Mosaic haplotypes underlie repeated adaptation to whole genome duplication in Arabidopsis lyrata and Arabidopsis arenosa

Authors:

Magdalena Bohutínská[123*], Eliška Petříková[1], Tom R. Booker[4], Cristina Vives Cobo[1], Jakub Vlček[15], Gabriela Šrámková[1], Alžběta Poštulková[1], Jakub Hojka[12], Karol Marhold[1], Levi Yant[6#], Filip Kolář[12#], Roswitha Schmickl[12#*]

# Equal contribution

* Correspondence:

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

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

Affiliations:

1 Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic

2 Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic

3 Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

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

5 Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic

6 Department of Life Sciences, University of Nottingham, Nottingham, UK.

Classification: Biological Sciences: Evolution

Keywords: adaptation, polyploidy, *Arabidopsis*, repeatability, polymorphism

Abstract

Polyploidy, resulting from whole genome duplication (WGD), is widespread across diversity, including plants, fungi, and fish. While some WGD lineages indeed thrive, WGD is often fatal due to meiotic challenges and changes in cell physiology. Polyploid adaptation may be facilitated by increased access to allelic variation resulting from gene flow with parental diploid and related tetraploid lineages, as well as elevated mutation rates due to doubled chromosomes. In this study, we aimed to deconstruct the composition and evolutionary origins of the haplotypes displaying the strongest selection signals in independent polyploid lineages. Specifically, we sought to investigate whether polyploid lineages utilize their greater diversity of allelic sources by assembling mosaic haplotypes from multiple such sources. We use two sister diploid/autotetraploid species (within-species polyploids, *Arabidopsis lyrata* and *Arabidopsis arenosa*), first demonstrating the existence of four independent autotetraploid lineages in these species. In each of these we see 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 983 Arabidopsis individuals (2924 haploid genomes in 504 diploids and 479 autotetraploids) found that these alleles were made up of fine-scaled mosaics of variants from remarkably diverse evolutionary sources, including primarily reassortments of trans-specific polymorphism from diploids, novel mutations, and inter-species hybridization. We speculate that such acquisition flexibility and re-shuffling of alleles enabled tetraploids to rapidly adapt to polyploidization, and may further promote their adaptation to environmental challenges.

Significance statement

Polyploidy, the result of genome doubling, 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 diploid relatives, 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 the evolutionary context of such variation: what are the sources of the adaptive genetic variation and how do these variants assemble into positively selected haplotypes.

From a genetic point of view, 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; Baduel et al. 2018; 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) providing a broader pool of standing variation. As a consequence, polyploids may generate and maintain genetic variability at increased rates compared to their diploid progenitors and possibly use multiple sources of genetic variability to adapt (Fig. 1).

Beneficial alleles of a particular gene are most commonly envisioned as originating from a single allelic source (as shown in Fig. 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 can accumulate over time in a genomic region 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 ask 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. 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 (i.e. ancient genetic variants whose origin predates speciation events, resulting in shared alleles) across *Arabidopsis* species. We thus do not know to which extent de novo vs. pre-existing variation served as a source of adaptive variation and, in particular, to which extent those processes may be jointly involved in adaptation via assembly of mosaic adaptive haplotypes.

Here, we overcome this by utilizing an exhaustive dataset of 983 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). We used this exhaustive sampling of four independent polyploid lineages, estalished which genes and cellular processes have been repeatedly under selection and what are the sources of adaptive alleles in these lineages (introgression, standing variation, de novo mutations). Using this knowledge we asked to which extent are haplotypes, which were repeatedly selected in tetraploids, formed from a fine-scale mosaic of multiple allelic sources.

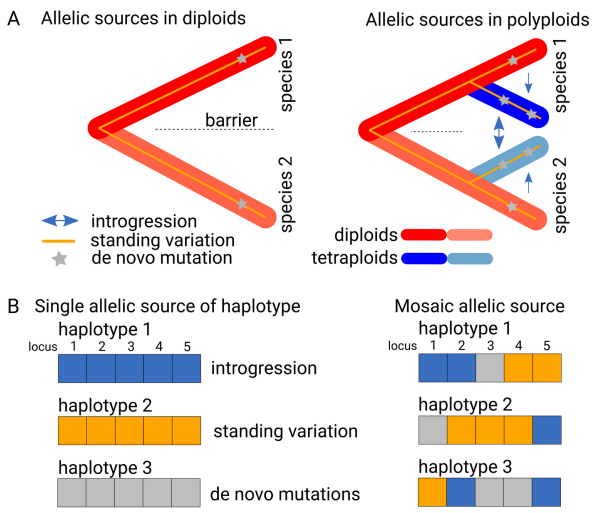

Fig. 1: *Hypotheses about sources of adaptive alleles in a diploid-polyploid system. A: As compared to diploids, polyploid lineages may acquire alleles via increased introgression potential, increased population-scaled mutation rate, and higher level of retainment of ancestral polymorphism, thus behaving 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

To be able to reconstruct allelic sources involved in adaptation to whole genome duplication, we compiled a collection of published and newly sequenced individual-level whole genome sequencing data of 818 individuals from 46 diploid and 61 tetraploid populations of *A. lyrata* and *A. arenosa,* as well as 165 individuals from outgroups. Our new sampling focused on Central European *A. lyrata* in order to cover all regions possibly harbouring the tetraploid cytotype (Ansell et al. 2010; 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). 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 and cytotypes 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).

In summary, we identified four distinct tetraploid lineages in Central European *Arabidopsis*, consistent with two to four independent WGD events (one in each species and possibly up to two additional WGD events in *A. lyrata*). These WGD events reduced between-species hybridization barriers, leading to interspecific gene flow between

tetraploids. This makes hybridization among polyploid lineages a plausible source of tetraploid-adaptive alleles in this system.

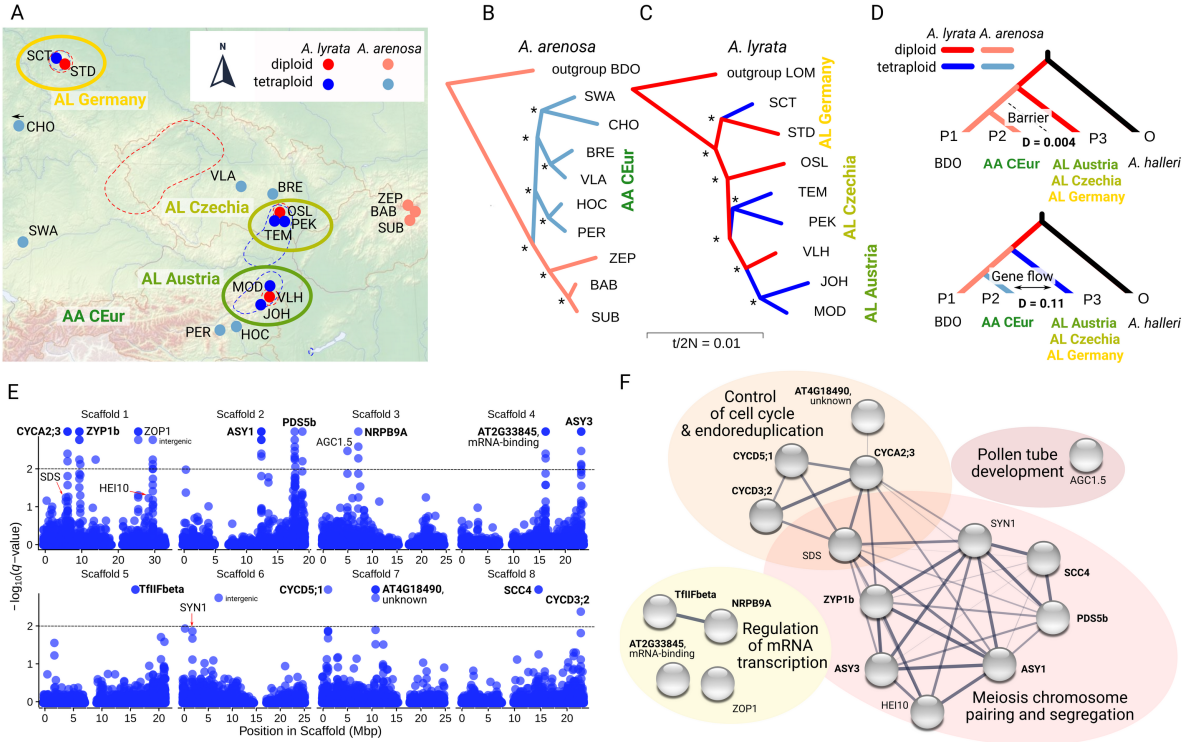

Fig. 2: *Evolutionary relationships and genomic signatures of repeated positive selection in tetraploid populations of A. lyrata and A. arenosa. A: Locations of 17 most proximal populations of diploid and tetraploid A. lyrata (AL Austria, AL Czechia, and AL Germany) and A. arenosa (AA CEur) in Central Europe. Red and blue dashed lines show the 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. Asterisks show bootstrap 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 A. halleri from Austria. E: Using PicMin, we identified a set of 14 unlinked genes showing significant evidence (p < 0.01) of positive selection in tetraploids (positively selected genes, PSGs). Additional three genes, SDS, HEI10 and SYN1, were identified using a screen for candidate SNPs. F: Functional characterisation of PSGs by STRING analysis. 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).*

Genome-wide signatures of repeated adaptation to whole genome duplication

To reconstruct allelic sources required for adaptation to whole genome duplication in our four polyploid lineages, we first needed to identify a reliable set of genes exhibiting repeated signals of positive selection. To do so, we compared our four pairs of diploid and tetraploid *A. lyrata* and *A. arenosa* populations (Fig. 2B, C). We identified candidate genes using PicMin, a novel method that leverages repeated adaptation to identify robust signatures of selection (Booker et al. 2022). We identified 54 significantly differentiated windows (PicMin

FDR-corrected *q*-value < 0.01; Dataset S2), overlapping 14 unlinked candidate genes (Fig. 2E, Table S2). 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. It confirmed all 14 PicMin candidate genes (Table S2) and identified three additional candidate genes (*HEI10*, *SYN1* and *SDS* with three, three and two candidate SNPs differentiated in all four tetraploid lineages, respectively; Dataset S3). Thus, in total, we identified 14 plus three candidate positively selected genes (PSGs), exhibiting strong signals of repeated selection in the tetraploid lineages. 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 understand in which molecular processes are the 17 PSGs involved, we predicted protein-protein interactions among them using STRING analysis (Szklarczyk et al. 2015). 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. 2F). 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), four of these six genes were found to be related to mRNA transcription via RNA polymerase II (Fig. 2F). 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). We provide further functional interpretations in Supplementary Text 1 and Fig. S3.

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. This is consistent with results of cytological studies reporting 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). Further, we used flow cytometry and found significant phenotype compensation in the cell cycle trait endoreduplication (Supplementary Text 2, Fig. S4, S5). Thus, the highly interacting meiosis and cell cycle regulation PSGs likely mediate beneficial phenotypic shifts resulting in the establishment of tetraploid lineages.

Mosaic tetraploid haplotypes are assembled from diverse allelic sources

Recent evidence suggests that polyploids may access a broader spectrum of sources of adaptive alleles than diploids (Ronfort 1999; Selmecki et al. 2015; Arnold et al. 2016; Monnahan et al. 2019). Here we inquired if these sources vary over fine genomic scales, possibly resulting in polyploid haplotypes representing a mosaic of different allelic sources (Fig. 1B).

We used our broad dataset of 2,924 sampled haploid *Arabidopsis* genomes (including exhaustive sampling of all known diploid lineages; Fig. 3A, Dataset S1) to deconstruct these

haplotypes and identify the specific allelic sources for every candidate tetraploid SNP in our PSGs. Removing five PSGs with insufficient variation (less than five candidate SNPs), we analysed 12 genes and 232 candidate SNPs (Dataset S4). For each gene, we reconstructed major diploid and tetraploid 'candidate haplotypes' (haplotypes hereafter) using these SNPs as haplotype markers (following (Bohutínská, Handrick, et al. 2021)). We also validated the haplotypes using PacBio HiFi reads from five diploid and five tetraploid individuals of *A. arenosa* (Dataset S4 – S6). We identified a single major tetraploid-specific haplotype for each of the 12 candidate genes that was shared across both *A. lyrata* and *A. arenosa* tetraploids (92-98% of their populations; 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 major 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. 3B). Other haplotypes were of minor frequency (Fig. 3B).

The tetraploid haplotypes for the 12 candidate genes were found at high frequency in tetraploid populations of both species, reinforcing the evidence that they have undergone strong selective sweeps (mean frequency = 0.62 across 61 populations; Fig. 3B, Table S4). Interestingly, none of these 12 tetraploid haplotypes were present in any diploids, despite exhaustive sampling including the putative source lineages of the polyploids. This indicates that they likely assembled upon the establishment of tetraploids rather than preexisting as standing variation in diploids.

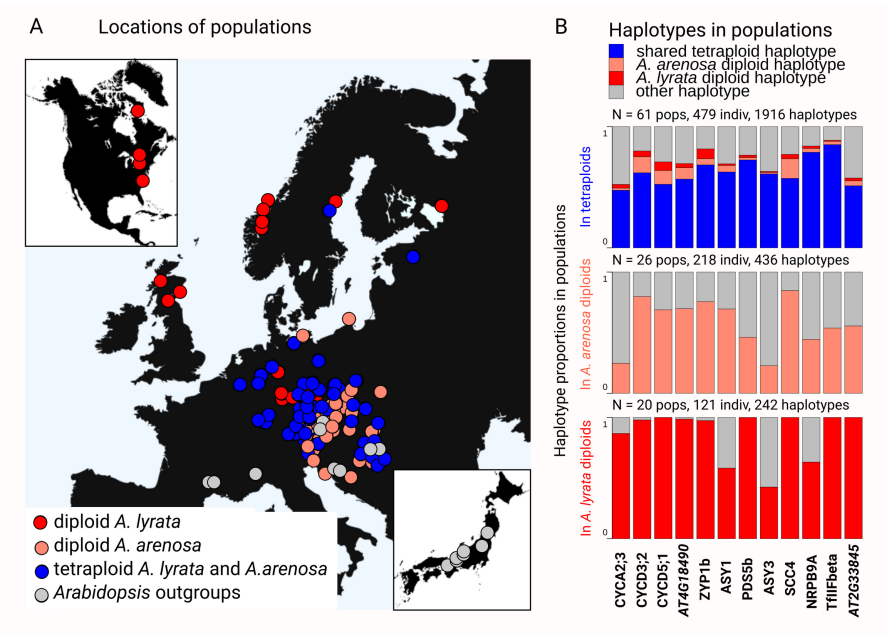

Fig. 3: Distribution of candidate haplotypes in diploids and tetraploids of A. lyrata and A. arenosa. A: Geographic distribution of populations used in this analysis (504 diploids and 479 tetraploids). Arabidopsis outgroups are A. halleri, A. croatica, A. cebennensis and A. pedemontana. B: Frequency of shared tetraploid (blue), A. arenosa diploid (light red), and A. lyrata diploid (red) haplotypes in each of the 12 candidate genes. All other haplotypes present in the populations (including possible recombinants of the above ones) are highlighted in grey.

While every complete tetraploid haplotype was missing in diploids, the sources of constitutive SNPs that made up these haplotypes were widespread within the diploids at both the species-level and often in multiple species (Fig. 4). 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* (alleles found in diploids of multiple *Arabidopsis* species (Novikova et al. 2016)), we worked with the full dataset of *A. arenosa, A. lyrata* and all other diploid *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 504 diploid individuals, we determined the most likely allelic source of each of the 232 haplotype-marking SNPs (Fig. 4, Dataset S7). Surprisingly, 65.5% of the SNPs forming tetraploid haplotypes were sourced from trans-specific polymorphism, segregating in diploids of multiple *Arabidopsis* species (Fig. 4A, scenarios 1-4). Note that this number can be still underestimated if the allele remained unsampled or went extinct in some (ghost or existing) diploid lineage. 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. 4A, 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. 4A, 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 possibly in diploids of other species, but not *A. lyrata*; Fig. 4A, scenarios 2, 5), while 3.4% likely came from diploid *A. lyrata* (Fig. 4A, 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. 4A, 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 is in line with signatures of hybridization among tetraploids (Fig. 2D, (Marburger et al. 2019)).

The different allelic sources forming tetraploid haplotypes varied genome-wide (when analysing all 12 PSGs together; Fig. 4A), but also within individual tetraploid haplotypes (3 – 6 scenarios per haplotype, median = 4; out of 7; Fig. 4B, 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 deconstruction of variants forming tetraploid haplotypes suggests that polyploids source alleles from diverse source pools, 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)).

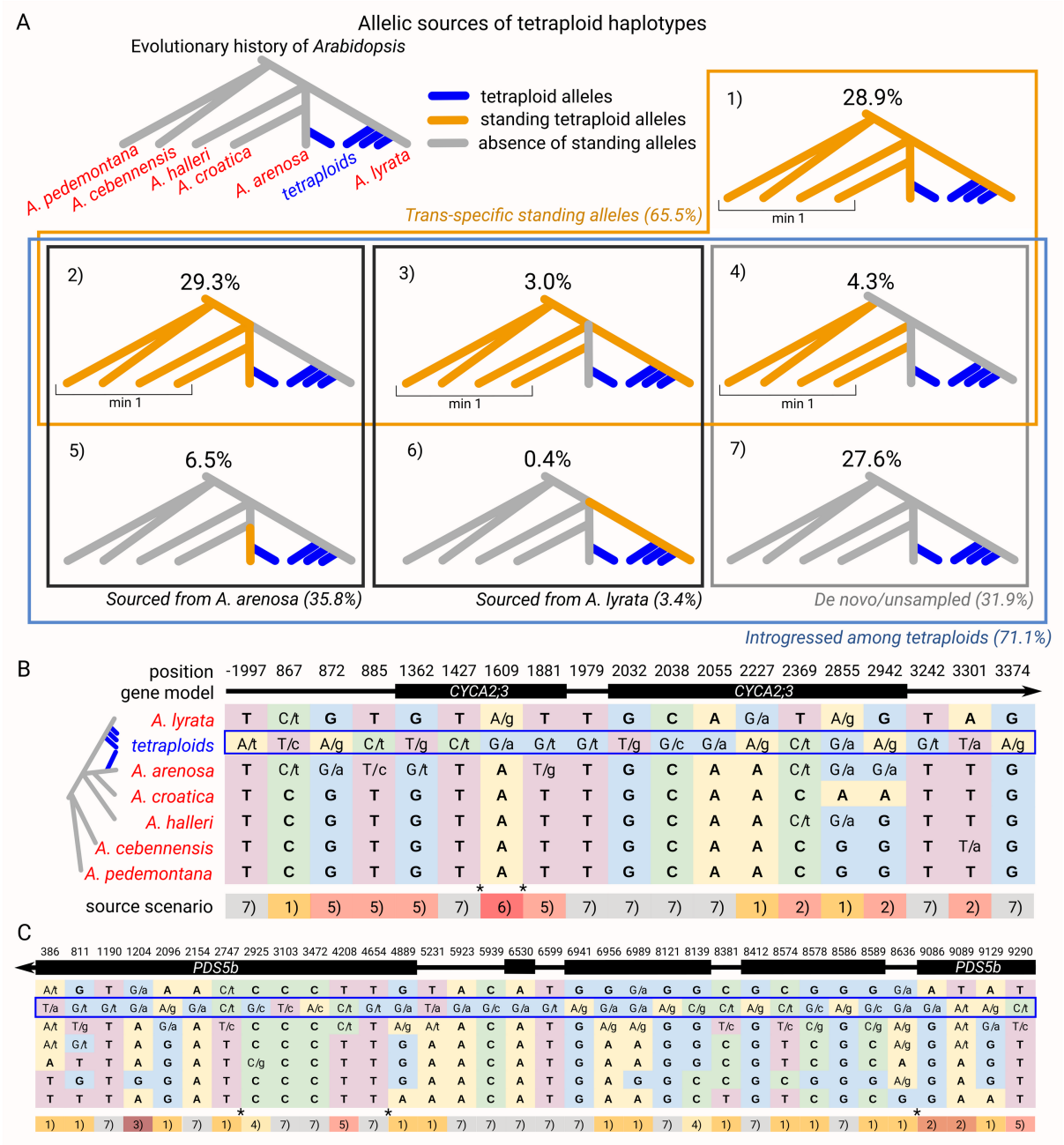

Fig. 4: Mosaic allelic sources of tetraploid haplotypes. A: Candidate SNPs forming tetraploid haplotypes were categorized into one of the seven source scenarios based on their allele distribution in the 983 samples. The most parsimonious origin of each pattern is provided in italics. Four of them (outlined by the orange frame) involve trans-specific standing variation shared among diploids of at least two Arabidopsis species. Six of them (outlined by a blue frame) 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 genetic variation in polyploids, leveraging our exhaustive sampling of all extant outcrossing *Arabidopsis* species in Europe. Through genomic analysis of four distinct polyploid lineages, we identified a set of genes involved in cell cycle, meiosis, and transcription that repeatedly exhibit the strongest signals of selection. We found that for all these candidate positively selected 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 polymorphisms in ancestral and sister diploid species 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, 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 rapidly and efficiently create new adaptive haplotypes. This haplotype assembly process, however, is not exclusive to 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 selected haplotypes for an entire suite of functionally and often physically interacting genes. Recent research in various plant and animal species has already shown that there is a significant variability in the extent to which standing, introgressed, and de novo allelic sources contribute to repeated adaptation at the level of entire genes (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 suggests that evolutionary pathways to adaptation may be even more variable if the assembly of individual variants into haplotypes is taken into account.

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 910 published (Novikova et al. 2016; Hämälä et al. 2017; Mattila et al. 2017; Novikova et al. 2017; Guggisberg et al. 2018; Hämälä et al. 2018; Marburger et al. 2019; Monnahan et al. 2019; Preite et al. 2019; Bohutínská, Vlček, et al. 2021; Konečná et al. 2021) whole genome sequences of these species and outgroup species, totaling 983 individuals and 129 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). *A. arenosa* was rooted with the diploid Pannonian population BDO and *A. lyrata* with the diploid Scandinavian population LOM. 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 (Booker et al. 2022; Bohutínská et al. 2023).

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 (86,249 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 0.01 and lower 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.

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).

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 218 / 121 diploids of *A. arenosa* and *A. lyrata*, respectively, and 479 tetraploids, we defined *Mi* to be the major allele frequency at the candidate SNP *i*. Given that the sample consists of 1916 tetraploid haplotypes and 436 / 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 of the 12 PSGs from five diploid and five tetraploid samples. The sequences of 12 PSGs (Dataset S5) were extracted from newly produced diploid and tetraploid assemblies by performing a BLASTn v2.10.0 on each assembly. Extracted diploid and tetraploid haplotype sequences were aligned using MUSCLE and visualised with IGV v2.11.9 to proceed with haplotype analysis. Further, long reads were aligned to the assemblies to validate if diploid and tetraploid candidate SNP alleles are physically linked on the same read, as determined by our short read-based analysis. Indeed, 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 – S6).

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* (339 diploid individuals) and their outcrossing relatives *A. halleri*, *A. croatica*, *A. cebennensis* and *A. pedemontana* (‘diploid outgroups’ hereafter, 165 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 (218 of *A. arenosa*, 121 of *A. lyrata*, and 165 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.

Data Availability

Sequence data that support the findings of this study are deposited in the NCBI (https://www.ncbi.nlm.nih.gov/bioproject/) under BioProjects PRJNA284572, PRJNA309929, PRJNA357693, PRJNA357372, PRJNA459481, PRJNA493227, PRJEB34247 (ENA), PRJNA506705, PRJNA484107, PRJNA592307, PRJNA667586, PRJNA929698. See Dataset S1 for individual codes.

Acknowledgements:

We greatly appreciate the constructive feedback from members of the Ecolgen team in Prague, Katie Peichel, Reto Burri, Alison Scott, Polina Novikova, Andrew MacColl, Tuomas Hämälä, John Brookfield, Emma Curran, Laura Dean, Sian Bray, Sarah Goodacre and Ana da Silvia. We further appreciate inspiration provided by the ForBio Polyploid course in Drobak, Norway. We thank Martin Čertner and Dorka Čertnerová for sharing their synthetic polyploid protocol, Bodo Schwarzberg (Untere Naturschutzbehörde Nordhausen) for sharing 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.), the European Research Council (project 850852 DOUBLEADAPT to F.K.) 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, Monnahan P, Kolář F. 2023. Population Genomic Analysis of Diploid-Autopolyploid Species. Methods Mol. Biol. [Internet] 2545:297–324. Available from: https://pubmed.ncbi.nlm.nih.gov/36720820/

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

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, Leinonen PH, Kuittinen H, Savolainen O. 2017. Role of seed germination in adaptation and reproductive isolation in Arabidopsis lyrata. Mol. Ecol. [Internet] 26:3484–3496. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/mec.14135

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, 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.

Mattila TM, Tyrmi J, Pyhäjärvi T, Savolainen O. 2017. Genome-Wide Analysis of Colonization History and Concomitant Selection in Arabidopsis lyrata. Mol. Biol. Evol. [Internet] 34:2665–2677. Available from: https://academic.oup.com/mbe/article/34/10/2665/3952783

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.

Novikova PY, Tsuchimatsu T, Simon S, Nizhynska V, Voronin V, Burns R, Fedorenko OM, Holm S, Säll T, Prat E, et al. 2017. Genome Sequencing Reveals the Origin of the Allotetraploid Arabidopsis suecica. Mol. Biol. Evol. [Internet] 34:957–968. Available from: https://pubmed.ncbi.nlm.nih.gov/28087777/

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/

Preite V, Sailer C, Syllwasschy L, Bray S, Ahmadi H, Krämer U, Yant L. 2019. Convergent evolution in Arabidopsis halleri and Arabidopsis arenosa on calamine metalliferous soils. Philos. Trans. R. Soc. B Biol. Sci. 374.

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. Available from: https://academic.oup.com/mbe/article/38/9/3567/6255756

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

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