Diploid origins and early genome stabilization in the allotetraploid *Arabidopsis suecica*

Robin Burns[1,*], Anna Glushkevich[2], Aboli Kulkarni[2], Uliana K Kolesnikova[2], Filip Kolář[3], Alison Dawn Scott[2], Polina Yu. Novikova[2,4,5,*]

[1]Department of Plant Sciences, University of Cambridge, Cambridge, UK
[2]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany
[3]Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic
[4]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium
[5]Flanders Institute for Biotechnology (VIB), VIB-UGent Center for Plant Systems Biology, Belgium

[*]co-corresponding authors: rb996@cam.ac.uk, pnovikova@mpipz.mpg.de

Summary

- **Polyploidization, followed by genome downsizing, is a recurrent evolutionary cycle that dramatically reshapes genome structure. Newly formed polyploids must quickly adjust their cell division machinery to maintain stable chromosome inheritance, while long-term stabilization involves rediploidization, returning the genome to a diploid-like state. Here, we investigate the origin and early genome evolution of *Arabidopsis suecica*, a hybrid polyploid derived from *A. thaliana* and *A. arenosa***
- **Leveraging a recent genome assembly for *A. suecica* along with population level whole-genome resequencing of all three species, we identify the closest progenitors to *A. suecica* and use this knowledge to examine genes under positive selection and to assess compensatory dynamics between homeologs carrying loss-of-function mutations.**
- **Our findings show that both parental species were diploid, including the paternal *A. arenosa* progenitor. We identify evidence for *de novo* adaptation to allopolyploidy within the *A. arenosa* sub-genome of *A. suecica*, including genes involved in homolog pairing and recombination. Although relaxed purifying selection is evident, likely due to the genome-wide redundancy, we observe functional compensation between homeologous gene pairs in *A. suecica*: when one copy loses function, the other maintains it**
- **Together, these findings revise the origin of *A. suecica* and identify early genome stabilization mechanisms, including evidence for meiotic adaptation and mutational buffering through homeologous gene compensation.**

Introduction

Cycles of polyploidy or recurrent whole-genome duplications (WGD) have profoundly shaped the genomes of most eukaryotes(Aköz & Nordborg, 2019; Hokamp et al., 2003; Van de Peer, 2023). In plants, allopolyploidy, WGD combined with interspecific hybridization, is both widespread(Soltis et al., 1993) and associated with important agronomic traits(Chen et al., 2020; Dubcovsky & Dvorak, 2007; A.-L. Li et al., 2015). Major crops including wheat, cotton, *Brassica*, strawberry, peanut, and tobacco trace their origins to allopolyploid events(Bertioli et al., 2019; Feldman & Levy, 2005; Liston et al., 2014; Renny-Byfield & Wendel, 2014; Sierro et al., 2024; Staudt, 2009; U, Nagaharu, 1935). Understanding the evolutionary success of allopolyploids requires resolving the mechanisms that stabilize their genomes, resolve meiotic incompatibilities, and manage genome-wide genetic redundancy.

The evolutionary fate of allopolyploids depends on the sequence of events that generate them. One key event is whether WGD occurs before or after hybridization because this can influence how readily a stable genome emerges. When WGD follows hybridization, as in *Primula kewensis(Newton & Pellew, 1929), Capsella bursa-pastoris(Bachmann et al., 2021)*, and experimental F_1 hybrids in *Mimulus(Wiiliam & Malcolm, 1971)* and *Arabidopsis(Nasrallah et al., 2000)*, the resulting neo-allotetraploid must adapt *de novo* to coordinate divergent genomes during basic cellular processes. In contrast, if WGD occurs before hybridization through the fusion of 2n gametes, autopolyploid parents can serve as a natural source, since their reduced gametes are equivalent to unreduced gametes from diploids. As a result, the resulting allopolyploid could inherit alleles already adapted to polyploidy, potentially easing genome stabilization. The extent to which inherited alleles are essential, or whether independent adaptations frequently arise in young allopolyploids, remains unclear. This question has important implications for both extinction risk and subsequent allopolyploid genome evolution.

Meiotic stabilization is one of the most immediate and consequential challenges facing newly formed allopolyploids. The formation of multivalents and improper pairing between homeologous chromosomes can disrupt segregation and compromise fertility(de Carvalho et al., 2021; Gonzalo, 2022; Katche, Schierholt, Becker, et al., 2023; Katche, Schierholt, Schiessl, et al., 2023). While divergence between sub-genomes can reduce these risks by limiting homeologous recombination(Ramsey & Schemske, 2002; Soltis et al., 1993; Stebbins, 1947), it does not eliminate it entirely(Gaeta & Chris Pires, 2010; Katche, Schierholt, Becker, et al., 2023; Katche, Schierholt, Schiessl, et al., 2023; Zhang et al., 2020). Experimental allopolyploids, including synthetic *A. suecica*, often exhibit meiotic defects despite a substantial sub-genomic divergence of 11.6%(Burns et al., 2021; Chéron et al., 2023; Nibau et al., 2024).

To achieve stable inheritance, polyploids evolve genetic mechanisms that modify crossover distribution or suppress homeologous recombination. In autopolyploids, this often involves polygenic adaptations in meiotic genes, including genes in the synaptonemal complex(Bohutínská et al., 2021; Bray et al., 2024; Hollister et al., 2012; Yant et al., 2013) and kinetochore(Bray et al., 2024), yet whether these same meiotic pathways are changed in allopolyploids remains an open question. Stable allopolyploids do appear to have evolved genetic control of sub-genome specific recombination. Notable examples include *Ph1* (*ZIP4*)(Martín et al., 2021) and *Ph2* (*MSH7-3D*)(Serra et al., 2021) loci in allohexaploid wheat, the *PrBn* locus in *Brassica napus(Liu et al., 2006)* and *BYS* in natural *A. suecica(Henry et al., 2014)*, all of which suppress homeologous recombination. In yeast hybrids, the mismatch

repair pathway has been shown to limit crossovers between divergent chromosomes, and loss of MSH2 can restore fertility in sterile hybrids(Greig et al., 2003). On a transcriptional level, meiotic-related genes were shown to be up-regulated on the *A. thaliana* sub-genome of *A. suecica*(Burns et al., 2021). However, it remains unclear whether meiotic stabilization in young allopolyploids arises primarily through changes in gene expression or protein-coding sequences, or both.

Polyploid genomes do not remain polyploid indefinitely. Ancient WGDs in the ancestry of most diploids suggest that current polyploids will eventually return to a diploid-like state through a process known as rediploidization(Van de Peer et al., 2017). This transition reshapes genome architecture via gene loss, regulatory divergence, and biased fractionation(Baduel et al., 2019; Conover & Wendel, 2022; Mandáková & Lysak, 2018; Monnahan et al., 2019; Redmond et al., 2023; Spoelhof et al., 2021), as seen in both ancient allopolyploids like maize(Sun et al., 2023), and in the diversification of distinct morphotypes in *Brassica*(Cheng et al., 2014). During rediploidization, relaxed purifying selection in redundant gene copies allows deleterious mutations to accumulate, while functional homeologs compensate for their non-functional counterparts, marking early stages of genome downsizing, as observed in *Brachypodium hybridum*(Scarlett et al., 2023). Functional compensation between homeologs may reflect a shift in selective constraint, from individual gene copies to gene pairs, suggesting that polyploid genomes are shaped not just by genetic redundancy, but by selection acting to preserve gene functional integrity. Several studies have shown that deleterious mutations, both point mutations and structural variants, accumulate more rapidly in polyploids than in diploids(Conover & Wendel, 2022; Hämälä et al., 2024; Paape et al., 2018), likely due to masking effects between redundant gene copies. The extent to which this type of mutational buffering between homeologs is shaped by selection, or if it is merely a neutral consequence of genetic redundancy, remains unresolved. Another uncertainty is whether patterns of early gene loss are predictable based on gene function or the sub-genome gene location. While genome dominance and biased fractionation have been reported in some allopolyploid species (Ma et al., 2024; Schnable et al., 2011; Alger & Edger, 2020), they are notably absent in several natural allopolyploids(Burns et al., 2021; Penin et al., 2024; Scarlett et al., 2023; Sun et al., 2023). Gene function has been shown to impact gene fractionation rates, with previous studies showing that genes related to genome integrity and organelle function revert to a single copy level faster than genes involved in signaling, transport, and metabolism, likely due to selective constraints acting on gene dosage levels(Z. Li et al., 2016).

The allotetraploid *A. suecica*, a post-glacial hybrid between *A. thaliana* and *A. arenosa*, is an excellent model for studying the early stages of allopolyploid genome evolution. However the origin of the WGD in *A. suecica*, whether it arose somatically in a diploid hybrid or was inherited via 2x gametes, remains unresolved. If the paternal *A. arenosa* was autotetraploid, WGD likely occurred through the fusion of its reduced gamete (n=2x) with an unreduced *A. thaliana* gamete (2n=2x), a parsimonious route(Jakobsson et al., 2006; Novikova et al., 2017). However, recent evidence from sequence similarity of meiotic genes point to a diploid *A. arenosa* parent(Nibau et al., 2022). A diploid *A. arenosa* parent would imply WGD likely occurred somatically in a diploid hybrid, and that *A. suecica* had to evolve *de novo* solutions to the challenges of polyploidy. Therefore distinguishing between diploid and tetraploid *A. arenosa* as the paternal progenitor is critical for understanding the origin of WGD in *A. suecica*, and for disentangling inherited versus novel mechanisms of meiotic stabilization.

With a high-quality long-read genome assembly of *A. suecica(Burns et al., 2021)* and expanded sequence data from both parental lineages, including the full diversity of *A. arenosa*

cytotypes(1001 Genomes Consortium, 2016; Bohutínská, Petříková, Booker, Cobo, et al., 2023b; Burns et al., 2021; Monnahan et al., 2019; Novikova et al., 2016) we address three central hypotheses about the origin and early genome evolution following allopolyploidy in *A. suecica*. First, we ask whether the *A. arenosa* parent of *A. suecica* was diploid or autotetraploid. The paternal *A. arenosa* identity allows us to infer whether WGD likely occurred somatically in a diploid hybrid or was inherited through the hybridization of 2x gametes. Establishing a diploid origin would support the hypothesis that *A. suecica* had to evolve *de novo* solutions to stabilize meiosis and manage genome-wide redundancy. Second, we test whether adaptation to allopolyploidy in *A. suecica* involves positive selection on genes that regulate meiosis and genome integrity. To do this, we identify genes under positive selection in the *A. suecica* genome and use protein-protein interaction networks to determine whether they represent important functional clusters. Finally, we ask how does *A. suecica* manage the mutational load associated with genome-wide genetic redundancy? By analyzing the distribution of inherited and *de novo* loss-of-function (LoF) mutations across sub-genomes, we test for a negative correlation between damaging variants in homeologous gene pairs. In other words we ask, when one copy (e.g., from *A. thaliana*) harbors a LoF mutation, does the other (e.g., from *A. arenosa*) remain intact? Evidence for homeolog compensation would suggest that selection operates at the level of gene pairs, potentially buffering young polyploid genomes against the mutational load that accompanies the polyploid-induced relaxed selective constraint.

Materials and Methods

DNA extraction of herbarium samples and sequencing

Herbarium samples were obtained from Moscow University Herbarium(Seregin, 2024). Extraction of DNA from herbarium samples of five *A. arenosa* and one *A. suecica* accession was carried out under a PCR workstation equipped with UV sterilization lights. The workstation, tubes and equipment were UV-sterilized before beginning the extraction. Herbarium samples were weighed and photographed. The material was then crushed using a sterile plastic pestle in a 2ml tube and then weighed again. PTB buffer was added and the material was kept at 37°C overnight on continuous rotation. We extracted the DNA following the protocol of(Gutaker et al., 2017). Libraries were sequenced in 50-bp paired-end mode on Illumina HiSeq 2000 Analyzers at the Vienna BioCenter Core Facilities (VBCF).

Read mapping and SNP calling

For *A. thaliana* accessions, 970 accessions of *A. thaliana* were used from the 1,001 genomes dataset(1001 Genomes Consortium, 2016), these accessions had a read length greater than 60bp and greater than 8X coverage. In addition to the herbarium samples aforementioned, we used 15 previously published *A. suecica* accessions(Novikova et al., 2017). For *A. arenosa* we used 158 previously published individuals(Bohutínská, Petříková, Booker, & Cobo, 2023; Monnahan et al., 2019) in addition to the 5 herbarium *A. arenosa* individuals sequenced here, giving a total of 163 *A. arenosa* individuals. The *A. arenosa* individuals used had a minimum of 5X coverage. A complete list of the *A. thaliana* and *A. suecica* accessions and the *A. arenosa* individuals used in this study is available as Supplementary File 1.

We mapped illumina paired-end short reads for the 16 *A. suecica* accessions, 970 *A. thaliana* accessions and 163 *A. arenosa* individuals to a chromosome-level genome assembly of *A. suecica*(Burns et al., 2021) using the BWA-MEM algorithm with an increased penalty for unpaired reads set to 15(H. Li & Durbin, 2009), version 0.7.17. The sub-genomes of the *A. suecica* accessions were treated separately after partitioning the mapped short reads. The *A. thaliana* accessions and *A. arenosa* individuals were mapped to the *A. thaliana* and *A. arenosa* sub-genome, respectively. PCR duplicates were removed using the rmdup function from Samtools(H. Li et al., 2009), version 1.10. SNPS were called using Genome Analysis Toolkit (GATK) (McKenna et al., 2010), version 3.8, using the HaplotypeCaller function. Generated gVCFs were combined into a single merged VCF using the combineGVCFs and GenotypeGVCFs functions.

Variant filtration

We removed multi-allelic sites and indels from our analysis and focused on bi-alleic SNPs. Bi-allelic SNPs were filtered to remove repetitive regions in the *A. suecica* sub-genomes (TEs sequences, centromere repeats, simple repeats and rDNA repeats). SNPs and non-variant sites were removed if the site was less than -0.5x log median coverage or greater than 0.5x log median coverage. The coverage thresholds were chosen to exclude sites with abnormally low or high coverage. Low coverage may indicate poor read mapping or insufficient sequencing depth, while high coverage can signal repetitive elements or cryptic duplications, or potential homeologous exchange events, all of which can lead to erroneous variant calls.

Heterozygous SNPs in *A. thaliana* have previously been shown to often result from segregating gene duplications rather than being residual heterozygosity from outcrossing(Jaegle et al., 2023). To avoid SNP-calling errors from segregating gene duplications in our VCF, we removed heterozygous SNPs that were present in *A. thaliana* and *A. suecica* at a population frequency greater than 5%, since both are highly selfing species. As *A. arenosa* is an obligate outcrossing species, to avoid segregating duplicates we removed heterozygous SNPs that were heterozygous in more than 90% of the individuals in the *A. arenosa* populations we examined. We allowed for 10% missing sites in the *A. arenosa* VCF. We allowed for 15% missing sites on the *A. thaliana* VCF.

The final VCFs contained 9,664,305 biallelic SNPs for the *A. arenosa* sub-genome, and 6,125,591 biallelic SNPs for the *A. thaliana* sub-genome.

Determining the ploidy level of *A. arenosa* individuals

The cytotype of the published *A. arenosa* individuals ploidy were previously confirmed using FACs(Bohutínská, Petříková, Booker, & Cobo, 2023; Monnahan et al., 2019). nQuire(Weiß et al., 2018) was run on raw fastq files of the 5 herbarium *A. arenosa* individuals, using the 'denoised' function and the 'lrdmodel'. Ploidy of the sample was determined by taking the ploidy that gave the maximum log-likelihood and lowest delta log-likelihood. Ploidy of the herbarium *A. arenosa* individuals was further confirmed by counting the number of different *S*-alleles for each individual with tetraploids having up to four and diploids up to two different *S*-alleles. The S-alleles were identified as described in(Kolesnikova et al., 2023) using short-read data and NGSGenotyp pipeline(Genete et al., 2020) with known *Arabidopsis* SRK sequences as a database.

Nucleotide divergence analysis

Pairwise nucleotide diversity (π) was calculated using custom scripts (see Code availability) in python. In order to carry out the network based on Nei's D, we read the VCF into R using the package vcfR(Knaus & Grünwald, 2017) and calculated Nei'D using StAMPP(Pembleton et al., 2013). To visualize the network we used the R package tanggle(Klaus Schliep, Marta Vidal-Garcia, Claudia Solis-Lemus, Leann Biancani, Eren Ada, L. Francisco Henao Diaz, Guangchuang Yu, 2021) and ggtree(Yu, 2020). Geographic maps were made using the ggspatial package in R (Dunnington, 2017)

Niche modeling of *A. suecica*

We modeled the niche of *A. suecica* using the current occurrence records and the selected Bioclim variables to project it onto the climate at LGM. We used workflow described in the ModleR package in R(Sánchez-Tapia et al., 2020) to the entire modeling (set up, evaluation and projection).

We obtained occurrence data of *A. suecica* from the GBIF database (gbif.org). There are a total of 8,451 records available of which we considered records only from preserved specimens having georeferencing data (1,772 records). This data was supplemented with the published records(Novikova et al., 2017). After cleaning the data for duplicates and occurrences in the same pixel 483 clean data points were obtained. Then, species thinning was carried out using the geo_filt function embdded in ModleR. Finally, 197 data points remained after applying the geographic filter of 0.5 km. These occurrence points were considered for further steps. We created 10000 background points using the setup_sdmdata function in the ModleR package.

We obtained 19 bioclim variables from Worldclim climate data (Fick and Hijmans, 2017) at a resolution of 30 arc seconds. We extracted the data for our occurrence points using the R package raster(Hijmans, 2018). Multicollinearity was checked for the bioclim variables in order to enhance the model consistency and remove the impact of highly correlated variables. Out of the 19 variables, eight uncorrelated variables (bio3, bio4, bio8, bio9, bio10, bio14, bio15) were then used for further analysis

We used maxent to fit our model, which uses the occurrence data and the corresponding climatic data to predict the probability of distribution of the target species. Twenty-five models were created by partitioning the data into five cross-validation partition types and five runs per partition. We evaluated the performance of these models using the true skill statistic (TSS), which is based on the confusion matrix(Allouche et al., 2006). The TSS scores span from 0 to 1, where 0 signifies a model performing no better than chance, and 1 denotes a flawless alignment with the observed data. We had set a TSS value greater than 0.7 to prepare the final model. Since all our models showed TSS values >0.9, we considered all 25 models.

The final model was projected onto the past climate (LGM). Further, occurrence points of *A. thaliana* and *A. arenosa* were overlaid on the projection to understand the origin and distribution of *A. suecica* environmentally and spatially.

Selection scans

To examine polymorphism patterns in genes we first ran SnpEff for annotation, and we used a modified version of the McDonald Kreitman (MK) test in order to find genes showing signatures of selection in *A. suecica.* We used a modified version of the MK test as *A. suecica* is evolutionarily young (~16 Kya), meaning if we relied on ratios of divergence to polymorphism

in the MK test we would often have division by zero. Therefore we focused instead on genes where nonsynonymous divergence (Dn) exceeded synonymous divergence (Ds), nonsynonymous polymorphisms (Pn) were fewer than synonymous polymorphisms (Ps) and where Pn was zero in *A. suecica.*

Gene ontology (GO) enrichment

We used the R package TopGO(Adrian Alexa, 2017) to perform our GO enrichment analysis, using the 'weight01' algorithm to account for the hierarchical structure in GO terms. We kept GO terms with a Fisher test p-value lower than 0.05. We used the gene annotations of *A. thaliana* one-to-one orthologs of *A. suecica* genes. Gene annotations for *A. thaliana* were obtained using the R package biomaRt from Ensembl 'biomaRt::useMart(biomart = 'plants_mart', dataset = 'athaliana_eg_gene', host = 'plants.ensembl.org').

SNP annotation, polarization and homeolog analysis

We annotated SNPs using SNPeff (version 5.2). We used 200 of the genetically closest *A. thaliana* accession to polarize annotated SNPs on the *A. arenosa* sub-genome. We used 29 individuals of the closest diploid *A. arenosa* to polarize annotated SNPs on the *A. thaliana* sub-genome. This resulted in 3,500,145 annotated and polarized SNPs for the *A. arenosa* sub-genome and 1,205,789 for the *A. thaliana* sub-genome. 11,812 and 11,983 SNPs, on the *A. arenosa* and *A. thaliana* sub-genome respectively, were annotated as predicted LoF polymorphisms that were later used in our homeolog analysis. Predicted LoF polymorphisms were selected by taking SNPs annotated by SNPeff with a high effect. We excluded SNPs with heterozygosity frequencies exceeding 1% in both *A. thaliana* and *A. suecica* to minimize the risk of bias in the LoF mutation analysis, which could be introduced by segregating gene duplicates leading to false signals of heterozygosity. We filtered for homeolog pairs where in *A. suecica* at least one homeolog in the pair carried a LoF mutation, amounting to 414 genes. In instances where a homeolog carried more than one LoF mutation, we selected the LoF mutation that was at a higher population frequency.

Permutation analysis

To assess the negative correlation of LoF mutations between the *A. thaliana* and *A. arenosa* sub-genomes in *A. suecica*, we conducted two complementary permutation analyses. First, we compared the observed number of homeologous gene pairs with LoF mutations in both sub-genomes to expectations from random pairings. Specifically, we focused on 414 homeolog pairs in which at least one gene copy carried a LoF polymorphism in *A. suecica*. These 414 homeolog pairs represent the subset of the genome where LoF variation exists and where compensatory dynamics can be meaningfully assessed. We did not include all homeologs in the permutation, as the rarity of LoF mutations would heavily skew the null distribution toward zero co-occurrence, limiting the sensitivity of the test. From this filtered set, we generated 1,000 permutations by randomly re-pairing homeologs, and compared the observed number of dual LoF pairs to the resulting distribution.

Secondly, we performed a permutation analysis using simulated *in-silico A. suecica* genomes. We generated 1,000 in-silico *A. suecica* by randomly combining genomes from 200 *A. thaliana*

304 and 29 diploid *A. arenosa* individuals that are genetically closest to natural *A. suecica*,
305 sampling with replacement to account for all possible combinations. For each simulated
306 genome, we assessed dual LoF occurrence in the same set of 414 homeolog pairs as in the
307 natural *A. suecica* dataset. The observed number of homeolog pairs with LoF mutations in
308 both sub-genomes was then compared to the distribution derived from the *in-silico* genomes,
309 providing an additional null expectation under neutral inheritance.

310 For both tests, statistical significance was determined by comparing the observed number of
311 dual LoF pairs in natural *A. suecica* to the empirical null distribution generated by the
312 permutation.

313 Purifying selection analysis

314 To evaluate evidence of purifying selection acting on intact homeologs relative to their LoF
315 carrying counterpart, we quantified pairwise nucleotide diversity (π) using coding sequences
316 from *A. suecica*. We further calculated the ratio of non-synonymous (Pn) to synonymous (Ps)
317 polymorphisms to assess selective constraint. Given the severe population bottleneck in *A.*
318 *suecica*, a pseudocount of 0.1 was added to both Pn and Ps to avoid division by zero.
319 Statistical significance was determined using a one-sided paired t-test, testing the hypothesis
320 that intact homeologs have reduced π and Pn/Ps compared to their LoF pairs.
321

322 Analysis of cell cycle genes in the *A. suecica A. arenosa* sub-genome

323 Biallelic SNPs from CDS regions of the genes were used to generate SNP-frequency
324 heatmaps. *A.arenosa* assemblies from(Bohutínská, Petříková, Booker, Cobo, et al., 2023a)
325 were scaffolded with RagTag(Alonge et al., 2022) on the *A.lyrata* NT1 genome
326 assembly(Kolesnikova et al., 2023) and annotated with Helixer(Stiehler et al., 2021). *A.*
327 *thaliana* protein sequences of target genes were used for protein BLAST(Camacho et al.,
328 2009) against annotated proteins extracted with AGAT(Dainat et al., 2024). Genes identified
329 by BLAST were aligned with L-INS-i algorithm of MAFFT(Katoh et al., 2005) and trees were
330 constructed with IQTree(Hoang et al., 2018; Nguyen et al., 2015) with fast bootstrap option
331 equals 1000. Protein alignments were visualized using snipit(O'Toole et al., 2024)
332

333 Single and multi copy gene analysis

334 Single copy and multi copy genes were obtained from(Z. Li et al., 2016) and hypergeometric
335 tests were performed to test for enrichment.
336

337 Expression analysis

338 To test whether transcription is biased toward the intact member of a homeologous pair, we
339 mined $\log_2$-fold-change ($\log_2$FC) estimates for rosette tissue from(Burns et al., 2021) which
340 include 15 natural *A. suecica* accessions and two synthetic lines. We first averaged $\log_2$FC
341 values within each group (natural vs. synthetic). We then restricted the analysis to homeolog
342 pairs in which the LoF allele segregates at high frequency in natural *A. suecica* (≥ 60 %). This

filter retained 64 intact homeologs, 16 on the *A. arenosa*-derived sub-genome and 48 on the *A. thaliana*-derived sub-genome, out of the 414 total pairs.

STRING Network Analysis

To investigate protein-protein association networks among genes identified in our selection scan of the *A. arenosa* sub-genome of *A. suecica*, we used the STRING database(Szklarczyk et al., 2023). To identify gene groups with shared function and protein-protein associations, we applied Markov Cluster Algorithm (MCL) clustering, using an inflation parameter of 2, resulting in 102 discrete clusters. Four of the 102 clusters were investigated further due to their association with important polyploid traits.

Results

Genetic origins of *A. suecica*

To reconstruct the evolutionary history of allotetraploid *A. suecica*, we first aimed to identify the closest parental accessions, using available data from the parental species, *A. thaliana* and *A. arenosa*, covering their vast geographic range. We mapped 970 *A. thaliana* accessions(1001 Genomes Consortium, 2016) and 163 *A. arenosa* individuals (93 tetraploids and 70 diploids), representing spatially close lineages of both diploid and autotetraploid cytotypes(Bohutínská et al., 2024; Monnahan et al., 2019; Novikova et al., 2016), to the *A. thaliana* and *A. arenosa* sub-genomes of *A. suecica*(Burns et al., 2021). Genomic positions were conservatively filtered to ensure reliable variant calling (see Materials and Methods). In summary, callable sites included ~62% (73,813,230) of the *A. thaliana* sub-genome and ~34% (47,950,105) of the *A. arenosa* sub-genome, with 6,125,591 biallelic single nucleotide polymorphisms (SNPs) in *A. thaliana* and 9,664,305 in *A. arenosa.* Throughout this study, results generated for the sub-genomes of *A. suecica* are compared with corresponding data from the parental species*, A. thaliana* and *A. arenosa.*

Previously, it was shown that the closest *A. thaliana* accessions to *A. suecica* are found in Central Eurasia(Novikova et al., 2017). Here, eliminating the possibility of a reference bias artifact, we calculated median pairwise diversity (π) for each *A. thaliana* accession to our 16 *A. suecica* natural accessions and confirmed that result (Figure 1A). To identify the *A. arenosa* individuals most closely related to *A. suecica* and to determine their cytotype (diploid or tetraploid), we expanded the previously released *A. arenosa* dataset(Bohutínská et al., 2024; Monnahan et al., 2019). We sequenced five additional autotetraploid *A. arenosa* individuals sampled from herbarium specimens collected in the southern Urals, representing the easternmost populations and those geographically closest to the parental *A. thaliana* accessions. The southern Ural *A. arenosa* individuals were identified as tetraploids, using the software nQuire(Weiß et al., 2018) and counting the number of different *S*-alleles (see Materials and Methods, similar to ploidy determination in *A. lyrata*[66]), while the ploidy of most other *A. arenosa* samples was previously determined using flow cytometry(Bohutínská et al., 2024; Monnahan et al., 2019).

We conducted a network analysis based on genetic distances (Nei's D(Nei, 1972)) that are suitable for mixed-ploidy datasets(Pembleton et al., 2013) (Figure 1B). We also calculated the median pairwise nucleotide diversity (π) for each *A. arenosa* individual against our 16 *A.*

suecica natural accessions and placed it on a geographical map (Figure 1C). Our results indicate that diploid *A. arenosa* from southern Baltic sea coast(Kolář et al., 2016) is genetically the closest to *A. suecica*, and certainly closer than any autotetraploid *A. arenosa*, including both the current sympatric tetraploid Ruderal lineage growing in Fennoscandia(Monnahan et al., 2019) and tetraploid populations from the southern Urals, i.e. the geographic region closest to putative ancestral *A. thaliana* lineage (Figure 1A). A full list of *A. thaliana* and *A. arenosa* accessions and their π values to *A. suecica* is available in Supplementary File 1. The vast majority of the SNPs (>99%) in *A. suecica* are shared with *A. arenosa* and *A. thaliana,* indicating that polymorphisms in *A. suecica* was largely inherited from standing variation in multiple, genetically distinct parental individuals, rather than arising from *de novo* mutation, in agreement with previous reports(Novikova et al., 2017) (Supplementary Figure 1).

Together our results suggest a diploid rather than tetraploid origin for the *A. arenosa* sub-genome within *A. suecica*. This is in agreement with data showing that meiotic genes on the *A. arenosa* sub-genome of *A. suecic*a carry sequence polymorphisms characteristic of diploid *A. arenosa,* and not the polymorphisms found in autotetraploid *A. arenosa* (Nibau et al., 2022). The diploid origin supports the original hypothesis by Hylander in 1957(Hylander, 1957) that *A. suecica* originated from hybridization between diploid *A. thaliana* and diploid *A. arenosa*, with autotetraploid *A. arenosa* later spreading across Fennoscandia, while the diploids remain restricted to southeastern Europe and isolated Baltic areas(Kolář et al., 2016). The lack of geographical overlap between the current distribution of the closest *A. thaliana* and *A. arenosa* progenitors raises the question of where the hybridisation could have occurred? To resolve this question we performed ecological niche modelling in the context of the estimated time of origin of *A. suecica*, that is, towards the end of the Last Glacial Maximum(Novikova et al., 2017), and we suggest that hybridization between *A. thaliana* and *A. arenosa* likely happened close to the border of the Fennoscandia ice sheet (see Supplementary Note 1, Supplementary Figure 7).

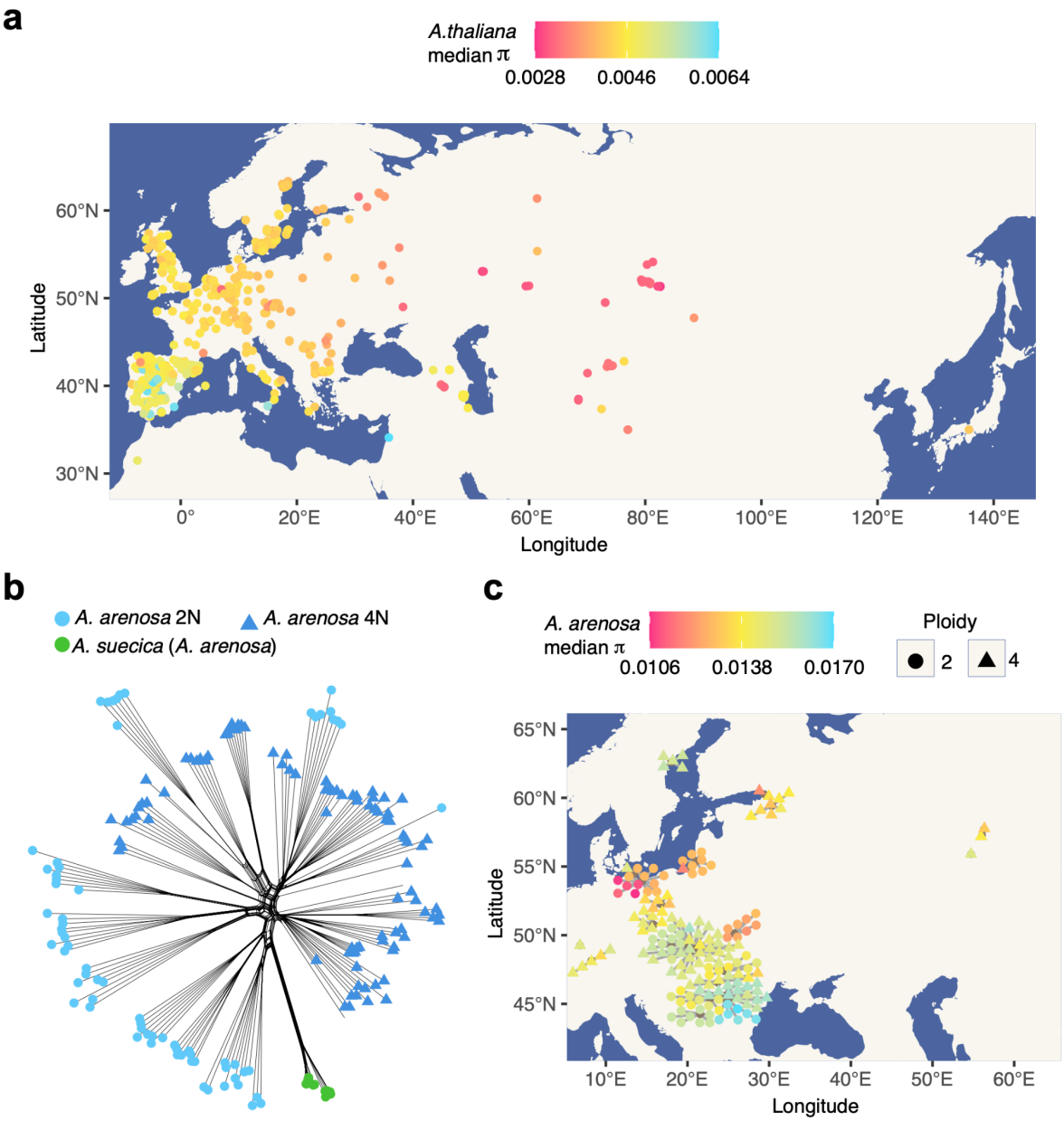


Figure 1: The genetic and geographic origin of *A. suecica*. a Median nucleotide diversity (π) of 970 *A. thaliana* accessions to 16 *A. suecica* accessions shows Central Eurasian *A. thaliana* are genetically closest **b** A genetic distance network (Nei's D) shows that *A. suecica* (green circles) is genetically closer to diploid *A. arenosa* (light blue circles) than autotetraploid *A. arenosa* (dark blue triangles). **c** Median nucleotide diversity (π) of *A. arenosa* individuals to 16 *A. suecica* accessions indicates that Southern Baltic *A. arenosa* diploids (red circles) are the genetically closest. Locations are represented by grey lines connecting *A. arenosa* individuals to their geographic origins.

Genomic adaptation to allopolyploidy in *A. suecica*

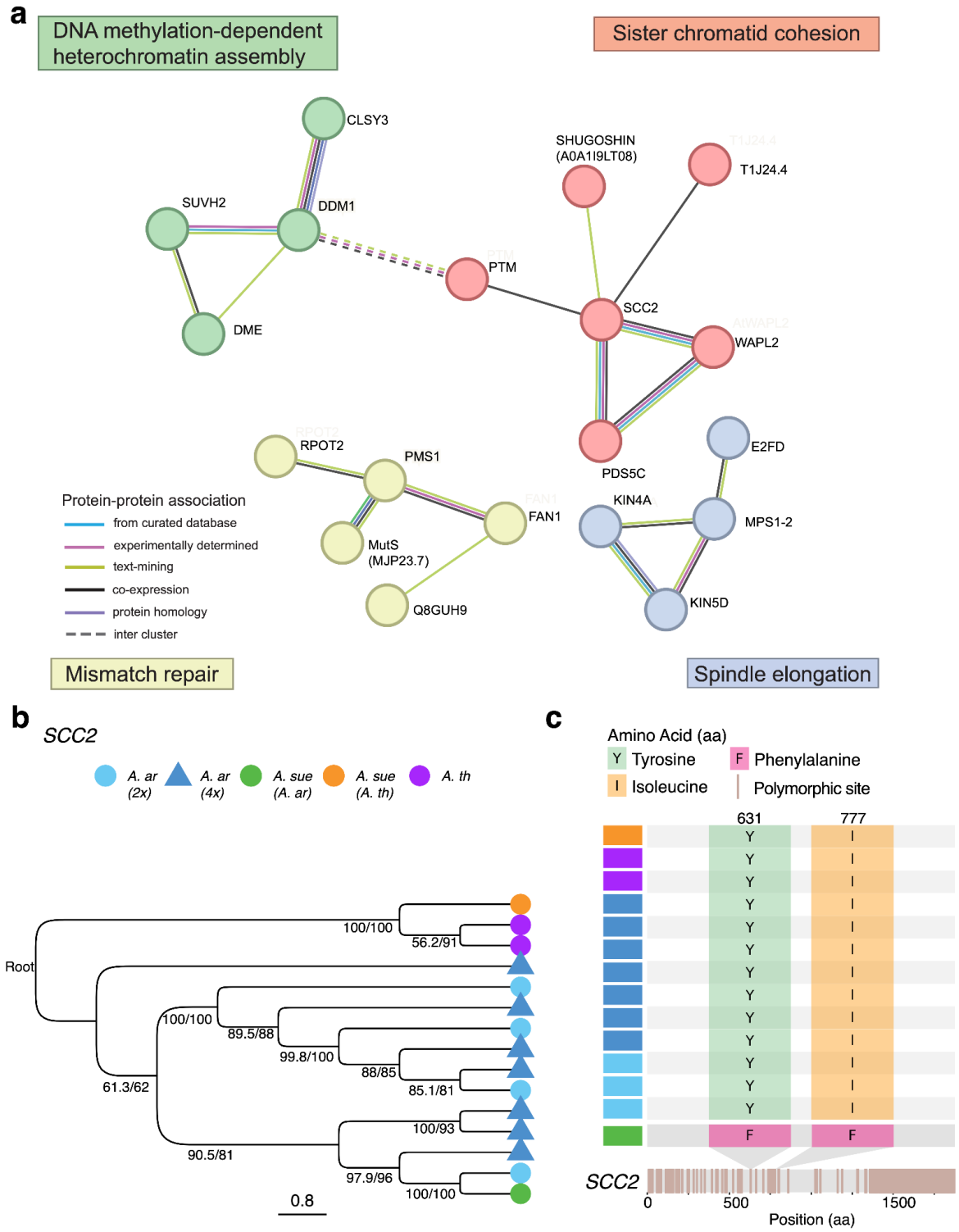
Given that the genetically closest parents of *A. suecica* were diploids, *A. suecica* could not inherit any of the potentially pre-adaptive meiotic alleles from autotetraploid *A. arenosa*. Previously, we reported that meiosis-related genes on the *A. thaliana* sub-genome of *A. suecica* are up-regulated compared to parental transcriptional levels(Burns et al., 2021). However, in the same study, we did not observe significant upregulation of meiosis-related genes on the *A. arenosa* sub-genome. Notably, the latter transcriptional comparison was made against tetraploid *A. arenosa*, the previously hypothesized parent of *A. suecica*, whereas we now establish here that diploid *A. arenosa* is closer. Therefore, to uncover how *A. suecica* adapted to allopolyploidy from diploid ancestry, we searched for signatures of selection on genes, performing McDonald-Kreitman test between *A. suecica* and the identified closest progenitors (see Methods). While this approach highlights strong candidates for selection on coding sequences, we note that it will miss selection acting on regulatory elements. Homeologous recombination events could also inflate divergence estimates at some loci; however, we applied coverage-based filters during SNP calling to avoid such events (see Materials and Methods). Additionally, HE events are rare in natural *A. suecica*, with only two homoeologous exchange junctions reported in (Burns et al., 2021), and nine in (Zhang et al., 2020), in a population of *A. suecica*. HE events are also distally biased towards subtelomeric regions(Nibau et al., 2024; Zhang et al., 2020). In contrast, our selection-scan genes do not show such a bias (see Supplemental Figure 2), indicating that HEs are unlikely to drive our selection-scan results. Additionally, gene conversion rates in *A. thaliana* are similarly low, estimated at ~3.6×10^{-6} per site per meiosis (Wijnker et al., 2013).

Using the MK test, we identified signatures of selection on 35 genes in *A. thaliana* and 585 genes in the *A. arenosa* sub-genomes of *A. suecica* (Supplementary File 2). Due to the low number of founders contributing to the origin of *A. suecica*(Novikova et al., 2017), regions of the genome are depleted for genetic diversity, however, we find that many of the genes under selection are located outside of these low-diversity regions (Supplementary Figure 2), and show GO enrichment, supporting their biological relevance. Enriched categories include meiotic cell cycle (GO:0061780) and mitotic cohesin loading (GO:0051321); however, some are driven by individual genes (Supplementary Figure 3, Supplementary File 2), for example, SCC2 alone drives the mitotic cohesin loading term and so this single-gene driven GO terms limit a broader interpretation. We note that genes under selection on the *A. arenosa* sub-genome do not overlap with the previously described BYS QTL locus for meiotic stability in *A. suecica*(Henry et al., 2014).

To better explore how positively selected genes might functionally interact, we used the STRING database(Szklarczyk et al., 2023) to construct a protein-protein association network (Figure 2A; see Methods). Among the 585 genes under selection in the *A. arenosa* sub-genome of *A. suecica*, 19 formed a connected subnetwork in STRING. These genes clustered into four distinct functional clusters identified using the Markov Cluster Algorithm (MCL), corresponding to sister chromatid cohesion, DNA methylation-dependent heterochromatin assembly, mismatch repair, and spindle elongation. The sister chromatid cohesion cluster contains 6 interacting proteins, including *SCC2*, *WAPL2*, and *SHUGOSHIN*. The DNA methylation cluster includes 4 genes, among them *SUVH2* and *DDM1*. The mismatch repair cluster (5 genes) contains *MutS* and *PMS1* and the spindle elongation cluster (4 genes) includes *MPS1-2.* The four functional clusters are supported by multiple protein-protein association evidence types, including experimentally determined, co-expression, protein homology, and curated databases (Figure 2A legend). In contrast, the 35 selected

467 genes from the *A. thaliana* sub-genome did not form a connected network or enriched
468 functional clusters.
469 Among the four clusters, the sister chromatid cohesion cluster stood out because of
470 the importance of cohesion and the cohesin complex in meiosis(H. Wang et al.,
471 2020)(Storchová et al., 2006). *SCC2* (SISTER CHROMATID COHESION2), which encodes
472 the cohesin loader adherin, a protein essential for cohesin loading onto chromosomes(H.
473 Wang et al., 2020), emerged as a particularly compelling candidate because *SCC2* has also
474 previously been implicated in adaptation in autotetraploid *Arabidopsis arenosa*(Yant et al.,
475 2013). To investigate whether SCC2 represents the same haplotype as that of tetraploid *A.*
476 *arenosa* or instead represents *de novo* evolution within *A. suecica* following
477 allopolyploidization, we performed a phylogenetic analysis which showed that *A. suecica*
478 clusters consistently with diploid *A. arenosa*(Figure 2B), and does not show evidence of
479 haplotype sharing with autotetraploid *A. arenosa*. Furthermore, *A. suecica* has private
480 polymorphisms on SCC2 (*A. arenosa* sub-genome), which are not present in either diploid or
481 tetraploid *A. arenosa* or *A. thaliana* (Figure 2C), suggesting *de novo* evolution.

482



483
484
485
486
487
488
489
490

Figure 2: Genomic adaptation to allopolyploidy in *A. suecica* a STRING protein-protein association network showing four functional clusters among 19 genes under positive selection on the *A. arenosa* sub-genome of *A. suecica*, as identified by McDonald–Kreitman tests. Colored clusters represent functional clusters including DNA methylation-dependent heterochromatin assembly (green), sister chromatid cohesion (red), mismatch repair (yellow), and spindle elongation (blue). Solid lines represent intra-cluster associations; dotted lines indicate inter-cluster associations **b** Phylogenetic tree from protein alignment of SCC2 showing that the *A. arenosa* homeolog in *A. suecica* clusters with diploid *A. arenosa*. Branch length is not drawn to scale to improve the analysis of the tree topology. Both the translated genomic and CDS sequence of the *SCC2* from *A. thaliana,* in addition to the *A. suecica A. thaliana* homeolog were used to root the tree. **b** Fixed amino acid differences in the *SCC2* between the *A. suecica A. arenosa* sub-genome and diploid *A. arenosa*. Tetraploid *A. arenosa* and *A. thaliana* share the same amino acid as diploid *A. arenosa*, the difference is unique to *A. suecica.*

Preservation of gene function via homeolog compensation in *A. suecica*

Polyploid organisms have a substantial amount of genetic redundancy, which can lead to relaxed purifying selection. This pattern of relaxed selection has been observed in autopolyploid *A. arenosa*(Baduel et al., 2019)(Vlček et al., 2025) and allopolyploid *A. kamchatica*(Paape et al., 2018). We investigated whether *A. suecica*, the youngest polyploid in the *Arabidopsis* genus, shows a similar signal of relaxed purifying selection and asked how patterns of LoF mutations are distributed across its sub-genomes.

Analyzing over 3.5 million SNPs in the *A. arenosa* sub-genome and 1.2 million SNPs in the *A. thaliana* sub-genome, we found that both sub-genomes in *A. suecica* are evolving under purifying selection, however purifying selection appears relaxed in *A. suecica* compared to the parental species (Figure 3A). This genome-wide signal suggests a reduced selective constraint following polyploidization, consistent with the buffering effects of homeolog redundancy.

A reduced selective constraint following polyploidization enables mutational robustness between the sub-genomes, since a LoF mutation in one sub-genome can be compensated by its homeolog on the alternate sub-genome. To investigate the compensatory relationship between homeologs we identified a total 414 homeologous gene pairs where at least one copy carries a segregating LoF mutation in *A. suecica*. To understand the nature of LoF mutations between homeologs we split them into inherited (standing variation) and *de novo* (occurring post-allopolyploidization) LoF mutations in *A. suecica*. While a slight excess of LoFs were found in *A. arenosa*-derived homeologs (269 vs. 142), the majority of these mutations are also present in the parental genomes and thus appear to be inherited rather than arising *de novo* (Figure 3B). Focusing only on *de novo* LoF mutations unique to *A. suecica*, the distribution appears more balanced (49 in *A. arenosa*, 62 in *A. thaliana*). To investigate if these LoF homeologs overlap with genes identified in our selection scans for genomic adaptation to allopolyploidy in *A. suecic*a (see previous section), we compared the genes involved. However, we find minimal overlap between the gene lists (Supplementary Figure 4), indicating that a LoF occurring on one homeolog is not driving the signals of positive selection we observed in the previous section.

We next asked if LoF mutations (both inherited and *de novo)* observed in 414 homeologous gene pairs are distributed randomly or whether having a LoF mutation in one homeolog would interfere with occurrence of LoF in the other homeolog. The latter would be explained by purifying selection acting to preserve at least one functional copy in *A. suecica*.

To test this, we compared the observed joint frequency spectrum of LoF mutations between the *A. suecica* sub-genomes to (1) spectra resulting from reshuffling of homeolog pairings and (2) to spectra of in-silico synthetic *A. suecica* genomes generated by combining natural diploid *A. arenosa* and *A. thaliana* individuals (Figure 3D). In both comparisons, we found that dual LoF mutations are significantly less frequent in natural *A. suecica* (Figure 3C–D; $p < 0.05$). The depletion of dual LoFs supports a model of compensatory homeolog retention, in which functional redundancy between homeologs allows relaxed selection at the level of individual genes, while purifying selection operates at the pair level to prevent the complete loss of gene function. To further test for evidence of purifying selection acting on the remaining intact homeolog, we compared pairwise nucleotide diversity (π) and the ratio of non-synonymous to synonymous polymorphisms (Pn/Ps) between our 414 homeolog pairs. We found that intact homeologs exhibited significantly lower π and Pn/Ps values than the paired homeolog carrying a LoF mutation, consistent with purifying selection acting on the intact copy ($p = 2.14 \times 10^{-26}$ for π; $p = 5.15 \times 10^{-10}$ for Pn/Ps; one-sided *t*-tests Figure 3E). Supporting this further, we also observed sub-genome-biased expression favoring the intact copy, a pattern absent in synthetic *A. suecica*, suggesting it is an evolved rather than inherited expression change (Supplementary Fig. 6). (Supplementary Figure 6). We observed no consistent bias in retention from either sub-genome, suggesting that functional constraint, rather than sub-genome dominance, is the primary force shaping homeolog retention in *A. suecica*.

Genes that consistently revert to a single-copy state following whole-genome duplication are thought to be under strong evolutionary constraint, likely because they function optimally at single-copy dosage and are sensitive to the effects of gene duplication(Edger & Pires, 2009; Makino & McLysaght, 2010). To test whether such dosage constraints influence homeolog retention in *A. suecica*, we asked whether genes previously identified as evolving toward single-copy status across angiosperms(Z. Li et al., 2016) are overrepresented among homeologs carrying LoF mutations. We also tested whether LoF mutations in these single-copy genes occur at higher frequencies in *A. suecica* populations compared to those in genes typically retained in multiple copies. Using the published gene set from(Z. Li et al., 2016), we observed no significant enrichment of single-copy genes among homeologs carrying a LoF mutation, nor do we see that LoF mutations within single copy genes exist at higher population frequencies than LoF mutations within multi-copy genes in *A. suecica* (Supplementary Figure 5). These results suggest that constraint on single-copy gene retention is not a major factor shaping homeolog loss of function in *A. suecica*, at least at this stage of genome evolution.

In the three *A. suecica* homeolog gene pairs where we do observe segregating dual LOF mutations, we tested whether functionally redundant copies from older whole-genome duplications (WGDs) might compensate for complete gene function loss. Using ancient homeolog pairs from *A. thaliana* (Guo et al., 2013), we found no overlap with our dual LOF genes, suggesting these losses are not buffered by retained duplicates from older WGDs. The low frequency of dual LOF mutations (3 out of 414 gene pairs) supports the action of strong purifying selection in the population against complete loss of the gene function.

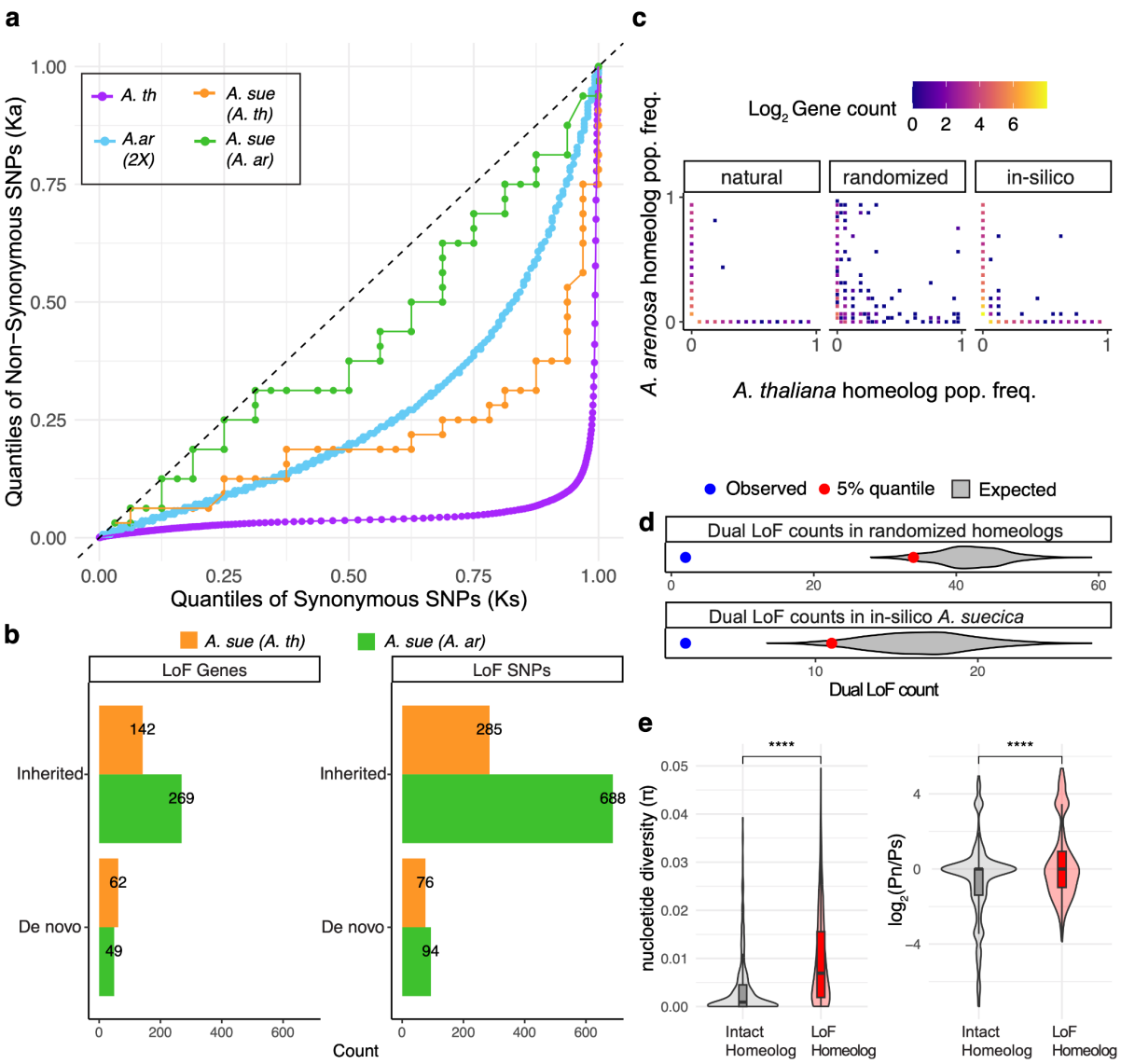


Figure 3: Relaxed purifying selection and homeolog compensation in *A. suecica*

a QQ plot of synonymous (Ks) and non-synonymous (Ka) SNP frequencies for *A. suecica* and its parent species (*A. thaliana* and diploid *A. arenosa*).The dotted line indicates neutral evolution. All species exhibit purifying selection, with the *A. suecica* sub-genomes showing relaxed purifying selection relative to the parents. **b** Bar plots showing the number of genes and total counts of LoF SNPs in *A. suecica*, the majority of which are inherited from the parent species. **c** Joint frequency spectra of homeolog pairs with at least one LoF SNP, displaying population frequencies for *A. thaliana* and *A. arenosa* homeologs on the x- and y-axes, respectively. From left to right: observed spectrum in natural *A. suecica*, example from permutations of randomized homeologs and in silico *A. suecica* controls (random combinations of *A. thaliana* and *A. arenosa* individuals). The number of homeologs analyzed are the same (414 homeologs), but the distribution of LoF mutations are different between the three panels. **d** Comparison of natural *A. suecica* with 1,000 randomized and in-silico controls, showing that *A. suecica* (blue dot) has fewer dual LoF SNPs in homeolog pairs than expected by chance (gray violin plot), a difference that is statistically significant (red dot denotes p-value = 0.05) **e** Evidence of purifying selection acting on the intact homeolog, which shows significantly lower nucleotide diversity (π) and a lower ratio of non-synonymous (Pn) to synonymous polymorphisms (Ps) than the LoF homeolog, within *A. suecica.* Statistical significance is indicated by stars (one-sided *t*-test).

Discussion

A diploid origin for *A. suecica*

A. suecica is a relatively young allopolyploid species that originated from the hybridization of *A. thaliana* and *A. arenosa*. Using population-level DNA sequencing for all three species, we revisited the genetic origins of *A. suecica* to understand the nature of WGD leading to the formation of the allopolyploid species and if the ancestry shaped adaptation to allopolyploidy.

Leveraging the recent genome assembly of *A. suecica*(Burns et al., 2021), we minimized reference bias and identified the genetically closest individuals of both progenitor species. Consistent with earlier reports(Novikova et al., 2017), we confirm that the *A. thaliana* parent was diploid. However, we now find that the *A. arenosa* lineage closest to *A. suecica* is also diploid, not tetraploid. Together, this supports a model in which *A. suecica* originated via somatic genome doubling in a diploid hybrid background, rather than through fusion of unreduced gametes. A similar pattern is also seen in *A. kamchatica*, another allopolyploid *Arabidopsis* species, which also derives from diploid progenitors(Kolesnikova et al., 2023; Scott et al., 2024). These findings suggest that somatic doubling of diploid hybrids may be a general mechanism for allopolyploid formation in *Arabidopsis*.

Evolution of meiosis and genome stability in *A. suecica*

Our STRING network revealed four functional clusters, sister chromatid cohesion, DNA methylation, spindle elongation, and mismatch repair that together suggest genomic adaptation to the challenges of recombination and chromosome segregation in allopolyploid meiosis.

The sister chromatid cohesion cluster includes *SCC2* and *WAPL2*, which are central regulators of cohesin dynamics. *SCC2* encodes the conserved cohesin loader required for deposition of cohesin complexes along chromosomes(H. Wang et al., 2020). In *A. thaliana*, SCC2 localizes to meiotic chromosome axes and is essential for axis formation, synapsis, and homolog pairing(H. Wang et al., 2020). *WAPL2*, conversely, promotes cohesin removal, allowing for the timely release of cohesion during meiotic prophase I(Srinivasan et al., 2019; H. Wang et al., 2020) and in *A. thaliana*, loss of *WAPL* leads to unresolved bivalents and uneven chromosome segregation(De et al., 2014). *Shugoshin* has an important role in protecting cohesion at centromeres at anaphase I(Watanabe & Kitajima, 2005). Together these genes support the functional importance of cohesion and the cohesin complex in polyploid meiosis. We note that SCC2 shows independent signatures of selection in both *A. arenosa* and *A. suecica*, suggesting its general importance in relation to polyploid genome stabilization.

The DNA methylation cluster contains *SUVH2* and *DDM1*. *SUVH2* has been previously shown to play an important role in suppressing the transcription of genes that are associated with meiotic double strand breaks (DSBs) (C. Wang et al., 2022) while a deficiency in *DDM1* has been shown to lead to an increase in meiotic crossovers in *A. thaliana*(Melamed-Bessudo & Levy, 2012). Their selection in *A. suecica* suggests the importance of chromatin in allopolyploids possibly by restricting DSBs and recombination to appropriate genomic regions.

637 The mismatch repair cluster contains the key genes *MutS* and *PMS1*. In allohexaploid wheat,
638 the *Ph2* locus encodes a *MutS* homolog *MSH7-3D* which has been shown to suppress
639 homeologous recombination by destabilizing homeologous interactions (Serra et al., 2021).
640 *MSH7-3D* is thought to influence recombination partner choice and favor homologous over
641 homeologous chromosomes by destabilizing recombination intermediates that contain
642 mismatches(Serra et al., 2021). In yeast, *PMS1* has been shown to reduce recombination
643 between homeolog chromosomes in meiosis(Chambers et al., 2023) and loss of MMR activity
644 in *A. thaliana* has been shown to increase homeologous recombination nine-fold(L. Li et al.,
645 2006).
646 The spindle elongation cluster which includes the cell cycle kinase *MPS1-2*. *MPS1-2* is
647 essential for spindle pole body duplication and the spindle assembly checkpoint(Jiang et al.,
648 2009), and polyploid yeast have been shown to be critically dependent on *MPS1-2* for
649 successful chromosome segregation(Meyer et al., 2023) suggesting a conserved role in
650 supporting meiotic stability in polyploid genomes.
651 Taken together, these four clusters suggest a coordinated mechanism by which *A.*
652 *suecica* has adapted its meiotic machinery to maintain genome stability following
653 allopolyploidization. First, cohesion regulators and cohesin-associated proteins, including
654 *SCC2*, *WAPL2*, and *Shugoshin*, help establish the meiotic axis and form the physical scaffold
655 necessary for proper homolog alignment. Second, DNA methylation and heterochromatin-
656 associated proteins, such as *DDM1* and *SUVH2*, contribute to shaping the chromatin
657 landscape to regulate the formation of DSBs, which are essential for recombination. Third,
658 MMR proteins, including *MutS* and *PMS1*, destabilize mismatched recombination
659 intermediates to favor homologous over homeologous pairing. Finally, spindle checkpoint
660 components, such as *MPS1-2*, ensure accurate chromosome orientation and segregation
661 during meiosis.
662

663 Barriers to autopolyploids as progenitors of allopolyploids

664 The limited contribution of autopolyploids to the origin of allopolyploids raises questions
665 about their suitability as allopolyploid progenitors. Despite harboring alleles pre-adapted to
666 polyploidy, we propose that autopolyploids may be constrained in two key ways. First,
667 autopolyploids with four homologous chromosome copies can accumulate deleterious
668 mutations due to relaxed purifying selection(Baduel et al., 2019; Bao et al., 2022). When an
669 autopolyploid gamete contributes to the origin of an allopolyploid, these mutations may
670 become exposed and, if not fully compensated by a homeolog, could lead to reduced fitness
671 and increased extinction risk. Second, autopolyploids like *A. arenosa* exhibit altered meiotic
672 crossover patterns(Bomblies et al., 2016). Whether these altered crossover patterns are
673 beneficial or reversible or could be disruptive for homeolog segregation in allopolyploids
674 remains unknown, but it may further reduce their viability as allopolyploid parents.
675

676 Homeolog compensation and early stages of rediploidization

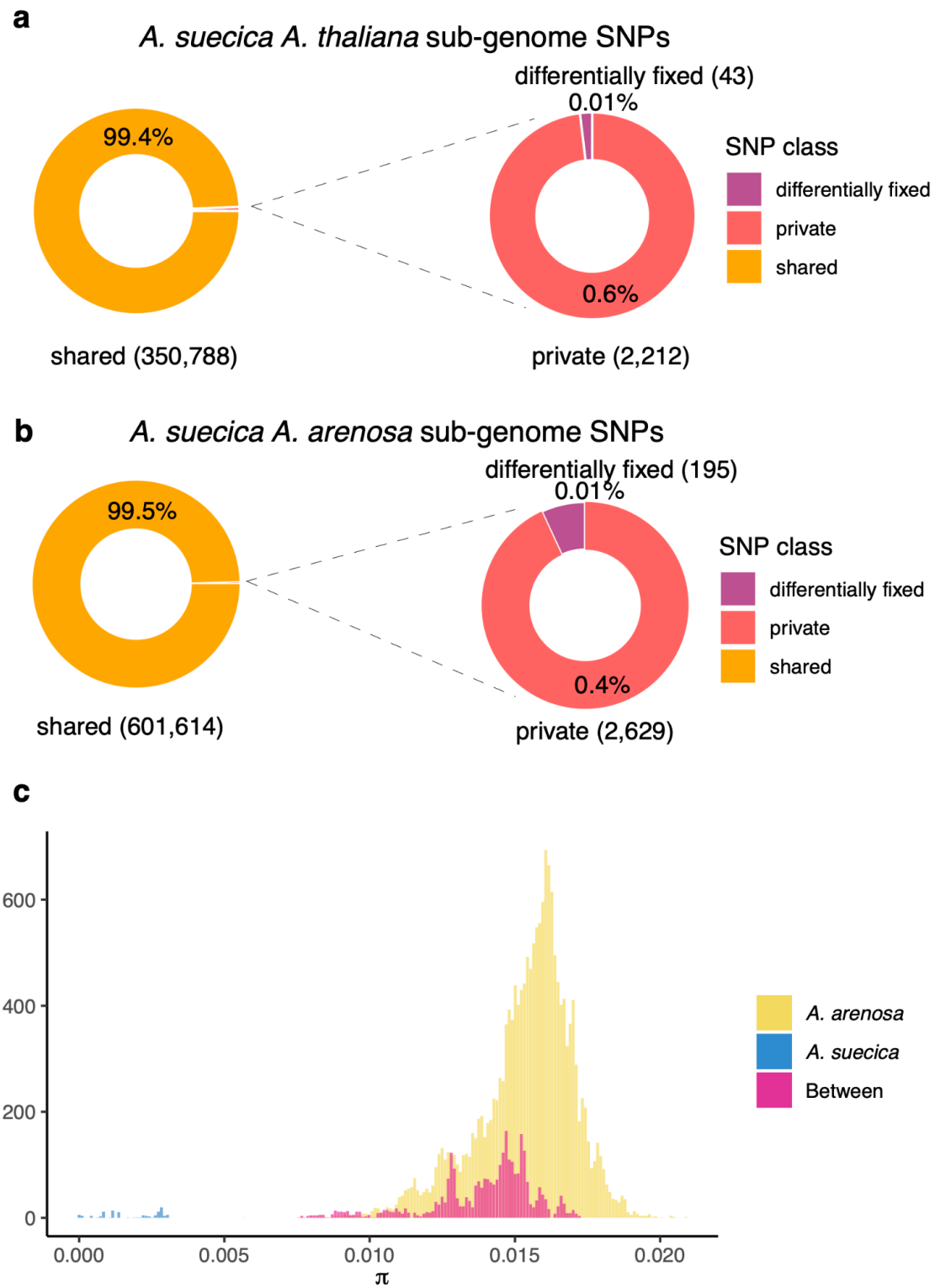
677 We find evidence of relaxed purifying selection in *A. suecica*, consistent with the idea
678 that polyploidy confers increased mutational robustness through genetic redundancy.
679 Previous studies have shown that diploid *A. thaliana* is more sensitive to mutagenesis than
680 both natural and synthetic *A. suecica*(Spoelhof et al., 2021), supporting the claim that *A.*

suecica has enhanced mutational robustness compared to its diploid parents. Despite the evidence of relaxed purifying selection, *A. suecica* has accumulated relatively few genes with a *de novo* LoF mutation suggesting that selection has still remained sufficiently strong to limit the accumulation of new LoF variants.

We observed bias in *A*. *suecica* toward LoF mutations in the *A*. *arenosa* sub-genome, however, most of these LoF mutations also segregate in *A. arenosa* and are likely inherited. This higher number of inherited mutations from the *A. arenosa* sub-genome are due to the higher genetic diversity and outcrossing nature of *A. arenosa*, which allows it to better mask deleterious mutations compared to the highly homozygous and predominantly selfing *A. thaliana*. *De novo* LoF mutations did not show sub-genome bias, consistent with our previous study, showing absence of sub-genome dominance on the transcriptional level(Burns et al., 2021).

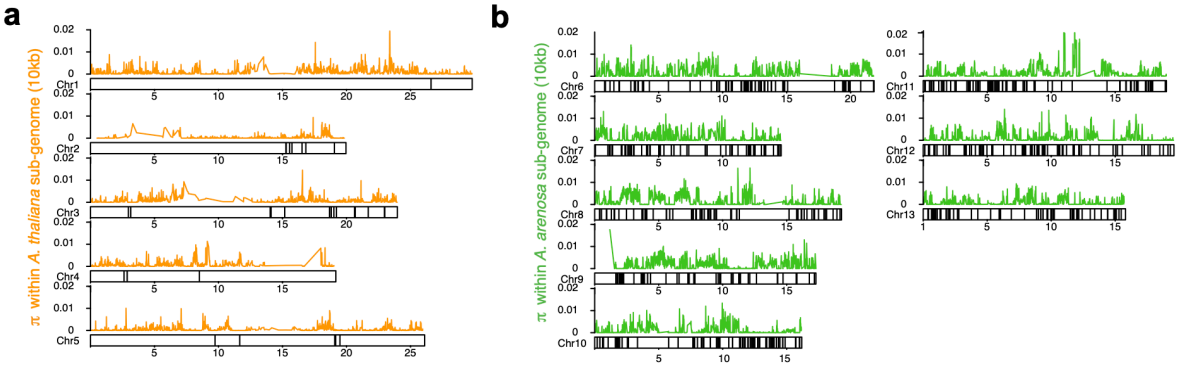
Our analysis of LoF mutations in *A. suecica* reveals a pattern of compensation between homeologs, whereby the retention of at least one functional copy mitigates the impact of deleterious mutations. Specifically, when a LoF mutation arises in one homeolog, the other tends to remain intact more often than expected by chance. We find that these retained homeologs are under purifying selection, suggesting a selective pressure to preserve essential gene functions. This pattern aligns with findings in yeast, where duplicate genes can enhance mutational robustness but are often rapidly inactivated by LoF mutations, rather than undergoing gradual subfunctionalization(Mihajlovic et al., 2024). The homeolog compensation pattern is symmetric across sub-genomes; we find no bias in the accumulation of *de novo* LoF mutations or in the representation of dosage-sensitive single-copy genes. These patterns suggest that, shortly after allopolyploidization, mutations occur at random across sub-genomes, with selection acting *post hoc* to retain at least one functional copy per gene pair. While relaxed purifying selection allows LoF mutations to accrue, the compensatory retention of one intact copy buffers against gene function loss. This buffering, leveraging both standing and *de novo* variation, likely facilitates early genome stabilization in polyploids and helps maintain overall functional integrity. Yet this same genetic robustness may come at a cost: by masking the effects of deleterious mutations, homeolog compensation facilitates their persistence, setting the stage for gene loss to occur over time. We propose that homeolog compensatory dynamics mark the very beginning of rediploidization, a process by which the initial robustness of a polyploid genome ultimately shapes its return to a diploid-like state.

Supplemental Figures

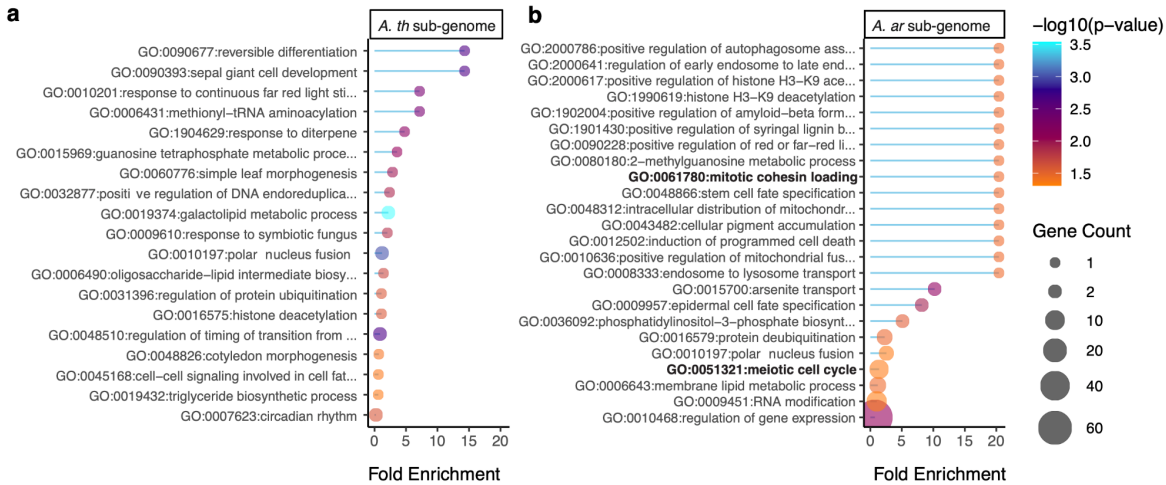


Supplementary Figure 1: SNP Sharing and Nucleotide Diversity (π) in *A. suecica*

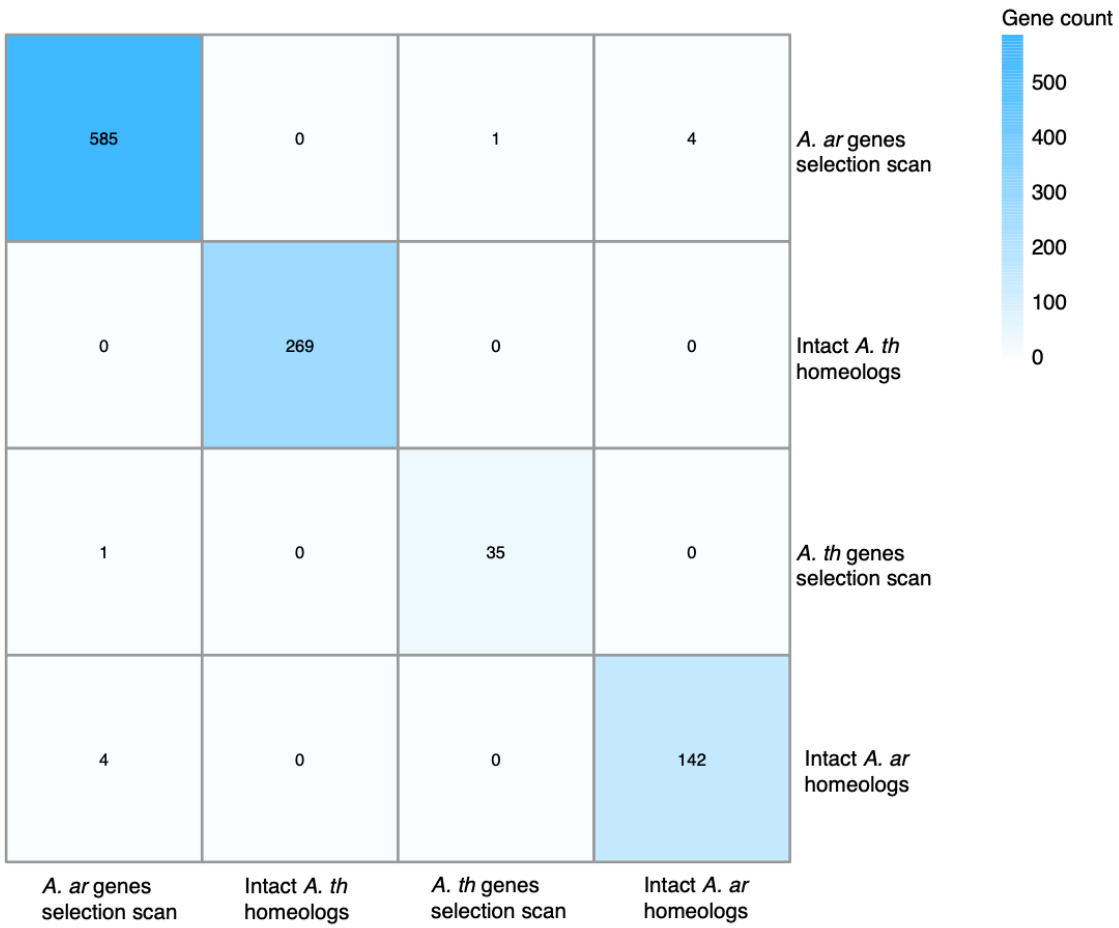
a, b Over 99% of *A. suecica* SNPs are shared with progenitor species (*A. thaliana* and *A. arenosa*), indicating contributions from multiple individuals of both progenitors. **c** Nucleotide diversity (π) is significantly lower in *A. suecica* (blue) compared to *A. arenosa* (yellow) due to the allopolyploidy bottleneck and self-compatibility. In contrast, the obligate outcrosser *A. arenosa* displays greater diversity within species than between species (pink).



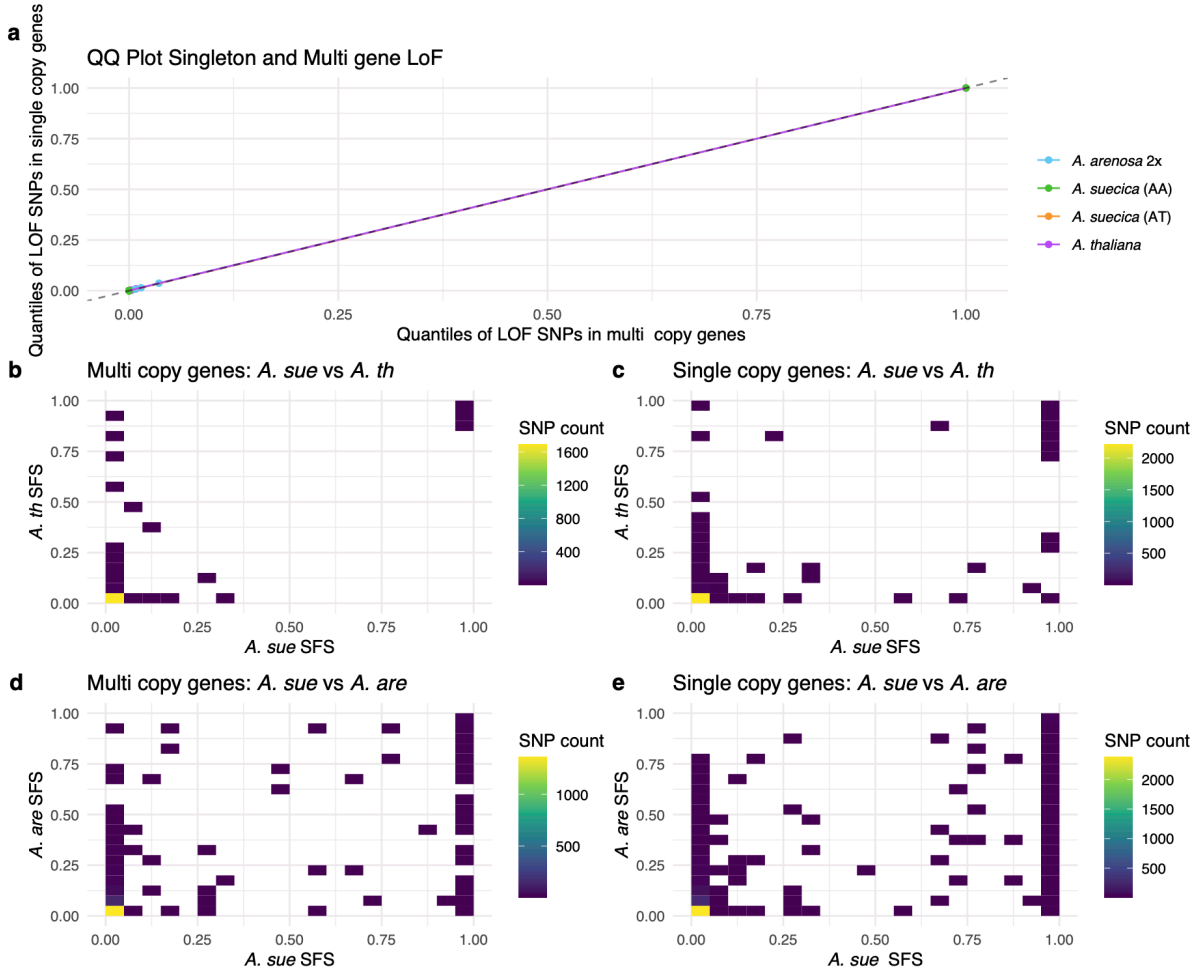
Supplementary Figure 2 a In the *A. thaliana* sub-genome (Chr1 to 5), 35 genes, and **b** in the *A. arenosa* sub-genome (Chr 6 to 13), 585 genes show signatures of positive selection. These genes are found outside of the areas of low nucleotide diversity (π) within *A. suecica*, meaning a founder effect or population bottleneck is not the likely explanation for the signature of positive selection.



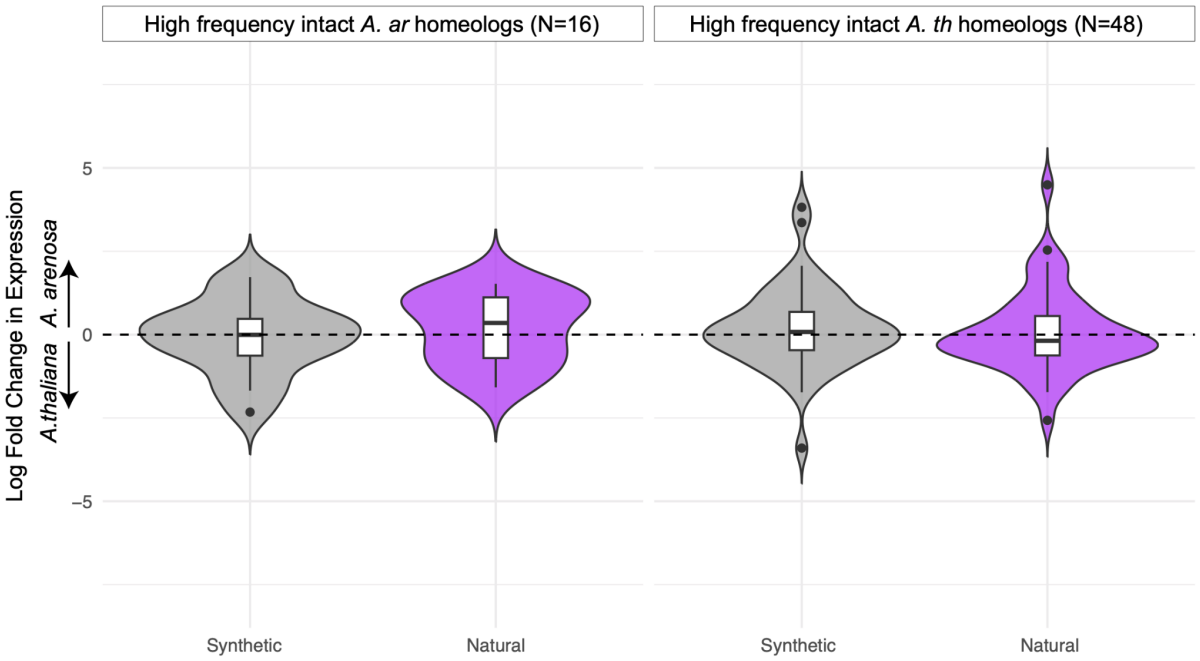
Supplementary Figure 3. GO analysis of selection scan genes in *A. suecica* GO enrichment for the **a** 35 genes on the A. thaliana sub-genome and **b** the 585 genes on the A. arenosa sub-genome. GO terms mentioned in the text are highlighted in bold.



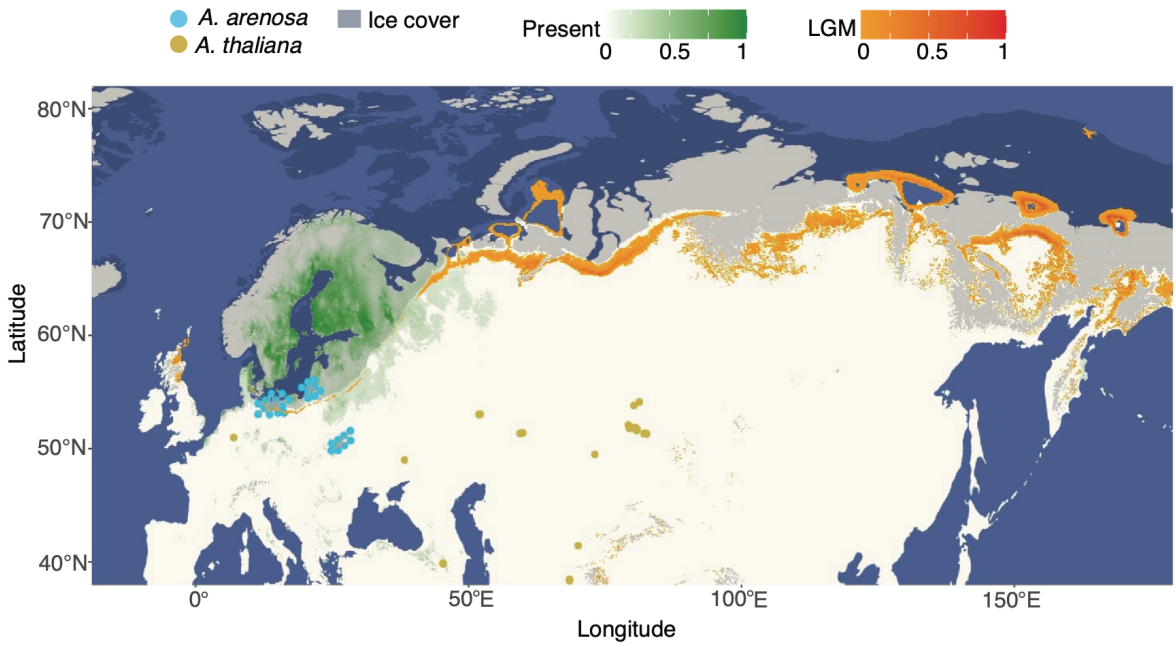
Supplementary Figure 4 Overlap of genes in selection scans and homeologous gene pairs carrying one LoF mutation. Minimal overlap was observed between genes from selection scans and homeologous gene pairs with one intact copy and one carrying a LoF mutation, indicating that LoF on one homeolog is not driving the likely adaptive genetic changes on the other.



Supplementary Figure 5. Lack of enrichment for LoF mutations in single-copy genes in *A. suecica.* **a** QQ plots comparing the population frequencies of LoF SNPs in single-copy genes versus multi-copy genes in *A. suecica*. No significant increase in the population frequency of LoF mutations is observed in *A. suecica* for single-copy genes, indicating that both single- and multi-copy genes are under similar selective pressures. **b-e** Joint allele frequency spectra for LoF SNPs, comparing *A. suecica* and the parental species for multi-copy and single-copy genes. The data show an enrichment of singleton SNPs for all species and gene copy status, consistent with strong purifying selection acting on these deleterious mutations.



Supplementary Figure 6. Sub-genome expression bias towards intact homeolog. Violin plots show the $\log_2$ fold change in gene expression between homeologs in synthetic versus natural *A. suecica* accessions. Expression bias is plotted for genes where one homeolog exhibits a high-frequency loss-of-function mutation (≥60% in the population) while the homeolog either on the *A. arenosa* sub-genome (left panel, N = 16) or in the *A. thaliana* sub-genome (right panel, N = 48) remains intact. Each violin represents the distribution of expression bias for either synthetic (gray) or natural (purple) *A. suecica*. A log fold change of zero (dashed line) indicates balanced expression between homeologs. Expression in natural *A. suecica* is biased toward the intact homeolog, consistent with purifying selection maintaining expression of the functional copy. This expression bias is absent in synthetic *A. suecica*, indicating that it is not an inherited regulatory pattern but rather an evolved expression change due to the presence of a nonfunctional allele in the other homeolog.



Supplementary Figure 7: The geographic origin of *A. suecica*. Niche modeling of *A. suecica* shows probability of occurrence for the current climate (green) and the LGM (orange). Closest parental populations include Southern Baltic and Ukrainian diploid *A. arenosa*, and the top 5% of *A. thaliana* accessions genetically closest to *A. suecica*.

Code availability

Custom scripts used in this study are available at https://github.com/burnsro/ArabidopsisSuecica/tree/main/origin

Data availability

Raw fastq files of the Illumina sequenced herbarium samples for 5 tetraploid *A. arenosa* and 1 *A. suecica* generated in this study were deposited to the European Nucleotide Archive (ENA) under the project number PRJEB79527.

Author contributions

R.B and P.Y.N conceived the study. Herbarium samples were aquired by P.Y.N and genomic DNA was extracted by R.B. R.B, P.Y.N, A.K, A.G, U.K.K, A.D.S analyzed the data. R.B, P.Y.N, F.K and A.D.S wrote the paper.

Acknowledgements

This work was supported by an EMBO long-term postdoctoral fellowship ALTF224-2022 to R.B. P.Y.N. acknowledges the European Union (ERC, HOW2DOUBLE, 101041354). The views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. F.K. and P.Y.N. acknowledge the joint Deutsche Forschungsgemeinschaft (DFG) and Czech Science Foundation (GACR) program, project number 490698526 and 22-29078K. F.K. and P.Y.N. also acknowledge the DAAD exchange travel grant, project number 57601834. We thank Heïdi Serra, Andrew Lloyd and Nicola Gorringe for helpful comments on the manuscript and discussions.

Competing interests

The authors declare no competing interests.

References

1001 Genomes Consortium. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, *166*(2), 481–491.

Adrian Alexa, J. R. (2017). *topGO*. Bioconductor. https://doi.org/10.18129/B9.BIOC.TOPGO

Aköz, G., & Nordborg, M. (2019). The Aquilegia genome reveals a hybrid origin of core eudicots. *Genome Biology*, *20*(1), 256.

Alger, E. I., & Edger, P. P. (2020). One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Current Opinion in Plant Biology*, *54*, 108–113.

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *The Journal of Applied Ecology*, *43*(6), 1223–1232.

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*, *23*(1), 1–19.

Bachmann, J. A., Tedder, A., Fracassetti, M., Steige, K. A., Lafon-Placette, C., Köhler, C., & Slotte, T. (2021). On the origin of the widespread self-compatible allotetraploid Capsella

bursa-pastoris (Brassicaceae). *Heredity*. https://doi.org/10.1038/s41437-021-00434-9

Baduel, P., Quadrana, L., Hunter, B., Bomblies, K., & Colot, V. (2019). Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nature Communications*, *10*(1), 5818.

Bao, Z., Li, C., Li, G., Wang, P., Peng, Z., Cheng, L., Li, H., Zhang, Z., Li, Y., Huang, W., Ye, M., Dong, D., Cheng, Z., VanderZaag, P., Jacobsen, E., Bachem, C. W. B., Dong, S., Zhang, C., Huang, S., & Zhou, Q. (2022). Genome architecture and tetrasomic inheritance of autotetraploid potato. *Molecular Plant*, *15*(7), 1211–1226.

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. C. M., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy, B., Agarwal, G., Ballén-Taborda, C., Cameron, C., Campbell, J., Chavarro, C., Chitikineni, A., Chu, Y., … Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut Arachis hypogaea. *Nature Genetics*, *51*(5), 877–884.

Bohutínská, M., Petříková, E., Booker, T. R., & Cobo, C. V. (2023). Repeated adaptation to whole genome duplication in outcrossing Arabidopsis is mediated by mosaic adaptive haplotypes. *bioRxiv*. https://www.biorxiv.org/content/10.1101/2023.01.11.523565.abstract

Bohutínská, M., Petříková, E., Booker, T. R., Cobo, C. V., Vlček, J., Šrámková, G., Poštulková, A., Hojka, J., Marhold, K., Yant, L., Kolář, F., & Schmickl, R. (2023a). *Novel reassortments of variation underlie repeated adaptation to whole genome duplication in two outcrossing Arabidopsis species*. https://doi.org/10.1101/2023.01.11.523565

Bohutínská, M., Petříková, E., Booker, T. R., Cobo, C. V., Vlček, J., Šrámková, G., Poštulková, A., Hojka, J., Marhold, K., Yant, L., Kolář, F., & Schmickl, R. (2023b). Repeated adaptation to whole genome duplication in outcrossing Arabidopsis is mediated by mosaic adaptive haplotypes. In *bioRxiv* (p. 2023.01.11.523565). https://doi.org/10.1101/2023.01.11.523565

Bohutínská, M., Petříková, E., Booker, T. R., Cobo, C. V., Vlček, J., Šrámková, G., Poštulková, A., Hojka, J., Marhold, K., Yant, L., Kolář, F., & Schmickl, R. (2024). Novel

reassortments of variation underlie repeated adaptation to whole genome duplication in two outcrossing Arabidopsis species. In *bioRxiv* (p. 2023.01.11.523565). https://doi.org/10.1101/2023.01.11.523565

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., Slotte, T., & Kolář, F. (2021). Genomic basis of parallel adaptation varies with divergence in and its relatives. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(21). https://doi.org/10.1073/pnas.2022713118

Bomblies, K., Jones, G., Franklin, C., Zickler, D., & Kleckner, N. (2016). The challenge of evolving stable polyploidy: could an increase in “crossover interference distance” play a central role? *Chromosoma*, *125*(2), 287–300.

Bray, S. M., Hämälä, T., Zhou, M., Busoms, S., Fischer, S., Desjardins, S. D., Mandáková, T., Moore, C., Mathers, T. C., Cowan, L., Monnahan, P., Koch, J., Wolf, E. M., Lysak, M. A., Kolar, F., Higgins, J. D., Koch, M. A., & Yant, L. (2024). Kinetochore and ionomic adaptation to whole-genome duplication in Cochlearia shows evolutionary convergence in three autopolyploids. *Cell Reports*, *43*(8), 114576.

Burns, R., Mandáková, T., Gunis, J., Soto-Jiménez, L. M., Liu, C., Lysak, M. A., Novikova, P. Y., & Nordborg, M. (2021). Gradual evolution of allopolyploidy in Arabidopsis suecica. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-021-01525-w

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421.

Chambers, S. R., Hunter, N., Louis, E. J., & Borts, R. H. (2023). The Mismatch Repair System Reduces Meiotic Homeologous Recombination and Stimulates Recombination-Dependent Chromosome Loss. *Molecular and Cellular Biology*, 6110–6120.

Cheng, F., Wu, J., & Wang, X. (2014). Genome triplication drove the diversification of Brassica plants. *Horticulture Research*, *1*, 14024.

Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., Ding, M., Ye, W., Kirkbride, R. C., Jenkins, J., Plott, C., Lovell, J., Lin, Y.-M., Vaughn, R., Liu, B., Simpson, S., Scheffler, B. E., Wen, L., Saski, C. A., … Schmutz, J. (2020).

Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. *Nature Genetics*, *52*(5), 525–533.

Chéron, F., Petiot, V., Lambing, C., White, C., & Serra, H. (2023). Incorrect recombination partner associations contribute to meiotic instability of neo-allopolyploid Arabidopsis suecica. *The New Phytologist*. https://doi.org/10.1111/nph.19487

Conover, J. L., & Wendel, J. F. (2022). Deleterious Mutations Accumulate Faster in Allopolyploid Than Diploid Cotton (Gossypium) and Unequally between Subgenomes. *Molecular Biology and Evolution*, *39*(2), msac024.

Dainat, J., Hereñú, D., Murray, K. D., Davis, E., Ugrin, I., Crouch, K., LucileSol, Agostinho, N., pascal-git, Zollman, Z., & tayyrov. (2024). *NBISweden/AGAT: AGAT-v1.4.1*. Zenodo. https://doi.org/10.5281/ZENODO.3552717

de Carvalho, J. F., Stoeckel, S., Eber, F., Lodé-Taburel, M., Gilet, M.-M., Trotoux, G., Morice, J., Falentin, C., Chèvre, A.-M., & Rousseau-Gueutin, M. (2021). Untangling structural factors driving genome stabilization in nascent Brassica napus allopolyploids. *New Phytologist*, *230*(5), 2072–2084.

De, K., Sterle, L., Krueger, L., Yang, X., & Makaroff, C. A. (2014). Arabidopsis thaliana WAPL Is Essential for the Prophase Removal of Cohesin during Meiosis. *PLoS Genetics*, *10*(7), e1004497.

Dubcovsky, J., & Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, *316*(5833), 1862–1866.

Dunnington, D. (2017). ggspatial: Spatial Data Framework for ggplot2 [Dataset]. In *CRAN: Contributed Packages*. The R Foundation. https://doi.org/10.32614/cran.package.ggspatial

Edger, P. P., & Pires, J. C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research*, *17*(5), 699–717.

Feldman, M., & Levy, A. A. (2005). Allopolyploidy--a shaping force in the evolution of wheat genomes. *Cytogenetic and Genome Research*, *109*(1-3), 250–258.

Gaeta, R. T., & Chris Pires, J. (2010). Homoeologous recombination in allopolyploids: the

894 polyploid ratchet. *New Phytologist*, *186*(1), 18–28.
895 Genete, M., Castric, V., & Vekemans, X. (2020). Genotyping and De Novo Discovery of
896 Allelic Variants at the Brassicaceae Self-Incompatibility Locus from Short-Read
897 Sequencing Data. *Molecular Biology and Evolution*, *37*(4), 1193–1201.
898 Gonzalo, A. (2022). All Ways Lead to Rome-Meiotic Stabilization Can Take Many Routes in
899 Nascent Polyploid Plants. *Genes*, *13*(1). https://doi.org/10.3390/genes13010147
900 Greig, D., Travisano, M., Louis, E. J., & Borts, R. H. (2003). A role for the mismatch repair
901 system during incipient speciation in Saccharomyces. *Journal of Evolutionary Biology*,
902 *16*(3), 429–437.
903 Guo, H., Lee, T. H., Wang, X., & Paterson, A. H. (2013). Function relaxation followed by
904 diversifying selection after whole-genome duplication in flowering plants. *Plant*
905 *Physiology*, *162*(2). https://doi.org/10.1104/pp.112.213447
906 Gutaker, R. M., Reiter, E., Furtwängler, A., Schuenemann, V. J., & Burbano, H. A. (2017).
907 Extraction of ultrashort DNA molecules from herbarium specimens. *BioTechniques*,
908 *62*(2), 76–79.
909 Hämälä, T., Moore, C., Cowan, L., Carlile, M., Gopaulchan, D., Brandrud, M. K., Birkeland,
910 S., Loose, M., Kolář, F., Koch, M. A., & Yant, L. (2024). Impact of whole-genome
911 duplications on structural variant evolution in Cochlearia. *Nature Communications*,
912 *15*(1), 1–13.
913 Henry, I. M., Dilkes, B. P., Tyagi, A., Gao, J., Christensen, B., & Comai, L. (2014). The BOY
914 NAMED SUE quantitative trait locus confers increased meiotic stability to an adapted
915 natural allopolyploid of Arabidopsis. *The Plant Cell*, *26*(1), 181–194.
916 Hijmans, R. J. (2018). raster: Geographic data analysis and modeling. *R Package Version*,
917 *2*, 8.
918 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2:
919 Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*,
920 *35*(2), 518–522.
921 Hokamp, K., McLysaght, A., & Wolfe, K. H. (2003). The 2R hypothesis and the human

genome sequence. *Genome Evolution*, 95–110.

Hollister, J. D., Arnold, B. J., Svedin, E., Xue, K. S., Dilkes, B. P., & Bomblies, K. (2012). Genetic adaptation associated with genome-doubling in autotetraploid Arabidopsis arenosa. *PLoS Genetics*, *8*(12), e1003093.

Hylander, N. (1957). Cardaminopsis suecica (fr.) Hiit., a Northern Amphidiploid species. *Bulletin Du Jardin Botanique de l'État a Bruxelles*, *27*(4), 591–604.

Jaegle, B., Pisupati, R., Soto-Jiménez, L. M., Burns, R., Rabanal, F. A., & Nordborg, M. (2023). Extensive sequence duplication in Arabidopsis revealed by pseudo-heterozygosity. *Genome Biology*, *24*(1), 1–19.

Jakobsson, M., Hagenblad, J., Tavaré, S., Säll, T., Halldén, C., Lind-Halldén, C., & Nordborg, M. (2006). A Unique Recent Origin of the Allotetraploid Species Arabidopsis suecica: Evidence from Nuclear DNA Markers. *Molecular Biology and Evolution*, *23*(6), 1217–1231.

Jiang, H., Wang, F.-F., Wu, Y.-T., Zhou, X., Huang, X.-Y., Zhu, J., Gao, J.-F., Dong, R.-B., Cao, K.-M., & Yang, Z.-N. (2009). MULTIPOLAR SPINDLE 1 (MPS1), a novel coiled-coil protein of Arabidopsis thaliana, is required for meiotic spindle organization. *The Plant Journal*, *59*(6), 1001–1010.

Katche, E. I., Schierholt, A., Becker, H. C., Batley, J., & Mason, A. S. (2023). Fertility, genome stability, and homozygosity in a diverse set of resynthesized rapeseed lines. *The Crop Journal*. https://www.sciencedirect.com/science/article/pii/S2214514122002082

Katche, E. I., Schierholt, A., Schiessl, S.-V., He, F., Lv, Z., Batley, J., Becker, H. C., & Mason, A. S. (2023). Genetic factors inherited from both diploid parents interact to affect genome stability and fertility in resynthesized allotetraploid Brassica napus. *G3* , *13*(8). https://doi.org/10.1093/g3journal/jkad136

Katoh, K., Kuma, K.-I., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, *33*(2), 511–518.

Klaus Schliep, Marta Vidal-Garcia, Claudia Solis-Lemus, Leann Biancani, Eren Ada, L.

Francisco Henao Diaz, Guangchuang Yu. (2021). *tanggle*. Bioconductor. https://doi.org/10.18129/B9.BIOC.TANGGLE

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53.

Kolář, F., Fuxová, G., Záveská, E., Nagano, A. J., Hyklová, L., Lučanová, M., Kudoh, H., & Marhold, K. (2016). Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model Arabidopsis arenosa. *Molecular Ecology*, *25*(16), 3929–3949.

Kolesnikova, U. K., Scott, A. D., Van de Velde, J. D., Burns, R., Tikhomirov, N. P., Pfordt, U., Clarke, A. C., Yant, L., Seregin, A. P., Vekemans, X., Laurent, S., & Novikova, P. Y. (2023). Transition to Self-compatibility Associated With Dominant S-allele in a Diploid Siberian Progenitor of Allotetraploid Arabidopsis kamchatica Revealed by Arabidopsis lyrata Genomes. *Molecular Biology and Evolution*, *40*(7). https://doi.org/10.1093/molbev/msad122

Li, A.-L., Geng, S.-F., Zhang, L.-Q., Liu, D.-C., & Mao, L. (2015). Making the Bread: Insights from Newly Synthesized Allohexaploid Wheat. *Molecular Plant*, *8*(6), 847–859.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

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

Li, L., Jean, M., & Belzile, F. (2006). The impact of sequence divergence and DNA mismatch repair on homeologous recombination in Arabidopsis. *The Plant Journal*, *45*(6), 908–916.

Liston, A., Cronn, R., & Ashman, T.-L. (2014). Fragaria: a genus with deep historical roots and ripe for evolutionary and ecological insights. *American Journal of Botany*, *101*(10), 1686–1699.

Liu, Z., Adamczyk, K., Manzanares-Dauleux, M., Eber, F., Lucas, M.-O., Delourme, R.,

Chèvre, A. M., & Jenczewski, E. (2006). Mapping PrBn and other quantitative trait loci responsible for the control of homeologous chromosome pairing in oilseed rape (Brassica napus L.) haploids. *Genetics*, *174*(3), 1583–1596.

Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., & De Smet, R. (2016). Gene Duplicability of Core Genes Is Highly Consistent across All Angiosperms. *The Plant Cell*, *28*(2), 326–344.

Makino, T., & McLysaght, A. (2010). Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proceedings of the National Academy of Sciences*, *107*(20), 9270–9274.

Mandáková, T., & Lysak, M. A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology*, *42*, 55–65.

Ma, P.-F., Liu, Y.-L., Guo, C., Jin, G., Guo, Z.-H., Mao, L., Yang, Y.-Z., Niu, L.-Z., Wang, Y.-J., Clark, L. G., Kellogg, E. A., Xu, Z.-C., Ye, X.-Y., Liu, J.-X., Zhou, M.-Y., Luo, Y., Yang, Y., Soltis, D. E., Bennetzen, J. L., … Li, D.-Z. (2024). Genome assemblies of 11 bamboo species highlight diversification induced by dynamic subgenome dominance. *Nature Genetics*, *56*(4), 710–720.

Martín, A. C., Alabdullah, A. K., & Moore, G. (2021). A separation-of-function ZIP4 wheat mutant allows crossover between related chromosomes and is meiotically stable. *Scientific Reports*, *11*(1), 21811.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Melamed-Bessudo, C., & Levy, A. A. (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16). https://doi.org/10.1073/pnas.1120742109

Meyer, R. E., Sartin, A., Gish, M., Harsha, J., Wilkie, E., Haworth, D., LaVictoire, R.,

Alberola, I., Chuong, H. H., Gorbsky, G. J., & Dawson, D. S. (2023). Polyploid yeast are dependent on elevated levels of Mps1 for successful chromosome segregation. In *bioRxiv* (p. 2023.01.09.523325). https://doi.org/10.1101/2023.01.09.523325

Mihajlovic, L., Iyengar, B. R., Baier, F., Barbier, I., Iwaszkiewicz, J., Zoete, V., Wagner, A., & Schaerli, Y. (2024). A direct experimental test of Ohno's hypothesis. *eLife*, *13*. https://doi.org/10.7554/eLife.97216.2

Monnahan, P., Kolář, F., Baduel, P., Sailer, C., Koch, J., Horvath, R., Laenen, B., Schmickl, R., Paajanen, P., Šrámková, G., Bohutínská, M., Arnold, B., Weisman, C. M., Marhold, K., Slotte, T., Bomblies, K., & Yant, L. (2019). Pervasive population genomic consequences of genome duplication in Arabidopsis arenosa. *Nature Ecology & Evolution*, *3*(3), 457–468.

Nasrallah, M. E., Yogeeswaran, K., Snyder, S., & Nasrallah, J. B. (2000). Arabidopsis species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiology*, *124*(4). https://doi.org/10.1104/pp.124.4.1605

Nei, M. (1972). Genetic Distance between Populations. *The American Naturalist*, *106*(949), 283–292.

Newton, nW C. F., & Pellew, N. (1929). Primula kewensis and its derivatives. *Journal of Genetics*, *20*(3), 405–467.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Nibau, C., Evans, A., King, H., Phillips, D. W., & Lloyd, A. (2024). Homoeologous crossovers are distally biased and underlie genomic instability in first-generation neo-allopolyploid Arabidopsis suecica. *The New Phytologist*. https://doi.org/10.1111/nph.20095

Nibau, C., Gonzalo, A., Evans, A., Sweet-Jones, W., Phillips, D., & Lloyd, A. (2022). Meiosis in allopolyploid Arabidopsis suecica. *The Plant Journal: For Cell and Molecular Biology*, *111*(4), 1110–1122.

Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., Guggisberg,

A., Paape, T., Schmid, K., Fedorenko, O. M., Holm, S., Sall, T., Schlotterer, C., Marhold, K., Widmer, A., Sese, J., Shimizu, K. K., Weigel, D., Kramer, U., … Nordborg, M. (2016). Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics*, *48*(9), 1077–1082.

Novikova, P. Y., Tsuchimatsu, T., Simon, S., Nizhynska, V., Voronin, V., Burns, R., Fedorenko, O. M., Holm, S., Säll, T., Prat, E., Marande, W., Castric, V., & Nordborg, M. (2017). Genome Sequencing Reveals the Origin of the Allotetraploid Arabidopsis suecica. *Molecular Biology and Evolution*, *34*(4), 957.

O'Toole, Á., Aziz, A., & Maloney, D. (2024). Publication-ready single nucleotide polymorphism visualization with snipit. *Bioinformatics (Oxford, England)*, *40*(8). https://doi.org/10.1093/bioinformatics/btae510

Paape, T., Briskine, R. V., Halstead-Nussloch, G., Lischer, H. E. L., Shimizu-Inatsugi, R., Hatakeyama, M., Tanaka, K., Nishiyama, T., Sabirov, R., Sese, J., & Shimizu, K. K. (2018). Patterns of polymorphism and selection in the subgenomes of the allopolyploid Arabidopsis kamchatica. *Nature Communications*, *9*(1), 3909.

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952.

Penin, A. A., Kasianov, A. S., Klepikova, A. V., Omelchenko, D. O., Makarenko, M. S., & Logacheva, M. D. (2024). Origin and diversity of Capsella bursa-pastoris from the genomic point of view. *BMC Biology*, *22*(1), 1–15.

Ramsey, J., & Schemske, D. W. (2002). Neopolyploidy in Flowering Plants. *Annual Review of Ecology and Systematics*, *33*(1), 589–639.

Redmond, A. K., Casey, D., Gundappa, M. K., Macqueen, D. J., & McLysaght, A. (2023). Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. *Nature Communications*, *14*(1), 1–15.

Renny-Byfield, S., & Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop

plants. *American Journal of Botany*, *101*(10), 1711–1725.

Sánchez-Tapia, A., Mortara, S. R., Bezerra Rocha, D. S., Mendes Barros, F. S., Gall, G., & de Siqueira, M. F. (2020). modleR: a modular workflow to perform ecological niche modeling in R. In *bioRxiv*. bioRxiv. https://doi.org/10.1101/2020.04.01.021105

Scarlett, V. T., Lovell, J. T., Shao, M., Phillips, J., Shu, S., Lusinska, J., Goodstein, D. M., Jenkins, J., Grimwood, J., Barry, K., Chalhoub, B., Schmutz, J., Hasterok, R., Catalán, P., & Vogel, J. P. (2023). Multiple origins, one evolutionary trajectory: gradual evolution characterizes distinct lineages of allotetraploid Brachypodium. *Genetics*, *223*(2). https://doi.org/10.1093/genetics/iyac146

Schnable, J. C., Springer, N. M., & Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(10), 4069.

Scott, A. D., Kolesnikova, U., Glushkevich, A., Steinmann, L., Tikhomirov, N., Pfordt, U., Bohutínská, M., Burns, R., Seregin, A. P., Kolar, F., Schmickl, R., & Novikova, P. Y. (2024). Multiple polyploidizations in Arabidopsis lyrata stabilized by long-range adaptive introgression across Eurasia. In *bioRxiv* (p. 2024.08.27.609292). https://doi.org/10.1101/2024.08.27.609292

Seregin, A. (2024). *Moscow University Herbarium (MW)* [Dataset]. Lomonosov Moscow State University. https://doi.org/10.15468/CPNHCC

Serra, H., Svačina, R., Baumann, U., Whitford, R., Sutton, T., Bartoš, J., & Sourdille, P. (2021). Ph2 encodes the mismatch repair protein MSH7-3D that inhibits wheat homoeologous recombination. *Nature Communications*, *12*(1), 803.

Sierro, N., Auberson, M., Dulize, R., & Ivanov, N. V. (2024). Chromosome-level genome assemblies of Nicotiana tabacum, Nicotiana sylvestris, and Nicotiana tomentosiformis. *Scientific Data*, *11*(1), 135.

Soltis, D. E., Soltis, P. S., & Rieseberg, L. H. (1993). Molecular Data and the Dynamic Nature of Polyploidy. *Critical Reviews in Plant Sciences*, *12*(3), 243–273.

Spoelhof, J. P., Soltis, D. E., & Soltis, P. S. (2021). Polyploidy and mutation in Arabidopsis. *Evolution; International Journal of Organic Evolution*, *75*(9), 2299–2308.

Srinivasan, M., Petela, N. J., Scheinost, J. C., Collier, J., Voulgaris, M., Roig M, B., Beckouët, F., Hu, B., & Nasmyth, K. A. (2019). Scc2 counteracts a Wapl-independent mechanism that releases cohesin from chromosomes during G1. *eLife*, *8*. https://doi.org/10.7554/eLife.44736

Staudt, G. (2009). *STRAWBERRY BIOGEOGRAPHY, GENETICS AND SYSTEMATICS*. International Society for Horticultural Science (ISHS), Leuven, Belgium.

Stebbins, G. L. (1947). Types of polyploids: their classification and significance. *Advances in Genetics*, *1*, 403–429.

Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A. P. M., & Denton, A. K. (2021). Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics* , *36*(22-23), 5291–5298.

Storchová, Z., Breneman, A., Cande, J., Dunn, J., Burbank, K., O'Toole, E., & Pellman, D. (2006). Genome-wide genetic analysis of polyploidy in yeast. *Nature*, *443*(7111), 541–547.

Sun, Y., Liu, Y., Shi, J., Wang, L., Liang, C., Yang, J., Chen, J., & Chen, M. (2023). Biased mutations and gene losses underlying diploidization of the tetraploid broomcorn millet genome. *The Plant Journal: For Cell and Molecular Biology*, *113*(4), 787–801.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(D1), D638–D646.

U, Nagaharu. (1935). Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. *Japanese Journal of Botany*, *7*, 389–452.

Van de Peer, Y. (2023). *Polyploidy: Methods and Protocols*. Springer Nature.

Van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews. Genetics*, *18*(7), 411–424.

Vlček, J., Hämälä, T., Cobo, C. V., Curran, E., Šrámková, G., Slotte, T., Schmickl, R., Yant, L., & Kolář, F. (2025). Whole-genome duplication increases genetic diversity and load in outcrossing Arabidopsis. In *bioRxiv* (p. 2025.01.12.632621). https://doi.org/10.1101/2025.01.12.632621

Wang, C., Huang, J., Zhang, J., Yu, Y., Copenhaver, G. P., You, C., & Wang, Y. (2022). DNA polymerase epsilon interacts with SUVH2/9 to repress the expression of genes associated with meiotic DSB hotspot in Arabidopsis. *Proceedings of the National Academy of Sciences*, *119*(41), e2208441119.

Wang, H., Xu, W., Sun, Y., Lian, Q., Wang, C., Yu, C., He, C., Wang, J., Ma, H., Copenhaver, G. P., & Wang, Y. (2020). The cohesin loader SCC2 contains a PHD finger that is required for meiosis in land plants. *PLOS Genetics*, *16*(6), e1008849.

Watanabe, Y., & Kitajima, T. S. (2005). Shugoshin protects cohesin complexes at centromeres. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1455), 515.

Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S., & Burbano, H. A. (2018). nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*, *19*(1), 122.

Wiiliam, M., & Malcolm, A. (1971). *Biosystematics, genetics, and physiological ecology of the Erythranthe section of Mimulus*.

Wijnker, E., James, G. V., Ding, J., Becker, F., Klasen, J. R., Rawat, V., Rowan, B. A., de Jong, D. F., de Snoo, C. B., Zapata, L., Huettel, B., de Jong, H., Ossowski, S., Weigel, D., Koornneef, M., Keurentjes, J. J. B., & Schneeberger, K. (2013). *The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana*. https://doi.org/10.7554/eLife.01426

Yant, L., Hollister, J. D., Wright, K. M., Arnold, B. J., Higgins, J. D., Franklin, F. C. H., & Bomblies, K. (2013). Meiotic adaptation to genome duplication in Arabidopsis arenosa.

Current Biology: CB, *23*(21), 2151–2156.

Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *69*(1), e96.

Zhang, Z., Gou, X., Xun, H., Bian, Y., Ma, X., Li, J., Li, N., Gong, L., Feldman, M., Liu, B., & Levy, A. A. (2020). Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proceedings of the National Academy of Sciences*, *117*(25), 14561–14571.