# Multiple polyploidizations in *Arabidopsis lyrata* stabilized by long-range adaptive introgression across Eurasia

**Alison D. Scott[1,\*], Uliana Kolesnikova[1], Anna Glushkevich[1], Laura Steinmann[1], Nikita 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,\*]**

[1]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, Cologne, 50829, Germany

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

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

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

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

[\*]co-corresponding authors: ascott@mpipz.mpg.de, pnovikova@mpipz.mpg.de

## Abstract

**Abundance of polyploidy varies across lineages, evolutionary time and geography, suggesting both genetics and environment play a role in polyploid persistence. *Arabidopsis lyrata* appears to be the most polyploidy-rich species-complex in the *Arabidopsis* genus, with multiple origins of autotetraploidy. This is revealed by genomic data from over 400 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 high 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 evolved, adaptation to polyploidy promotes the establishment of new polyploid lineages through adaptive inter- and intraspecific introgression.**

## 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 polyploids are observed in colder regions closer to the poles[1,2], which is paralleled with

1

dating of ancient WGD events at environmentally harsh times[3]. Whole genome duplications are recurrent; both monocot and dicot lineages experienced multiple polyploidy events detected in the period between 75 and 55 Mya[4], a time range of drastic climatic changes. Polyploidy is more abundant in plants[3,5] compared to animals[6], in Angiosperms vs Gymnosperms[7], and in Amphibia vs Mammals[6,8]. 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, a function of time and space, can both trigger polyploidy through unreduced gamete production[9–11] and impact extinction rates[3]. Genetic background also affects polyploid birth rates directly through production rates of unreduced gametes[12–14], and can provide a route to frequent polyploid emergence.

The emergence of polyploids does not guarantee their survival. Neopolyploid establishment includes adaptation to external environments[15,16], but adaptation to the new internal environment is even more critical. Whole genome duplication in an individual (or single lineage; autopolyploidy) results in multiple sets of homologous chromosomes. Challenges in executing faithful segregation of the doubled chromosomes can hinder the survival of neopolyploids[17,18]. Artificially induced polyploids have reduced fertility and show high levels of aneuploidy in subsequent generations[19–22], suggesting that natural polyploids have evolved adaptations to the polyploid state[23,24] or may be established only from certain preadaptive combinations of alleles in the diploid pool[22].

The genus *Arabidopsis* is a prime example of the non-random nature of polyploidy. Multiple lineages are either fully polyploid (*A. suecica*, *A. kamchatica*) or harbor polyploid populations (*A. arenosa*, *A. lyrata*). *Arabidopsis* polyploids emerged in times of climatic upheaval[25]. *A. 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[26,27]. Establishing tetraploid *A. arenosa* involved polygenic adaptation of reproductive machinery, including pollen tube growth[28] and meiosis. The former is probably to restore the ion homeostasis, the latter is to prevent entanglement between three or more chromosomal copies and ensure faithful segregation[18,20,23].

We focus our study on the most Northern *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[29–32]. Tetraploid *A. lyrata* populations in Central Europe form three distinct lineages occupying three distinct regions[33], which may suggest multiple origins of polyploidy in contrast to *A. arenosa*. Diploid and tetraploid *A. lyrata* and *A. arenosa* have overlapping, yet non-sympatric ranges in Europe, and while diploids are reproductively isolated[34] and do not show signs of ongoing interspecific gene flow[29,32,35,36], tetraploid *A. lyrata* and *A. arenosa* often hybridize[30,32,37–40]. 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[39,41].

There are indications of a polyploid population in Yakutia[25,42], 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[43–46], 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[47], 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 (Figure 1a, Supplementary Figure 1, Supplementary Data 1). Third, we included whole-genome sequencing data of samples from different *A. lyrata* lineages which have been previously published[39,45,48–51]. All re-sequencing data were mapped to the reference genome of the self-compatible Siberian *A. lyrata* NT1 strain[52].

We used a statistical framework based on relative allelic counts implemented in nQuire software[53] to get a first inference of ploidy from sequencing data (See *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 (Figure 1b, Supplementary Figure 2). To complement the initial inference, we genotyped and counted *S*-alleles in each sample using NGSgenotyp[54]. Due to the extreme divergence and diversity of *S*-alleles controlling self-incompatibility in *Arabidopsis*[55–61], 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 (Figure 1b). Finally, we directly counted chromosome numbers and measured genome sizes in the live samples grown in the greenhouse from seeds collected in the field (Supplementary Figure 1, Supplementary Data 2) and confirmed our bioinformatic inferences. When bioinformatic inference yielded unclear results and cytogenetics was not possible (i.e. for herbarium specimens), we indicate the ploidy as "unsure" and exclude these accessions from downstream analyses (Figure 1b, Supplementary Data 1). 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 1). Following the assignment of ploidy by heterozygosity, *S*-allele genotype, and cytological confirmation, we re-called 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) (Figure 1b, Supplementary Figure 2). Tetraploid individuals had a significantly higher proportion of heterozygous sites compared to diploid selfers and outcrossers, and up to four different *S*-alleles in each individual, which points to their outcrossing nature (Figure 1b, Supplementary Figure 2), confirmed by greenhouse observations. Nucleotide diversity (π) calculated on putatively neutral four-fold degenerate sites is comparable between diploid and tetraploid (~0.005) *A. lyrata* geographic populations (individuals sampled in the same collection site) (Supplementary Table 1), showing no evidence of genetic bottleneck in tetraploids.
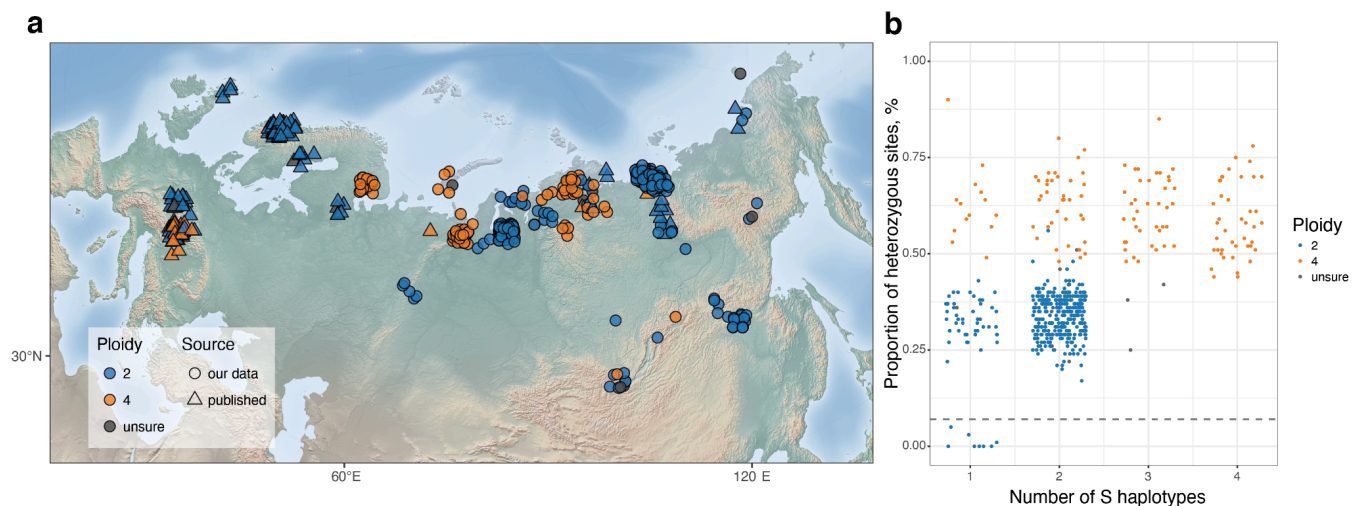


**Figure 1** Distribution of diploid and tetraploid *A. lyrata* across Eurasia. (**a**) Geographical distribution of *A. lyrata* samples used in this study. For better visualization in cases of overlaps we randomly shifted the points within 100 km from the actual sampling location, all true coordinates can be found in the Supplementary Data 1. 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[62]. Herbarium individuals with ambiguous assignments of ploidy between two inference methods are shown in gray dots.

## Independent autopolyploidization events of *A. lyrata* in Northern Ural and Central Siberia around LGM

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 (Figure 2a), showing that the newly described tetraploids cluster within two distinct groups, separate from the Central Europe tetraploids. Northern Ural (NU, highlighted in pink, Figure 2a) populations sampled around the White Sea and in Polar Ural mountains form a cluster containing diploid and tetraploid individuals. This NU cluster is genetically close to the

4

West Siberian cluster (WS, highlighted in blue, Figure 2a) comprised purely of diploid populations sampled from Taz Estuary. Central Siberian cluster (CS, highlighted in yellow, Figure 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, Figure 2a). Tetraploids are genetically closer to their spatially close diploid populations from their corresponding clusters (NU 4x to NU 2x and CS 4x to CS 2x, CE4x) than each other, suggesting their independent origin.

Using population-level data, we estimated site frequency spectra for the two newly discovered tetraploid *A. lyrata* lineages (NU 4x and CS 4x) separately and compared them to theoretical expectations for auto- and allo-polyploids. In both cases, the observed site frequency spectra of Siberian *A. lyrata* tetraploids are consistent with autopolyploidy, indicated by an absence of sites at an intermediate allele frequency (Supplementary Figure 3). 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[63]. Allotetraploid *Arabidopsis kamchatica*, originating through hybridizations of Siberian *A. lyrata* and *Arabidopsis halleri*[62,64,65], is also included in the network and clusters separately from NU and CS autotetraploid *A. lyrata* (Figure 2a).

We then used TreeMix[66] to describe relationships among diploid populations. The maximum likelihood model (Supplementary Figure 4a) groups Karelia, Northern Ural, Western Siberia, Central Siberia, Eastern Siberia, and Amur Basin 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 Figure 4b). Clustering of individuals using ADMIXTURE[67] with the optimal number of groups K=8 (Supplementary Figure 5) assigned tetraploid and diploid NU populations into a single cluster (pink, Figure 2b with high levels of admixture with diploid WS populations (blue, Figure 2b); tetraploids and diploids from CS were also assigned into a single cluster (yellow, Figure 2b) with high levels of admixture with diploid ES cluster (orange, Figure 2b). We used topology weighting with Twisst[68] to further explore the relationships between *A. lyrata* lineages when variation in topologies across the genome is taken into account (Figure 2c, Supplementary Figure 6). The top three highest-weighted topologies are consistent with the independent origin of the two autotetraploid lineages. Northern Ural tetraploids and diploids form a monophyletic clade in the most frequent topology, as do Central Siberian tetraploids and diploids.
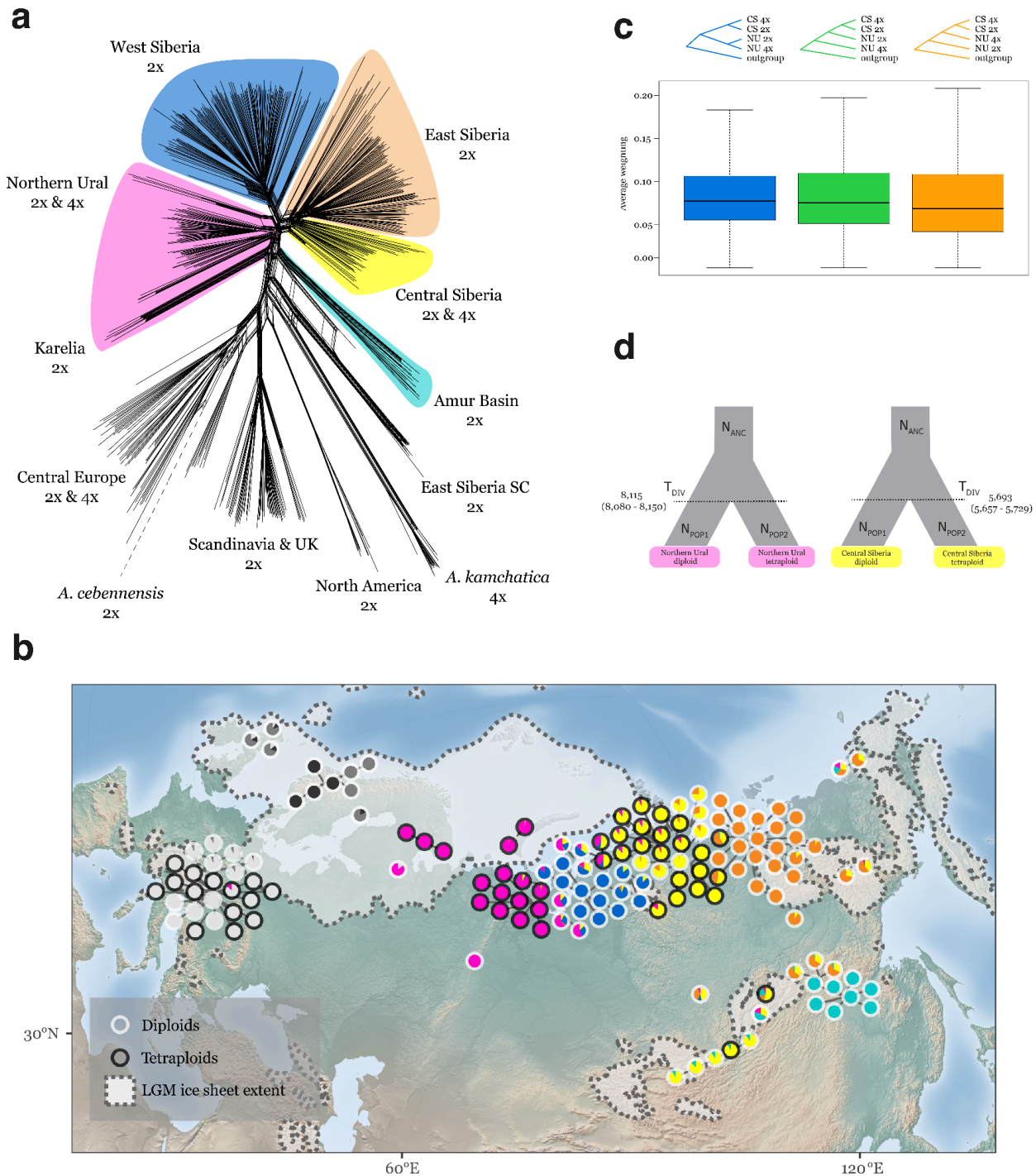
**Figure 2** Population structure of *A. lyrata* diploid-tetraploid species complex. (**a**) Network representation of genetic pairwise distances between individuals. *A. lyrata* accessions form several clusters outside of Central Europe, named here as Northern Ural (NU; pink, diploid/tetraploid), West Siberian (WS; blue, diploid), Central Siberian (CS; yellow, diploid/tetraploid), East Siberian (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

6

*A. kamchatica*. (**b**) Admixture across the Eurasian *A. lyrata* distribution. Colors correspond to labeling in (**a**) 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 Last Glacial Maximum[69] shown in white. (**c**) The three highest weighted topologies of populations show that diploid and tetraploid Northern Ural populations form a clade; the same for diploid and tetraploid Central Siberian populations. (**d**) The estimates of the time of origins for the two tetraploid lineages as divergence time estimates from the closest diploid progenitors using Fastsimcoal2[70].

To estimate the timing of tetraploid origin in Northern Ural and Central Siberia, we used a demographic modeling approach based on fitting simulated and observed site frequency spectra[70]. For each pair of diploid-tetraploid populations (NU 2x and NU 4x, CS 2x and CS 4x) we tested a simple divergence model and a simple model with unidirectional gene flow from diploids into tetraploids, assuming gene flow can be realized through unreduced gametes formed by diploids. Likelihood comparisons and AIC tests suggested that the gene flow model better fit the data in both cases. For the origin of Northern Ural tetraploids, we estimated 8,115 generations ago (95% CI 8,080-8,150), which we date at roughly 16 Kya (assuming 2 year generation time), and for the Central Siberian tetraploids we estimated an origin at 5,693 generations ago or approximately 11 Kya (95% CI 5,693-5,729) (Figure 2d, Supplementary Table 2).

## Tetraploid-specific adaptive introgression of genes involved in meiotic stabilization
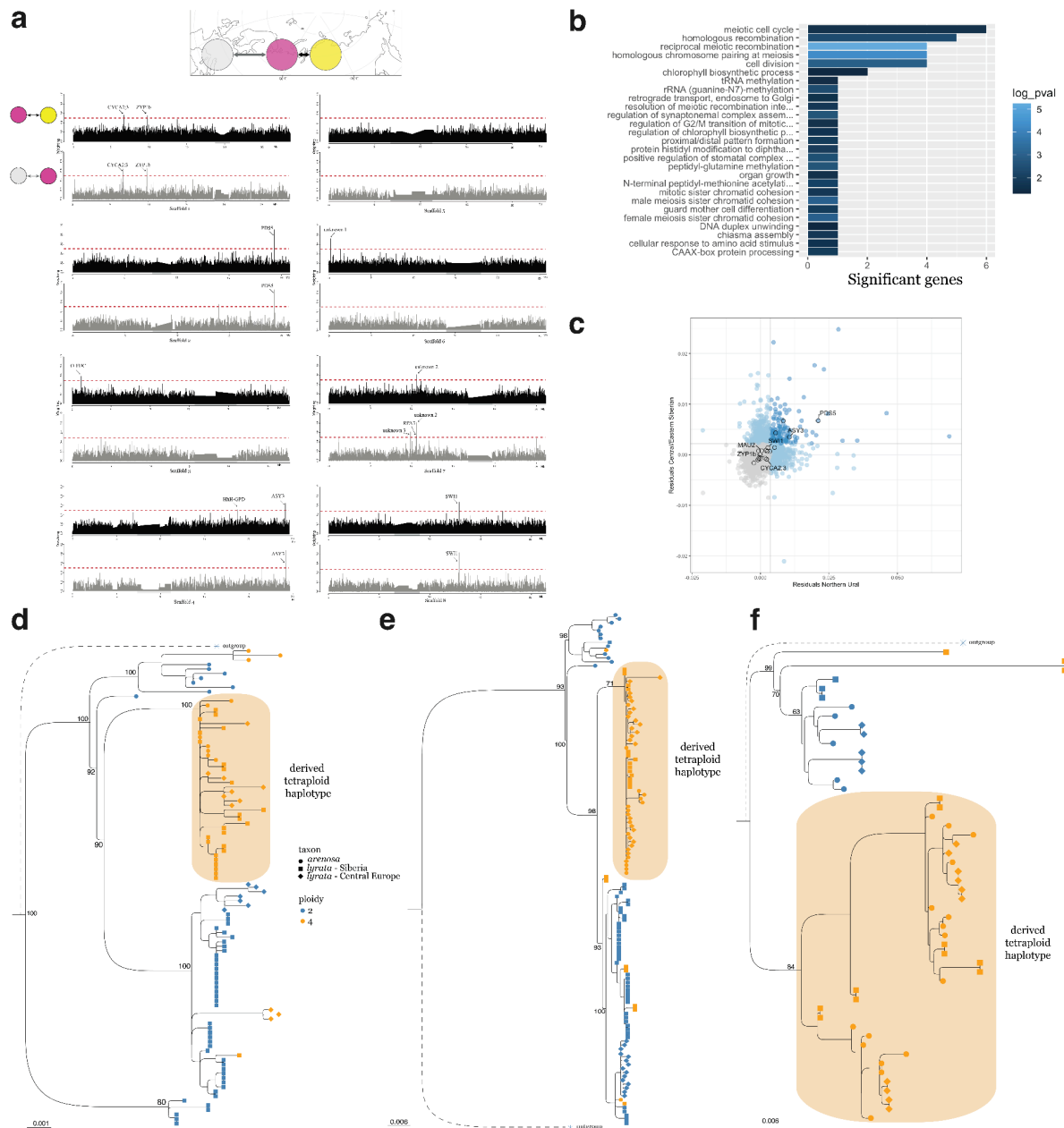
While topologies consistent with the population tree and the network analysis were the most frequent (Figure 2c), a smaller but considerable fraction indicated introgression between the Northern Ural and Central Siberian tetraploids (topologies 4,5,6, Supplementary Figure 6). We combine the topology weighting results for all introgression topologies (4,5,6) and show them along the chromosomes (Figure 3a, top panels). In total, 11 windows with a high weighting of gene flow topologies are apparent. Comparing these window boundaries with annotation, we identified 37 gene models (Supplementary Data 4) located within introgression windows. These genes are enriched for gene ontology terms related to meiosis and chromosome organization (Figure 3b, Supplementary Data 4). Performing the same analysis now including Central European and Northern Ural diploid and tetraploid populations, we find striking overlap of the introgression windows found in Northern Ural and Central Siberian comparison (Figure 3a, bottom panels). For each gene in introgression windows, we look at SNP frequencies across diploid and tetraploid lineages including Central Europe, Northern Ural and Central Siberia (Supplemental Figure 7). We find SNP sharing across the entire geographical range: in some cases SNPs appear to be tetraploid-specific (e.g. *ASY3*, *MAU2*, *PDS5*, and *ZYP1b*) and other SNPs are shared among diploids and tetraploids (e.g in *CYCA2;3* and *PAPP2C*).

The strong enrichment of specific gene ontology terms (Figure 3b) suggests introgressed alleles are undergoing positive selection. To verify this, we calculated nucleotide diversity (π) per gene within Northern Ural tetraploids and between Northern Ural diploids and tetraploids and between Central Siberian diploids and tetraploids using piawka (https://github.com/novikovalab/piawka). Under neutrality, these values should be correlated,

7

and outliers should indicate selection (Supplementary Figure 8). Four introgressed genes deviated from neutral expectations and instead fall among outliers in both comparisons (Figure 3c), concluding that genes prone to introgression between the Northern Ural and Central Siberian tetraploids are also under selection in both tetraploid lineages.

To validate haplotype sharing across the range, we obtained full haplotype sequences from synthetic long reads (single-tube long fragment reads; stLFR), PacBio long-read assemblies (see Methods), and published haplotype sequences of Sanger sequenced PCR products[23,71]. 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 phylogenetic trees. Tetraploid-specific haplotypes from *A. arenosa*, Central European, Northern Ural and Central Siberian *A. lyrata* form a monophyletic clade at these loci (Figure 3d-f). This corresponds well with SNP frequency heatmaps at the same loci (Supplemental Figure 7).

**Figure 3** Genes present in introgression topologies. (**a**) Distribution of introgression topology windows and their weighting across the genome, per scaffold. Height of peaks indicates weight, and introgression blocks are annotated. (**b**) Results of GO term enrichment for genes within introgression windows, with number of significant genes indicated on the x axis and log p-value by color of bars. (**c**) Scatter plot of residuals from linear model fit to nucleotide diversity (π) within tetraploids and divergence (dxy) between diploids and tetraploids. Nucleotide diversity is calculated per-gene, so that each dot represents a single gene. Outlier genes colored light blue for each comparison (Northern Ural diploids and tetraploids on x axis; Central Siberian diploids and tetraploids on y axis), darker blue indicates outliers in both comparisons. ASY3 and PDS5 annotated within outliers. (**d, e, f**) 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

9

Siberia, diamonds for *A. lyrata* in Central Europe, and circles for *A. arenosa*. (**d**) Haplotype tree for *PDS5* (**e**) Haplotype tree for *ASY3* (**f**) Haplotype tree for *ZYP1b*.

## Discussion

Our range-wide sampling of *A. lyrata* across Eurasia reveals the extent of *A. lyrata* tetraploids (Figure 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* (Figure 2) and the absence of fixed heterozygosity due to subgenome divergence (Supplementary Figure 3). The autotetraploid populations from Northern Ural and Central Siberia regions group together with local diploids rather than with each other (Figure 2, Supplementary Figure 4b), suggesting different source populations for their origin. Considering that autotetraploid *A. lyrata* populations in Europe also form distinct lineages[33] and that allotetraploid *A. kamchatica* originated from multiple hybridizations between *A. lyrata* and *A. halleri*[44,65], 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* (1) has a propensity for whole genome duplication, perhaps by an increased likelihood of producing unreduced gametes or increased sporadic somatic WGD, (2) establishment of *A. lyrata* tetraploids is facilitated by special environmental conditions or genetic background.

*A. lyrata* is the northernmost species among *Arabidopsis*, and its evolutionary history has been largely affected by temperature fluctuations during the Pleistocene. Expansion of *A. lyrata* eastwards from Europe has been proposed to predate the second to last Ice Age[35], 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[44]. 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 Northern Ural and the latter from Central Europe, both occurring after the ice sheet retreat post-LGM[48,72]. Such a pattern of distinct re-colonization routes is repeated in *Arabidopsis thaliana*[73], and other Arctic species including keystone species such as *Dryas octopetala, Vaccinium vitis-idaea* and *Betula nana*[74].

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 Northern Ural and Central Siberia arrive close to the peak of the last Ice Age (Figure 2) and follow a general pattern of co-occurrence with glaciations[25]. This suggests that environment may play a crucial role in polyploid birth rate, triggering whole-genome duplications via production of 2x 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 or mitotic cell division[12,75].

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[76,77], we see strong evidence of the impact of genetic adaptation. Topology weighting analysis and haplotype sharing from *de novo* assemblies reveal introgression among distinct tetraploid lineages over long distances (Figure 3). Given that the tetraploid populations in Europe were dated at around 160 Kya[39], considerably older than Siberian tetraploid lineages dated around 16 and 11 Kya, we infer the direction of the gene flow from Central European *A. lyrata* all the way to Central Siberian *A. lyrata* (Figure 3). The interspecific introgression from *A. arenosa* to *A. lyrata* in Central Europe has been shown previously[39]. Regions introgressed from Central Europe to Northern Ural and to Central Siberia are narrow and highly enriched in meiotic genes (Figure 3a,b). Among introgressed haplotypes in Siberia, we find the *ASY3* tetraploid-derived allele which has an established functional role[23]. 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[23], leading to more stable meiosis. Additional research[71] 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[78,79]. In tetraploids, axis length may be especially important if it serves to limit crossovers to one per chromosome, reducing the risk of multivalent formation[80]. Patterns of diversity and divergence in Siberian lineages at the introgressed region harboring *ASY3* gene are consistant 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[81]. While experimental work in *A. thaliana* shows that PDS5 mutants have minimal meiotic phenotypes[82], suggesting its role in *Arabidopsis* diverges from that in yeast systems, more recent work shows that PDS5 impacts axis length. 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*[20,83,84], then in Central European *A. lyrata*[39], and in *Cardamine amara*[85], 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 Northern Ural and Central Siberian 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 cross-overs and establish proper bivalents[86–88]. Tetraploid alleles of *ZYP1* paralogs are likely adaptive as they appear on introgression and selection scans[20,39,84]. However, in Siberia, although tetraploid Central European alleles of *ZYP1b* are already introduced and detected by introgression scans, their frequency is still low as *ZYP1b* does not appear among outliers in diversity and divergence

patterns (Figure 3). Meiotic adaptation to polyploidy is polygenic without a clear key player but rather many players of cumulative effects[17,89], the dynamics of which we probably observe here.

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 at first. 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[90]. 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. 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 Urals and South Yakutia (Supplementary Data 1). We then grew these seeds in the greenhouse (21 C°, 16 hours of daylight) and sampled leaf tissue for sequencing.

### Short-read sequencing for population analyses

Plant material was processed in two different ways, indicated by types I and II in Supplementary Data 1. Processing protocols for each type are described in detail in [62]. 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. Full list of short-read sequenced samples used in the study is in the Supplementary Data 1.

### 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 was performed by BGI as described in [91]. 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 100bp.

## Short read mapping and variant calling for population analysis

We analyzed short paired-end reads (2 x 150 bp) the same way as in Kolesnikova et al[62]. The reads were filtered for adapters with bbduk.sh script from BBMap (38.20)[92] with ktrim=r k=23 mink=11 hdist=1 tbo tpe qtrim=rl trimq=15 minlen=70. Then, we mapped reads to the NT1 *A. lyrata* reference genome using bwa mem (0.7.17)[93] with -M parameter. We used picard MarkDuplicates[94] to mark PCR duplicated reads and then used samtools[95] to sort and index the bam files. We called variants with HaplotypeCaller algorithm from GATK (3.8)[96] 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 (HaplotypeCaller with --sample-ploidy 2 or --sample-ploidy 4).

## 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[97] with --style=pseudohap2 option to get two haplotypes for each assembly. The haplotypes were scaffolded on the NT1 *A.lyrata* genome[62] with ragtag[98]. GALBA pipeline[99–103] was used for protein annotation of the assemblies with proteins from Phytozome[104] *A.lyrata* v.2.1 release[105] as a protein database. The annotated genes were used for gene tree construction. Protein BLAST[106] was used to identify genes of interest in the annotation. Then, we used L-INS-i algorithm of MAFFT[107] for multiple sequence alignment of the gene regions. The alignments were manually curated and used for tree construction with iqtree[108,109], fast bootstrap=1000.

## Ploidy estimation and *S*-allele genotyping from short-read sequencing data

To estimate ploidy we used nQuire[53] with "create", "denoise" and "lrdmodel" steps. This ploidy estimation method is based on allelic read depth distributions, also previously described[110]. We identified *S*-alleles as described in Kolesnikova et al[62] using short-read sequencing data and *S*-locus genotyping pipeline NGSgenotyp[54] with two first steps: kmerfilter (-k 20 for *SRK* and -k 15 for *SCR*) and genotyp – with both *SRK* and *SCR* sequence reference databases. In addition to the databases used previously[62], we have assembled four *SRK* alleles from short-read data with assembly option and provide them together with 6 more published[59,111] *SRK* alleles in Supplementary Table 3. The sequences of *S*-alleles are provided in the Supplementary Data 3 in .fasta format. The number of different *S*-alleles in each individual together with proportion of heterozygous sites were used to confirm the estimated ploidy level.

## Visualization

13

We used R v4.2.2[112], primarily with the tidyverse v2.0.0 package collection[113], for making the plots. Maps were created using Lambert Asimuthal Equal Area projection centered at N 40° E 90°, packages sf v1.0-14[114] and raster v3.6-23[115] and a background map by NaturalEarth; the ggrepel algorithm[116] was used to adjust the positioning of the pie charts. The Last Glacial Maximum (LGM) ice sheet extents[69] 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 (3:1; Ethanol: Acetic Acid) for 2h on a horizontal shaker at room temperature, changed to fresh fixative, and stored them at -20°C until further usage.

Fixed buds were rinsed twice for 5min in 10mM 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 10mM Citrate buffer) at 37°C for 1h 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 1-2 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 2ug/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 2.

## Estimation of nucleotide diversity

We calculated π[117] 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[118] compatible with mixed-ploidy datasets. We also calculated π within each tetraploid lineage and between diploid-tetraploid lineages on a per-gene basis using piawka.

## Twisst

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). Genealogies were inferred in windows of 50 SNPs along the genome using PHYML[119] and a phasing approach implemented in genomics_general. Previously published[45] samples of *Arabidopsis cebennensis* and *Arabidopsis pedemontana* were used as outgroups. We used Twisst[68] for topology

weighting. GO enrichment analysis using g:Profiler was performed for genes intersecting the regions with >50% weighting of introgression topologies.

## Demographic Modeling

To generate Site Frequency Spectra (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 1. Using the chosen SNPs we generated bootstrap replicates by dividing the genome into 10kb 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 boot-strapped replicates and a joint site frequency spectrum for each pseudoobservation using the python package dadi[120]. The SFS spectra generated by dadi were used as input into fastsimcoal2 (fsc27)[70]. 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 fastsimcoal 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 (CIs) in R.

## 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 number PRJEB67879 and PRJEB60410. The list of all the samples used in the study with individual ENA numbers is in Supplementary Data 1. 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 1 column "accession". Both raw stLFR data and resulting assemblies are also accessible in the ENA database, with individual ENA numbers in Supplementary Data 5. Supplementary Data 1-5 and the scripts used in the analysis are available at https://github.com/novikovalab/Siberian_Alyrata

## Acknowledgments

15

## References

1. Rice, A. *et al.* The global biogeography of polyploid plants. *Nat Ecol Evol* **3**, 265–273 (2019).

2. David, K. T. Global gradients in the distribution of animal polyploids. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2214070119 (2022).

3. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).

4. Lohaus, R. & Van de Peer, Y. Of dups and dinos: evolution at the K/Pg boundary. *Curr. Opin. Plant Biol.* **30**, 62–69 (2016).

5. Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z. & Levin, D. A. On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* **210**, 391–398 (2016).

6. Gregory, T. R. & Mable, B. K. CHAPTER 8 - Polyploidy in Animals. in *The Evolution of the Genome* (ed. Gregory, T. R.) 427–517 (Academic Press, Burlington, 2005).

7. Khoshoo, T. N. Polyploidy in Gymnosperms. *Evolution* **13**, 24–39 (1959).

8.  Schmid, M., Evans, B. J. & Bogart, J. P. Polyploidy in Amphibia. *Cytogenet. Genome Res.* **145**, 315–330 (2015).

9.  Mason, A. S. & Pires, J. C. Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends Genet.* **31**, 5–10 (2015).

10. Zhou, X. *et al.* Cytological, molecular mechanisms and temperature stress regulating production of diploid male gametes in Dianthus caryophyllus L. *Plant Physiol. Biochem.* **97**, 255–263 (2015).

11. De Storme, N., Copenhaver, G. P. & Geelen, D. Production of diploid male gametes in Arabidopsis by cold-induced destabilization of postmeiotic radial microtubule arrays. *Plant Physiol.* **160**, 1808–1826 (2012).

12. Mason, A. S., Nelson, M. N., Yan, G. & Cowling, W. A. Production of viable male unreduced gametes in Brassica interspecific hybrids is genotype specific and stimulated by cold temperatures. *BMC Plant Biol.* **11**, 103 (2011).

13. Kreiner, J. M., Kron, P. & Husband, B. C. Frequency and maintenance of unreduced gametes in natural plant populations: associations with reproductive mode, life history and genome size. *New Phytol.* **214**, 879–889 (2017).

14. Kreiner, J. M., Kron, P. & Husband, B. C. Evolutionary Dynamics of Unreduced Gametes. *Trends Genet.* **33**, 583–593 (2017).

15. Manzaneda, A. J. *et al.* Environmental aridity is associated with cytotype segregation and polyploidy occurrence in Brachypodium distachyon (Poaceae). *New Phytol.* **193**, 797–805 (2012).

16. Chao, D. Y. *et al.* Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis. *Science* **341**, 658–659 (2013).

17. Bomblies, K., Higgins, J. D. & Yant, L. Meiosis evolves: adaptation to external and internal environments. *New Phytol.* **208**, 306–323 (2015).

18. 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* **125**, 287–300 (2016).

19. Comai, L. *et al.* Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids. *Plant Cell* **12**, 1551–1568 (2000).

20. Yant, L. *et al.* Meiotic adaptation to genome duplication in Arabidopsis arenosa. *Curr. Biol.* **23**, 2151–2156 (2013).

21. Burns, R. *et al.* Gradual evolution of allopolyploidy in Arabidopsis suecica. *Nat Ecol Evol* (2021) doi:10.1038/s41559-021-01525-w.

22. Katche, E. I. *et al.* Genetic factors inherited from both diploid parents interact to affect genome stability and fertility in resynthesized allotetraploid Brassica napus. *G3* (2023) doi:10.1093/g3journal/jkad136.

23. Morgan, C., Zhang, H., Henry, C. E., Franklin, F. C. H. & Bomblies, K. Derived alleles of two axis proteins affect meiotic traits in autotetraploid Arabidopsis arenosa. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8980–8988 (2020).

24. Bohutínská, M. *et al.* De Novo Mutation and Rapid Protein (Co-)evolution during Meiotic Adaptation in Arabidopsis arenosa. *Mol. Biol. Evol.* **38**, 1980–1994 (2021).

25. Novikova, P. Y., Hohmann, N. & Van de Peer, Y. Polyploid Arabidopsis species originated around recent glaciation maxima. *Curr. Opin. Plant Biol.* **42**, 8–15 (2018).

26. Arnold, B., Kim, S. T. & Bomblies, K. Single Geographic Origin of a Widespread Autotetraploid Arabidopsis arenosa Lineage Followed by Interploidy Admixture. *Mol. Biol. Evol.* **32**, 1382–1395 (2015).

27. Monnahan, P. *et al.* Pervasive population genomic consequences of genome duplication in Arabidopsis arenosa. *Nat Ecol Evol* **3**, 457–468 (2019).

28. Westermann, J., Srikant, T., Gonzalo, A., Tan, H. S. & Bomblies, K. Defective pollen tube tip

growth induces neo-polyploid infertility. *Science* **383**, eadh0755 (2024).

29. Ansell, S. W. *et al.* Population structure and historical biogeography of European Arabidopsis lyrata. *Heredity* **105**, 543–553 (2010).

30. Jorgensen, M. H., Ehrich, D., Schmickl, R., Koch, M. A. & Brysting, A. K. Interspecific and interploidal gene flow in Central European Arabidopsis (Brassicaceae). *BMC Evol. Biol.* **11**, 346 (2011).

31. Polatschek, A. Cytotaxonomische Beiträge zur Flora der Ostalpenländer, I. *Österreichische Botanische Zeitschrift* **113**, 1–46 (1966).

32. Schmickl, R. & Koch, M. A. Arabidopsis hybrid speciation processes. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14192–14197 (2011).

33. Bohutínská, M. *et al.* Mosaic haplotypes underlie repeated adaptation to whole genome duplication in Arabidopsis lyrata and Arabidopsis arenosa. *bioRxiv* 2023.01.11.523565 (2023) doi:10.1101/2023.01.11.523565.

34. Lafon-Placette, C. *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.* **114**, E1027–E1035 (2017).

35. Ross-Ibarra, J. *et al.* Patterns of polymorphism and demographic history in natural populations of Arabidopsis lyrata. *PLoS One* **3**, e2411 (2008).

36. Clauss, M. J. & Mitchell-Olds, T. Population genetic structure of Arabidopsis lyrata in Europe. *Mol. Ecol.* **15**, 2753–2766 (2006).

37. Lafon-Placette, C. & Kohler, C. Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol. Ecol.* **25**, 2620–2629 (2016).

38. Hohmann, N. & Koch, M. A. An Arabidopsis introgression zone studied at high spatio-temporal resolution: interglacial and multiple genetic contact exemplified using whole nuclear and plastid genomes. *BMC Genomics* **18**, 810 (2017).

39. Marburger, S. *et al.* Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* **10**, 5218 (2019).

40. Schmickl, R., Marburger, S., Bray, S. & Yant, L. Hybrids and horizontal transfer: introgression allows adaptive allele discovery. *J. Exp. Bot.* **68**, 5453–5470 (2017).

41. Schmickl, R. & Yant, L. Adaptive introgression: how polyploidy reshapes gene flow landscapes. *New Phytol.* **230**, 457–461 (2021).

42. Zhukova, P. G., Petrovsky, V. V. & Plieva, T. V. The chromosome numbers and taxonomy of some plant species from Siberia and Far East. *Bot. Z.* **58**, 1331–1342 (1973).

43. Dart, S., Kron, P. & Mable, B. K. Characterizing polyploidy in Arabidopsis lyrata using chromosome counts and flow cytometry. *Can. J. Bot.* **82**, 185–197 (2004).

44. Schmickl, R., Jorgensen, M. H., Brysting, A. K. & Koch, M. A. The evolutionary history of the Arabidopsis lyrata complex: a hybrid in the amphi-Beringian area closes a large distribution gap and builds up a genetic barrier. *BMC Evol. Biol.* **10**, 98 (2010).

45. Novikova, P. Y. *et al.* Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nat. Genet.* **48**, 1077–1082 (2016).

46. Walden, N., Lucek, K. & Willi, Y. Lineage-specific adaptation to climate involves flowering time in North American Arabidopsis lyrata. *Mol. Ecol.* **29**, 1436–1451 (2020).

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

48. Mattila, T. M., Tyrmi, J., Pyhajarvi, T. & Savolainen, O. Genome-Wide Analysis of Colonization History and Concomitant Selection in Arabidopsis lyrata. *Mol. Biol. Evol.* **34**, 2665–2677 (2017).

49. Hamala, T., Mattila, T. M., Leinonen, P. H., Kuittinen, H. & Savolainen, O. Role of seed germination in adaptation and reproductive isolation in Arabidopsis lyrata. *Mol. Ecol.* **26**,

3484–3496 (2017).

50. Takou, M. *et al.* Maintenance of Adaptive Dynamics and No Detectable Load in a Range-Edge Outcrossing Plant Population. *Mol. Biol. Evol.* **38**, 1820–1836 (2021).

51. 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.* **13**, 7564 (2022).

52. Kolesnikova, U. K. *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. *bioRxiv* 2022.06.24.497443 (2023) doi:10.1101/2022.06.24.497443.

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

54. 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.* **37**, 1193–1201 (2020).

55. Mable, B. K., Schierup, M. H. & 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* **90**, 422–431 (2003).

56. Charlesworth, D., Mable, B. K., Schierup, M. H., Bartolomé, C. & Awadalla, P. Diversity and linkage of genes in the self-incompatibility gene family in Arabidopsis lyrata. *Genetics* **164**, 1519–1535 (2003).

57. Castric, V. & Vekemans, X. Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Mol. Ecol.* **13**, 2873–2889 (2004).

58. Mable, B. K., Beland, J. & Di Berardo, C. Inheritance and dominance of self-incompatibility

alleles in polyploid Arabidopsis lyrata. *Heredity* **93**, 476–486 (2004).

59. Castric, V., Bechsgaard, J., Schierup, M. H. & Vekemans, X. Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genet.* **4**, e1000168 (2008).

60. Llaurens, V. *et al.* Does frequency-dependent selection with complex dominance interactions accurately predict allelic frequencies at the self-incompatibility locus in Arabidopsis halleri? *Evolution* **62**, 2545–2557 (2008).

61. Le Veve, A. *et al.* Long-Term Balancing Selection and the Genetic Load Linked to the Self-Incompatibility Locus in Arabidopsis halleri and A. lyrata. *Mol. Biol. Evol.* **40**, (2023).

62. Kolesnikova, U. K. *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.* **40**, (2023).

63. Blischak, P. D., Sajan, M., Barker, M. S. & Gutenkunst, R. N. Demographic history inference and the polyploid continuum. *Genetics* **224**, (2023).

64. Shimizu, K. K., 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* **56**, 163–172 (2005).

65. Shimizu-Inatsugi, R. *et al.* The allopolyploid Arabidopsis kamchatica originated from multiple individuals of Arabidopsis lyrata and Arabidopsis halleri. *Mol. Ecol.* **18**, 4024–4048 (2009).

66. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).

67. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

68. Martin, S. H. & Van Belleghem, S. M. Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics* **206**, 429–438 (2017).

69. *Quaternary Glaciations - Extent and Chronology: Volume 15: A Closer Look*. (Elsevier Science, London, England, 2011).

70. Excoffier, L. *et al.* fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885 (2021).

71. Seear, P. J. *et al.* A novel allele of ASY3 is associated with greater meiotic stability in autotetraploid Arabidopsis lyrata. *PLoS Genet.* **16**, e1008900 (2020).

72. Muller, M.-H., Leppälä, J. & Savolainen, O. Genome-wide effects of postglacial colonization in Arabidopsis lyrata. *Heredity (Edinb.)* **100**, 47–58 (2008).

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

74. Eidesen, P. B. *et al.* Genetic roadmap of the Arctic: plant dispersal highways, traffic barriers and capitals of diversity. *New Phytol.* **200**, 898–910 (2013).

75. Sora, D., Kron, P. & Husband, B. C. Genetic and environmental determinants of unreduced gamete production in Brassica napus, Sinapis arvensis and their hybrids. *Heredity* **117**, 440–448 (2016).

76. Brochmann, C. *et al.* Polyploidy in arctic plants. *Biol. J. Linn. Soc. Lond.* **82**, 521–536 (2004).

77. Parisod, C., Holderegger, R. & Brochmann, C. Evolutionary consequences of autopolyploidy. *New Phytol.* **186**, 5–17 (2010).

78. Wang, S. *et al.* Per-nucleus crossover covariation and implications for evolution. *Cell* **177**, 326–338.e16 (2019).

79. Song, M. *et al.* Interplay between Pds5 and Rec8 in regulating chromosome axis length and crossover frequency. *Sci. Adv.* **7**, eabe7920 (2021).

80. Morgan, C. *et al.* Evolution of crossover interference enables stable autopolyploidy by ensuring pairwise partner connections in Arabidopsis arenosa. *Curr. Biol.* **31**, 4713–4726.e4 (2021).

81. Zhang, Z. *et al.* Budding yeast PDS5 plays an important role in meiosis and is required for sister chromatid cohesion: PDS5 and meiotic chromatids cohesion. *Mol. Microbiol.* **56**, 670–680 (2005).

82. Pradillo, M. *et al.* Involvement of the cohesin cofactor PDS5 (SPO76) during meiosis and DNA repair in Arabidopsis thaliana. *Front. Plant Sci.* **6**, 1034 (2015).

83. Hollister, J. D. *et al.* Genetic adaptation associated with genome-doubling in autotetraploid Arabidopsis arenosa. *PLoS Genet.* **8**, e1003093 (2012).

84. Wright, K. M. *et al.* Selection on meiosis genes in diploid and tetraploid Arabidopsis arenosa. *Mol. Biol. Evol.* **32**, 944–955 (2015).

85. Bohutínská, M. *et al.* Novelty and convergence in adaptation to whole genome duplication. *Mol. Biol. Evol.* (2021) doi:10.1093/molbev/msab096.

86. Higgins, J. D., Sanchez-Moran, E., Armstrong, S. J., Jones, G. H. & Franklin, F. C. H. The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. *Genes Dev.* **19**, 2488–2500 (2005).

87. Barakate, A. *et al.* The synaptonemal complex protein ZYP1 is required for imposition of meiotic crossovers in barley. *Plant Cell* **26**, 729–740 (2014).

88. Capilla-Pérez, L. *et al.* The synaptonemal complex imposes crossover interference and heterochiasmy in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

89. Bomblies, K. Learning to tango with four (or more): the molecular basis of adaptation to polyploid meiosis. *Plant Reprod.* **36**, 107–124 (2023).

90. Morgan, C. H., 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.* **372**, (2017).

91. Wang, O. *et al.* Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.* **29**, 798–808 (2019).

92. Bushnell, B. BB tools software package, URL. Preprint at (2014).

93. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* (2009).

94. Picard. http://broadinstitute.github.io/picard/.

95. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

96. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

97. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).

98. Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 1–19 (2022).

99. Hoff, K. J. & Stanke, M. Predicting Genes in Single Genomes with AUGUSTUS. *Curr. Protoc. Bioinformatics* **65**, e57 (2019).

100. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).

101. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, (2023).

102. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-Genome Annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).

103. 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* **7**, 62 (2006).

104. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–86 (2012).

105. Rawat, V. *et al.* Improving the Annotation of Arabidopsis lyrata Using RNA-Seq Data. *PLoS One* **10**, e0137391 (2015).

106. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

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

108. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

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

110. Scott, A. D., Van de Velde, J. D. & Novikova, P. Y. Inference of Polyploid Origin and Inheritance Mode from Population Genomic Data. *Methods Mol. Biol.* **2545**, 279–295 (2023).

111. Neuffer, B. *et al.* S-alleles and mating system in natural populations of Capsella grandiflora (Brassicaceae) and its congeneric relatives. *Flora* **299**, 152206 (2023).

112. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.R-project.org/ (2022).

113. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* vol. 4 1686 Preprint at https://doi.org/10.21105/joss.01686 (2019).

114. Pebesma, E. Simple features for R: Standardized support for spatial vector data. *R J.* **10**, 439 (2018).

115. Hijmans, R. J. raster: Geographic Data Analysis and Modeling. Preprint at https://CRAN.R-project.org/package=raster (2023).

116. Slowikowski, K. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. Preprint at https://CRAN.R-project.org/package=ggrepel (2023).

117. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 5269–5273 (1979).

118. Korunes, K. L. & Samuk, K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).

119. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

120. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).