OXFORD GENETICS

Rapid large-scale genomic introgression in *Arabidopsis suecica* via an autoallohexaploid bridge

Vidya Oruganti,[1] Helena Toegelová (ID) ,[2] Aleš Pečinka (ID) ,[2] Andreas Madlung (ID) ,[3,*,†] Korbinian Schneeberger (ID) [1,4,5,*,†]

[1]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany,
[2]Institute of Experimental Botany of the Czech Academy of Sciences, 77900 Olomouc, Czech Republic,
[3]Department of Biology, University of Puget Sound, Tacoma, WA 98416, USA,
[4]Department of Genetics, Faculty of Biology, Ludwig Maximilian Universität München, 82152 Planegg-Martinsried, Germany,
[5]Cluster of Excellence on Plant Sciences, Heinrich-Heine University, Düsseldorf, Germany

*Corresponding author: University of Puget Sound, Department of Biology, 1500 N Warner St, CMB 1088, Tacoma, WA, 98416, USA. Email: amadlung@pugetsound.edu; *Corresponding author: Faculty of Biology, Ludwig Maximilian Universität München, 82152 Planegg-Martinsried, Germany. Email: k.schneeberger@lmu.de
[†]These authors contributed equally to this work.

Abstract

Gene flow between species in the genus Arabidopsis occurs in significant amounts, but how exactly gene flow is achieved is not well understood. Polyploidization may be one avenue to explain gene flow between species. One problem, however, with polyploidization as a satisfying explanation is the occurrence of lethal genomic instabilities in neopolyploids as a result of genomic exchange, erratic meiotic behavior, and genomic shock. We have created an autoallohexaploid by pollinating naturally co-occurring diploid *Arabidopsis thaliana* with allotetraploid *Arabidopsis suecica* (an allotetraploid composed of *A. thaliana* and *Arabidopsis arenosa*). Its triploid offspring underwent spontaneous genome duplication and was used to generate a multigenerational pedigree. Using genome resequencing, we show that 2 major mechanisms promote stable genomic exchange in this population. Legitimate meiotic recombination and chromosome segregation between the autopolyploid chromosomes of the 2 *A. thaliana* genomes occur without any obvious bias for the parental origin and combine the *A. thaliana* haplotypes from the *A. thaliana* parent with the *A. thaliana* haplotypes from *A. suecica* similar to purely autopolyploid plants. In addition, we repeatedly observed that occasional exchanges between regions of the homoeologous chromosomes are tolerated. The combination of these mechanisms may result in gene flow leading to stable introgression in natural populations. Unlike the previously reported resynthesized neoallotetraploid *A. suecica*, this population of autoallohexaploids contains mostly vigorous, and genetically, cytotypically, and phenotypically variable individuals. We propose that naturally formed autoallohexaploid populations might serve as an intermediate bridge between diploid and polyploid species, which can facilitate gene flow rapidly and efficiently.

Keywords: polyploidy; autoallopolyploidy; hexaploid; gene flow; introgression; homoeologous exchange; aneuploidy; evolution; Arabidopsis

Introduction

Reproductive isolation is central to the theory of speciation. Hybridization transcends this theory and while often hybrids are evolutionary dead ends. Hybridization transcends this theory and while often hybrids are evolutionary dead ends, in some cases, especially in plants, hybridization events can lead to speciation (Mallet 2007). A special case of hybrid speciation is that of allopolyploidy, where genome duplication simultaneous with hybridization allows the newly formed genome to overcome the challenges of imperfect homolog pairing of diploid hybrids during meiosis (Comai 2005). Because allopolyploids have paired homologous chromosomes and in principle are therefore able to undergo normal meiosis, allopolyploidy has been described as a mechanism for instant speciation (Linder and Rieseberg 2004; Pearse *et al.* 2006). Although genomic stability and vigor are commonly found in established allopolyploids, early generation

allopolyploids (neoallopolyploids), frequently display phenotypic and cytological instabilities and are often characterized by low fitness (Comai *et al.* 2000; Xiong *et al.* 2011; Chester *et al.* 2012; Zhang *et al.* 2013).

Allopolyploids may benefit from various genomic and epigenetic changes following allopolyploidization (Chen 2007; Madlung and Wendel 2013) but genomic change also brings dangers and genomic conflicts. The similarity between homoeologous chromosomes can, for example, lead to exchanges of chromosome arms and rearrangements between nonhomologous chromosomes. In addition, multivalent chromosome pairing can result in abnormalities during cell division and the formation of aneuploids, a danger that is particularly frequent in autopolyploids, which arise from whole-genome duplication. How the resolution of these conflicts between 2 genomes in a neopolyploid is achieved and how genomic change in response to polyploidy

affects the fitness of the new species is still controversial and largely unknown (Mayrose *et al.* 2011; Soltis *et al.* 2014). Experiments with synthetic neoallopolyploids have suggested that from an interspecies polyploid cross only few individuals stabilize their genomes throughout the subsequent generations, whereas the majority of neopolyploids is either inviable or infertile (Comai *et al.* 2000; Gaeta *et al.* 2007; Gaeta and Pires 2010). Unless genomic stability is restored after allopolyploidization, genomic instability, such as in the form of aneuploidy, can accumulate from generation to generation. This spiraling process has been termed the "polyploid ratchet" (Gaeta and Pires 2010), where genomic errors accumulate over generations until fitness is compromised to the point of extinction.

While genomic rearrangements and aneuploidy constitute genetic errors that can lead to detrimental mutations, legitimate meiotic recombination may be viewed as a genetic mechanism that could facilitate intergenomic gene flow in neoallopolyploids. Whether or not, and if so to what extent, this occurs is not well understood. Strict pairing of homologous chromosomes during meiosis reduces the risk of monovalents during cell division and thus of aneuploid gamete formation. Genetic loci that reduce the threat of intergenomic (homoeologous) pairing have been described in allohexaploid wheat (Griffiths *et al.* 2006), allotetraploid *Brassica napus* (Jenczewski *et al.* 2003), and in allotetraploid *Arabidopsis suecica* (Henry *et al.* 2014). It is possible that loss or gain of these (and maybe other) loci in allopolyploids play a role in stabilizing neoallopolyploids. A study of the genus Arabidopsis showed that gene flow can occur between species inconsistent with the notion that speciation gives rise to new species exclusively in a linear manner or via bifurcating phylogenetic trees (Novikova *et al.* 2016). In their study, Novikova and colleagues (2016) showed that admixture is frequent in the genus Arabidopsis and suggested that evolution in this genus cannot be explained by mutation and divergence alone.

To shed more light on the question what mechanisms might explain how gene flow and introgression in the genus Arabidopsis occurs, we worked with a pedigree that was created by pollinating a diploid (*Arabidopsis thaliana*, Col-0, $2n = 2x = 10$) with an allotetraploid (*A. suecica*, Sue-1, $2n = 4x = 26$), whose genome itself is composed of those of *A. thaliana* and *Arabidopsis arenosa*. Here, we show that in this autoallohexaploid controlled meiotic recombination can promote rapid integration of genomic material from *A. thaliana* into *A. suecica* (introgression), while out-of-control mechanisms that result in aneuploidy, instability, and the eventual removal of the line from the gene pool effectively act against evolutionarily stable introgression and gene flow. Our data further suggest that an autoallohexaploid might efficiently serve as a bridge between different, related species and might thus provide a mechanism for rapid admixture between genomes.

Materials and methods
Material
The auto-allohexaploid pedigree was created as previously described (Matsushita *et al.* 2012). Briefly, an individual of the allotetraploid *A. suecica* Sue-1 (ABRC accession numbers CS22505) was used as the pollen donor in a cross with diploid *A. thaliana* (Col-0). The resulting F1 offspring was ascertained to be triploid using flow cytometry. Two branches of the otherwise infertile plant became fertile, probably after spontaneous somatic doubling. F2 individuals were shown to be mostly hexaploid using flow cytometry. F2 lines were propagated by selfing and single-seed descent (Fig. 1) down to the F_{14} to F_{17} (depending on line). For this study, seeds saved from early generations were planted alongside
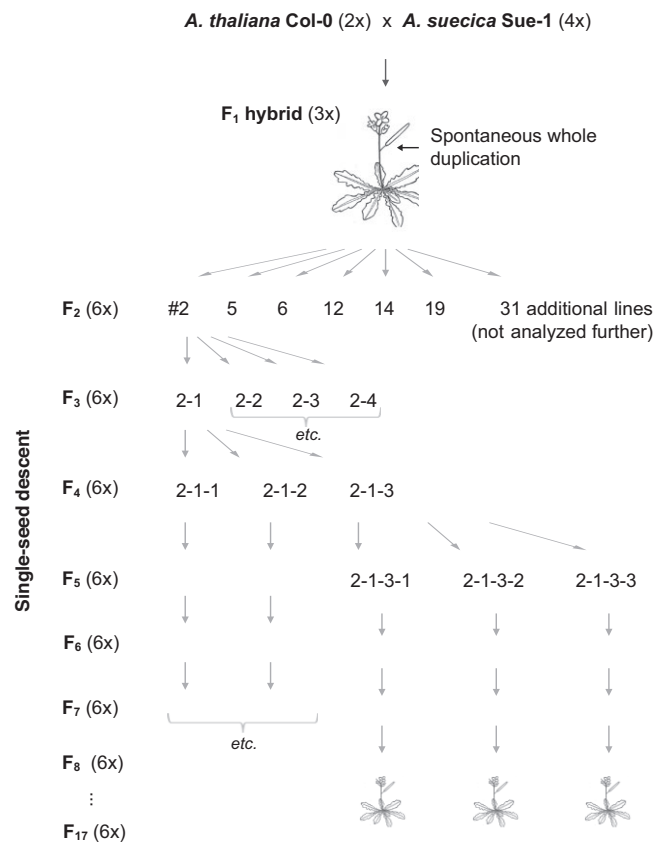

Fig. 1. Pedigree of the experimental population. Diploid *A. thaliana* (Col-0) was used as maternal parent in a cross with allotetraploid *A. suecica* (Sue-1). The triploid F1 individual was sterile with the exception of 2 sectors that arose from cells that had undergone somatic doubling. A total of 37 seeds were recovered and grown (F2). These seeds founded each of the numbered lines 1–37, of which 6 (numbered 2, 5, 6, 12, 14, and 19) were propagated further by single-seed descent. These plants were ascertained by flowcytometry to be generally hexaploid (Matsushita *et al.* 2012). The lines were propagated through the current 17th generation. Seeds from each individual used for line propagation were set aside. At the time of analysis for this study, 3 seeds from selected lines and generations were grown at the same time in greenhouse-controlled conditions. This figure only shows a small subset of the complete pedigree. The entire list of participating individuals can be found in Supplementary Tables 1 and 2. The pedigree figure is reproduced from Matsushita *et al.* (2012).

mid- and late generations. To do so, multiple seeds saved from individuals that had been used for propagating the line were planted to produce same-generation sisters. These sisters thus represent short side branches to the descending line. During the production of the pedigree, lines that readily produced seed were preferred, resulting in the selection of fertility traits.

Autoallohexaploids as well as *A. suecica* plants were grown on soil in 10-cm diameter plastic pots in a greenhouse with supplemental lighting under long-day conditions (16 h light/8h dark) or in an incubator with the same light conditions and a temperature of $20°C$. For phenotypic analysis, same-age plants were photographed and their rosette diameter measured using ImageJ. Seeds of same-age plants were later harvested, dried, and weighed to estimate fitness. For short-read sequencing, leaf material was harvested when plants were 7 weeks of age, snap frozen in liquid nitrogen, and stored at $-80°C$ until used. DNA was extracted using a Qiagen Plant DNeasy kit and sequenced at the sequencing center of the Max Planck Institute for Plant Breeding Research.

Assembly and annotation of *A. suecica* Sue-1

For the genome assembly, DNA was extracted from a single *A. suecica* Sue-1 individual provided to the Max Planck Genome Centre Cologne, which extracted the DNA that was used for PacBio sequencing on a Sequel II instrument. Sequencing reads were generated at 50× coverage comprising a total length of ~23.8 Gb. Using Canu software v1.5 (Koren *et al.* 2017), the *A. suecica* long reads were corrected, trimmed, and assembled into unitigs using the parameters corOutCoverage = 200, batOptions=-dg 3 -db 3 -dr 1 -ca 500 -cp 50. The assembly was corrected with Pilon v1.22 (Walker *et al.* 2014) using *A. suecica* Illumina short paired-end reads. For this, the short reads were first aligned against the assembly and the subsequent BAM file was sorted and indexed using SAMtools (Li *et al.* 2009), and subsequently used as input for the Pilon tool. Optical mapping was performed using the nicking enzyme *Nb.BssSI*. The high molecular weight DNA was labeled and stained with the Bionano Prep™ DNA Labeling Kit and analyzed on the Irys platform (Bionano Genomics). Molecules greater than 100 kb were used for the de novo optical map assembly run in the Irys Solve software version 2.1.1. To align the long-read assembly with the optical consensus maps, the PacBio assembly was in silico digested with *Nb.BssSI* to generate a reference consensus map. Hybrid scaffolding was executed by using the set of the perl scripts integrated in the automated scaffolding pipeline in the Bionano Solve software package. The resulting contigs were aligned against the pseudoreference genome of the concatenated *A. thaliana* Col-0 and *Arabidopsis lyrata* (https://plants.ensembl.org; accessed 2022 September 8, v1.0; Hu *et al.* 2011) genomes using NUCmer from the MUMmer package v3.5 (Kurtz *et al.* 2004). The parameters for NUCmer were set to—maxmatch and a minimum cluster size of 100. When *A. suecica* short reads were aligned against the pseudoreference using Bowtie2 v2.2.8 (Langmead and Salzberg 2012), a significant coverage increase was observed at the end of *At2* which was not reciprocated by the syntenic region at the end of *A. lyrata* chromosome 4 and thus was considered to be a duplication. In order to rectify this, the end of the *At2* was copied onto the end of the *Aa4* in the assembly. The centromeres of the *A. arenosa* subgenome could not be assembled due to the lack of centromeres in the *A. lyrata* v1.0 sequence.

The syntenic relationships between the 2 subgenomes in the final assembly were established using SyRI v1.2 (Goel *et al.* 2019). The subgenome chromosome sequences were first combined in blocks according to the syntenic relationship between *A. thaliana* Col-0 and *A. lyrata* (Hu *et al.* 2011). *Arabidopsis arenosa* chromosomes 1 and 2 were then combined into 1 sequence and aligned using NUCmer of the MUMmer package with similar settings as before (–maxmatch and minimum cluster size of 100) against *A. thaliana* subgenome chromosome 1. Chromosomes 3, 4, and 5 of *A. arenosa* were combined and aligned against chromosomes 2 and 3 of *A. thaliana*. Similarly, chromosomes 6, 7, and 8 of *A. arenosa* were aligned against chromosomes 4 and 5 of *A. thaliana*. These alignments were then used as input for the SyRI software and the resulting synteny between the subgenomes was plotted with SyRI's plotsr. An inversion was reported between the end of *A. thaliana* subgenome chromosome 4 and the whole of the *A. arenosa* subgenome chromosome 6 as well as part of *A. arenosa* subgenome chromosome 7. As this inversion conflicted with a duplication between the *A. thaliana* and *A. arenosa* subgenomes reported in part of the same region, the inversion was shortened to be only between the end of *A. thaliana* subgenome chromosome 4 and the beginning part of *A. arenosa* subgenome chromosome 7. This observation is similar to what has been reported in

the synteny analysis between *A. thaliana* Col-0 and *A. lyrata* (Hu *et al.* 2011).

The annotation of the assembly was carried out as follows: Protein hints were first generated using prior peptide information from *A. thaliana* Col-0 and *A. lyrata* (v1.0). The program Exonerate v2.2.0 (Slater and Birney 2005) was used to align the protein sequences to the assembly. RNA sequence data from previous work (Carlson *et al.* 2017) was used to generate RNA hints and aid in the annotation. For this, the RNA sequence accessions SRR3676009, SRR3676011, and SRR3676012 (BioProject PRJNA321210) were downloaded from the NCBI. The programs Stringtie v2.1.5 (Pertea *et al.* 2015) and HiSat2 v2.1.0 (Kim *et al.* 2019) were used to align the RNA sequences against the assembly and assemble the alignments into potential transcripts. Ab initio runs of the annotation software Augustus 3.4.0 (Stanke *et al.* 2006), Semi-HMM-based Nucleic Acid Parser (SNAP) version 2006-07-28 (Johnson *et al.* 2008), and GlimmerHMM (Majoros *et al.* 2004) were carried out. The EVidenceModeler (Haas *et al.* 2008) software combined with the Program to Assemble Spliced Alignments (Haas *et al.* 2003) was then used to combine the annotations predicted based on user-defined weights provided in a configuration file.

After annotating the protein-coding genes, the program RepeatMasker version open-4.0.9 (http://www.repeatmasker.org) was used to identify transposable elements (TEs) in the assembly sequence. The tool intersectBed from the BEDtools suite v2.21.0 (Quinlan and Hall 2010) was then used to find overlaps between the coordinates of the predicted TEs and those of the protein-coding genes. If TEs overlapped 20% or more of the protein-coding gene, the region was annotated as TE-related. The program LiftOff v1.6.1 (Shumate and Salzberg 2021) was used to correct the strand specificity in the annotation file for a contig inversion in chromosome 3 of the assembly. This inversion was also corrected in the actual assembly sequence.

Here, the 13 chromosomes of the *A. suecica* genome are indicated by the labels "Chr1" to "Chr13," where the first 5 chromosomes refer to the *A. thaliana* subgenome and the last 8 chromosomes to the *A. arenosa* subgenome.

Analysis of meiotic recombination in the autoallohexaploid genomes

Short paired-end reads generated using Illumina technology from individual progeny genomes were aligned against the Sue-1 assembly using Bowtie2 with the default parameters (in end-to-end mode). The BAM files were then sorted and indexed using SAMtools sort and SAMtools index, respectively. The per-base depth of coverage for each alignment was obtained using genomecov from BEDtools, which gave an output file of the cumulative read count per position of each chromosome in each sample. The length of each chromosome was divided into bins of 100 to 500 kb (with stepsizes of 10 kb for 100 kb windows) and the average coverage per base calculated per window normalized by the chromosome length. R ggplot2 was then used to plot the sliding window coverages normalized on a scale between 0 and 100. The only noise was observed in the centromeric regions which stems exclusively from the *A. thaliana* chromosomes as the centromeres are not sequenced in the *A. arenosa* chromosomes.

To generate a reference marker list of SNP alleles, short reads of *A. thaliana* Col-0 obtained from the public Arabidopsis database TAIR were aligned against the *A. suecica* assembly as well as to the other samples, and variant calling and further filtering of the VCF file were performed. A marker list of Col-0 parental SNP alleles (that differ from the *A. suecica* SNP alleles) was generated

to ensure that the variants not matching the SNP alleles in the *A. suecica* parent matched the SNP alleles derived from the Col-0 parent. Only biallelic sites in the samples were considered. Variant calling was performed for all the samples using SAMtools mpileup and BCFtools call from SAMtools v1.3.1 (Li *et al.* 2009). Using SAMtools mpileup anomalous read pairs were retained with the parameter -A and the minimum base quality threshold -Q was set to 0. BCFtools call was run with the option -A to report all possible alternate alleles at variant sites. The variants were then filtered to generate high-quality SNPs by selecting only those whose quality scores were at least 10, and using BCFtools to filter out indels. A sliding window variant analysis was carried out for each individual with windows ranging from 100 to 300 kb and stepsizes of 10 to 30 kb. For each window, the proportion of SNP alleles deriving from the *A. suecica* parent vs the *A. thaliana* Col-0 parent was calculated by dividing the number of *A. suecica* SNP alleles over the total number of SNP alleles. As the genome is comprised of an *A. thaliana* subgenome and an *A. arenosa* subgenome SNP alleles derived from the Col-0 parent were only seen in the sliding window analysis of the *A. thaliana* subgenome. R ggplot2 was used to plot the SNP allele frequencies on a scale of 0 to 1 and overlayed onto the coverage plots. Those windows for which the total number of SNP alleles did not exceed 300 were excluded as these may not give an accurate representation of the ratio of the parental alleles. A frequency or proportion of 1.0 indicates that all the SNP alleles are derived from *A. suecica* while 0 indicates all of them are derived from the Col-0 parent.

The sliding window analyses of the SNP allele frequencies were used to estimate the extent of homologous recombination across the 4 chromosome copies within the *A. thaliana* subgenome as follows: For each chromosome sudden, sustained shifts in SNP allele frequencies along the length of the chromosome were counted using SNP frequency graphs with a 300-kb window size rendered in R with ggplot2. The total number of shifts indicates the frequency with which cross-overs occur along a given chromosome, not the accumulated number of recombination events since the original cross.

Identification and validation of aneuploidies

In order to investigate whole chromosome loss or gain within the *A. thaliana* subgenome, a coverage-based optimization method was used. In this method, a reference with a known karyotype (here *A. suecica*) was used as a base for the prediction of copy numbers in the sample. Based on the aligned *A. suecica* reads, we established a reference coverage per copy of each chromosome (*i*), r_i, which was subsequently used to calculate an expected coverage for each chromosome if the copy number (x_i) is varied. As the search space of different combinations of copy numbers is traversed, the proportion of the expected coverage over the updated total is updated. Due to the condition of the total number of reads needing to stay the same, the updated proportion is normalized by the original total to get the actual expected coverage per chromosome. Each time the copy combination is varied and expected coverage updated, the deviation from the observed proportion of reads in the sample over the total observed reads for that particular chromosome is calculated. These differences are then summed up over the total number of chromosomes, and the optimal copy combination *x* that minimizes the sum of the differences is chosen as the predicted karyotype of the sample.

The general feasible region for *x* is [0, inf] can give rise to invariant or multiple optimal solutions. In order to prevent this, in the naïve method for this particular dataset, the feasible region was shrunk to [2, 6] as the most likely number of copies is no

than 6 in this case. To further narrow down the optimal solutions, an upper limit of 2 copies was assumed for the *A. arenosa* chromosomes. Using the coverage information for each sample, a method selecting the optimal copy combination which minimizes the sum of the differences between the expected (for a known karyotype) and observed coverages was used. Placing an upper limit of 2 copies on the *A. arenosa* chromosome copies reduced the number of invariant or multiple optimal solutions and thus focused us on the aneuploidy predictions in the *A. thaliana* subgenome. This assumption allowed the following approach: Let there be *n* chromosomes in a particular sample and let the observed number of reads be y_i for each chromosome *i* in the sample. The sum of the observed reads over the *n* chromosomes is given by *Y*. For a reference genome with a known karyotype, let the coverage per copy for each chromosome *i* be given by r_i and let the total number of reads for the reference over all the chromosomes be *T*. For a given number of copies for a particular chromosome *i*, x_i, the expected number of reads for chromosome *i* is: $(r_i \cdot x_i/N) \cdot T$ where *N* refers to the updated total amount of expected reads after changing the copy number of chromosome *i*. However, as the total number of reads for the reference should stay the same, the coverage computed for chromosome *i* was normalized against the original total for the reference (*T*) to give the actual expected number of reads. The copy combination that minimized the sum of the differences between the proportion of total observed reads and the proportion of total expected reads belonging to chromosome *i* was computed as follows: $argmin_x \sum_{i=1}^{n} \left| \frac{y_i}{Y} - \frac{r_i \cdot x_i}{N} \right|$. The method was implemented in Python.

To verify the results from this method, we also used the allele frequency we had generated for the analysis of meiotic recombination frequencies. As, for example, a partental allele frequency of 0.25 or 0.75 unambiguously indicates a parental ratio of 1:3 or 3:1 copies (and thus the expected copy number of 4), if we ignore extremely unlikely chromosome counts like 6:2 or 12:4, we used histograms of the sliding windows to determine the frequency of windows occurring at each allele frequency from 0.0 to 1.0. Peaks in the number of sliding windows corresponding to certain allele frequency ratio were used to validate the aneuploidy prediction for that particular sample. In cases where the allele frequency was 0.0, 0.5, or 1.0, these could not be unambiguously validated as these frequencies could result in more than one possible copy number.

Identification of homoeologous exchanges between *A. thaliana* and *A. arenosa* subgenomes

Significant coverage shifts occurring in the syntenic regions between *A. thaliana* and *A. arenosa* were determined using sliding window analysis of the coverage. The mean coverage of all the windows in a particular sample along a particular chromosome was computed. Coverage shifts of one copy or more at the chromosome arm ends were deemed as significant. The read density was plotted to observe the changes in this region and identify those regions where shifts and possible exchanges occurred. The approximate regions of the breakpoints were visualized with the Integrative Genome Viewer (Robinson *et al.* 2011).

Results and discussion
A multigenerational hexaploid pedigree of a single autoallopolyploidization event shows phenotypic diversity

To find mechanisms that might explain gene flow between congeners resulting in stable introgressions in just a few generations, we

created a multigenerational pedigree of autoallohexaploids (Fig. 1). The mother plant resulted from a single cross of diploid *A. thaliana* (Col-0) whose genome we designated AA, and the natural allotetraploid *A. suecica* (Sue-1) whose genome we designated A'A'BB. Seeds from the original cross were mostly inviable, except for 1 viable plant. This resulting triploid F_1 plant was sterile except for 2 late-developing side-branches that had undergone spontaneous mitotic whole-genome duplication. These branches bore several siliques with a total of 37 viable seeds. Flow cytometry was used to verify ploidy levels and ascertained additive genome sizes of all analyzed F_2 individuals with the genomic constitution AAA'A'BB (Matsushita *et al.* 2012). Individuals within and between the lines showed a high degree of phenotypic variability with respect to flowering time, leaf serration, rosette diameter, and fecundity, but we observed no obvious phenotypic patterns that changed with increasing generation number (Fig. 2, Supplementary Fig. 1a). Fitness in this pedigree was assessed by measuring seed mass in a subset of the plants. Only 1 plant (4B in Fig. 2) was completely sterile, all other plants produced plump seeds. Aside from a few particularly fertile plants in F_{16}, individuals in all other generations produced seeds that were nonsignificantly different in weight from each other (Supplementary Fig. 1b).

For the current study, we advanced 6 of these lines via single seed descent over a total of 17 generations. We then grew multiple sisters from several generations of each line in the greenhouse to observe their phenotypes and to extract DNA for whole-genome sequencing. Each sample was sequenced with Illumina short-read sequencing with an average of 20× to 40× genome coverage per sample.

A genome assembly of *A. suecica* Sue-1 reveals a high degree of synteny between its subgenomic chromosomes

To ask questions about the genomic constitution of individual hexaploid plants, we first created a de novo chromosome-level assembly of the paternal parent, *A. suecica* (Sue-1). Initially, PacBio long reads were assembled into 482 contigs with a total length of 271.6 Mb, which were then polished with Illumina short reads. We also generated 453 optical consensus maps with a total length of 278.4 Mb to correct errors in the assembly and combine the contigs to scaffolds. Upon hybrid scaffolding using the optical consensus maps, the N50 of the assembly improved from 3.2 to 5.7 Mb and the number of PacBio contigs was reduced to 78 scaffolds with a maximum length of more than 15 Mb. The total length of the final assembly sequence including the nonscaffolded contigs was 275.3 Mb, closely resembling in length 2 recently reported assemblies of *A. suecica* measuring 272.4 Mb for accession As9502 (aka CS22509 or Sue-5; Wang *et al.* 2006; Jiang *et al.* 2021), and 262 Mb for accession ASS3 (Burns *et al.* 2021). As has been done before (Henry *et al.* 2014; Novikova *et al.* 2016), we generated a pseudoreference allopolyploid genome by concatenating the *A. thaliana* (Col-0) and *A. lyrata* reference sequences and used it as a guide for the generation of a final chromosome-level assembly. The final assembly (N50 = 19.06 Mb) included the sequences for the 13 chromosomes as well as nonscaffolded sequences and scaffolds that were not placed on chromosomes. The annotation of the *A. suecica* genome resulted in 60,688 protein-coding genes while the number of annotated TEs was 115,872.

Some inversions and other structural variations remained between the final assembly and the pseudoreference, especially between the *A. arenosa* subgenome and the *A. lyrata* reference, which we assumed to represent genuine differences between the 2 closely related species. In the alignment of the new assembly

with the Col-0 accession, an inversion in chromosome 4 could be seen adjacent to the centromere. This inversion was already described (Fransz 2000) and has previously been reported to be present in other *A. thaliana* accessions (Zapata *et al.* 2016). Some of the structural variations between the *A. thaliana* and *A. arenosa* subgenomes of *A. suecica* (shown in Fig. 3) were similar to those observed between *A. thaliana* Col-0 and *A. lyrata* (Hu *et al.* 2011). These included an inversion between *A. thaliana* subgenome chromosome 1 (*At1*) and *A. arenosa* subgenome chromosome 2 (*Aa2*), and the inversion between *At5* and *Aa8*. The translocation identified between the beginning of *At3* and the beginning of *Aa3* as well as the duplications between the beginning of *At5* and the beginning of *Aa6* are also found between *A. thaliana* Col-0 and *A. lyrata* (Hu *et al.* 2011).

In summary, comparing our assembly of Sue-1 with the recently published *A. suecica* assemblies (Burns *et al.* 2021; Jiang *et al.* 2021) and genomic comparisons within the Arabidopsis genus (Hu *et al.* 2011) gave us confidence that this assembly, representing the paternal parent in our pedigree, would serve as a high-quality reference for the analysis of the autoallohexaploid population.

Meiotic recombination occurs between all 4 *A. thaliana* haplotypes

The autoallohexaploid used in this study contained the allotetraploid genome of *A. suecica* and the diploid genome of *A. thaliana*. The allotetraploid *A. suecica* was itself formed by a hybridization of *A. thaliana* and *A. arenosa* and it thus contains both homologous and homoeologous chromosomes (i.e. 4 haplotypes of *A. thaliana* and 2 of *A. arenosa*). Since the *A. thaliana* haplotypes (AA and A'A') in the hexaploid progeny were derived from 2 different parental *A. thaliana* genotypes, the *A. thaliana* subgenome was heterozygous in itself (A vs A') including the Col-0 accession (from the *A. thaliana* parent) and an accession similar to Borsk-2 from the Sue-1 parent (Carlson *et al.*, 2017). The 2 *A. thaliana*-derived genomes were both homozygous implying that the 2 chromosomes from the same parent could not be distinguished. However, the haplotypes from different parents could easily be distinguished using SNP markers. Such genomic state suggests that both disomic and tetrasomic segregation could occur (Tate *et al.* 2005).

To determine the allelic makeup of the *A. thaliana* subgenomes in each of the hexaploid individuals, we aligned the Illumina short reads of the autoallohexaploids against the pseudoreference and the Sue-1 assembly. To determine if recombination had occurred between different *A. thaliana* haplotypes, we assessed the allele frequencies of Col-0 (AA) and Sue-1 (A'A') along the chromosomes using read counts supporting either of the alleles at a chromosome-wide set of SNP markers within sliding windows (Fig. 4).

If both subgenomes, A and A', contribute exactly half of the hexaploid's *A. thaliana* complement we would expect an allele frequency of 50% for both of the alleles. Any deviation suggests that entire chromosomes or some chromosomal regions were more frequently derived from Col-0 or from Sue-1, for example allele frequencies of 0.25 or 0.75 would indicate that one of the parental haplotypes was represented in 1 or 3 copies, respectively. Chromosome segregation during meiosis, for example, can increase or decrease the number of one or the other parental haplotypes. Moreover, shifts in the allele frequency along a single chromosome are an indication of meiotic recombination (i.e. a crossover) between A and A', examples of which (along with a cartoon explaining how the analysis was performed) are shown

Fig. 2. Phenotypes of plant material. Individuals shown in the grid are sorted in rows (1–9) and columns (a–h) corresponding to the entries in Supplementary Tables 1 and 2.

in Fig. 4. Our analysis showed that within the F_2 generation, all individuals displayed shifts of allele frequencies along most of their *A. thaliana* chromosomes (Fig. 4c) indicating frequent meiotic recombination between the interparental chromosomes already in the early generations. Even though tetraploid genomes need more generations for allelic fixation as compared to diploids, allelic fixation was common in the later generations, and

thereby outlines the possibility of stable introgression of one of the alleles into all 4 *A. thaliana*-derived chromosomes (Fig. 4).

As the 4 *A. thaliana* chromosomal copies are structurally identical this can lead to the formation of chiasmata with more than 1 homolog, creating multivalents, which may, or may not segregate the 4 chromosomes properly during meiosis (Bomblies *et al.* 2016) and can lead to aneuploidy. Comparisons of natural diploid
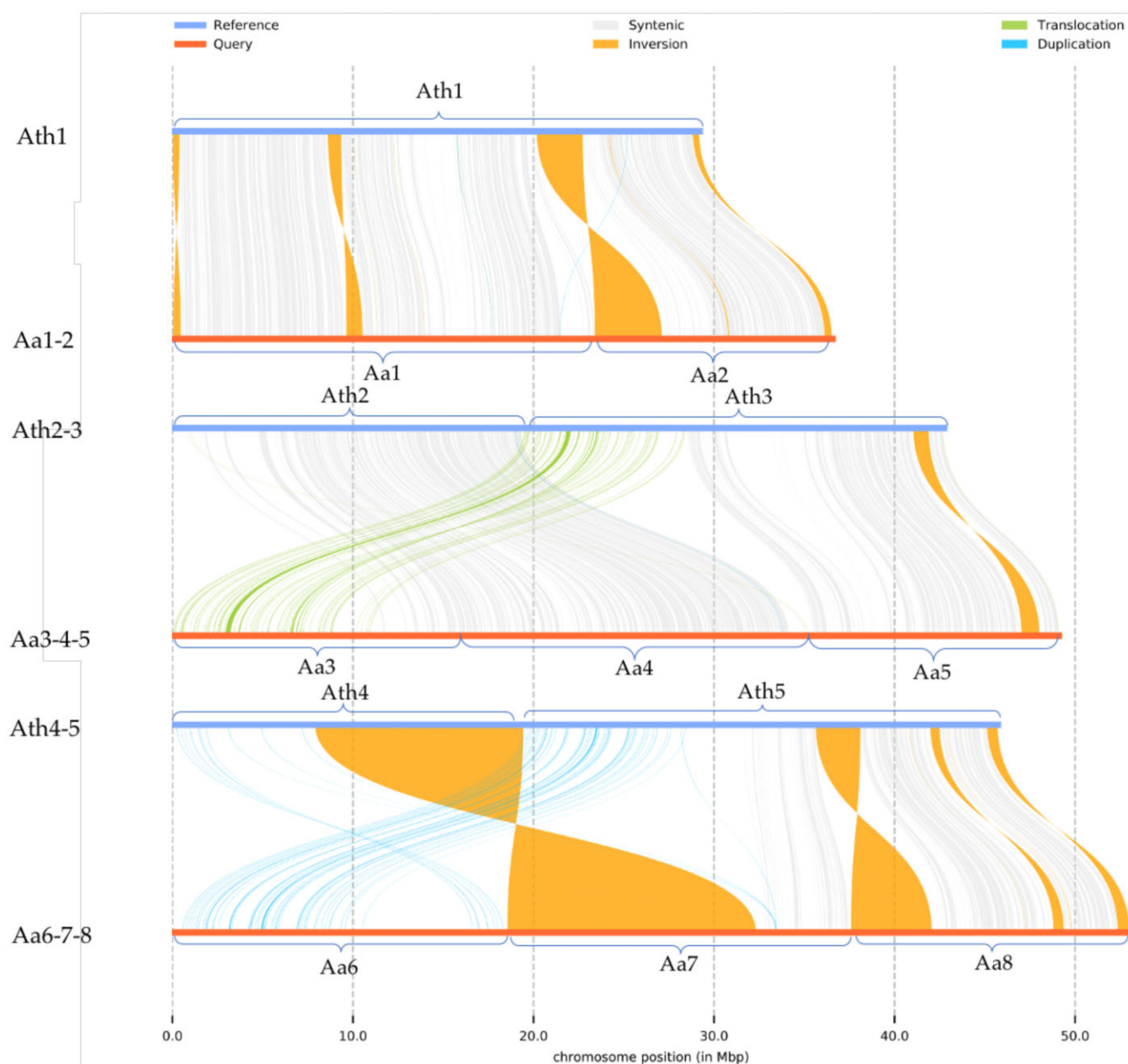
Fig. 3. The *A. suecica* genome shows a high degree of synteny within its subgenomes. The chromosomes of the *A. thaliana* (Ath) and *A. arenosa* (Aa) subgenomes of *A. suecica* were compared to each other. Translocations, inversions, and duplications between subgenomes are indicated. The syntenic blocks are similar to those previously reported between *A. thaliana* and *A. lyrata* (Hu *et al.* 2011). Blue lines represent the *A. thaliana* subgenome, orange lines the *A. arenosa* subgenome. The figure was generated with SyRI (Goel *et al.* 2019) and plotsr (Goel and Schneeberger 2022).

and autotetraploid populations of *A. arenosa* and *Cardamine amara* showed that chromosomes in some diploids are not inherently adapted to successfully undergo both diploid and tetraploid meiosis (Yant *et al.* 2013). Instead, a number of genes appeared to undergo selection during autopolyploidization, thereby stabilizing functions such as chromosome cohesion and segregation in the neo-autotetraploid (Hollister *et al.* 2012; Yant *et al.* 2013; Bohutínská *et al.* 2021).

Our data showed that in the hexaploid meiotic recombination occurred between the 2 *A. thaliana* genomes immediately after genome merger and continued for at least 17 generations (Fig. 4). Analogous to the observations in established autotetraploids, not restricting meiotic recombination across the 2 *A. thaliana* genomes might serve as one mechanism to allow for

introgressions between haplotypes and, in addition, the selection of alleles best suited for survival in the new genomic environment.

Aneuploidy accumulates in the *A. thaliana* subgenomes

Cytological analysis of early generation (F_2–F_7) individuals within the hexaploid pedigree was previously performed and showed that chromosome number in autoallohexaploid individuals varied from 27 to 39 (Matsushita *et al.*, 2012), while the expected euploid number in this population is 36 (Col 2x, $N = 10$, Sue-1 4x, $N = 26$). Most of the variation in these early generations occurred in the *A. thaliana* subgenomes (Matsushita *et al.* 2012). In the current study, we developed a read coverage-based algorithm to
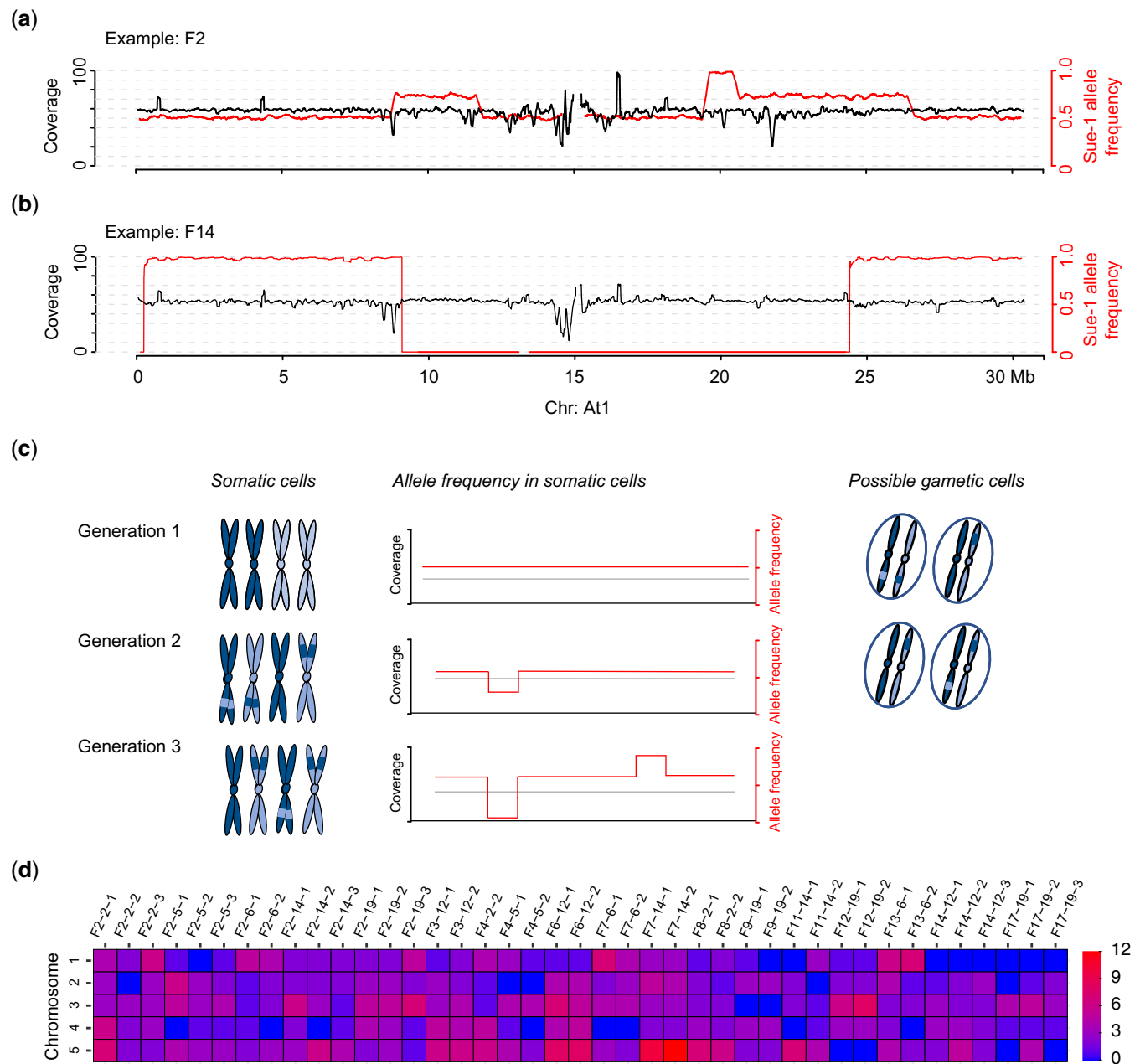
(a)

(b)

(c)

(d)

Fig. 4. Evidence for frequent meiotic recombination between the Col-0 and Sue-1 *A. thaliana* haplotypes in the autoallohexaploid. a) An example of the sliding window analysis estimating the local sequencing coverage (black) and Sue-1 allele frequencies (red). Here, the individual shows 5 major shifts in the allele frequencies on chromosome At1. The chromosome features 3 regions with 2 haplotypes of Col-0 and Sue-1 (beginning, middle, and end) as well as 2 regions where the genome consists of 3 haplotypes of Sue-1 and only one of Col-0 (~9–12 and ~21–27 Mb), as well as a short region (~19–21 Mb), where the genome fixed for the Sue-1 haplotypes. b) In this example, chromosome At1 shows fixation of the Sue-1 alleles at the beginning and at end of the chromosome (i.e. all 4 haplotypes are Sue-1 derived), and zero copies of the Col-0 haplotypes. In the middle of the chromosome, there are no Sue-1 alleles, thus indicating fixation of the Col-0 allele in that area. In both a) and b) the centromeric region (at around 15 Mb shows artifactual noise due to unreliable short-read alignments. c) Cartoon illustrating how homologous (meiotic) exchange can lead to fixed introgressions. Chromosomes on the left correspond to allele frequency plots in the middle and possible gametes arising from the respective generations on the right. d) Number of recombination events for each *A. thaliana* chromosome for each chromosome based on number of allele frequency shifts as exemplified in Fig. 4, a and b.

calculate the number of chromosomes in each of the sequenced samples (see *Methods* for details), which we used in lieu of conventional cell-based karyotyping. Using this algorithm, we analyzed a total of 210 *A. thaliana* chromosomes (42 lines times 5 *At* chromosomes) in the hexaploid individuals (Fig. 5, Supplementary Table 3). Of these, the copy number of 28 chromosomes (~13%) deviated from the expected number of 4 copies. Within this group, 20 chromosomes showed aneuploidy only in

one of the 2 or 3 sister individuals of the same generation. In the other 8 cases, multiple same-generation sisters were found to have similar copy numbers, suggesting possible karyotype destabilization through the earlier occurrence of aneuploidy. Out of the total 42 lines analyzed, only 20 individuals (~48%) had a euploid *A. thaliana* chromosome complement. Fourteen of the individuals (~33%) had a total chromosome count that was less than the expected 20 *A. thaliana* chromosomes, and 8 (~19%) had more
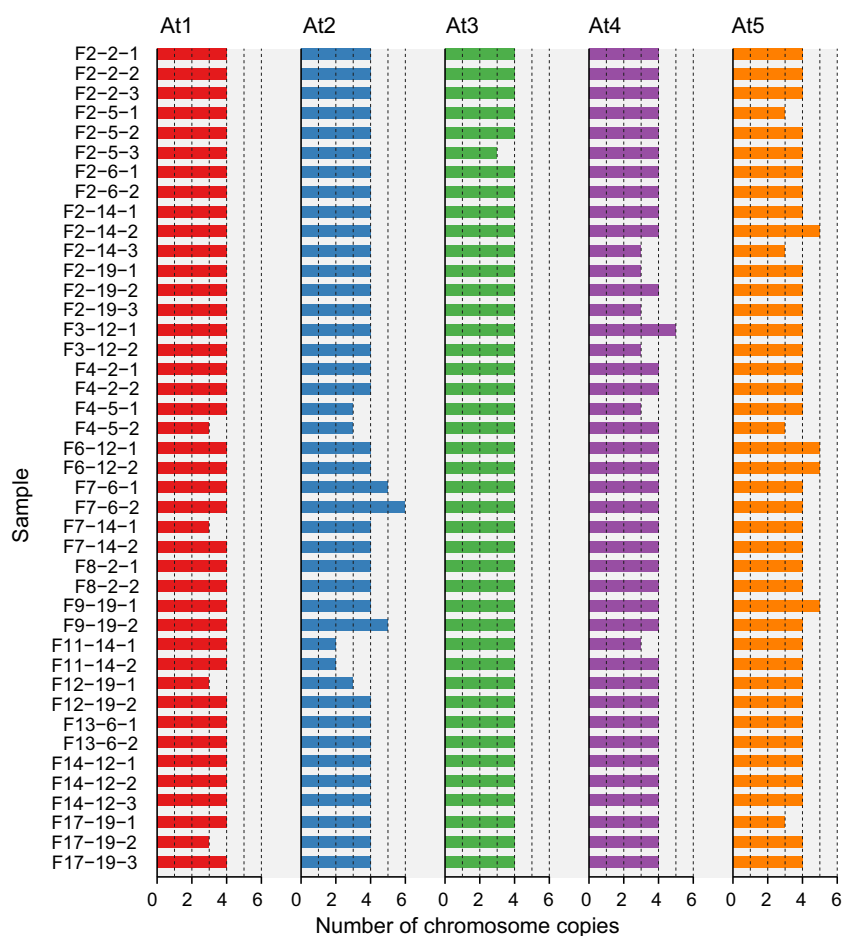
Fig. 5. Read count analysis predicts no obvious pattern of aneuploidy or diploidization in the allohexaploid pedigree. Sample numbers are ordered from earliest to latest generation (y-axis). The number of haplotype copies for each chromosome of the *A. thaliana* subgenome is shown on the x-axes. The euploid haplotype number of each chromosome of the *A. thaliana* subgenome is 4.

than 20 *A. thaliana* copies. To validate the copy number predictions from our count-based algorithm (Fig. 5, Supplementary Table 3), we again used the allele frequencies of the parental alleles along the chromosomes (Fig. 4). A frequency of 0.75, for example, indicates that the ratio of chromosome copies from the Sue-1 parent to those from the Col-0 parent is 3:1, suggesting a total number of 4 copies. An allele frequency of, for example, 0.6 or 0.8, on the other hand, would indicate a ratio of 3:2 or 4:1, respectively, and hence a total copy number of 5. In all cases, where validation was unambiguously possible (in 150 out of 210), the algorithm-based chromosome copy calculations were validated by the allele frequencies. An example of this validation is shown in Supplementary Fig. 2.

In some lines seemingly stable chromosome numbers in an earlier generation became unstable again in later generations (Fig. 5, Supplementary Table 3). For example, in line 12 in the F_6 generation both sisters (F_6 12-1 and F_6 12-2) appear to have 5 copies of chromosome 5 while all 3 sisters of that same line in the F_{14} generation show the expected count of 4 copies for chromosomes 5. Similarly, we observed only 3 copies of chromosome 4 in 2 sisters in the F_2 generation of line 19, yet in the F_{12} and F_{17} generations of this line, all individuals had the full complement of chromosomes 4. While this might at first sight seem surprising, it is important to point out that our pedigree analysis did not sample direct descendants in subsequent generations but those descending from a separate, unsampled sister of any sampled

individual. We therefore only see snapshots in time of the generation sampled. In some cases, genome fractionation (Wendel *et al.* 2018) appeared to occur in later generations. For example, in line 14 our analysis found only 2 copies of chromosome 2 in both sisters in the F_{11} generation (Fig. 5). Overall, the chromosome copy numbers did not appear to follow any obvious pattern or trend toward chromosome stabilization or diploidization.

Our analysis suggested that aneuploidy is frequent in all generations of this autoallohexaploid population and is reminiscent of findings in the resynthesized allotetraploids *B. napus* (Xiong *et al.* 2011), and *Tragopogon miscellus* (Chester *et al.* 2012). Similarly, in a population of resynthesized allohexaploid wheat, aneuploidy was rampant in the first 10 generations (Zhang *et al.* 2013). Even in resynthesized lines that were specifically bred by euploid selection and selfing to propagate the line to the 10th generation, these researchers found equally as many aneuploids in euploidy-selected lines as in lines that had descended from aneuploids (Zhang *et al.* 2013). In our study, as in those in *Brassica*, *Tragopogon*, and wheat, the analyzed samples do not directly reflect the karyotype of the gametes that the analyzed samples used to produce their offspring. It is thus possible that selection for euploidy occurs at the stage of zygote formation and that postzygotic accumulation of aneuploidy in somatic tissues or in unsuccessful gametes simply does not interfere with evolution to present a severe enough problem to be selected against on the individual level. In fact, the observation described above that later generations "regain"

chromosomes toward the full euploid complement, seems to suggest that aneuploid individuals sampled by us in earlier generations were unlikely to stably pass on a new karyotype to subsequent generations. Likewise, it is possible that by selecting fertile individuals during the construction of the pedigree we inadvertently selected for euploid individuals, but when we regrew plants from our seed stocks for sequencing leaf material when fertility was not required, we were more likely to grow aneuploid individuals. Thus, rather than introducing stable variation into the population as could be the case during recombination, aneuploidy might be responsible for selection against certain individuals, thus preventing any previous introgression events (e.g. by recombination) from remaining in the population. Given that aneuploids in this population were most likely to either produce (still unstable) aneuploid offspring or to become a dead-end in the gene pool, aneuploidy could be regarded as a counter-mechanism to stable introgression that essentially prevents individuals with excessive introgression or detrimental loss of whole chromosomes to take a foothold in the evolving population.

Diploidization is a commonly found process in which polyploids slowly return to a diploid state. During this process, repetitive DNA is eliminated via unequal and illegitimate recombination, and entire chromosomes can be lost due to aneuploidy (Soltis *et al.* 2015). In this context, it is of interest to note that we initially had also observed several lines within our pedigree that had reverted back to a karyotype resembling allotetraploidy much more so than autoallohexaploidy (Supplementary Fig. 3). Upon closer examination, these lines contained cryptic SNPs that did not stem from the parental genomes of Col-0 or Sue-1, but instead likely indicated cross-pollination with different natural *A. suecica* lines that were growing in the greenhouse at the same time as our advanced autoallohexaploids. We removed these plants from subsequent analyses in this study. It seems likely that the backcross of the hexaploids to natural tetraploid *A. suecica* produced pentaploid offspring, which then naturally decayed to tetraploids, however, we cannot exclude the possibility that the reverted lines had acquired favorable, stabilizing alleles from the natural *A. suecica* populations (Henry *et al.* 2014). But whatever the reason for the reversion to allotetraploidy might have been, the effect is a fixation of the introgressed Col-0 DNA into *A. suecica*, presenting a possible mechanism for the occurrence of new haplotype DNA in the parental species. Interestingly, some studies have reported natural chromosomal variation in *A. suecica* (Harmaja and Pellinen 1990), which potentially could indicate dysploidy resulting from naturally occurring hybridization events of *A. suecica* with *A. thaliana* and thereby could indicate that such introgressions are also occurring in nature. In this context, it may be of note that introgressions of foreign *A. thaliana* alleles could result in genome variation in *A. suecica* that would look like genomic patterns supporting multiple origins (despite resulting from introgressions; Novikova *et al.* 2017).

Interestingly, while genome evolution in autopolyploids is often the result of problems associated with multivalent formation (Lloyd and Bomblies 2016; Bohutínská *et al.* 2021), genomic change in allopolyploids is often caused by chromosomal translocations, transcriptome shock, and epigenetic changes (Hegarty *et al.* 2013; Wendel *et al.* 2016). This population of autoallohexaploids, although riddled with aneuploidy, appeared cytogenetically relatively stable. While over the 17 generations some sub-lines would die out occasionally (see for example individual F_5 19-1, B4 in Fig. 2) chromosome number fluctuations did not appear to commonly create a ratchet effect of successively more aneuploidy as was previously reported in resynthesized allopolyploid *B. napus*

(Gaeta and Pires 2010; Xiong *et al.* 2011). It is possible that the larger hexaploid auto- and allopolyploid genome constitution in our population exerted some sort of genomic buffering function maintaining enough fitness despite repeated chromosome gains or losses to allow for continued intergenomic exchange in subsequent generations before a possible collapse back to allotetraploidy.

Homoeologous exchanges occur between the *A. thaliana* and *A. arenosa* subgenomes

Given the large amount of genetic exchange within the autopolyploid subgenome and the wide stretches of syntenic regions between the *A. thaliana* and *A. arenosa* subgenomes in the autoallohexaploid (Fig. 3, Hu *et al.* 2011), we asked if, and to what degree, homoeologous exchange could occur. We hypothesized that homoeologous exchanges would be visible in coverage plots under the following conditions: If homoeologs pair, chromosome arms could be swapped between the *A. thaliana* and *A. arenosa* subgenomes and the intrasubgenomic recombinant chromosomes would then segregate according to their centromere identity like any other chromosome within the subgenome. In the next generation, the exchanged chromosome arm would then be subject to segregation. This could lead to an overall gain or loss of copies of the exchanged material, which subsequently could be observed as change in the coverage in both subgenomes. We therefore inspected the coverage plots of each individual in regions of high intrasubgenomic sequence homology. If read coverage changed on one chromosome in a highly homologous area of one subgenome then one would expect to see evidence of a reciprocal change also in the homologous area of the other subgenome (Fig. 6). Over time, segregation could then lead to some of the homoeologous exchanges to become fixed, and could lead to genomic loss or gain during the evolution of the population.

Overall, we observed a total of 19 individuals where such homoeologous exchanges occurred (Table 1). In several of these individuals more than one exchange, usually on different chromosomes, was seen. For example, in line 14, the genetic exchange was initially observed in the F_2 generation in only one of the sisters (Fig. 6). The pattern of variable copy numbers in this genomic region is maintained throughout the subsequent generations. In lines 6 and 14, coverage shifts of at least 1 copy were seen in the syntenic regions at the ends of At1 and Aa2, as well as at the beginning of At1 and Aa1. In line 6 (Supplementary Fig. 4), we observed a pattern of homoeologous exchanges at the ends of At1 and Aa2, where an increase of 1 copy is observed at the end of the *A. thaliana* subgenome for both tested sisters in generations F_2 and F_7 with a reciprocal loss of DNA in the corresponding region of the *A. arenosa* subgenome (Supplementary Fig. 4). In the F13 generation of this line, a copy number increase of 2 copies is seen on chromosome At1 while complete loss (i.e. a coverage of zero) of the reciprocal area on Aa2 suggests fixation of this exchange in that line as early as in the F_7. The coverage increase on At1 in one of the F13 sister lines suggested fixation of the genomic material from *A. arenosa* on the chromosome of *A. thaliana* (Supplementary Fig. 4).

Another example was seen in line 19 (Supplementary Fig. 5) where homoeologous exchange between At1 and Aa1 occurred at the top of the chromosomes in 2 regions. In this case, both sisters showed exchange in opposite directions which is reciprocated between the 2 subgenomes. The 3 sisters in the F_{17} of this line exemplified a case where each of the 3 individuals varies in copy number within this region of chromosome Aa1 (Supplementary Fig. 5). A similar example in the same line showed homoeologous exchange between chromosomes At5 and Aa8 (Supplementary Fig. 6). With the exception of line 19, the general genomic regions
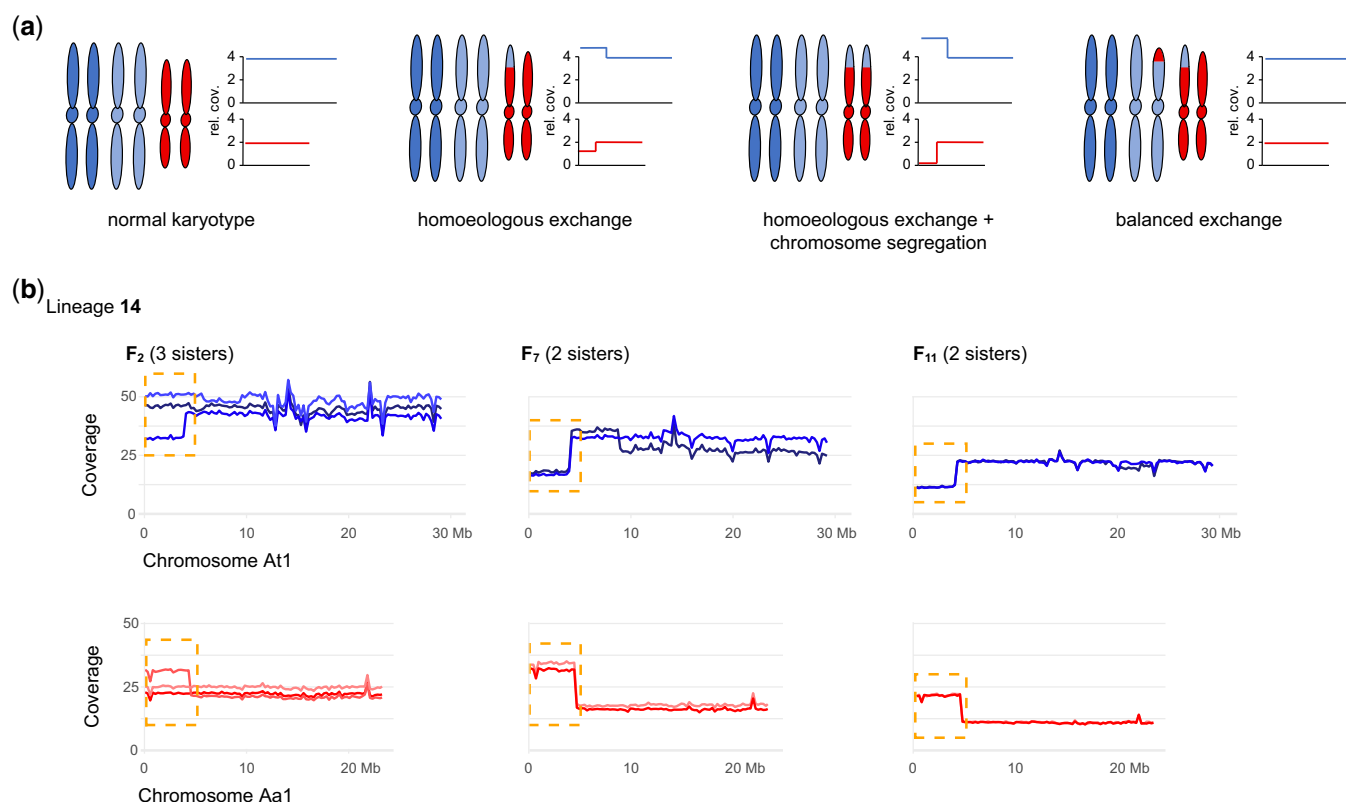
Fig. 6. Identification of homoeologous exchanges. a) Chromosomes from the *A. thaliana* (blue) and *A. arenosa* (red) subgenomes share syntenic homologous regions. The coverage plot next to the chromosomes shows the expected relative coverage for both chromosomes during short-read sequencing. Four possible groups of chromosomes (karyotypes) are shown illustrating (1) the normal karyotype, (2) chromosomes after a homoeologous exchange occurred, (3) the chromosomes of the next generation when a homoeologous exchange got fixed due to chromosome segregation, and (4) a reciprocal exchange, which—when balanced—cannot be found in the coverage plots of the short-read resequencing. b) Example of a homoeologous exchange between chromosomes At1 and Aa1 in line 14 and how it is propagated through generations. Sliding window coverage plots of generation 2 (3 sisters corresponding to the 3 lines in the graph), generation 7 (2 sisters), and generation 11 (2 sisters). In the F_2 generation, one of the 3 sisters shows a homoeologous exchange between the beginning of *A. thaliana* subgenome chromosome 1 (top) and the beginning of *A. arenosa* subgenome chromosome 1 (bottom, highlighted by the orange box). The same exchange (loss of region in At1 and gain of region in Aa1) is observed in all samples of all later generations (F_7 and F_{11}).

Table 1. Summary of homoeologous exchanges between subgenomes.

Line	Generation	Regions	Approximated breakpoint locations (Mb)
5	F_2 (1)	Start At5 (↓)/Start Aa6 (↑)	1.5/1.7
6	F_2 (2), F_7 (2), F_{13} (2)	End At1 (↑)/End Aa2 (↓)	27.0/10.8
14	F_2 (1), F_7 (2), F_{11} (2)	Start At1 (↓)/Start Aa1 (↑)	3.8/4.4
19	F_{12} (1), F_{17} (1)	Start At1 (↑)/Start Aa1 (↓)	4.4/5.0
	F_{12} (1)	Start At1 (↓)/Start Aa1 (↑)	1.9/2.3
	F_{17} (1)	Start At1 (↓)/Start Aa1 (↑)	4.4/5.0
	F_{17} (2)	End At5 (↓)/End Aa8 (↑)	24.9/13.5
	F_{17} (1)	End At5 (↑)/End Aa8 (↓)	24.9/13.5

Arrows indicate whether the coverage (i.e. copy number) increased or decreased in these regions. Fixation of the subgenome exchange between the end of At1 and the end of Ae7 can be observed in all individuals sequenced in line 6 from F_2 to F_{13}. Line 14 shows propagation of the subgenome exchange pattern found in one of the sisters of the F_2 generations to the later generations. In line 19, diverging patterns of coverage shifts are observed in the same regions between same-generation sisters in F_{12} and F_{17} generations.
Line: line number; Generation: generation within the line in which the homoeologous exchange was observed (parentheses indicate the number of individual sisters in the given generation where the exchanges occurred); Regions: regions where exchanges occurred; Breakpoint Locations: indicates the approximated genomic location where homoeologous exchanges occurred.

where the breakpoints were located are consistent between sisters and within each line. Although genetic exchanges were found in similar syntenic regions in both lines 14 and 19, the breakpoints are in slightly different locations (~500 kb apart).

Overall, our data suggested that homoeologous exchanges appeared at a comparatively high rate in the autoallohexaploid offspring (Table 1), in stark contrast to the established allotetraploid *A. suecica* Sue-1 where we only found one instance of a short homoeologous exchange at the end of chromosome 2. Burns and colleagues (2021) recently reported homoeologous exchanges in 3 of 15 analyzed *A. suecica* lines, in which a total of 2 different types of homoeologous exchanges were documented (Burns *et al.* 2021). Homoeologous exchange is thus much more frequent in autoallopolyploids than in established allotetraploids

although these types of exchanges apparently can either still occur in the established polyploid or can become fixed in natural populations.

In summary, our data showed how segregation can lead to the fixation of homoeologous exchanges in the autoallohexaploid, and suggests, in addition to intergenomic meiotic recombination discussed earlier, a second mechanism of how genetic material from one species can introgress into the genome of another species while still leading to viable offspring without any apparent effects on fitness.

Conclusions

As in several other genera, there is incomplete reproductive isolation between the species in the Arabidopsis genus, and pervasive gene flow exists between species in this genus (Schmickl and Koch 2011; Novikova *et al.* 2016). Polyploids have been implicated as one way to facilitate interspecific gene flow, specifically due to the reduced barriers to gene exchange in their compound genome, and for that reason have been referred to "allelic sponges" (Schmickl and Yant 2021) that can increase genetic diversity. Furthermore, hybridization of a reduced and an unreduced gamete via a triploid intermediate, also known as a triploid bridge, has been described as a frequent way to produce tetraploids (Ramsey and Schemske 1998; Schatlowski and Köhler 2012). As our analysis was based only a single cross, it is difficult to extrapolate relevance or frequency of this pathway in nature. However, our data did demonstrate that an intermediate autoallohexaploid arising via an F_1 triploid bridge can provide a mechanism for rapid gene flow between species. Such gene flow can occur using intergenomic legitimate meiotic homologous recombination as a tool for positive or neutral change. Aneuploidy, on the other hand, appears to act as a mechanism preventing evolutionarily stable gene flow and introgression. Homoeologous recombination, as a third mechanism in the response to interspecific genome merger, seems to be able to act both as a mechanism preventing gene flow when it results in instability, or as an agent of facilitating introgression if it occurs against the background of a big enough "genomic buffer" where loss or gain of chromosomal regions does not lead to detrimental change.

Diploidization and genome fractionation are major forces in the constant cycle of polyploidy and diploidy (Soltis *et al.* 2015). It is thus likely only a matter of time when this autoallohexaploid population will fully revert back to an allotetraploid or diploid state (Baduel *et al.* 2018). While it is unknown at this point if or how quickly this reversion might occur naturally, the structural differences and vast introgressions seen at this stage would likely remain as a reminder of its former intermediate polyploid state.

Data availability

The read data for both the allohexaploid samples as well as *A. suecica* Sue-1 were submitted to the National Center for Biotechnology Information (NCBI) under the BioProject accession number PRJNA753993, where also the assembly of Sue-1 can be found. The names for the hexaploid samples submitted begin with "Hex" followed by the generation, line number, and sister. For example, HexF2-2-1 refers to sister 1 in the F2 generation of line 2. The scripts for the chromosome count predictions are found on the GitHub website (https://github.com/schneeberger lab/Hexaploid_MA_lines).

Supplemental material is available at GENETICS online.

Conflicts of interest

None declared.

Author contributions

AM and KS conceived the original research experiment. AM created the pedigree and isolated the DNA. HT and AP provided the optical maps. VO, AM, and KS analyzed the data; VO, AM, and KS wrote the article with input from all coauthors.

Literature cited

Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. The "Polyploid Hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. Front Ecol Evol. 2018;6:117. doi: 10.3389/fevo.2018.00117.

Bohutínská M, Alston M, Monnahan P, Mandáková T, Bray S, Paajanen P, Kolář F, Yant L. Novelty and convergence in adaptation to whole genome duplication. Mol Biol Evol. 2021;38(9): 3910–3924. doi:10.1093/molbev/msab096.

Bomblies K, Jones G, Franklin C, Zickler D, Kleckner N. The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role? Chromosoma. 2016; 125(2):287–300. doi:10.1007/s00412-015–0571-4.

Burns R, Mandáková T, Gunis J, Soto-Jiménez LM, Liu C, Lysak MA, Novikova PY, Nordborg M. Gradual evolution of allopolyploidy in *Arabidopsis suecica*. Nat Ecol Evol. 2021;5(10):1367–1381. doi: 10.1038/s41559-021–01525-w.

Carlson KD, Fernandez-Pozo N, Bombarely A, Pisupati R, Mueller LA, Madlung A. Natural variation in stress response gene activity in the allopolyploid *Arabidopsis suecica*. BMC Genomics. 2017;18(1): 653. doi:10.1186/s12864-017–4067-x.

Chen Z. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol. 2007;58:377–406. doi:10.1146/annurev.arplant.58.032806.103835.

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proc Natl Acad Sci U S A. 2012; 109(4):1176–1181. doi:10.1073/pnas.1112041109.

Comai L. The advantages and disadvantages of being polyploid. Nat Rev Genet. 2005;6(11):836–846. doi:10.1038/nrg1711.

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. Phenotypic instability and rapid gene silencing in newly formed Arabidopsis allotetraploids. Plant Cell. 2000;12(9):1551–1568.

Fransz PF, Armstrong S, de Jong JH, Parnell LD, van Drunen C, Dean C, Zabel P, Bisseling T, Jones GH. Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. Cell. 2000;100(3): 367–376.

Gaeta RT, Pires JC. Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol. 2010;186(1):18–28. doi:10.1111/ j.1469–8137.2009.03089.x.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell. 2007;19(11):3403–3417.

Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics. 2022;38(10):2922–2926. doi:10.1093/bioinformatics/btac196.

Goel M, Sun H, Jiao W-B, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 2019;20(1):277. doi:10.1186/s13059-019–1911-0.

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G. Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature. 2006;439(7077): 749–752.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31(19):5654–5666. doi: 10.1093/nar/gkg770.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9(1):R7. doi:10.1186/gb-2008-9-1-r7.

Harmaja H, Pellinen K. Three different chromosome numbers from Finnish Arabidopsis suecica (Brassicaceae). Annales Botanici Fennici. 1990;27(4).

Hegarty M, Coate J, Sherman-Broyles S, Abbott R, Hiscock S, Doyle J. Lessons from natural and artificial polyploids in higher plants. Cytogenet Genome Res. 2013;140(2–4):204–225. doi:10.1159/ 000353361.

Henry IM, Dilkes BP, Tyagi A, Gao J, Christensen B, Comai L. The BOY NAMED SUE quantitative trait locus confers increased meiotic stability to an adapted natural allopolyploid of Arabidopsis. Plant Cell. 2014;26(1):181–194. doi:10.1105/tpc.113.120626.

Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. PLoS Genet. 2012;8(12):e1003093. doi: 10.1371/journal.pgen.1003093.

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, *et al.* The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet. 2011;43(5):476–481. doi:10.1038/ng.807.

Jenczewski E, Eber F, Grimaud A, Huet S, Lucas MO, Monod H, Chèvre AM. PrBn, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. Genetics. 2003;164(2):645–653.

Jiang X, Song Q, Ye W, Chen ZJ. Concerted genomic and epigenomic changes accompany stabilization of Arabidopsis allopolyploids. Nat Ecol Evol. 2021;5(10):1382–1393. doi:10.1038/s41559-021–01523-y.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008;24(24): 2938–2939. doi:10.1093/bioinformatics/btn564.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–915. doi:10.1038/s41587-019–0201-4.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long- read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5): 722–736. doi:10.1101/gr.215087.116.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):R12. doi:10.1186/gb-2004-5-2-r12.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359. doi:10.1038/nmeth.1923.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

Linder CR, Rieseberg LH. Reconstructing patterns of reticulate evolution in plants. Am J Bot. 2004;91:1700–1708. doi:10.3732/ajb.91.10.1700.

Lloyd A, Bomblies K. Meiosis in autopolyploid and allopolyploid Arabidopsis. Curr Opin Plant Biol. 2016;30:116–122. doi:10.1016/ j.pbi.2016.02.004.

Madlung A, Wendel JF. Genetic and epigenetic aspects of polyploid evolution in plants. Cytogenet Genome Res. 2013;140(2–4): 270–285. doi:10.1159/000351430.

Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20(16):2878–2879. doi:10.1093/bioinformatics/bth315.

Mallet J. Hybrid speciation. Nature. 2007;446(7133):279–283. doi: 10.1038/nature05706.

Matsushita SC, Tyagi AP, Thornton GM, Pires JC, Madlung A. Allopolyploidization lays the foundation for evolution of distinct populations: evidence from analysis of synthetic Arabidopsis allohexaploids. Genetics. 2012;191(2):535–547. doi:10.1534/genetics.112.139295.

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. Recently formed polyploid plants diversify at lower rates. Science. 2011;333(6047):1257. doi:10.1126/science.1207205.

Novikova PY, Tsuchimatsu T, Simon S, Nizhynska V, Voronin V, Burns R, Fedorenko OM, Holm S, Säll T, Prat E, *et al.* Genome Sequencing Reveals the Origin of the Allotetraploid Arabidopsis suecica. Mol Biol Evol. 2017;34(4):957–968.

Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, Guggisberg A, Paape T, Schmid K, Fedorenko OM, *et al.* Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nat Genet. 2016;48(9):1077–1082. doi:10.1038/ng. 3617.

Pearse IS, Krügel T, Baldwin IT. Innovation in anti-herbivore defense systems during neopolyploidy—the functional consequences of instantaneous speciation. Plant J. 2006;47(2):196–210. doi:10.1111/ j.1365-313X.2006.02776.x.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3): 290–295. doi:10.1038/nbt.3122.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842. doi: 10.1093/bioinformatics/btq033.

Ramsey J, Schemske DW. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst. 1998; 29(1):467–501.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–26. doi:10.1038/nbt.1754.

Schatlowski N, Köhler C. Tearing down barriers: understanding the molecular mechanisms of interploidy hybridizations. J Exp Bot. 2012;63(17):6059–6067. doi:10.1093/jxb/ers288.

Schmickl R, Koch MA. Arabidopsis hybrid speciation processes. Proc Natl Acad Sci U S A. 2011;108(34):14192–14197. doi:10.1073/pnas.1104212108.

Schmickl R, Yant L. Adaptive introgression: how polyploidy reshapes gene flow landscapes. New Phytol. 2021;230(2):457–461. doi:10.1111/nph.17204.

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021;37(12):1639–1643. doi:10.1093/bioinformatics/btaa1016.

Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31. doi:10.1186/1471–2105-6–31.

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). New Phytol. 2014;202(4):1105–1117. doi:10.1111/nph.12756.

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. Polyploidy and genome evolution in plants. Curr Opin Genet Dev. 2015;35:119–125. doi:10.1016/j.gde.2015.11.003.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34(Web Server issue):W435–W439.

Tate JA, Soltis DE, Soltis PS. Chapter 7—polyploidy in plants. In: The Evolution of the Genome, Gregory TR, editor. Burlington: Academic Press; 2005. p. 371–426.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963. doi:10.1371/journal.pone.0112963.

Wang J, Tian L, Lee HS, Chen ZJ. Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in Arabidopsis allopolyploids. Genetics. 2006;173(2):965–974.

Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. Genome Biol. 2016;17:37. doi:10.1186/s13059-016–0908-1.

Wendel JF, Lisch D, Hu G, Mason AS. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Curr Opin Genet Dev. 2018;49:1–7. doi:10.1016/j.gde.2018.01.004.

Xiong Z, Gaeta RT, Pires JC. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Natl Acad Sci U S A. 2011;108(19):7908–7913. doi:10.1073/pnas.1014138108.

Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, Franklin FCH, Bomblies K. Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. Curr Biol. 2013;23(21):2151–2156. doi:10.1016/j.cub.2013.08.059.

Zapata L, Ding J, Willing E-M, Hartwig B, Bezdan D, Jiao W-B, Patel V, Velikkakam James G, Koornneef M, Ossowski S, *et al.* Chromosome-level assembly of *Arabidopsis thaliana* Ler reveals the extent of translocation and inversion polymorphisms. Proc Natl Acad Sci U S A. 2016;113(28):E4052–E4060. doi:10.1073/pnas.1607532113.

Zhang H, Bian Y, Gou X, Zhu B, Xu C, Qi B, Li N, Rustgi S, Zhou H, Han F, *et al.* Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. Proc Natl Acad Sci U S A. 2013;110(9):3447–3452. doi:10.1073/pnas.1300153110.

Communicating editor: J. Birchler