

Multiple Autopolyploid *Arabidopsis lyrata* Populations Stabilized by Long-Range Adaptive Introgression Across Eurasia

Alison D. Scott ^{1,*} Uliana K. Kolesnikova ¹ Anna Glushkevich ¹ Laura Steinmann ¹
Nikita P. Tikhomirov ¹ Ursula Pfordt ¹ Magdalena Bohutínská ^{2,3} Robin Burns ⁴
Alexey P. Seregin ⁵ Filip Kolar ^{2,3} Roswitha Schmickl ^{2,3} Polina Yu. Novikova ^{1,*}

¹Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne 50829, Germany

²Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic

³Institute of Botany, The Czech Academy of Sciences, Průhonice, Czech Republic

⁴Department of Plant Sciences, University of Cambridge, Cambridge, UK

⁵Herbarium (MW), Faculty of Biology, M. V. Lomonosov Moscow State University, Moscow 119991, Russia

*Corresponding authors: E-mails: ascott@mpipz.mpg.de; pnovikova@mpipz.mpg.de.

Associate editor: Kathryn Hodgins

Abstract

Abundance of established polyploid lineages varies across lineages, evolutionary time, and geography, suggesting both genetics and environment play a role in polyploid persistence. We show *Arabidopsis lyrata* is the most polyploid-rich species complex in the *Arabidopsis* genus, with multiple origins of autotetraploidy. This is revealed by genomic data from over 400 *A. lyrata* samples across Eurasia. We found over 30 previously undescribed autotetraploid populations in Siberia with a minimum of two separate origins, independent of those previously reported in Central Europe. The establishment of Siberian tetraploids is mediated by meiotic adaptation at the same genes as in European tetraploid *A. lyrata* and *Arabidopsis arenosa*, despite their genomic divergence and geographical separation. Haplotype analysis based on synthetic long-read assemblies supports the long-range introgression of adaptive alleles from the tetraploid interspecific pool of European *A. lyrata* and *A. arenosa* to tetraploid Siberian *A. lyrata*. Once adaptations to polyploidy emerge, they promote the establishment of new polyploid lineages through adaptive inter- and intraspecific introgression.

Keywords: *Arabidopsis lyrata*, introgression, polyploid, adaptation

Introduction

Polyploidization cycles are recurrent and abundant with non-random distribution in space, through time, and across the Tree of Life. Polyploid frequency has a latitudinal gradient: more polyploid lineages are observed in colder regions closer to the poles (Rice et al. 2019; David 2022), which is paralleled with dating of ancient whole-genome duplication (WGD) events at climatically harsh time periods (Van de Peer et al. 2017). Whole genome duplications are recurrent; most major lineages of flowering plants experienced multiple polyploidy events detected in the period between 75 and 55 Mya (Lohaus and Van de Peer 2016), a time range of drastic climatic changes. Polyploidy is more abundant in plants (Barker et al. 2016; Van de Peer et al. 2017) compared to animals (Gregory and Mable 2005), in angiosperms versus gymnosperms (Khoshoo 1959), and in amphibians versus mammals (Gregory and Mable 2005; Schmid et al. 2015). What explains the abundance of polyploids in certain lineages, their emergence at certain times and climates? The explanation of the non-random distribution of polyploidy must lie in the balance between the birth and death rate of whole genome duplications, which differs depending on the context. For example, the environment, which varies across both time and space, can trigger polyploidy through unreduced gamete production

(De Storme et al. 2012; Mason and Pires 2015; Zhou et al. 2015) and impact extinction rates (Van de Peer et al. 2017). Genetic background also affects polyploid birth rates directly through production rates of unreduced gametes (Mason et al. 2011; Kreiner et al. 2017a, 2017b), and can provide a route to frequent polyploid emergence, as suggested by theoretical modeling (Kauai et al. 2023). Studies from numerous plant systems show multiple origins of polyploidy, further suggesting particular genetic backgrounds provide ample opportunity for polyploid emergence (Soltis et al. 2009), including in autopolyploids (Servick et al. 2015).

The emergence of polyploids does not guarantee their survival. Neopolyploid establishment includes adaptation to external environments (Manzaneda et al. 2012; Chao et al. 2013), but adaptation to the new environment within the nucleus is even more critical. WGD in an individual (or single lineage; autopolyploidy) results in multiple sets of homologous chromosomes. Challenges in executing faithful segregation of the doubled chromosomes during meiosis can hinder the survival of neopolyploids (Bomblies et al. 2015, 2016). Artificially induced polyploids have reduced fertility and show high levels of aneuploidy in subsequent generations (Comai et al. 2000; Yant et al. 2013; Burns et al. 2021; Katche et al. 2023), suggesting that natural polyploids have

Received: September 25, 2024. Revised: May 5, 2025. Accepted: June 6, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

evolved adaptations to the polyploid state (Morgan et al. 2020; Bohutínská et al. 2021b) or may be established only from certain preadaptive combinations of alleles in the diploid pool (Katche et al. 2023).

The genus *Arabidopsis* is a prime example of the non-random nature of polyploidy, as the majority of diploid *Arabidopsis* species contribute to either allo- or auto-polyploid lineages. There are allopolyploid species *Arabidopsis suecica* (parents *Arabidopsis arenosa* and *Arabidopsis thaliana*) and *Arabidopsis kamchatica* (parents *A. lyrata* and *Arabidopsis halleri*), and both *A. lyrata* and *A. arenosa* harbor autotetraploid populations.

Arabidopsis polyploids emerged in times of climatic upheaval (Novikova et al. 2018). *Arabidopsis arenosa* polyploids have been described to have a single geographical origin in the western Carpathian Mountains and spread across Europe after a whole-genome duplication (Arnold et al. 2015; Monnahan et al. 2019). Establishing tetraploid *A. arenosa* involved polygenic adaptation of reproductive machinery, including pollen tube growth (Westermann et al. 2024) and meiosis (Yant et al. 2013; Bombliès et al. 2016; Morgan et al. 2020). The former may be required to restore ion homeostasis, and the latter to prevent entanglement between three or more chromosomal copies and ensure faithful segregation.

We focus our study on the northernmost *Arabidopsis* species with a wide geographical distribution—*Arabidopsis lyrata*. Multiple tetraploid *A. lyrata* populations have been described in Europe in several areas in the Czech Republic and Austrian Alpine foothills (Polatschek 1966; Ansell et al. 2010; Jørgensen et al. 2011; Schmickl and Koch 2011). Tetraploid *A. lyrata* populations in Central Europe form three lineages occupying three distinct regions (Bohutínská et al. 2023), which may suggest multiple origins of polyploidy, in contrast to a single tetraploid origin in *A. arenosa* (Arnold et al. 2015). Diploid and tetraploid *A. lyrata* and *A. arenosa* have overlapping, yet nonsympatric ranges in Europe, and while diploids are reproductively isolated (Lafon-Placette et al. 2017) and do not show signs of ongoing interspecific gene flow (Clauss and Mitchell-Olds 2006; Ross-Ibarra et al. 2008; Ansell et al. 2010; Schmickl and Koch 2011), tetraploid *A. lyrata* and *A. arenosa* often hybridize (Jørgensen et al. 2011; Schmickl and Koch 2011; Lafon-Placette and Kohler 2016; Hohmann and Koch 2017; Schmickl et al. 2017; Marburger et al. 2019). Interestingly, interspecific gene flow between European tetraploid *A. lyrata* and *A. arenosa* has been adaptive: several introgressed alleles of meiotic genes from *A. arenosa* are also under selection in tetraploid *A. lyrata*, as they increase the fitness of tetraploids by stabilizing chromosomal segregation (Marburger et al. 2019; Schmickl and Yant 2021).

There are indications of a polyploid population in Yakutia (Zhukova et al. 1973; Novikova et al. 2018), suggesting that ploidy variation could still be massively underestimated in this species due to a lack of extensive sampling, particularly in Northern Eurasia. Despite the wide distribution of *A. lyrata* across the Northern Hemisphere (Dart et al. 2004; Schmickl et al. 2010; Novikova et al. 2016; Walden et al. 2020), a large part of the described range from Eastern Europe to Eastern Siberia lacks genetic data. We generate a whole-genome re-sequencing dataset from this understudied region and ask if polyploids are abundant there. Here we report multiple new tetraploid *A. lyrata* populations in the areas of White Sea shore, Polar Ural, and Central Siberia. We describe the nature,

origins, and relationships of these WGDs. We ask what explains the abundance of polyploids in *A. lyrata* in these remote geographical regions since the distribution of previously described adapted polyploids is limited to European *A. arenosa* and *A. lyrata*. We investigate if any of the alleles known to stabilize polyploid meiosis could reach from Europe to Eastern Siberia or adaptation to polyploidy is purely convergent.

Results

Tetraploid *A. lyrata* is Abundant in Siberia

We assembled a dataset of whole-genome sequences from several sources. First, we sequenced 116 herbarium samples from the Moscow University Herbarium (Seregin 2025), spanning a wide geographical area from the White Sea to the Bering Strait. Second, we sequenced several individuals per population (283 samples from 32 locales) grown from seeds collected during expeditions to the White Sea, Polar Urals, Gydan Peninsula, Putorana Plateau, Lower Lena, and Amur basin (Fig. 1a, supplementary fig. S1, Supplementary Material online, supplementary data S1, Supplementary Material online). Third, we included whole-genome sequencing data of samples from different *A. lyrata* lineages which have been previously published (Novikova et al. 2016; Hämälä et al. 2017; Mattila et al. 2017; Marburger et al. 2019; Takou et al. 2021; Willi et al. 2022). All re-sequencing data were mapped to the reference genome of the self-compatible Siberian *A. lyrata* NT1 strain (Kolesnikova et al. 2023).

We used a statistical framework based on relative allelic counts implemented in nQuire software (Weiß et al. 2018) to get a first inference of ploidy from sequencing data (see Materials and Methods). We initially called variants for all samples as diploids. The proportion of heterozygous sites in tetraploid individuals was generally higher than in the diploids (Fig. 1b, supplementary fig. S2, Supplementary Material online). To complement the initial inference, we genotyped and counted *S*-alleles in each sample using NGSgenotyp (Genete et al. 2020). Due to the extreme divergence and diversity of *S*-alleles controlling self-incompatibility in *Arabidopsis* (Charlesworth et al. 2003; Mable et al. 2003, 2004; Castric and Vekemans 2004; Castric et al. 2008; Llaurens et al. 2008; Le Veve et al. 2023), it was often possible to find up to two different *S*-alleles in diploid and up to four different *S*-alleles in tetraploid individuals. These independent assessments of ploidy levels allowed us to infer ploidy for most of the samples (Fig. 1b). Finally, we directly counted chromosome numbers in 18 live samples grown in the greenhouse from seeds collected in the field (representative images in supplementary fig. S1, Supplementary Material online and supplementary data S2, Supplementary Material online) and confirmed our bioinformatic inferences. When bioinformatic inference yielded unclear results and cytogenetics was not possible (i.e. for herbarium specimens), we indicated the ploidy as “unsure” and excluded these accessions from downstream analyses—overall, 11 samples were excluded, representing ~2.5% of new accessions (Fig. 1b, supplementary data S1, Supplementary Material online). Also, while tetraploid and diploid populations sometimes occur nearby, we found no evidence of mixed-ploidy populations: all individuals sampled from the same population had consistent ploidy levels (supplementary data S1, Supplementary Material online). Following the assignment of ploidy by heterozygosity, *S*-allele genotype, and cytological confirmation, we re-called

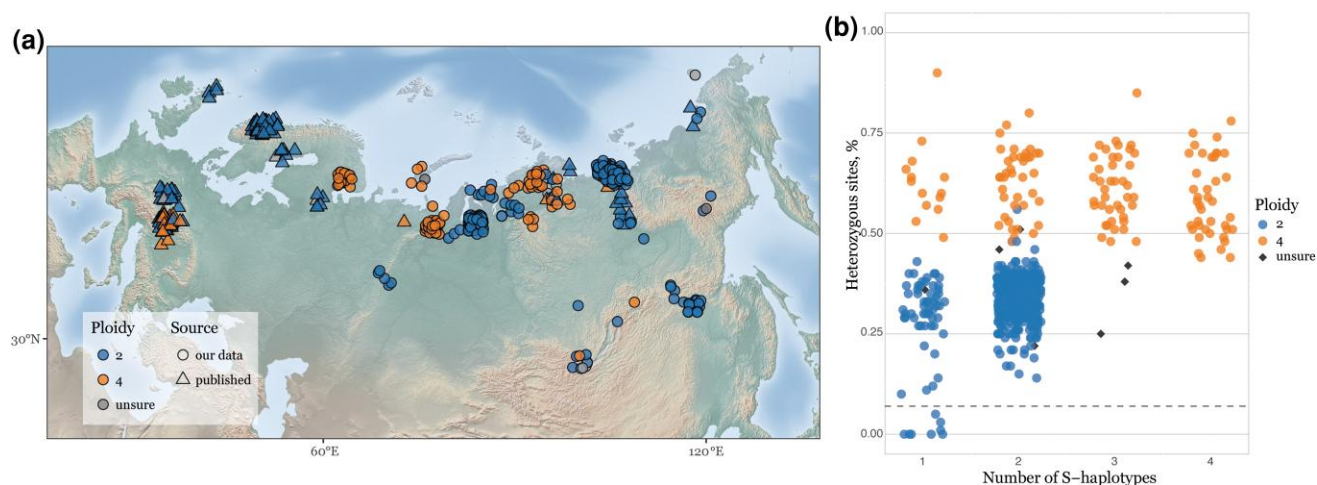


Fig. 1. Distribution of diploid and tetraploid *A. lyrata* across Eurasia. a) Geographical distribution of *A. lyrata* samples used in this study. For better visualization in cases of overlaps we randomly shifted the points within 100 km from the actual sampling location, all true coordinates can be found in the [supplementary data S1, Supplementary Material](#) online. Newly found tetraploid *A. lyrata* populations are represented in orange circles. b) Relationships between proportions of heterozygous biallelic SNPs and the number of different S-alleles in each individual shown in (a). Tetraploid individuals (orange dots) show a higher proportion of heterozygous sites and up to four different S-alleles. Diploid individuals (blue dots) show a lower proportion of heterozygous sites and up to two different S-alleles; individuals with one S-allele below the dashed gray line correspond to recently described Siberian selfing *A. lyrata* lineage with nonfunctional S-locus (Kolesnikova et al. 2023). Herbarium individuals with ambiguous assignments of ploidy between two inference methods are shown in gray dots.

variants according to the inferred ploidy state using the designated ploidy-aware approach in GATK.

Our live collections included confirmed self-compatible and self-incompatible diploid *A. lyrata* individuals, which differ in the levels of heterozygosity (Kolesnikova et al. 2023) (Fig. 1b, [supplementary fig. S2, Supplementary Material](#) online). Tetraploid individuals had a higher proportion of heterozygous sites compared to diploid selfers and outcrossers, and up to four different S-alleles in each individual (Fig. 1b, [supplementary fig. S2, Supplementary Material](#) online), pointing to their outcrossing nature, consistent with self-incompatibility observed in the greenhouse. While nucleotide diversity (π) calculated on putatively neutral four-fold degenerate sites is comparable (~ 0.004) between diploid and tetraploid *A. lyrata* lineages in Siberia ([supplementary fig. S3, Supplementary Material](#) online), a *t*-test comparing nucleotide diversity between diploid and tetraploid lineage pairs shows that π is significantly higher in Northern Ural (NU) tetraploids versus diploids ([supplementary fig. S4, Supplementary Material](#) online). The same comparison for Central Siberian (CS) diploids and tetraploids yields no significant difference. While the theoretical expectation is that π in tetraploids should be twice as high as their diploid ancestors, our observations are consistent with recent work showing that nucleotide diversity in tetraploids takes millions of generations to reach twice that of diploids (Vlček et al. 2025). We additionally calculated Tajima's D (Tajima 1989), which is more strongly negative in tetraploid populations than diploid populations ([supplementary fig. S4, Supplementary Material](#) online), suggesting tetraploid lineages may still be rebounding from a historical bottleneck.

Independent Autopolyploidization Events of *A. lyrata* in NU and Central Siberia Around Last Glacial Maximum

To characterize the population structure of *A. lyrata* across Eurasia, we calculated pairwise genetic distances from the

four-fold degenerate sites as a proxy for nearly neutral sites and built a network from the resulting distance matrix (Fig. 2a), showing that the newly described tetraploids cluster within two distinct groups, separate from the Central Europe tetraploids. NU (highlighted in pink, Fig. 2a) populations sampled around the White Sea and in the Polar Ural Mountains form a cluster containing diploid and tetraploid individuals. This NU cluster is genetically close to the West Siberian (WS) cluster (WS, highlighted in blue, Fig. 2a) comprised purely of diploid populations sampled from the Taz Estuary. The CS cluster (highlighted in yellow, Fig. 2a) also contains both diploids and tetraploids and, in turn, is closer to diploids from the lower Lena river area in East Siberia (ES, highlighted in orange, Fig. 2a). We also used the distance matrix to estimate a UPGMA clustering tree to validate the identity of *A. lyrata* samples ([supplementary fig. S5, Supplementary Material](#) online), which confirmed our taxon assignments.

We estimated gene trees for representatives of the Siberian *A. lyrata* lineages and the resulting phylogenetic summary from ASTRAL (Fig. 2b) clarifies relationships among diploids and tetraploids. Tetraploids from Central Siberia (yellow) are most closely related to diploids from ES (orange), and these two lineages form a clade with WS diploids (blue). NU tetraploids (pink) are sister to the clade of Central Siberia-East Siberia-West Siberia. The self-compatible Eastern Siberia diploid lineage falls outside the rest of the Siberian clade, consistent with its position on the distance matrix in Fig. 2a. Diploid and tetraploid accessions of *A. arenosa* form an independent clade, separate from *A. lyrata*. Overall, the tree topology supports two independent tetraploid origins of *A. lyrata* in Siberia and suggests the Northern Ural (NU) tetraploids are equally closely related to WS diploids, East Siberian diploids, and CS tetraploids.

Using population-level data, we estimated site frequency spectra (SFS) for the two newly discovered tetraploid *A. lyrata* lineages (NU 4 $\times$ and CS 4 $\times$) separately and compared them to theoretical expectations for auto- and allo-polyploids. In both cases, the observed SFS of Siberian *A. lyrata* tetraploids are

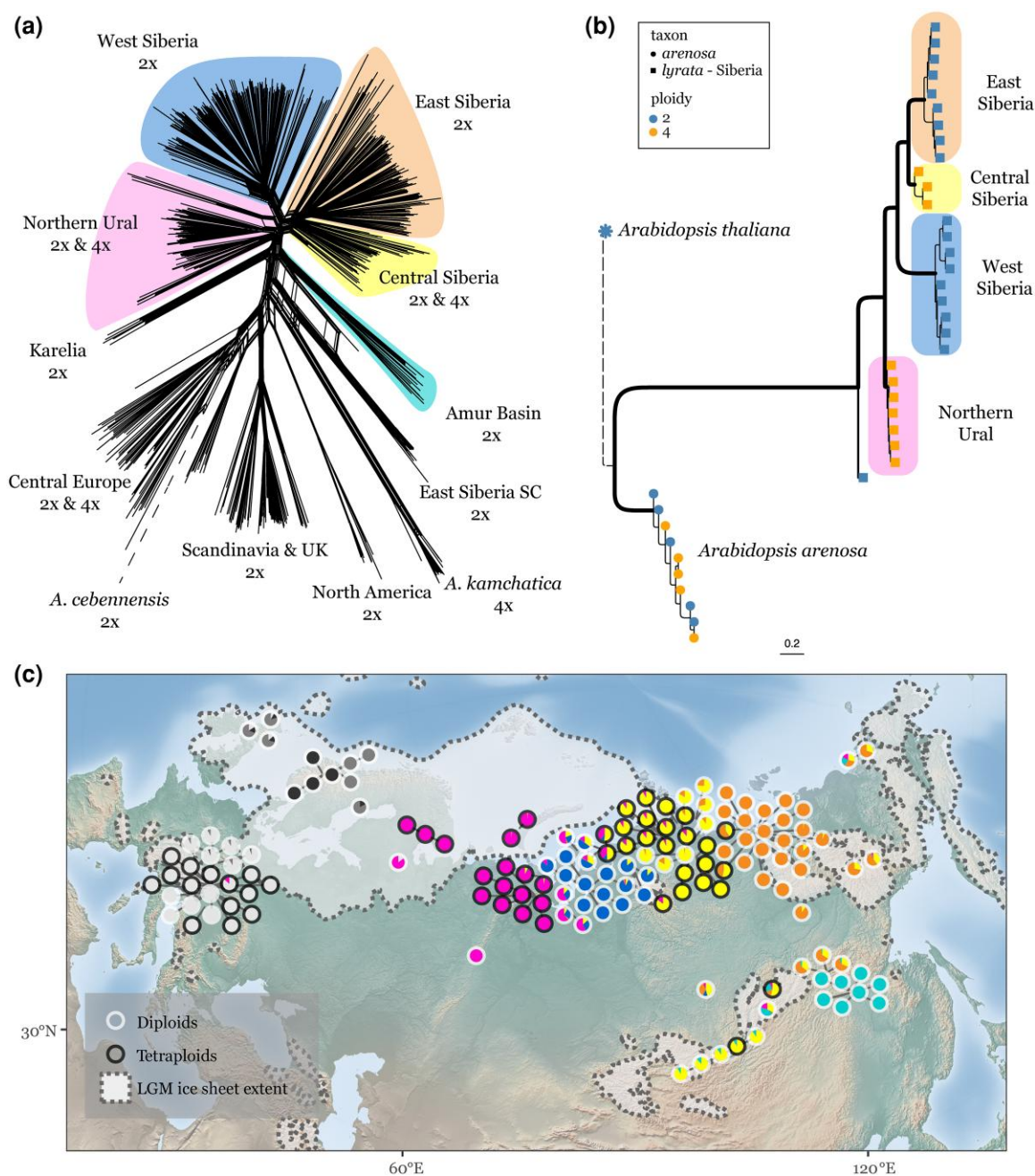


Fig. 2. Population structure of *A. lyrata* diploid-tetraploid species complex. a) Network representation of genetic pairwise distances (Nei's D) between individuals. *Arabidopsis lyrata* accessions form several clusters outside of Central Europe, named here as NU (pink, diploid/tetraploid), WS (blue, diploid), CS (yellow, diploid/tetraploid), ES (orange, diploid), and Amur Basin (AB; turquoise, diploid). Note that while network clusters may contain multiple ploidy levels, geographic populations are not mixed ploidy. Newly described autotetraploids are found in two distinct clusters (NU/pink and CS/yellow), distinct from Central European autotetraploids (gray), and allotetraploid *A. kamchatica*. b) ASTRAL summary phylogeny of 4,000 gene trees from stLFR assemblies of *A. lyrata* and PacBio assemblies of *A. arenosa*, with *A. thaliana* as outgroup. Each tip is an accession, shape reflects taxon (square = *A. lyrata*, circle = *A. arenosa*) and color denotes ploidy (blue = diploid, orange = tetraploid). Thicker branches have a quartet support of 1.00. Branch lengths reported in coalescent units. Dotted branch subtending the outgroup has been shortened for easier plotting. Color of each clade is consistent with labeling in (a). c) Admixture across the Eurasian *A. lyrata* distribution. Colors correspond to labeling in (a) and (b) and reflect average admixture proportions per sampling site. Tetraploid accessions are found across the range (denoted by black outlines). Extent of the ice sheet during the LGM (Ehlers et al. 2011) shown in white.

consistent with autopolyploidy, indicated by an absence of sites at an intermediate allele frequency (supplementary fig. S6, Supplementary Material online). Allotetraploid populations exhibit a pronounced peak at the intermediate frequency when mapped to a diploid reference due to subgenome divergence, while autotetraploid populations do not show such a peak (Blischak et al. 2023); however, allotetraploids may

lack a peak due to, e.g. multisomic inheritance or widespread homeologous exchange. Allotetraploid *Arabidopsis kamchatica*, originating through hybridizations of Siberian *A. lyrata* and *A. halleri* (Shimizu et al. 2005; Shimizu-Inatsugi et al. 2009; Kolesnikova et al. 2023), is also included in the network and clusters separately from NU and CS autotetraploid *A. lyrata* (Fig. 2a).

We then used TreeMix (Pickrell and Pritchard 2012) to describe relationships among diploid populations. The maximum likelihood model (supplementary fig. S7a, Supplementary Material online) groups Karelia, NU, Western Siberia, Central Siberia, Eastern Siberia, and AB together with suggested migration edge between Karelia and Central Europe, while UK and Scandinavian populations grouped with Central European populations. In the maximum likelihood model including all diploid and tetraploid populations, neither European nor Siberian tetraploids grouped together; in both cases instead grouping with geographically close diploids (supplementary fig. S7b, Supplementary Material online). We also ran TreeMix using the diploid and tetraploid Siberian lineages from phylogenetic analysis in Fig. 2b, and the resulting tree recapitulates the ASTRAL summary phylogeny, with the addition of a migration edge between the tetraploid populations (supplementary fig. S7c, Supplementary Material online). Clustering of individuals using ADMIXTURE (Alexander et al. 2009) with the optimal number of groups $K=8$ (supplementary fig. S8, Supplementary Material online) assigned tetraploid and diploid NU populations into a single cluster (pink, Fig. 2c) with high levels of admixture with diploid WS populations (blue, Fig. 2c); tetraploids and diploids from CS were also assigned into a single cluster (yellow, Fig. 2c) with high levels of admixture with diploid ES cluster (orange, Fig. 2c). We used topology weighting with Twisst (Martin and Van Belleghem 2017) to further explore the relationships between *A. lyrata* lineages when variation in topologies across the genome is taken into account (supplementary fig. S9, Supplementary Material online). The top three highest-weighted topologies are consistent with the independent origin of the two autotetraploid lineages.

To test whether the tetraploids in Siberia have a single origin or independent origins, we used a demographic modeling approach based on fitting simulated and observed SFS (Excoffier et al. 2021). We used four diploid and tetraploid populations (NU 2 $\times$ and NU 4 $\times$, CS 2 $\times$ and CS 4 $\times$) and tested simple divergence models with different topologies, alongside models allowing gene flow between tetraploids and from diploids into tetraploids (supplementary fig. S10, Supplementary Material online). Among all the tested models, likelihood comparisons and AIC supported an independent origin model which recapitulates the ASTRAL summary tree in Fig. 2b (supplementary fig. S10, model c, Supplementary Material online; supplementary table S3, Supplementary Material online). While we exercise caution in the interpretation of parameter estimates, the origin of CS tetraploids is estimated to be much younger than NU tetraploids (~4,000 generations vs. ~30,000 generations; supplementary table S4, Supplementary Material online). We also tested demographic models to compare relative origin times of Central European and NU tetraploids, with resulting divergence time estimates suggesting Central European tetraploids are older by at least an order of magnitude than NU tetraploids (supplementary table S5, Supplementary Material online).

Tetraploid-Specific Adaptive Introgression of Genes Involved in Meiotic Stabilization

Topology weighting results showed that topologies consistent with the ASTRAL tree and demographic analysis were the most frequent (supplementary fig. S9, Supplementary Material online), but a smaller but considerable fraction

indicated introgression between the NU and CS tetraploids (topologies 7, 8, 9, supplementary fig. S9, Supplementary Material online). We combine the topology weighting results for all introgression topologies and show them along the chromosomes (Fig. 3a, upper panels in black). In total, there are 11 windows with a high weighting of gene flow topologies (introgression windows), visible as sharp peaks. Comparing these window boundaries with gene annotation, we identified 37 gene models (supplementary data S4, Supplementary Material online) located within introgression windows. These genes are enriched for gene ontology terms related to meiosis and chromosome organization (supplementary fig. S11 and data S4, Supplementary Material online). Performing the same analysis to identify introgression windows using Central European and NU diploid and tetraploid populations, we find striking overlap of the introgression windows found in the NU and CS comparison (Fig. 3a, lower panels in gray). For each gene within introgression windows, we calculate SNP frequencies across diploid and tetraploid lineages including Central Europe, NU, and CS and visualize frequencies as heatmaps (supplementary fig. S12, Supplementary Material online). We see SNP sharing across the entire sampled range, from Central Europe to Eastern Siberia: in some cases SNPs appear to be tetraploid-specific (e.g. *ZYP1b*, *ASY3*, *MAU2*, and *PDS5*, supplementary fig. S12, d, m, ee, ff, Supplementary Material online) and other SNPs are shared among diploids and tetraploids (e.g. in *CYCA2;3* and *PAPP2C*, supplementary fig. S12, a and c, Supplementary Material online).

Upon seeing which genes fell within introgression windows, we tested enrichment for gene ontology terms among the 37 introgressed genes. The strong enrichment of meiosis-related gene ontology terms (supplementary fig. S11, Supplementary Material online) raises the question of whether introgressed alleles may be undergoing positive selection. To explore this idea, we calculated nucleotide diversity (π) per gene within NU tetraploids and between NU diploids and tetraploids and between CS diploids and tetraploids (dxy) using piawka (<https://github.com/novikovalab/piawka>) and calculated residuals. This “diversity versus differentiation” metric allows us to identify genes that are more differentiated than expected given their within-population diversity (Yant et al. 2013), which could indicate positive selection acting in the tetraploid population (supplementary figs. S13 and S14, Supplementary Material online). Four genes from the introgression windows also fall among outliers in the residuals analysis (supplementary fig. S14, Supplementary Material online), including meiosis genes *ASY3* and *PDS5*, suggesting these loci may be experiencing selection.

Given that introgression genes are enriched for meiotic functions and some of these loci may be under selection, we moved beyond the SNP dataset to examine haplotype sharing across the range of autopolyploid *Arabidopsis lyrata* and *A. arenosa*. We used full haplotype sequences from synthetic long reads (single-tube long fragment reads; stLFR), PacBio long-read assemblies (see Materials and Methods), and published haplotype sequences of Sanger sequenced PCR products (Morgan et al. 2020; Seear et al. 2020) in *A. lyrata* and *A. arenosa*. We focused on three introgressed genes (*ASY3*, *PDS5*, and *ZYP1b*) with substantial published sequence data across *Arabidopsis* genus allowing haplotype-scale alignments for further analyses. We aligned assembled haplotypes from diploid and tetraploid *A. arenosa* and *A. lyrata* and estimated

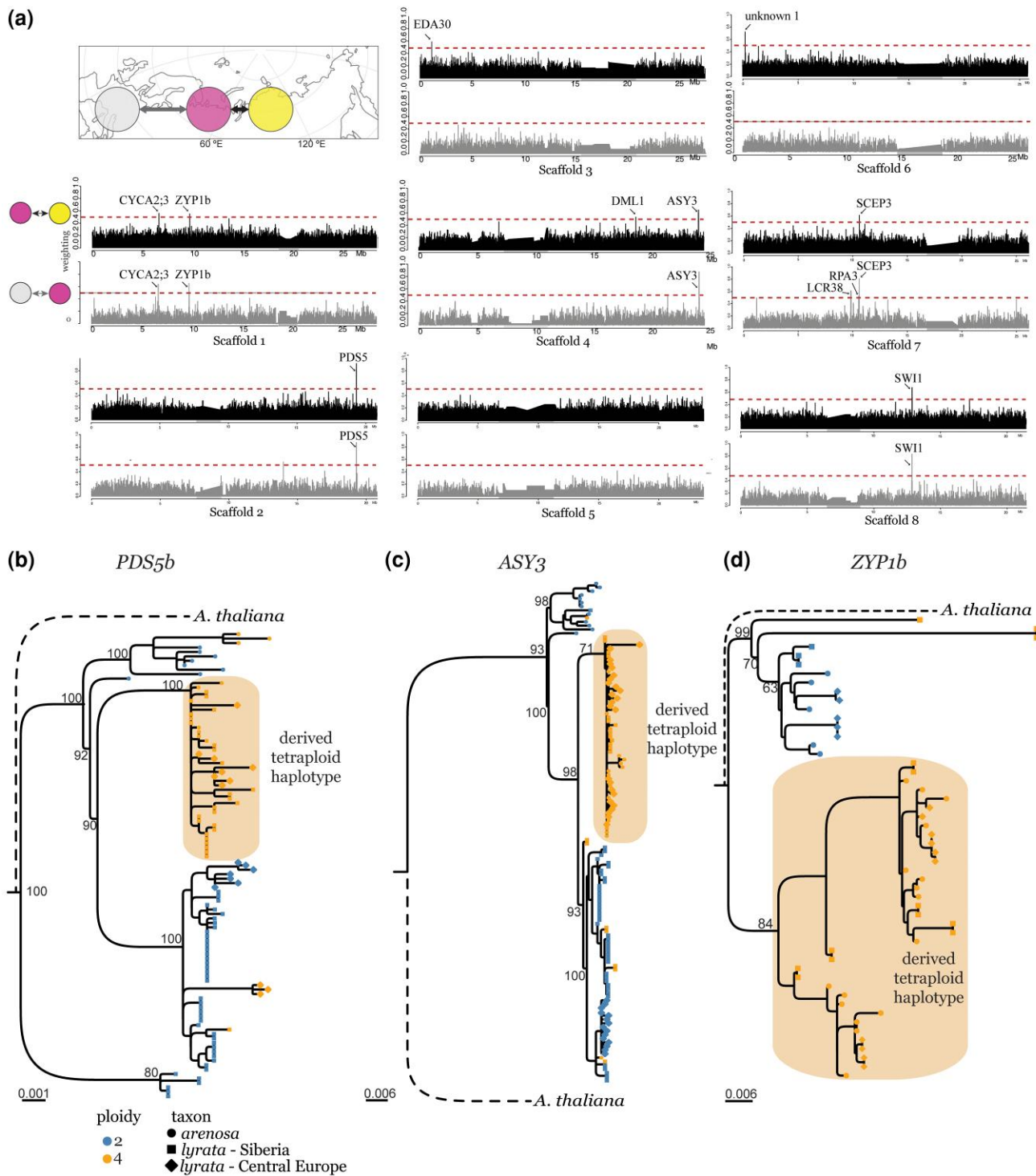


Fig. 3. Genes present in introgression topologies. a) Map showing tetraploid lineages as circles, Grey = Central Europe, Pink = Northern Ural, Yellow = Central Siberian. Each plot is a distribution of introgression topology windows along a chromosome, with upper panels (black) showing introgression windows between NU and CS tetraploids, and lower panels (gray) showing introgression windows between Central Europe and NU tetraploids. Height of peaks indicates topology weighting, and introgression blocks are annotated. Blocks may contain multiple gene models, see [supplementary data S4, Supplementary Material](#) online for full list. Red dashed line indicates 50% weight. b to d) gene trees for haplotypes from *A. lyrata* in Siberia, *A. lyrata* in Central Europe, and *A. arenosa*. Each tip is a haplotype, orange tips are haplotypes from tetraploid individuals, and blue tips from diploids. Shape indicates taxon: squares for *A. lyrata* in Siberia, diamonds for *A. lyrata* in Central Europe, and circles for *A. arenosa*. Orange box highlights derived tetraploid haplotype clade (b) Haplotype tree for *PDS5b* (c) Haplotype tree for *ASY3* (d) Haplotype tree for *ZYP1b*.

phylogenetic trees. Tetraploid-specific haplotypes from *A. arenosa*, Central European, NU, and CS *A. lyrata* form a monophyletic clade at these loci (Fig. 3b-d, orange clade). While the derived tetraploid haplotype is only found in tetraploids and never in diploids, tetraploids can harbor diploid

haplotypes as well. This corresponds well with SNP frequency heatmaps at the same loci ([supplementary fig. S11, d, m, and ff, Supplementary Material](#) online), but is a different topology than the ASTRAL summary tree which is comprised of loci lacking introgression signatures (Fig. 2b).

Discussion

Our range-wide sampling of *A. lyrata* across Eurasia reveals the extent of *A. lyrata* tetraploids (Fig. 1), which was dramatically underestimated until now. Here, we report tetraploids from 30 locations and classify them as autopolyploids, which is apparent from both their deep clustering within *A. lyrata* (Fig. 2) and the absence of fixed heterozygosity due to subgenome divergence (supplementary fig. S6, Supplementary Material online). Autotetraploid populations from the NU region cluster separately, sister to the rest of the Siberian samples, while CS tetraploids are most closely related to local diploids (Fig. 2, supplementary fig. S4b, Supplementary Material online), suggesting different source populations for tetraploid origins. Considering that autotetraploid *A. lyrata* populations in Europe form distinct lineages, perhaps from multiple origins (Bohutínská et al. 2023), and that allotetraploid *A. kamchatica* originated from multiple hybridizations between *A. lyrata* and *A. halleri* (Shimizu-Inatsugi et al. 2009; Schmickl et al. 2010), we conclude that *A. lyrata* is the most polyploid-rich species complex within *Arabidopsis* genus in both auto- and allopolyploid context. The recurrent formation of tetraploids may imply *A. lyrata* (i) has a propensity for WGD, perhaps by an increased likelihood of producing unreduced gametes or increased sporadic somatic WGD and (ii) establishment of *A. lyrata* tetraploids is facilitated by special environmental conditions or genetic background.

Arabidopsis lyrata is the northernmost species in the genus, and its evolutionary history shaped by temperature fluctuations during the Pleistocene (Schmickl et al. 2010). Expansion of *A. lyrata* eastwards from Europe may predate the second to last Ice Age (Ross-Ibarra et al. 2008), which peaked at 130 Kya. Pleistocene temperature oscillations and glaciation cycles explain the main migration routes and secondary contacts in refugia areas, such as Beringia (Schmickl et al. 2010). TreeMix analysis grouped Siberian and Karelian populations together, while UK and Scandinavian populations grouped with Central European populations, consistent with previously suggested different colonization routes for Karelia and Scandinavia, with the former being colonized from NU and the latter from Central Europe, both occurring after the ice sheet retreat post-LGM (Muller et al. 2008; Mattila et al. 2017). Such a pattern of distinct recolonization routes has also been observed in *A. thaliana* (The 1001 Genomes Consortium 2016), and other Arctic species including keystone species such as *Dryas octopetala*, *Vaccinium vitis-idaea*, and *Betula nana* (Eidesen et al. 2013).

Given that *A. lyrata* appears to be the most polyploid-rich taxon in the *Arabidopsis* genus, we again ask: what explains the abundance of polyploids in certain lineages, their emergence at certain times and climates? Our estimations of the divergence times between the diploids and tetraploids in NU and Central Siberia arrive close to the peak of the last Ice Age (supplementary table S4, Supplementary Material online) and follow a general pattern of co-occurrence with glaciations (Novikova et al. 2018). This suggests that environment may play a crucial role in polyploid birth rate, triggering whole-genome duplications via production of 2× gametes: either unreduced gametes formed in a diploid, or normal gametes of a somatically doubled branch. Environmental stress can disturb microtubules and prevent the completion of meiotic cell division (Mason et al. 2011; Sora et al. 2016).

However, as the emergence of polyploids does not guarantee their survival, what contributes to the reduced death rate in

Siberian tetraploid lineages? While environmental disturbance may also facilitate establishment of polyploids, e.g. by opening novel space for colonization after the ice-sheet retreat (Brochmann et al. 2004; Parisod et al. 2010), we see strong evidence of the impact of genetic adaptation. Topology weighting analysis and haplotype sharing from de novo assemblies reveal introgression among distinct *A. lyrata* tetraploid lineages over long distances (Fig. 3). Tetraploid populations in Europe were dated at around 160 Kya (Marburger et al. 2019) and in our demographic modeling results are older than Siberian tetraploid lineages by at least an order of magnitude. Given the colonization history of *A. lyrata* moving from west to east, we infer the direction of the gene flow from Central European *A. lyrata* all the way to Central Siberian *A. lyrata* (Fig. 3). The interspecific introgression from tetraploid *A. arenosa* to tetraploid *A. lyrata* in Central Europe has been shown previously (Marburger et al. 2019). Here, we show that genomic regions introgressed between Central Europe/NU and NU/Central Siberia are narrow and highly enriched in meiotic genes (Fig. 3a, supplementary fig. S11, Supplementary Material online). Among introgressed haplotypes in Siberia, we find the *ASY3* tetraploid-derived allele which has an established functional role (Morgan et al. 2020). In natural tetraploid *A. arenosa*, plants with the derived tetraploid *ASY3* allele in conjunction with a derived *ASY1* allele have fewer multivalents and shorter meiotic axes than plants with the diploid alleles (Morgan et al. 2020), leading to more stable meiosis. Additional research (Seear et al. 2020) on *ASY3* as an adaptation candidate showed an increase in meiotic stability associated with the derived tetraploid *ASY3* haplotype. During meiosis, the length of the chromosomal axis dictates crossover frequency and position (Song et al. 2021; S. Wang et al. 2019). In tetraploids, axis length may be especially important if it serves to limit crossovers to one per chromosome, reducing the risk of multivalent formation (Morgan et al. 2021). Patterns of diversity and divergence in Siberian lineages at the introgressed region harboring *ASY3* gene are consistent with positive selection, suggesting that introgression was adaptive.

Similar patterns of both long-range introgression and positive selection in *A. lyrata* polyploids across Eurasia are observed in the region harboring the cohesin cofactor *PDS5*. In budding yeast, *PDS5* supports sister chromatid cohesion, as mutants suffer from early separation of sister chromatids (Zhang et al. 2005). While experimental work in *A. thaliana* shows that *PDS5* mutants have minimal meiotic phenotypes (Pradillo et al. 2015), suggesting its role in *Arabidopsis* diverges from that in yeast systems, recent work shows that *PDS5* impacts axis length in multiple model systems (Viera et al. 2020; Song et al. 2021; Joo et al. 2022). As described above for *ASY3*, regulation of axis length in tetraploids can help to stabilize meiosis, as a shorter axis can limit crossover number. *PDS5* and its paralogs were repeatedly found in selection scans between diploid and tetraploid lineages in Brassicaceae: first in *A. arenosa* (Hollister et al. 2012; Yant et al. 2013; Wright et al. 2015), then in Central European *A. lyrata* (Marburger et al. 2019), and in *Cardamine amara* (Bohutínská et al. 2021a), where tetraploids harbor a private derived allele, absent from diploid gene pools. We find a high frequency of described adaptive tetraploid alleles in *ASY3* and *PDS5* meiotic genes in NU and CS tetraploids, which suggests that meiotic stabilization is required for polyploid establishment, and that this stabilization is primarily achieved through adaptive introgression.

Siberian *A. lyrata* tetraploids are younger compared to European and it is possible that meiotic stabilization there is incomplete. For example, we find synaptonemal complex protein ZYP1b in introgression scans across Eurasia. The paralogs *ZYP1a* and *ZYP1b* encode transverse filament proteins which are central elements of the synaptonemal complex, helping to “zip” together chromosomal axes of homologs, distribute crossovers and establish proper bivalents (Higgins et al. 2005; Barakate et al. 2014; Capilla-Pérez et al. 2021). Tetraploid alleles of *ZYP1* paralogs are likely adaptive as they appear on introgression and selection scans (Yant et al. 2013; Wright et al. 2015; Marburger et al. 2019). However, in Siberia, although tetraploid Central European alleles of *ZYP1b* are already introduced and detected by introgression scans, their frequency is intermediate (supplementary fig. S11d, Supplementary Material online) and *ZYP1b* does not appear among outliers in diversity and divergence patterns (supplementary fig. S13, Supplementary Material online). The varying patterns across meiotic loci is consistent with previous work showing meiotic adaptation to polyploidy is polygenic without a single key player but rather many players of cumulative effects (Bomblies et al. 2015; Bomblies 2023). Emphasizing the polygenic nature of meiotic adaptation to polyploidy, our introgression windows contain a gene model (SCEP3) recently described as a member of the synaptonemal complex which may work in concert with *ZYP1* in crossover mediation (Heckmann et al. 2024).

It seems that, while environmental fluctuation may yield increased polyploid formation in *A. lyrata*, the persistence of these neopolyploids is due to genetic adaptation involving long-range introgression. This raises the question of how neopolyploids survived initially, without beneficial haplotypes in the population. In contrast to *A. arenosa*, *A. lyrata* is known to propagate clonally (via root sprouts), which could help local neopolyploid persistence in spite of initial unstable meiosis. Environmental conditions could also play a role in early stabilization. For example, it is known that temperature impacts meiosis by modifying the number of crossovers per chromosome (Morgan et al. 2017). It is plausible that vegetative propagation and environmental conditions allowed the initial survival of neopolyploid *A. lyrata* in Siberia. Then, gene flow into the nascent tetraploid populations from the established Central European gene pool introduced the adaptive alleles, enhancing further polyploid establishment and ultimately resulting in further persistence of polyploids in Siberia.

Materials and Methods

Fieldwork and Plant Collection

We used a combination of herbarium specimens and seeds collected from the field for this study. Herbarium samples were provided by the Moscow University Herbarium (MW), the Herbarium of Irkutsk State University (IRKU) and the herbarium of the Siberian Institute of Plant Physiology and Biochemistry. For herbarium accessions, we sampled a small amount of dried tissue from at least one individual plant on a herbarium sheet. Samples from a single collection locale were named using the herbarium accession appended by a number indicating multiple plants that were sequenced from a sheet (e.g. MW0079552_1, MW0079552_2 and MW0079552_3). We collected seeds from wild populations of *A. lyrata* in expeditions to the Lena River Valley, Taz Estuary, Polar Ural, and South Yakutia (supplementary data

S1, Supplementary Material online). We then grew these seeds in the greenhouse (21 °C, 16 h of daylight) and sampled leaf tissue for sequencing.

Short-Read Sequencing for Population Analyses

We used dried leaves either sampled from herbaria, the field, or live plants in the greenhouse. Plant material was processed in two different ways, indicated by types I and II in supplementary data S1, Supplementary Material online. Processing protocols for each type are described in detail in Kolesnikova et al. (2023). Briefly, we isolated DNA from type I herbarium samples using the DNeasy Plant Mini Kit (Qiagen) followed by library prep with the NEBNext DNA Library Prep Kit. For type II samples, we isolated DNA using the NucleoMag Plant” kit from Macherey and Nagel (Düren, Germany) on the KingFisher 96Plex device (Thermo), followed by a TPase-based library prep. Libraries were pooled and sequenced on the NovaSeq 6000 S4 flow cell Illumina system with paired-end 150 bp reads. The Full list of short-read sequenced samples used in the study is in the supplementary data S1, Supplementary Material online.

Linked-Read Sequencing

DNA from 2 g of fresh material was isolated with a NucleoBond HMW DNA kit (Macherey Nagel). Quality, library preparation, and sequencing were performed by BGI as described in O. Wang et al. (2019). Library preparation process starts from random insertions of a transposon sequence into the long fragment DNA, then the product is combined with a magnetic bead carrying a multi-copy molecular barcode. Each DNA molecule receives a unique barcode. After that, fragments are amplified using PCR, and single-stranded cyclized products are used for pair-end sequencing DNBSEQ with read length of 100 bp.

Short Read Mapping and Variant Calling for Population Analysis

We analyzed short paired-end reads (2×150 bp) the same way as in Kolesnikova et al. (2023). The reads were filtered for adapters with bbdut.sh script from BBMap (38.20) (Bushnell 2014) with ktrim=r k=23 mink=11 hdist=1 tbo tpe qtrim=rl trimq=15 minlen=70. Then, we mapped all reads to the NT1 *A. lyrata* reference genome (Kolesnikova et al. 2023) using bwa mem (0.7.17) (Li and Durbin 2009) with -M parameter. We used Picard MarkDuplicates (Broad Institute) to mark PCR duplicated reads and then used Samtools (Li et al. 2009) to sort and index the BAM files. We called variants with HaplotypeCaller algorithm from GATK (3.8) (McKenna et al. 2010) and then combined individual vcf files using CombineGVCFs and created final vcf using GenotypeGVCF from GATK including non-variant sites with—includeNonVariantSites parameter. We estimated heterozygosity on a vcf with variants called for all samples as for diploids by calculating proportion of heterozygous sites among all called sites. For the further analysis, we created a vcf with variant and non-variant sites called according to the ploidy of the samples such that tetraploid individuals are genotyped as tetraploids mapped to a diploid reference (HaplotypeCaller with—sample-ploidy 2 or—sample-ploidy 4). Autotetraploids (*A. lyrata*) and allotetraploids alike (*A. kamchatica*) were mapped in full to the diploid reference and not split by subgenome.

Synthetic Long-Read Assemblies and Tree Inference

StLFR reads were assembled using the stlfr2supernova pipeline (https://github.com/BGI-Qingdao/stlfr2supernova_pipeline) which is based on Supernova (Weisenfeld et al. 2017) with—style=pseudohap2 option to get two haplotypes for each assembly. The haplotypes were scaffolded on the NT1 *A. lyrata* genome (Kolesnikova et al. 2023) with ragtag (Alonge et al. 2022). Long-read *A. arenosa* assemblies from (Bohutínská et al. 2024) were also scaffolded on *A. lyrata* reference and further used together with stLFR assemblies. The GALBA pipeline (Stanke et al. 2006; Buchfink et al. 2015; Hoff et al. 2019; Hoff and Stanke 2019; Li 2023) was used for protein annotation of the assemblies with proteins from Phytozome (Goodstein et al. 2012) *A. lyrata* v.2.1 release (Rawat et al. 2015) as a protein database. The annotated genes were used for gene tree construction. Protein BLAST (Camacho et al. 2009) was used to identify genes of interest in the annotation. Then, we used L-INS-i algorithm of MAFFT (Katoh et al. 2005) for multiple sequence alignment of the gene regions. The alignments were manually curated in UGENE (<https://doi.org/10.1093/bioinformatics/bts091>) and used for tree construction with iqtree (Nguyen et al. 2015; Hoang et al. 2018), fast bootstrap=1,000.

For species tree inference from gene trees we used full-length BUSCO gene sequences (Simão et al. 2015). Predictions with the same ID from all the assemblies were aligned with L-INS-i after filtering for length to get rid of incomplete models. Gene trees produced by IQ-TREE were then summarized with wASTRAL (<https://doi.org/10.1093/molbev/msac215>) from the ASTER package. This method takes into account tree support and branch length values.

Ploidy Estimation and S-Allele Genotyping From Short-Read Sequencing Data

To estimate ploidy, we used nQuire (Weiß et al. 2018) with “create,” “denoise,” and “lrdmodel” steps. This ploidy estimation method is based on allelic read depth distributions, also previously described (Scott et al. 2023). We identified S-alleles as described in Kolesnikova et al. (2023) using short-read sequencing data and S-locus genotyping pipeline NGSgenotyp (Genete et al. 2020) with two first steps: kmerfilter (−k 20 for SRK and −k 15 for SCR) and genotyp—with both SRK and SCR sequence reference databases. The databases used for genotyping also contain known SRK- and SCR-like, but unlinked genetic regions, that are used to normalize the coverage over the S-alleles and are later excluded from the genotyping results as false-positives. For every individual we used normalized S-allele coverage to make sure that all S-alleles were genotyped and count copy numbers for each S-allele. The number of different S-alleles in each individual together with proportion of heterozygous sites were used to confirm the estimated ploidy level. In tetraploid individuals with only two S-alleles the normalized depth over the S-alleles was sometimes unequal (3:1). This fact provided evidence that there were three copies of one allele and one copy of another and was also used to confirm ploidy. In addition to the databases used previously (Kolesnikova et al. 2023), we have assembled four SRK alleles from short-read data with assembly option and provide them together with 6 more published (Castric et al. 2008; Neuffer et al. 2023) SRK alleles in [supplementary table S6, Supplementary Material](#) online. The sequences of S-alleles are provided in

the [supplementary data S3, Supplementary Material](#) online in .fasta format.

Network Analysis and Population Admixture

To obtain a network of genetic distances between individuals, we followed Kolesnikova et al. (2023). Briefly, we filtered our biallelic vcf down to only fourfold degenerate sites. We then calculated Nei’s genetic distance (D) (Nei 1972) on the filtered vcf in R (using StAMPP; Pembleton et al. 2013), and used SplitsTree4 (Huson and Bryant 2006) to visualize the resulting distance matrix. We ran a series of ADMIXTURE (Gutenkunst et al. 2009) analyses with $K=2$ up to $K=15$. Based on cross-validation error ([supplementary fig. S8, Supplementary Material](#) online), $K=8$ was the optimal number of groups. To get population-level clustering, we combined clustering proportions for accessions from the same geographical coordinate and took the mean proportion. To accommodate for uneven sampling where some populations are represented by more individuals than others, we took three random representatives from populations with more than three accessions. Note that ADMIXTURE is not designed for mixed-ploidy datasets, and while there is the risk of spurious clustering by ploidy level rather than population (see Stift et al. 2019 for a thorough discussion), we do not see such a pattern with our dataset, instead observing clustering based on geography.

Geographical Map

We used R v4.2.2 (R Core Team 2022), primarily with the tidyverse v2.0.0 package collection (Wickham et al. 2019), for making the plots. Maps were created using Lambert Azimuthal Equal Area projection centered at N 40° E 90°, with packages sf v1.0-14 (Pebesma 2018) and raster v3.6-23 (Hijmans 2023) and a background map by NaturalEarth; the ggrepel algorithm (Słowikowski 2023) was used to adjust the positioning of the pie charts. The Last Glacial Maximum (LGM) ice sheet extents (Ehlers et al. 2011) were simplified for plotting.

Cytogenetic Validation of Ploidy

To confirm ploidy levels estimated by bioinformatic methods, we used meiotic spreads. We first harvested immature buds (prior to pollen being formed) and fixed them in Carnoy fixative (3:1 ethanol:acetic acid) for 2 h on a horizontal shaker at room temperature, changed to fresh Carnoy fixative, and stored them at −20 °C until further usage.

Fixed buds were rinsed twice for 5 min in 10 mM citrate buffer (tri-sodium citrate, pH 4.5 with HCl), followed by a digestion with an enzymatic solution (0.3% Pectolyase Y-23, 0.3% Driselase, 0.3% Cellulase Onozuka R 10 and 0.1% sodium azide in 10 mM citrate buffer) at 37 °C for 1 h in a humid chamber. After another two washes with citrate buffer, the anthers were separated from surrounding tissue and put onto a slide. One edge of an anther was cut open and the cells within were squeezed out with the tip of a needle. After processing all anthers of one to two buds, cells were mixed into a suspension by stirring with a needle. A small amount of freshly prepared 60% acetic acid was added to the suspension from the side and the whole drop gently homogenized by again stirring with a needle. The slide was placed on a hot plate at 45 °C and cells were spread with the help of a bent needle until almost all liquid evaporated. Around the spread area, a

boundary of ice cold fresh fixative solution was created and allowed to enter, followed by a jet of ice cold fixative directly into the center of the area. Then the slide was allowed to air dry in an upright position. Chromosomal spreads were stained with DAPI in a solution of 2 µg/ml in the antifade mounting medium Vectashield and covered with a coverslip. Imaging was done on a Zeiss Axio Imager Z2 Microscope using the ZEN software. Ploidy validation data shown in [supplementary data S2, Supplementary Material](#) online.

Estimation of Nucleotide Diversity

We calculated π (Nei and Li 1979) for diploid and tetraploid live populations with more than five individuals in each (excluding siblings) using piawka (<https://github.com/novikovalab/piawka>), our implementation of the site weighting strategy introduced in pixy (Korunes and Samuk 2021) compatible with mixed-ploidy datasets. We also calculated π within each tetraploid lineage and between diploid-tetraploid lineages on a per-gene basis with piawka.

Introgression Analysis

Starting with a VCF containing only biallelic SNPs, we filtered the dataset further using genomics_general-0.4 toolkit (minimum quality = 20, minimum depth = 5, maximum missing data = 0.2, maximum heterozygosity = 0.75, minimum variant count = 5) (https://github.com/simonhmartin/genomics_general). Genealogies were inferred in windows of 50 SNPs along the genome using PHYML (Guindon et al. 2010) and a phasing approach implemented in genomics_general. Previously published (Novikova et al. 2016) samples of *Arabidopsis cebennensis* and *Arabidopsis pedemontana* were used as outgroups. We used Twisst (Martin and Van Belleghem 2017) for topology weighting. GO enrichment analysis using g:Profiler (Kolberg et al. 2023) was performed for genes intersecting the regions with >50% weighting of introgression topologies.

TreeMix (Pickrell and Pritchard 2012) was used with 500 SNPs per block, 4 allowed migration events, and 1,000 bootstrap replicates.

Demographic Modeling

To generate SFS, we utilized biallelic SNPs and invariant sites. We excluded genic, centromeric, and pericentromeric regions, retaining only SNPs found in intergenic regions of chromosome arms. Individuals of the two main lineages were selected as outlined in [supplementary data S1, Supplementary Material](#) online. Using the chosen SNPs, we generated bootstrap replicates by dividing the genome into 10 kb segments. These segments were resampled with replacement until the equivalent size of the total amount of filtered SNPs in the genome. We computed site frequency spectrum (SFS) for each population of the 200 bootstrapped replicates and a joint site frequency spectrum for each pseudo-observation using the Python package dadi (Gutenkunst et al. 2009). The SFS spectra generated by dadi were used as input into fastsimcoal2 (fsc27) (Excoffier et al. 2021). We first assessed a simple divergence model for both lineages through 50 fastsimcoal2 runs, then compared with a model allowing migration from the diploid into tetraploid lineage. Each fastsimcoal2 run was allowed 40 optimization cycles and 100,000 iterations. Based on likelihood and AIC criteria we continued with the divergence mode with migration for the confidence interval calculation. For each of the

200 bootstrap replicates we initiated 50 fastsimcoal2 runs. The resulting parameter estimates from the 200 pseudo-observations were used to calculate the 95% confidence intervals in R.

Supplementary Material

Supplementary material is available at [Molecular Biology and Evolution](#) online.

Acknowledgments

We thank the numerous collectors of the herbarium material we used and the herbarium curators A.V. Verkhovina (IRK) and N.V. Stepanova (IRKU) who kindly provided access to some collections. We are grateful to Igor Pospelov (FGBU “Zapovedniki Taimyra”) and Elena Pospelova for the live samples from the Putorana massif and to Daiana Zhernova for the live samples from the White Sea coast. Our fieldwork was additionally supported by the Austrian Science Fund (FWF grant P30208-B-29) and held within the state assignment of the Papanin Institute for Biology of Inland Waters Russian Academy of Sciences (theme 121051100099-5). We also owe its success to many locals and other friends who provided us shelter and transport. Special thanks to Tatiana Kirillina, Matryona Achikaso-Koryakina, Djulustan Nikitin, and Chekurovka camp fishermen, Aleksandr Markevich, Manchaary Vladimirov, Stolz weather station staff, Khokhuccha camp fishermen, staff of Zhigansk and Bykovsky schools, Leonid Shaglanov, Tiksi Arctic Gymnasium staff, Grigorii from Sagastyr, Aleksandr Gukov, Andrey Kuzkin, and Yevgeny Serebrennikov (Lena river area, 2019 to 2020); Vladimir Segoy together with Samburg firefighters, Prokopy Tesida, staff of the Messo-Yakha protected area, Aleksey Yar, and nenets communities of Salyakaptan, Salyapayuta and Yabtasalya areas (West Siberia, 2021); and Faina and her family from the Kharp town (Polar Ural, 2021). We thank the Max Planck-Genome-Centre Cologne, Germany (<http://mpgc.mpiiz.mpg.de/home/>) for performing short-read libraries in this study. We thank Levi Yant and Andrew C. Clarke for performing DNA extractions and sequencing of a subset of the herbarium samples.

Funding

The work of A.P.S. on the curation of dry plant material was supported by the Russian Science Foundation (project #21-77-20042). The sequencing and data analysis were funded by the European Union (ERC, HOW2DOUBLE, 101041354) and by the joint Deutsche Forschungsgemeinschaft (DFG) and GACR program—project number 490698526. This project was supported by the Deutscher Akademischer Austauschdienst (DAAD) fund 57601834 to P.Y.N. This research was also funded by the Czech Science Foundation (project 22-29078K) to F.K. P.Y.N., A.D.S., and N.P.T. thank the DFG-funded research consortium TRR341 for support and discussion. Funded by the European Union. The views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability

The whole-genome re-sequencing short reads data for the samples which was generated in this study were submitted to the ENA database under study numbers PRJEB67879 and PRJEB60410. The list of all the samples used in the study with individual ENA numbers is in [supplementary data S1, Supplementary Material](#) online. The images of the herbarium vouchers can be found online using an example link <https://plant.depo.msu.ru/open/public/item/MW0079552> and replacing MW0079552 with the individual number from [supplementary data S1, Supplementary Material](#) online column “accession”. Both raw stLFR data and resulting assemblies are also accessible in the ENA database, with individual ENA numbers in [supplementary data S5, Supplementary Material](#) online. [supplementary data S1 to S5, Supplementary Material](#) online and the scripts used in the analysis are available at https://github.com/novikovalab/Siberian_Alyrata.

References

- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 2022;23(1):258. <https://doi.org/10.1186/s13059-022-02823-7>.
- Ansell SW, Stenøien HK, Grundmann M, Schneider H, Hemp A, Bauer N, Russell SJ, Vogel JC. Population structure and historical biogeography of European *Arabidopsis lyrata*. *Heredity (Edinb.)*. 2010;105(6):543–553. <https://doi.org/10.1038/hdy.2010.10>.
- Arnold B, Kim ST, Bomblies K. Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Mol Biol Evol.* 2015;32(6):1382–1395. <https://doi.org/10.1093/molbev/msv089>.
- Barakate A, Higgins JD, Vivera S, Stephens J, Perry RM, Ramsay L, Colas I, Oakley H, Waugh R, Franklin FCH, et al. The synaptonemal complex protein ZYP1 is required for imposition of meiotic crossovers in barley. *Plant Cell.* 2014;26(2):729–740. <https://doi.org/10.1105/tpc.113.121269>.
- Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 2016;210(2):391–398. <https://doi.org/10.1111/nph.13698>.
- Blischak PD, Sajjan M, Barker MS, Gutenkunst RN. Demographic history inference and the polyploid continuum. *Genetics.* 2023;224(4):iyad107. <https://doi.org/10.1093/genetics/iyad107>.
- Bohutínská M, Alston M, Monnahan P, Mandáková T, Bray S, Paajanen P, Kolář F, Yant L. Novelty and convergence in adaptation to whole genome duplication. *Mol Biol Evol.* 2021a;38(9):3910–3924. <https://doi.org/10.1093/molbev/msab096>.
- Bohutínská M, Handrick V, Yant L, Schmickl R, Kolář F, Bomblies K, Paajanen P. De novo mutation and rapid protein (co-)evolution during meiotic adaptation in *Arabidopsis arenosa*. *Mol Biol Evol.* 2021b;38(5):1980–1994. <https://doi.org/10.1093/molbev/msab001>.
- Bohutínská M, Petříková E, Booker TR, Cobo CV, Vlček J, Šrámková G, Poštulková A, Hojka J, Marhold K, Yant L, et al. Mosaic haplotypes underlie repeated adaptation to whole genome duplication in *Arabidopsis lyrata* and *Arabidopsis arenosa*. *bioRxiv* 2023. 01.11.523565. <https://www.biorxiv.org/content/10.1101/2023.01.11.523565v2>, March 02, 2023, preprint: not peer reviewed.
- Bohutínská M, Petříková E, Booker TR, Vives Cobo C, Vlček J, Šrámková G, Poupětová A, Hojka J, Marhold K, Yant L, et al. Polyploids broadly generate novel haplotypes from trans-specific variation in *Arabidopsis arenosa* and *Arabidopsis lyrata*. *PLoS Genet.* 2024;20(12):e1011521. <https://doi.org/10.1371/journal.pgen.1011521>.
- Bomblies K. Learning to tango with four (or more): the molecular basis of adaptation to polyploid meiosis. *Plant Reprod.* 2023;36(1):107–124. <https://doi.org/10.1007/s00497-022-00448-1>.
- Bomblies K, Higgins JD, Yant L. Meiosis evolves: adaptation to external and internal environments. *New Phytol.* 2015;208(2):306–323. <https://doi.org/10.1111/nph.13499>.
- Bomblies K, Jones G, Franklin C, Zickler D, Kleckner N. The challenge of evolving stable polyploidy: could an increase in “crossover interference distance” play a central role? *Chromosoma.* 2016;125(2):287–300. <https://doi.org/10.1007/s00412-015-0571-4>.
- Broad Institute. Picard. [accessed 2024 Aug 27]. Available from: <http://broadinstitute.github.io/picard/>
- Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen AC, Elven R. Polyploidy in Arctic plants. *Biol J Linn Soc Lond.* 2004;82(4):521–536. <https://doi.org/10.1111/j.1095-8312.2004.00337.x>.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.
- Burns R, Mandáková T, Gunis J, Soto-Jiménez LM, Liu C, Lysak MA, Novikova PY, Nordborg M. Gradual evolution of allopolyploidy in *Arabidopsis suecica*. *Nat Ecol Evol.* 2021;5(10):1367–1381. <https://doi.org/10.1038/s41559-021-01525-w>.
- Bushnell B. BB tools software package; 2014 [accessed 2024 Jan 8]. Available from: <http://btools.jgi.doe.gov>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10(1):421. <https://doi.org/10.1186/1471-2105-10-421>.
- Capilla-Pérez L, Durand S, Hurel A, Lian Q, Chambon A, Taochy C, Solier V, Grelon M, Mercier R. The synaptonemal complex imposes crossover interference and heterochiasmy in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2021;118(12):e2023613118. <https://doi.org/10.1073/pnas.2023613118>.
- Castric V, Bechsgaard J, Schierup MH, Vekemans X. Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genet.* 2008;4(8):e1000168. <https://doi.org/10.1371/journal.pgen.1000168>.
- Castric V, Vekemans X. Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Mol Ecol.* 2004;13(10):2873–2889. <https://doi.org/10.1111/j.1365-294X.2004.02267.x>.
- Chao DY, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science.* 2013;341(6146):658–659. <https://doi.org/10.1126/science.1240561>.
- Charlesworth D, Mable BK, Schierup MH, Bartolomé C, Awadalla P. Diversity and linkage of genes in the self-incompatibility gene family in *Arabidopsis lyrata*. *Genetics.* 2003;164(4):1519–1535. <https://doi.org/10.1093/genetics/164.4.1519>.
- Clauss MJ, Mitchell-Olds T. Population genetic structure of *Arabidopsis lyrata* in Europe. *Mol Ecol.* 2006;15(10):2753–2766. <https://doi.org/10.1111/j.1365-294X.2006.02973.x>.
- Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell.* 2000;12(9):1551–1568. <https://doi.org/10.1105/tpc.12.9.1551>.
- Dart S, Kron P, Mable BK. Characterizing polyploidy in *Arabidopsis lyrata* using chromosome counts and flow cytometry. *Can J Bot.* 2004;82(2):185–197. <https://doi.org/10.1139/b03-134>.
- David KT. Global gradients in the distribution of animal polyploids. *Proc Natl Acad Sci U S A.* 2022;119(48):e2214070119. <https://doi.org/10.1073/pnas.2214070119>.
- De Storme N, Copenhaver GP, Geelen D. Production of diploid male gametes in *Arabidopsis* by cold-induced destabilization of postmeiotic radial microtubule arrays. *Plant Physiol.* 2012;160(4):1808–1826. <https://doi.org/10.1104/pp.112.208611>.
- Ehlers J, Gibbard PL, Hughes PDM, editors. *Quaternary glaciations—extent and chronology: volume 15: a closer look*. London (England): Elsevier Science; 2011.

- Eidesen PB, Ehrich D, Bakkestuen V, Alsos IG, Gilg O, Taberlet P, Brochmann C. Genetic roadmap of the Arctic: plant dispersal highways, traffic barriers and capitals of diversity. *New Phytol.* 2013;200(3):898–910. <https://doi.org/10.1111/nph.12412>.
- Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. Fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics.* 2021;37(24):4882–4885. <https://doi.org/10.1093/bioinformatics/btab468>.
- Genete M, Castric V, Vekemans X. Genotyping and de novo discovery of allelic variants at the Brassicaceae self-incompatibility locus from short-read sequencing data. *Mol Biol Evol.* 2020;37(4):1193–1201. <https://doi.org/10.1093/molbev/msz258>.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012;40(D1):D1178–D1186. <https://doi.org/10.1093/nar/gkr944>.
- Gregory TR, Mable BK. CHAPTER 8—polyploidy in animals. In: Gregory TR, editor. *The evolution of the genome*. Burlington: Academic Press; 2005. p. 427–517.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–321. <https://doi.org/10.1093/sysbio/syq010>.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 2009;5(10):e1000695. <https://doi.org/10.1371/journal.pgen.1000695>.
- Hämälä T, Mattila TM, Leinonen PH, Kuittinen H, Savolainen O. Role of seed germination in adaptation and reproductive isolation in *Arabidopsis lyrata*. *Mol Ecol.* 2017;26(13):3484–3496. <https://doi.org/10.1111/mec.14135>.
- Heckmann S, Feng C, Lorenz J, Dreissig S, Schubert V, Wang B, Hartmann F, Cuacos M, Fernández-Jiménez N, Cámara AS, et al. The novel synaptonemal complex central element SCEP3 interlinks synapsis initiation and crossover formation in *Arabidopsis thaliana*; 2024 [accessed 2025 Mar 25]. Available from: <https://www.researchsquare.com/article/rs-5394998/latest.pdf>
- Higgins JD, Sanchez-Moran E, Armstrong SJ, Jones GH, Franklin FCH. The *Arabidopsis* synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. *Genes Dev.* 2005;19(20):2488–2500. <https://doi.org/10.1101/gad.354705>.
- Hijmans RJ. raster: Geographic data analysis and modeling; 2023 [accessed 2024 Jan 8]. Available from: <https://CRAN.R-project.org/package=raster>
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 2018;35(2):518–522. <https://doi.org/10.1093/molbev/msx281>.
- Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. *Methods Mol Biol.* 2019;1962:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.
- Hoff KJ, Stanke M. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc Bioinformatics.* 2019;65(1):e57. <https://doi.org/10.1002/cpbi.57>.
- Hohmann N, Koch MA. An *Arabidopsis* introgression zone studied at high spatio-temporal resolution: interglacial and multiple genetic contact exemplified using whole nuclear and plastid genomes. *BMC Genomics.* 2017;18(1):810. <https://doi.org/10.1186/s12864-017-4220-6>.
- Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet.* 2012;8(12):e1003093. <https://doi.org/10.1371/journal.pgen.1003093>.
- Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006;23(2):254–267. <https://doi.org/10.1093/molbev/msj030>.
- Joo JH, Kang HA, Kim KP, Hong S. Meiotic prophase roles of Pds5 in recombination and chromosome condensation in budding yeast. *J Microbiol.* 2022;60(2):177–186. <https://doi.org/10.1007/s12275-022-1635-9>.
- Jørgensen MH, Ehrich D, Schmickl R, Koch MA, Brysting AK. Interspecific and interploidal gene flow in central European *Arabidopsis* (Brassicaceae). *BMC Evol Biol.* 2011;11(1):346. <https://doi.org/10.1186/1471-2148-11-346>.
- Katche EI, Schierholt A, Schiessl S-V, He F, Lv Z, Batley J, Becker HC, Mason AS. Genetic factors inherited from both diploid parents interact to affect genome stability and fertility in resynthesized allotetraploid *Brassica napus*. G3 (Bethesda, Md.). 2023;13(8):jkad136. <https://doi.org/10.1093/g3journal/jkad136>.
- Katoh K, Kuma K-I, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005;33(2):511–518. <https://doi.org/10.1093/nar/gki198>.
- Kauai F, Mortier F, Milosavljevic S, Van de Peer Y, Bonte D. Neutral processes underlying the macro eco-evolutionary dynamics of mixed-ploidy systems. *Proc Biol Sci.* 2023;290(1995):20222456. <https://doi.org/10.1098/rspb.2022.2456>.
- Khosho TN. Polyploidy in gymnosperms. *Evolution.* 1959;13(1):24–39. <https://doi.org/10.2307/2405943>.
- Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. G: Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* 2023;51(W1):W207–W212. <https://doi.org/10.1093/nar/gkad347>.
- Kolesnikova UK, Scott AD, Van de Velde JD, Burns R, Tikhomirov NP, Pfordt U, Clarke AC, Yant L, Seregin AP, Vekemans X, et al. Transition to self-compatibility associated with dominant S-allele in a diploid Siberian progenitor of allotetraploid *Arabidopsis kamchatica* revealed by *Arabidopsis lyrata* genomes. *Mol Biol Evol.* 2023;40(7):msad122. <https://doi.org/10.1093/molbev/msad122>.
- Korunes KL, Samuk K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour.* 2021;21(4):1359–1368. <https://doi.org/10.1111/1755-0998.13326>.
- Kreiner JM, Kron P, Husband BC. Frequency and maintenance of unreduced gametes in natural plant populations: associations with reproductive mode, life history and genome size. *New Phytol.* 2017a;214(2):879–889. <https://doi.org/10.1111/nph.14423>.
- Kreiner JM, Kron P, Husband BC. Evolutionary dynamics of unreduced gametes. *Trends Genet.* 2017b;33(9):583–593. <https://doi.org/10.1016/j.tig.2017.06.009>.
- Lafon-Placette C, Johannessen IM, Hornslien KS, Ali MF, Bjerkan KN, Bramsiepe J, Glockle BM, Rebernik CA, Brysting AK, Grini PE, et al. Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc Natl Acad Sci U S A.* 2017;114(6):E1027–E1035. <https://doi.org/10.1073/pnas.1615123114>.
- Lafon-Placette C, Kohler C. Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol Ecol.* 2016;25(11):2620–2629. <https://doi.org/10.1111/mec.13552>.
- Le Veve A, Burghgraeve N, Genete M, Lepers-Blassiau C, Takou M, De Meaux J, Mable BK, Durand E, Vekemans X, Castric V. Long-term balancing selection and the genetic load linked to the self-incompatibility locus in *Arabidopsis halleri* and *A. lyrata*. *Mol Biol Evol.* 2023;40(6):msad120. <https://doi.org/10.1093/molbev/msad120>.
- Li H. Protein-to-genome alignment with miniprot. *Bioinformatics.* 2023;39(1):btad014. <https://doi.org/10.1093/bioinformatics/btad014>.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Llaurens V, Billiard S, Leducq J-B, Castric V, Klein EK, Vekemans X. Does frequency-dependent selection with complex dominance interactions accurately predict allelic frequencies at the self-incompatibility locus in *Arabidopsis halleri*? *Evolution.* 2008;62(10):2545–2557. <https://doi.org/10.1111/j.1558-5646.2008.00469.x>.

- Lohaus R, Van de Peer Y. Of dups and dinos: evolution at the K/Pg boundary. *Curr Opin Plant Biol*. 2016;30:62–69. <https://doi.org/10.1016/j.pbi.2016.01.006>.
- Mable BK, Beland J, Di Berardo C. Inheritance and dominance of self-incompatibility alleles in polyploid *Arabidopsis lyrata*. *Heredity (Edinb)*. 2004;93(5):476–486. <https://doi.org/10.1038/sj.hdy.6800526>.
- Mable BK, Schierup MH, Charlesworth D. Estimating the number, frequency, and dominance of S-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity (Edinb)*. 2003;90(6):422–431. <https://doi.org/10.1038/sj.hdy.6800261>.
- Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman C, Raskin E, Mitchell-Olds T. Environmental aridity is associated with cytotype segregation and polyploidy occurrence in *Brachypodium distachyon* (Poaceae). *New Phytol*. 2012;193(3):797–805. <https://doi.org/10.1111/j.1469-8137.2011.03988.x>.
- Marburger S, Monahan P, Seear PJ, Martin SH, Koch J, Paajanen P, Bohutínská M, Higgins JD, Schmickl R, Yant L. Interspecific introgression mediates adaptation to whole genome duplication. *Nat Commun*. 2019;10(1):5218. <https://doi.org/10.1038/s41467-019-13159-5>.
- Martin SH, Van Belleghem SM. Exploring evolutionary relationships across the genome using topology weighting. *Genetics*. 2017;206(1):429–438. <https://doi.org/10.1534/genetics.116.194720>.
- Mason AS, Nelson MN, Yan G, Cowling WA. Production of viable male unreduced gametes in Brassica interspecific hybrids is genotype specific and stimulated by cold temperatures. *BMC Plant Biol*. 2011;11(1):103. <https://doi.org/10.1186/1471-2229-11-103>.
- Mason AS, Pires JC. Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends Genet*. 2015;31(1):5–10. <https://doi.org/10.1016/j.tig.2014.09.011>.
- Mattila TM, Tyrmi J, Pyhajarvi T, Savolainen O. Genome-wide analysis of colonization history and concomitant selection in *Arabidopsis lyrata*. *Mol Biol Evol*. 2017;34(10):2665–2677. <https://doi.org/10.1093/molbev/msx193>.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytisky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303. <https://doi.org/10.1101/gr.107524.110>.
- Monahan P, Kolář F, Baduel P, Sailer C, Koch J, Horvath R, Laenen B, Schmickl R, Paajanen P, Šrámková G, et al. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat Ecol Evol*. 2019;3(3):457–468. <https://doi.org/10.1038/s41559-019-0807-4>.
- Morgan C, White MA, Franklin FCH, Zickler D, Kleckner N, Bomblies K. Evolution of crossover interference enables stable autopolyploidy by ensuring pairwise partner connections in *Arabidopsis arenosa*. *Curr Biol*. 2021;31(21):4713–4726.e4. <https://doi.org/10.1016/j.cub.2021.08.028>.
- Morgan C, Zhang H, Henry CE, Franklin FCH, Bomblies K. Derived alleles of two axis proteins affect meiotic traits in autotetraploid *Arabidopsis arenosa*. *Proc Natl Acad Sci U S A*. 2020;117(16):8980–8988. <https://doi.org/10.1073/pnas.1919459117>.
- Morgan CH, Zhang H, Bomblies K. Are the effects of elevated temperature on meiotic recombination and thermotolerance linked via the axis and synaptonemal complex? *Philos Trans R Soc Lond B Biol Sci*. 2017;372(1736):20160470. <https://doi.org/10.1098/rstb.2016.0470>.
- Muller M-H, Leppälä J, Savolainen O. Genome-wide effects of postglacial colonization in *Arabidopsis lyrata*. *Heredity (Edinb)*. 2008;100(1):47–58. <https://doi.org/10.1038/sj.hdy.6801057>.
- Nei M. Genetic distance between populations. *Am Nat*. 1972;106(949):283–292. <https://doi.org/10.1086/282771>.
- Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A*. 1979;76(10):5269–5273. <https://doi.org/10.1073/pnas.76.10.5269>.
- Neuffer B, Bechsgaard J, Paetsch M, Titel C, Wesse C, Bona E, Schimpf R, Žerdoner Čalasan A, Hurka H. S-alleles and mating system in natural populations of *Capsella grandiflora* (Brassicaceae) and its congeneric relatives. *Flora*. 2023;299:152206. <https://doi.org/10.1016/j.flora.2022.152206>.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–274. <https://doi.org/10.1093/molbev/msu300>.
- Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, Guggisberg A, Paape T, Schmid K, Fedorenko OM, et al. Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nat Genet*. 2016;48(9):1077–1082. <https://doi.org/10.1038/ng.3617>.
- Novikova PY, Hohmann N, Van de Peer Y. Polyploid *Arabidopsis* species originated around recent glaciation maxima. *Curr Opin Plant Biol*. 2018;42:8–15. <https://doi.org/10.1016/j.pbi.2018.01.005>.
- Parisod C, Holderegger R, Brochmann C. Evolutionary consequences of autopolyploidy. *New Phytol*. 2010;186(1):5–17. <https://doi.org/10.1111/j.1469-8137.2009.03142.x>.
- Pebesma E. Simple features for R: standardized support for spatial vector data. *R J*. 2018;10(1):439. <https://doi.org/10.32614/RJ-2018-009>.
- Pembleton LW, Cogan NOI, Forster JW. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol Ecol Resour*. 2013;13(5):946–952. <https://doi.org/10.1111/1755-0998.12129>.
- Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012;8(11):e1002967. <https://doi.org/10.1371/journal.pgen.1002967>.
- Polatschek A. Cytotaxonomische Beiträge zur Flora der Ostalpenländer, I. *Österreichische Botanische Zeitschrift*. 1966;113(1):1–46. <https://doi.org/10.1007/BF01373246>.
- Pradillo M, Knoll A, Oliver C, Varas J, Corredor E, Puchta H, Santos JL. Involvement of the cohesin cofactor PDS5 (SPO76) during meiosis and DNA repair in *Arabidopsis thaliana*. *Front Plant Sci*. 2015;6:1034. <https://doi.org/10.3389/fpls.2015.01034>.
- Rawat V, Abdelsamad A, Pietzenek B, Seymour DK, Koenig D, Weigel D, Pecinka A, Schneeberger K. Improving the annotation of *Arabidopsis lyrata* using RNA-Seq data. *PLoS One*. 2015;10(9):e0137391. <https://doi.org/10.1371/journal.pone.0137391>.
- R Core Team. *R: a language and environment for statistical computing*; 2022. Available from: <https://www.R-project.org/>
- Rice A, Šmarda P, Novosolov M, Drori M, Glick L, Sabath N, Meiri S, Belmaker J, Mayrose I. The global biogeography of polyploid plants. *Nat Ecol Evol*. 2019;3(2):265–273. <https://doi.org/10.1038/s41559-018-0787-9>.
- Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS One*. 2008;3(6):e2411. <https://doi.org/10.1371/journal.pone.0002411>.
- Schmickl R, Jørgensen MH, Brysting AK, Koch MA. The evolutionary history of the *Arabidopsis lyrata* complex: a hybrid in the amph-Beringian area closes a large distribution gap and builds up a genetic barrier. *BMC Evol Biol*. 2010;10(1):98. <https://doi.org/10.1186/1471-2148-10-98>.
- Schmickl R, Koch MA. *Arabidopsis* hybrid speciation processes. *Proc Natl Acad Sci U S A*. 2011;108(34):14192–14197. <https://doi.org/10.1073/pnas.1104212108>.
- Schmickl R, Marburger S, Bray S, Yant L. Hybrids and horizontal transfer: introgression allows adaptive allele discovery. *J Exp Bot*. 2017;68(20):5453–5470. <https://doi.org/10.1093/jxb/erx297>.
- Schmickl R, Yant L. Adaptive introgression: how polyploidy reshapes gene flow landscapes. *New Phytol*. 2021;230(2):457–461. <https://doi.org/10.1111/nph.17204>.
- Schmid M, Evans BJ, Bogart JP. Polyploidy in Amphibia. *Cytogenet Genome Res*. 2015;145(3-4):315–330. <https://doi.org/10.1159/000431388>.
- Scott AD, Van de Velde JD, Novikova PY. Inference of polyploid origin and inheritance mode from population genomic data. *Methods Mol*

- Biol.* 2023;2545:279–295. https://doi.org/10.1007/978-1-0716-2561-3_15.
- Seear PJ, France MG, Gregory CL, Heavens D, Schmickl R, Yant L, Higgins JD. A novel allele of ASY3 is associated with greater meiotic stability in autotetraploid *Arabidopsis lyrata*. *PLoS Genet.* 2020;16(7):e1008900. <https://doi.org/10.1371/journal.pgen.1008900>.
- Seregin A. Moscow University Herbarium (MW). Version 1.391. Lomonosov Moscow State University. Occurrence dataset; 2025 [accessed 2025 Jul 21]. <https://doi.org/10.15468/cpnhcc> accessed via GBIF.org
- Servick S, Visger CJ, Gitzendanner MA, Soltis PS, Soltis DE. Population genetic variation, geographic structure, and multiple origins of autopolyploidy in *Galax urceolata*. *Am J Bot.* 2015;102(6):973–982. <https://doi.org/10.3732/ajb.1400554>.
- Shimizu KK, Fujii S, Marhold K, Watanabe K, Kudoh H. *Arabidopsis kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh and *A. kamchatica* subsp. *kawasakiana* (Makino) K. Shimizu & Kudoh, New Combinations. *Acta Phytotaxonomica et Geobotanica.* 2005;56:163–172. <https://doi.org/10.18942/apg.KJ00004623241>.
- Shimizu-Inatsugi R, Lihová J, Iwanaga H, Kudoh H, Marhold K, Savolainen O, Watanabe K, Yakubov VV, Shimizu KK. The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Mol Ecol.* 2009;18(19):4024–4048. <https://doi.org/10.1111/j.1365-294X.2009.04329.x>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Slowikowski K. ggrepel: Automatically Position Non-Overlapping Text Labels with “ggplot2”; 2023 [accessed 2024 Jan 8]. Available from: <https://CRAN.R-project.org/package=ggrepel>
- Soltis DE, Buggs RJA, Barbazuk WB, Schnable PS, Soltis PS. On the origins of species: does evolution repeat itself in polyploid populations of independent origin? *Cold Spring Harb Symp Quant Biol.* 2009;74:215–223. <https://doi.org/10.1101/sqb.2009.74.007>.
- Song M, Zhai B, Yang X, Tan T, Wang Y, Yang X, Tan Y, Chu T, Cao Y, Song Y, et al. Interplay between Pds5 and Rec8 in regulating chromosome axis length and crossover frequency. *Sci Adv.* 2021;7(11):eabe7920. <https://doi.org/10.1126/sciadv.abe7920>.
- Sora D, Kron P, Husband BC. Genetic and environmental determinants of unreduced gamete production in *Brassica napus*, *Sinapis arvensis* and their hybrids. *Heredity (Edinb).* 2016;117(6):440–448. <https://doi.org/10.1038/hdy.2016.69>.
- Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics.* 2006;7(1):62. <https://doi.org/10.1186/1471-2105-7-62>.
- Stift M, Kolář F, Meirmans PG. STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity (Edinb).* 2019;123(4):429–441. <https://doi.org/10.1038/s41437-019-0247-6>.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123(3):585–595. <https://doi.org/10.1093/genetics/123.3.585>.
- Takou M, Hämälä T, Koch EM, Steige KA, Dittberner H, Yant L, Genete M, Sunyaev S, Castric V, Vekemans X, et al. Maintenance of adaptive dynamics and no detectable load in a range-edge outcrossing plant population. *Mol Biol Evol.* 2021;38(5):1820–1836. <https://doi.org/10.1093/molbev/msaa322>.
- The 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell.* 2016;166(2):481–491. <https://doi.org/10.1016/j.cell.2016.05.063>.
- Van de Peer Y, Mizrahi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet.* 2017;18(7):411–424. <https://doi.org/10.1038/nrg.2017.26>.
- Viera A, Berenguer I, Ruiz-Torres M, Gómez R, Guajardo A, Barbero JL, Losada A, Suja JA. PDS5 proteins regulate the length of axial elements and telomere integrity during male mouse meiosis. *EMBO Rep.* 2020;21(6):e49273. <https://doi.org/10.15252/embr.201949273>.
- Vlček J, Hämälä T, Cobo CV, Curran E, Šrámková G, Slotte T, Schmickl R, Yant L, Kolář F. Whole-genome duplication increases genetic diversity and load in outcrossing *Arabidopsis*. *bioRxiv* 2025.01.12.632621. <https://www.biorxiv.org/content/10.1101/2025.01.12.632621v1.abstract>, 15 January 2025, preprint: not peer reviewed.
- Walden N, Lucek K, Willi Y. Lineage-specific adaptation to climate involves flowering time in north American *Arabidopsis lyrata*. *Mol Ecol.* 2020;29(8):1436–1451. <https://doi.org/10.1111/mec.15338>.
- Wang O, Chin R, Cheng X, Wu MKY, Mao Q, Tang J, Sun Y, Anderson E, Lam HK, Chen D, et al. Efficient and unique cobarcode of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.* 2019a;29(5):798–808. <https://doi.org/10.1101/gr.245126.118>.
- Wang S, Veller C, Sun F, Ruiz-Herrera A, Shang Y, Liu H, Zickler D, Chen Z, Kleckner N, Zhang L. Per-nucleus crossover covariation and implications for evolution. *Cell.* 2019b;177(2):326–338.e16. <https://doi.org/10.1016/j.cell.2019.02.021>.
- Weiβ CL, Pais M, Cano LM, Kamoun S, Burbano HA. Nquire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics.* 2018;19(1):122. <https://doi.org/10.1186/s12859-018-2128-z>.
- Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. *Genome Res.* 2017;27(5):757–767. <https://doi.org/10.1101/gr.214874.116>.
- Westermann J, Srikant T, Gonzalo A, Tan HS, Bomblies K. Defective pollen tube tip growth induces neo-polyploid infertility. *Science.* 2024;383(6686):eadh0755. <https://doi.org/10.1126/science.adh0755>.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemond G, Hayes A, Henry L, Hester J, et al. Welcome to the tidyverse. *J Open Source Softw.* 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>.
- Willi Y, Lucek K, Bachmann O, Walden N. Recent speciation associated with range expansion and a shift to self-fertilization in North American *Arabidopsis*. *Nat Commun.* 2022;13(1):7564. <https://doi.org/10.1038/s41467-022-35368-1>.
- Wright KM, Arnold B, Xue K, Šurinová M, O’Connell J, Bomblies K. Selection on meiosis genes in diploid and tetraploid *Arabidopsis arenosa*. *Mol Biol Evol.* 2015;32(4):944–955. <https://doi.org/10.1093/molbev/msu398>.
- Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, Franklin FCH, Bomblies K. Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Curr Biol.* 2013;23(21):2151–2156. <https://doi.org/10.1016/j.cub.2013.08.059>.
- Zhang Z, Ren Q, Yang H, Conrad MN, Guacci V, Kateneva A, Dresser ME. Budding yeast PDS5 plays an important role in meiosis and is required for sister chromatid cohesion: PDS5 and meiotic chromatids cohesion. *Mol Microbiol.* 2005;56(3):670–680. <https://doi.org/10.1111/j.1365-2958.2005.04582.x>.
- Zhou X, Mo X, Gui M, Wu X, Jiang Y, Ma L, Shi Z, Luo Y, Tang W. Cytological, molecular mechanisms and temperature stress regulating production of diploid male gametes in *Dianthus caryophyllus* L. *Plant Physiol Biochem.* 2015;97:255–263. <https://doi.org/10.1016/j.plaphy.2015.10.003>.
- Zhukova PG, Petrovsky VV, Plieva TV. The chromosome numbers and taxonomy of some plant species from Siberia and Far East. *Bot Z.* 1973;58:1331–1342. <https://doi.org/10.17581/bp.2017.06103>.