A simple method to efficiently generate structural variation in plants

Lindsey L. Bechen[1], Naiyara Ahsan[2,4], Alefiyah Bahrainwala[2], Mary Gehring[1,3,5*], PRV Satyaki[1,2,4*]

1 Whitehead Institute for Biomedical Research, Cambridge, MA, USA

2 Dept. of Biological Sciences, University of Toronto Scarborough, Toronto, ON, Canada

3 Dept. of Biology, Massachusetts Institute of Technology, Cambridge, MA USA

4 Dept. of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada

5 Howard Hughes Medical Institute

*co-corresponding authors

mgehring@wi.mit.edu

satyaki.rajavasireddy@utoronto.ca

Key Words: structural variation, etoposide, Arabidopsis, mutagenesis

Abstract

Phenotypic variation is essential for the selection of new traits of interest. Structural variants, consisting of deletions, duplications, inversions, and translocations, have greater potential for phenotypic consequences than single nucleotide variants. Pan-genome studies have highlighted the importance of structural variation in the evolution and selection of novel traits. Here, we describe a simple method to induce structural variation in plants. We demonstrate that a short period of growth on the topoisomerase II inhibitor etoposide induces heritable structural variation and altered phenotypes in *Arabidopsis thaliana* at high frequency. Using long-read sequencing and genetic analyses, we identified deletions and inversions underlying semi-dominant and recessive phenotypes. This method requires minimal resources, is potentially applicable to any plant species, and can replace irradiation as a source of induced large-effect structural variation.

Introduction

Genomic structural variation (SV) – insertions, deletions, duplications, translocations and inversions – are an important source of genetic and phenotypic novelty. Structural variants create genetic novelty by multiple means, including fusing the coding regions of two genes, altering cis-regulatory environments and gene expression patterns, and by altering or suppressing recombination, among other mechanisms. In addition, deletions and duplications alter gene or chromatin dosage[1–3]. Novel structural variants have larger phenotypic effects per mutation than single nucleotide variants [4] and have been selected for when they underlie traits beneficial to the organism or desirable to agriculturists [5,6].

Recent pan-genome sequencing of multiple plant species has underscored the versatility and significance of structural variation to phenotypic variation, modern crop traits, and crop improvement. Lower seed sets in watermelons and grapes, important for the development of popular “seedless” varieties, have been linked to large inversions that cause meiotic defects [7,8]. SVs have also been linked to the domestication of broomcorn millet [9] and sorghum [10]. In pearl millets, gain of heat tolerance has been linked to SVs at multiple loci [11]. SVs are also associated with variation in immunity and plant defense responses [12–14]. In Arabidopsis, a multi-enzyme pathway synthesizes the glucosinolate family of defense compounds. SVs that either delete or fuse enzyme-encoding paralogs modify enzymatic pathways and are responsible for ecotype-specific variation in glucosinolates [15]. Studies of *O. sativa* provide numerous examples of how SVs influence phenotypes. Change in copy number of *VIL1* is associated with changes to flowering time and grain number variation [16]. In addition, duplications of the *KALA4* gene – a regulator of the anthocyanin biosynthesis pathway – created a novel cis-regulatory region that promoted ectopic expression of anthocyanins and caused a “black pericarp” phenotype [17]. In peaches, fruit flesh color around the stone and fruit shape are dependent on a deletion and a mega-base scale inversion [18]. Modern varieties of sweet

corn were created by selecting for an inversion that creates a loss-of-function mutation in the *shrunken-2* gene, which encodes the first enzyme of the starch biosynthesis pathway [19,20]. In tomatoes, a duplication of a cytochrome P450 gene was linked to increased fruit weight [5].

Extant crop genetic diversity is unlikely to provide all possible SV diversity that could impact crop improvement. Induced structural variation can, therefore, provide: novel germplasm for breeding programs, information that can be used for targeted CRISPR-mediated restructuring of crop genomes, and novel genetic variants for understanding fundamental aspects of plant biology. The most common method employed to induce random structural variation is exposure to ionizing radiation, including X-rays, gamma irradiation, and heavy ions [21–26]. Since the 1960's, irradiation has been used to induce structural variation in model organisms and in numerous crops including, more recently, *B. napus* [27], wheat [28], cotton [29], rice [23,30], poplar [22,31], soybean [32,33], and buckwheat [34]. Populations carrying such structural variation have been used for the isolation of mutations that abolish hybrid sterility in rice [35] and the creation of seeds differing in oil composition [21,27].

Although radiation-induced structural variation remains an important element of modern plant breeding, there are logistical challenges. In contrast to the induction of single nucleotide variants that researchers can conveniently pursue in their own labs using ethyl methanesulfonate (EMS), the induction of structural variants requires access to radiation sources. Radiation sources are heavily regulated and access to nuclear reactors, particle accelerators, and other sources of radiation is a bottleneck in the creation of structural variant libraries [26]. Additionally, application of irradiation can be tedious. For example, the creation of a mutant population in poplar required irradiation of collected and dried pollen [22]. One alternative is the extended TAQing approach, in which a transgene expresses a restriction enzyme to induce conditional double-stranded DNA breaks, which then leads to shuffling of the plant genome [36]. However,

this approach can be challenging as the commercial varieties of many crops are not easily transformed.

To surmount the current challenges of creating SVs in plants, we sought a simple method that could be applied to any plant species of interest to induce SVs with high frequency. We drew inspiration from the ubiquity and ease of EMS chemical mutagenesis to create an inexpensive, drug-based system that can be used in labs with minimal resources to induce SVs in plants. Here, we provide proof-of-principle that the chemotherapeutic drug etoposide efficiently generates novel genomic structural variation in *Arabidopsis thaliana*.

Results

Treatment with etoposide results in mutants with a variety of heritable phenotypes

We tested whether a DNA topoisomerase II (Topo II) inhibitor could induce structural variation in *Arabidopsis thaliana*. Topoisomerase II relaxes torsional stress from DNA supercoiling generated during DNA replication or transcription by transiently breaking both strands and then ligating them after passing a DNA segment through the break. Between strand breakage and ligation, Topo II is covalently linked to DNA via a tyrosine residue, forming a topoisomerase cleavage complex [37]. This complex is stabilized by the inhibitor etoposide. A collision between covalently-linked Topo II and DNA polymerases during DNA replication, or with RNA polymerases during transcription, leads to removal of the Topo II enzyme, which results in the generation of double-stranded breaks (DSBs) [38–41]. The imprecise repair of DSBs leads to genomic rearrangements and structural variation in mouse spermatocytes, fibroblasts, and in human cells [42–44]. Previously, it was shown that treatment with etoposide inhibits plant growth[45,46] and causes fragmentation of chromosomes during meiosis in Arabidopsis [45]. However, its potential as a mutagen that can induce structural variation has not been

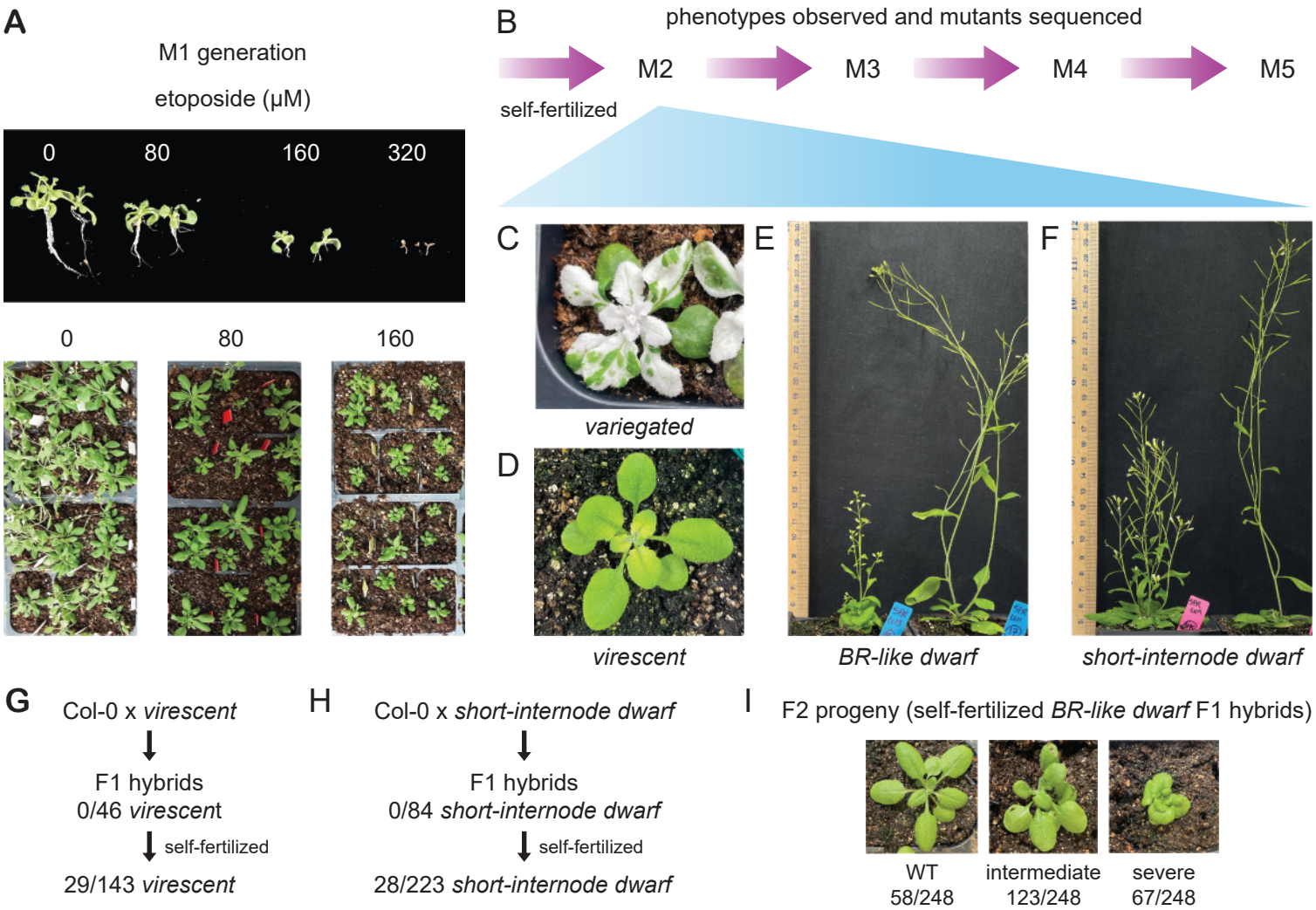

Figure 1: Etoposide induces novel heritable phenotypes. (A) Arabidopsis seeds germinated on increasing concentrations of etoposide show a dose-responsive growth defect. Seedlings grown on 320 µM etoposide did not exhibit true leaves. This generation is referred to as M1. Growth defects persisted after transplantation to soil. **(B)** M1 was self-fertilized to give rise to M2. Subsequent roundsof self-fertilization produced M3, M4, and M5 generations. A visual examination of M2 generation identified multiple phenotypes, including: **(C)** a variegated phenotype, **(D)** a virescent phenotype with yellowish leaves, **(E)** a brassinosteroid-like dwarf phenotype, and **(F)** a fertile dwarf with short internodes. **(G-I)** describe crosses to the wild-type and subsequent selfing to identify the nature of inheritance of the mutant phenotypes and the number of contributing loci. **(G)** The *virescent* phenotype is recessive and caused by a single locus. **(H)** The *short-internode dwarf* phenotype is recessive and is likely caused by more than one locus. **(I)** The *BR-like dwarf* phenotype is incompletely dominant and displays three phenotypic classes. F2 segregation data suggests that this phenotype is caused by mutation at a single locus.

investigated. To test whether etoposide can induce heritable structural variation in plants, we examined the effects of prolonged exposure to etoposide on Arabidopsis. Wild-type Col-0 seeds were germinated and grown on MS media supplemented with sucrose and differing concentrations of etoposide (0 µM (DMSO only), 40 µM, 80 µM, and 160 µM) for up to two weeks (Figure 1A). Plants grown on DMSO alone showed no observable growth defects. Two-week-old seedlings grown on 40 µM etoposide exhibited gnarled leaves but no differences in size. Seedlings grown on 80 µM etoposide showed marginally reduced growth. Of the seedlings grown on 160 µM of etoposide, less than half (75/184) developed true leaves. In addition, these plants displayed stunted root development (Figure 1A). Seedlings grown on DMSO only, 80 µM, and 160 µM etoposide (referred to as M1 plants) were transplanted to soil and compared at maturity. Those exposed to 160 µM of etoposide exhibited significantly more abnormal phenotypes than DMSO only or 80 µM etoposide plants, including loss of apical dominance, gnarled leaves, reduced plant size, seed abortion, and lower seed number at maturity (Figure 1A). To test if these abnormal plants produced progeny with mutant phenotypes, we scored for six visible phenotypes – dwarf architecture or loss of apical dominance, seed size or abortion, leaf shape and size, flowering time, and leaf pigmentation – among the progeny (the M2 generation) of plants treated with 160 µM etoposide. Of M2 progeny derived from 42 different M1 parents, 29 lines exhibited at least one obvious visible phenotype (Figure 1B, Table S1). These phenotypes included: variegated or albino plants, altered flowering time, altered leaf shape and size (Figure S1, Table S1), sterility or seed abortion, and dwarfism. The large proportion of plants showing visible phenotypes suggested that etoposide is an excellent mutagen for efficiently creating large-effect mutations.

To assess the heritability of the mutant phenotypes and to establish genotype-phenotype relationships, we closely examined four M1 lines with visible phenotypes. These included: a line with a *variegated* phenotype (line 5A) (Figure 1C); a line with delayed greening, or *virescent*

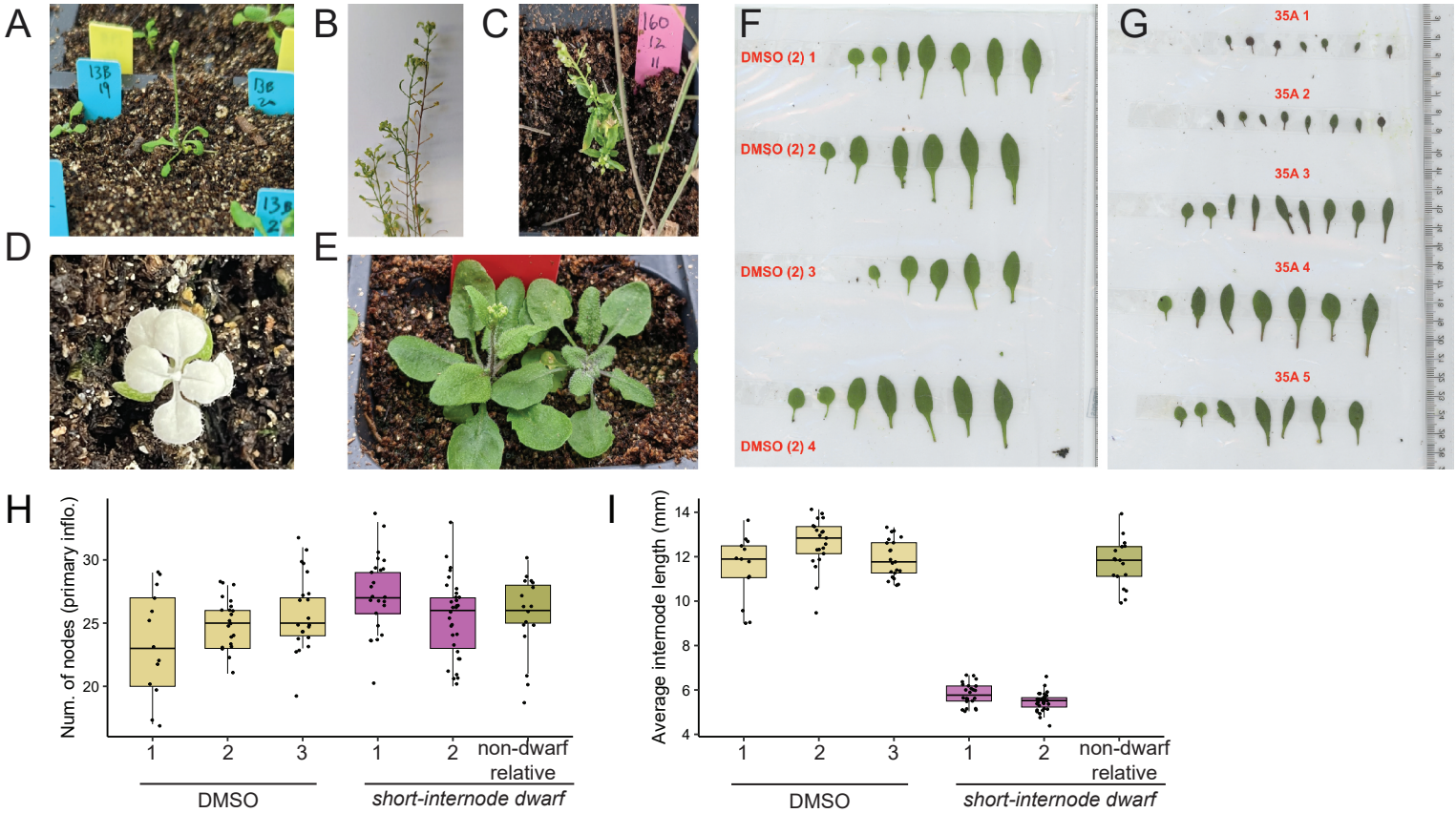

Figure S1: Additional phenotyping of progeny of etoposide-treated plants. Visual observation of M2 plants arising from different etoposide-treated M1 progenitors identified multiple lines with mutant phenotypes, including: early flowering **(A)**, sterility **(B)**, dwarfism **(C)**, variegation **(D)**, altered leaf shape **(E-G) .** Lines with altered leaf shape or size were visually identified. These differences were confirmed by measuring leaf size (blade area, length, perimeter, width, and petiole length) and shape (cirularity) of leaves with the LeafJ plugin for ImageJ software. Leaves from multiple M2 siblings of line 35 are shown in **(G)**. 35A_1 and 35A_2 show a dramatic decrease in size parameters while 35A_3 shows an intermediate decrease. 35A_4 and 5 show no change compared to DMSO-treated lineages. **(H)** *Short-internode dwarf* plants have the same number of nodes as DMSO-treated lineages and a genetic relative that lacks the phenotype, but **(I)** decreased distance between nodes, as measured on the primary inflorescence.

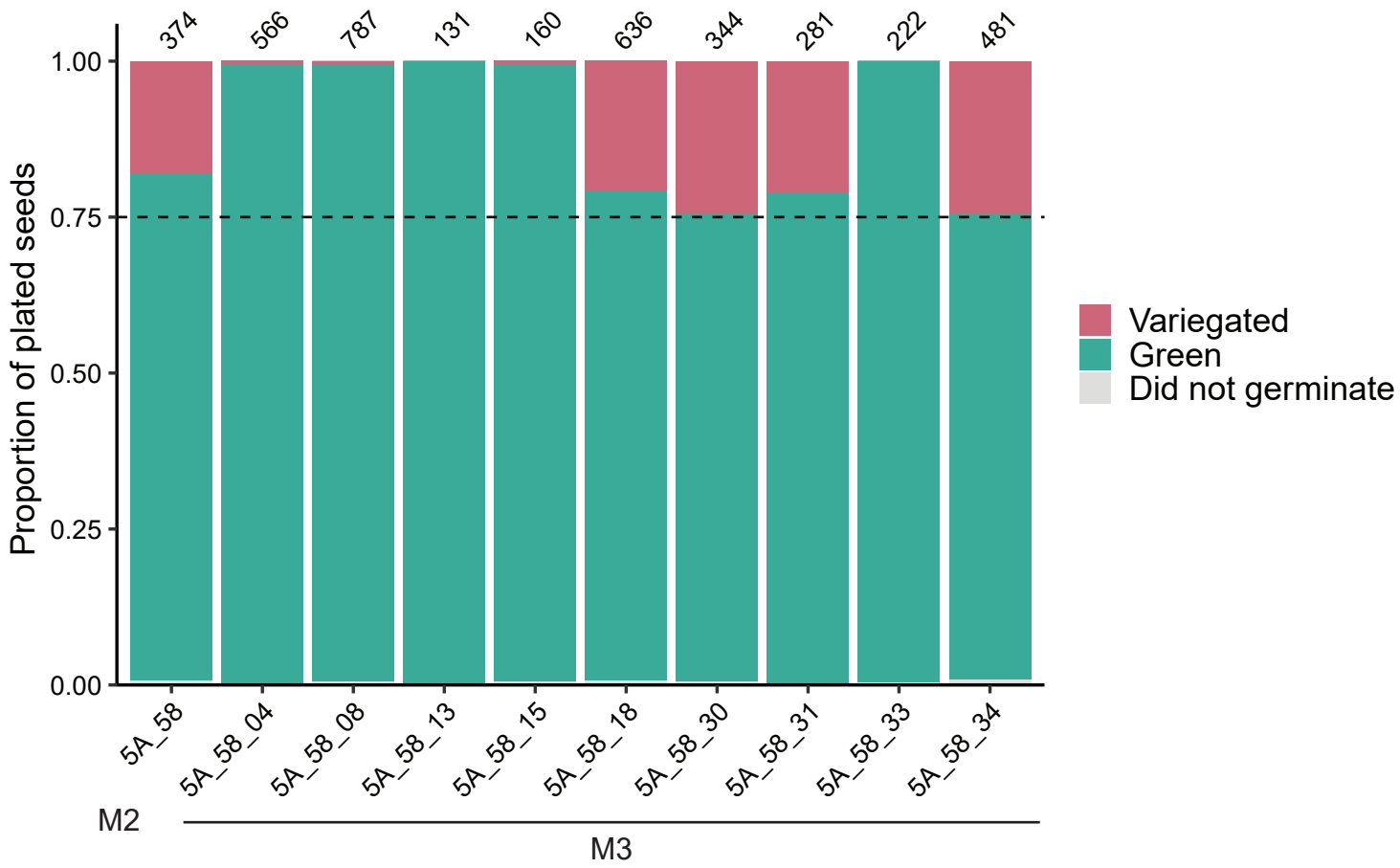

Figure S2: Segregation of the *variegated* phenotype from green parents. Green plants from variegated line 5A produce either approximately 25% variegated progeny or all green progeny. Number of seedlings assesed is annotated above each bar. M2 plant 5A_58 is the parent of the M3 progeny assayed.

(line 13B) (Figure 1D); a dwarf line reminiscent of weak brassinosteroid-insensitive mutants (Figure 1E) [47], referred to as *brassinosteroid-like (BR-like) dwarf* (line 1A); and a dwarf line with short internodes, or *short-internode dwarf* (line 34C) (Figure 1F). *Variegated* mutants displayed light- and temperature-dependent variations in the proportions of green and white sectors on cotyledons and true leaves (Figure 1C). In the *virescent* line, juvenile plants exhibited reduced chlorophyll pigmentation, whereas adult plants exhibited normal pigmentation (Figure 1D). The semi-sterile *BR-like dwarf* line (Figure 1E) exhibited short, thick stems that bore fleshy leaves. In contrast, the *short-internode dwarf* mutant, though short in stature, was fertile, produced the same number of nodes on the primary inflorescence as WT plants (Figure S1H,I), and lacked the fleshy leaf phenotype observed in the *BR-like dwarf* (Figure 1E,F).

The *BR-like dwarf, short-internode dwarf,* and *virescent* phenotypes were transmitted from self-fertilized parents to offspring for at least three additional monitored generations, indicating that the phenotype was true-breeding and the underlying mutant genotype was stable (Figure 1B). To assess whether these phenotypes were dominant or recessive, mutant lines were crossed to wild-type. The resultant F1 progeny from the *virescent* line and the *short-internode dwarf* displayed a wild-type phenotype, indicating that the mutant phenotypes were recessive (Figure 1G, H). The F1 progeny from the *BR-like dwarf* exhibited intermediate dwarfing, suggesting that the phenotype was incompletely dominant (Figure 1I). *Variegated* plants did not flower readily, so this phenotype was maintained as a heterozygous stock for the same number of generations. We observed that when self-fertilized, plants from this line either produced all green progeny or one-quarter variegated progeny, indicating that this phenotype is recessive (Figure S2).

To estimate the number of loci contributing to the *BR-like dwarf*, *short-internode dwarf*, and *virescent* phenotypes, we examined F2 progeny obtained by self-fertilizing the F1 plants. For the *virescent* line, 20.9% of the F2 progeny displayed the mutant phenotype (Figure 1G). This is consistent with the phenotype being caused by a mutation at a single locus (29:114

virescent:wild-type, H_0=1:3, χ^2= 1.69; df=1, 0.25>*p*>0.1). The *short-internode dwarf* phenotype was observed in 1/8 of progeny obtained by self-fertilizing F1 plants (28:195 *short-internode dwarf*:wild-type) (Figure 1H). This ratio is consistent with the recessive phenotype being caused by mutation in two linked loci, among other possibilities. Of F2 progeny of *BR-like dwarf* plants with an intermediate dwarfing phenotype, 23.4% had a wild-type phenotype, 49.6% exhibited an intermediate phenotype, and 27.0% had a severe mutant phenotype (Figure 1I). This result suggested that a single incompletely dominant locus causes the *BR-like dwarf* phenotype (58:123:67 severe:intermediate:wild-type, H_0=1:2:1 ratio, χ^2=0.669; df=2, *p*>0.5). These observations indicate that etoposide mutagenesis can efficiently create novel recessive and dominant phenotypes caused by large-effect alteration to one or more loci.

Etoposide treatment induces a spectrum of structural variation types

To characterize the molecular nature of etoposide-induced mutations, we employed Illumina short-read sequencing technology to sequence the genomes of 32 M2 or M3 progeny of 15 etoposide-treated M1 plants (Table S2). Sequenced plants included those with no scored mutant phenotype and those expressing a phenotype, including *virescent*, *BR-like dwarf*, and *short-internode dwarf*, among others. We also sequenced near-isogenic genetic relatives of these three mutants that displayed wild-type phenotypes (see Table S3 for relationships). To enhance the accuracy of our structural variation (SV) calls, we identified potential SVs present in lab Col-0 populations by sequencing four M2 plants descended from four control M1 lines treated only with DMSO.

To identify etoposide-induced SVs, we first adapted a computational approach that uses genome-wide coverage to identify large segmental deletions or duplications [48,49]. Using this technique, we identified 10 duplications and 1 deletion that were between 100 kb and 1 Mb in

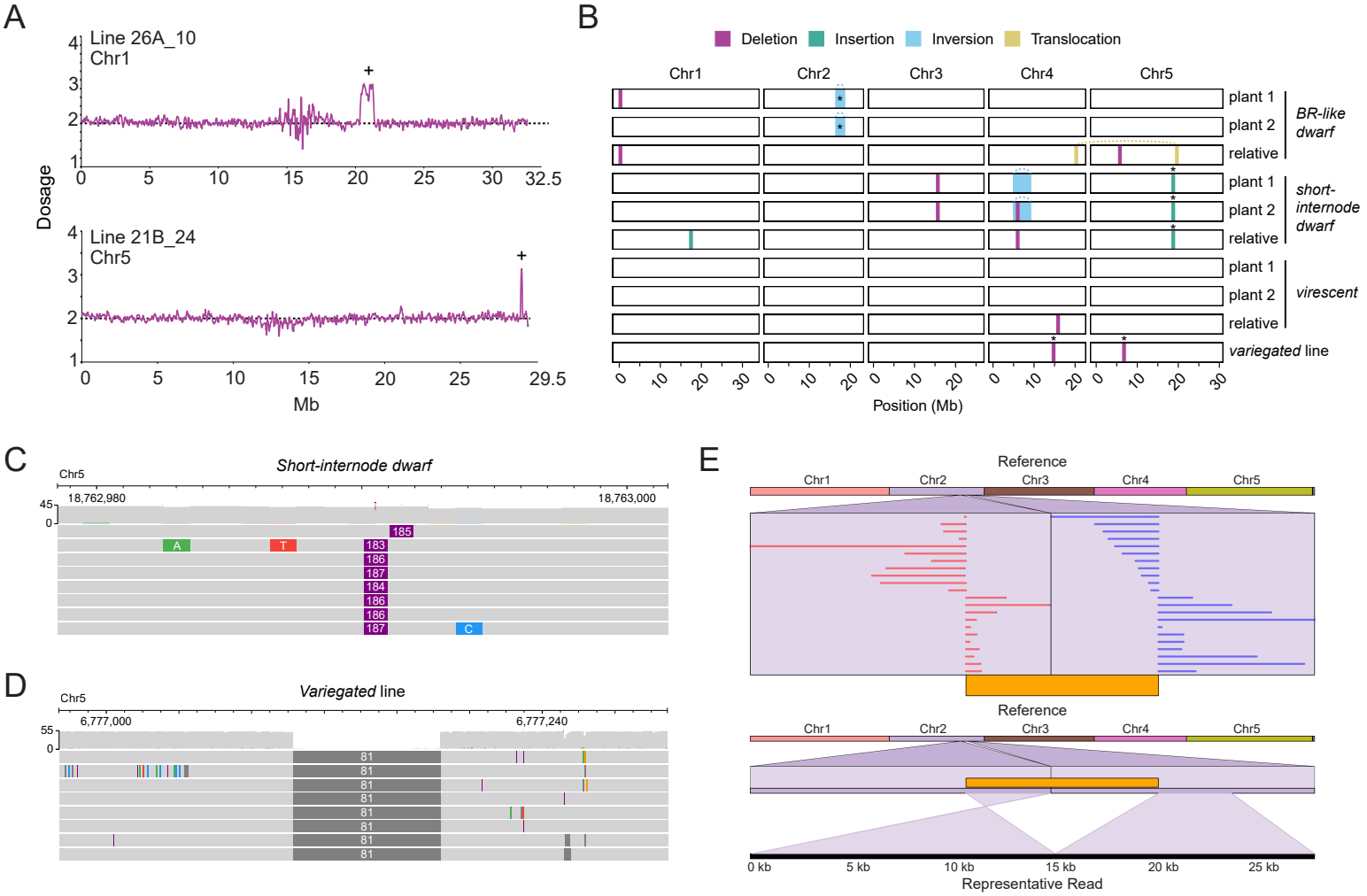

Figure 2: Whole genome sequencing identifies structural variation in progeny of etoposide treated plants. (A) Duplications in etoposide-treated lines. Dosage across the indicated chromosome is shown. Chromosomal segments with one extra copy are indicated with +. **(B)** Nanopore long-read sequencing identifies structural variation in lines that show the *BR-like dwarf*, *short-internode dwarf*, *virescent* and *variegated* phenotypes. SVs verified via PCR are indicated with *. Putative causal mutations were identified by sequencing two plants with a phenotype and one relative without a phenotype. Only one mutant plant was sequenced for the *variegated* phenotype. This plant was green but produced ~25% *variegated* progeny. Three PCR-verified mutations in (B) are described in (C-E). **(C)** A genome browser snapshot with reads identifying a 184 bp insertion on Chr5 in the *short-internode dwarf* line. **(D)** Reads in the genome browser snapshot indicate an 81 bp deletion on Chr5 in the *variegated* line. **(E)** A ribbon plot with alignments indicating a 1.52 Mb chromosome 2 inversion in the *BR-like dwarf* line (top) and a detailed view of a representative read spanning the right breakend.

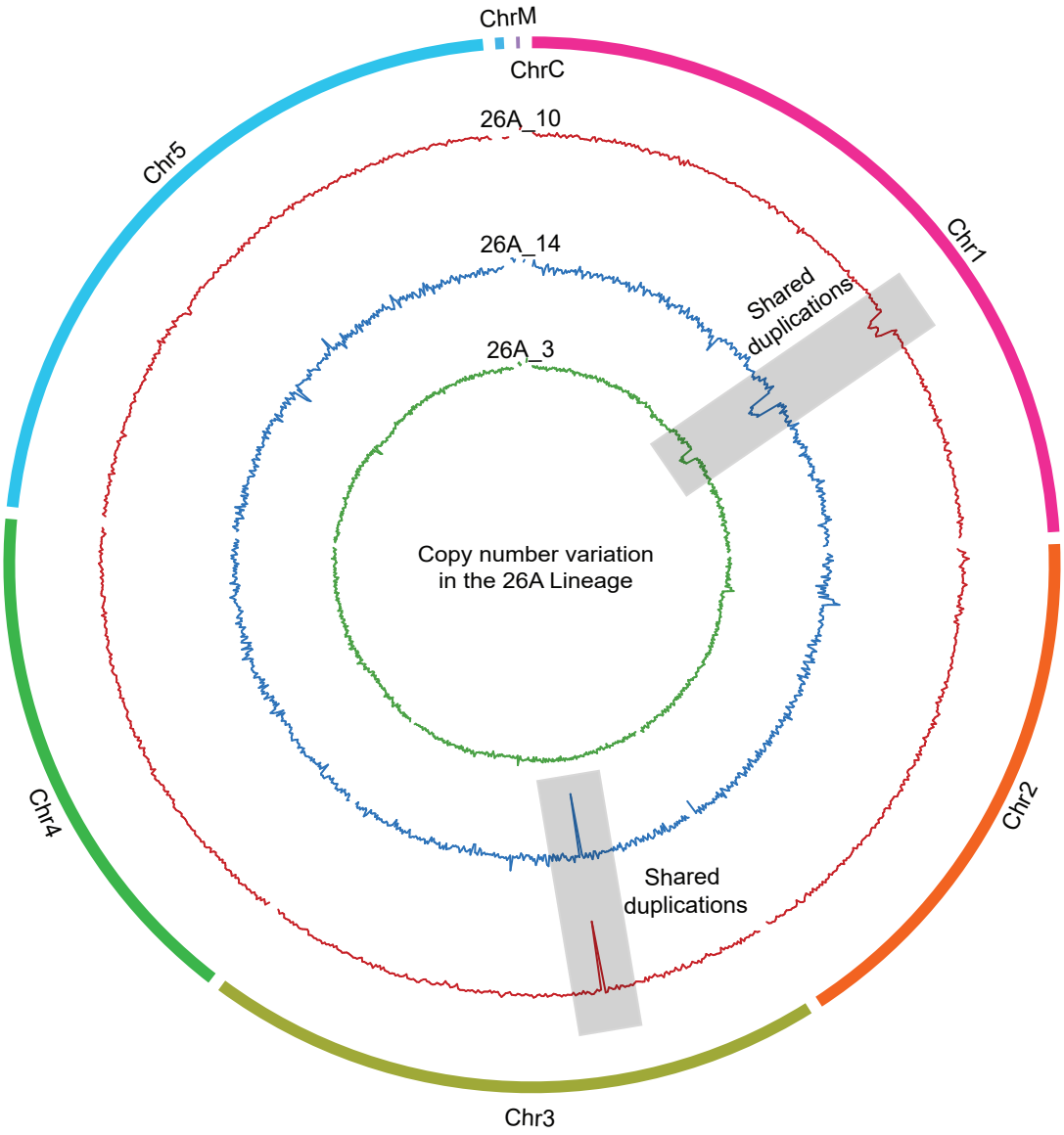

Figure S3: Duplications detected in the 26A lineage. Each track of the circos plot represents genome-wide read coverage analysis for an M2 sibling generated by selfing the 26A M1 mutant. Chromosome 1 duplications are shared by three M2 siblings while a Chromosome 3 duplication is shared by two of three siblings.

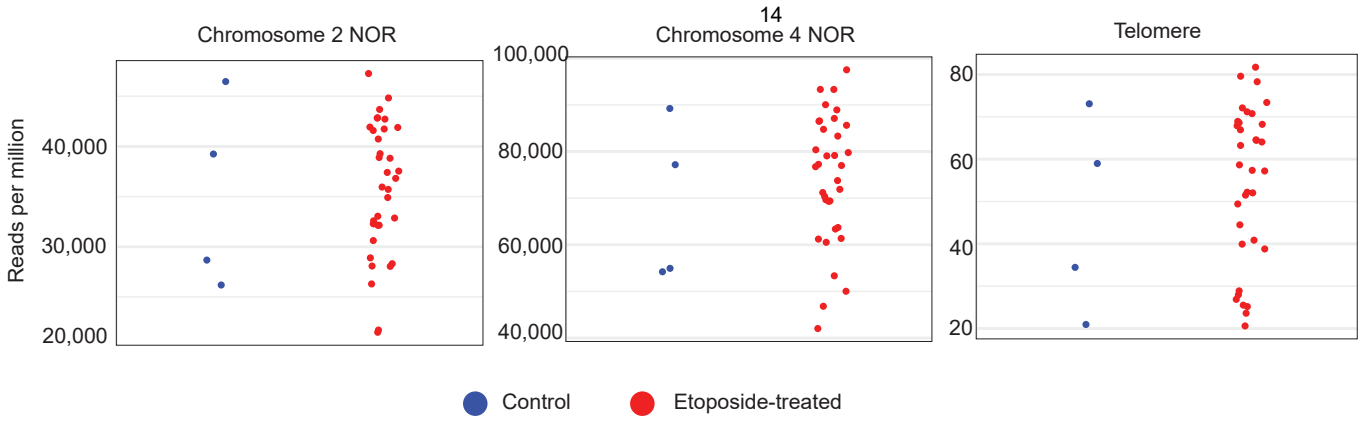

Figure S4: Variation in the copy number of rDNA and telomere repeats is similar in both etoposide-treated lines and wild-type lines. Reads aligning to NOR2, NOR4 or the telomere (CCCTAAA) were normalized to total aligned reads for each library. Each data point represents a value for an Illumina library derived from one individual line.

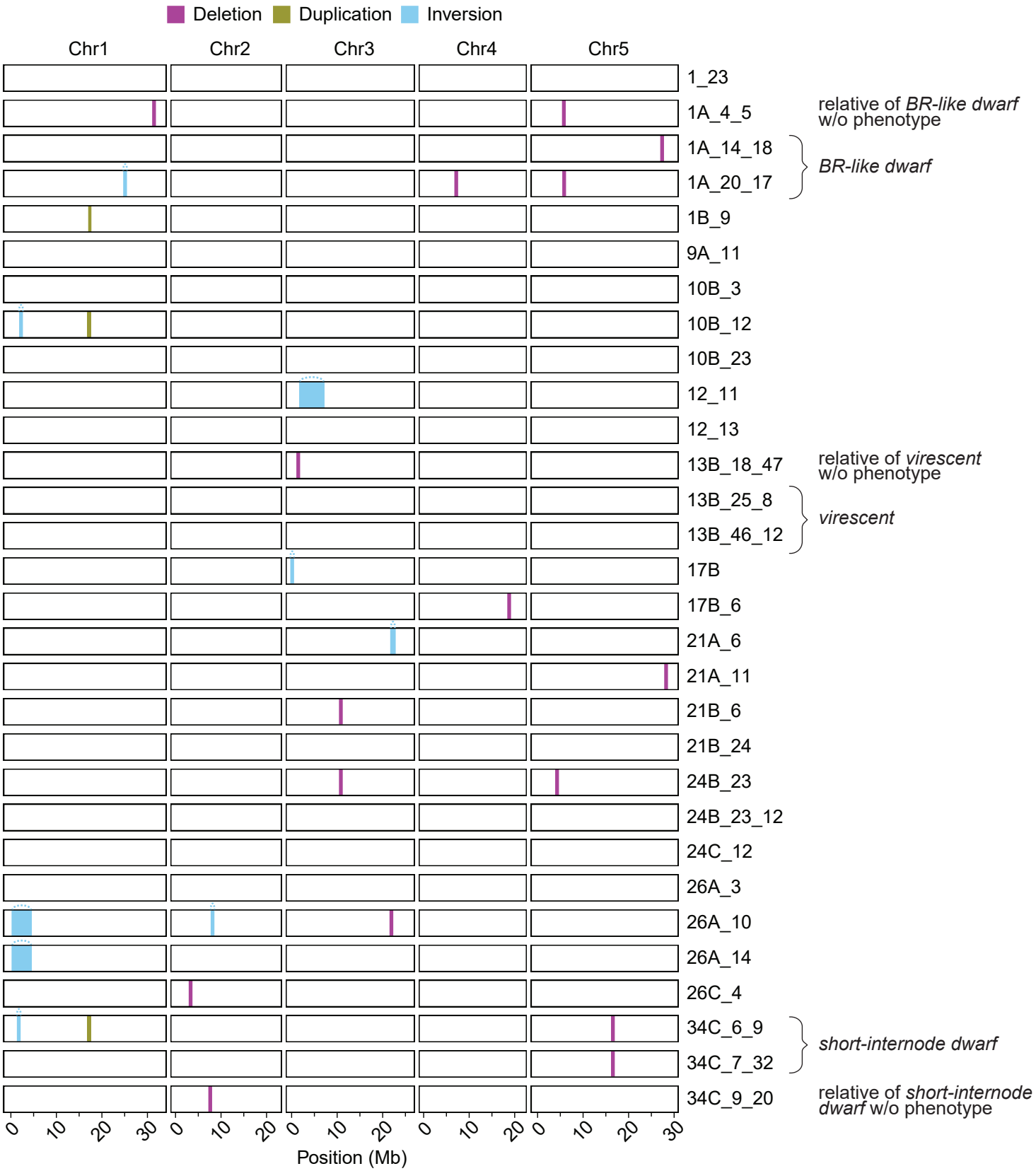

Figure S5: Structural variants identified by Lumpy Express in short-read data. A total of 27 SVs (16 deletions, 3 duplications, and 8 inversions) were identified. Eleven sequenced samples did not have any detected SVs.

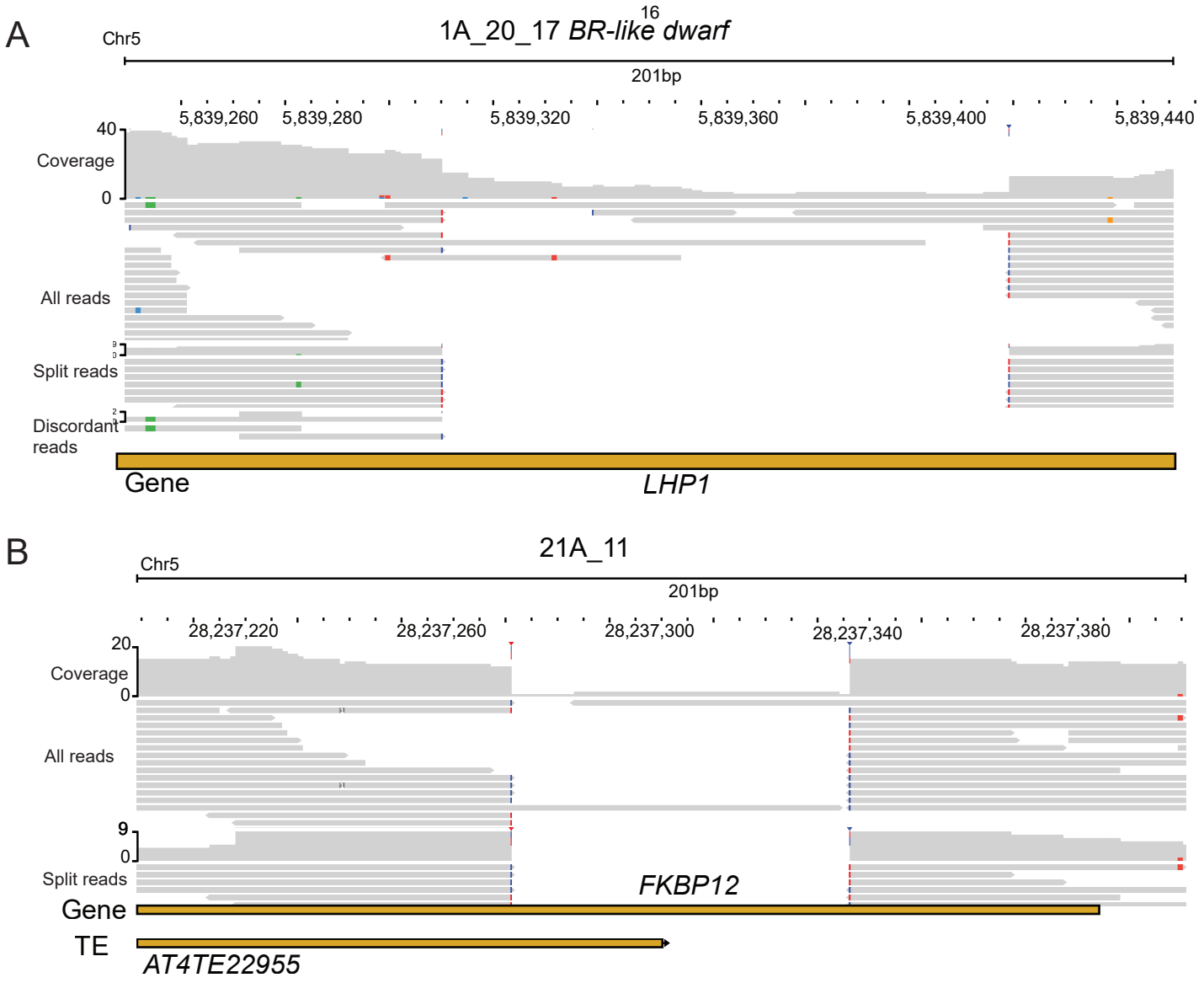

Figure S6: Examples of deletions in an etoposide-treated line identified by Lumpy Express using short reads. JBrowse snapshots of split reads and all reads overlapping *LHP1* **(A)** and *FKBP12* **(B)** identifies short deletions in two independent lines.

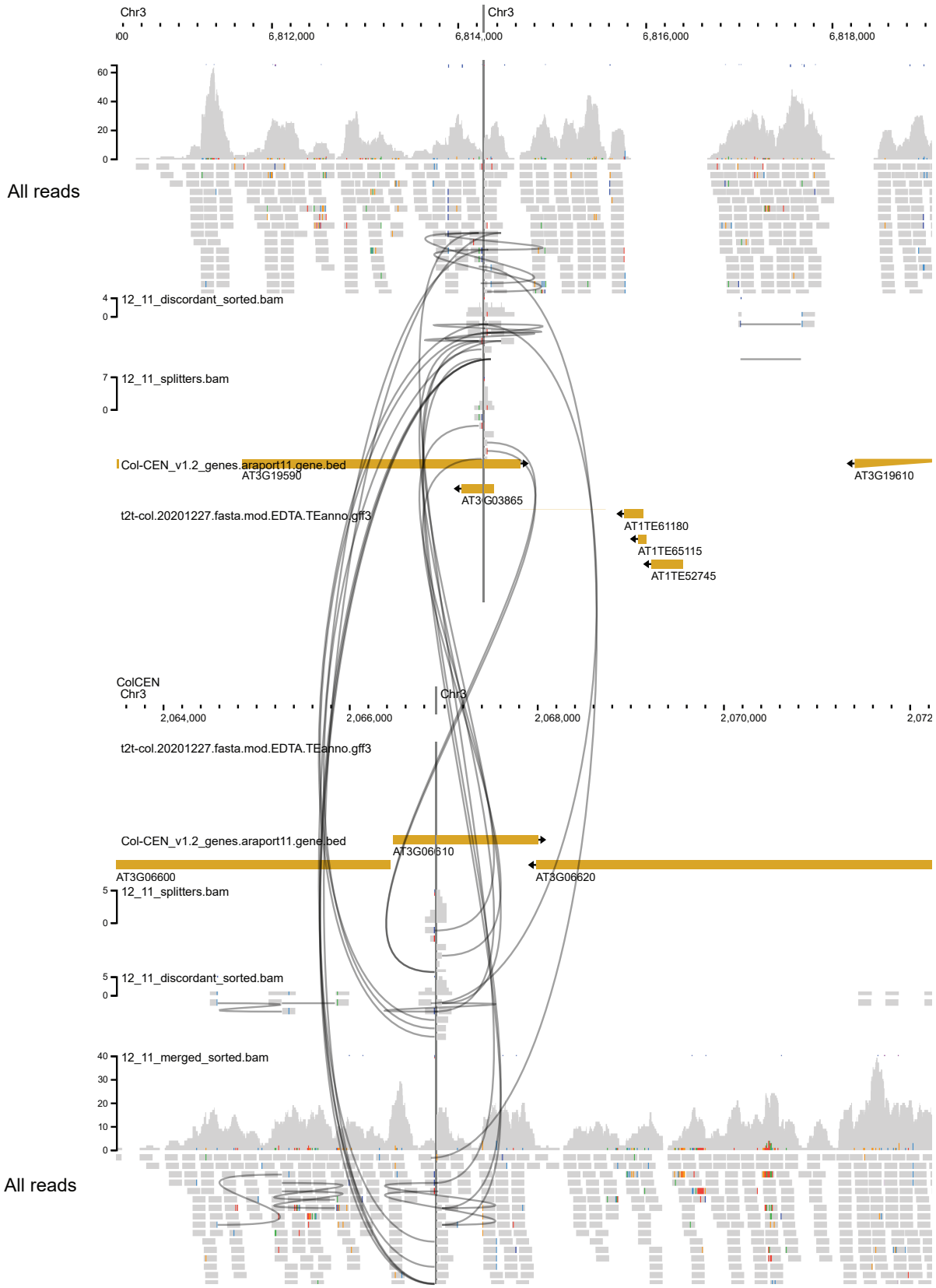

Figure S7: An example of an intrachromosomal inversion in an etoposide-treated line detected by Lumpy Express using short reads. Split reads and discordant reads indicate an intra-chromosomal inversion on Chr3. Black lines link split and discordant reads mapping to to two different parts of the chromosome.

length (Figure 2A, Table S4). The robustness of this analytical approach is underscored by detecting the same duplications in M2 siblings (Figure S3, Table S4). Next, we employed Lumpy Express [50] to identify structural variants (SVs) in mutant and control plants. This pipeline identified 2,224 variants that included inversions, deletions, duplications, and break ends, which were then filtered for high-confidence events. Break ends were excluded from further analyses because of the challenges of identifying their molecular nature. A high proportion of predicted SVs mapped to Nucleolar Organizer Regions, plastid genomes, and a small subset of other genomic loci. These SVs were also found in control plants, suggesting that these SVs represent mapping artifacts or naturally occurring SVs extant in wild-type plants. To further test the potential for the induction of SVs in difficult-to-map repetitive regions, we measured read depth at rDNA and telomeres and assessed if etoposide treatment was linked to a higher degree of repeat copy number variation. We found that the variation in read depth among etoposide-treated lines was comparable to that in control lines, suggesting that etoposide treatment likely did not trigger additional genomic instability in repetitive DNA (Figure S4). Based on these observations, we excluded SVs mapping to these regions from further analysis. To further filter for etoposide-independent SVs, those that were found in four or more etoposide-derived lineages were excluded, although we cannot exclude the possibility that these represent common fragile sites. After applying these stringent filters, we identified 27 unique SVs including 16 deletions and three duplications that ranged between 35 bp and 950 bp in size (Figure S5, Figure S6, Table S5). We also identified eight inversions that span between 124 bp and 4 Mb in length (Figure S5, Figure S7, Table S5). In sum, the coverage-based approach and Lumpy express analysis suggest that etoposide treatment induces structural variation in plants.

As the inhibition of topoisomerase II activity increases double-stranded DNA breaks in the genome [51], the short-read data was used to test if the repair of these breaks results in increased single nucleotide variation (SNV)). However, SNV analysis identified a similar number

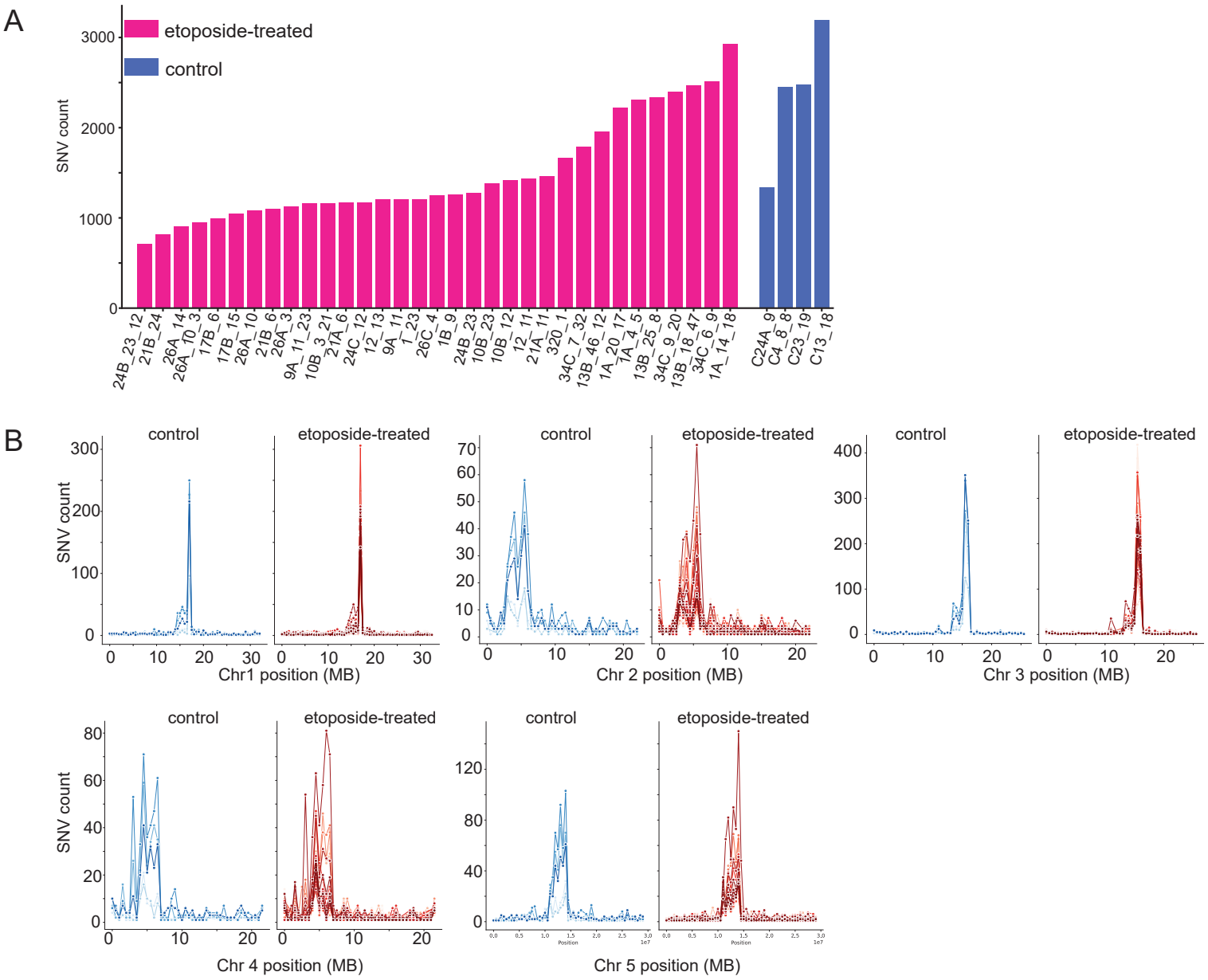

Figure S8: Etoposide treatment does not increase single-base substitutions. Single nucleotide variants per genome were identified using BCFtools. **(A)** Total number of single nucleotide variants, genome-wide, identified in etoposide-treated and control lines. **(B)** Frequency of single nucleotide variants in 100 Kb bins across the genome.

of SNVs in etoposide-treated and control lines (Figure S8), suggesting that etoposide did not induce excess SNVs.

Long-read sequencing can identify structural variants that are not easily resolved or cannot be identified by short-read sequencing[52]. Using Nanopore sequencing technology, we sequenced the genomes of: two M3 plants with the *BR-like dwarf*, *short-internode dwarf,* or *virescent* phenotypes, an M3 relative of each corresponding line that lacked the phenotype, and a single M3 plant from the *variegated* line that was green but produced a proportion of variegated progeny (Table S6). To control for structural variation already present in our wild-type lab Col-0 stock compared to the reference Col-0 genome, we also sequenced two M2 plants that were the progeny of M1 plants grown on DMSO only, and an untreated Col-0 plant. Reads were aligned to the Col-CEN v1.2 reference genome [53] and SVs were called using cuteSV [54,55]. Structural variants present in the DMSO and/or Col-0 controls, or present in more than one M1 line, or with a quality less than 20 were discarded. In total we identified 12 SVs: seven deletions (ranging from 37 bp to 176 bp), two insertions (71 bp and 184 bp), two inversions (1.52 Mb and 3.54 Mb), and one translocation (Figure 2B; Table S7). We also chose to confirm several of the SVs via PCR (Figure S9): a 184 bp insertion with homology to an intergenic region of chromosome 2 present in the *short-internode dwarf* line (Figure 2C); a 37 bp and an 81 bp deletion present in the *variegated* line (Figure 2D); and a 1.52 Mb inversion present on chromosome 2 in the *BR-like dwarf* line (Figure 2E). Together, these data indicate that etoposide treatment generates a range of SV types and sizes.

RNA-Seq identifies genes that are associated with structural variation and mutant phenotypes

To quantify the changes in gene expression for each of the four phenotypes studied in depth, we performed RNA sequencing for M3 plants of *BR-like dwarf, short-internode dwarf, virescent,*

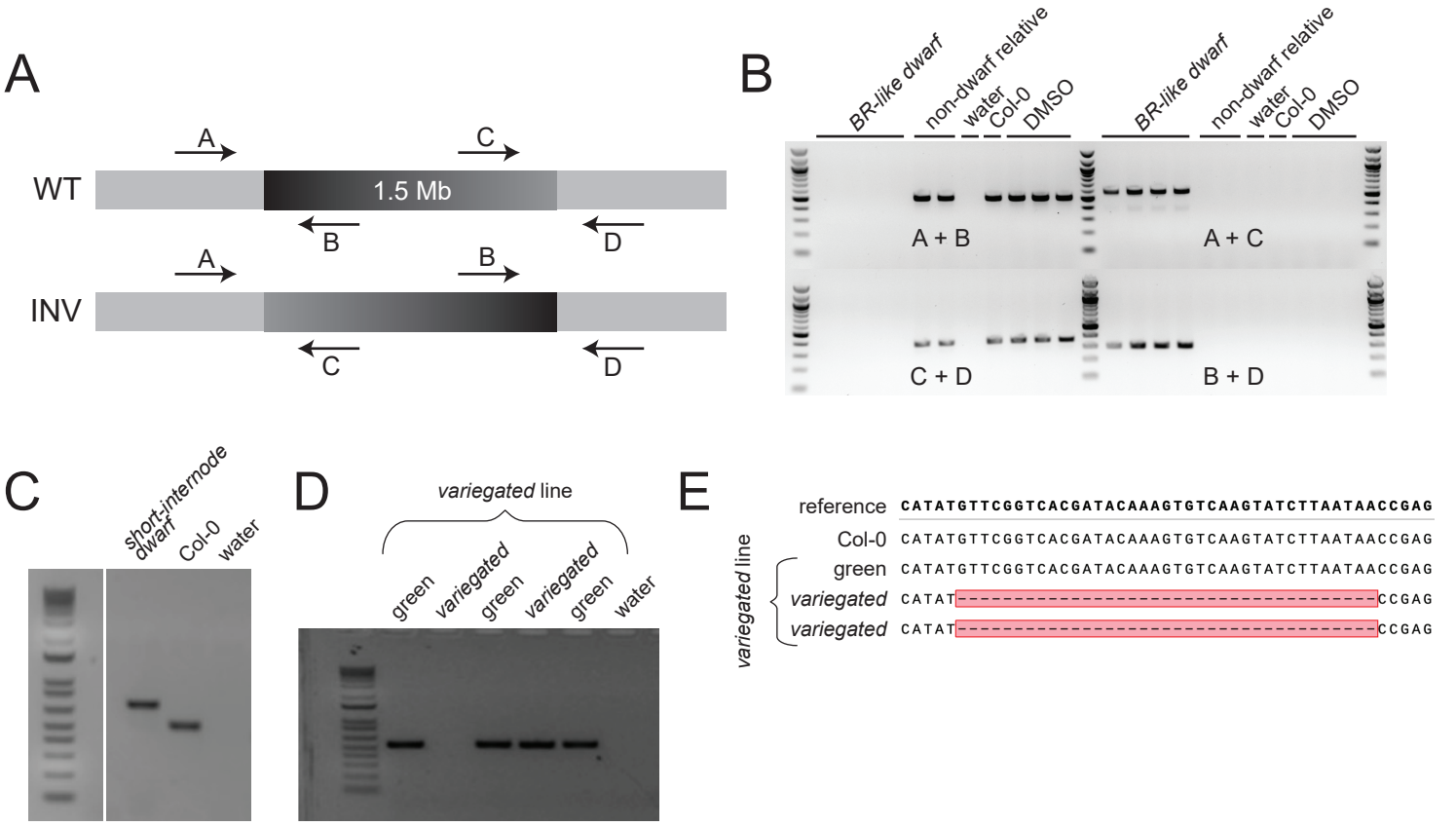

Figure S9: PCR verification of four strucutral variants identified via long-read sequencing. All plants sampled for PCR verification were relatives of the plants sequenced. **(A)** Diagram depicting primer design for verifcation of 1.52 Mb inversion identified in *BR-like dwarf* plants. **(B)** Image of gel electrophoresis of PCR products generated with different combinations of primers in (A); the inversion is present only in plants exhibiting the *BR-like dwarf* phenotype. **(C)** Gel image for verification of 184 bp insertion identified in *short-internode dwarf* plants. Primers were designed to flank the insertion. **(D)** Gel image for verification of 81 bp deletion identified in the plant sequenced from the *variegated* line. Primers were designed to flank the deletion; all plants tested are homozygous for the deletion, except for one plant (lane 3) for which the PCR failed. **(E)** Alignment of sequenced PCR products to *IMMUTANS* reference sequence. Primers were designed to flank the deletion. Individuals in the line exhibiting the *variegated* phenotype are homozygous for the deletion while green plants are either heterozygous or homozygous for the WT allele.

and *variegated* lines (Table S8). As controls, we included genetic relatives that lacked the phenotype and M3 progeny of DMSO-treated plants. To obtain a broad overview of changes in gene expression, we performed gene set enrichment analysis (GSEA) for biological process gene ontology (GO) on the full ranked list of genes (ranked by $\log_2$ fold change) for each phenotypic line. We also identified differentially expressed genes for each phenotype, defined as genes with an adjusted *p*-value < 0.01 and a $|\log_2 \text{fold change}| > 1$.

For the *BR-like dwarf* phenotype, the five most significant GO terms that were overrepresented among upregulated genes were: meristem development, anatomical structure formation involved in morphogenesis, response to auxin, plant organ formation, and post-embryonic plant morphogenesis (Figure S10). A total of 309 genes were differentially expressed in *BR-like dwarf* plants compared to non-phenotypic plants (114 downregulated and 195 upregulated; Figure 3A, Table S9). Both sequenced *BR-like dwarf* plants contain a 1.52 Mb inversion on chromosome 2 that is not present in non-phenotypic control plants of the same line (Figure 2B) and we investigated gene expression surrounding this locus further. We binned genes into 30 kb sliding windows (10 kb overlap between windows), regardless of whether genes had been identified as differentially expressed, and plotted average $\log_2$ fold change for each window within 1.5 Mb of the inversion breakpoints. Genes in windows ~500 kb prior to the left break end of the inversion displayed lower transcript abundance, whereas windows that were closer to or intersecting the left break end showed increased transcript abundance (Figure 3B). Genes in windows surrounding the right break end also displayed lower transcript abundance compared to wild-type (Figure 3B). Genes in windows greater than 1 Mb distal to either break end showed no change in average expression. The 1.52 Mb inversion itself was significantly enriched for differentially expressed genes, as determined by a permutation test (Figure S11). Examining the inversion breakpoints more closely, the right break end separated the first 42 bp of the 5'UTR of *ASYMMETRIC LEAVES 1 (AS1; AT2G37630*) from the rest of the gene (Figure 3C). *AS1* is a

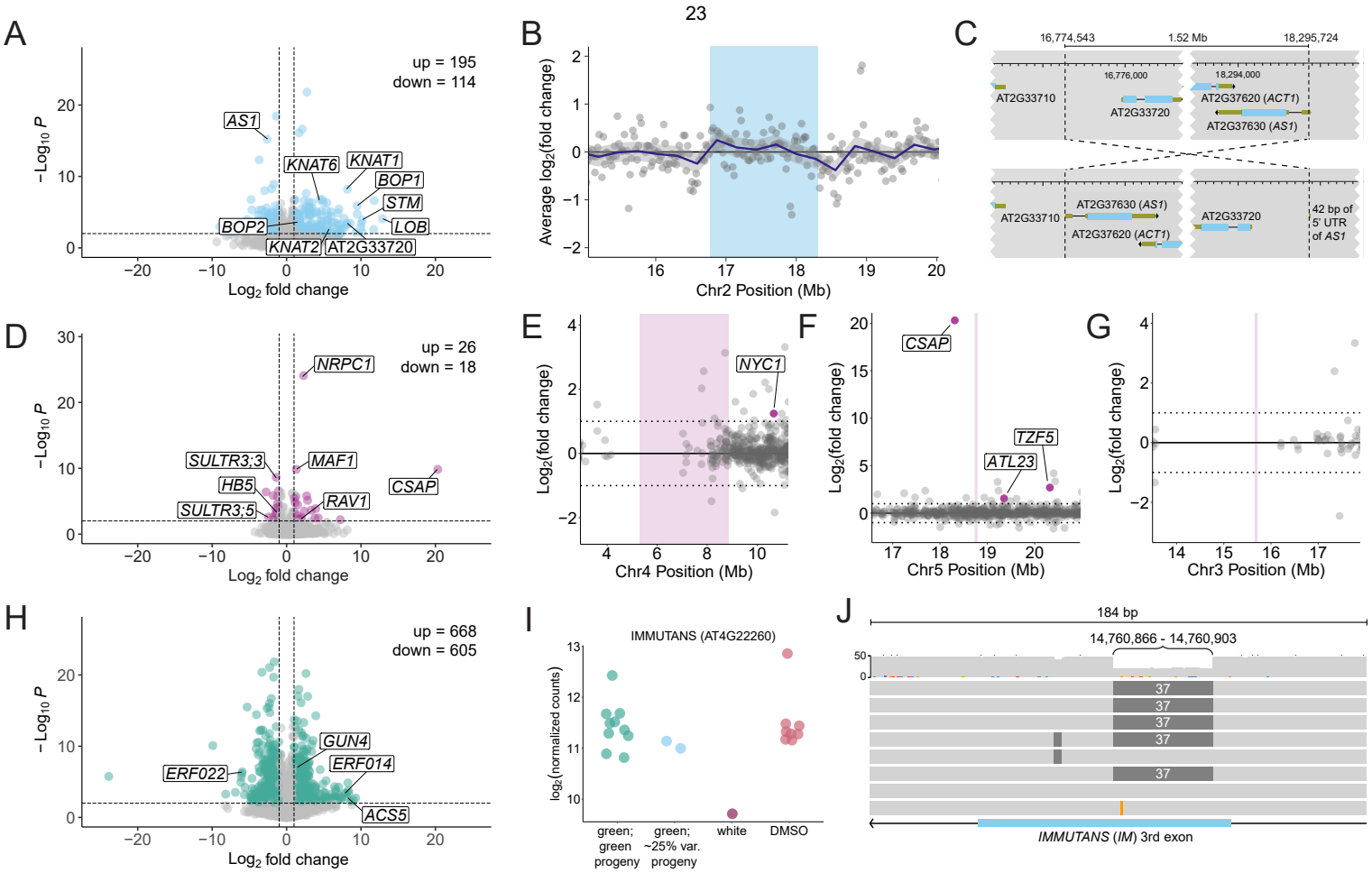

Figure 3: RNA-Seq identifies changes in gene expression that are associated with structural variation and mutant phenotypes. (A) Volcano plot summarizing differential expression analysis of *BR-like dwarf* vs. non-dwarf plants. Significantly differentially expressed genes ($|\log_2 FC| > 1$ and adjusted *p*-value < .01) are highlighted in blue. Key genes are labeled. **(B)** Average $\log_2$ fold change of genes within 30 kb sliding windows across chromosome 2 for *BR-like dwarf* vs. non-dwarf plants. Line represents locally estimated scatterplot smoothing to aid pattern visualization. The 1.52 Mb inversion on chromosome 2 in *BR-like dwarf* plants is shaded in blue. **(C)** Diagram of 1.52 Mb inversion in *BR-like dwarf* plants compared to reference depicting the genes that are disrupted. **(D)** Volcano plot summarizing differential expression analysis of *short-internode dwarf* vs. non-dwarf plants. Significantly differentially expressed genes ($|\log_2 FC| > 1$ and adjusted *p*-value < .01) are highlighted in pink. Select genes are labeled. **(E-G)** Log_2 fold change of genes 2 Mb outside candidate structural variants in *short-internode dwarf.* DEGs are in pink and SVs are shaded in light pink. **(H)** Volcano plot summarizing differential expression analysis of *virescent* vs. non-virescent plants. Significantly differentially expressed genes ($|\log_2 FC| > 1$ and adjusted *p* value < .01) are highlighted in yellow. Key genes are labeled. **(I)** Normalized counts of *IMMUTANS* transcripts in variegated line. **(J)** Snapshot of long-read alignments to reference *IMMUTANS*, showing a 37 bp deletion in the third exon. Deleted sequence shown in grey and coverage is shown above read alignments. The deletion is heterozygous in this particular plant, which was green and produced ~25% *variegated* progeny.

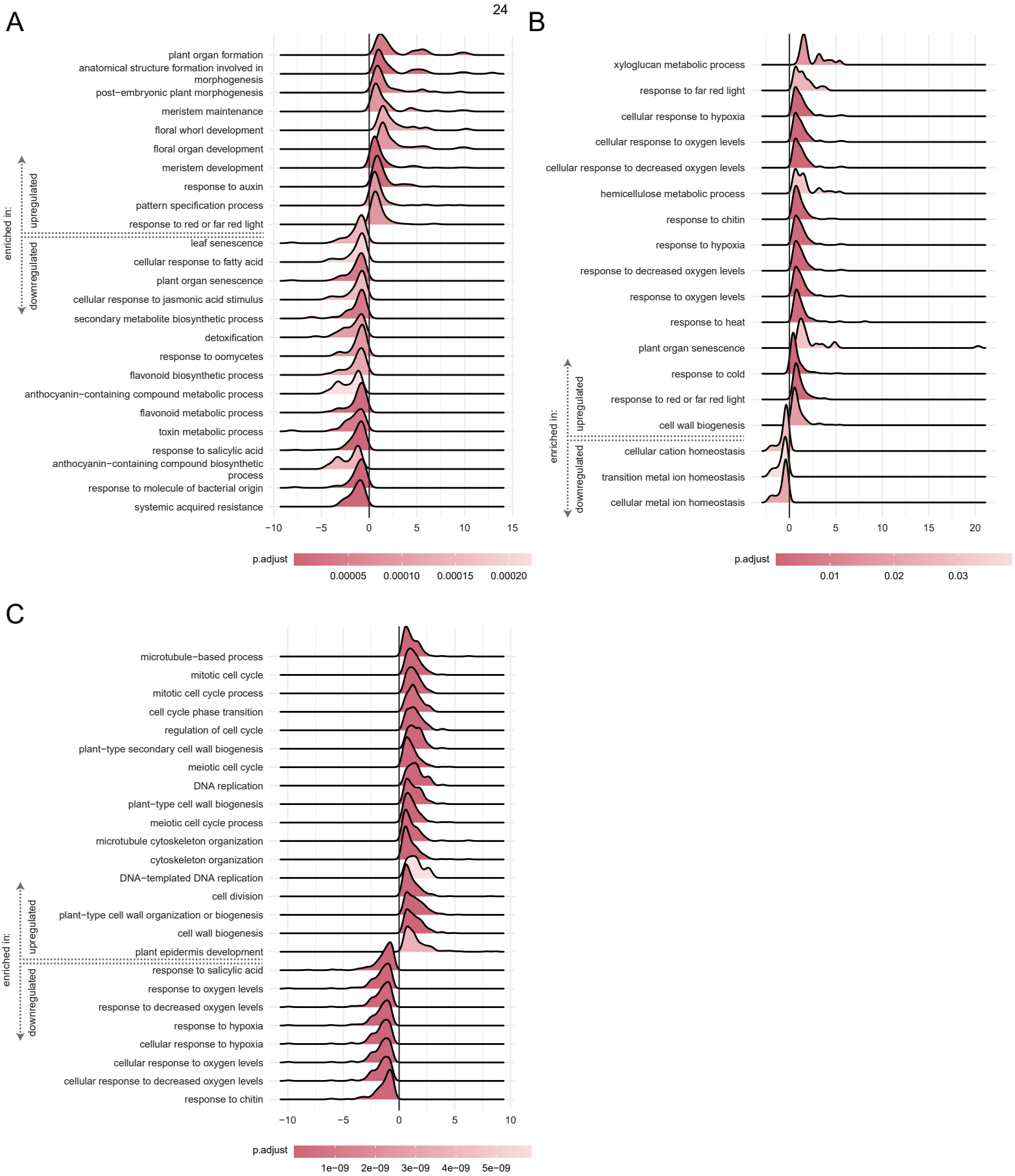

Figure S10: GO term enrichment for genes differentially expressed between plants exhibiting a phenotype and a genetic relative without a phenotype. GSEA was used to identify enriched GO terms. Ridge plots depict significant terms for **(A)** *BR-like dwarf*, **(B)** *short-internode dwarf*, and **(C)** *virescent* phenotypes.

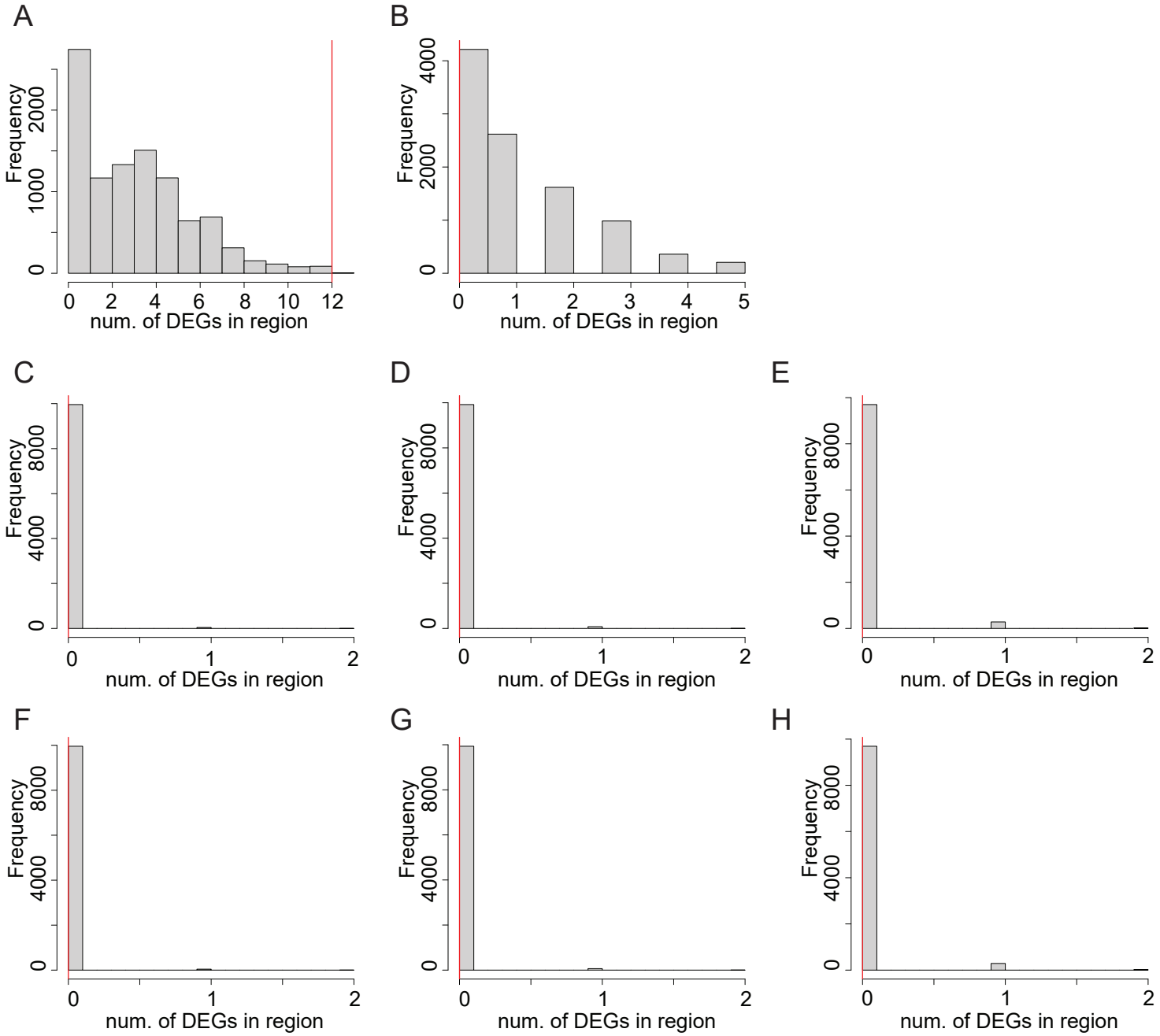

Figure S11: Permutation tests to determine if structural variants are associated with location of differentially expressed genes. For each test, 10,000 genomic regions of an equivalent size to the region of interest were chosen at random. Histograms plots the distribution of number of DEGs contained within the randomly selected regions. Red line marks the number of DEGs in the region of interest. P-values calculated by the probability that permuted values exceed the observed value. Regions tested for association with corresponding phenotype's DEGs were: **(A)** 1.52 Mb inversion in *BR-like dwarf* plants (p−value = 0.0091**)**; **(B)** 3.54 Mb inversion in *short-internode dwarf* plants (p-value = 1); **(C)** 5 kb, **(D)** 10 kb, and **(E)** 50 kb flanking the insertion on Chr3 in *short-internode dwarf* plants (*p*-value = 1 for all); and **(F)** 5 kb, **(G)** 10 kb, and **(H)** 50 kb flanking the deletion on Chr5 in *short-internode dwarf* plants (*p*-value = 1 for all).

differentially expressed gene ($\log_2$FC = -2.62, adjusted *p*-value = 6.37 x 10^{-16}), along with several genes it is known to interact with either directly or indirectly such as *KNAT1*, *KNAT2*, *KNAT6*, *STM*, *BOP1*, *BOP2*, and *LOB* (Figure 3A, Table S9)[56–59]. The phenotypes we observed in the *BR-like dwarf* plants are strikingly similar to those described for *as1* mutants, including a compact rosette with lobed and curled leaves [56,60]. Near the left break end of the inversion, the expression of an AP2/B3-like transcriptional factor family protein (*AT2G33720*) was increased in leaves (Figure 3A,C). Taken together, these data suggest that the 1.52 Mb inversion resulted in an altered *AS1*-mediated transcript abundance, influencing leaf patterning in *BR-like dwarf* plants.

In contrast to *BR-like dwarf* plants, *short-internode dwarf* plants showed considerably fewer gene expression changes in leaves. GSEA yielded primarily stress-associated GO terms (Figure S11). A total of 44 genes (18 down and 26 up) were differentially expressed in *short-internode dwarf* leaves compared to control leaves (Figure 3D, Table S10). Interestingly, both the largest subunit of RNA polymerase III, *NRPC1,* and the repressor of RNA polymerase III, *MAF1*, were upregulated. Three SVs were common to the two *short-internode dwarf* plants that were sequenced via Nanopore: a 176 bp deletion on chromosome 3, a 3.54 Mb inversion on chromosome 4, and a 184 bp insertion on chromosome 5 (Figure 2B, Table S7). The large inversion on chromosome 4 was not enriched for DEGs, and neither were regions surrounding the deletion on chromosome 3 or the insertion on chromosome 5 (Figure 3E-G, S10). It is possible that *short-internode dwarf* plants have changes in gene expression in other plant organs besides the leaf or perhaps exhibit a global change in gene expression that cannot be detected with standard RNA-seq approaches.

Virescent plants had more dysregulated genes than either dwarf line, with a total of 1273 differentially expressed genes (668 up and 605 down; Figure 3H, Table S11). The most significant GO terms overrepresented amongst genes with increased expression include cell

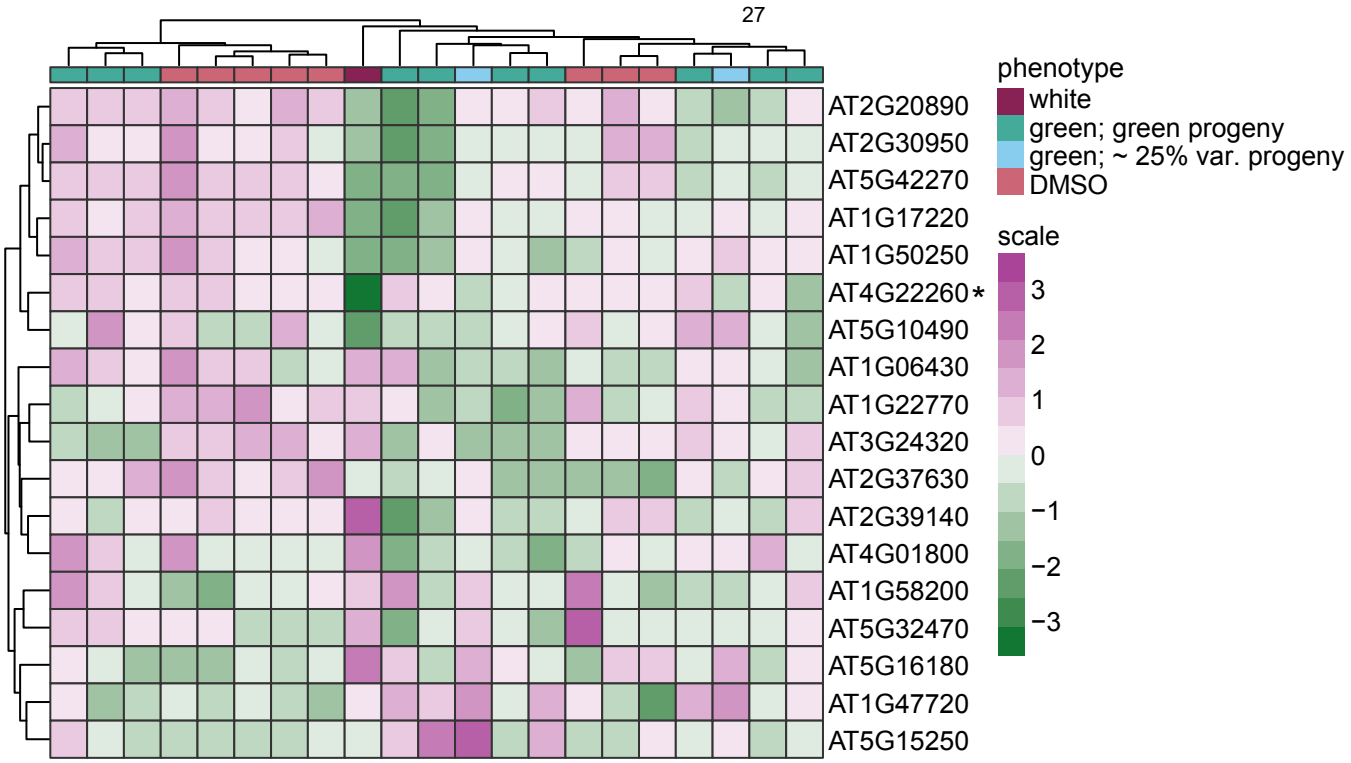

Figure S12: *IMMUTANS* is downregulated in *variegated* plant. Heatmap of expression of genes known to be associated with variegation. *IMMUTANS* is annotated with a *.

division, plant-type cell wall organization/biogenesis, mitotic cell cycle, microtubule-based process, and meiotic cell cycle (Figure S11). The most significantly overrepresented GO terms among genes with decreased expression include response to chitin, response to hypoxia, and response to oxygen levels (Figure S11). Interestingly, genes related to ethylene synthesis and signaling including *ERF022*, *ERF014*, and *ACS5* were among the most differentially expressed genes (Table S11).The mis-regulation of these genes might be associated with the *virescent* phenotype, as ethylene is involved in greening of etiolated seedlings after light exposure, and conversely degradation of chlorophyll during leaf senescence [61,62]. Other differentially expressed genes of note include several subunits of photosystem I and II and GUN4 (Table S11).

Differential expression analysis could not reliably be conducted for the *variegated* line, due to the few number of variegated plants recovered. As an alternative strategy, we examined the expression of 18 genes known to be associated with variegation (Figure S12). In the one variegated plant we recovered and performed RNA-seq on, the *IMMUTANS* (*IM*; *AT4G22260*) gene stood out as highly downregulated (Figure 3I, Figure S12). Long-read sequencing identified a 37 bp deletion in the third exon of *IM*; this caused a frame-shift resulting in a premature stop codon (Figure 3J). Crossing plants of the *variegated* phenotype to known *immutans* mutants failed to complement the variegated phenotype (32/32 variegated when crossed to *im-spotty* CS3639 [63], 34/34 variegated when crossed to *im-spotty* CS73029 [64] and 42/42 variegated when crossed to *im* (CS3157), indicating that the 37 bp deletion in *IM* is the causal locus for variegation.

Discussion

We describe a simple and convenient technique that enables researchers or breeders with limited resources to induce structural variation. We found that early exposure of *A. thaliana* to the topo II inhibitor etoposide induces heritable, large-effect mutations at high frequency. Examination of M2 plants found that 24/37 mutant lineages had at least one out of six assessed phenotypes. By short- and long-read genomic sequencing of mutagenized lines – including those that displayed novel phenotypes – we identified the full spectrum of structural variations. Analysis of short-read data identified 2.53 SVs per mutagenized M1 line while long-read data identified 3.25 SVs per mutagenized M1 line, with one event being detected in both short and long-read data. The lower number of SVs identified by short-read data is consistent with the poorer performance of short-read SV callers relative to callers using long-reads [52].

A further increase in the number of mutations per M1 parent could perhaps be attained either by prolonging exposure to etoposide beyond two weeks or by germinating seeds on media containing a cocktail of etoposide and other drugs that induce genomic instability such as topoisomerase I inhibitors [65] or the DNA damage-inducing agent cisplatin [66].

Analysis of all detected SVs suggests that etoposide disproportionately impacts genes, although our analytical approaches cannot conclusively identify all SVs in heterochromatin. The SVs that we identified disrupt one or more genes simultaneously or create novel cis-regulatory regions. The effect of this genic bias is striking. Even in our relatively small mutant population, multiple loss-of-function mutations were identified in a single line. This suggests that etoposide treatment can be used to create novel polygenic traits.

What explains this bias for mutations in genes? Topoisomerase II has been shown to associate with actively transcribed regions and open chromatin regions [41]. This association with actively transcribed regions suggests a future method to enrich for mutations in condition-specific genetic pathways by simultaneously exposing seeds to etoposide and another condition. For

example, etoposide might promote mutagenesis of immunity genes whose transcription is triggered when seeds are germinated in the presence of microbial pathogens. In another scenario, seeds germinated in the presence of a hormone and etoposide might accumulate mutations in genes that respond to hormones.

The inversions and translocations generated by this technique might also disrupt meiosis and distort segregation, but this possibility has not been tested. Etoposide also created duplications and deletions that could potentially alter gene dosage over mega-base scale regions. In allopolyploids, such segmental deletions can provide a deficiency chromosome that can be used to assess the relative contribution of homeologs to different traits.

The analysis of mutants generated in our study suggests that mutation rate and type induced by etoposide exposure is comparable to mutant populations generated by irradiation. Analysis of Arabidopsis seeds irradiated with γ rays or carbon ions showed mutations occurring at a rate of 0.8-1.6 SVs per mutant lineage[67]. Thus, etoposide treatment out-performed irradiation by producing more than 2.5 events per mutagenized plant. Sequencing and close analysis of the genomes and transcriptomes of *BR-like dwarf, short-internode dwarf, virescent,* and *variegated* plants allowed us to map two of these phenotypes to strong candidate mutations. The *variegated* phenotype is caused by a 37 bp deletion in *IMMUTANS* and the *BR-like dwarf* phenotype is likely caused by an inversion that disrupts the expression of *ASYMMETRIC LEAVES 1.* Strikingly, nearly 60 years ago, the *immutans* and *asymmetric leaves* phenotypes were amongst the first X-ray induced mutations isolated in *A. thaliana* [68,69]. These results suggest that our easy-to-use chemical mutagenesis approach provides plant biologists and breeders with a new tool that can replace irradiation and democratize access to the creation of structural variation.

Materials and Methods

Drug treatment

Etoposide (Abcam AB120227) was dissolved in DMSO to obtain a 100 mM solution that was added to 0.5X MS media (supplemented with 2% sucrose and Phytoagar) to final concentrations of 0, 20, 40, 80,160, 320, and 640 µM. We noticed decreased solubility of etoposide in aqueous solutions at concentrations beyond 160 µM. Wild-type Col-0 seeds were sterilized by incubating in 2% (v/v) PPM (Plant Cell Technologies) for three days at 4°C and plated on MS media containing etoposide or DMSO. Seeds were germinated and grown on this media for two weeks, and seedlings were examined for signs of stunted root or shoot growth. Two- to three-week-old seedlings (M1 plants), including several showing stunted growth, were transferred to soil and allowed to set seed. M2 seeds were collected from mutagenized M1 plants.

Plant Growth Conditions

Arabidopsis thaliana was grown in a Conviron growth chamber at 22°C with 16 h of light per day (120 µmol).

Phenotyping

Between 25 and 75 seeds from each M1 line were planted over three phases and M2 plants with visible phenotypes were manually identified. To identify mutants with leaf-shape differences, we measured leaf shape and size in mutant lineages where M2/M3 populations included three or more individuals displaying similar mutant phenotypes. Leaves were excised at the base of the petiole, arranged in decreasing size order on white paper, and scanned at 600 dpi on an Epson V600 scanner. Six automated leaf shape measurements were obtained (blade area, circularity, length, perimeter, width, and petiole length) using the LeafJ plugin for

ImageJ software [70]. Leaf shape was manually traced in cases where leaf boundaries were not distinguished by ImageJ.

Illumina sequencing and identification of structural variants

DNA was extracted from the rosette leaves of M2 and M3 plants using a CTAB protocol. Up to 15 ng of genomic DNA was used by the BioMicroCenter at MIT to construct DNA sequencing libraries using the Nextera DNA Flex library preparation kit. Libraries were multiplexed and sequenced in 150 bp paired-end format using an Illumina NovaSeq 6000 S4. The first 10 bp of each read as well as any adaptor sequence was removed using Trimmomatic [71] and aligned to the Col-0 Centromere assembly [53] using BWA-MEM [72] (*bwa mem -R "@RG\tID:id\tSM:sample\tLB:lib" reference.fasta sample.R1.fastq sample.R2.fastq \ | samblaster --excludeDups --addMateTags --maxSplitCount 2 --minNonOverlap 20*). These alignments were used to identify structural variants by two separate methods. Larger deletions and duplications were identified using a method proposed by the Comai Lab [48,49]. Coverage was calculated for each 100 kb window for properly paired and aligned reads using Bedtools [73] (*bedtools coverage -a 100kb_windows.bed -b sample_paired_sorted.bam -counts*) and read-counts/window for each line was normalized to library size as reads per million. The median normalized read-counts for all lines for each window was calculated. As most lines are not expected to share a structural variant, this median read depth is considered to be the wild-type genome dosage and is set to a ploidy or chromosomal dosage of two. The chromosomal dosage of any window for a given line is then calculated as (*normalized read-counts / median read-counts) x2.* Windows with a coverage score ≥3 were considered heterozygous duplications, ≥4 to be homozygous duplications, ≤1 to be heterozygous deletion and ≤0 to be a homozygous deletion.

In addition, we also identified structural variation using a benchmarked Lumpy Express pipeline [12,50]. Briefly, Samtools [74,75] was used to identify discordant and split reads from the SAM

file that was output by BWA [72]. Files containing discordant reads, split reads, and all reads were used by Lumpy Express to identify putative structural variants. Due to mapping artifacts, Lumpy Express and other variant callers often identify false-positive structural variant candidates at repetitive regions [76]. To address this problem, we filtered out variants from ChrC, ChrM, and the NOR on chromosome 2. To remove SVs present in lab wild-type Col-0 stocks, we removed SVs within 2 kb of any SV found in lines only treated with DMSO, or SVs shared by four or more unrelated etoposide-treated mutant lineages. Finally, only those deletions and duplications that were supported by split-reads were retained in the final list of SVs.

To assess if etoposide causes single nucleotide variants, we used BWA-MEM to align 32 reads from M2/M3 progeny of etoposide-treated plants, and four M2 progeny of control plants treated with DMSO to the Col-0 centromere assembly [53]. Aligned reads were sorted and converted to variant calling format (*bcftools mpileup -O b -o file.bcf -f ref.fasta file.sorted.bam*) followed by SNP calling (*bcftools call --ploidy 2 -m -v -o file.vcf file.bcf*) [74]. Variants that were called by bcftools were filtered to retain only calls with a quality score higher than or equal to 40 (*bcftools view -O z -o filtered.vcf.gz -e 'QUAL<=40' in.vcf.gz*). Variants shared by all untreated lines were extracted (*bcftools isec -n=5 -p common_positions.vcf input_files_filtered.vcf.gz*), and then excluded from all other treated lines (*bcftools view -T common_positions.vcf filtered_files.vcf -O z -o untreated_filtered_files.vcf*).

Nanopore sequencing and identification of structural variants

Two mature leaves from 32-day-old M3 plants were flash frozen in liquid nitrogen. Frozen tissue was ground into a powder with a Tissuelyser II (QIAGEN; 30 beats/sec for 1 min). High molecular weight DNA was extracted using a Wizard HMW DNA Extraction Kit (Promega). DNA was sheared by passing through a 23G needle twice. Contaminating RNA and proteins were removed using RNase A (New England Biolabs) and Proteinase K (New England Biolabs). DNA

was further cleaned and small fragments removed using a modified SPRI bead mixture in which the buffer of Agencourt AMPure XP beads was exchanged with a custom buffer of 10 mM Tris-HCl, 1mM EDTA pH8, 1.6M NaCl, 11% w/v PEG 8000 [77]. To perform the size selection, in brief, 0.7x volume of modified bead solution was added to each DNA sample, then incubated on a nutator for 10 min. Samples were placed on a magnetic stand to bind the beads to the side of the tube, then washed twice with 70% ethanol. Beads were not allowed to dry before adding nuclease-free water to elute the DNA. To increase yield and remove residual ethanol, samples were incubated at 50°C for 2 min and then at room temperature for two hours with the lids open.

All library preparation and sequencing was performed at the MIT BioMicroCenter. For two samples showing the *BR-like dwarf* phenotype, high molecular weight DNA extractions had low yield. For these samples and corresponding controls (individual from sibling line without dwarf phenotype and individual from a line treated with DMSO), library preparation was performed utilizing a Rapid Barcoding Kit V10 (Oxford Nanopore Technologies; SQK-RBK110-96). These four samples were multiplexed and sequenced on a single PromethION R9.4.1 flow cell (Oxford Nanopore Technologies; FLO-PRO002). For all other samples, library preparation was performed with a Native Barcoding Kit V14 (Oxford Nanopore Technologies; SQK-NBD114-24). These samples were multiplexed into groups of three and each group was sequenced on its own PromethION R10.4.1 flow cell (Oxford Nanopore Technologies; FLO-P1RO14M). Both runs utilized the following software versions for sequencing and base-calling: MinKNOW 23.04.6, Bream 7.5.10, Configuration 5.5.14, Guppy 6.5.7, and MinKNOW Core 5.5.5.
Nanopore reads were filtered utilizing Filtlong v0.2.1 [78] to select the best reads in terms of length and quality, up to a max coverage of 50x(*--min_length 1000 --min_window_q 40 --trim --split 1500 --target_bases 6605000000*). Filtered reads were mapped to the Col-CEN v1.2 genome [53] using Vulcan dual-mode alignment pipeline [79], which leverages both minimap2 [80] and NGMLR [81], using default settings for ONT reads. Structural variants in each sample

were detected with cuteSV *(--max_cluster_bias_INS 100 --diff_ratio_merging_INS 0.3 --max_cluster_bias_DEL 100 --diff_ratio_merging_DEL 0.3 --min_size 30 --max_size -1 –genotype;* [55,82]). SVs identified in all samples were merged with SURVIVOR (*merge 1000 1 1 1 0 30;* [83]), then samples were genotyped with cuteSV for all SVs identified (*--min_mapq 20 --max_cluster_bias_INS 100 --diff_ratio_merging_INS 0.3 --max_cluster_bias_DEL 100 --diff_ratio_merging_DEL 0.3 --min_size 30 --max_size -1*). Structural variants present in the DMSO and/or Col-0 controls, or present in more than one phenotypic line, or with a quality less than 20 were removed. Data was visualized in R using ggplot2.

RNA sequencing and analysis

RNA sequencing was conducted on M3 generation plants. Six plants of each *BR-like dwarf, short-internode dwarf*, and *virescent* phenotypes, as well as four non-phenotypic individuals corresponding to sibling lines of each phenotype were sequenced. For the fraction variegated line, thirteen green plants (three of which produced ~25% variegated progeny) and one white plant were sampled. Nine M3 progeny of DMSO-treated plants were included as controls. Mature leaf tissue from 32-day-old plants was flash frozen in liquid nitrogen. Frozen tissue was ground into a powder with a Tissuelyser II (QIAGEN; 30 beats/sec for 30 sec). RNA was extracted using an RNeasy Plant Mini Kit (QIAGEN). Library preparation was performed at the MIT BioMicro Center using an NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Samples were multiplexed on one lane of an Illumina NovaSeq 6000 S4 flow cell and sequenced at the MIT BioMicro Center with 150 bp paired end reads.

Trim Galore [84] was used to remove adapters and low-quality ends from reads (*--paired --retain_unpaired -q 25*). Salmon [85] was used to quantify the expression of transcripts, utilizing the Araport 11 transcriptome as the reference and the Col-CEN v1.2 genome [53] as decoy sequences. Five samples were identified as outliers and removed from the dataset, based upon the PCAgrid function in the R package rrcov [86,87] and preliminary analyses. DESeq2 [88] was used to quantify differential expression; for a gene to be considered it must have at least 10

normalized counts in four or more samples. Each phenotype of interest was analyzed independently; DESeq2 design (~ *treatment* + *phenotype*) incorporated both treatment (etoposide or DMSO) and phenotype (present or absent). The R package clusterProfiler [89] was used to conduct gene set enrichment analysis. Data was visualized utilizing R and packages ggplot2, EnhancedVolcano[90] , and pheatmap [91].

Data Availability

Short-read DNA sequencing, long-read DNA sequencing, and mRNA-seq data are available in GEO records GSE284455, GSE284456 and GSE284457.

Declaration of Interest

Whitehead Institute has filed a patent application related to this work.

Acknowledgements

We thank Aysha Hussain, Jaza Alam, and Lingfeng Shi for assistance with phenotyping, Dyna Louis for assistance with verifying structural variants, and Xinlei Gao for assistance with statistical testing. This research was funded by grants to MG from The Abdul Latif Jameel Water and Food Systems Lab at the Massachusetts Institute of Technology, from the Professor Amar G. Bose Research Grants, and from a gift to MG from the Dr. Vincent J. Ryan Orphan Plant Project. MG is an Investigator of the Howard Hughes Medical Institute.

Supplemental Information

Figure S1: Additional phenotyping of progeny of etoposide-treated plants.

Figure S2: Segregation of the *variegated* phenotype from green parents.

Figure S3: Duplications detected in the 26A lineage.

Figure S4: Variation in the copy number of rDNA and telomere repeats is similar in both etoposide-treated and wild-type lines.

Figure S5: Structural variants identified by Lumpy Express in short-read data.

Figure S6: Examples of deletions in an etoposide-treated line identified by Lumpy Express using short reads

Figure S7: An example of an intrachromosomal inversion in an etoposide-treated line identified by Lumpy Express using short reads.

Figure S8: Etoposide exposure does not increase single-base substitutions.

Figure S9: PCR verification of four structural variants identified via long-read sequencing.

Figure S10: GO term enrichment for genes differentially expressed between plants exhibiting a phenotype and a genetic relative without a phenotype.

Figure S11: Permutation tests to determine if structural variants are associated with location of differentially expressed genes.

Figure S12: *IMMUTANS* is downregulated in the *variegated* plant.

Table S1: List of phenotypes observed in M2 progeny of etoposide-treated plants

Table S2: List of etoposide-treated lines sequenced with Illumina.

Table S3: Relationship between the sequenced lines, treatments, and nature of sequencing.

Table S4: Deletions and duplications identified by read coverage analysis.

Table S5: List of structural variants detected by Lumpy Express using short-read sequencing

Table S6: List of etoposide-treated lines sequenced with Nanopore.

Table S7: Structural variants identified utilizing long-read data.

Table S8: List of RNA-Seq libraries.

Table S9: DESeq2 output identifying mis-regulated genes in *BR-like dwarf.*
Table S10: DESeq2 output identifying mis-regulated genes in *short-internode dwarf.*

Table S11: DESeq2 output identifying mis-regulated genes in *virescent.*

References

1. Brown EJ, Nguyen AH, Bachtrog D. The Drosophila Y Chromosome Affects Heterochromatin Integrity Genome-Wide. Molecular Biology and Evolution. 2020;37: 2808–2824. doi:10.1093/molbev/msaa082

2. Henikoff S. Dosage-dependent modification of position-effect variegation in Drosophla. BioEssays. 1996;18: 401–409. doi:10.1002/bies.950180510

3. Murphy PJ, Berger F. The chromatin source-sink hypothesis: a shared mode of chromatin-mediated regulations. Development. 2023;150: dev201989. doi:10.1242/dev.201989

4. Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, et al. The impact of structural variation on human gene expression. Nat Genet. 2017;49: 692–699. doi:10.1038/ng.3834

5. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, et al. Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell. 2020;182: 145-161.e23. doi:10.1016/j.cell.2020.05.021

6. Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T, et al. The population genetics of structural variants in grapevine domestication. Nat Plants. 2019;5: 965–979. doi:10.1038/s41477-019-0507-8

7. Tian S, Ge J, Ai G, Jiang J, Liu Q, Chen X, et al. A 2.09 Mb fragment translocation on chromosome 6 causes abnormalities during meiosis and leads to less seed watermelon. Horticulture Research. 2021;8: 256. doi:10.1038/s41438-021-00687-9

8. Wang X, Liu Z, Zhang F, Xiao H, Cao S, Xue H, et al. Integrative genomics reveals the polygenic basis of seedlessness in grapevine. Current Biology. 2024;34: 3763-3777.e5. doi:10.1016/j.cub.2024.07.022

9. Chen J, Liu Y, Liu M, Guo W, Wang Y, He Q, et al. Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. Nat Genet. 2023;55: 2243–2254. doi:10.1038/s41588-023-01571-z

10. Tao Y, Luo H, Xu J, Cruickshank A, Zhao X, Teng F, et al. Extensive variation within the pan-genome of cultivated and wild sorghum. Nat Plants. 2021;7: 766–773. doi:10.1038/s41477-021-00925-x

11. Yan H, Sun M, Zhang Z, Jin Y, Zhang A, Lin C, et al. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. Nat Genet. 2023;55: 507–518. doi:10.1038/s41588-023-01302-4

12. Göktay M, Fulgione A, Hancock AM. A New Catalog of Structural Variants in 1,301 A. thaliana Lines from Africa, Eurasia, and North America Reveals a Signature of Balancing Selection at Defense Response Genes. Molecular Biology and Evolution. 2021;38: 1498–1511. doi:10.1093/molbev/msaa309

13. Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science. 2021;373: 655–662. doi:10.1126/science.abg5289

14. Mizuno H, Katagiri S, Kanamori H, Mukai Y, Sasaki T, Matsumoto T, et al. Evolutionary dynamics and impacts of chromosome regions carrying R-gene clusters in rice. Sci Rep. 2020;10: 872. doi:10.1038/s41598-020-57729-w

15. Katz E, Li J-J, Jaegle B, Ashkenazy H, Abrahams SR, Bagaza C, et al. Genetic variation, environment and demography intersect to shape Arabidopsis defense metabolite variation across Europe. Schuman MC, Korte A, editors. eLife. 2021;10: e67784. doi:10.7554/eLife.67784

16. Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell. 2021;184: 3542-3558.e16. doi:10.1016/j.cell.2021.04.046

17. Oikawa T, Maeda H, Oguchi T, Yamaguchi T, Tanabe N, Ebana K, et al. The Birth of a Black Rice Gene and Its Local Spread by Introgression. Plant Cell. 2015;27: 2401–2414. doi:10.1105/tpc.15.00310

18. Guo J, Cao K, Deng C, Li Y, Zhu G, Fang W, et al. An integrated peach genome structural variation map uncovers genes associated with fruit traits. Genome Biol. 2020;21: 258. doi:10.1186/s13059-020-02169-y

19. Hu Y, Colantonio V, Müller BSF, Leach KA, Nanni A, Finegan C, et al. Genome assembly and population genomic analysis provide insights into the evolution of modern sweet corn. Nat Commun. 2021;12: 1227. doi:10.1038/s41467-021-21380-4

20. Kramer V, Shaw JR, Senior ML, Hannah LC. The sh2-R allele of the maize shrunken-2 locus was caused by a complex chromosomal rearrangement. Theor Appl Genet. 2015;128: 445–452. doi:10.1007/s00122-014-2443-3

21. Du Y, Luo S, Li X, Yang J, Cui T, Li W, et al. Identification of Substitutions and Small Insertion-Deletions Induced by Carbon-Ion Beam Irradiation in Arabidopsis thaliana. Front Plant Sci. 2017;8. doi:10.3389/fpls.2017.01851

22. Henry IM, Zinkgraf MS, Groover AT, Comai L. A System for Dosage-Based Functional Genomics in Poplar. The Plant Cell. 2015;27: 2370–2383. doi:10.1105/tpc.15.00349

23. Ichida H, Morita R, Shirakawa Y, Hayashi Y, Abe T. Targeted exome sequencing of unselected heavy-ion beam-irradiated populations reveals less-biased mutation characteristics in the rice genome. The Plant Journal. 2019;98: 301–314. doi:10.1111/tpj.14213

24. Ishii K, Kazama Y, Hirano T, Fawcett JA, Sato M, Hirai MY, et al. Genomic view of heavy-ion-induced deletions associated with distribution of essential genes in Arabidopsis thaliana. Front Plant Sci. 2024;15. doi:10.3389/fpls.2024.1352564

25. Jankowicz-Cieslak J, Hofinger BJ, Jarc L, Junttila S, Galik B, Gyenesei A, et al. Spectrum and Density of Gamma and X-ray Induced Mutations in a Non-Model Rice Cultivar. Plants (Basel). 2022;11: 3232. doi:10.3390/plants11233232

26. Ma L, Kong F, Sun K, Wang T, Guo T. From Classical Radiation to Modern Radiation: Past, Present, and Future of Radiation Mutation Breeding. Front Public Health. 2021;9: 768071. doi:10.3389/fpubh.2021.768071

27. Havlickova L, He Z, Berger M, Wang L, Sandmann G, Chew YP, et al. Genomics of predictive radiation mutagenesis in oilseed rape: modifying seed oil composition. Plant Biotechnol J. 2024;22: 738–750. doi:10.1111/pbi.14220

28. Xiong H, Guo H, Fu M, Xie Y, Zhao L, Gu J, et al. A large-scale whole-exome sequencing mutant resource for functional genomics in wheat. Plant Biotechnol J. 2023;21: 2047–2056. doi:10.1111/pbi.14111

29. Zhao Z, Liu Z, Zhou Y, Wang J, Zhang Y, Yu X, et al. Creation of cotton mutant library based on linear electron accelerator radiation mutation. Biochem Biophys Rep. 2022;30: 101228. doi:10.1016/j.bbrep.2022.101228

30. Li F, Shimizu A, Nishio T, Tsutsumi N, Kato H. Comparison and Characterization of Mutations Induced by Gamma-Ray and Carbon-Ion Irradiation in Rice (Oryza sativa L.) Using Whole-Genome Resequencing. G3 Genes|Genomes|Genetics. 2019;9: 3743–3751. doi:10.1534/g3.119.400555

31. Guo W, Comai L, Henry IM. Chromoanagenesis from radiation-induced genome damage in Populus. PLOS Genetics. 2021;17: e1009735. doi:10.1371/journal.pgen.1009735

32. Bolon Y-T, Stec AO, Michno J-M, Roessler J, Bhaskar PB, Ries L, et al. Genome resilience and prevalence of segmental duplications following fast neutron irradiation of soybean. Genetics. 2014;198: 967–981. doi:10.1534/genetics.114.170340

33. Feng Z, Du Y, Chen J, Chen X, Ren W, Wang L, et al. Comparison and Characterization of Phenotypic and Genomic Mutations Induced by a Carbon-Ion Beam and Gamma-ray Irradiation in Soybean (Glycine max (L.) Merr.). International Journal of Molecular Sciences. 2023;24: 8825. doi:10.3390/ijms24108825

34. Yasui Y, Mori M, Aii J, Abe T, Matsumoto D, Sato S, et al. S-LOCUS EARLY FLOWERING 3 Is Exclusively Present in the Genomes of Short-Styled Buckwheat Plants that Exhibit Heteromorphic Self-Incompatibility. PLOS ONE. 2012;7: e31264. doi:10.1371/journal.pone.0031264

35. Koide Y, Ogino A, Yoshikawa T, Kitashima Y, Saito N, Kanaoka Y, et al. Lineage-specific gene acquisition or loss is involved in interspecific hybrid sterility in rice. Proceedings of the National Academy of Sciences. 2018;115: E1955–E1962. doi:10.1073/pnas.1711656115

36. Tanaka H, Muramoto N, Sugimoto H, Oda AH, Ohta K. Extended TAQing system for large-scale plant genome reorganization. The Plant Journal. 2020;103: 2139–2150. doi:10.1111/tpj.14888

37. Wu C-C, Li T-K, Farh L, Lin L-Y, Lin T-S, Yu Y-J, et al. Structural Basis of Type II Topoisomerase Inhibition by the Anticancer Drug Etoposide. Science. 2011;333: 459–462. doi:10.1126/science.1204117

38. Ban Y, Ho C-W, Lin R-K, Lyu YL, Liu LF. Activation of a Novel Ubiquitin-Independent Proteasome Pathway when RNA Polymerase II Encounters a Protein Roadblock. Molecular and Cellular Biology. 2013;33: 4008–4016. doi:10.1128/MCB.00403-13

39. Canela A, Maman Y, Huang SN, Wutz G, Tang W, Zagnoli-Vieira G, et al. Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. Molecular Cell. 2019;75: 252-266.e8. doi:10.1016/j.molcel.2019.04.030

40. Tammaro M, Barr P, Ricci B, Yan H. Replication-Dependent and Transcription-Dependent Mechanisms of DNA Double-Strand Break Induction by the Topoisomerase 2-Targeting Drug Etoposide. PLOS ONE. 2013;8: e79202. doi:10.1371/journal.pone.0079202

41. Uusküla-Reimand L, Wilson MD. Untangling the roles of TOP2A and TOP2B in transcription and cancer. Science Advances. 2022;8: eadd4920. doi:10.1126/sciadv.add4920

42. Amoiridis M, Verigos J, Meaburn K, Gittens WH, Ye T, Neale MJ, et al. Inhibition of topoisomerase 2 catalytic activity impacts the integrity of heterochromatin and repetitive DNA and leads to interlinks between clustered repeats. Nat Commun. 2024;15: 5727. doi:10.1038/s41467-024-49816-7

43. Marchetti F, Bishop JB, Lowe X, Generoso WM, Hozier J, Wyrobek AJ. Etoposide induces heritable chromosomal aberrations and aneuploidy during male meiosis in the mouse. Proceedings of the National Academy of Sciences. 2001;98: 3952–3957. doi:10.1073/pnas.061404598

44. Quispe-Tintaya W, Gorbacheva T, Lee M, Makhortov S, Popov VN, Vijg J, et al. Quantitative detection of low-abundance somatic structural variants in normal cells by high throughput sequencing. Nat Methods. 2016;13: 584–586. doi:10.1038/nmeth.3893

45. Parra-Nunez P, Cooper C, Sanchez-Moran E. The Role of DNA Topoisomerase Binding Protein 1 (TopBP1) in Genome Stability in Arabidopsis. Plants. 2021;10. doi:10.3390/plants10122568

46. Sugimoto-Shirasu K, Roberts GR, Stacey NJ, McCann MC, Maxwell A, Roberts K. RHL1 is an essential component of the plant DNA topoisomerase VI complex and is required for ploidy-dependent cell growth. Proceedings of the National Academy of Sciences. 2005;102: 18736–18741. doi:10.1073/pnas.0505883102

47. Noguchi T, Fujioka S, Choe S, Takatsuto S, Yoshida S, Yuan H, et al. Brassinosteroid-Insensitive Dwarf Mutants of Arabidopsis Accumulate Brassinosteroids1. Plant Physiology. 1999;121: 743–752. doi:10.1104/pp.121.3.743

48. Tan E, Comai L, Henry I. Chromosome Dosage Analysis in Plants Using Whole Genome Sequencing. BIO-PROTOCOL. 2016;6. doi:10.21769/BioProtoc.1854

49. Tan EH, Henry IM, Ravi M, Bradnam KR, Mandakova T, Marimuthu MP, et al. Catastrophic chromosomal restructuring during genome elimination in plants. eLife. 2015;4: e06516. doi:10.7554/eLife.06516

50. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biology. 2014;15: R84. doi:10.1186/gb-2014-15-6-r84

51. Bromberg KD, Burgin AB, Osheroff N. A Two-drug Model for Etoposide Action against Human Topoisomerase IIα *. Journal of Biological Chemistry. 2003;278: 7406–7412. doi:10.1074/jbc.M212056200

52. Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. Genome Biology. 2019;20: 246. doi:10.1186/s13059-019-1828-7

53. Naish M, Alonge M, Wlodzimierz P, Tock AJ, Abramson BW, Schmücker A, et al. The genetic and epigenetic landscape of the Arabidopsis centromeres. Science. 2021;374: eabi7489. doi:10.1126/science.abi7489

54. Jiang T, Cao S, Liu Y, Liu S, Liu B, Wang G, et al. Regenotyping structural variants through an accurate force-calling method. bioRxiv; 2023. p. 2022.08.29.505534. doi:10.1101/2022.08.29.505534

55. Jiang T, Liu Y, Jiang Y, Li J, Gao Y, Cui Z, et al. Long-read-based human genomic structural variation detection with cuteSV. Genome Biology. 2020;21: 189. doi:10.1186/s13059-020-02107-y

56. Byrne ME, Simorowski J, Martienssen RA. ASYMMETRIC LEAVES1 reveals knox gene redundancy in Arabidopsis. Development. 2002;129: 1957–1965. doi:10.1242/dev.129.8.1957

57. Ha CM, Jun JH, Nam HG, Fletcher JC. BLADE-ON-PETIOLE1 and 2 Control Arabidopsis Lateral Organ Fate through Regulation of LOB Domain and Adaxial-Abaxial Polarity Genes. The Plant Cell. 2007;19: 1809–1825. doi:10.1105/tpc.107.051938

58. Ikezaki M, Kojima M, Sakakibara H, Kojima S, Ueno Y, Machida C, et al. Genetic networks regulated by ASYMMETRIC LEAVES1 (AS1) and AS2 in leaf development in Arabidopsis thaliana: KNOX genes control five morphological events. The Plant Journal. 2010;61: 70–82. doi:10.1111/j.1365-313X.2009.04033.x

59. Jun JH, Ha CM, Fletcher JC. BLADE-ON-PETIOLE1 coordinates organ determinacy and axial polarity in arabidopsis by directly activating ASYMMETRIC LEAVES2. Plant Cell. 2010;22: 62–76. doi:10.1105/tpc.109.070763

60. Sun Y, Zhou Q, Zhang W, Fu Y, Huang H. ASYMMETRIC LEAVES1, an Arabidopsis gene that is involved in the control of cell differentiation in leaves. Planta. 2002;214: 694–702. doi:10.1007/s004250100673

61. Qiu K, Li Z, Yang Z, Chen J, Wu S, Zhu X, et al. EIN3 and ORE1 Accelerate Degreening during Ethylene-Mediated Leaf Senescence by Directly Activating Chlorophyll Catabolic

Genes in Arabidopsis. PLOS Genetics. 2015;11: e1005399. doi:10.1371/journal.pgen.1005399

62. Zhong S, Zhao M, Shi T, Shi H, An F, Zhao Q, et al. EIN3/EIL1 cooperate with PIF1 to prevent photo-oxidation and to promote greening of Arabidopsis seedlings. Proceedings of the National Academy of Sciences. 2009;106: 21431–21436. doi:10.1073/pnas.0907670106

63. Fu A, Park S, Rodermel S. Sequences Required for the Activity of PTOX (IMMUTANS), a Plastid Terminal Oxidase: In vitro and in planta mutagenesis of iron-binding sites and a conserved sequence that corresponds to exon 8 *. Journal of Biological Chemistry. 2005;280: 42489–42496. doi:10.1074/jbc.M508940200

64. Wetzel CM, Jiang C-Z, Meehan LJ, Voytas DF, Rodermel SR. Nuclear—organelle interactions: the immutans variegation mutant of Arabidopsis is plastid autonomous and impaired in carotenoid biosynthesis. The Plant Journal. 1994;6: 161–175. doi:10.1046/j.1365-313X.1994.6020161.x

65. Pommier Y. Topoisomerase I inhibitors: camptothecins and beyond. Nat Rev Cancer. 2006;6: 789–802. doi:10.1038/nrc1977

66. Weimer AK, Biedermann S, Harashima H, Roodbarkelari F, Takahashi N, Foreman J, et al. The plant-specific CDKB1-CYCB1 complex mediates homologous recombination repair in Arabidopsis. The EMBO Journal. 2016;35: 2068–2086. doi:10.15252/embj.201593083

67. Hase Y, Satoh K, Kitamura S. Comparative analysis of seed and seedling irradiation with gamma rays and carbon ions for mutation induction in Arabidopsis. Front Plant Sci. 2023;14. doi:10.3389/fpls.2023.1149083

68. Rédei GP. Non-Mendelian Megagametogenesis in Arabidopsis. Genetics. 1965;51: 857–872.

69. Rodermel S. Arabidopsis Variegation Mutants. Arabidopsis Book. 2002;1: e0079. doi:10.1199/tab.0079

70. Maloof JN, Nozue K, Mumbach MR, Palmer CM. LeafJ: An ImageJ Plugin for Semi-automated Leaf Shape Measurement. JoVE. 2013; 50028. doi:10.3791/50028

71. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170

72. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv; 2013. doi:10.48550/arXiv.1303.3997

73. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

74. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10: giab008. doi:10.1093/gigascience/giab008

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

76. Zverinova S, Guryev V. Variant calling: Considerations, practices, and developments. Hum Mutat. 2022;43: 976–985. doi:10.1002/humu.24311

77. Schalamun M, Nagar R, Kainer D, Beavan E, Eccles D, Rathjen JP, et al. Harnessing the MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant tissue from Eucalyptus pauciflora. Molecular Ecology Resources. 2019;19: 77–89. doi:10.1111/1755-0998.12938

78. Wick R. rrwick/Filtlong. 2024. Available: https://github.com/rrwick/Filtlong

79. Fu Y, Mahmoud M, Muraliraman VV, Sedlazeck FJ, Treangen TJ. Vulcan: Improved long-read mapping and structural variant calling via dual-mode alignment. GigaScience. 2021;10: giab063. doi:10.1093/gigascience/giab063

80. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34: 3094–3100. doi:10.1093/bioinformatics/bty191

81. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018;15: 461–468. doi:10.1038/s41592-018-0001-7

82. Jiang T, Cao S, Liu Y, Liu S, Liu B, Wang G, et al. Regenotyping structural variants through an accurate force-calling metho. 2022. doi:10.1101/2022.08.29.505534

83. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun. 2017;8: 14061. doi:10.1038/ncomms14061

84. Krueger F. FelixKrueger/TrimGalore. 2024. Available: https://github.com/FelixKrueger/TrimGalore

85. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14: 417–419. doi:10.1038/nmeth.4197

86. Chen X, Zhang B, Wang T, Bonni A, Zhao G. Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. BMC Bioinformatics. 2020;21: 269. doi:10.1186/s12859-020-03608-0

87. Todorov V, Filzmoser P. An Object-Oriented Framework for Robust Multivariate Analysis. Journal of Statistical Software. 2010;32: 1–47. doi:10.18637/jss.v032.i03

88. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550. doi:10.1186/s13059-014-0550-8

89. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2: 100141. doi:10.1016/j.xinn.2021.100141

90. Blighe K, Rana S, Lewis M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. 2018. Available: https://github.com/kevinblighe/EnhancedVolcano

91. Kolde R. pheatmap: Pretty Heatmaps. 2019. Available: https://cran.r-project.org/web/packages/pheatmap/index.html