adaptors containing 5-bp barcode sequences were used.

Genomic libraries were produced from all other samples, as follows. For each sample, 500 ng of starting DNA was used. Water was added to the DNA to reach a total of 50.4 μL. The DNA was fragmented by adding 6 μL of 10× fragmentase buffer, 0.6 μL of 100× BSA, and 3 μL of DNA fragmentase (New England Biolabs). The mixture was incubated for 45 min at 37°C before 5 μL of 0.5 M EDTA was added to terminate the reaction. The fragmented DNA was purified using AMPure beads (Beckman Coulter) and a 1:1 ratio of beads to sample volumes. End repair and "A"-base addition were performed using the Next DNA Sample Prep kit (New England Biolabs) according to the manufacturer's recommendations, except that half of the recommended volumes were used for the end-repair reactions. Reactions were purified using AMPure beads and a 1:1 ratio of beads to sample volumes, with a final elution volume of 13 μL. Adaptors for Illumina sequencing were ligated by combining the following: 13 μL of DNA, 15 μL of quick ligation reaction buffer, 1 μL of adaptor mix (50 mm), and 1 μL of quick T4 DNA ligase and 20 μL of water for a total of 50 μL. Adaptors containing 6-bp index sequences (Bioo Scientific) or 8-bp index sequences (in-house design) were used. Size selection of the resulting fragments was performed using a two-step AMPure purification, as described previously (Monson-Miller et al., 2012), with ratios of 1:0.6 and 1: 0.7 volumes of samples to AMPure for the upper and lower cuts, respectively, and eluted in a final volume of 20 μL. The resulting libraries were amplified by PCR by mixing 10 μL of template DNA, 12.5 μL of Phusion High-Fidelity PCR Master Mix (Finnzymes), and 5 μL of 5 mM paired-end primer mix, with the following protocol: 30 s at 98°C followed by 10 cycles of 10 s at 98°C, 30 s at 65°C, and 30 s at 72°C, ending with 5 min at 72°C. The PCR products were purified using AMPure beads and a 1:0.8 ratio of beads to sample volumes, with a final elution volume of 15 μL.

For both types of libraries, samples were quantified using a Qubit 2.0 fluorometer and the High Sensitivity Quantification kit (Invitrogen), and equimolar amounts of each sample library were pooled prior to sequencing. Pools of up to 96 samples were sequenced on an Illumina HiSeq2000 or HiSeq2500 sequencer (Illumina), and ~1 million, 50-bp reads were obtained per sample.

Read Processing

Sequencing reads were divided into individual pools according to their barcode or index sequence. Barcode sequences, when present, were removed from each read. Reads were further filtered for quality, the presence of ambiguous (N) bases, length, and adaptor/primer contamination using custom Python scripts available online (http://comailab.genomecenter.ucdavis.edu/index.php/Barcoded_data_preparation_tools). Reads were aligned to the current version (v3.0) of the *P. trichocarpa* genome sequence (Tuskan et al., 2006) using the Burrows-Wheeler Alignment tool (Li and Durbin, 2009) and further processed using a combination of SAMtools (Li et al., 2009) and custom Python scripts (see above). An alignment (.sam) file was created for each sample, which was used for downstream analyses.

Dosage Analysis

For each individual, reads were pooled into nonoverlapping 100-kb bins covering the 19 largest scaffolds of the reference genome (corresponding to the 19 chromosomes of poplar) using custom Python scripts available online on our laboratory website (http://comailab.genomecenter.ucdavis.edu/index.php/Bin-by-sam), as described previously (Henry et al., 2010). Raw coverage across the genome was derived by counting the number of reads that mapped to each bin. Reads that contained a maximum of five mismatches with the reference sequence were selected from the alignment files. The mean percentage values across all individuals was used as control, and relative coverage values were derived for each bin such that coverage values along chromosomes oscillated around 2.0 for chromosome or chromosome fragments present in two copies and deviated up or down in individuals carrying increased or decreased dosage of chromosomal fragments. For 11 of the samples originating from the gamma-irradiated population, higher coverage (on average 15 million 50-bp reads per samples) was obtained, and more precise dosage plots were derived, using bins of 10 kb. Using bins of smaller size results in a high number of bins with higher coverage variation, depending on the sequence context. This results in noisier dosage traces and makes the detection of indel mutations more challenging. To address this problem, outliers were removed by limiting the analysis to bins for which the total number of reads for all samples combined ranged between 500 and 3000. For the five samples used for point mutation detection, and high-coverage indel detection, even higher coverage was obtained, and reads were trimmed prior to alignment in order to obtain exactly 115 million reads, each 50 bp long. This step insures that the samples are fully comparable and that mapping biases associated with read lengths will be similar in all samples. For indel detection, bins of 5 kb were used and bins were selected that exhibited a total coverage ranging from 1000 to 5000. For all bin sizes, the criteria used for identifying indels were as follows: At least three consecutive bins had to exhibit low (deletion) or high (insertion) relative coverage values. Next, data for the corresponding bins were assessed in at least 10 other individuals to verify that the indel was specific to the sample at hand or common to several samples. If so, the indel was discarded because it was assumed to be due to variation inherent to the sequence context.

Triploid Detection and Determination of the Parental Origin of the Indels

Positions polymorphic between the *P. deltoides* and the *P. nigra* genomes were identified using parental sequencing data. Separate SNP lists were created for the two female *P. deltoides* trees and alleles were assigned to one or the other genome (Supplemental Files 2 and 3). Next, mpileup files were created, each containing all base calls in all samples sharing the same maternal genome, using the SAMtools package (Li et al., 2009) and the Python script named run-mpileup.py (available at http://comailab.genomecenter.ucdavis.edu/index.php/Mpileup for download and documentation). This mpileup file was further processed to obtain the percentages of each base call at each position and each sample using the script called mpileup-parser.py (see http://comailab.genomecenter.ucdavis.edu/index.php/Mpileup for download and documentation). Using this parsed mpileup file, base calls at the polymorphic positions were recorded for each sample. Finally, ratios of parental alleles were calculated either for each indel or for complete chromosomes. From these ratios, we inferred copy number variations as follows: bins or chromosomes exhibiting a 1:1 ratio of *P. deltoides* to *P. nigra* alleles are diploid, while triploids exhibit a 1:2 or 2:1 ratio (Figure 3). The direction of the allelic bias was used to determine the origin of each indel as well (Figure 4). If all chromosomes displayed a similar allelic bias consistent with the expected 1:2 or 2:1 ratios, the individual was identified as triploid (Figures 3 and 4; Supplemental Figure 2).

Mutation Detection

Point mutations and short indels (1 to 3 bp) were detected using the MAPS pipeline, as described previously (Henry et al., 2014) and using software available from our laboratory website (http://comailab.genomecenter.ucdavis.edu/index.php/MAPS). In short, five individuals were sequenced at high sequencing read coverage and used as controls for each other to detect mutations specific to a single individual. All reads were cut down to 50 bp, and the number of reads from each sample was subsampled to match the number of reads present in the smallest sample (115 million). To avoid false positives originating from heterozygosity in one or the other parent, only positions that were well covered, homozygous, and identical in the parental reads were retained. The MAPS parameters used were set to the default values except for the following: -l 3, -v 30, and -c 1500. For the first part of MAPS (MAPS 1), the minimum percentage for considering that a position is heterozygous was set to 5% (-i 5). For the second part of MAPS (MAPS 2), additional parameters were used: -d 6, -s 5, and -p 5. Mutation density was calculated as previously described (Henry et al., 2014).

Data Sharing

All dosage plots and a description of the chromosomal regions exhibiting increased or decreased dosage have been deposited online on the poplar Genome Integrative Explorer (www.popgenie.org; Sjödin et al., 2009) and are readily accessible there through a user-friendly Web interface dedicated to this resource (PopIndels) that allows users to search for indels covering specific genes. This resource will be updated as we characterize new variants.

Accession Numbers

Sequence data from this article can be found in the NCBI SRA database under accession number SRP040492 and BioProject ID PRJNA241273.

Supplemental Data

Supplemental Figure 1. Example of unusual dosage patterns.

Supplemental Figure 2. Percentage of the genome in triploid state per individual.

Supplemental Figure 3. High coverage data from 11 samples reveals few additional indels.

Supplemental Figure 4. Potential chromosomal rearrangements consistent with the observed dosage plots.

Supplemental Table 1. Outcome of *P. deltoides* × *P. nigra* interspecific crosses, using gamma-irradiated pollen.

Supplemental Table 2. Summary of the number of F1 hybrid progeny tested and the number of triploids detected.

Supplemental Table 3. Outcome of mutation detection.

Supplemental File 1. List of indel mutations in the F1 population.

Supplemental File 2. List of polymorphic positions between the *P. nigra* pollen parent (SO361SL) and the *P. deltoides* female parent (SO598SL).

Supplemental File 3. List of polymorphic positions between the *P. nigra* pollen parent (SO361SL) and the *P. deltoides* female parent (SO546SL).

ACKNOWLEDGMENTS

We thank Brian Stanton, Kathy Haiby, and Richard Shuren of Greenwood Resources for providing germplasm, performing controlled crosses, and providing extensive technical advice on poplar breeding and cultivation. We thank Annie Mix, Courtney Castle, Shani Sarena, and Suzanne Gerttula at the U.S. Forest Service Institute of Forest Genetics and Brian Watson at the University of Washington for assistance in poplar breeding and cultivation. We thank Clara Williams, Helen Tsai, and Jian Gao for technical assistance and Meric Lieberman and Kathie Ngo for bioinformatics assistance. We thank Karen Shaw at the Center for Comparative Medicine (UC Davis) for assistance with the irradiator. This work used the UC Davis DNA Technologies Core and the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303 for sequencing services. This research was supported by the DOE Office of Science, Office of Biological and Environmental Research, Grant DE-SC0007183.

AUTHOR CONTRIBUTIONS

L.C., A.T.G., and I.M.H. designed the experiments. A.T.G. and M.S.Z. performed the crosses and generated and maintained the F1 population. I.M.H. coordinated and/or performed the preparation of sequencing libraries and analysis of the next-generation sequencing data. M.S.Z. performed the phenotypic analysis. I.M.H. drafted the article. All authors read, reviewed, and approved the final article.

Received April 25, 2015; revised July 28, 2015; accepted August 5, 2015; published August 28, 2015.

REFERENCES

Alvarez, C.E., and Akey, J.M. (2012). Copy number variation in the domestic dog. Mamm. Genome **23:** 144–163.

Birchler, J.A., Bhadra, U., Bhadra, M.P., and Auger, D.L. (2001). Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. Dev. Biol. **234:** 275–288.

Birchler, J.A., and Veitia, R.A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol. **186:** 54–62.

Birchler, J.A., and Veitia, R.A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc. Natl. Acad. Sci. USA **109:** 14746–14753.

Birchler, J.A., and Veitia, R.A. (2014). The Gene Balance Hypothesis: dosage effects in plants. Methods Mol. Biol. **1112:** 25–32.

Birchler, J.A., Yao, H., and Chudalayandi, S. (2007). Biological consequences of dosage dependent gene regulatory systems. Biochim. Biophys. Acta **1769:** 422–428.

Bradshaw, H.D.J., Jr., and Stettler, R.F. (1993). Molecular genetics of growth and development in Populus. I. Triploidy in hybrid poplars. Theor. Appl. Genet. **86:** 301–307.

Blakeslee, A.F., Belling, J., and Farnham, M.E. (1922). Variation in Datura due to changes in chromosome number. Am. Nat. **56:** 16–31.

Brewbaker, J.L., and Emery, G.C. (1962). Pollen radiobotany. Radiat. Bot. **1:** 101–154.

Bruce, M., Hess, A., Bai, J., Mauleon, R., Diaz, M.G., Sugiyama, N., Bordeos, A., Wang, G.L., Leung, H., and Leach, J.E. (2009). Detection of genomic deletions in rice using oligonucleotide microarrays. BMC Genomics **10:** 129.

Carlsbecker, A., et al. (2010). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature **465:** 316–321.

Cook, D.E., et al. (2012). Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science **338:** 1206–1209.

Craddock, N., et al.; Wellcome Trust Case Control Consortium (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature **464:** 713–720.

Cruz, C., and Houseley, J. (2014). Endogenous RNA interference is driven by copy number. eLife **3:** e01581.

Da Silva Aquino, K.A. (2012). Sterilization by gamma irradiation. In Gamma Radiation, F. Adrovic, ed (Rijeka, Croatia: InTech), pp. 171–206.

De Storme, N., Copenhaver, G.P., and Geelen, D. (2012). Production of diploid male gametes in Arabidopsis by cold-induced destabilization of postmeiotic radial microtubule arrays. Plant Physiol. **160:** 1808–1826.

DeWoody, J., Rowe, C.A., Hipkins, V.D., and Mock, K.E. (2008). Pando lives: Molecular genetic evidence of a giant aspen clone in Central Utah. West. N. Am. Naturalist **68:** 493–497.

Díaz, A., Zikhali, M., Turner, A.S., Isaac, P., and Laurie, D.A. (2012). Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). PLoS One **7:** e33234.

Dillen, S.Y., Storme, V., Marron, N., Bastien, C., Neyrinck, S., Steenackers, M., Ceulemans, R., and Boerjan, W. (2009). Genomic regions involved in productivity of two interspecific poplar families in Europe. 1. Stem height, circumference and volume. Tree Genet. Genomes **5:** 147–164.

Farbos, I., Veuskens, J., Vyskot, B., Oliveira, M., Hinnisdaels, S., Aghmir, A., Mouras, A., and Negrutiu, I. (1999). Sexual dimorphism in white campion: deletion on the Y chromosome results in a floral asexual phenotype. Genetics **151:** 1187–1196.

Gao, W., Chen, Z.J., Yu, J.Z., Kohel, R.J., Womack, J.E., and Stelly, D.M. (2006). Wide-cross whole-genome radiation hybrid mapping of the cotton (*Gossypium barbadense* L.) genome. Mol. Genet. Genomics **275:** 105–113.

Geraldes, A., Hefer, C.A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R.Y., Stanton, B., Guy, R.D., Mansfield, S.D., Douglas, C.J., and Cronk, Q.C. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (Populus). Mol. Ecol. **24:** 3243–3256.

Henrichsen, C.N., Chaignat, E., and Reymond, A. (2009). Copy number variants, diseases and gene expression. Hum. Mol. Genet. **18:** R1–R8.

Henry, I.M., Dilkes, B.P., and Comai, L. (2006). Molecular karyotyping and aneuploidy detection in *Arabidopsis thaliana* using quantitative fluorescent polymerase chain reaction. Plant J. **48:** 307–319.

Henry, I.M., Dilkes, B.P., Miller, E.S., Burkart-Waco, D., and Comai, L. (2010). Phenotypic consequences of aneuploidy in *Arabidopsis thaliana.* Genetics **186:** 1231–1245.

Henry, I.M., Dilkes, B.P., Young, K., Watson, B., Wu, H., and Comai, L. (2005). Aneuploidy and genetic variation in the *Arabidopsis thaliana* triploid response. Genetics **170:** 1979–1988.

Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T.H., and Comai, L. (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell **26:** 1382–1397.

Horiguchi, G., Gonzalez, N., Beemster, G.T., Inzé, D., and Tsukaya, H. (2009). Impact of segmental chromosomal duplications on leaf size in the grandifolia-D mutants of *Arabidopsis thaliana.* Plant J. **60:** 122–133.

Huettel, B., Kreil, D.P., Matzke, M., and Matzke, A.J. (2008). Effects of aneuploidy on genome structure, expression, and interphase organization in *Arabidopsis thaliana.* PLoS Genet. **4:** e1000226.

IAEA (1977). Manual on Mutation Breeding, 2nd ed. Technical Report Series No. 119 (Vienna, Austria: International Atomic Energy Agency).

Induri, B.R., Ellis, D.R., Slavov, G.T., Yin, T., Zhang, X., Muchero, W., Tuskan, G.A., and DiFazio, S.P. (2012). Identification of quantitative trait loci and candidate genes for cadmium tolerance in Populus. Tree Physiol. **32:** 626–638.

Jiang, K., Liberatore, K.L., Park, S.J., Alvarez, J.P., and Lippman, Z.B. (2013). Tomato yield heterosis is triggered by a dosage sensitivity of the florigen pathway that fine-tunes shoot architecture. PLoS Genet. **9:** e1004043.

Kersten, B., Pakull, B., Groppe, K., Lueneburg, J., and Fladung, M. (2014). The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on P. trichocarpa chromosome 19. Plant Biol. (Stuttg.) **16:** 411–418.

Khush, G.S. (1973). Cytogenetics of Aneuploids. (New York: Academic Press).

Khush, G.S., and Rick, C.M. (1968). Cytogenetic analysis of the tomato genome by means of induced deficiencies. Chromosoma **23:** 452–484.

Kumar, A., et al. (2012). Physical mapping resources for large plant genomes: radiation hybrids for wheat D-genome progenitor Aegilops tauschii. BMC Genomics **13:** 597.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078–2079.

Li, Y., Xiao, J., Wu, J., Duan, J., Liu, Y., Ye, X., Zhang, X., Guo, X., Gu, Y., Zhang, L., Jia, J., and Kong, X. (2012). A tandem segmental duplication (TSD) in green revolution gene Rht-D1b region underlies plant height variation. New Phytol. **196:** 282–291.

Liharska, T.B., Hontelez, J., van Kammen, A., Zabel, P., and Koornneef, M. (1997). Molecular mapping around the centromere of tomato chromosome 6 using irradiation-induced deletions. Theor. Appl. Genet. **95:** 969–974.

Lu, M., Zhang, P., and Kang, X. (2013). Induction of 2n female gametes in Populus adenopoda Maxim by high temperature exposure during female gametophyte development. Breed. Sci. **63:** 96–103.

Ma, T., et al. (2013). Genomic insights into salt adaptation in a desert poplar. Nat. Commun. **4:** 2797.

Mallory, A.C., Dugas, D.V., Bartel, D.P., and Bartel, B. (2004). MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. Curr. Biol. **14:** 1035–1046.

Marron, N., Storme, V., Dillen, S.Y., Bastien, C., Ricciotti, L., Salani, F., Sabatti, M., Rae, A.M., Ceulemans, R., and Boerjan, W. (2010). Genomic regions involved in productivity of two interspecific poplar families in Europe. 2. Biomass production and its relationships with tree architecture and phenology. Tree Genet. Genomes **6:** 533–554.

McClintock, B. (1929). A cytological and genetical study of triploid maize. Genetics **14:** 180–222.

Mock, K.E., Callahan, C.M., Islam-Faridi, M.N., Shaw, J.D., Rai, H.S., Sanderson, S.C., Rowe, C.A., Ryel, R.J., Madritch, M.D., Gardner, R.S., and Wolf, P.G. (2012). Widespread triploidy in Western North American aspen (*Populus tremuloides*). PLoS One **7:** e48406.

Mock, K.E., Rowe, C.A., Hooten, M.B., Dewoody, J., and Hipkins, V.D. (2008). Clonal dynamics in western North American aspen (*Populus tremuloides).* Mol. Ecol. **17:** 4827–4844.

Monson-Miller, J., Sanchez-Mendez, D.C., Fass, J., Henry, I.M., Tai, T.H., and Comai, L. (2012). Reference genome-independent assessment of mutation density using restriction enzyme-phased sequencing. BMC Genomics **13:** 72.

Morita, R., Kusaba, M., Iida, S., Yamaguchi, H., Nishio, T., and Nishimura, M. (2009). Molecular characterization of mutations induced by gamma irradiation in rice. Genes Genet. Syst. **84:** 361–370.

Muntzing, A. (1936). The chromosomes of a giant *Populus tremula.* Hereditas **21:** 383–393.

Naito, K., Kusaba, M., Shikazono, N., Takano, T., Tanaka, A., Tanisaka, T., and Nishimura, M. (2005). Transmissible and non-transmissible mutations induced by irradiating *Arabidopsis thaliana* pollen with gamma-rays and carbon ions. Genetics **169:** 881–889.

Novaes, E., et al. (2009). Quantitative genetic analysis of biomass and wood chemistry of Populus under different nitrogen levels. New Phytol. **182:** 878–890.

Papp, B., Pál, C., and Hurst, L.D. (2003). Dosage sensitivity and the evolution of gene families in yeast. Nature **424:** 194–197.

Pearce, S., et al. (2011). Molecular characterization of Rht-1 dwarfing genes in hexaploid wheat. Plant Physiol. **157:** 1820–1831.

Pécrix, Y., Rallo, G., Folzer, H., Cigna, M., Gudin, S., and Le Bris, M. (2011). Polyploidization mechanisms: temperature environment can induce diploid gamete formation in Rosa sp. J. Exp. Bot. **62:** 3587–3597.

Pfahler, P.L. (1967). Fertilization ability of maize pollen grains. III. Gamma irradiation of mature pollen. Genetics **57:** 523–530.

Rohde, A., et al. (2011). Bud set in poplar–genetic dissection of a complex trait in natural and hybrid populations. New Phytol. **189:** 106–121.

Rudolph, T.D. (1978). Seed yield and quality in *Populus tremuloides* following pollination with gamma-irradiated pollen. Can. J. Bot. **56:** 2967–2972.

Sato, Y., Shirasawa, K., Takahashi, Y., Nishimura, M., and Nishio, T. (2006). Mutant selection from progeny of gamma-ray-irradiated rice by DNA heteroduplex cleavage using *Brassica* petiole extract. Breed. Sci. **56:** 179–183.

Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana.* Development **125:** 3329–3341.

Shirley, B.W., Hanley, S., and Goodman, H.M. (1992). Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis transparent testa mutations. Plant Cell **4:** 333–347.

Sjödin, A., Street, N.R., Sandberg, G., and Gustafsson, P. (2009). The Populus Genome Integrative Explorer (PopGenIE): a new resource for exploring the Populus genome. New Phytol. **182:** 1013–1025.

Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N., and Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat. Biotechnol. **23:** 75–81.

Stadler, L.J., and Roman, H. (1948). The effect of x-rays upon mutation of the gene A in maize. Genetics **33:** 273–303.

Stankiewicz, P., and Lupski, J.R. (2010). Structural variation in the human genome and its role in disease. Annu. Rev. Med. **61:** 437–455.

Stanton, B.J., and Villar, M. (1996). Controlled reproduction of *Populus*. In Biology of *Populus* and Its Implications for Management and Conservation, R.F. Stettler, H.D. Bradshaw, Jr., P.E. Heilman, and T.M. Hinckley, eds (Ottawa, Canada: NRC Research Press), pp. 113–138.

Tsai, H., Missirian, V., Ngo, K.J., Tran, R.K., Chan, S.R., Sundaresan, V., and Comai, L. (2013). Production of a high-efficiency TILLING population through polyploidization. Plant Physiol. **161:** 1604–1614.

Tuskan, G.A., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596–1604.

Tuskan, G.A., et al. (2012). The obscure events contributing to the evolution of an incipient sex chromosome in Populus: a retrospective working hypothesis. Tree Genet. Genomes **8:** 559–571.

Veitia, R.A., Bottani, S., and Birchler, J.A. (2013). Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. Trends Genet. **29:** 385–393.

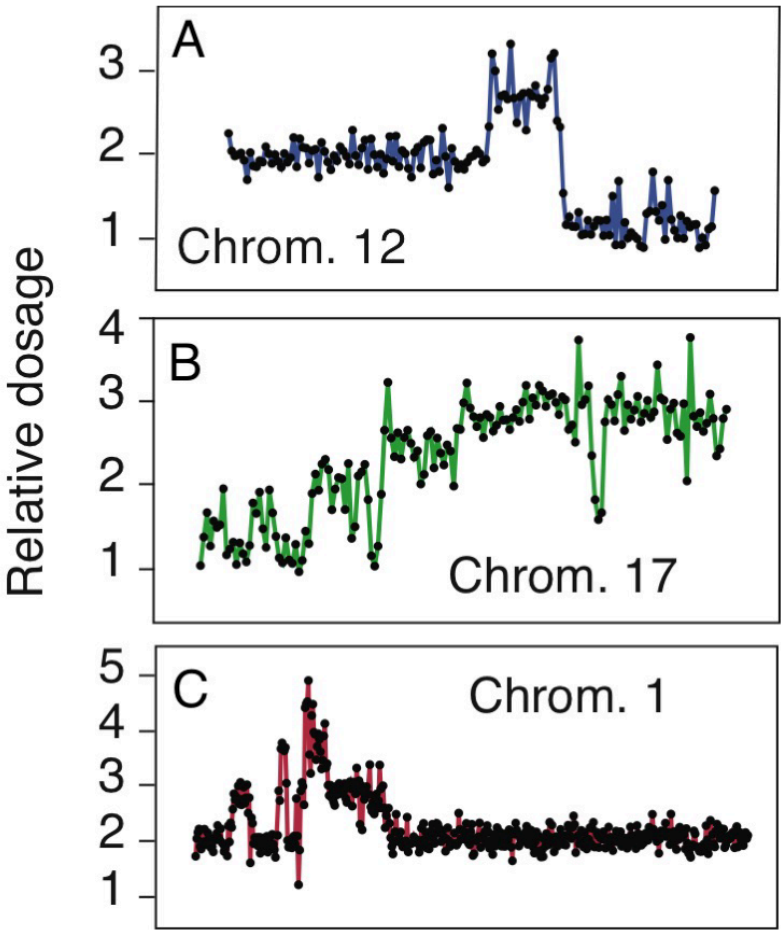
Vizir, I.Y., Anderson, M.L., Wilson, Z.A., and Mulligan, B.J. (1994). Isolation of deficiencies in the Arabidopsis genome by gamma-irradiation of pollen. Genetics **137:** 1111–1119.

Vizir, I.Y., and Mulligan, B.J. (1999). Genetics of gamma-irradiation-induced mutations in *Arabidopsis thaliana*: large chromosomal deletions can be rescued through the fertilization of diploid eggs. J. Hered. **90:** 412–417.

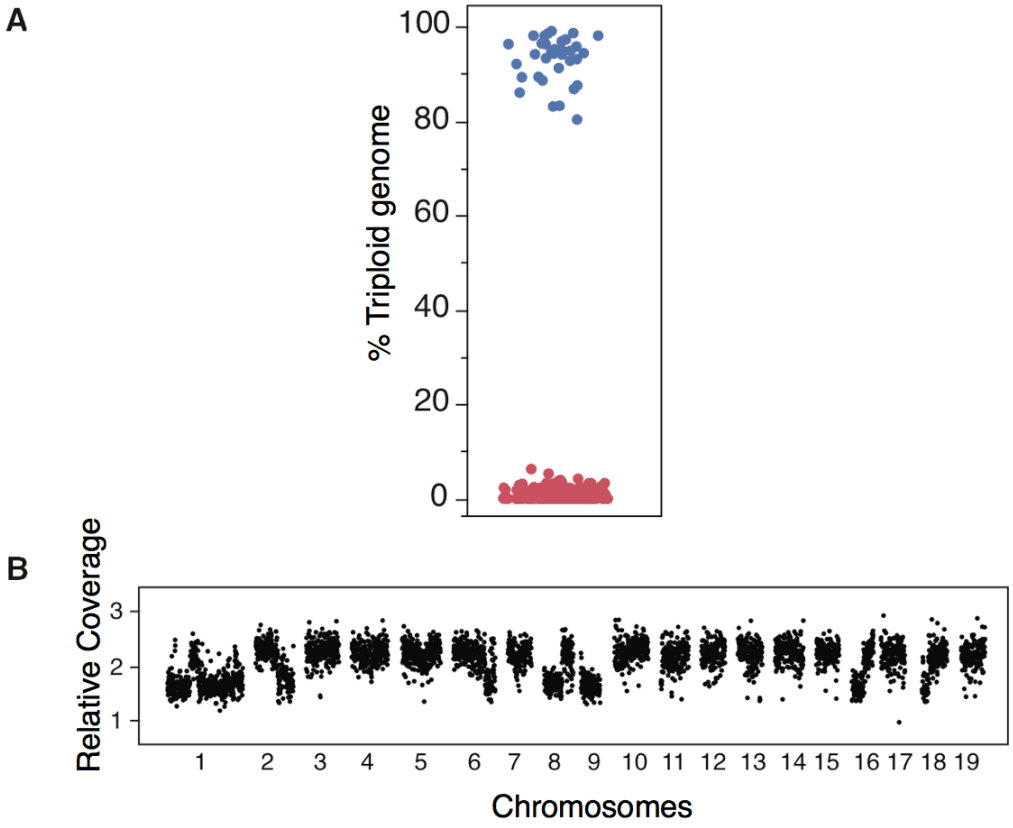
Wingen, L.U., Münster, T., Faigl, W., Deleu, W., Sommer, H., Saedler, H., and Theißen, G. (2012). Molecular genetic basis of pod corn (Tunicate maize). Proc. Natl. Acad. Sci. USA **109:** 7115–7120.

Yao, H., Dogra Gray, A., Auger, D.L., and Birchler, J.A. (2013). Genomic dosage effects on heterosis in triploid maize. Proc. Natl. Acad. Sci. USA **110:** 2665–2669.

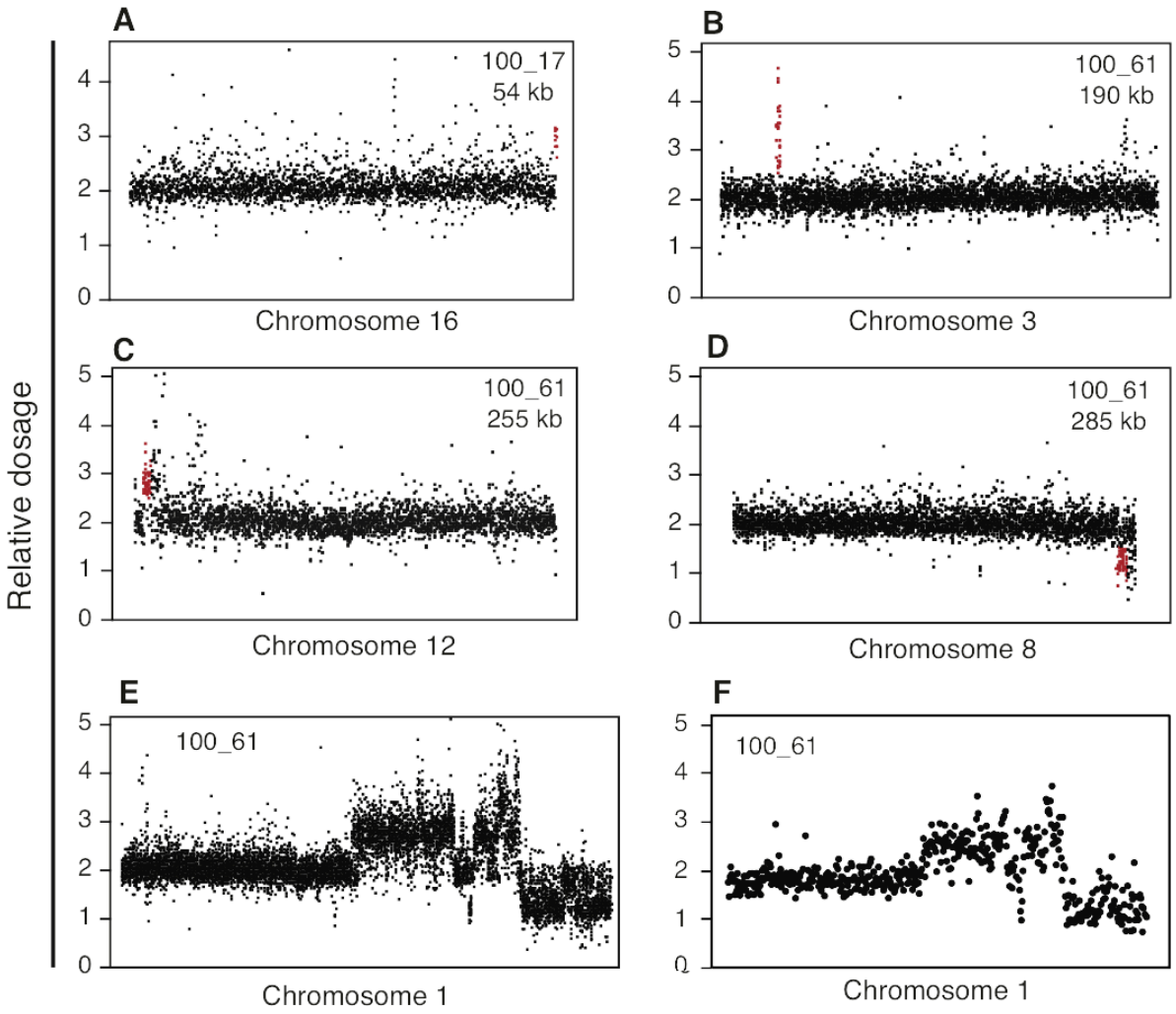
Yuan, T.Q., Sun, S.N., Xu, F., and Sun, R.C. (2011). Structural characterization of lignin from triploid of *Populus tomentosa* Carr. J. Agric. Food Chem. **59:** 6605–6615.



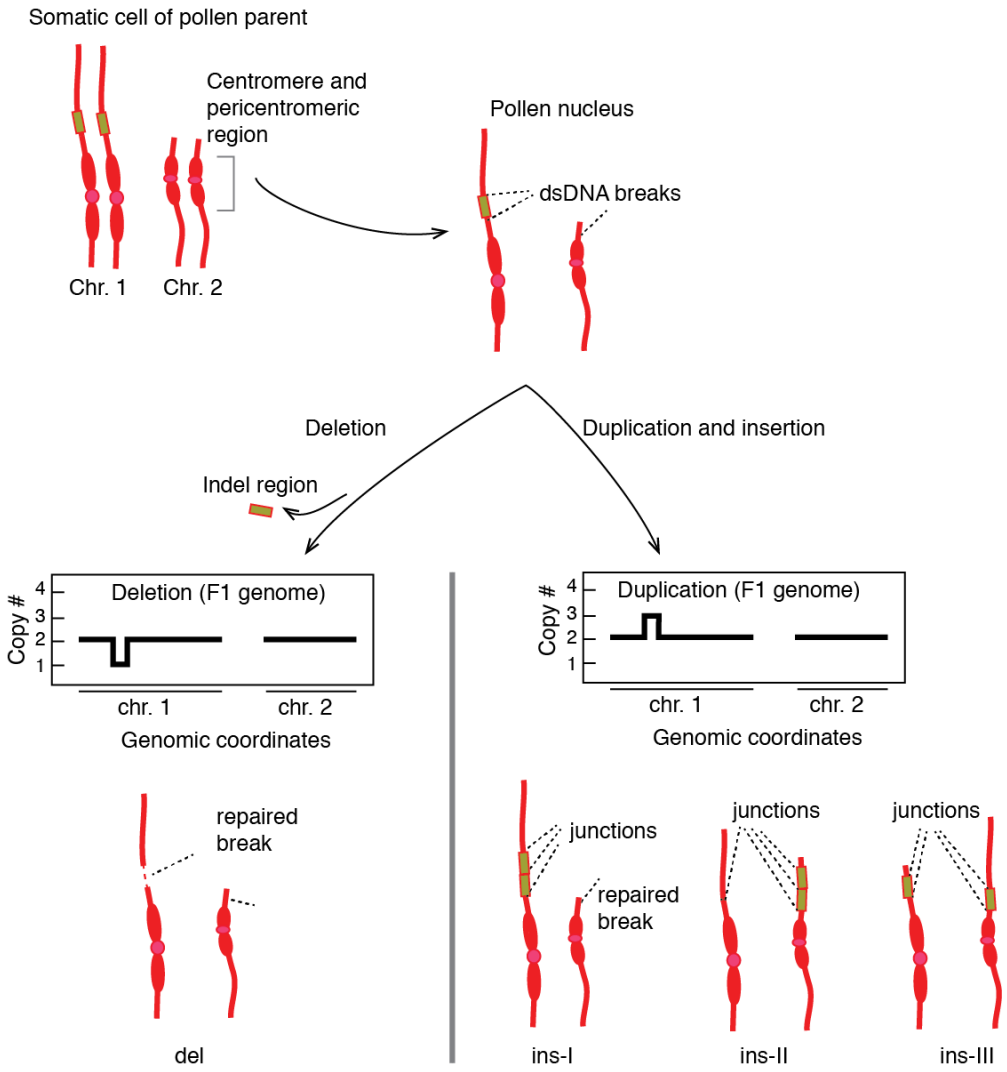
Supplemental Figure 1. Example of unusual dosage patterns. Sequencing reads were aligned to the reference sequences and sorted into consecutive bins of 100 kbs, along the 19 chromosomes present in the current version of the poplar genome. For each of three samples, only one chromosome is represented. Each dot represents the average sequence read coverage for a 100 kb piece of chromosome. For ease of visualization, relative sequence read coverage values are normalized such that fragments present in two copies oscillate around 2.0. Changes up or down in relative sequence read coverage indicate copy number variation and correspond to the presence of an additional copy (insertion) or the deletion of a copy of a particular chromosome fragment, respectively. **A.** Direct switch between increased and decreased dosage within chromosome 12. **B and C:** For both, high frequency of dosage changes within a short chromosomal section and exhibiting unexpected copy numbers.



Supplemental Figure 2. Percentage of the genome in triploid state per individual. A. For each individual exhibiting indel(s), the percentage of the genome present in three copies was calculated. The samples fell into two clear categories: < 10% triploid or > 80% triploid. **B.** Dosage analysis of the triploid individuals with the highest percentage of genome in two copies (80.3% triploid genome) is consistent with several deletions superimposed onto a triploid background. These results support the hypothesis that the individuals with high percentage of triploid genome originated from whole-genome duplication and not from the presence of many insertions events in a diploid background. Whether the whole-genome duplication originated from a meiotic error forming a 2N gamete or resulted from a mitotic error during or after gamma irradiation is unknown.



Supplemental Figure 3. High sequence read coverage data from eleven samples reveals few additional indel mutations. High coverage of sequencing reads (115 million 50 bp reads per sample) were aligned to the reference sequences and sorted into consecutive bins of 5 kbs, along the 19 chromosomes present in the current version of the poplar genome (v3.0). Each dot represents the normalized sequence read coverage for a 5 kb piece of chromosome. For ease of visualization, relative sequence read coverage values are normalized such that fragments present in two copies oscillate around 2.0. Changes up or down in relative sequence read coverage indicate copy number variation and correspond to the presence of an additional copy (insertion) or the deletion of a copy of a particular chromosome fragment respectively. **A-D**. New indel mutations, not previously identified at low sequence read coverage, are depicted in red. **E-F**. A complex series of dosage changes was visible at low sequence read coverage (F) but is better defined at high sequence read coverage (E).



Supplemental Figure 4. Potential chromosomal rearrangements consistent with the observed dosage plots. Upon gamma irradiation of the *P. nigra* pollen grains, double stranded breaks are created throughout the genome. Two breaks close to each other can create the isolation and subsequent loss of a copy of the intervening fragment (deletion mutation). Alternatively, the intervening fragment can be replicated via a mechanism to be determined, possibly when the germinative cell undergoes one last mitotic division during pollen tube growth, allowing the pollen to transition from binucleate to trinucleate. In this case, each of the two resulting copies can be re-inserted either at the original locus or at a different locus. Several of these scenarios are consistent with each observed dosage plot and the data currently available does not allow us to distinguish between them.

Supplemental Table 1: Outcome of *P. deltoides* x *P. nigra* interspecific crosses, using gamma-irradiated pollen.

Irradiation dose (Gy)* **	**0**	**50**	**100**	**125**	**150**	**200**	**250**
Viable seedlings	76/196 (39%) **	22/98 (22%) *	57/196 (29%) **	No seed *	No seed *	No seed * / **	No seed *

Crosses were performed using two different female parents in two different locations: **P. deltoides* SO598SL x *P. nigra* SO361SL performed at Greenwood Resources, OR. ** *P. deltoides* SO546SL x *P. nigra* SO361SL performed at the Institute of Forest Genetics, CA. Pollen was irradiated at various levels (expressed in Grays). For crosses for which seeds were obtained, the following statistics are indicated: number of viable seedlings / total number of seeds collected (% viable seedlings). Crosses that did not produce any viable seeds are labelled as "No seed". ***: Irradiation doses were measured in Grays (Gy). 1 Gy corresponds to the absorption of one joule of radiation energy per kilogram of matter.

Supplemental Table 2: Summary of the number of F1 hybrid progeny tested and the number of triploids detected.

Sample Type	DDN	DN	DNN	Total	% DD gamete	% NN gamete
GWR_100*	2	298	34	334	0.6	10.2
GWR_50*	0	61	0	61	0.0	0.0
IFG_0**	0	14	0	14	0.0	0.0
IFG_100**	1	117	2	120	0.8	1.7
Total	**3**	**490**	**36**	**529**	**0.6**	**6.8**

Crosses were performed using two different female parents in two different locations: **P. deltoides* SO598SL x *P. nigra* SO361SL performed at Greenwood Resources, OR. ** *P. deltoides* SO546SL x *P. nigra* SO361SL performed at the Institute of Forest Genetics, CA.

Supplemental Table 3: Outcome of mutation detection.

	IFG_0_1 (Non-irradiated)	IFG_100_28	IFG_100_17	IFG_100_49	IFG_100_61
		100 Gy irradiated samples			
Homozygous mutations					
Space Assayed (Mb)	92.45	91.79	92.64	91.6	93.38
Mutations	0	0	0	0	0
Mutation rate	0	0	0	0	0
Heterozygous mutations					
Space Assayed (Mb)	30.23	29.59	27.30	31.90	31.61
Mutations	5	3	1	3	3
Mutation rate (Mut / Mb)	0.17	0.10	0.04	0.58	0.58
Exp. Number of mutations with functional effect (genome-wide)*	1	1	0	1	1

* The number of expected deleterious mutations per genome was calculated based on the observed mutation rate, an estimated exon space of 47.8 Mb (based on the annotation of the *P. trichocarpa* genomic reference v3.0) and the expectation that, on average, 15.4% of mutations within exons are expected to functionally affect a gene, based on previous studies (Henry et al., 2014).

Supplemental File 1: List of indel mutations in the F1 population.

Supplemental File 2: List of polymorphic positions between the *P. nigra* pollen parent *(*SO361SL) and the *P. deltoides* female parent (SO598SL. Columns represent the chromosome, position (bp) and reference sequence genotype followed by the *P. deltoides* allele and finally the *P. nigra* allele at that position.

Supplemental File 3: List of polymorphic positions between the *P. nigra* pollen parent *(*SO361SL) and the *P. deltoides* female parent (SO546SL. Columns represent the chromosome, position (bp) and reference sequence genotype followed by the *P. deltoides* allele and finally the *P. nigra* allele at that position.

Supplemental references

Henry, I.M., Nagalakshmi, U., Lieberman, M.C. *et al.* (2014) Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. *The Plant Cell Online*

Slavov, G.T., DiFazio, S.P., Martin, J. *et al.* (2012) Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree Populus trichocarpa. *New Phytol* **196**, 713–725.