Repeated adaptation to whole genome duplication in outcrossing Arabidopsis is mediated by mosaic adaptive haplotypes

Authors:

Magdalena Bohutínská[12*], Eliška Petříková[1], Tom R. Booker[3], Cristina Vives Cobo[1], Jakub Vlček[14], Gabriela Šrámková[1], Alžběta Poštulková[1], Jakub Hojka[12], Karol Marhold[1], Levi Yant[5#], Filip Kolář[12#], Roswitha Schmickl[12#*]

# Equal contribution

* Correspondence:

Magdalena Bohutínská: magdalena.holcova@natur.cuni.cz

Roswitha Schmickl: roswitha.schmickl@natur.cuni.cz

Affiliations:

1 Department of Botany, Faculty of Science, Charles University, Benátská 2, 128 01, Prague, Czech Republic.

2 Czech Academy of Sciences, Institute of Botany, Zámek 1, 252 43, Průhonice, Czech Republic.

3 Biodiversity Research Center and Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada

4 Biology Centre CAS, Branišovská 31, 37005, České Budějovice, Czech Republic

5 Department of Life Sciences, University of Nottingham, Nottingham NG7 2RD, UK.

Classification: Biological Sciences: Evolution

Keywords: adaptation, polyploidy, *Arabidopsis*, repeatability, trans-specific polymorphism

Abstract

Polyploidy, the doubling of an entire genome, is common in many organisms, including plants, yeast, and fish. While it can often be fatal due to the sudden transformation of DNA management and cell physiology, some lineages are able to adapt to this challenge. One proposed mechanism for this adaptation is the increased access to allelic variation in polyploids, reflecting both enhanced gene flow and higher rates of mutation accumulation. Here, we investigated whether this increased allelic diversity is utilized by polyploids to form mosaic adaptive haplotypes made up of diverse allelic sources. To do so, we studied two sister diploid/autotetraploid species (within-species polyploids, *Arabidopsis lyrata* and *Arabidopsis arenosa*). We first demonstrated the existence of two novel, independent autotetraploid lineages in these species, bringing the total number known to four. These lineages exhibited consistent changes in probable adaptive traits, as well as strong signatures of selection on the same set of 17 candidate genes involved in cell cycle, meiosis, and transcription. Interestingly, candidate adaptive haplotypes of these genes were completely absent in their natural diploid progenitors. Instead, our analysis of 995 individuals (2970 haploid genomes in 504 diploids and 491 autotetraploids) found that these alleles were made up of fine-scaled mosaics of variants from diverse evolutionary sources, including primarily reassortments of trans-specific polymorphism from diploids, novel mutations, and inter-species hybridization. We speculate that such flexibility in acquisition and re-shuffling of adaptive alleles enabled tetraploids to rapidly adapt to polyploidization, and may further promote their rapid adaptation to environmental challenges.

Significance statement

Polyploidy, the doubling of the genome, is a massive mutation. Despite this, it is found in many species, showing that some can rapidly adapt to this challenge. To fuel such adaptation, polyploids may access and maintain adaptive alleles more efficiently than diploids, acting as 'allelic sponges'. In two *Arabidopsis* plant species, we discovered four different polyploid lineages that showed repeated changes in likely adaptive traits and signatures of selection on the same set of genes. The candidate adaptive haplotypes for these genes were not present in their natural diploid ancestors, but instead were a mosaic assembled from diverse allelic sources. We speculate that such increased input of alleles helped polyploids to adapt to the process that caused this increase – polyploidy – and may also help them adapt to external environments.

Introduction

Whole genome duplication (WGD) is a dramatic mutation, widespread across eukaryotes, especially in plants. It comes with significant costs, such as meiotic instability and cell cycle changes, both of which require immediate adaptation (Doyle and Coate 2019; Bomblies 2020). Recent work has shown that within-species WGD lineages (autopolyploids) occasionally adapt to specific challenges involving meiotic crossover organization by evolutionary shifts in meiosis genes (Yant et al. 2013; Bray et al. 2020; Bohutínská, Handrick, et al. 2021; Bohutínská, Alston, et al. 2021). Additionally, cyclin genes have been observed to mediate tolerance to tetraploidization in WGD tumours and in proliferating *Arabidopsis* tissues (Imai et al. 2006; Sterken et al. 2012; Potapova et al. 2016; Crockford et al. 2017).

While some genes mediating adaptation to WGD have been revealed and their function has begun to be studied (Morgan et al. 2020; Morgan et al. 2022), we still lack an understanding of the exact sources of the underlying adaptive variation. From a genetic point of view, we know that polyploids may enjoy enhancement of particular characteristics to aid rapid adaptation. Across studies of adaptive alleles, primary sources include gene flow/hybridization, de novo mutation and ancestral standing variation (Fig. 1A). Recent work points to enhancement of all of these sources after WGD (Fig. 1B). First, polyploid lineages have reduced hybridization barriers (Marhold and Lihová 2006; Lafon-Placette et al. 2017; Schmickl and Yant 2021), leading to increased interspecific allelic exchange (Arnold et al. 2016; Marburger et al. 2019). Polyploids can also acquire additional diploid alleles via introgression from their diploid sisters, whereas introgression in the opposite direction is much less likely due to the nature of the diploid-tetraploid barrier (Te Beest et al. 2012; Kolář et al. 2017; Baduel et al. 2018; Monnahan et al. 2019; Morgan et al. 2021). Second, polyploids have an elevated mutation input, due to the doubled number of chromosomes in a population of a given census size, leading to more novel mutations per generation (Selmecki et al. 2015). Finally, polyploids can maintain allelic diversity more efficiently due to increased masking of recessive alleles (Ronfort 1999; Monnahan et al. 2019). As a consequence, polyploids may generate and maintain genetic variability at increased rates compared to their diploid progenitors (Fig. 1).

Beneficial alleles of a particular gene are most commonly envisioned as originating from a single allelic source (as shown in Figure 1 (Lee and Coop 2019)), such as introgression, standing variation, or de novo mutation (Oziolor et al. 2019; Bohutínská, Vlček, et al. 2021; Konečná et al. 2021; Wang et al. 2021). However, recent evidence suggests that adaptive polymorphisms in a genomic region can accumulate over time to form a finely tuned haplotype (McGregor et al. 2007; Archambeault et al. 2020; Seear et al. 2020; Roberts Kingman et al. 2021). This indicates that such adaptive haplotypes may be the result of multiple allelic sources, rather than just one. Here, we test if polyploids use their expanded allelic sources to construct particularly fine-scaled mosaic WGD-adaptive haplotypes.

The sister species *Arabidopsis arenosa* and *Arabidopsis lyrata* have emerged as premiere models for understanding adaptation to WGD as they naturally exhibit variation in ploidy (Yant and Bomblies 2017). Recent studies in *A. arenosa* and *A. lyrata* showed that both species adapt via a similar set of genes to WGD, conspicuously genes involved in meiosis (Yant et al. 2013; Marburger et al. 2019; Seear et al. 2020; Bohutínská, Handrick, et al. 2021). Yet, limited sampling in these studies had not permitted a comprehensive investigation of the evolutionary origin of adaptive alleles across all involved species,

lineages and loci. We thus lack an understanding of the sources of this remarkable adaptive variation.

While previous work revealed provocative indications that adaptive alleles, taken as entire evolutionary units, may have originated in one source population or another (Marburger et al. 2019), or that allelic chimerism may be involved in the construction of adaptive alleles (Seear et al. 2020), they suffered from limited sampling (52 (Seear et al. 2020) or 92 (Marburger et al. 2019) individuals from a fraction of each species range) and no consideration of trans-specific polymorphism across *Arabidopsis* species. Here, we overcome this by utilizing an exhaustive dataset of 995 sequenced individuals encompassing all known lineages of natural autotetraploids in European *Arabidopsis*, backed by genome-wide diversity of all diploid outcrossing *Arabidopsis* species. Apart from the tetraploid *A. lyrata* lineage from the eastern Austrian Forealps ('Austria' hereafter; (Schmickl and Koch 2011; Marburger et al. 2019)), we newly analysed two tetraploid lineages of *A. lyrata* from Central Europe: south-eastern Czechia ('Czechia' hereafter) and Harz in Germany ('Germany' hereafter). We performed a joint analysis of these three *A. lyrata* tetraploid lineages and tetraploid *A. arenosa*, which derived from a single WGD event (Arnold et al. 2015). Leveraging this exhaustive sampling of four independent polyploid lineages we ask: 1. What genes and cellular processes have been repeatedly under selection in independent tetraploid *Arabidopsis* lineages? 2. What are the sources of adaptive alleles in these lineages (introgression, standing variation, de novo mutations)? 3. Do adaptive tetraploid haplotypes arise from a fine-scale mosaic of multiple allelic sources?

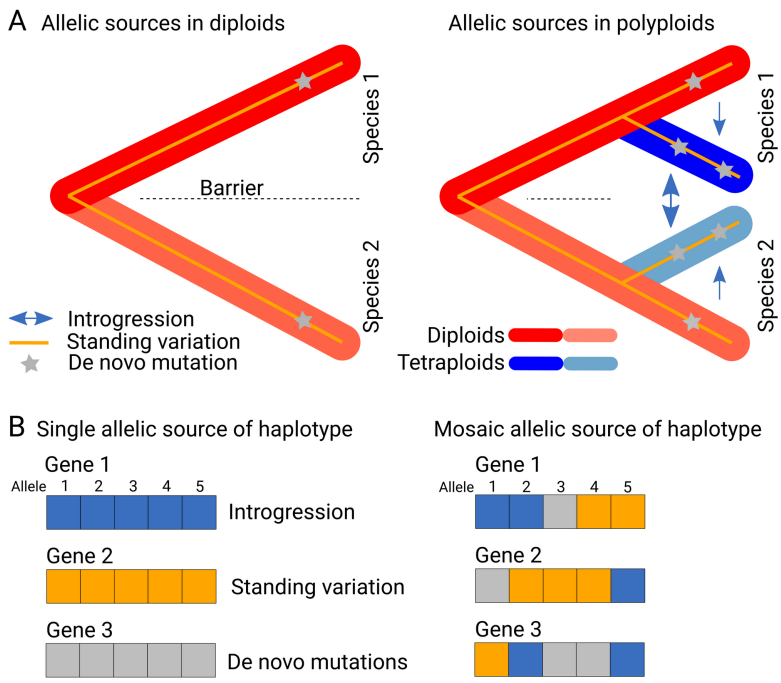

Fig. 1: *Hypotheses about sources of adaptive alleles in a diploid-polyploid system. A: As compared to diploids, polyploid lineages may acquire alleles via increased introgression potential, increased population-scaled mutation rate, and higher level of retainment of ancestral polymorphism, thus behaving as 'allelic sponges'. For the left scenario, we show two diploid species (red), each of which gave rise to an autotetraploid (blue) lineage (right scenario). B: Higher variability in the possible sources of adaptive alleles suggests that adaptive polyploid haplotypes might be more likely to form from a mosaic of allelic sources (right scenario), in contrast to the typically assumed homogeneous (single-source) scenario (left scenario).*

Results and discussion

Hybridization among tetraploid lineages of A. lyrata and A. arenosa

For this study, we compiled a collection of published and newly sequenced individual-level whole genome sequencing data of 823 individuals from 46 diploid and 64 tetraploid populations of *A. lyrata* and *A. arenosa,* as well as 172 individuals from outgroups for a total of 995. Our new sampling focused on Central European *A. lyrata* in order to cover all regions possibly harbouring the tetraploid cytotype, beyond the well-described Austrian contact zone (Schmickl and Koch 2011; Marburger et al. 2019; Seear et al. 2020) as well as representative sampling of all European diploid outcrossing *Arabidopsis* species. We determined ploidy using flow cytometry (Dataset S1) and observed that autotetraploids ('tetraploids' hereafter) occur at geographically distinct locations throughout the *A. lyrata* species range (Fig. 2 (Ansell et al. 2010)). To obtain comparable and representative samples of the Central European *A. lyrata* and *A. arenosa* populations for selection scans, we first analyzed 154 individuals from 17 proximal diploid and tetraploid populations (Fig. 2). First, we estimated the population structure and potential admixture of these populations (Fig. 2). Using bootstrapped allele covariance trees (TreeMix; Fig. 2B, C), neighbour-joining networks and Bayesian clustering (FastStructure; Fig. S1), we showed that tetraploids consistently cluster with nearby diploid populations within each species. We observed a single Central European tetraploid lineage in *A. arenosa,* in line with previous studies (Arnold et al. 2015; Monnahan et al. 2019). We newly found that *A. lyrata* formed three diploid-tetraploid lineages located in distinct regions – in Austria, Czechia, and Germany. This suggests either independent formation and establishment of tetraploid populations in each region or a single origin of autotetraploid *A. lyrata* followed by a vicariance event and, possibly, local introgression from parapatric diploid lineages. Both species maintained high genetic diversity (*A. arenosa*: mean nucleotide diversity over four-fold degenerate sites (4d-π) = 0.026 / 0.024 for diploids / tetraploids; *A. lyrata*: mean 4d-π = 0.012 / 0.016 for diploids / tetraploids, respectively; Table S1). Although Central European *A. lyrata* has a more scattered distribution and slightly lower nucleotide diversity than *A. arenosa*, it did not show signatures of bottlenecks (Tajima's D < 2; mean Tajima's D in *A. arenosa* = 0.01 / 0.20 for diploids / tetraploids; mean Tajima's D in *A. lyrata* = 0.27 / 0.23 for diploids / tetraploids; Table S1). Altogether, our analyses found three different tetraploid lineages in Central European *A. lyrata*, in addition to the previously well-described tetraploid lineage in its sister species *A. arenosa*.

Next, we tested for introgression between both species using D statistics (ABBA-BABA, (Martin et al. 2013)). While insignificant among diploids, we identified evidence of introgression (D between = 0.106 – 0.108; Fig. 2D) between tetraploids of *A. arenosa* and each of the three tetraploid *A. lyrata* lineages. These results are in line with previous experimental studies: while diploids of both species exhibit strong postzygotic barriers, polyploidy-mediated hybrid seed rescue enables hybridization between tetraploid *A. arenosa* and *A. lyrata* (Lafon-Placette et al. 2017). The strongest introgression was apparent in eastern Austria, the region reported as a hybrid zone between tetraploid *A. lyrata* and *A. arenosa* ((Schmickl and Koch 2011; Marburger et al. 2019), D = 0.108), followed by the newly analysed tetraploid lineages from Czechia and Germany.

In summary, we identified four independent tetraploid lineages in Central European *Arabidopsis*, consistent with two to four independent WGD events (one in each species or

additional WGDs events in *A. lyrata*). These WGD events reduced between-species hybridization barriers, leading to interspecific gene flow between tetraploids.

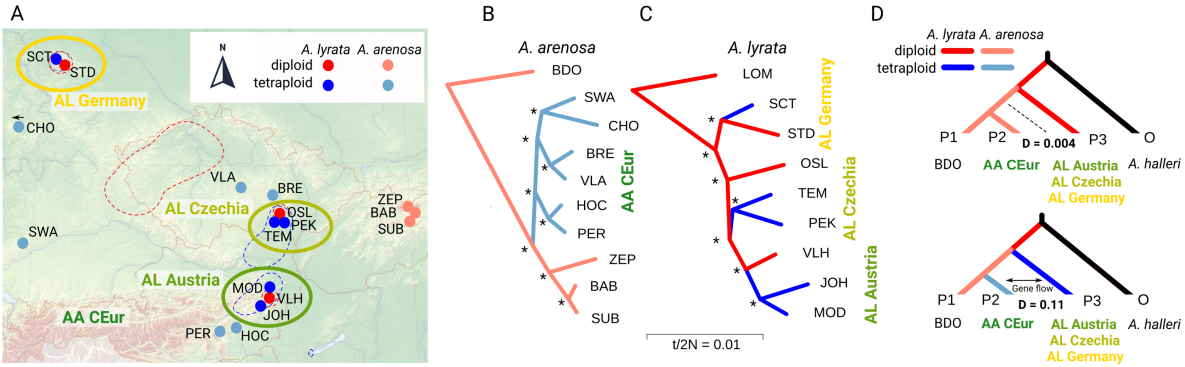

Fig. 2: *Evolutionary relationships of diploid and autotetraploid populations of A. lyrata and A. arenosa. A: Locations of 17 most proximal natural populations of diploid and tetraploid A. lyrata (AL Austria, AL Czechia, and AL Germany) and A. arenosa (AA CEur) in Central Europe. Red and blue dashed lines show the natural ranges of diploid and tetraploid A. lyrata, respectively; tetraploids of A. arenosa occur throughout the entire area. B, C: Phylogenetic relationships among the focal populations of A. arenosa (B) and A. lyrata (C) inferred by TreeMix analysis. A. arenosa is rooted with the diploid Pannonian population BDO and A. lyrata with the diploid Scandinavian population LOM. The x-axis shows the drift parameter, corresponding to the number of generations separating the lineages (t), and effective population size (N) (Pickrell and Pritchard 2012). Asterisks show bootstram support = 100. D: Introgression among tetraploid but not diploid populations of both* Arabidopsis *species. ABBA-BABA analysis demonstrating excess allele sharing between tetraploids (bottom tree), but not diploids (top tree), of A. arenosa and each of the three A. lyrata tetraploid lineages. P1 is BDO, the earliest diverging and spatially isolated diploid population of A. arenosa (Kolář, Fuxová, et al. 2016) and outgroup is Arabidopsis halleri from Austria.*

Genome-wide signatures of repeated adaptation to whole genome duplication

To identify genes exhibiting signals of positive selection repeatedly in different polyploid lineages, we compared our four pairs of diploid and tetraploid *A. lyrata* and *A. arenosa* lineages (Fig. 2B, C). We identified candidate genes using PicMin, a novel method based on order statistics that tests for signatures of repeated adaptation (Booker et al. 2022). We identified 59 significantly differentiated windows (PicMin FDR-corrected q-value < 0.1; 43 windows with $q < 0.05$; Dataset S2), overlapping 14 unlinked candidate genes (13 of them with $p < 0.05$; Fig. 3A, Table S2). Functional annotation of the 14 candidate genes by gene ontology (GO) and UniProt Keywords enrichment analyses revealed two candidate processes: meiosis (specifically homologous chromosome pairing and segregation) and cyclin-mediated control of the cell cycle and of endoreduplication (i.e. the replication of chromosomes without mitotic cell division; both represented by multiple enriched GO terms and UniProt Keywords; $p < 0.05$ for each; Table S3).

PicMin assumes that selection signatures affect entire genomic windows. Therefore, to cover narrower peaks of differentiation we also performed a dedicated search for genes harbouring highly differentiated SNPs found repeatedly in different lineages (see methods). Such a ‘candidate SNP’ approach is able to identify genes experiencing positive selection in tetraploids on a subset of SNPs while the evolution of the rest of the gene sequence is constrained. The results of the candidate SNPs approach confirmed all 14 PicMin candidate genes (Table S2). Additionally, it identified three candidate genes related to meiosis and cell cycle regulation (HEI10, SYN1 and SDS with three, three and two candidate SNPs

differentiated in all four tetraploid lineages, respectively; Fig. 3B, Dataset S3). Thus, in total, we identified 14 plus three candidate positively selected genes (PSGs), exhibiting strong signals of selection in the tetraploid lineages. Sixty five percent of the PSGs were identified in all four tetraploid lineages and 94% in at least three lineages, showing broadly shared differentiation signals among independent tetraploids (Fig. 3B, significant overlap of candidate SNPs gene lists, $p < 0.001$, Fisher's exact test; Dataset S3). We note that these loci do not cluster in regions with extreme values of recombination rate per gene, suggesting low bias due to the recombination landscape (Fig. S2 (Burri 2017)).

To refine our functional annotation of the 17 PSGs, we predicted protein-protein interactions among them using STRING analysis (Szklarczyk et al. 2015). In agreement with GO term and UniProt Keyword enrichment analyses, we retrieved two interconnected clusters: cell cycle regulation by cyclins, and chromosome pairing and segregation during meiosis ($p < 0.001$, Fisher's exact test; Fig. 3C). Despite the high overall functional connectivity, six of the 17 PSGs were not connected in this interaction network. Interestingly, after manual annotation using the TAIR database (Berardini et al. 2015) and literature searches, four of these six genes were found to be related to mRNA transcription via RNA polymerase II (Fig. 3C). The fifth tetraploid PSG (*AT4G18490*) encodes an unknown protein which is strongly expressed in young *Arabidopsis thaliana* flower buds (Klepikova et al. 2016) and was reported to be co-expressed with CYCA2;3 (STRING database, last accessed 12/21/2021). The sixth gene, AGC1.5, was not connected to any of the above-mentioned processes, but regulates pollen tube growth (Zhang et al. 2009). To summarise, the 17 tetraploid PSGs mediate processes of homologous chromosome pairing during prophase I of meiosis, cell cycle timing and regulation of endoreduplication via different classes of cyclins, and mRNA transcription via RNA polymerase II. We provide further functional interpretation in Supplementary Text 1 and Fig. S3.

Phenotypic shifts in polyploids have already been associated with some of these functions. Specifically, cytological studies reported stable meiotic chromosome segregation in established tetraploids of *A. arenosa* and *A. lyrata*, compared to neo-tetraploids of *A. arenosa*, which suggests a compensatory shift in the meiotic stability phenotype (Yant et al. 2013; Marburger et al. 2019; Morgan et al. 2020; Seear et al. 2020; Morgan et al. 2022). Here, we extended the inquiry of polyploidy-related trait shifts to cell cycle phenotypes, asking if established tetraploids reduce their level of endoreduplication, as reported in *A. thaliana* and alpine lineages of *A. arenosa* (Imai et al. 2006; Sterken et al. 2012; Wos et al. 2022). To do so, we grew plants from each of the four established tetraploid lineages, their closely related diploids, and synthetic (oryzalin-induced) neo-tetraploids. We found that the level of endoreduplication (the fraction of endoreduplicated nuclei) and the number of endoreduplication cycles were consistently higher in diploids compared to tetraploids (Fig. 3D). Notably, established tetraploids showed reduced endoreduplication compared to neo-tetraploids. Thus, the decrease of endoreduplication observed in tetraploids is likely due to post-WGD adaptation, rather than a mere phenotypic consequence of WGD itself. We found this decrease of tetraploid endoreduplication in each of the four lineages (Fig. 3D, Fig. S4 and p-values herein), and also when comparing all diploids with all neo-tetraploids and established tetraploids (mean level of endoreduplication = 0.82 / 0.74 / 0.61 for diploids / neo-tetraploids / tetraploids, $p < 0.001$; median maximum number of endoreduplication cycles = 3 / 2.5 / 2 for diploids / neo-tetraploids / tetraploids, $p < 0.001$; Wilcoxon rank sum test). In diploids we observed a nonsignificant difference in the number of endoreduplication

cycles (median = 3 in all four diploid lineages, p > 0.05, Wilcoxon rank sum test) and only a subtle difference in the level of endoreduplication in one lineage (Fig. S5). This suggests that the regulation of endoreduplication in *Arabidopsis* is generally conserved, as long as the intrinsic cell conditions do not change, for example as a result of WGD.

In summary, these results point to an adaptive compensation for DNA content per nucleus: while nearly double in neo-tetraploids, it returns toward diploid levels in established tetraploids. This suggests 'compensation' over 'trait change' scenarios, as have been suggested for other polyploidy-specific traits following WGD (Bomblies 2020). Importantly, some of our strongest PSGs are associated with endoreduplication level regulation in *A. thaliana* or genome doubled tumours (genes CYCA2;3 (Imai et al. 2006), CYCD5;1 (Sterken et al. 2012)). Thus, these PSGs are possibly responsible for the reduction of endoreduplication in tetraploid *A. arenosa* and *A. lyrata*.

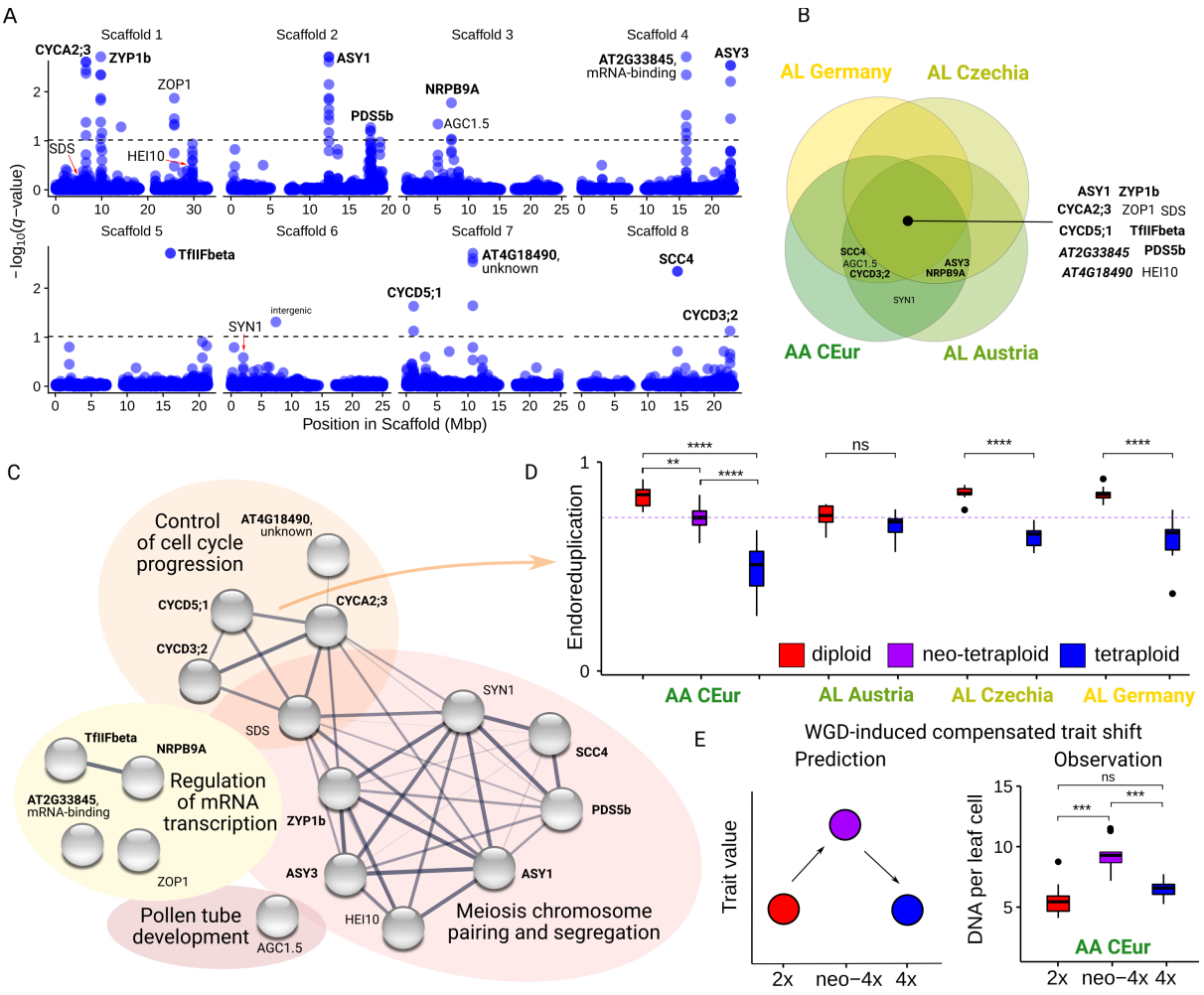

Fig. 3: *Genomic signatures of repeated positive selection in tetraploids of A. lyrata and A. arenosa. A: Using PicMin, we identified a set of 14 unlinked genes showing significant evidence of positive selection in tetraploids (positively selected genes, PSGs). Additional three genes, SDS, HEI10 and SYN1, were identified using a screen for candidate SNPs (see text and Dataset S3). B: Venn diagramme showing the tetraploid lineages for which PSGs were identified, supporting a high degree of repeated differentiation. C: Functional characterisation of PSGs. The network shows predicted protein-protein interactions among the 17 PSGs. The width of each line corresponds to the confidence of the interaction prediction. PSGs were*

*annotated into four processes, each represented by a bubble of different colour. The 12 PSGs with names written in bold had a sufficient number of candidate SNPs for the reconstruction of tetraploid haplotypes (see the main text). D, E: Phenotypic shift associated with the establishment of tetraploids of A. arenosa and A. lyrata in the form of a decrease in the level of DNA endoreduplication. D: Proportion of endoreduplicated nuclei in leaves of 10 individuals per each lineage and ploidy. **: $p < 0.01$, ****: $p < 0.0001$, ns: nonsignificant, Wilcoxon rank sum test. The horizontal violet line shows the level of endoreduplication under a scenario of no post-WGD adaptation (i.e. estimated from values of synthetic neo-tetraploids of A. arenosa). E: Prediction for the compensatory evolution of polyploid traits from (Bomblies 2020) and observed average DNA content per leaf cell (calculated as a mean number of homologous chromosomes per nucleus) for diploid, neo-tetraploid and tetraploid A. arenosa.*

Mosaic tetraploid haplotypes are assembled from variants from different sources

Recent evidence suggests that polyploids may access broader sources of adaptive alleles than diploids (Ronfort 1999; Selmecki et al. 2015; Arnold et al. 2016; Monnahan et al. 2019). We speculated that if these sources vary over fine genomic scales, the resultant polyploid haplotypes may represent a mosaic of different allelic sources (Fig. 1B). Some empirical support for this was observed in an earlier study, where chimerism of tetraploid adaptive alleles was observed in a sampling of 52 individuals(Seear et al. 2020).

We used our dataset of 2,970 sampled haploid *Arabidopsis* genomes (including exhaustive sampling of all known diploid lineages; Fig. 4A, Dataset S1) to deconstruct these haplotypes and identify the specific allelic sources for every candidate tetraploid SNP in our PSGs. We specifically focused on PSGs that contained more than five candidate SNPs, analysing 12 out of the 17 genes and 232 candidate SNPs (Dataset S4, S6). We reconstructed major diploid and tetraploid ‘candidate haplotypes’ (haplotypes hereafter) for each species using these SNPs as haplotype markers (following (Bohutínská, Handrick, et al. 2021)). We validated the haplotypes using PacBio HiFi assemblies from ten diploid and tetraploid individuals of *A. arenosa* (Dataset S4, S5). We identified a single major tetraploid haplotype for each of the 12 candidate genes in the majority of tetraploid populations of both *A. lyrata* and *A. arenosa* (91-98%; Table S4). Based on estimates of the age of the tetraploids, which range from approximately 20,000 to 230,000 generations (Arnold et al. 2015; Marburger et al. 2019), this suggests that the tetraploid haplotype quickly spread throughout Europe. We further identified two major diploid haplotypes for each gene, one specific to *A. lyrata* and one specific to *A. arenosa* (Fig. 4B). Other haplotypes were of minor frequency (Fig. 4B).

The tetraploid haplotypes for the 12 candidate genes were found at high frequency in tetraploid populations of both species, suggesting that they have undergone a selective sweep (mean frequency = 0.62 across 64 populations; Fig. 4B, Table S4). Interestingly, none of these 12 tetraploid haplotypes were present in diploids, despite exhaustive sampling of diploid lineages of both species in Europe. This suggests that they may have assembled upon the establishment of tetraploids rather than preexisting as standing variation in diploids.

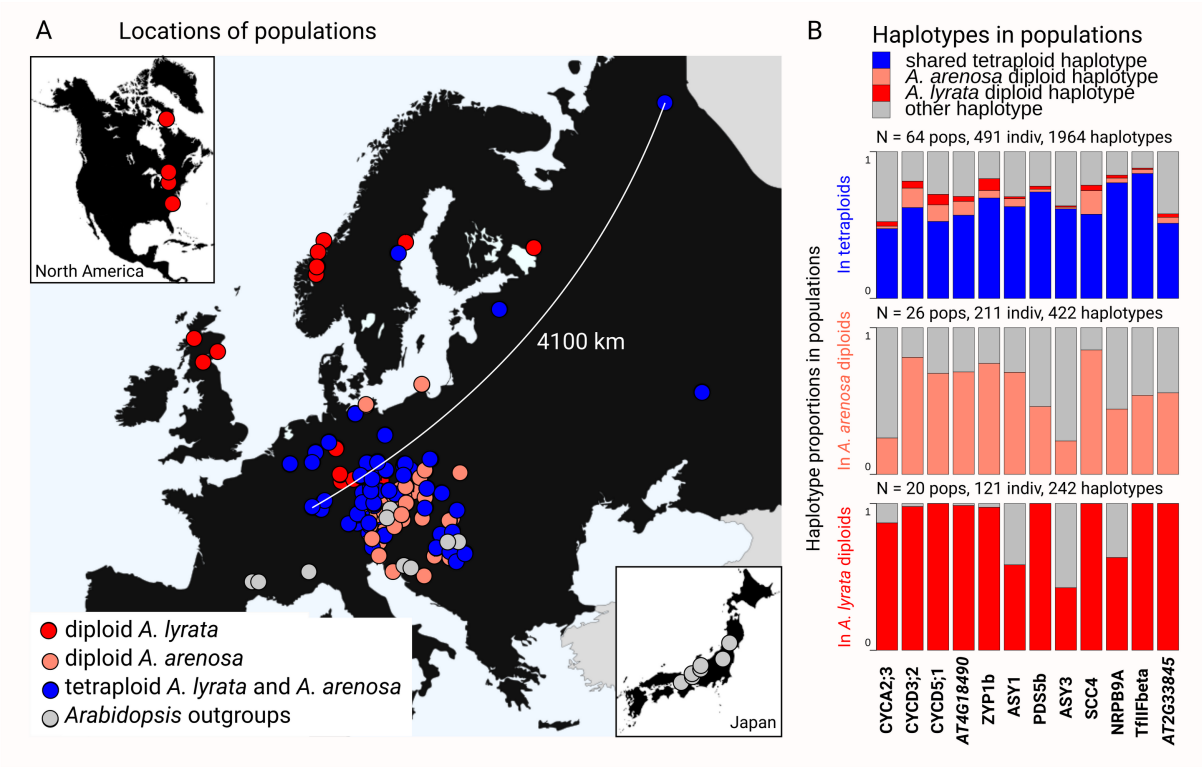

Fig. 4: Distribution of candidate haplotypes in diploids and tetraploids of A. lyrata and A. arenosa. A: Geographic distribution of populations used in this analysis (504 diploids and 491 tetraploids). Arabidopsis outgroups are A. halleri, A. croatica, A. cebennensis and A. pedemontana. B: Frequency of shared tetraploid (blue), A. arenosa (light red), and A. lyrata diploid (red) haplotypes in each of the 12 candidate genes. All other haplotypes present in the populations are highlighted in grey.

While every complete tetraploid haplotype was missing in diploids, the sources of constitutive SNPs that made up these haplotypes were much more complex (Fig. 5). We therefore investigated the possibility that the tetraploid haplotypes might be assembled from multiple allelic sources. To do this, we determined the likely source for each candidate SNP forming these haplotypes. Considering the extent of adaptive trans-specific polymorphism in *Arabidopsis* (i.e. alleles found in diploids of multiple *Arabidopsis* species (Novikova et al. 2016)), we combined our full dataset of *A. arenosa* and *A. lyrata* with all other *Arabidopsis* outcrossers (*A. halleri*, *A. croatica*, *A. cebennensis* and *A. pedemontana*) (Dataset S1). Using the phylogenetic relationships among these species (Novikova et al. 2018) and the presence/absence information about each SNP across all 995 individuals, we determined the most likely allelic source of each of the 232 haplotype-marking SNPs (Fig. 5, Dataset S6). Surprisingly, 65.5% of the SNPs forming tetraploid haplotypes were sourced from trans-specific polymorphism, segregating in diploids of multiple *Arabidopsis* species (Fig. 5A, scenarios 1-4). This supports the growing recognition of the role of trans-specific standing polymorphism in adaptation (Guggisberg et al. 2018; Marques et al. 2019)). Only 6.9% of the candidate SNPs came as standing variation from a single diploid progenitor (*A. arenosa* or *A. lyrata*; Fig. 5A, scenarios 5, 6), and the remaining 27.6% of SNPs possibly accumulated de novo in tetraploids or remained unsampled in our dataset (absent in any of the 504 diploid individuals; Fig. 5A, scenario 7). We further quantified that 35.8% of the SNPs forming the tetraploid haplotypes were likely contributed from diploid *A. arenosa* (present in

this species and potentially diploids of other species, but not *A. lyrata*; Fig. 5A, scenarios 2, 5), while 3.4% likely came from diploid *A. lyrata* (present in this species and potentially diploids of other species, but not *A. arenosa*; Fig. 5A, scenarios 3, 6). Finally, during the accumulation of SNPs in tetraploids via these source scenarios, 71.1% of SNPs were likely shared among all four tetraploid lineages through interspecific hybridization (note that this scenario is nested with others; Fig. 5A, scenarios 2-7). This was suggested by their presence in tetraploids and maximum one of the diploid progenitors (diploid *A. arenosa* or *A. lyrata*) and genome-wide signatures of hybridization among tetraploids (Fig. 2C, (Marburger et al. 2019)).

The different allelic sources forming tetraploid haplotypes varied genome-wide (when analysing all 12 PSGs together; Fig. 5A), but also within individual tetraploid haplotypes (3 – 6 scenarios per haplotype, median = 4; out of 7; Fig. 5B, C, Fig. S6). This suggests that tetraploid haplotypes represent a combination of reuse of species-specific as well as trans-specific SNPs, with additional input of de novo mutations after WGD (Fig. S6). Such composite 'mosaic' haplotypes were likely spread among tetraploid lineages by introgression. The rearrangement of diverse allelic sources into mosaic tetraploid haplotypes may be facilitated by high recombination rates reported for these *Arabidopsis* species (Hämälä and Savolainen 2019; Dukić and Bomblies 2022) and the observation of enhanced recombination rates in neo-tetraploid *A. arenosa* and *A. thaliana* (Pecinka et al. 2011; Morgan et al. 2021). The mosaic scenario was further supported by reticulation in neighbour-joining networks of the 12 PSGs (Fig. S7), in which tetraploids form a single lineage that is connected by multiple splits to two or more diploid *Arabidopsis* lineages. We provide hypotheses about the spatio-temporal context of the origin of these mosaic haplotypes in Supplementary Text 2.

To conclude, our screen for variants forming tetraploid haplotypes suggests that polyploids source alleles from multiple pools of diversity, combine them during post-WGD adaptation, and further redistribute this variation by gene flow among polyploid lineages. Our findings indicate that interspecific introgression plays a role in shaping genome-wide adaptive variability in polyploid species complexes, a complementary evidence to earlier studies (reviewed in (Marhold and Lihová 2006; Schmickl and Yant 2021)).

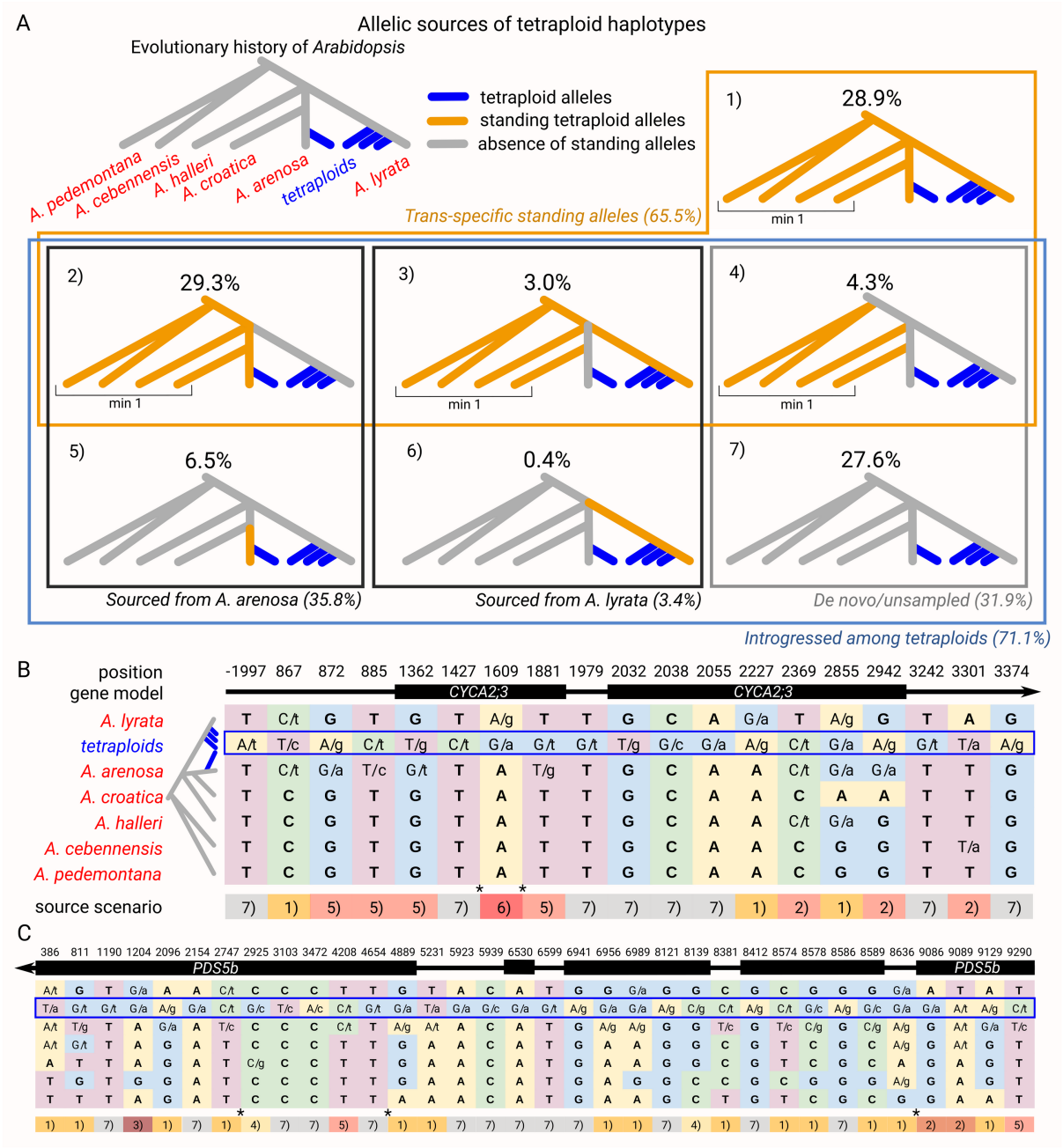

Fig. 5: Mosaic allelic sources of tetraploid haplotypes. A: Candidate SNPs forming tetraploid haplotypes were categorized into one of the seven source scenarios based on their distribution in the 996 samples. The most parsimonious origin of each pattern is provided in italics. Four of them (orange background) involve trans-specific standing variation shared among diploids of at least two Arabidopsis species. Six of them (outlined by a blue box) require introgression between tetraploid lineages. Boxes 'min 1' show that the tetraploid standing variation is present in at least one outgroup species. Phylogenetic relationships according to (Novikova et al. 2018). B, C: Example alignment of inferred diploid and tetraploid haplotypes. B: Endoreduplication gene CYCA2;3. Shown are the 19 candidate SNPs used as markers to detect the haplotype (distributed over 5371 bp). Alleles defining the haplotype are shown in capitals, two letters at a position mark the presence of an alternative minor allele. Upper lines give position relative to the transcription start site and gene model, with coding sequences shown as boxes (only elements overlapping with candidate SNPs are depicted). Bottom line shows SNP assignment into its source scenario from panel A. Possible recombination breakpoints needed to assemble the haplotype from the mosaic of sources are shown as asterisks. C: Meiosis gene PDS5b. Shown are candidate SNPs marking the 8904 bp-long haplotype. Description same as in panel B.

Conclusions

Here, we focused on the role of evolutionary scenarios governing the availability of adaptive variation, leveraging our exhaustive sampling of extant outcrossing *Arabidopsis* species in Europe. Through selection scans of four independent polyploid lineages, we identified a set of genes involved in cell cycle, meiosis, and transcription that were repeatedly subjected to selection. We found that for all these candidate genes, all four tetraploid lineages of both species brought the same haplotypes to fixation. By deconstructing the evolutionary origins of the independent SNPs forming these haplotypes, we revealed that each haplotype reflects a fine-scale mosaic of allelic sources: trans-specific and species-specific diploid polymorphisms as well as likely de novo mutations in tetraploids. Once formed in the young tetraploids, these 'evolutionary mosaic' haplotypes were shared by introgression among the tetraploids, likely further re-assembled and spread across Europe.

In sum, we demonstrated that widespread re-shuffling of trans-specific, species-specific and novel variation was rapidly leveraged in response to severe intracellular challenges accompanying WGD. Our case illustrates how polyploid lineages can use their increased capacity to accumulate genetic variation from different sources to create new adaptive haplotypes. This haplotype assembly process, however, is not specific for polyploids, as accumulation of multiple adaptive changes into finely tuned haplotypes was also reported for diploids (McGregor et al. 2007; Archambeault et al. 2020; Roberts Kingman et al. 2021). Yet, the tetraploid *Arabidopsis* case is exceptional by its magnitude, as we found mosaic haplotypes for an entire suite of interacting candidate adaptive genes. Recent research has already shown that there is a significant amount of variation in the contribution of standing, introgressed, and de novo allelic sources to repeated adaptation at the level of entire genes in various plant and animal species (Bohutínská, Vlček, et al. 2021; Konečná et al. 2021; Wang et al. 2021; Moran et al. 2022). The magnitude of the mosaic haplotypes found in tetraploid *Arabidopsis* system suggests that there may be even more variation to be discovered.

Methods

Sampling

The genus *Arabidopsis*, including the most researched plant model *A. thaliana* and multiple other model species, is primarily diploid. However, autotetraploids were found in *A. arenosa* and *A. lyrata* at several places throughout their natural range in Central Europe. While diploids and tetraploids of *A. arenosa* are widespread and form large populations, *A. lyrata* occupies a narrow, refugial ecological niche in Central Europe (Schmickl and Koch 2011). It was proposed that introgression between tetraploids of *A. arenosa* and *A. lyrata* enabled tetraploids of *A. lyrata* to increase the genetic differentiation from their diploid progenitors and escape this narrow niche (Schmickl and Koch 2011; Marburger et al. 2019).

Here, we sampled and re-sequenced genomes of diploid as well as adjacent tetraploid populations from all known diploid-tetraploid lineages in Central European *Arabidopsis*. In total, we sequenced genomes of 73 diploid and tetraploid individuals of both species and complemented them with additional 750 published whole genome sequences of these species and 172 outgroup species, totaling 995 individuals and 132 populations (Dataset S1, Dataset S7). Ploidy of each sequenced individual was checked using flow cytometry following (Kolář, Lučanová, et al. 2016).

Population genetic structure

Samples were sequenced on Illumina HiSeq X Ten, mapped to the reference genome of *A. lyrata* (Hu et al. 2011), and processed following (Monnahan et al. 2019; Bohutínská, Vlček, et al. 2021).

We inferred relationships between populations using allele frequency covariance graphs implemented in TreeMix v. 1.13 (Pickrell and Pritchard 2012). We ran TreeMix allowing a range of migration events and presented two additional migration edges for gene flow between *A. arenosa* and *A. lyrata*, as they represented points of log-likelihood saturation. To obtain confidence in the reconstructed topology of *A. arenosa* and *A. lyrata* trees (Fig. 1B, C), we bootstrapped the trees, choosing a bootstrap block size of 1000 bp, equivalent to the window size in our selection scans (see below), and 100 replicates. Further, we used the model-based method FastStructure (Raj et al. 2014). We randomly sampled two alleles per tetraploid individual, using a custom script. This approach does not appear to bias clustering in autotetraploid samples based on (Stift et al. 2019). Finally, we displayed genetic relatedness among individuals using principal component analysis (PCA) as implemented in adegenet (Jombart 2008). We calculated genome-wide four-fold degenerate (4d) within-population metrics (nucleotide diversity (π) and Tajima's D (Tajima 1989)) using the python3 ScanTools_ProtEvol pipeline (available at github.com/mbohutinska/ScanTools_ProtEvol, (Bohutínská, Vlček, et al. 2021)).

To test for introgression, we used the ABBA-BABA test as described in (Martin et al. 2013) and in an associated pipeline (available at github.com/simonhmartin/tutorials/tree/master/ABBA_BABA_whole_genome).

Only four-fold degenerate sites genotyped by an individual read depth >8 were considered, and sites genotyped in fewer than 80% of individuals were excluded, resulting in a dataset of ~75,000 genome-wide SNPs. We tested for introgression from tetraploid *A. arenosa* to each *A. lyrata* tetraploid lineage separately. For P1 (the putative non-admixed population), we used the early diverging and spatially isolated diploid Pannonian lineage of

A. arenosa (Kolář, Fuxová, et al. 2016), population BDO. Allele frequencies were polarised using the diploid outgroup *A. halleri*, population GUN.

Genome-wide scans for positive selection

To identify candidate genes that underwent positive selection in tetraploids (positively selected genes, PSGs), we used two selection scan methods based on population allele frequencies and applicable to both diploid and autopolyploid populations (Bohutínská, Vlček, et al. 2021; Booker et al. 2022).

First, we used the PicMin software (available at github.com/TBooker/PicMin) to analyze population genomic data for all four tetraploid lineages, calculating Fst (Hudson et al. 1992) between diploid and tetraploid population pairs in 1 kb windows across the genome. We performed PicMin on all windows that had data for at least three lineages (13,579 windows in total). We then applied a genome-wide false discovery rate correction to the resulting p-values, with α_{Adapt} set to 0.05, considering windows with FDR-corrected p-values of 0.05 and 0.1 as significant. If a selection sweep spanned multiple adjacent windows, we retained only the window with the lowest p-value and highest overall Fst.

Second, we used a ‘candidate SNP’ approach. We calculated SNP-based Fst between diploid and tetraploid population pairs within each lineage, and used the 1% outlier threshold to identify the most ploidy-differentiated SNPs genome-wide. Then we calculated the density of such candidate SNPs per gene, using *A. lyrata* gene models (Rawat et al. 2015). We identified the upper quartile of genes with the highest density of outlier SNPs as candidate genes, and used Fisher's exact test (SuperExactTest (Wang et al. 2015) package in R) to identify repeatedly selected candidates (assigned to Venn diagram in Fig. 3B) and to test for a higher than random number of such genes.

All genes identified with PicMin were also identified using the candidate SNP approach and showed distinct peaks of increased differentiation in the genome. Further, they were not biased towards regions with low recombination rate, as estimated based on the available *A. lyrata* recombination map (Hämälä et al. 2018). This corresponds well with outcrossing in both species and high nucleotide diversity that aids adaptive gene detection (Yant and Bomblies 2017).

Functional annotation

To infer functions significantly associated with tetraploid PSGs, we performed gene ontology (GO) and UniProt Keywords enrichment analyses using the STRING database (https://string-db.org/, last accessed 12/21/2021, (Szklarczyk et al. 2015)). We used *A. thaliana* orthologs of *A. lyrata* genes. Only categories with FDR < 0.05 were considered. We also manually searched for functional descriptions of each gene using the TAIR database and associated literature (Berardini et al. 2015). Finally, to identify potential protein-protein interactions among our 17 candidate genes, we used the "multiple proteins" search in the STRING database (Szklarczyk et al. 2015), including text mining, experiments, databases, co-expression, neighbourhood, gene fusion, and co-occurrence as information sources. We retained only 1st shell interactions (proteins directly associated with the candidate protein).

Measuring endoreduplication: plant material and growth conditions

To investigate whether genetic changes in cyclin genes may lead to a reduction in the level of endoreduplication, we grew 10 plants from each of the four lineages , from a diploid

and tetraploid population each per lineage (totaling 80 plants), and we cultivated 9 oryzalin-induced neo-tetraploid plants of *A. arenosa*, generated from the Western Carpathian diploid lineages which was shown to be a direct progenitor of tetraploids (Arnold et al. 2015). We determined the endoreduplication profile of these plants using flow cytometric analysis on nuclei isolated from fully grown leaves (11th leaf in the rosette, (Wos et al. 2022)).

Specifically, we collected seeds from both diploid and tetraploid populations of *A. lyrata* and *A. arenosa* corresponding to their four tetraploid lineages and stored them at 4°C for vernalization. We germinated the seeds on moist filter paper in petri dishes. For induction of synthetic polyploids, we applied a drop of oryzalin solution (5% aqueous DMSO) to the shoot apical meristems of young seedlings at the stage of fully developed cotyledons. After 48 hours, we rinsed the seedlings with distilled water and transplanted them to soil, cultivating them under controlled conditions (20°C/16h day, 15°C/8h night).

Measuring endoreduplication: flow cytometry

To estimate the success rate of the oryzalin treatment, the ploidy level of each plant was tested by flow cytometry following (Doležel et al. 2007). Specifically, 0.5 cm^2 of leaf tissue (without middle vein) was chopped, using a razor blade, in Otto I buffer simultaneously with the same amount of *Carex acutiformis* as internal standard. Samples were then dyed with a staining solution consisting of Otto II buffer, DAPI (4',6-diamidino-2-phenylindole) and β-mercaptoethanol, and analysed on a CytoFLEX S flow cytometer (Beckman Coulter, Inc) equipped with a UV laser (375 nm, 60 mW) using the Beckman Coulter CytExpert Acquisition and Analysis Software v. 2.5. Only the plants with no or significantly low percentage (<5% of all events) of diploid nuclei (also showing no distinguishable population of nuclei) were considered good candidates for synthetic polyploids, and nine of them were chosen for further analysis.

To assess the level of endopolyploidy, the amount of DNA was estimated using a Partec CyFlow SL flow cytometer equipped with a 532 nm solid-state laser (Cobolt Samba 150 mW). Samples were processed using the same simplified two-step method with Otto buffers and stained with propidium iodide. Samples were analysed without internal standard. Resulting histograms/dot plots were evaluated using Partec FloMax v. 2.4d.

We calculated the level of endoreduplication as the number of endoreduplicated nuclei compared to the number of all nuclei in the analysis. We also quantified the maximum number of endoreduplication cycles which nuclei of each leaf underwent, considering only peaks supported by at least five nuclei of a given genome size. Finally, we calculated the endoreduplication intensity, measured as the weighted number of endoreduplication cycles per nucleus [EI = (0 × %2C)+ (1 × %4C)+ (2 × %8C)+ (3 × %16C)+ (4 × %32C)] (Sterken et al. 2012). We tested for the difference between diploid, neo-tetraploid and tetraploid endoreduplication levels and numbers using the Wilcoxon test in R, package stats.

Haplotype analysis

We searched for the presence of diploid and tetraploid haplotypes in *A. lyrata* and *A. arenosa* populations using a procedure from (Bohutínská, Handrick, et al. 2021). Briefly, this involved reconstructing lineage-specific haplotypes and their allele frequencies for a set of linked candidate SNPs within each tetraploid PSG (subsetted to genes with five or more candidate SNPs; 12 genes with 232 candidate SNPs in total). For each PSG, with *n* candidate SNPs in the dataset of 211 / 121 diploids of *A. arenosa* and *A. lyrata*, respectively,

and 491 tetraploids, we defined *Mi* to be the major allele frequency at the candidate SNP *i*. Given that the sample consists of 1964 tetraploid haplotypes and 422 / 242 diploid haplotypes, this major allele frequency corresponds to the tetraploid allele, thus we define the derived (i.e. tetraploid) haplotype frequency as minimum over *n* allele frequencies (HAFd=min{Mi}), and consequently, we define the ancestral (i.e. either *A. arenosa* or *A. lyrata* diploid) haplotype allele frequency as HAFa=1 – max{Mi}. We further define the frequency of all other haplotypes, which result from recombination of the two previous, as HAFr=1-HAFa-HAFd. All calculations were performed using our in-house R script (available at github.com/mbohutinska/repeatedWGD, section "Haplotype AF").

We used this allele frequency-based approach because standard phasing procedures were not reliable with our short read data and tetraploid samples (Kyriakidou et al. 2018). We also verified our haplotype analysis using long read assemblies from five diploid and five tetraploid samples. Diploid and tetraploid haplotypes assembled from long reads were consistent with the diploid and tetraploid haplotypes combined based on allele frequencies at candidate SNPs (Dataset S4, S5).

Allelic sources of tetraploid haplotypes

To determine whether the tetraploid haplotypes in our study originated from a mosaic of allelic sources or from a single source, we used two approaches: (i) reconstructing the allelic sources of tetraploid SNPs across the ancestral diploid lineages (Novikova et al. 2016) and (ii) reconstructing networks of genetic distances at the PSG regions (Jombart 2008).

First, to identify the allelic sources of tetraploid haplotypes, we searched for tetraploid variants among diploid individuals. We worked with a complete genetic dataset of *A. lyrata* and *A. arenosa* (332 diploid individuals) and their outcrossing relatives *A. halleri*, *A. croatica*, *A. cebennensis* and *A. pedemontana* (‘diploid outgroups’ hereafter, 172 individuals; Dataset S1). We filtered out singletons, i.e. the variants which occurred only once in the species, which could be potential sequencing errors (note that results did not change significantly when we included singletons in the analysis; Fig. S6). Then we used the phylogenetic relationships among these species (Fig. 5A, (Novikova et al. 2018)) and the presence/absence information about each SNP across all 504 diploid individuals to determine the most likely allelic source of each of the 232 candidate SNPs. Specifically, we distinguished between seven different scenarios (Fig. 5A): 1) trans-specific polymorphism in both *A. lyrata* and *A. arenosa*, 2) trans-specific polymorphism in *A. arenosa* only (tetraploid SNP occurs in *A. arenosa* but not *A. lyrata* diploids and in one to all diploid outgroups), 3) trans-specific polymorphism in *A. lyrata* only (tetraploid SNP occurs in *A. lyrata* but not *A. arenosa* diploids and in one to all diploid outgroups), 4) trans-specific polymorphism in the outgroup species only, 5) *A. arenosa*-specific polymorphism, 6) *A. lyrata*-specific polymorphism, 7) de novo origin in tetraploids/unsampled in diploids. It is noteworthy that a rarefaction analysis in (Bohutínská, Handrick, et al. 2021) suggests that as little as 40 diploid individuals sampled across the *A. arenosa* species range is enough to cover most of its diploid diversity. Thus, our comprehensive dataset of 504 diploid individuals (211 of *A. arenosa*, 121 of *A. lyrata*, and 172 of outgroups) should be sufficient to estimate the complete natural diversity of diploids. Finally, here we distinguish between trans-specific and species-specific variability at the diploid level, which reflect ancestral allele sharing rather than recent introgression. Still, it is important to note that, ultimately, any tetraploid SNP

identified here represents trans-specific polymorphism, as it is shared between tetraploids of two species, *A. arenosa* and *A. lyrata*.

As an independent evidence for the evlutionary history of tetraploid haplotypes, we also reconstructed networks based on genetic distance (Nei's distance, (Nei 1972)) using the adegenet package for R, and we displayed the resulting neighbour joining networks using SplitsTree.

Acknowledgements:

We are grateful for the support and comments from the Ecolgen team in Prague and the inspiration provided by the Polyploid course in Drobak, Norway, organized by ForBio. We greatly appreciate the constructive feedback from Katie Peichel and Reto Burri, as well as the discussions with Alison Scott and Polina Novikova. We also thank Martin Čertner and Dorka Čertnerová for sharing their synthetic polyploid protocol, Bodo Schwarzberg (Untere Naturschutzbehörde Nordhausen) for collecting seeds of the STD population, and Bertram Preuschhof (Untere Naturschutzbehörde Göttingen) for a collecting permit for the SCT population. This work was supported by the Czech Science Foundation (project 20-22783S to F.K., project 19-06632S to K.M.), Leverhulme Trust award (no. RPG-2020-367 to L.Y.), the PRIMUS Research Programme of Charles University (PRIMUS/17/SCI/23 to R.S.), the European Union's research and innovation programme under the Marie Skłodowska-Curie (project 101062703 to M.B.) and long-term research development project No. RVO 67985939 of the Czech Academy of Sciences. Sequencing was performed by the Norwegian Sequencing Centre, University of Oslo. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, under the program Projects of Large Research, Development, and Innovations Infrastructures.

References

Ansell SW, Stenøien HK, Grundmann M, Schneider H, Hemp A, Bauer N, Russell SJ, Vogel JC. 2010. Population structure and historical biogeography of European Arabidopsis lyrata. Hered. 2010 1056 [Internet] 105:543–553. Available from: https://www.nature.com/articles/hdy201010

Archambeault SL, Bärtschi LR, Merminod AD, Peichel CL. 2020. Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback *Eda* locus. Evol. Lett. [Internet] 4:282–301. Available from: https://onlinelibrary.wiley.com/doi/10.1002/evl3.175

Arnold B, Kim S-T, Bomblies K. 2015. Single Geographic Origin of a Widespread Autotetraploid Arabidopsis arenosa Lineage Followed by Interploidy Admixture. Mol. Biol. Evol. [Internet] 32:1382–1395. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25862142

Arnold BJ, Lahner B, DaCosta JM, Weisman CM, Hollister JD, Salt DE, Bomblies K, Yant L. 2016. Borrowed alleles and convergence in serpentine adaptation. Proc. Natl. Acad. Sci. U. S. A. 113:8320–8325.

Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. 2018. The “Polyploid Hop”: Shifting challenges and opportunities over the evolutionary lifespan of genome duplications. Front. Ecol. Evol. 6:117.

Te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P. 2012. The more the better? The role of polyploidy in facilitating plant invasions. Ann. Bot. [Internet] 109:19–45. Available from: https://academic.oup.com/aob/article/109/1/19/154024

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E. 2015. The Arabidopsis Information Resource: Making and Mining the “Gold Standard” Annotated Reference Plant Genome. genesis [Internet] 53:474–485. Available from: https://www.arabidopsis.

Bohutínská M, Alston M, Monnahan P, Mandáková T, Bray S, Paajanen P, Kolář F, Yant L. 2021. Novelty and convergence in adaptation to whole genome duplication.Purugganan M, editor. Mol. Biol. Evol. [Internet]. Available from: https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab096/6203814

Bohutínská M, Handrick V, Yant L, Schmickl R, Kolář F, Bomblies K, Paajanen P. 2021. *De-novo* mutation and rapid protein (co-)evolution during meiotic adaptation in *Arabidopsis arenosa*.Lu J, editor. Mol. Biol. Evol. [Internet]. Available from: https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab001/6120800

Bohutínská M, Vlček J, Yair S, Laenen B, Konečná V, Fracassetti M, Slotte T, Kolár F. 2021. Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. Proc. Natl. Acad. Sci. U. S. A. [Internet] 118:e2022713118. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.2022713118

Bomblies K. 2020. When everything changes at once: finding a new normal after genome duplication. Proc. R. Soc. B Biol. Sci. [Internet] 287:20202154. Available from: https://royalsocietypublishing.org/doi/10.1098/rspb.2020.2154

Booker TR, Yeaman S, Whitlock MC. 2022. Using genome scans to identify genes used repeatedly for adaptation. bioRxiv:1–28.

Bray SM, Wolf EM, Zhou M, Busoms S, Bohutinska M, Paajanen P, Monnahan P, Koch J, Fischer S, Koch MA, et al. 2020. Convergence and novelty in adaptation to whole genome duplication in three independent polyploids. bioRxiv [Internet]:2020.03.31.017939. Available from: http://biorxiv.org/content/early/2020/04/01/2020.03.31.017939.abstract

Burri R. 2017. Interpreting differentiation landscapes in the light of long-term linked selection. Available from: https://onlinelibrary.wiley.com/doi/10.1002/evl3.14

Crockford A, Zalmas LP, Grönroos E, Dewhurst SM, McGranahan N, Cuomo ME, Encheva V, Snijders AP, Begum J, Purewal S, et al. 2017. Cyclin D mediates tolerance of genome-doubling in cancers with functional p53. Ann. Oncol. [Internet] 28:149–156. Available from: http://www.annalsofoncology.org/article/S0923753419319350/fulltext

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nat. Protoc. [Internet] 2:2233–2244. Available from: https://pubmed.ncbi.nlm.nih.gov/17853881/

Doyle JJ, Coate JE. 2019. Polyploidy, the nucleotype, and novelty: The impact of genome doubling on the biology of the cell. Int. J. Plant Sci. 180:1–52.

Dukić M, Bomblies K. 2022. Male and female recombination landscapes of diploid Arabidopsis arenosa. Genetics [Internet] 220. Available from: https://academic.oup.com/genetics/article/220/3/iyab236/6499271

Guggisberg A, Liu X, Suter L, Mansion G, Fischer MC, Fior S, Roumet M, Kretzschmar R, Koch MA, Widmer A. 2018. The genomic basis of adaptation to calcareous and siliceous soils in Arabidopsis lyrata. Mol. Ecol. 27:5088–5103.

Hämälä T, Mattila TM, Savolainen O. 2018. Local adaptation and ecological differentiation under selection, migration, and drift in *Arabidopsis lyrata* *. Evolution (N. Y). [Internet] 72:1373–1386. Available from: http://doi.wiley.com/10.1111/evo.13502

Hämälä T, Savolainen O. 2019. Genomic Patterns of Local Adaptation under Gene Flow in Arabidopsis lyrata.Purugganan M, editor. Mol. Biol. Evol. [Internet] 32:2557–2571. Available from: https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msz149/5522663

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al. 2011. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat. Genet. [Internet] 43:476–481. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21478890

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589.

Imai KK, Ohashi Y, Tsuge T, Yoshizumi T, Matsui M, Oka A, Aoyama T. 2006. The A-Type Cyclin CYCA2;3 Is a Key Regulator of Ploidy Levels in Arabidopsis Endoreduplication W OA. Plant Cell [Internet] 18:382–396. Available from: www.plantcell.org/cgi/doi/10.1105/tpc.105.037309.

Jombart T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Klepikova A V., Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. 2016. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. Plant J. [Internet] 88:1058–1070. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/tpj.13312

Kolář F, Čertner M, Suda J, Schönswetter P, Husband BC. 2017. Mixed-Ploidy Species: Progress and Opportunities in Polyploid Research. Trends Plant Sci. [Internet] 22:1041–1055. Available from: http://www.cell.com/article/S1360138517302078/fulltext

Kolář F, Fuxová G, Záveská E, Nagano AJ, 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*. Mol. Ecol. [Internet] 25:3929–3949. Available from: http://doi.wiley.com/10.1111/mec.13721

Kolář F, Lučanová M, Záveská E, Fuxová G, Mandáková T, Španiel S, Senko D, Svitok M, Kolník M, Gudžinskas Z, et al. 2016. Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae). Biol. J. Linn. Soc. [Internet] 119:673–688. Available from: https://academic.oup.com/biolinnean/article-lookup/doi/10.1111/bij.12479

Konečná V, Bray S, Vlček J, Bohutínská M, Požárová D, Roy Choudhury R, Bollmann-Giolai A, Flis P, Salt DE, Parisod C, et al. 2021. Parallel adaptation in autopolyploid Arabidopsis arenosa is dominated by repeated recruitment of shared alleles. Nat. Commun. [Internet]. Available from: https://doi.org/10.1038/s41467-021-25256-5

Kyriakidou M, Tai HH, Anglin NL, Ellis D, Strömvik M V. 2018. Current strategies of polyploid plant genome sequence assembly. Front. Plant Sci. 871.

Lafon-Placette C, Johannessen IM, Hornslien KS, Ali MF, Bjerkan KN, Bramsiepe J, Glöckle BM, Rebernig CA, Brysting AK, Grini PE, et al. 2017. Endosperm-based hybridization barriers explain the pattern of gene flow between Arabidopsis lyrata and Arabidopsis arenosa in Central Europe. Proc. Natl. Acad. Sci. [Internet] 114:E1027–E1035. Available from: https://www.pnas.org/content/114/6/E1027

Lee KM, Coop G. 2019. Opinion piece Population genomics perspectives on convergent adaptation. Philos. Trans. R. Soc. B [Internet]:20180236. Available from: http://dx.doi.org/10.1098/rstb.2018.0236

Marburger S, Monnahan P, Seear PJ, Martin SH, Koch J, Paajanen P, Bohutínská M, Higgins JD, Schmickl R, Yant L. 2019. Interspecific introgression mediates adaptation to whole genome duplication. Nat. Commun. [Internet] 10:5218. Available from: http://www.nature.com/articles/s41467-019-13159-5

Marhold K, Lihová J. 2006. Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. Plant Syst. Evol.:143–174.

Marques DA, Meier JI, Seehausen O. 2019. A Combinatorial View on Speciation and Adaptive Radiation. Trends Ecol. Evol. 34:531–544.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Res. 23:1817–1828.

McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, Payre F, Stern DL. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. Nat. 2007 4487153 [Internet] 448:587–590. Available from: https://www.nature.com/articles/nature05988

Monnahan P, Kolář F, Baduel P, Sailer C, Koch J, Horvath R, Laenen B, Schmickl R, Paajanen P, Šrámková G, et al. 2019. Pervasive population genomic consequences of genome duplication in Arabidopsis arenosa. Nat. Ecol. Evol. [Internet] 3:457–468. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30804518

Moran RL, Richards EJ, Ornelas-García CP, Gross JB, Donny A, Wiese J, Keene AC, Kowalko JE, Rohner N, Mcgaugh SE. 2022. Selection-driven trait loss in independently evolved cavefish populations 1 2 Authors 3. bioRxiv [Internet]. Available from: https://doi.org/10.1101/2022.11.28.518185

Morgan C, Knight E, Bomblies K. 2022. The meiotic cohesin subunit REC8 contributes to multigenic adaptive evolution of autopolyploid meiosis in Arabidopsis arenosa. PLOS Genet. [Internet] 18:e1010304. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010304

Morgan C, Zhang H, Henry CE, Franklin FCH, Bomblies K. 2020. Derived alleles of two axis proteins affect meiotic traits in autotetraploid *Arabidopsis arenosa*. Proc. Natl. Acad. Sci. [Internet]:201919459. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1919459117

Morgan EJ, Čertner M, Lučanová M, Deniz U, Kubíková K, Venon A, Kovářík O, Lafon Placette C, Kolář F. 2021. Disentangling the components of triploid block and its fitness consequences in natural diploid-tetraploid contact zones of Arabidopsis arenosa. New Phytol. [Internet] 232:1449–1462. Available from: https://pubmed.ncbi.nlm.nih.gov/33768528/

Nei M. 1972. Genetic Distance between Populations. Am. Nat. 106:283–292.

Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, Guggisberg A, Paape T, Schmid K, Fedorenko OM, et al. 2016. Sequencing of the genus Arabidopsis

identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nat. Genet. 48:1077–1082.

Novikova PY, Hohmann N, Van de Peer Y. 2018. Polyploid Arabidopsis species originated around recent glaciation maxima. Curr. Opin. Plant Biol. 42:8–15.

Oziolor EM, Reid NM, Yair S, Lee KM, Guberman VerPloeg S, Bruns PC, Shaw JR, Whitehead A, Matson CW. 2019. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. Science [Internet] 364:455–457. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31048485

Pecinka A, Fang W, Rehmsmeier M, Levy AA, Mittelsten Scheid O. 2011. Polyploidization increases meiotic recombination frequency in Arabidopsis. BMC Biol. [Internet] 9:1–7. Available from: https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-9-24

Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet. 8.

Potapova TA, Seidel CW, Box AC, Rancati G, Li R. 2016. Transcriptome analysis of tetraploid cells identifies cyclin D2 as a facilitator of adaptation to genome doubling in the presence of p53. Mol. Biol. Cell [Internet] 27:3065–3084. Available from: https://pubmed.ncbi.nlm.nih.gov/27559130/

Raj A, Stephens M, Pritchard JK. 2014. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics [Internet] 197:573–589. Available from: /pmc/articles/PMC4063916/

Rawat V, Abdelsamad A, Pietzenuk B, Seymour DK, Koenig D, Weigel D, Pecinka A, Schneeberger K. 2015. Improving the Annotation of Arabidopsis lyrata Using RNA-Seq Data.Provart NJ, editor. PLoS One [Internet] 10:e0137391. Available from: https://dx.plos.org/10.1371/journal.pone.0137391

Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, et al. 2021. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. Sci. Adv. [Internet] 7:5285–5303. Available from: https://www.science.org/doi/10.1126/sciadv.abg5285

Ronfort J. 1999. The mutation load under tetrasomic inheritance and its consequences for the evolution of the selfing rate in autotetraploid species. Genet. Res., Camb [Internet] 74:31–42. Available from: https://doi.org/10.1017/S0016672399003845

Schmickl R, Koch MA. 2011. Arabidopsis hybrid speciation processes. Proc. Natl. Acad. Sci. U. S. A. [Internet] 108:14192–14197. Available from: www.pnas.org/lookup/suppl/doi:10.

Schmickl R, Yant L. 2021. Adaptive introgression: how polyploidy reshapes gene flow landscapes. New Phytol. [Internet] 230:457–461. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/nph.17204

Seear PJ, France MG, Gregory CL, Heavens D, Schmickl R, Yant L, Higgins JD. 2020. A novel allele of ASY3 is associated with greater meiotic stability in autotetraploid Arabidopsis lyrata. PLOS Genet. [Internet] 16:e1008900. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008900

Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, Sorenson AL, De S, Kishony R, Michor F, Dowell R, et al. 2015. Polyploidy can drive rapid adaptation in yeast. Nat. 2015 5197543 [Internet] 519:349–352. Available from: https://www.nature.com/articles/nature14187

Sterken R, Kiekens R, Boruc J, Zhang F, Vercauteren A, Vercauteren I, De Smet L, Dhondt S, Inzé D, De Veylder L, et al. 2012. Combined linkage and association mapping reveals CYCD5;1 as a quantitative trait gene for endoreduplication in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. [Internet] 109:4678–4683. Available from: www.pnas.org/cgi/doi/10.1073/pnas.1120811109

Stift M, Kolář F, Meirmans PG. 2019. STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. Heredity (Edinb). [Internet] 123:429–441. Available from: https://pubmed.ncbi.nlm.nih.gov/31285566/

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. [Internet] 43:D447–D452. Available from: http://tissues.jensenlab.org

Tajima F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics 123:585–595.

Wang L, Josephs EB, Lee KM, Roberts LM, Rellán-Álvarez R, Ross-Ibarra J, Hufford MB. 2021. Molecular Parallelism Underlies Convergent Highland Adaptation of Maize Landraces. Mol. Biol. Evol. [Internet] 38:3567–3580.

Wang M, Zhao Y, Zhang B. 2015. Efficient Test and Visualization of Multi-Set Intersections. Sci. Rep. 5.

Wos G, Macková L, Kubíková K, Kolář F. 2022. Ploidy and local environment drive intraspecific variation in endoreduplication in Arabidopsis arenosa. Am. J. Bot. [Internet] 109:259–271. Available from: https://pubmed.ncbi.nlm.nih.gov/35137947/

Yant L, Bomblies K. 2017. Genomic studies of adaptive evolution in outcrossing Arabidopsis species. Curr. Opin. Plant Biol. [Internet] 36:9–14. Available from: http://www.sciencedirect.com/science/article/pii/S1369526616301492

Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, Franklin FCH, Bomblies K. 2013. Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa. Available from: http://www.sciencedirect.com/science/article/pii/S0960982213011251

Zhang Y, He J, McCormick S. 2009. Two Arabidopsis AGC kinases are critical for the polarized growth of pollen tubes. Plant J. [Internet] 58:474–484. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-313X.2009.03792.x