DOI: 10.1038/s41559-019-0807-4. https://www.nature.com/articles/s41559-019-0807-4

SUPPLEMENTARY MATERIALS

Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*

Patrick Monnahan, Filip Kolář, Pierre Baduel, Christian Sailer, Jordan Koch, Robert Horvath, Benjamin Laenen, Roswitha Schmickl, Pirita Paajanen, Gabriela Šrámková, Magdalena Bohutínská, Brian Arnold, Caroline M. Weisman, Karol Marhold, Tanja Slotte, Kirsten Bomblies, and Levi Yant

Supplemental notes

Supplementary Text 1: Evidence for single origin of tetraploids followed by admixture.

Joint analyses of diploid and tetraploid populations reveal that, in two parallel cases (Southern Carpathians and Baltic coast), local tetraploids clustered genetically with locally co-occurring diploids (Fig. 1, Fig S3, Fig. S4B), while still retaining a sizeable portion of polymorphisms shared with the widespread tetraploid lineages (*W. Carp-4x.*, *C. Europe-4x*, Fig 1b). Although such a general pattern could result from a scenario of multiple tetraploid origins followed by widespread gene flow among tetraploids, we find multiple lines of evidence that support a single origin of tetraploids followed by interploidy gene flow with geographically proximate diploids as most parsimonious:

First, coalescent simulations of population quartets involving populations from all tetraploid lineages consistently favour scenarios with a single tetraploid ancestor (~20 – 31 k generations ago; Table S9) followed by admixture (Fig 4a,b, Fig S9 - 10), in agreement with previous work [1].

Second, frequencies of alleles diagnostic of the putative diploid ancestor of all tetraploids (*W. Carp.-2x* lineage) are elevated (Fig. S23) and positively correlated (Fig. S11) across all tetraploid populations. In contrast, alleles specific to *S. Carp.-2x* or *Baltic-2x* were only frequent in geographically proximal tetraploid lineages (*S. Carp.-4x* and *Ruderal-4x*, respectively; Fig. 4c, Table S13). Under independent origins followed by admixture, we would expect independent sampling from the *W. Carp. 2x* allele pool by the different tetraploid lineages. Instead we observe widespread sharing of these alleles across multiple tetraploid lineages (Fig. S11).

The final source of evidence comes from the observed sharing of alleles at key meiosis genes among all tetraploids. These loci were first identified based on excess divergence between diploids and tetraploids in a small subset of populations outside the contact zones [2]. With our additional sampling, we confirm the majority of these loci present a characteristic pattern of elevated divergence between ploidies and reduced divergence among all tetraploid populations (Fig. S12). Moreover, even in areas where geographically proximate diploids and tetraploids cluster together (i.e. *S. Carp.-2x/S. Carp.-4x* and *Baltic-2x/Ruderal-4x*), sliding window phylogenetic tree topology weighting analyses (Twisst) indicate that windows containing these meiosis genes show highest support for topologies grouping tetraploids together, despite equivocal support for each topology genome-wide (Fig. 5; S24). Although it is formally possible that the *S. Carp.-4x* and *Ruderal-4x* tetraploids represent two additional, independent origins of tetraploids, this would also require two independent, parallel introgressions of the identical suite of meiotic adaptations among different tetraploid populations. In addition, such a scenario would require alternative meiosis stabilizing

mechanisms in the early formed tetraploid lineages in each contact zone, prior to introgression, as these genes are very likely essential for stability of meiosis and thus tetraploid fitness [2].

Supplementary Text 2: Functional annotation of the introgressed and introgression-resistant loci.

We found multiple interploidy introgressed regions indicative of recent selection. For example, in the Southern Carpathian contact zone, among the top outliers for the introgression topology (topology 3), two overlapped with an outlier for Fay and Wu's H. The first outlier is a novel region containing several gene coding loci (Table S10), including a root-expressed extensin-like family protein (AL8G22520) and the second contains 2 candidate genes: LSF1 (AL3G10660), a putative phosphatase involved in nocturnal starch degradation and betaCA1 (AL3G10670), a carbonic anhydrase that regulates stomatal movements in guard cells. Outliers in the Baltic-Ruderal contact zone identified 2 regions (Table S10), containing 4 gene-coding loci (AL7G24200, AL7G24210, AL7G24220, and AL8G17090). We further searched for potential functional relevance of introgressed regions by performing gene ontology (GO) enrichment analyses on the outlier genes for the introgression topology from each contact zone as well as the overlap. There was no significant enrichment, following Bonferroni correction, for the introgression topology outliers in either contact zone, individually. However, the overlapping outliers for both contact zones (4 windows encompassing 11 genes) was significantly enriched for genes involved in 'reproductive process' (GO: 0022414).

We also find multiple cases of the opposite pattern, i.e., excessive weight for the topology clustering them with other tetraploids while keeping divergence from all diploids. In these candidate introgression-resistant loci, the elevated divergence from the co-occurring diploids may reflect selection against such interploidy gene flow. Such introgression-resistant outliers exhibited a unique set of enriched functional categories in each contact zone. For the South Carpathian contact zone, the following GO terms were enriched: nucleic acid metabolic process (GO:0090304), cellular macromolecule metabolic process (GO:0044260), and RNA processing (GO:0006396). For the Baltic-Ruderal contact zone, genes involved in monovalent inorganic cation transport were enriched (GO:0015672). Outliers in common for both contact zones were enriched for both 7-methylguanosine mRNA capping (GO:0006370) and RNA capping (GO:0036260).

Supplementary Text 3: Considering the impacts of non-selective processes on the estimation of the adaptive substitution rate (DFE-alpha).

A number of non-selective processes can lead to misestimation of parameters within the DFE-alpha and similar methods [3-5] (see Supplementary Text 3 for further discussion). First, demographic history can have a large impact on the allele frequency spectrum, on which DFE-associated methods are based [3, 4]. For this reason, we chose the updated version [3], which accommodates a single population size change, over the original DFE method [6], which assumed constant population sizes. However, it was recently shown that, while this method is robust to recent population size changes, it is not robust to ancient, multiply-repeated fluctuations [4]. Fortunately, the time scale investigated (in [4]) is on the order of several hundreds of thousands of generations, involving a large number of fluctuations (~9 based on their Figure 1), and the origin of tetraploids is recent enough (~30,000 generations) that any overestimation due to ancient fluctuations should not differentially influence the rate of adaptive substitutions.

It has also recently been shown that multi-nucleotide substitutions, wherein DNA

polymerases generate mutations at immediately adjacent sites, can disproportionately affect nonsynonymous sites and lead to false conclusions of lineage-specific positive selection [5]. Although this phenomenon is expected to impact a very small proportion of mutations (estimated at 0.4% of all mutations in *Drosophila*), any differential bias of nonsynonymous over synonymous sites will positively bias tests for positive selection. Given enough time, this bias can be substantial.

We investigated the potential for these multi-nucleotide substitutions (MNS) occurring differentially between diploids and tetraploids by comparing the probability that two variants are adjacent to one another in each population. We used the data for our LD analysis, as it equilibrates sample size across populations. For diploids, the probability that two sites were adjacent to each other was 0.044 versus 0.053 in tetraploids. However, when we consider that tetraploids have double the number of chromosomes as diploids, the per-chromosome probability becomes 0.022 for diploids vs 0.013 for tetraploids. Given these slight differences and the short time scale over which any differential MNS could occur between ploidies (Venkat et al 2018 was comparing highly divergent species on a phylogenetic time scale), we believe that it is unlikely that MNS could be driving the highly significant difference that we observe in 2E.

1. Arnold B, Kim S-T, Bomblies K. Single Geographic Origin of a Widespread Autotetraploid Arabidopsis arenosa Lineage Followed by Interploidy Admixture. Molecular Biology and Evolution. 2015;32(6):1382-95. doi: 10.1093/molbev/msv089.
2. Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, Franklin FCH, et al. Meiotic adaptation to genome duplication in Arabidopsis arenosa. Current biology. 2013;23(21):2151-6.
3. Eyre-Walker A, Keightley PD. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Molecular biology and evolution. 2009;26(9):2097-108.
4. Rousselle M, Mollion M, Nabholz B, Bataillon T, Galtier N. Overestimation of the adaptive substitution rate in fluctuating populations. Biology Letters. 2018;14(5).
5. Venkat A, Hahn MW, Thornton JW. Multinucleotide mutations cause false inferences of lineage-specific positive selection. Nat Ecol Evol. 2018;2(8):1280-8. doi: papers3://publication/doi/10.1038/s41559-018-0584-5.
6. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nature Reviews Genetics. 2007;8(8):610-8.

Supplemental Figures

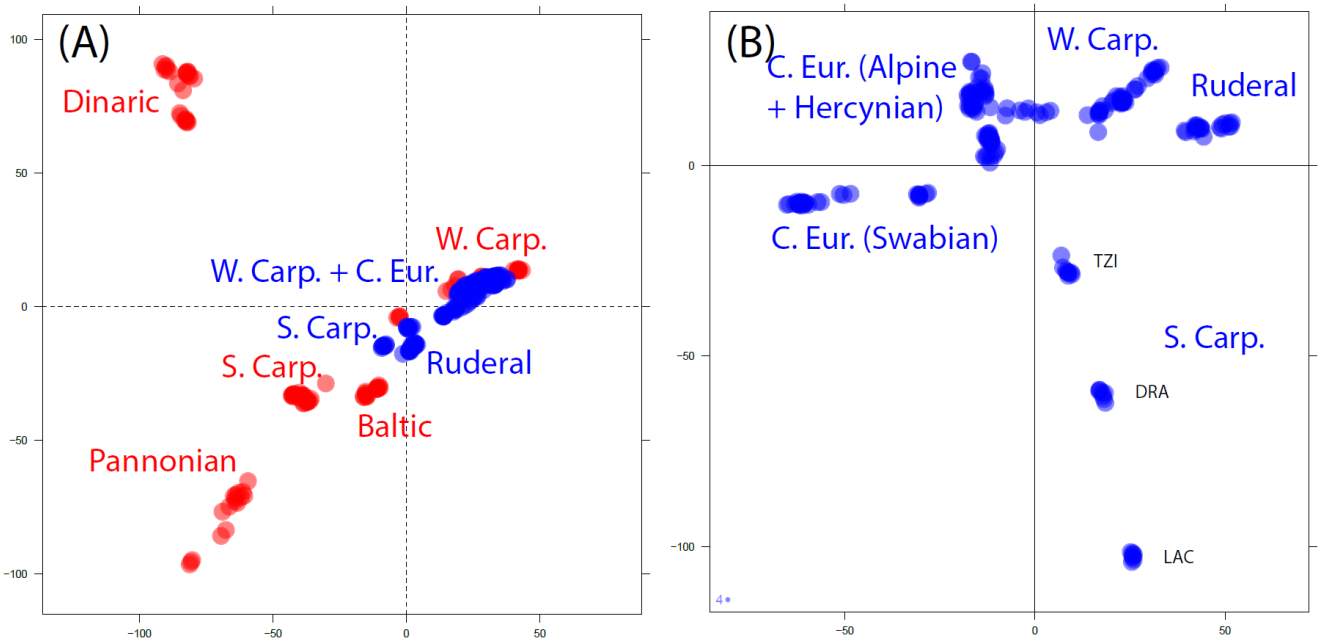


Figure S1 Principal component analysis of *A. arenosa* diploid (red) and tetraploid (blue) individuals calculated with (A) complete *A. arenosa* dataset and (B) tetraploid individuals only.

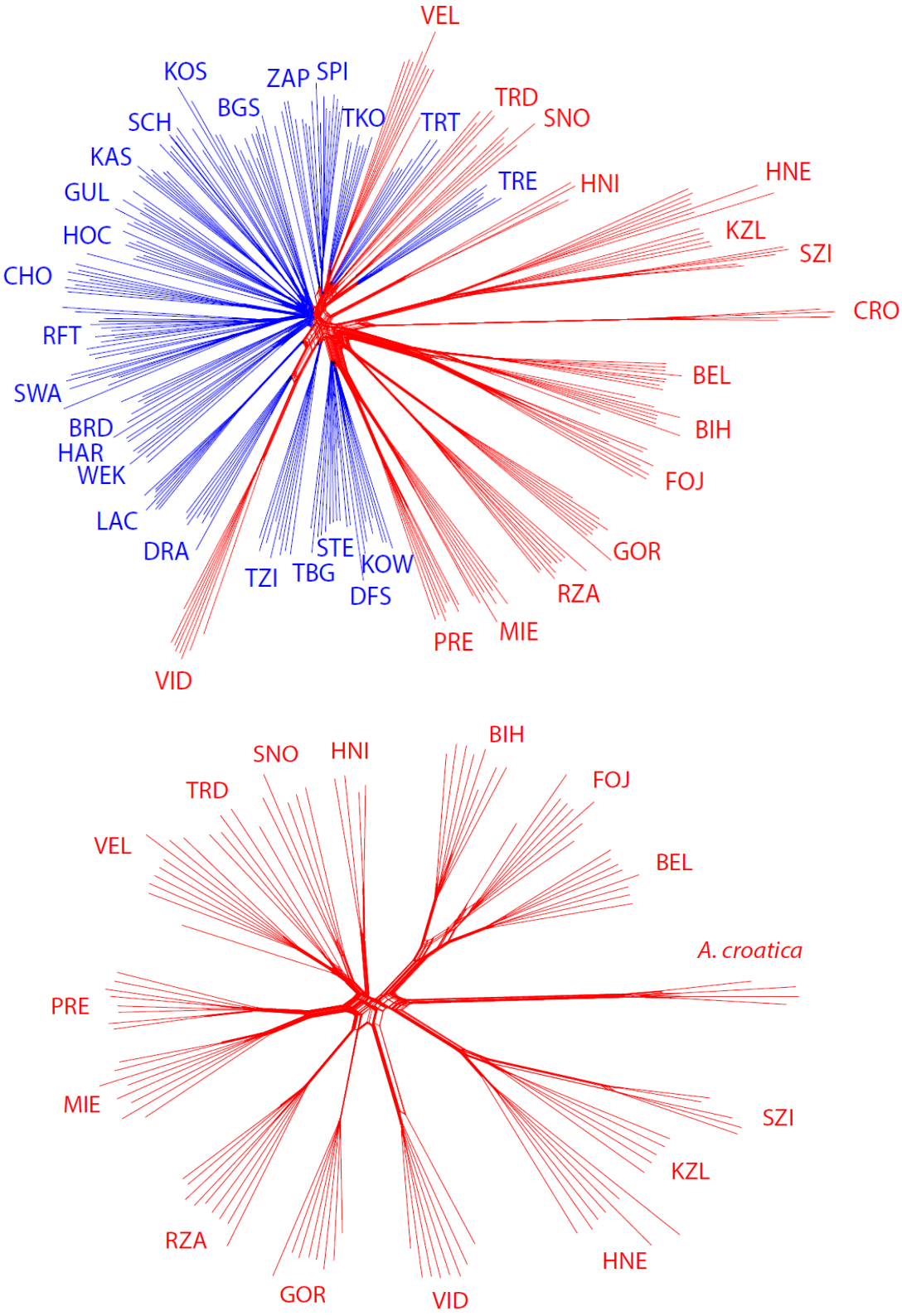


Figure S2 Neighbor-joining networks based on Nei's genetic distances among (top) all the 287 diploid (red) and tetraploid (blue) *A. arenosa* and four *A. croatica* individuals and (bottom) only the 109 diploid individuals.

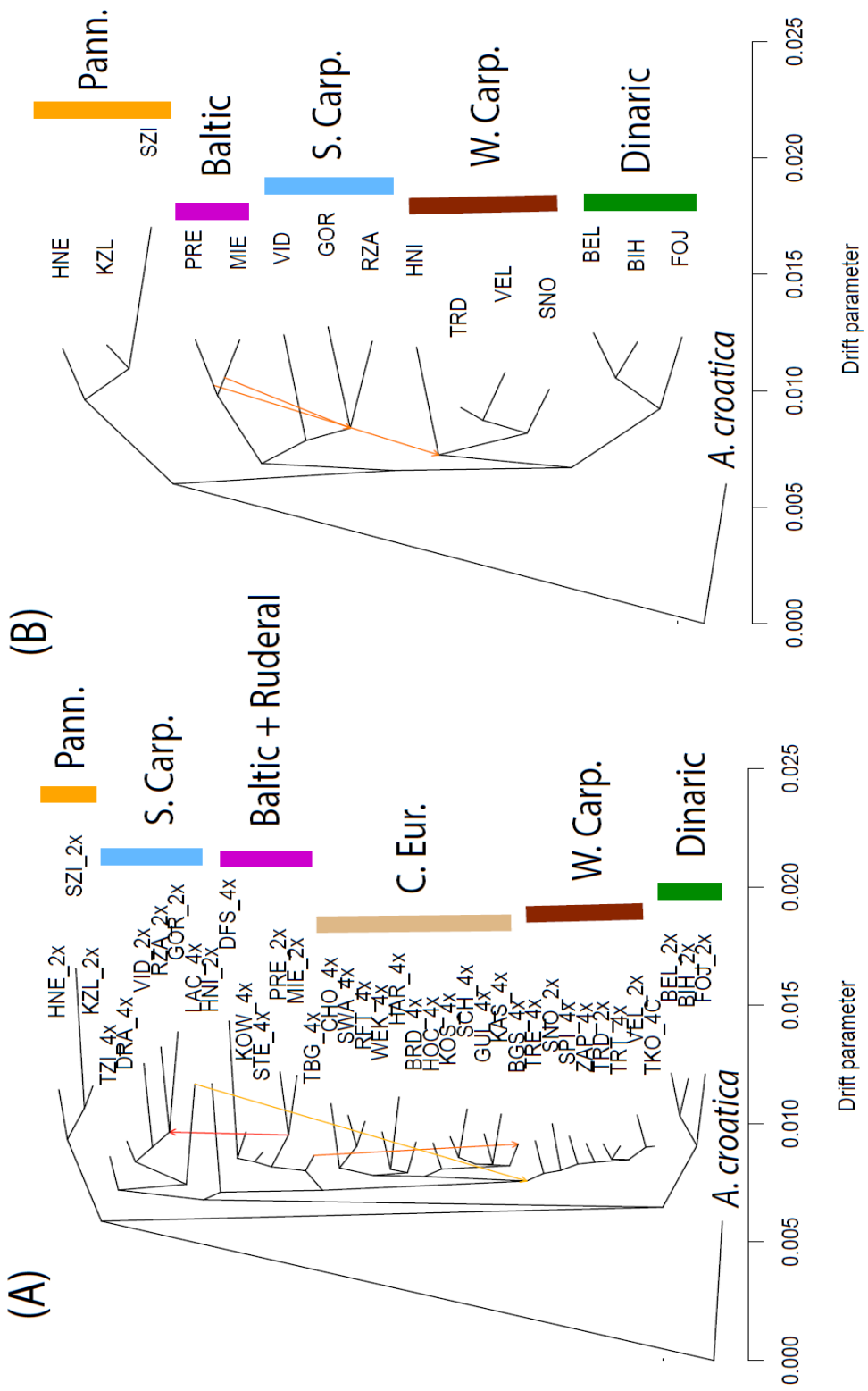
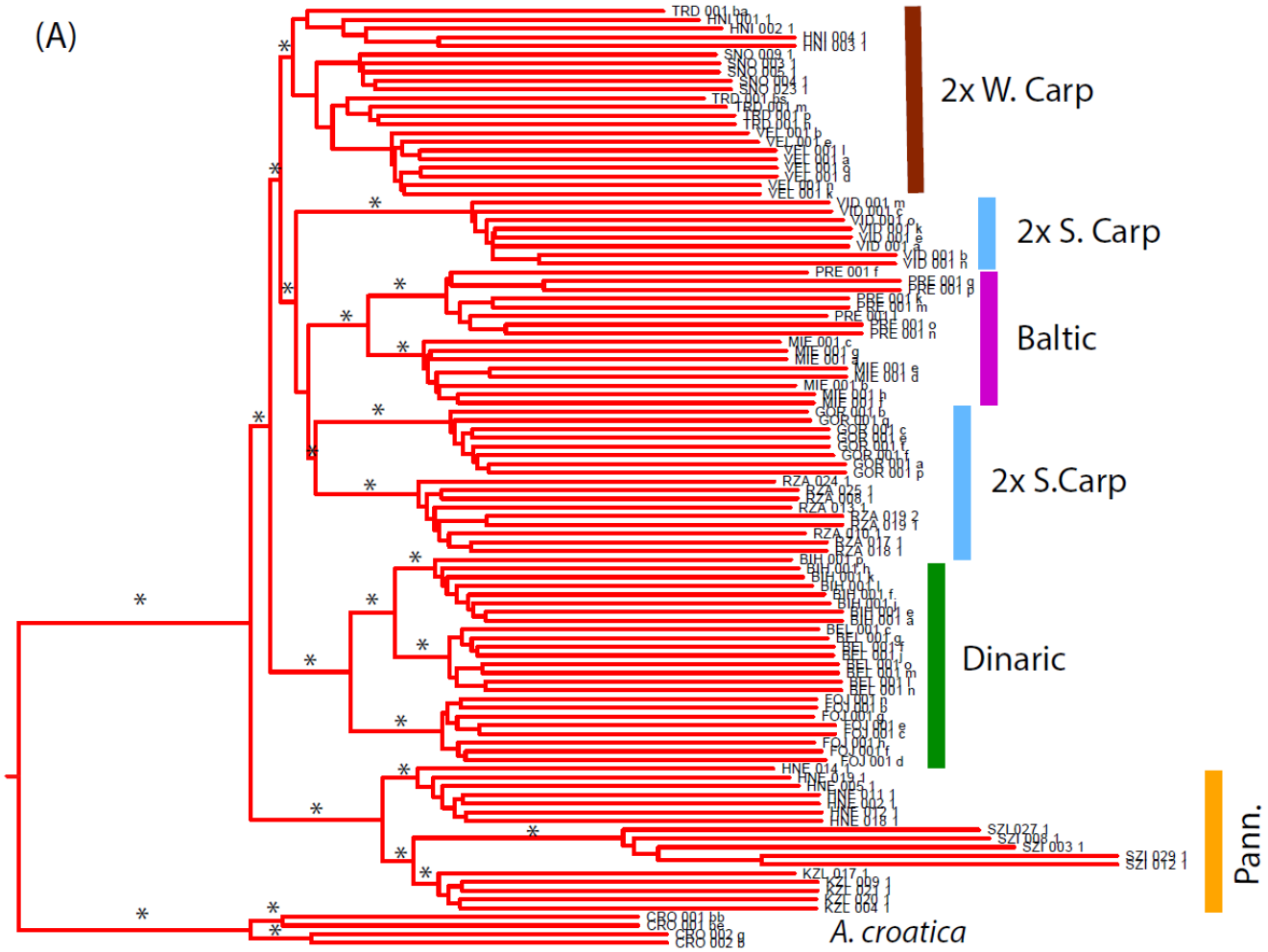


Figure S3 Maximum likelihood graph of allele frequency covariance inferred by Treemix analysis of (A) all and (B) diploid populations only. The colour bars correspond to major lineages as in Fig 1.


(A)
2x W. Carp
2x S. Carp
Baltic
2x S.Carp
Dinaric
Pann.
A. croatica

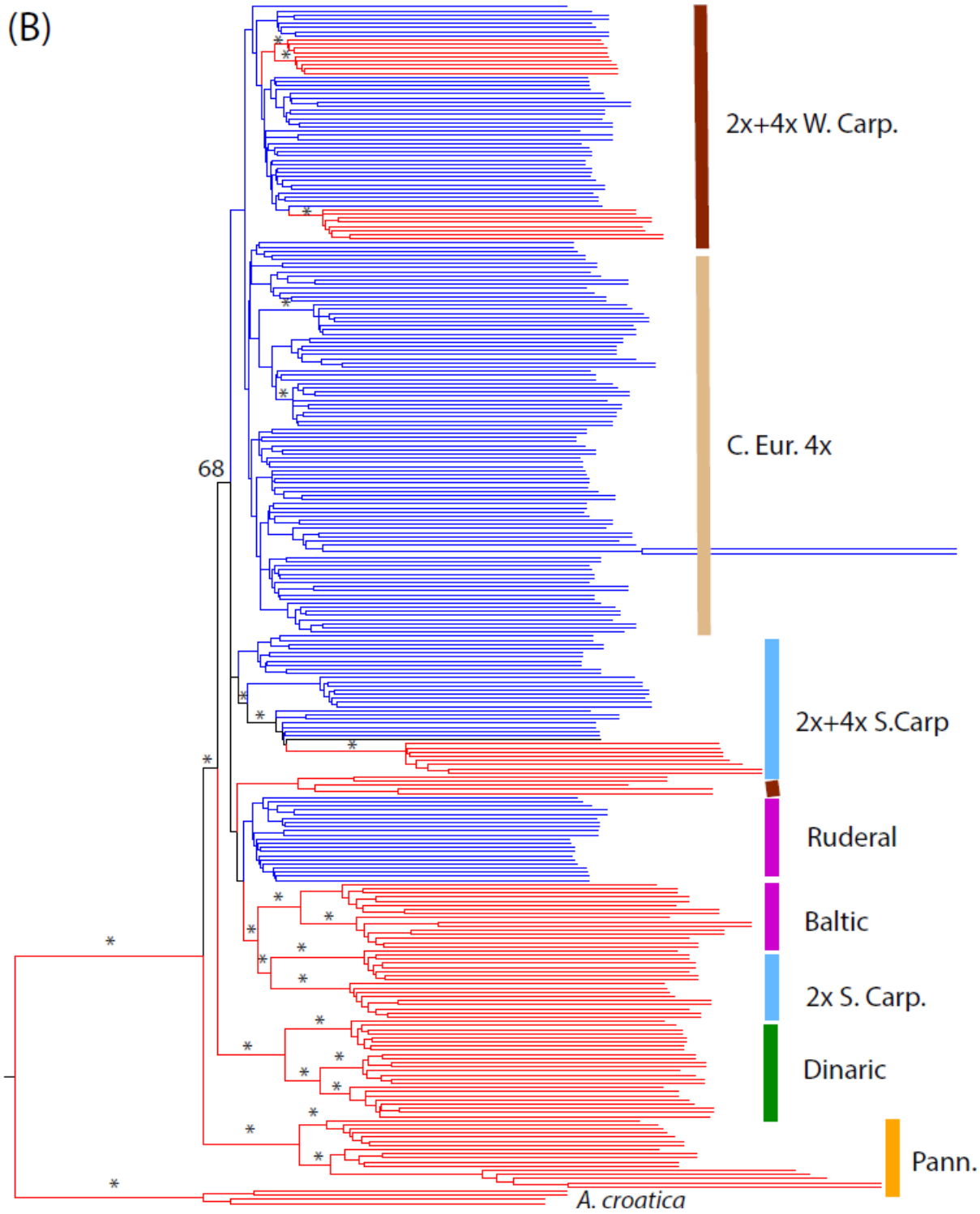
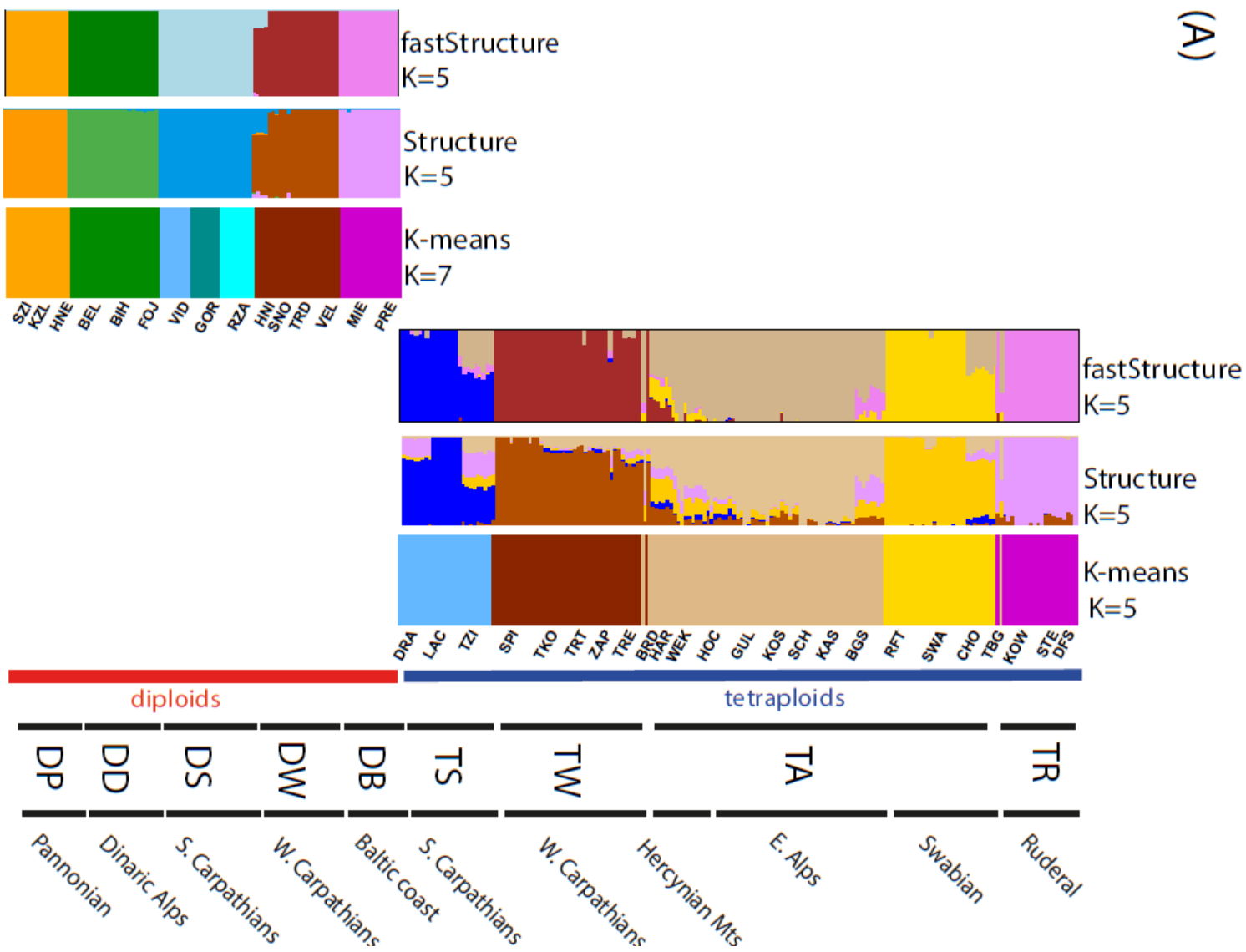


Figure S4 Phylogeny of diploid only (A) and all (B) *A. arenosa* individuals inferred from multispecies coalescent reconstruction of 503 (A) and 553 (B) gene trees using Astral. Branches are coloured according to ploidy level, the major lineages are marked with colours corresponding to Fig. 1, nodes with bootstrap values above 75% are denoted by an asterisk.


(A)
fastStructure
K=5
Structure
K=5
K-means
K=7
SZI KZL HNE BEL BIH FOJ VID GOR RZA HNI SNO TRD VEL MIE PRE
fastStructure
K=5
Structure
K=5
K-means
K=5
DRA LAC TZI SPI TKO TRT ZAP TRE BRD HAR WEK HOC GUL KOS SCH KAS BGS RFT SWA CHO TBG KOW STE DFS
diploids
tetraploids
DP
DD
DS
DW
DB
TS
TW
TA
TR
Pannonian
Dinaric Alps
S. Carpathians
W. Carpathians
Baltic coast
S. Carpathians
W. Carpathians
Hercynian Mts
E. Alps
Swabian
Ruderal

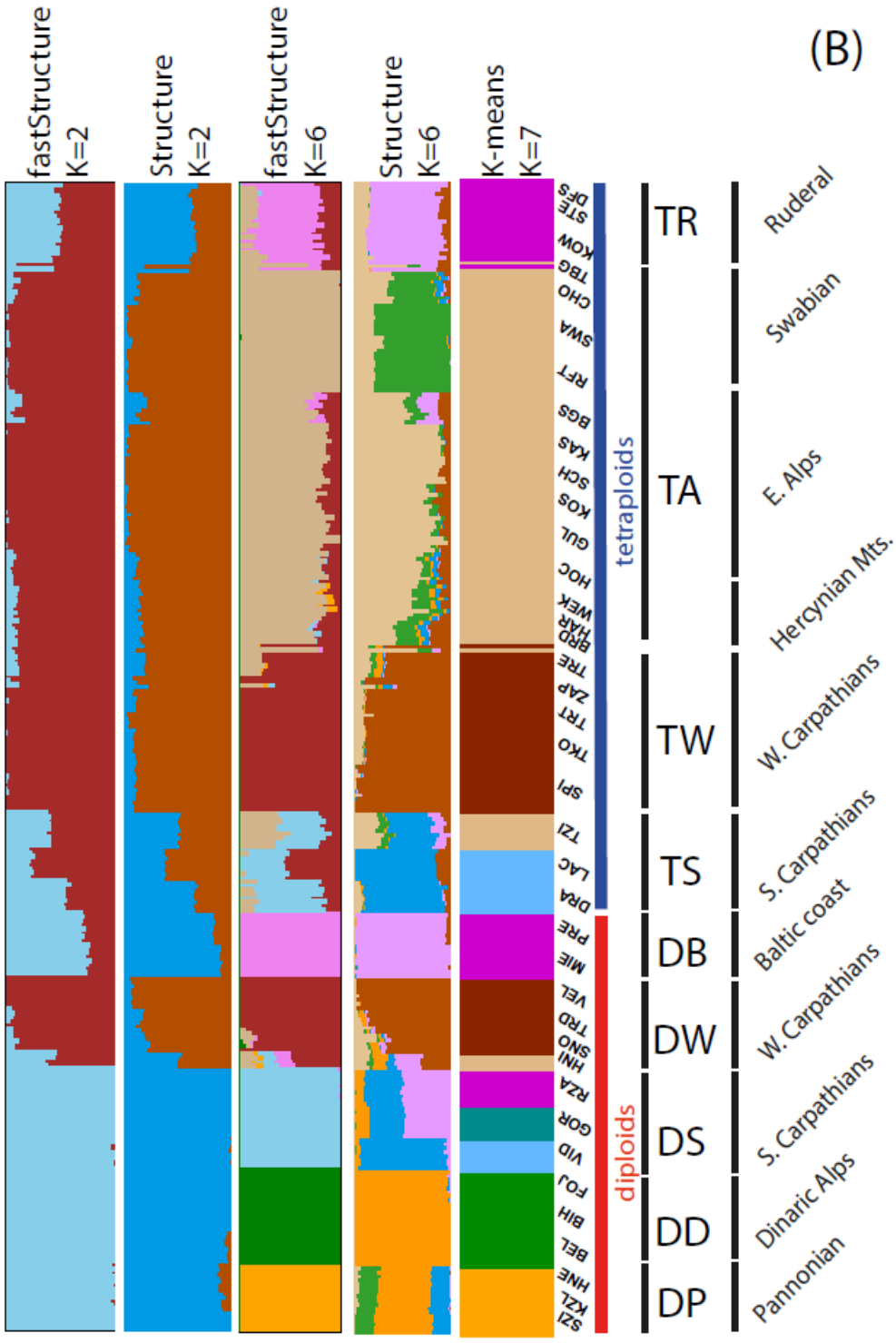


Figure S5 Comparison of group assignment of the 287 *A. arenosa* individuals under different model-based (Structure, fastStructure) and non-parametric methods (K-means clustering). Diploid and tetraploid individuals were analysed separately (A) and together (B). The plausible number of clusters (K) was selected based on convergence to similar results over five replicate runs (fastStructure) and Bayesian information criterion (K-means clustering).

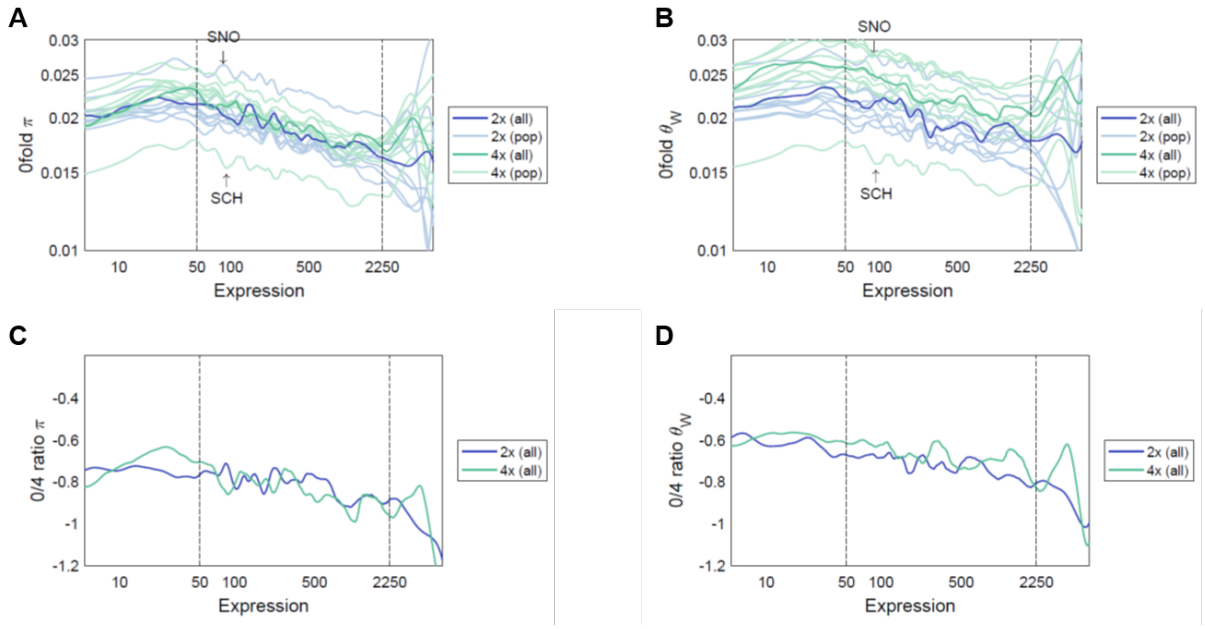


Figure S6. Increased non-synonymous diversity across expression levels. (A, B) LOWESS plots of gene-wise estimates of non-synonymous diversity ($\theta\pi$ and θ_W at 0-dg sites) against their respective log expression for genes non-differentially expressed (NDE) between ploidies for each population (faint lines) and ploidy (bold lines). (C, D) LOWESS of gene-wise 0-dg/4-dg log ratios of $\theta\pi$ and θ_W against their respective log expression for NDE genes. Note, that the trend seems to break for very high expressions (>2250 i.e. top 0.35%) possibly due to the low coverage of this expression range (67 genes).

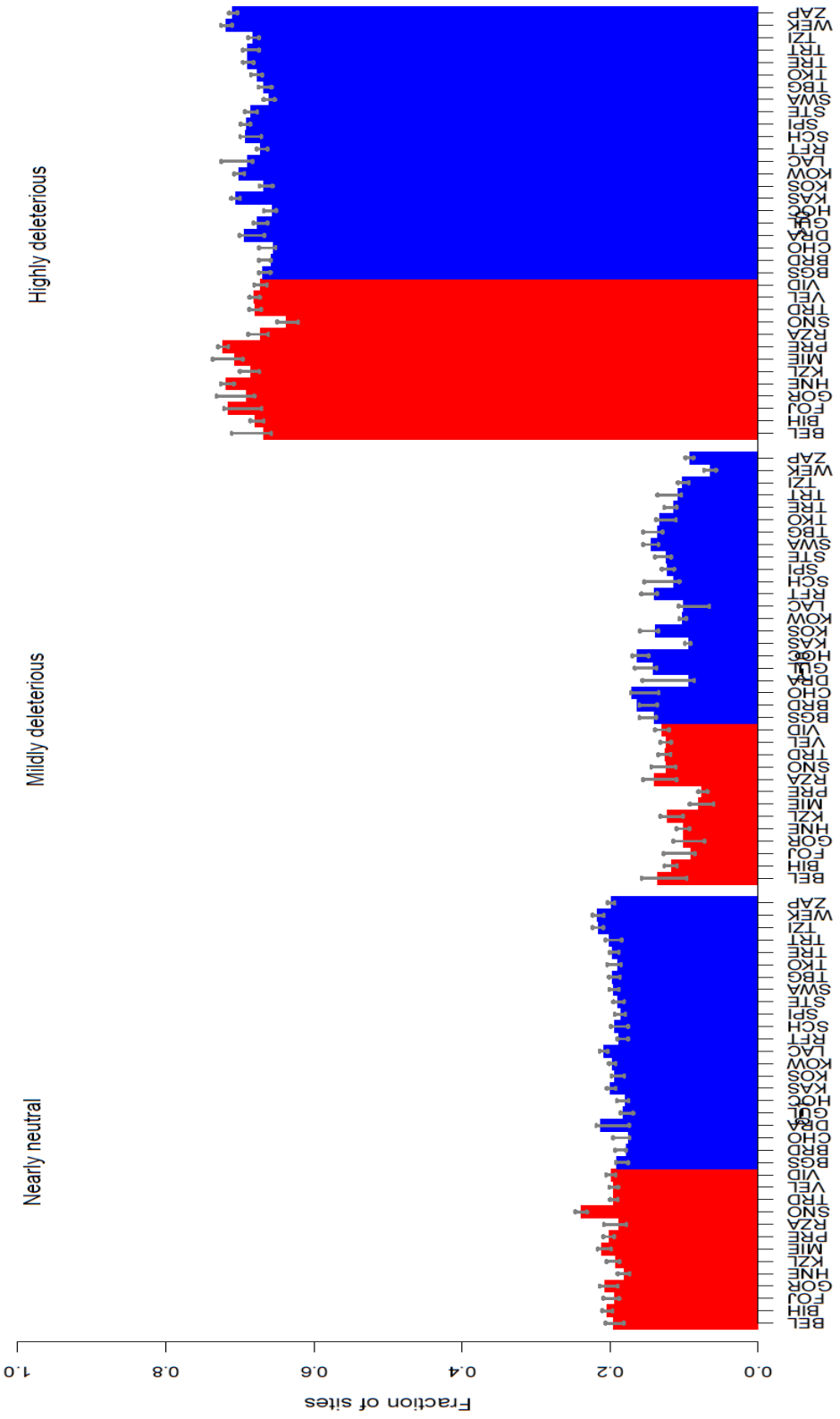


Figure S7 Distribution of fitness effect (DFE) of all populations based on folded site frequency spectra (diploid red, tetraploid blue). For each population, the DFE is presented in three bins of increasing strength of purifying selection. Errors bars represent 95% confidence interval based on 200 bootstraped replicates.

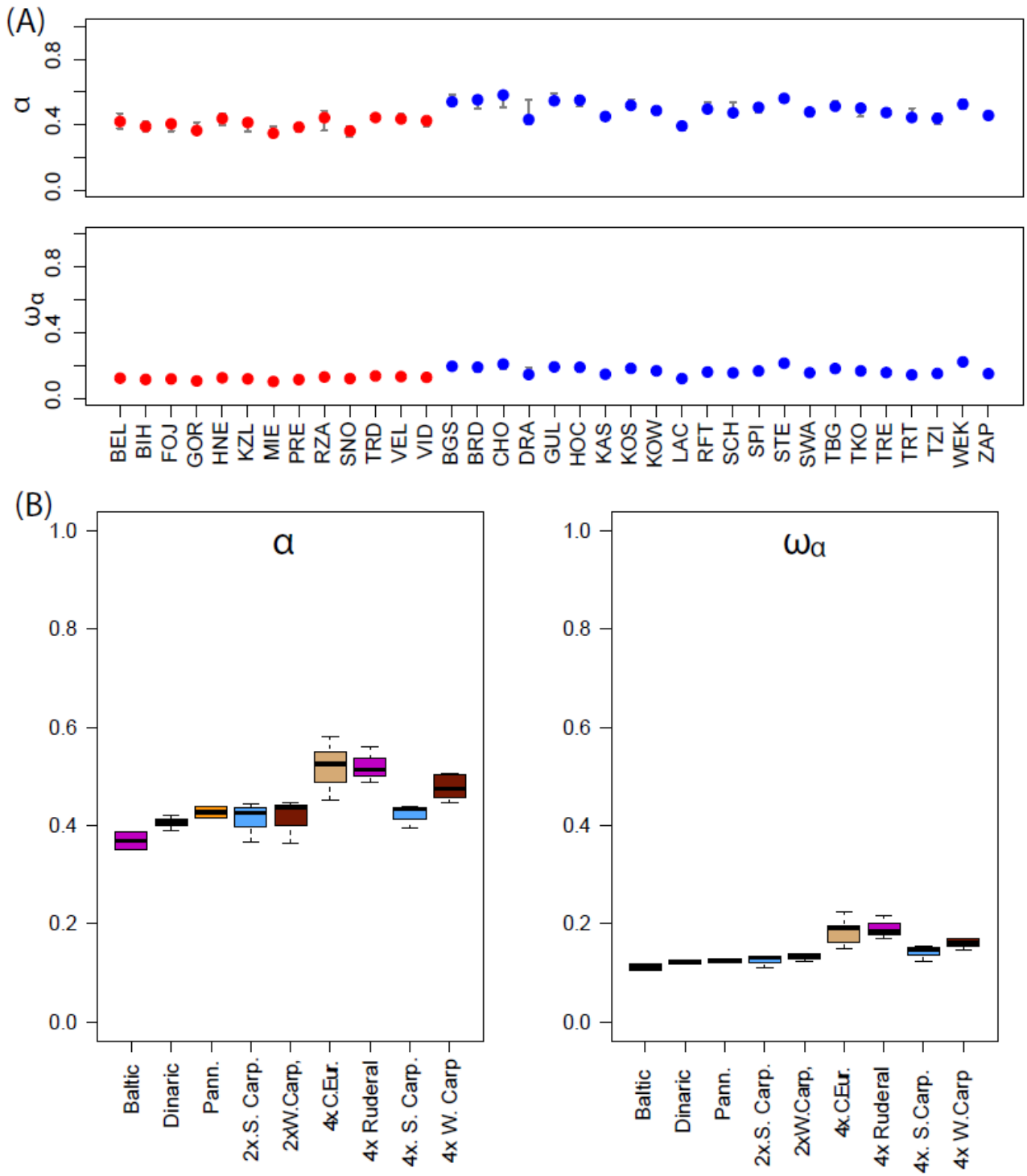


Figure S8. Proportion of adaptive substitution (α) and proportion of adaptive substitution relative to neutral (ω_α) of all populations (A) and populations grouped to major lineages (B) based on folded site frequency spectra. Errors bars represent 95% confidence interval based on 200 bootstraps replicates.

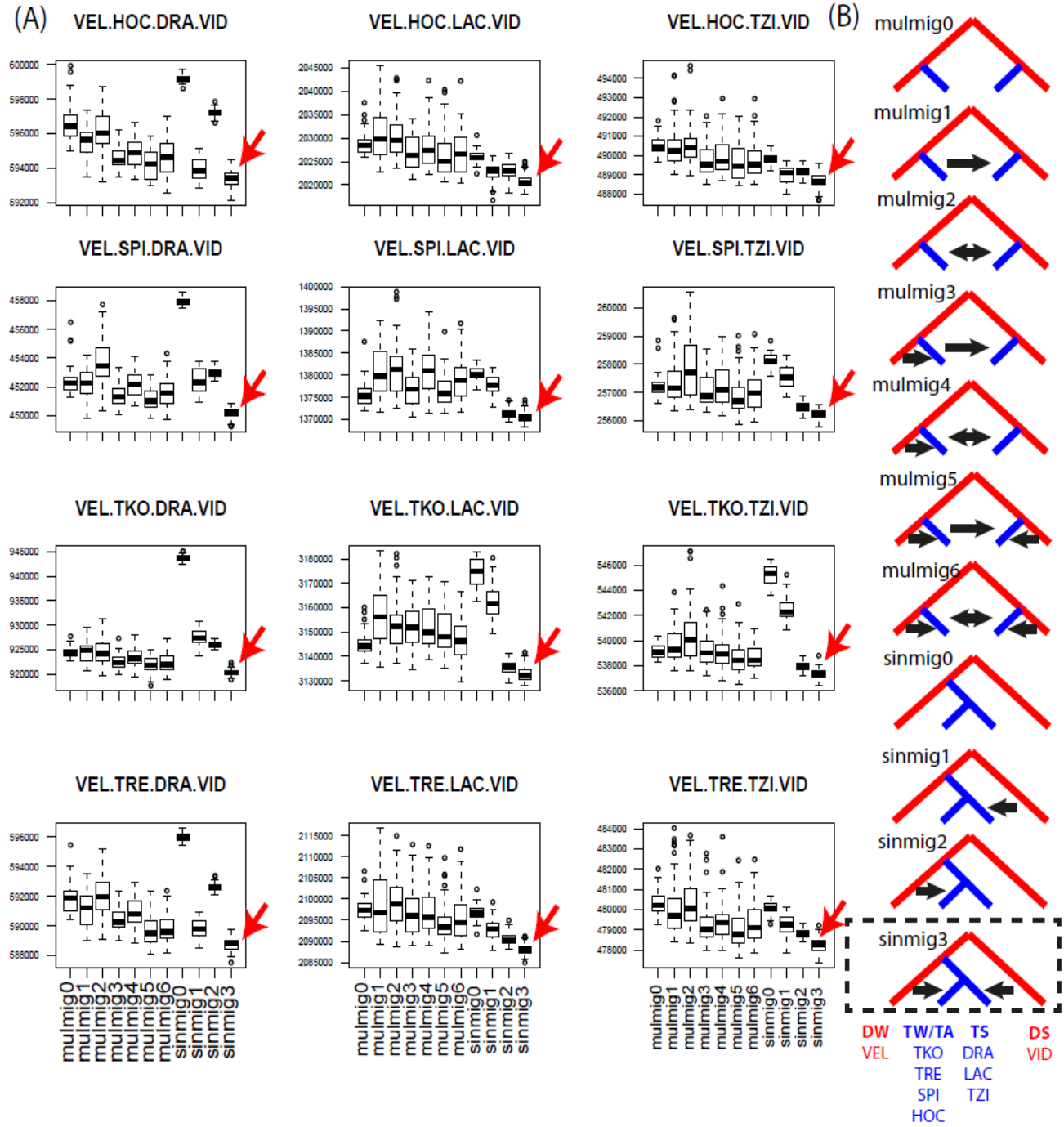


Figure S9 (A) Comparison Akaike information criteria (AIC) across 11 scenarios approximating the origin of the *S. Carp.-4x* tetraploids (populations DRA, LAC, TZI) either assuming independent origin from the local *S. Carp.-2x* diploids (*mulmig*) or divergence form a single common ancestor of the tetraploid populations (*sinmig*). Each scenario was simulated by 50 independent runs of fastsimcoal2 for various combination of population quartets, the corresponding distribution of the AIC values over these 50 runs are shown by the boxplots. Red arrow highlights the consistently most likely scenario (*sinmig 3* in all cases). (B) Topologies and direction of migration (forward in time) of the evaluated scenarios.

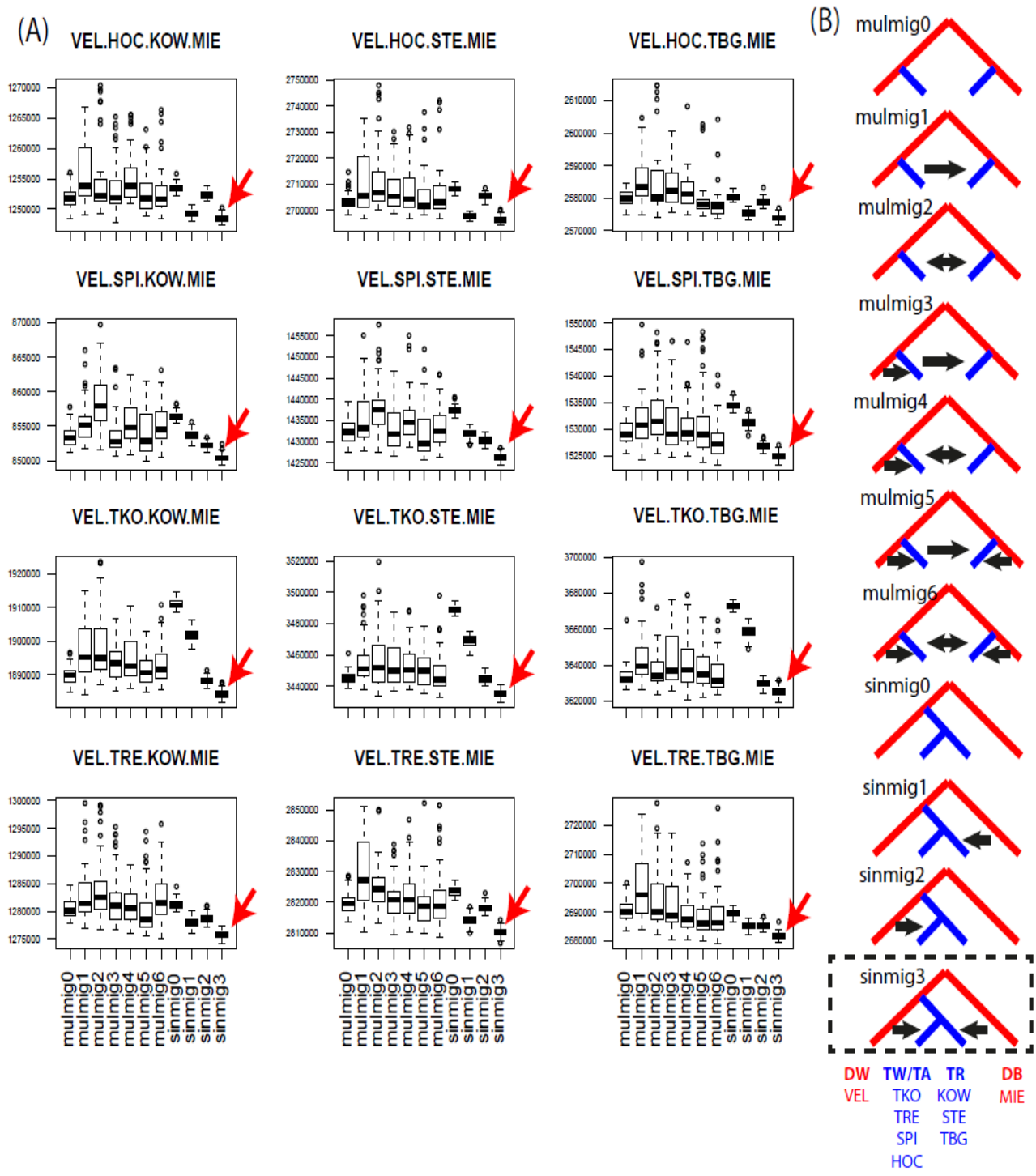


Figure S10 (A) Comparison Akaike information criteria (AIC) across 11 scenarios approximating the origin of the *Ruderal-4x* (populations KOW, STE, TBG) either assuming independent origin from the *Baltic-2x* (*mulmig*) or divergence form a single common ancestor of the tetraploid populations (*sinmig*). Each scenario was simulated by 50 independent runs of fastsimcoal2 for various combination of population quartets, the corresponding distribution of the AIC values over these 50 runs are shown by the boxplots. Red arrow highlights the most likely scenario (*sinmig 3*; the selection against scenario *sinmig 1* was equivocal in quartets involving HOC). (B) Topologies and direction of migration (forward in time) of the evaluated scenarios.

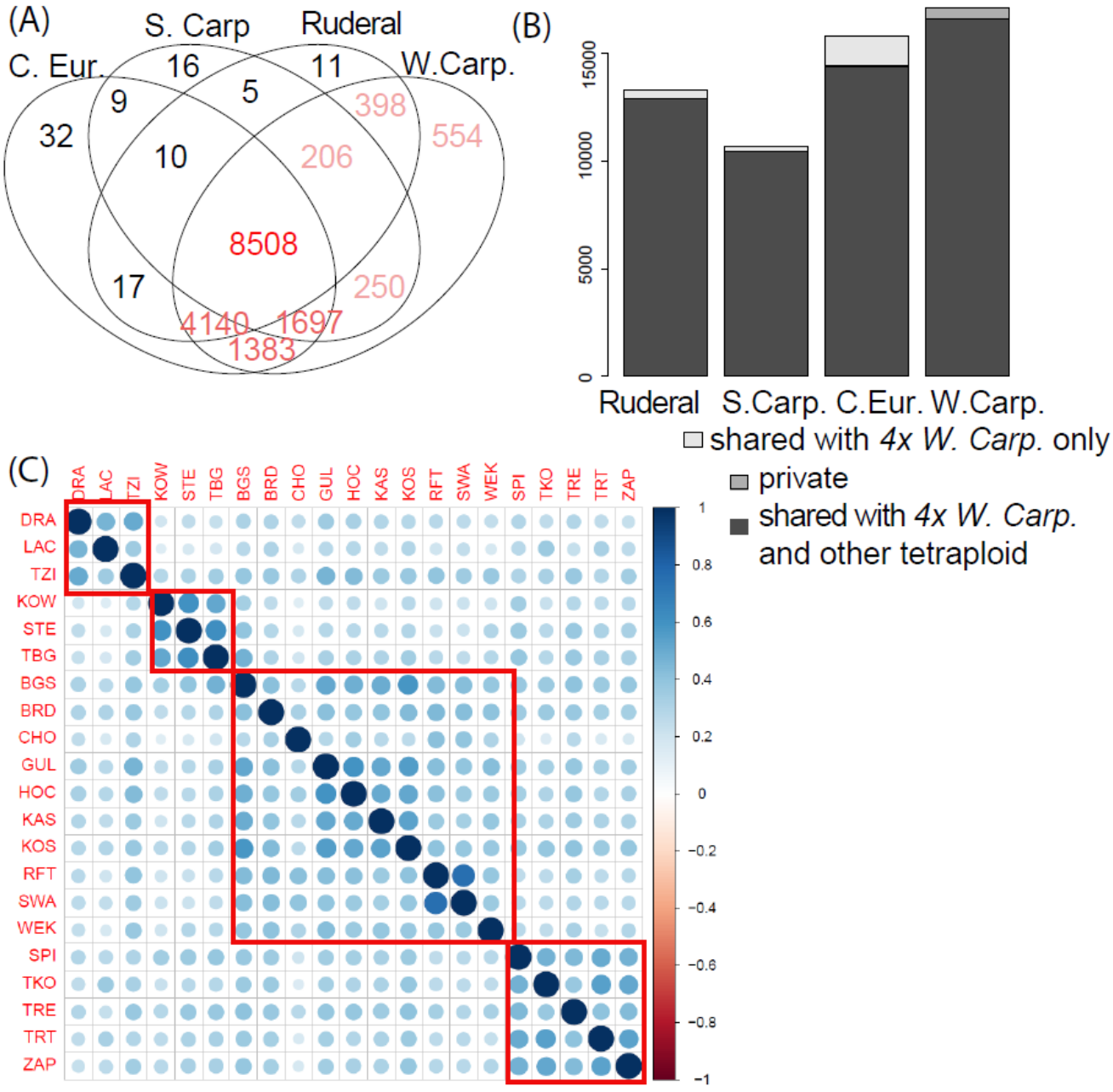


Figure S11 Frequency at which the *W. Carp.-2x* specific alleles were shared among tetraploid populations and lineages. (A) Venn diagram showing number of shared alleles among populations belonging to each tetraploid lineage, (B) Virtual absence of sharing of *2x W. Carpathian*-specific polymorphisms exclusively between the *W. Carp-4x* group and each other tetraploid lineage, (C) positive correlation of *W. Carp-2x* specific allele frequencies both in populations from the same (red square) and different (outside the square) tetraploid groups. The size and shade of the dot is proportional to the correlation coeffiecient. NB: Note that alleles specifically diagnostic to the putatively ancestral diploid *W. Carp.-2x* group are present at high frequencies in all tetraploid populations and their frequencies are correlated. Under independent origins, we would expect independent sampling from the *W. Carp.-2x* allele pool by the different tetraploid lineages as alleles would be separately introgressed into separate tetraploid lineages in different geographic regions.

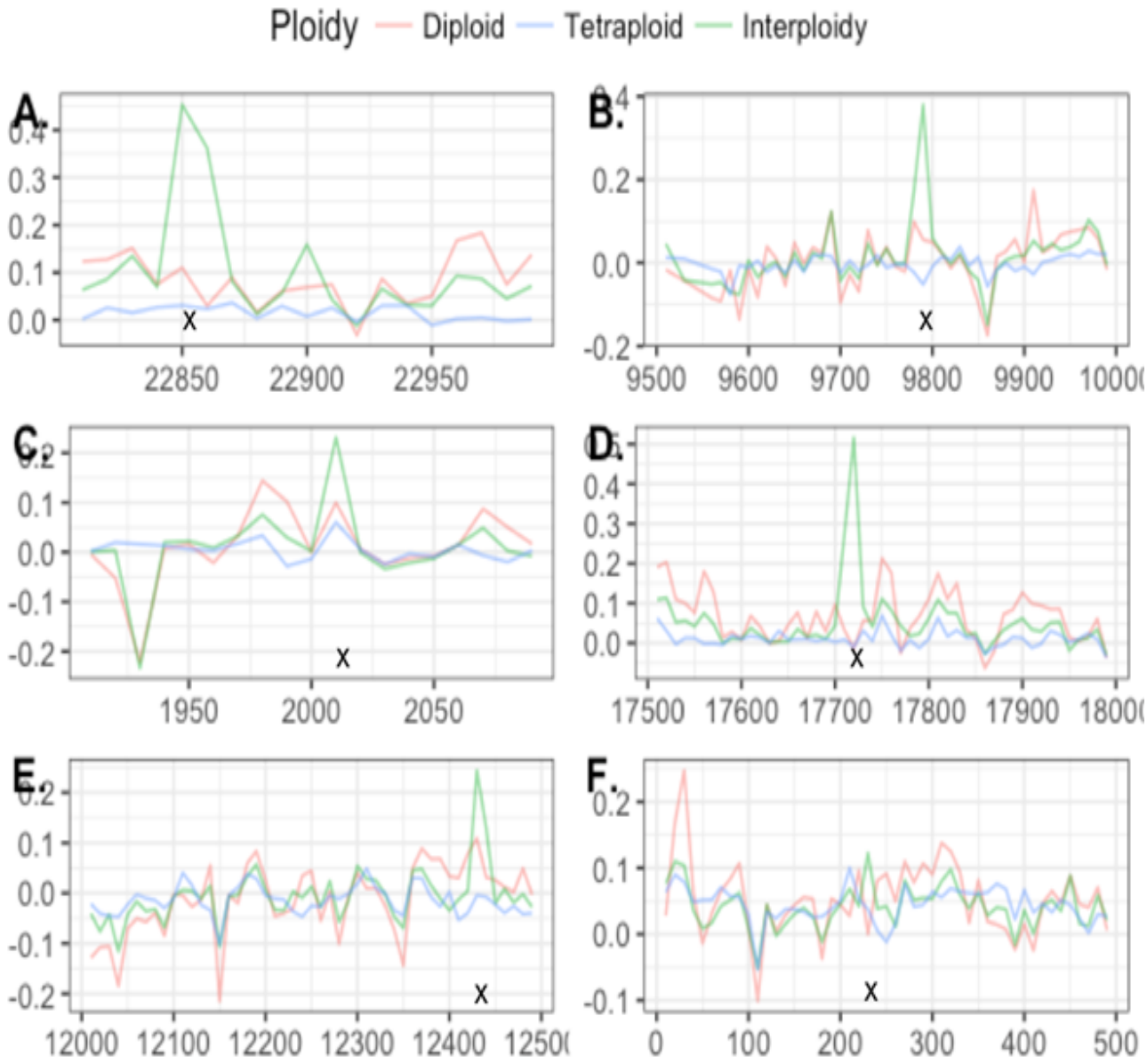


Figure S12. Excessive among-ploidy divergence in six key genes involved in the meiosis pathway. Standardized Fst is on the y-axis and genomic position in kb is on the x-axis. Actual location of the meiosis gene is demarcated with an 'x'. (A) ASY3, (B) ZYP1a/b, (C) SYN1, (D) PDS5B, (E) ASY1, (F) SMC3.

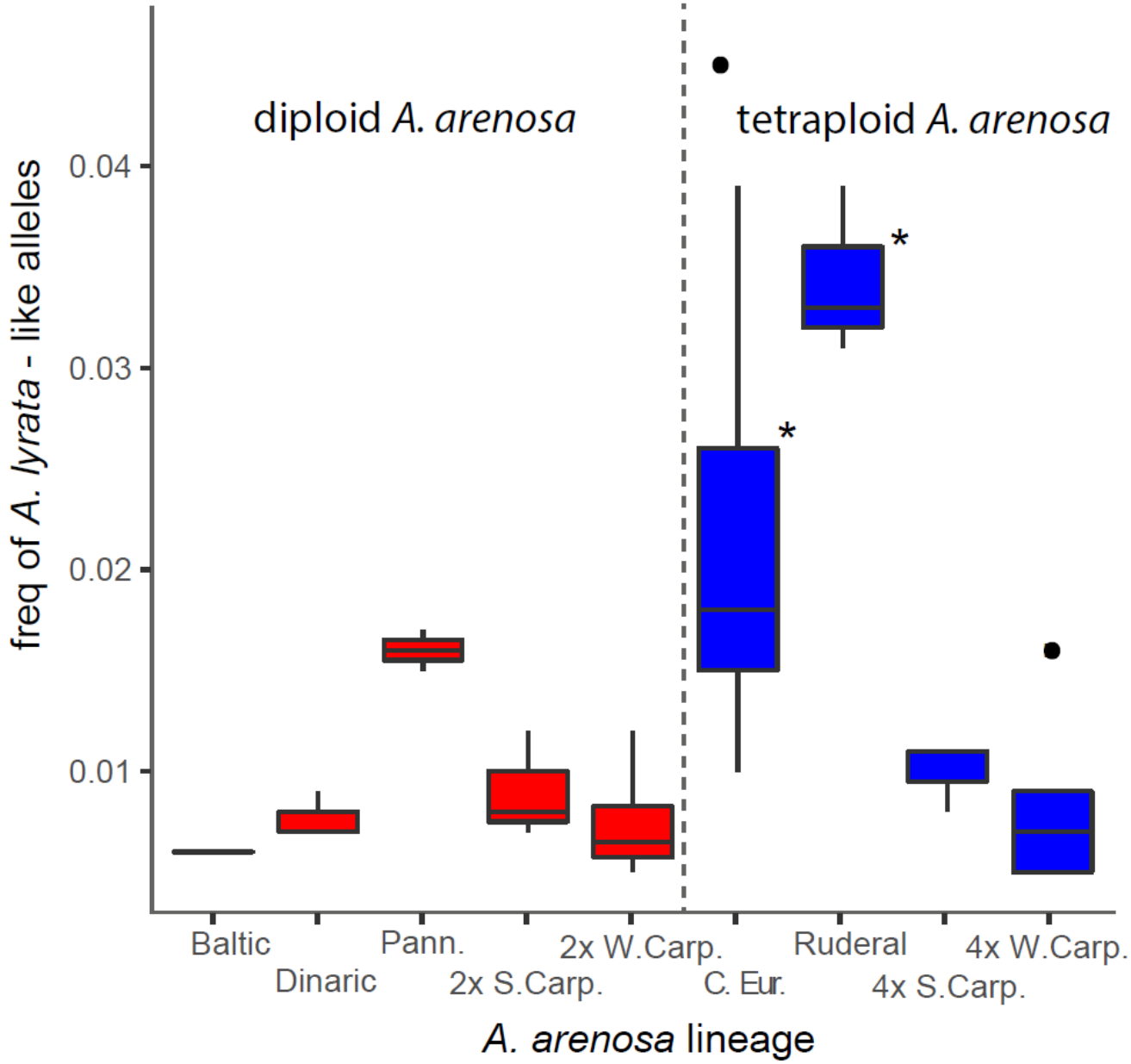


Figure S13 Frequency of rare reference *A. lyrata* alleles (max 0.07 frequency in total *A. arenosa* dataset) in different diploid and tetraploid *A. arenosa* lineages. Lineages with significantly ($\alpha = 0.05$) increased levels of *A. lyrata*-like alleles are denoted by an asterisk.

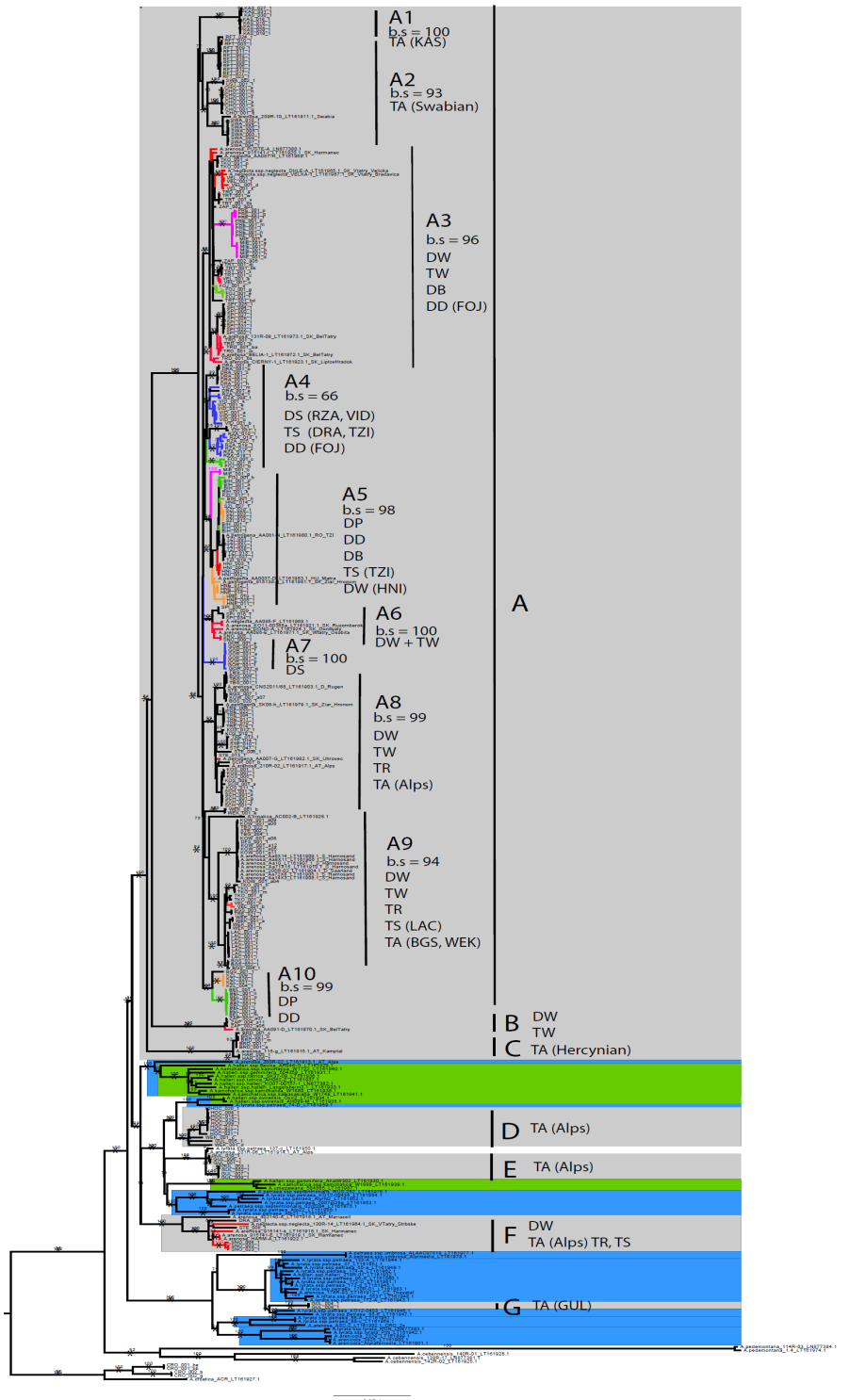
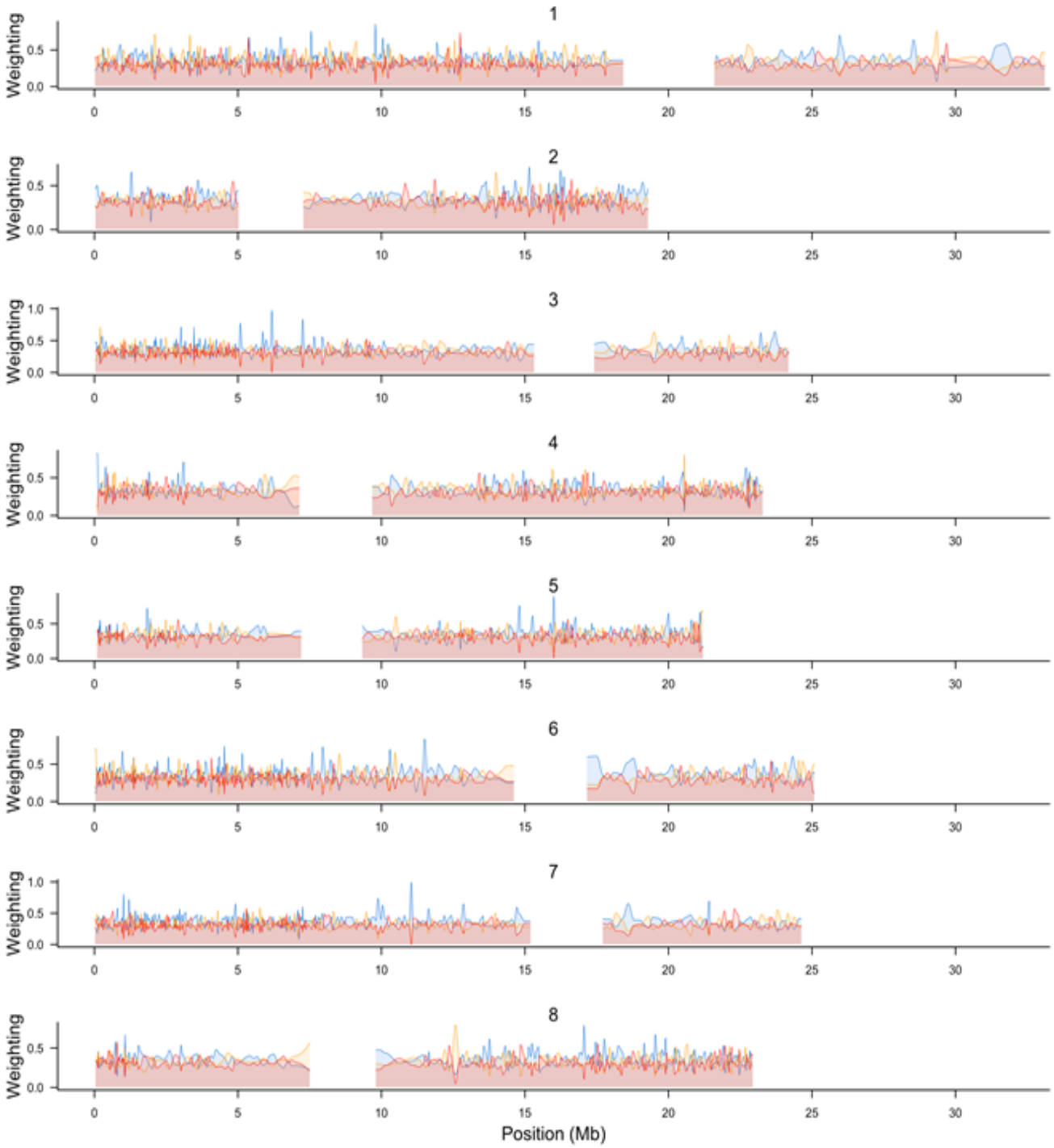


Figure S14 Maximum likelihood phylogeny of *Arabidopsis* plastomes from our study and of Novikova et al. (2016) rooted with *A. croatica* (*A. thaliana* was excluded to aid visibility). Background colour correspond to the species (blue: *A. lyrata*, green: *A. halleri*, gray: *A. arenosa*), branch colour to diploid lineages of *A. arenosa* (orange: *Pannonian-2x*, green: *Dinaric-2x*, violet: *Baltic-2x*, blue: *S. Carp.-2x*, red: *W. Carp.-2x*) and tetraploids (black). Asterisk denote bootstrap support above 75 %. See supplementary File S6 for the full resolution pdf file.


1
2
3
4
5
6
7
8
Weighting
Position (Mb)
DS
TS
TW
DD
TS
DS
TW
DD
DS
TS
TW
DD

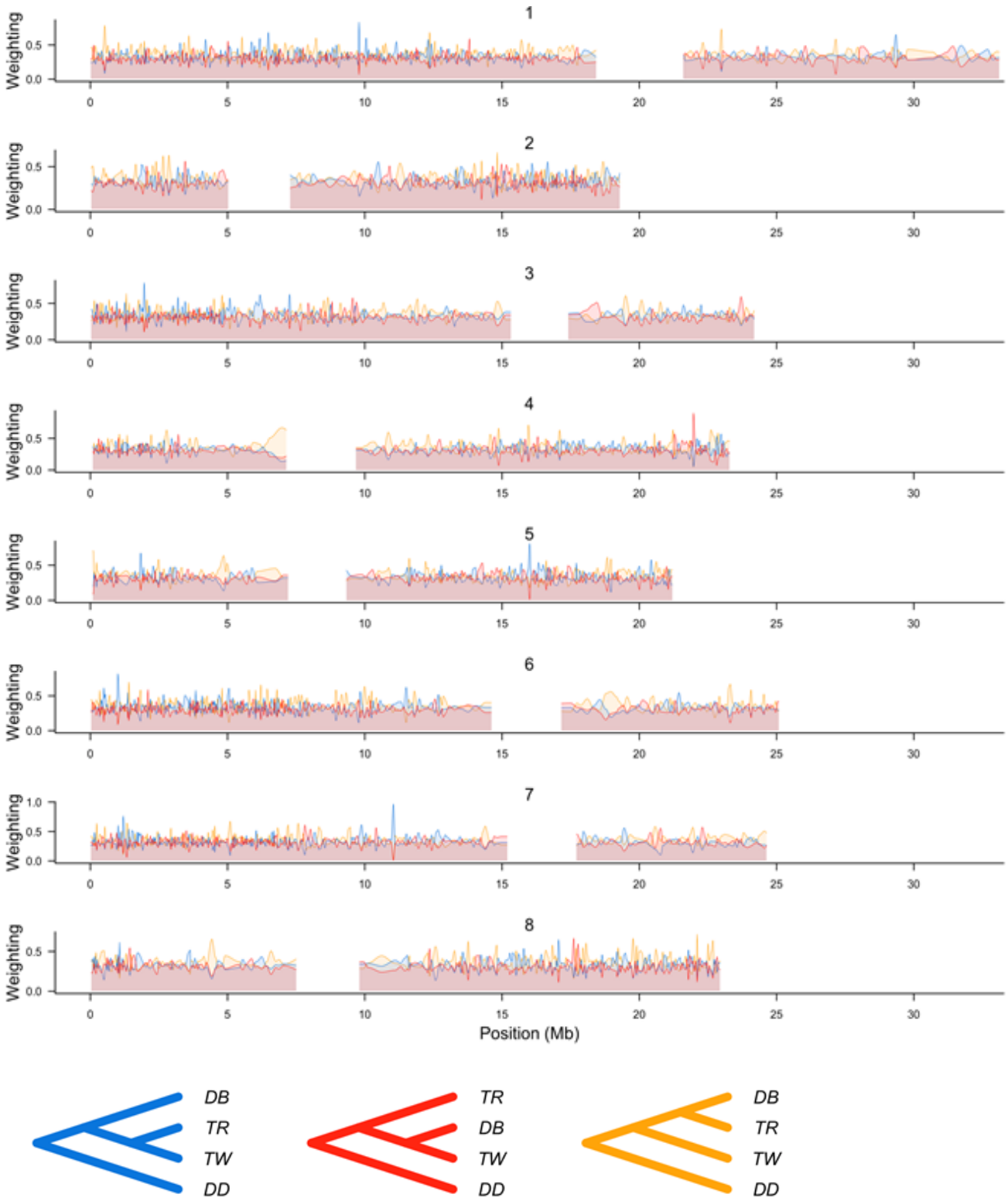


Figure S15. Topology weightings along the eight reference chromosomes (in 100 SNP windows) painted according to colour code beneath chromosome plots. Weights of populations lumped in lineages show pattern of admixture in Southern Carpathian (upper) and Baltic-Ruderal (lower) contact zones. DS = *S. Carp.2x*, TS = *S. Carp.-4x*, TW = *W. Carp.-4x*, TR = *Ruderal-4x*, DB = *Baltic-2x*, DD = *Dinaric-2x* lineages.

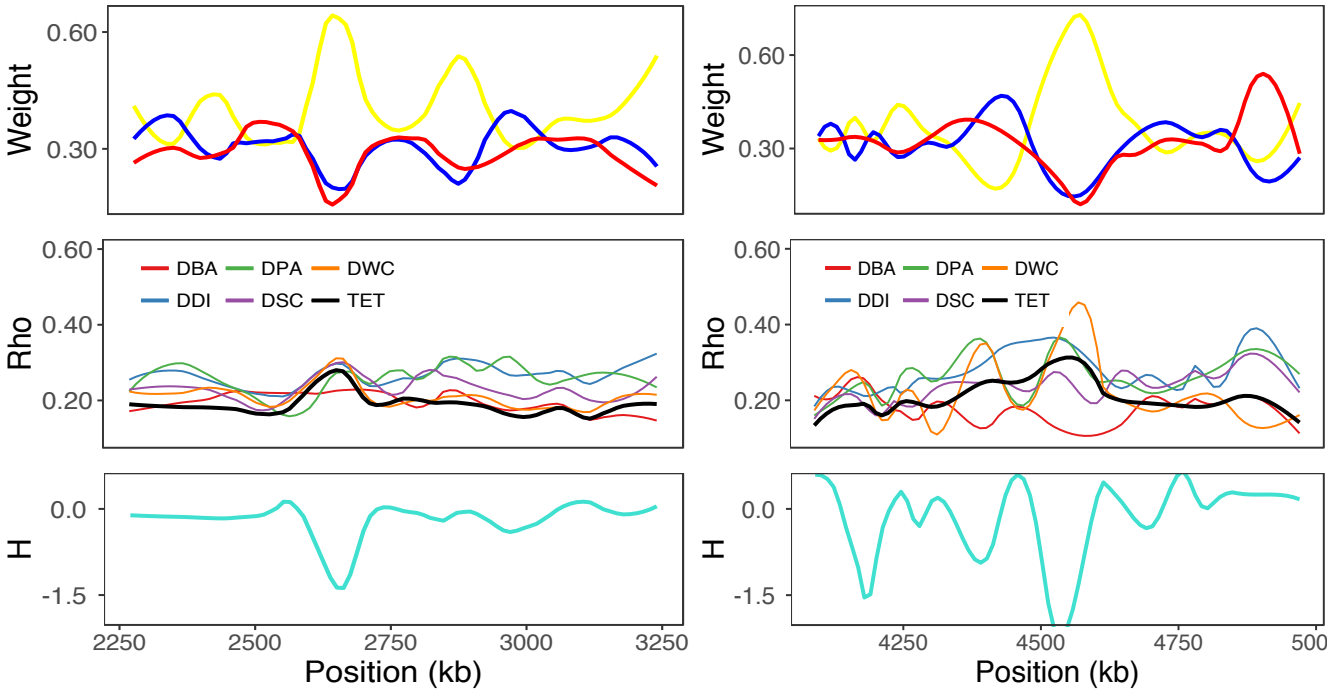


Figure S16. Examples of potential adaptive introgression for the Baltic-Ruderal contact zone. The top two plots are on chromosome 1, whereas the bottom two plots are from chromosome 2 and 8, respectively. Top panels give topology weightings: Yellow = topo1, red = topo2, and blue = topo3 from Figure 5. DBA = *Baltic-2x*, DPA = *Pannonian-2x*, DWC = *W. Carp.-2x*, DDI = *Dinaric-2x,* DSC = *S. Carp.2x* lineage, TET = all other tetraploid lineages. Right panel is the same outlier identified in Table S10.

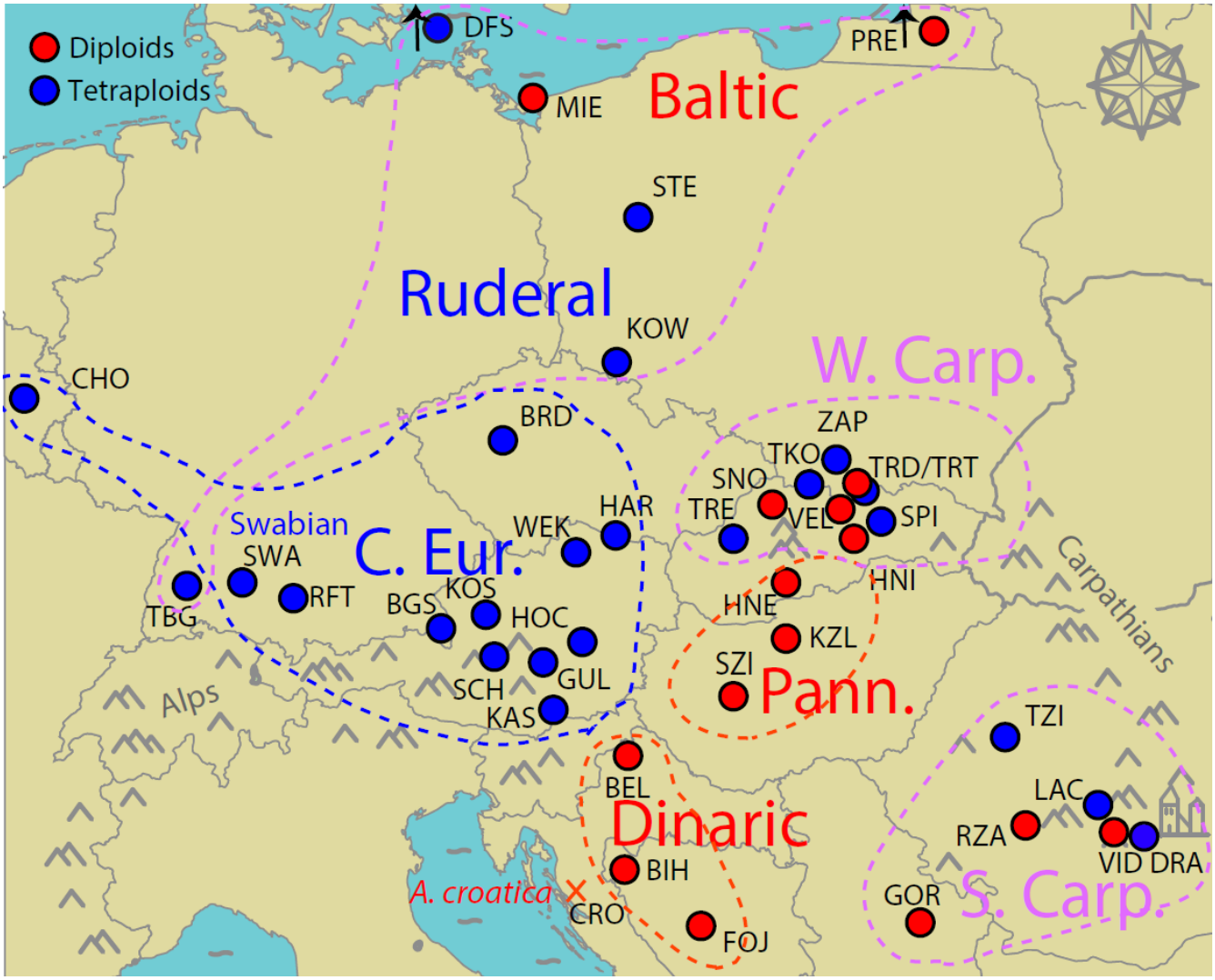


Figure S17 Location of the 15 diploid (red) and 24 tetraploid (blue) *A. arenosa* and one diploid *A. croatica* (cross) populations investigated. The geographical groups (corresponding with Bayesian clustering) are delimited by dashed lines with color corresponding to ploidy (violet – mixed-ploidy grouping).

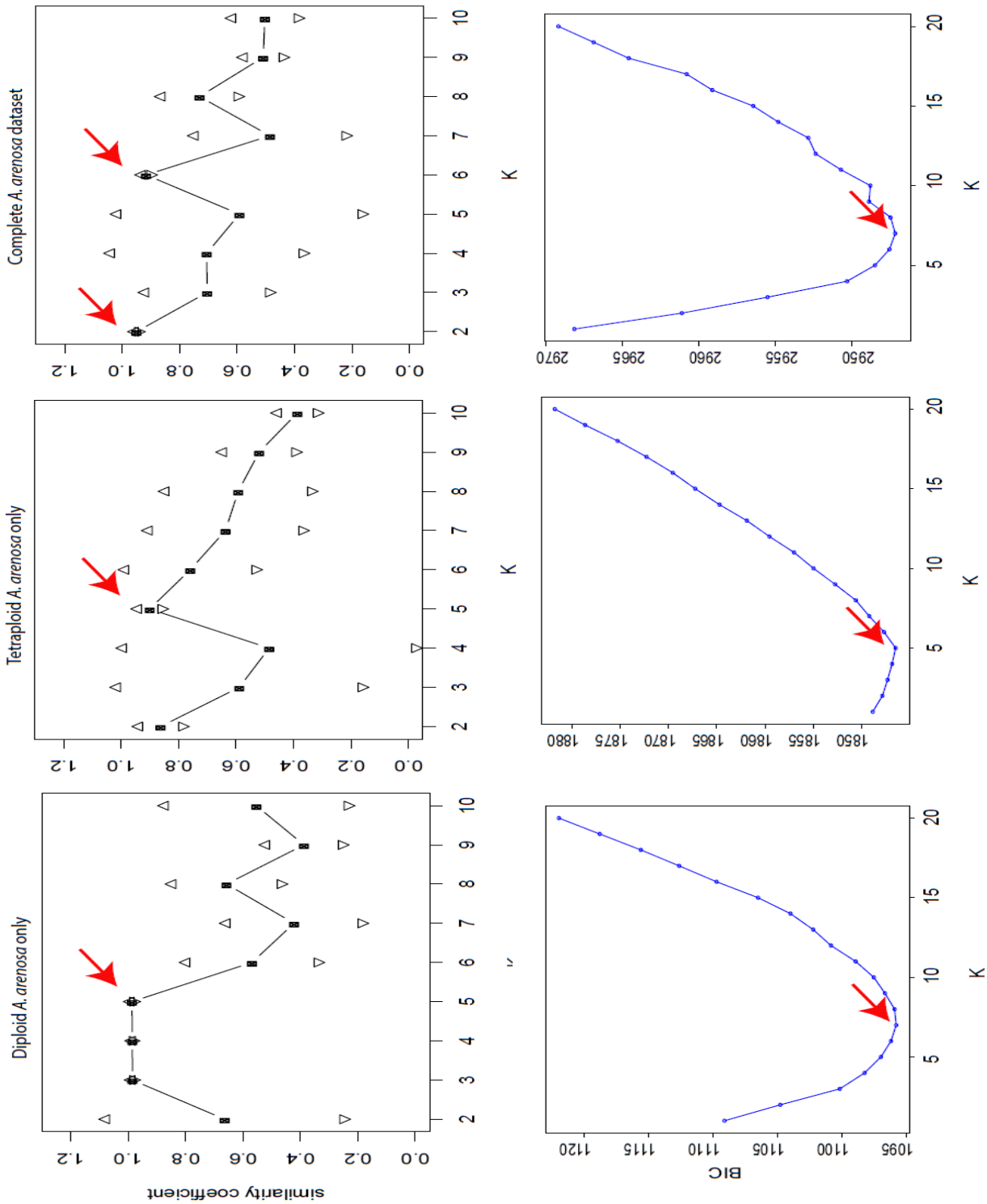


Figure S18 Selection of plausible number of clusters (K) in *fastStructure* analyses (upper) and k-means clustering (lower) in analyses of fourfold degenerate SNPs from 39 wild *A. arenosa* populations. Upper row: a similarity coefficient calculated among clustering results of five replicated runs for each K using fastStructure; lower row: Bayesian information criterion (BIC) of k-means groupings summarized over 1000 random starts.

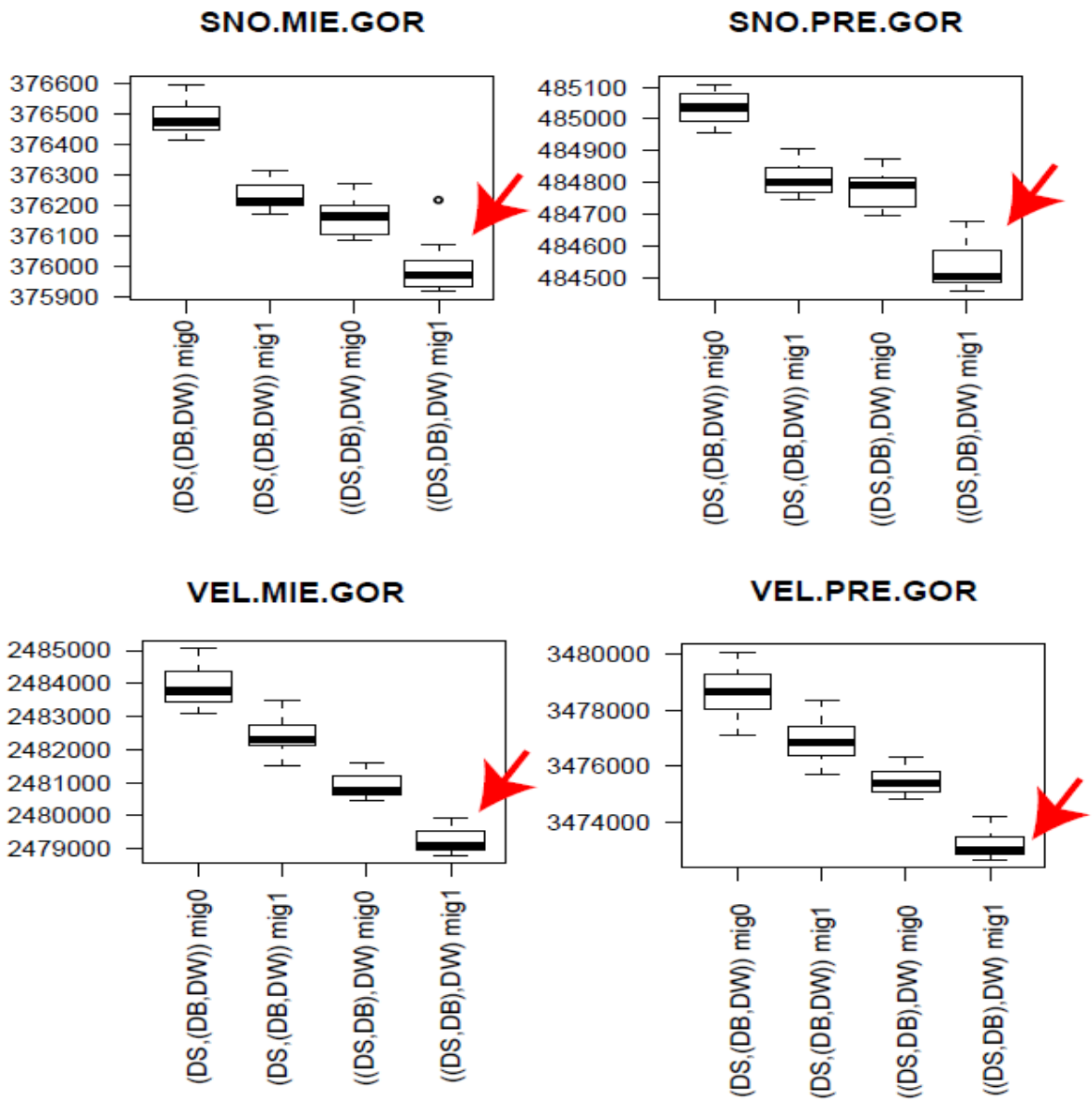


Figure S19 Comparison of Akaike information criteria (AIC) across four scenarios approximating the origin of the *Baltic-2x* (DB) from the *S. Carp.-2x* (DS) and *W. Carp.-2x* (DW) diploids, either assuming presence (*mig1*) or absence (*mig0*) of admixture from the non-sister lineage. Each scenario was simulated by 50 independent runs of fastsimcoal2 for various combination of population trios, the corresponding distribution of the AIC values over these 50 runs are shown by the boxplots. Red arrow highlights the most likely scenario.

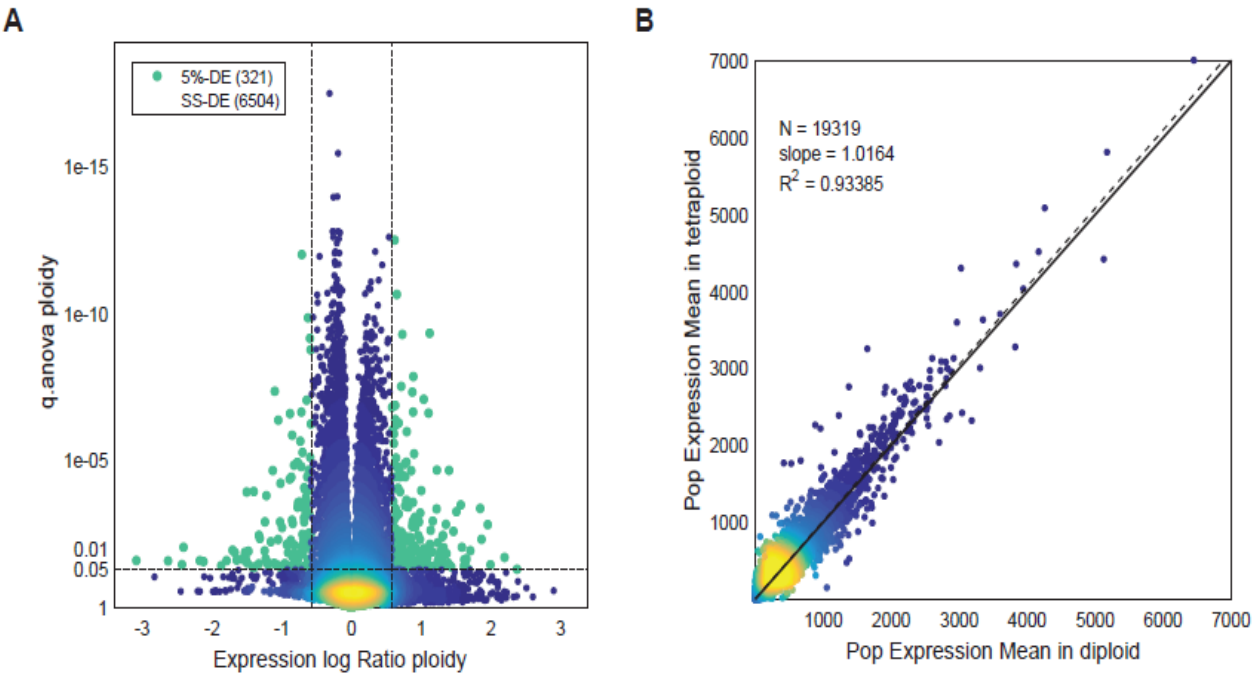


Figure S20 Weak differential expression across ploidies. (A) Volcano plot of q-values for cross-ploidy differential gene expression (ANOVA) against expression log ratios. Dotted lines represent 5% significance thresholds and based on this we identified 321 genes as differentially expressed (green markers) leaving 18,998 genes as non-differentially expressed between ploidies (NDE). Based on q-values only 6,183 additional genes show significant differential expression but with low absolute expression differences (ratio < 3.8x). (B) Mean population expression in tetraploids against their respective mean population expression in diploids. The y = x line (plain line) is represent along the linear regression (dotted line).

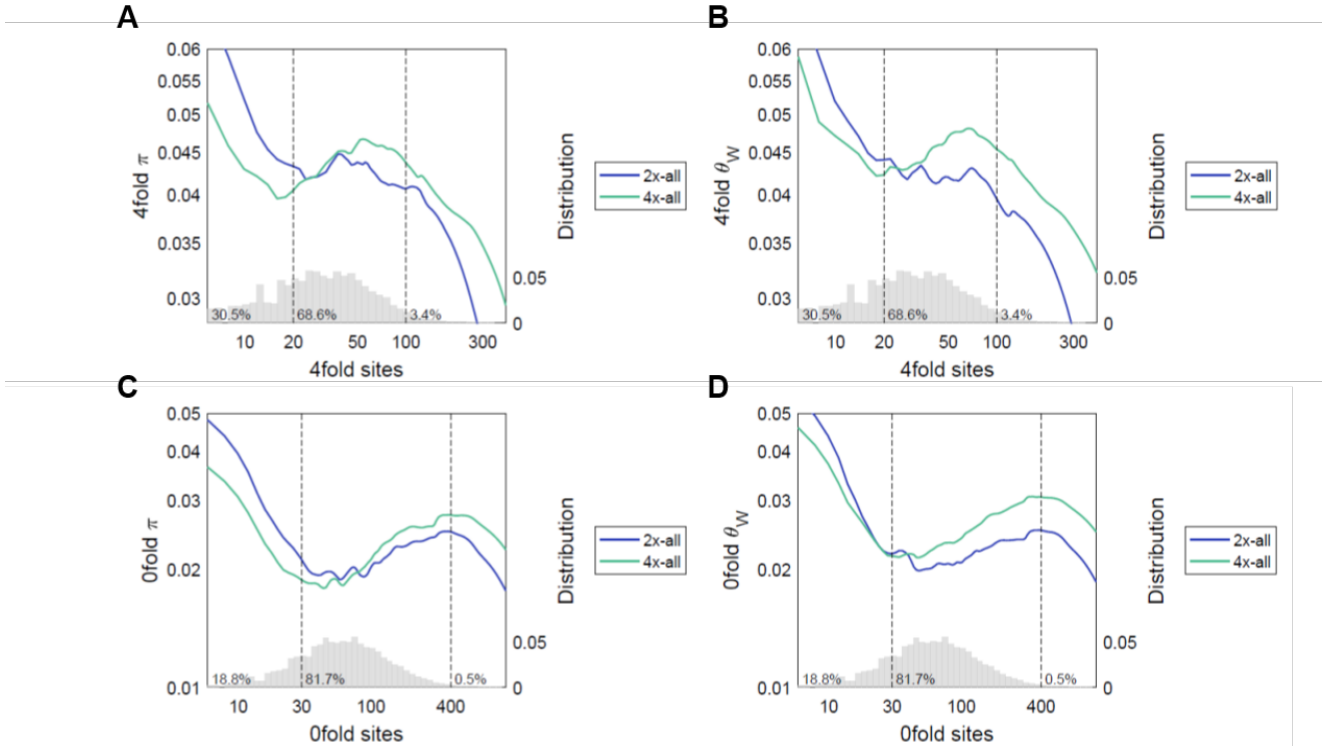
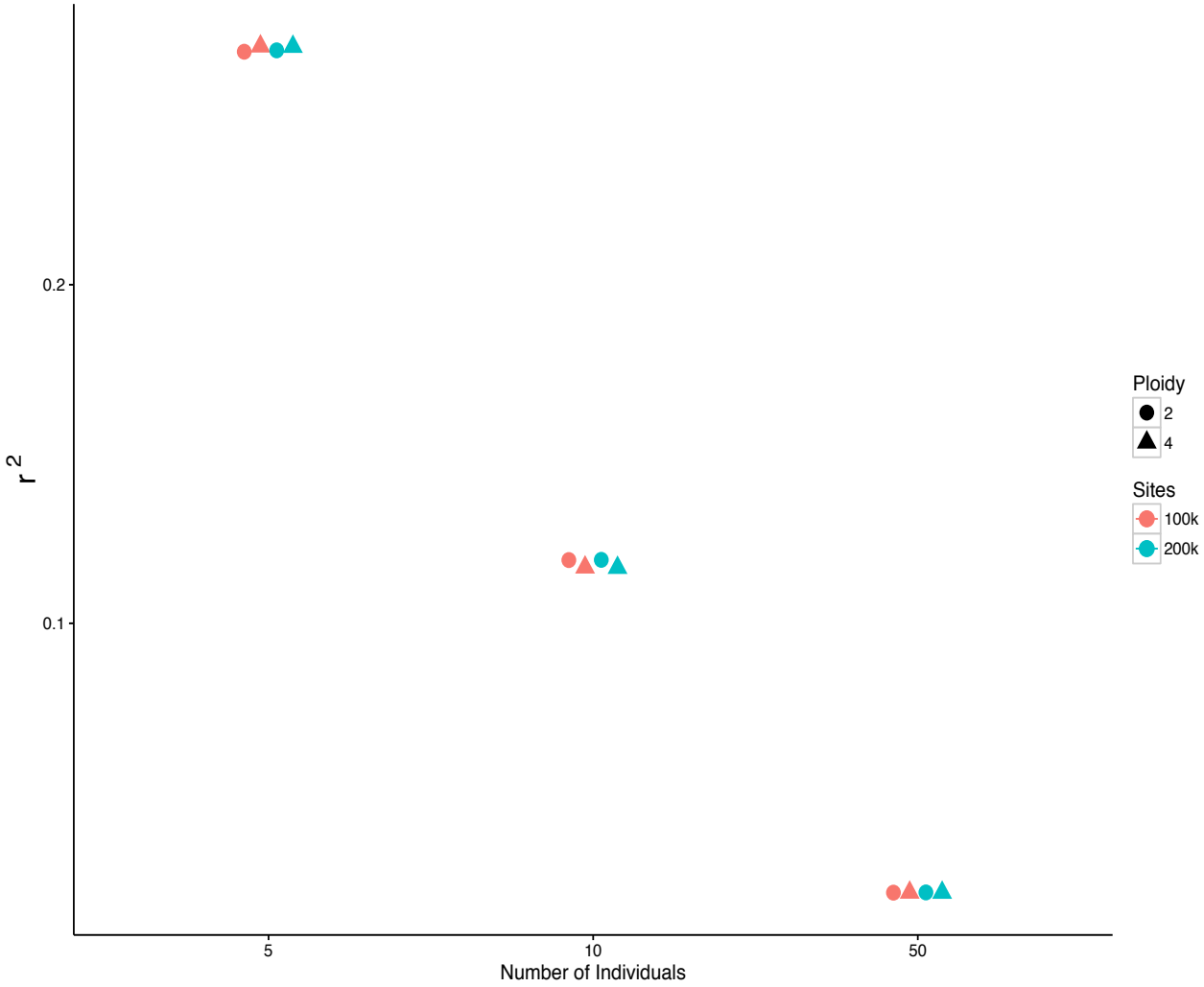
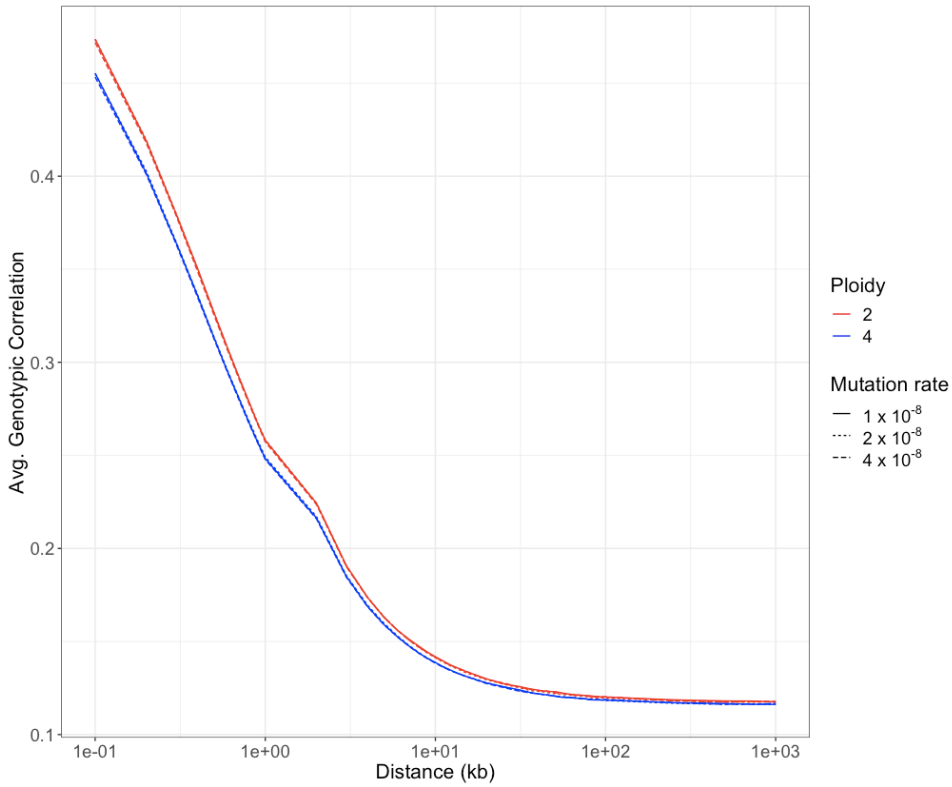
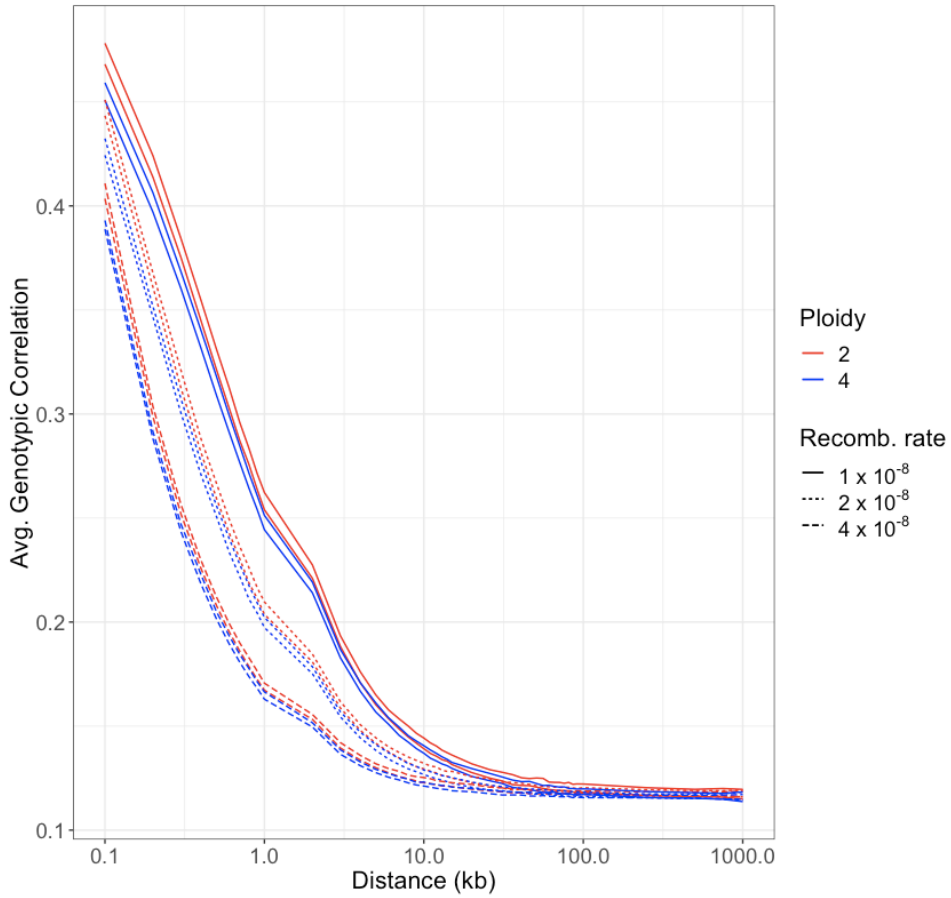


Figure S21 Impact of site number on diversity metrics. Locally weighted linear regressions (LOWESS) of 4-dg and 0-dg gene-wise estimates of $\theta\pi$ and θ_w against the number of corresponding sites per locus. The distribution of loci across the range of x-axis is represented in grey with the percentages of the 3 regimes.


r^2
0.2
0.1
5
10
50
Number of Individuals
Ploidy
2
4
Sites
100k
200k


Avg. Genotypic Correlation
0.4
0.3
0.2
0.1
1e-01
1e+00
1e+01
1e+02
1e+03
Distance (kb)
Ploidy
2
4
Mutation rate
1 x 10-8
2 x 10-8
4 x 10-8


Avg. Genotypic Correlation
0.4
0.3
0.2
0.1
0.1
1.0
10.0
100.0
1000.0
Distance (kb)
Ploidy
2
4
Recomb. rate
1 x 10-8
2 x 10-8
4 x 10-8

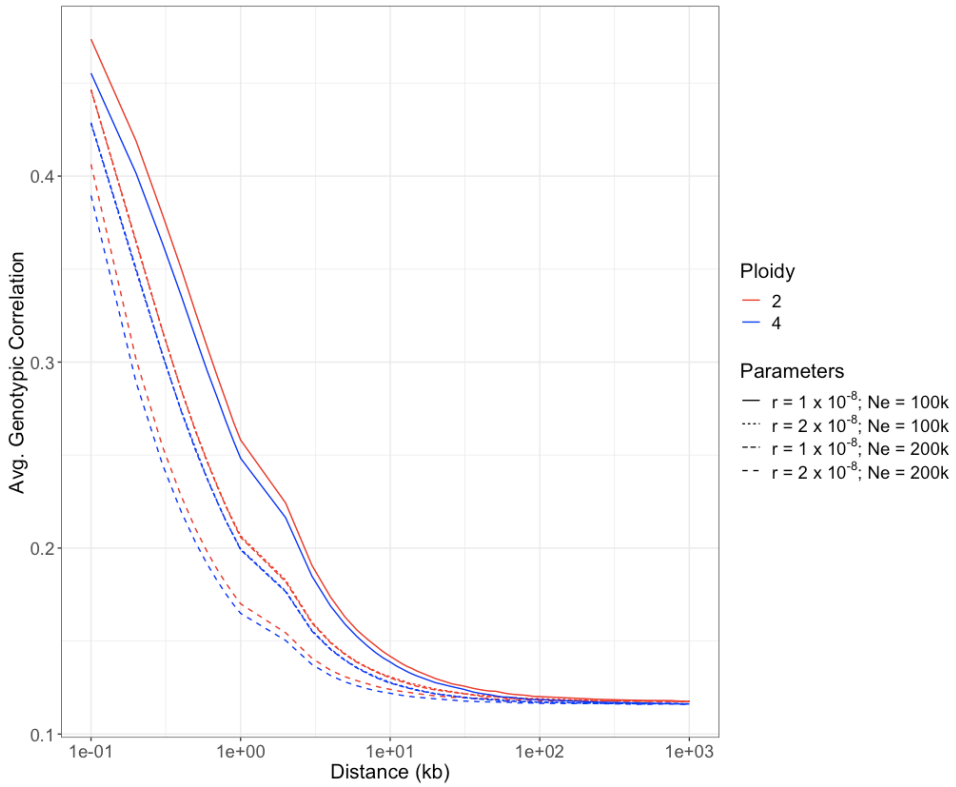


Figure S22 Effect of various factors on correlation approximation to LD. *Top*: Simulations of independent genotype data among loci demonstrate no inherent tendency for ploidy to affect the genotypic correlation. *Second*: Simulations of neutral linked data using msprime. Lines represent the mean value across simulations. Note that all lines overlap perfectly for different values of mutation rate. For all simulations, Ne = 100,000 and recombination rate = 1 x 10^{-8}. *Third*: Simulations as in *Second*, but varying recombination rate and keeping mutation rate constant (1 x 10^{-8}). The paired lines (by color and linetype) represent the 25th and 75th percentile across 100 simulations for each set of parameters. *Last:* Effect of varying population size on LD. Note that the lines corresponding to the second and third parameter sets overlap perfectly.

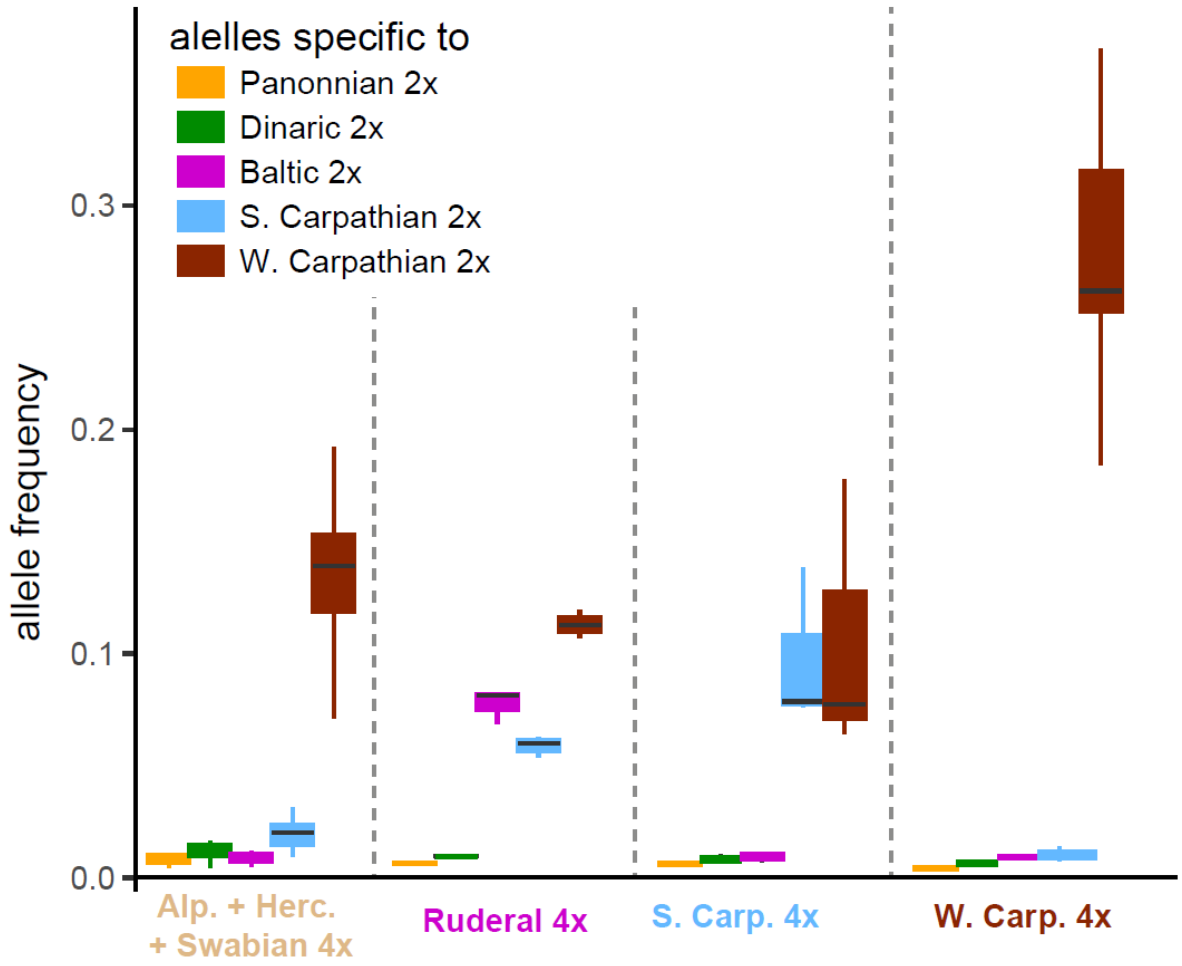


Figure S23 Genome-wide average frequencies of alleles diagnostic to particular diploid *A. arenosa* lineages present in each tetraploid lineage.

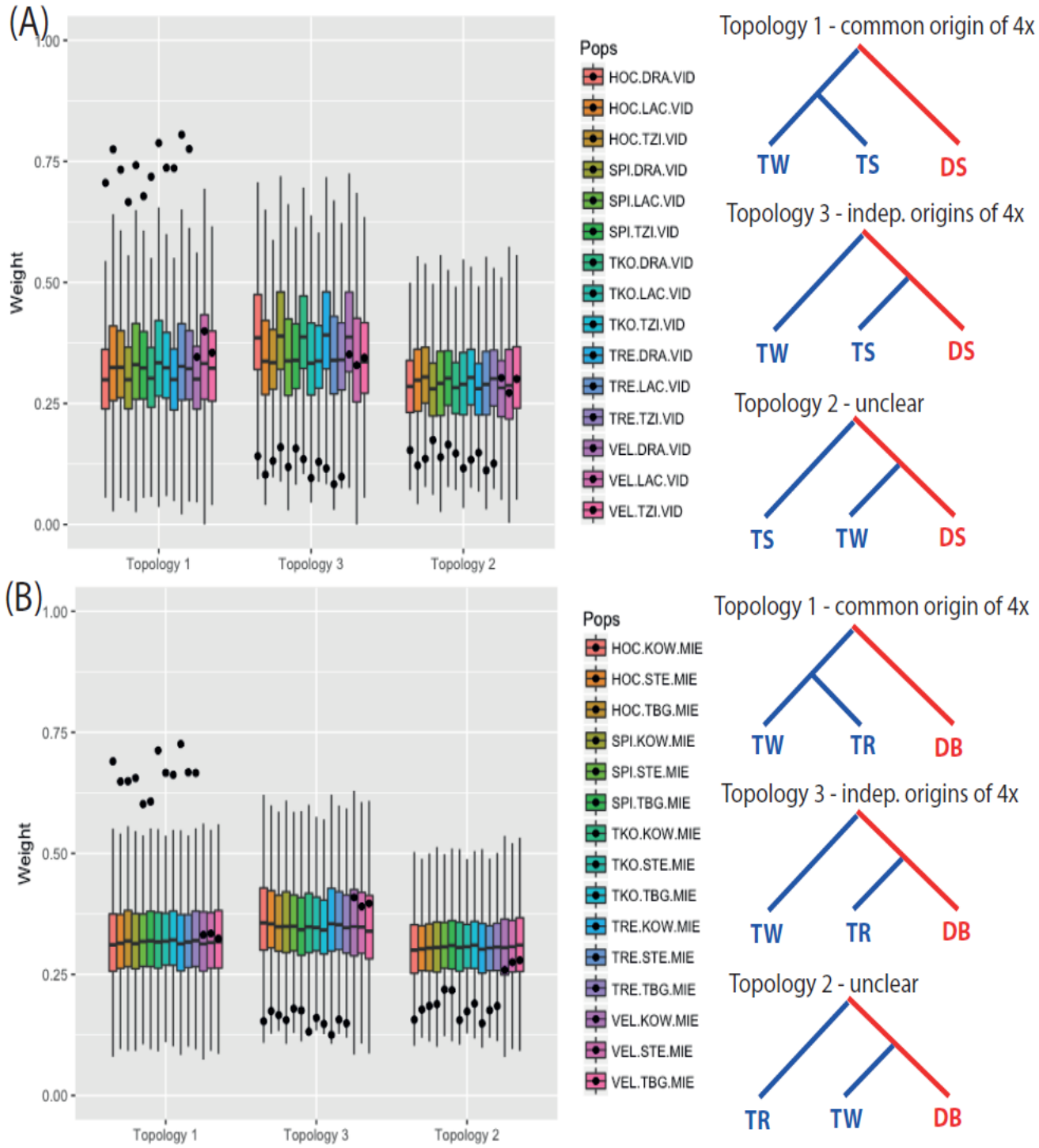


Figure S24. Distribution of weightings of three possible topologies over 100 SNP windows in population trios involving (A) *S. Carp.-4x* (TS) and *S. Carp.-2x* (DS) and (B) *Ruderal-4x* (TR) and *Baltic-2x* (DB). Note that the meiosis genes (dots) show highest support for a taxon topology that groups tetraploids as sister, despite the genome-wide distribution (boxplots) supported topology that grouped together the geographically proximate diploid/tetraploid pair. The three rightmost boxplots in each set involved diploid (VEL, *W. Carp.-2x* lineage) population instead its tetraploid (*W. Carp.-4x*) counterpart from the same area.

Supplemental Tables

Table S1 Measures of within-population diversity in diploid and tetraploid *A. arenosa* calculated over all individuals from particular ploidy (before slash) and averaged across populations (after slash).

group	pairwise diversity (θ_π)			Watterson's θ (θ_W)			Tajima's D			π_{NS}/π_S	θ_{NS}/θ_S
sites	all	4-dg	NS (0-dg)	all	4-dg	NS (0-dg)	all	4-dg	NS (0-dg)	-	-
Diploids (14 pops)	0.014/ 0.016	0.026 / 0.022	0.0058 / 0.0054	0.022 / 0.015	0.034 / 0.022	0.011 / 0.005	-1.18 / 0.03	-0.82 / 0.16	-1.05 / -0.09	0.223 / 0.242	0.324 / 0.255
Tetraploids (22 pops)	0.010 / 0.015	0.019 / 0.023	0.0045 / 0.0055	0.024 / 0.016	0.035 / 0.023	0.0139/ 0.006	-1.83 / -0.23	-1.44 / 0.00	-2.12 / -0.41	0.237 / 0.237	0.397 / 0.263

Populations with < 5 individuals were excluded. The statistics calculated (i) over all individuals from particular ploidy (downsampled to 50 and 80 individuals at each site for diploids and tetraploids, respectively) and (ii) averaged across populations (sites randomly downsampled to 5 individuals) are shown before and after the slash, respectively.

Table S2 Measures of within-population diversity in a subset of *A. arenosa* populations: (i) all diploid and a subset of tetraploid populations excluding those highly admixed by divergent diploid lineages (i.e. *S. Carp.-4x* and *Ruderal-4x*) (upper pane) and (ii) diploid and tetraploid populations from area of sympatry of genetically and ecologically very close populations in Western Carpathians.

	Diversity[1]										
	pairwise diversity (θ_π)			Watterson's θ (θ_W)			Tajima's D			π_{NS}/π_S	θ_{NS}/θ_S
Sites	all	4-dg	NS (0-dg)	all	4-dg	NS (0-dg)	all	4-dg	NS (0-dg)	-	-
All diploids (14 pops)	0.016 (0.003)	0.022 (0.003)	0.0054 (0.0007)	0.015 (0.004)	0.022 (0.003)	0.005 (0.0009)	0.03 (0.21)	0.16 (0.18)	-0.09 (0.23)	0.242 (0.017)	0.255 (0.017)
Tetraploids outside contact zones (16 pops)	0.015 (0.004)	0.023 (0.006)	0.0054 (0.0014)	0.016 (0.004)	0.023 (0.006)	0.006 (0.0013)	-0.27 (0.27)	-0.03 (0.26)	-0.45 (0.25)	0.238 (0.008)	0.264 (0.007)
Difference[2]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	**	.	***	n.s.	***
W. Carp. diploids (3 pops)	0.018	0.025	0.0062	0.019	0.025	0.007	-0.23	-0.06	-0.35	0.255	0.269[3]
W. Carp. tetraploids (5 pops)	0.017	0.025	0.0057	0.018	0.025	0.007	-0.32	-0.09	-0.51	0.233	0.260

Populations with < 5 individuals were excluded; for populations with > 5 individuals, sites were randomly downsampled to 5 to facilitate comparison across populations.

[1] values averaged across populations within the ploidy, standard deviation is in brackets in the upper comparison.

[2] Wilcoxon rank sum test; n.s. – non significant, $p \leq 0.07$ * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. Populations from W Carpathians were not tested due to low number of observations

[3]The value was dragged upwards by exceptionally high value of the SNO population; when SNO was excluded the average of diploid populations dropped to 0.249

Table S3 Measures of plastome and nuclear synonymous diversity and divergence within and among populations, ploidy levels, and major lineages of *A. arenosa* across 4-fold degenerate sites.

group	N pops	% among pops.[1]	average Rho / Fst[2]	N of plastome haplogroups	Nucleotide diversity (θ_π) (4-dg)[3]	Watterson's θ (4-dg)[3]	Tajima's D (4-dg)[3]
Diploids	14	70.6	0.30 / 0.29	10	0.0261 / 0.0224 ± 0.003	0.0299 / 0.0218 ± 0.0032	-0.527 / 0.1567 ± 0.1808
Tetraploids	22	48.2	0.20 / 0.11	15	0.0187 / 0.0232 ± 0.0056	0.0253 / 0.0231 ± 0.0052	-0.960 / 0.0005 ± 0.2685
S. Carp 2x + W. Carp. 2x + Baltic 2x	8	63.2	0.27 / 0.23	9	0.0270 / 0.0239 ± 0.0018	0.0296 / 0.0234 ± 0.0021	-0.366 / 0.1142 ± 0.1897
Baltic 2x	2	38.0	0.22 / 0.17	2	0.0268 / 0.0246 ± 0.0013	0.0244 / 0.0231 ± 0.0012	0.416 / 0.3331 ± 0.0066
Dinaric 2x	3	46.5	0.22 / 0.18	4	0.0227 / 0.0214 ± 0.0013	0.0225 / 0.0204 ± 0.0007	0.044 / 0.2263 ± 0.1292
Pannonian 2x	3	47.5	0.23 / 0.20	2	0.0251 / 0.0195 ± 0.0047	0.0248 / 0.0189 ± 0.0051	0.055 / 0.2002 ± 0.2294
S. Carp 2x	3	58.5	0.29 / 0.26	2	0.0255 / 0.0228 ± 0.002	0.0261 / 0.0222 ± 0.002	-0.096 / 0.1447 ± 0.1478
W. Carp 2x	3	33.8	0.14 / 0.09	7	0.0270 / 0.0246 ± 0.0018	0.0284 / 0.025 ± 0.002	-0.204 / -0.0621 ± 0.0898
C. Europe 4x	11	43.2	0.19 / 0.10	9	0.0178 / 0.0219 ± 0.0069	0.0231 / 0.0218 ± 0.0063	-0.856 / -0.0098 ± 0.2696
Ruderal 4x	3	22.2	0.11 / 0.05	3	0.0282 / 0.0273 ± 0.0021	0.0285 / 0.0267 ± 0.0028	-0.035 / 0.1024 ± 0.1934
S. Carp. 4x	3	43.0	0.23 / 0.14	4	0.0191 / 0.0210 ± 0.0019	0.0220 / 0.0206 ± 0.0001	-0.481 / 0.0818 ± 0.4021
W. Carp. 4x	5	27.9	0.09 / 0.04	5	0.0262 / 0.0246 ± 0.0044	0.0280 / 0.025 ± 0.0033	-0.240 / -0.0869 ± 0.2735

Populations with < 5 individuals were excluded, and for populations with > 5 individuals the data were downsampled to 5 to facilitate comparison across populations.

[1]% of explained variance among populations (compared to variance within populations) in Analysis of Molecular Variance (AMOVA)

[2]values averaged over pairwise comparisons of populations belonging to that group

[3]The statistics calculated (i) over all individuals from particular lineage and (ii) averaged ± SD across populations within the lineage are before and after slash, respectively.

Table S4 Maximum likelihood estimates of model parameters of the preferred scenarios addressing the admixed origin of the *Baltic-2x* (DB) from the *S. Carp.-2x* (DS) and *W. Carp.-2x* (DW) populations. The numbers refer to median and 95% confidence intervals, inferred from 100 bootstrapped input AFS of 4-fold degenerated sites. The analyses are based on trios of the best covered populations.

Pop trio	N sites	Ne DW	Ne DB	Ne DS	DIv DS-DB	Div DS-DW	Adm DW->DB
VEL(MIE,GOR)	1062773	(176583) **187176** (199806)	(4646) **19760** (28895)	(5538) **23563** (33593)	(1563) **6928** (10217)	(43744) **46941** (49880)	(0.28) **0.35** (0.4)
VEL(PRE,GOR)	1424748	(169122) **179390** (191240)	(5577) **20181** (29316)	(6607) **23510** (32910)	(2053) **7286** (10544)	(41419) **45139** (48824)	(0.29) **0.37** (0.43)

Shown are the population sizes of the three lineages (Ne DW, Ne DB, Ne DS), the divergence times between the *Baltic-2x* and *S. Carp.-2x* (Div DS-DB) and between the two Carpathian diploid lineages (Div DS-DW) and the admixture proportions (Adm) from *W. Carp.-2x* (DW) to the *Baltic-2x* (DB) going forward in time. Divergence times are expressed in terms of generations, population sizes are in haploid numbers of chromosomes.

Table S5. Standardized estimates of predictor effects, standard errors, and *F*-test p-values in multiple linear regression model of 0-dg θ_W and of 0-fold/4-fold θ_W log ratio against log-expression and ploidy.

		All populations			**No-Admixed**	**W-Carp-only**
0-dg θ_W	**Estimates**	**SE**	***p*-value**		***p*-value**	***p*-value**
intercept	0.018	1.1e-4	< 1E-100	***	< 1E-100	< 1E-100
α_{tet} (ploidy effect)	1.4e-3	3.6e-5	< 1E-100	***	< 1E-100	< 1E-9
β (expression effect)	-9.0e-4	1.1e-4	< 1E-14	***	< 1E-17	< 1E-5
γ_{tet} (expression by ploidy)	-1.1e-4	3.6e-5	0.0013	**	0.31	0.70
		All populations			**No-Admixed**	**W-Carp-only**
0-dg/4-dg θ_W log ratio	**Estimates**	**SE**	***p*-value**		***p*-value**	***p*-value**
intercept	-0.73	3.1e-3	< 1E-100	***	< 1E-100	< 1E-100
α_{tet} (ploidy effect)	0.049	4.3e-3	< 1E-29	***	< 1E-38	< 1E-9
β (expression effect)	-0.034	3.1e-3	< 1E-27	***	< 1E-26	< 1E-6
γ_{tet} (expression by ploidy)	-5.9e-3	4.3e-3	0.17		0.82	0.68

Table S6. Standardized estimates of predictor effects, standard errors, and *F*-test p-values in multiple linear regression model of 0-dg θ_W and of 0-fold/4-fold θ_W log ratio against log-expression, ploidy, and N_g.

		All populations			**No-Admixed**	**W-Carp-only**
0-dg θ_W	**Estimates**	**SE**	***p*-value**		***p*-value**	***p*-value**
intercept	0.021	5.3E-05	< 1E-100	***	< 1E-100	< 1E-100
β (expression effect)	-1.2e-03	3.4e-05	< 1E-100	***	< 1E-100	< 1E-60
α_{tet} (ploidy effect)	1.2e-03	7.3e-05	< 1E-57	***	**< 1E-69**	< 1E-17
δ (N_g effect)	2.4e-03	5.7e-05	< 1E-100	***	< 1E-100	< 1E-87
γ_N (expression by N_g)	-2.2e-04	7.4e-05	< 1E-09	***	< 1E-05	0.16
ε_p (ploidy by N_g)	6.2e-04	3.5e-05	< 1E-17	***	< 1E-29	< 1E-07
		All populations			**No-Admixed**	**W-Carp-only**
0-dg/4-dg θ_W log ratio	**Estimates**	**SE**	***p*-value**		***p*-value**	***p*-value**
intercept	-0.73	3.2e-3	< 1E-100	***	< 1E-100	< 1E-100
β (expression effect)	-0.034	3.1e-3	< 1E-27	***	< 1E-26	< 1E-06
α_{tet} (ploidy effect)	0.048	4.5e-3	< 1E-26	***	**< 1E-36**	< 1E-08
δ (N_g effect)	6.8e-4	2.2e-3	0.76		0.0089	0.062
γ_p (expression by ploidy)	-5.9e-3	4.3e-3	0.17		0.82	0.69

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table S7 Differences in fitness effect of purifying selection (categorized in three bins of increasing strength of purifying selection) and proportion of adaptive substitutions (α and ω_α) between diploids and tetraploids assessed using the DFE based on folded or unfolded site frequency spectra and tested using a Wilcoxon rank-sum test. Significant values after Bonferroni correction (p = 0.003) are in bold. The differences were tested for (i) all diploid and tetraploid populations and (ii) all diploid and all but admixed tetraploid populations from contact zones (i.e., excluding populations from *S. Carp.-4x* and *Ruderal-4x* lineages).

	All populations				**Excluding admixed 4x**			
	Folded		Unfolded		Folded		Unfolded	
	W	*P*-value	W	*P*-value	W	*P*-value	W	*P*-value
Nearly neutral (-$N_e s$ 0-1)	171	0.353	197	0.067	144	0.083	155	0.0251
Mildly deleterious (-$N_e s$ 1-10)	106	0.216	93	0.091	66	0.101	61	0.062
Highly deleterious (-$N_e s$ > 10)	161	0.555	163	0.511	125	0.374	121	0.475
α	14	**<0.0001**	40	**0.0002**	0	**<0.0001**	20	**<0.0001**
ω_α	6	**<0.0001**	18	**<0.0001**	0	**<0.0001**	6	**<0.0001**

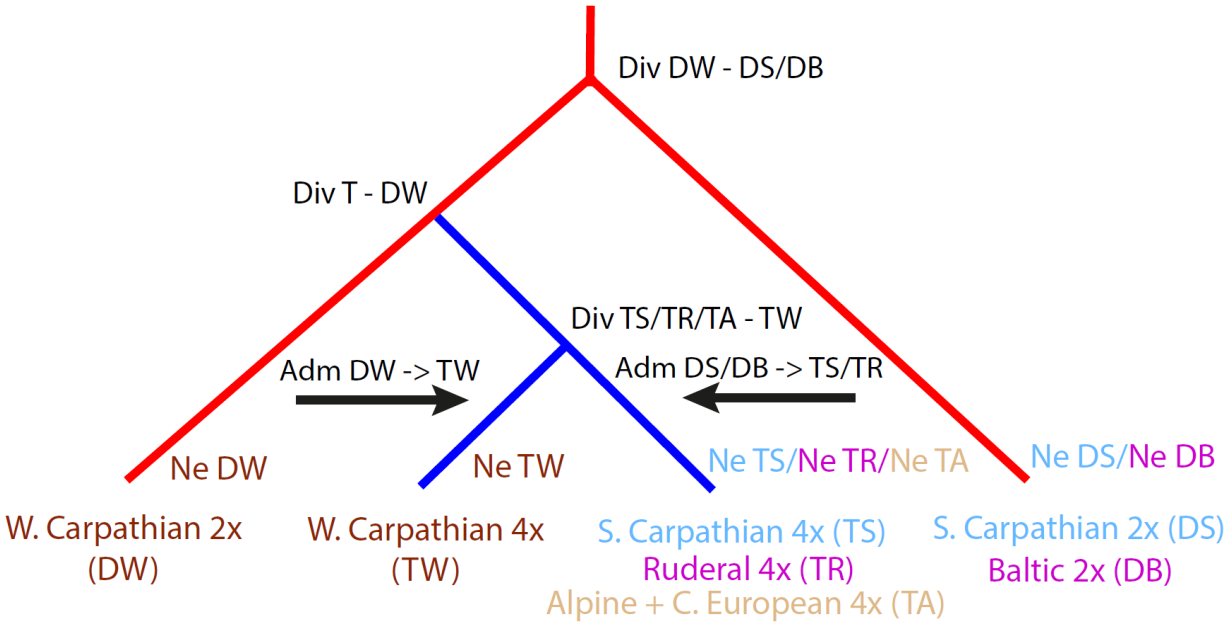
Table S8. Coefficient estimates and significance of terms from linked selection analysis. Top: Analysis performed after removing windows with fewer than 20 SNPs and populations with fewer than 2,000 non-missing windows. Middle: Analysis re-performed following additional removal of highly admixed tetraploid populations (KOW, TBG, STE, LAC, DRA, and TZI). Bottom: Analysis re-performed following inclusion of only sympatric populations in the Western Carpathians that pass filter (TKO, TRT, SPI, TRE, VEL, HMI, and TRD).

	sqrt[4dg-Diversity]	
Predictors	*Estimates*	*p*
E-NS	-0.01	**<0.001**
GDM	-0.10	**<0.001**
Ploidy	-0.00	0.667
E-NS^2	-0.00	**<0.001**
E-NS * GDM	-0.01	**<0.001**
E-NS * Ploidy	0.00	0.797
GDM * Ploidy	0.03	**<0.001**
E-NS * E-NS^2	-0.00	**<0.001**
GDM * E-NS^2	-0.00	0.239
Ploidy * E-NS^2	-0.00	**0.010**
E-NS * Ploidy * GDM	-0.00	**0.002**
Observations	71827	
Marginal R^2 / Conditional R^2	0.296 / 0.372	

	sqrt[4dg-Diversity]	
Predictors	*Estimates*	*p*
E-NS	-0.01	**<0.001**
GDM	-0.10	**<0.001**
Ploidy	-0.00	0.906
E-NS^2	-0.00	**<0.001**
E-NS * GDM	-0.01	**<0.001**
E-NS * Ploidy	0.00	0.967
GDM * Ploidy	0.02	**<0.001**
E-NS * E-NS^2	-0.00	**<0.001**
GDM * E-NS^2	-0.00	0.433
Ploidy * E-NS^2	-0.00	**0.038**
E-NS * Ploidy * GDM	-0.00	**0.005**
Observations	58426	
Marginal R^2 / Conditional R^2	0.300 / 0.367	

	sqrt[4dg-Diversity]	
Predictors	*Estimates*	*p*
E-NS	-0.01	**<0.001**
GDM	-0.10	**<0.001**
Ploidy	-0.00	0.547
E-NS^2	-0.00	**<0.001**
E-NS * GDM	-0.01	**0.008**
E-NS * Ploidy	0.00	0.127
GDM * Ploidy	0.02	**<0.001**
E-NS * E-NS^2	-0.00	**0.016**
GDM * E-NS^2	-0.00	0.966
Ploidy * E-NS^2	0.00	0.825
E-NS * Ploidy * GDM	-0.01	0.062
Observations	16153	
Marginal R^2 / Conditional R^2	0.293 / 0.416	

Table S9. Maximum likelihood estimates of model parameters of the preferred scenarios addressing the admixed origin of the *S. Carp.-4x* (TS) and *Ruderal-4x* (TR) tetraploids as well as the origin of the Alpine and western central European populations (*C. Europe-4x*, TA). The numbers refer to median and 95% confidence intervals, inferred from 100 bootstrapped input AFS of 4-fold degenerated sites. The analyses are based on quartets of the best covered populations from the target tetraploid (TS or TR) and *W. Carp.-4x* (TW) lineages, presumably ancestral *W. Carp.-2x* diploid (DW) and the admixture donor diploid population (*S. Carp-2x* for *S. Carp.-4x* and *Baltic-2x* for *Ruderal-4x*). The scheme explains individual parameters**,** red lines = diploid, blue = tetraploid populations, black arrow = migration forward in time (admixture).



S Carp-4x (TS)	N sites	Ne TS	Ne TW	Ne DW	Ne DS	Div TS-TW	Div T-DW	Div DW-DS	Adm DW -> TW	Adm DS -> TS
((VEL,(SPI,**DRA**)),VID)	167372	(59632) **98340** (148650)	(53035) **74876** (99664)	(101896) **141719** (191576)	(106696) **133826** (167355)	(10235) **18680** (27297)	(22200) **30739** (41177)	(36345) **41547** (50238)	(2.14) **3.06** (4.11)	(1.68) **2.37** (3.09)
((VEL,(SPI,**LAC**)),VID)	460462	(37912) **63933** (91208)	(52937) **74750** (99766)	(101353) **144113** (199854)	(89554) **121229** (162769)	(11484) **20574** (28603)	(22170) **30986** (40447)	(33053) **40519** (50718)	(2.27) **3.04** (4.04)	(0.05) **0.22** (0.41)
((VEL,(SPI,**TZI**)),VID)	91125	(65448) **115165** (193808)	(44866) **73328** (106894)	(90614) **134448** (183183)	(94247) **127726** (167063)	(8186) **15702** (24623)	(16723) **26538** (36349)	(30408) **39701** (54214)	(1.99) **2.82** (3.63)	(0.07) **0.51** (1.05)
((VEL,(TRE,**DRA**)),VID)	213380	(56525) **95515** (130214)	(53933) **92397** (127198)	(88121) **133886** (174300)	(103206) **134450** (174380)	(10567) **17780** (24621)	(18124) **28664** (37036)	(33660) **43288** (56221)	(0.97) **1.83** (2.53)	(1.73) **2.35** (3.26)
((VEL,(TRE,**LAC**)),VID)	674243	(39707) **61577** (90500)	(64252) **94447** (125505)	(100597) **132481** (168993)	(92707) **117593** (158583)	(13630) **21378** (29222)	(21694) **30847** (38964)	(33143) **40481** (50902)	(1.32) **1.86** (2.52)	(0.07) **0.22** (0.48)
((VEL,(TRE,**TZI**)),VID)	163302	(50708) **112544** (160094)	(46017) **91609** (130814)	(77064) **121009** (165891)	(95264) **125379** (174182)	(7173) **16831** (23130)	(17822) **26030** (35498)	(31603) **40465** (56868)	(0.72) **1.74** (2.7)	(0.15) **0.51** (0.99)

Ruderal-4x (TR)	N sites	Ne TR	Ne TW	Ne DW	Ne DB	Div TR-TW	Div T-DW	Div DW-DB	Adm DW -> TW	Adm DS -> TS
((VEL,(SPI,**KOW**)),MIE)	298671	(47362) **79919** (118587)	(43047) **70802** (101522)	(80900) **115260** (158939)	(60924) **83740** (106193)	(8519) **14840** (20939)	(16074) **21764** (28612)	(20574) **28415** (33561)	(2.16) **2.84** (3.67)	(0.50) 1.20 (1.78)
((VEL,(SPI,**STE**)),MIE)	478333	(79972) **125751** (180850)	(51018) **70253** (102357)	(78881) **107246** (138666)	(54232) **74529** (97273)	(8002) **14369** (19701)	(14869) **19790** (26296)	(19188) **23948** (30597)	(1.96) **2.79** (3.55)	(1.35) **2.30** (3.31)

		(76931)	(46548)	(77680)	(57895)	(9359)	(15328)	(21077)	(1.97)	(0.89)
((VEL,(SPI,**TBG**)),MIE)	536155	**147771**	**73370**	**110049**	**79242**	**15877**	**21354**	**26635**	**2.69**	**1.59**
		(199488)	(96418)	(137027)	(98004)	(21123)	(26897)	(31152)	(3.53)	(2.44)
		(49452)	(61063)	(75732)	(60665)	(9656)	(15498)	(20673)	(0.68)	(0.66)
((VEL,(TRE,**KOW**)),MIE)	431843	**78790**	**92948**	**104952**	**77530**	**15431**	**21153**	**26126**	**1.56**	**1.11**
		(118053)	(130691)	(136054)	(97651)	(22882)	(28383)	(32135)	(2.28)	(1.70)
		(76425)	(57430)	(75966)	(56068)	(9279)	(15078)	(20257)	(0.69)	(1.67)
((VEL,(TRE,**STE**)),MIE)	870653	**134021**	**88919**	**98089**	**71661**	**16162**	**20768**	**24758**	**1.49**	**2.34**
		(177954)	(125523)	(122855)	(88267)	(20299)	(25891)	(29899)	(2.49)	(3.30)
		(83322)	(68192)	(71667)	(51840)	(9617)	(15046)	(19661)	(0.61)	(0.85)
((VEL,(TRE,**TBG**)),MIE)	880692	**133778**	**94597**	**102058**	**71517**	**16428**	**21123**	**25257**	**1.38**	**1.56**
		(188135)	(131737)	(123291)	(91444)	(21377)	(26399)	(30646)	(2.02)	(2.54)

C. Europe-4x (TA)									
	N sites	Ne TA	Ne TW	Ne DW	Ne DS	Div TA-TW	Div T-DW	Div DW-DS	Adm DW -> TW
		(100716)	(53236)	(95021)	(87043)	(10574)	(18833)	(32365)	(1.98)
((VEL,(SPI,**BRD**)),VID)	453550	**170979**	**80369**	**131300**	**113722**	**18238**	**27660**	**40123**	**2.70**
		(253100)	(99778)	(169587)	(170510)	(27105)	(35827)	(56100)	(3.56)
		(78419)	(47934)	(92910)	(82884)	(8588)	(17362)	(29180)	(1.63)
((VEL,(SPI,**HOC**)),VID)	355269	**156167**	**75346**	**128912**	**107552**	**15462**	**24627**	**36974**	**2.43**
		(216605)	(109488)	(174170)	(148352)	(23203)	(31954)	(49297)	(3.45)
		(58175)	(51294)	(94077)	(88179)	(10055)	(19875)	(32027)	(1.9)
((VEL,(SPI,**KAS**)),VID)	301462	**93147**	**75042**	**135827**	**115325**	**17679**	**26986**	**40472**	**2.70**
		(135325)	(103604)	(182575)	(153536)	(26184)	(35823)	(51889)	(3.68)
		(75150)	(50671)	(104935)	(90604)	(11232)	(21413)	(32717)	(1.88)
((VEL,(SPI,**SWA**)),VID)	303454	**116701**	**75672**	**140999**	**114683**	**19128**	**29070**	**40666**	**2.74**
		(179661)	(108529)	(188258)	(153448)	(29853)	(38509)	(52428)	(3.64)
		(97292)	(62168)	(84993)	(81162)	(10440)	(18069)	(29619)	(0.72)
((VEL,(TRE,**BRD**)),VID)	665570	**159525**	**94949**	**121256**	**111171**	**17376**	**24800**	**39433**	**1.46**
		(220565)	(137013)	(156623)	(156533)	(24683)	(34949)	(51335)	(2.24)
		(86500)	(65904)	(71805)	(82045)	(8302)	(14402)	(29266)	(0.17)
((VEL,(TRE,**HOC**)),VID)	673649	**159079**	**107780**	**119238**	**109741**	**15848**	**23843**	**39685**	**1.26**
		(221072)	(162444)	(155499)	(160544)	(23780)	(31172)	(55334)	(1.99)
		(55469)	(70473)	(83949)	(79452)	(10884)	(17000)	(29613)	(0.34)
((VEL,(TRE,**KAS**)),VID)	642689	**88875**	102049	**121768**	**111787**	**17012**	**24857**	**38909**	**1.52**
		(135489)	(158456)	(174467)	(158140)	(26254)	(35165)	(53403)	(2.61)
		(64984)	(67237)	(80211)	(86099)	(11341)	(17948)	(30976)	(0.92)
((VEL,(TRE,**SWA**)),VID)	433371	**107574**	**97484**	**119597**	**112237**	**17929**	**27235**	**39686**	**1.64**
		(162353)	(131506)	(164215)	(145548)	(27189)	(35851)	(50927)	(2.35)

Shown are the population sizes of the diploid (D) and tetraploid (T) populations, the divergence times between the tetraploids (DivTS-TW / Div TR-TW / Div TA-TW), between the ancestral tetraploid and diploid (Div T-DW), and between the two diploids (Div DW-DS/DW-DB) and the admixture proportions (Adm) from diploids (D) to tetraploids (T) going forward in time. Divergence times are expressed in terms of generations, population sizes are in haploid numbers of chromosomes.

Table S10. *A. lyrata* gene identifiers associated with putatively adaptive introgression events.

Region	A. lyrata gene identifiers
Fig. 5 – introgressed region example. Southern Carpathian contact zone.	AL8G22360, AL8G22370, AL8G22380, AL8G22390, AL8G22400, AL8G22410, AL8G22420, AL8G22430, AL8G22440, AL8G22450, AL8G22460, AL8G22470, AL8G22480, AL8G22490, AL8G22500, AL8G22510, AL8G22520, AL8G22530, AL8G22540, AL8G22550, AL8G22560, AL8G22570, AL8G22580, AL8G22590, AL8G22600, AL8G22610
Region 1 – Baltic Contact zone	AL7G24200, AL7G24210, AL7G24220
Region 2 – Baltic Contact zone	AL8G17090

Table S11 Locality details of the populations included in the study.

Pop.	N ind.	Ploidy	Lineage	Altitude	Latitude	Longitude	Country	Locality	Published in	tissue
BEL	8	2x	DINARIC	550	46.16167	16.115	Cro	Belecgrad, open sites in the forest and walls of the castle ruin, limestone	This study	wild
BGS	8	4x	C.EUROPE	570	47.62806	13.00167	D	Berchtesgaden, railway, secondary gravel	Hollister et al. 2012 (BGS)	cultiv.
BIH	8	2x	DINARIC	217	44.88181	15.89882	BH	Bihać, rocks, limestone	This study	wild
BRD	5	4x	C.EUROPE	350	50.04967	13.89081	CZ	Brdatka, open forest in river canyon, silicate	This study	wild
CHO	8	4x	C.EUROPE	103	50.59298	5.44383	Bel	Chokier, rocky slopes, limestone/chalk	This study	wild
CRO	4	2x	A. croat.	1076	44.53147	15.19402	Cro	Ljubičko Brdo + Zavižan, limestone rocks	2 indivs from 44.81825+14.936217	wild
DFS	1	4x	RUDERAL	330	62.922218	15.95832	Swe	Deadfalls	This study	cultiv.
DRA	8	4x	S.CARP.-4x	858	45.44164	25.22394	Rom	Dambovicioara Dracula's limestone gorge	This study	wild
FOJ	8	2x	DINARIC	754	43.97502	17.82446	BH	Fojnica, shady rocky slope, silicate - acidic schists	This study	wild
GOR	8	2x	S.CARP.-2x	184	44.26528	21.54271	Srb	Gornjak, shady rocks, limestone	This study	wild
GUL	10	4x	C.EUROPE	820	47.29	14.93167	At	Gulsen, serpentine rocks, serpentine	Arnold et al. 2016 (GU)	cultiv.
HAR	2	4x	C.EUROPE	400	48.85166	15.85833	At	Hardegg, silicate	This study	cultiv.
HNE	7	2x	PANN.	280	48.26694	19	Hun	Hontianske Nemce, siliceous - volcanites	This study	cultiv.
HNI	4	2x	W.CARP.-2x	836	48.8775	20.5275	Sk	Hnilčík, siliceous - schists and sandstones	Yant et al. 2013 (CA)	cultiv.
HOC	8	4x	C.EUROPE	580	47.37	15.38667	At	Hochlantsch	Arnold et al. 2016 (HO)	cultiv.
KAS	8	4x	C.EUROPE	660	46.68833	14.87167	At	Kasparstein, limestone rocks around the castle ruin	Hollister et al. 2012 (KA)	cultiv.
KOS	7	4x	C.EUROPE	467	47.74694	13.68972	At	Kößlbach, mixed rocky riverbed and adjacent railway bank, calcareous rocks + railway ballast	This study	cultiv.
KOW	8	4x	RUDERAL	670	50.763153	15.8439	Pol	Kowary, secondary gravel	This study	wild
KZL	5	2x	PANN.	330	47.72444	18.77917	Hun	Kesztölc, calcareous	This study	cultiv.
LAC	8	4x	S.CARP.-4x	2092	45.59535	24.63458	Rom	Lacul Capra, alpine scree, base enriched silicate	This study	wild
MIE	8	2x	BALTIC	5	53.92109	14.42157	Pol	Miedzyzdroje, coastal sands, sand	This study	wild
PRE	8	2x	BALTIC	1	55.37821	21.03231	Lit	Preila, coastal sands, sand	This study	wild
RFT	12	4x	C.EUROPE	790	48.10104	9.049581	D	Reiftal, partly sunny, S-facing rock in valley above Neidingen, calcareous	This study	cultiv.
RZA	9	2x	S.CARP.-2x	850	45.37778	22.75833	Rom	Retezat, siliceous rocks	This study	cultiv.
SCH	7	4x	C.EUROPE	2240	47.27767	14.3219	At	Schießeck, screes, amphibolite	This study	wild
SNO	5	2x	W.CARP.-2x	390	49.17417	18.86167	Sk	Strečno, limestone	Yant et al. 2012 (SN)	cultiv.
SPI	13	4x	W.CARP.-4x	550	48.98889	20.775	Sk	Dreveník, limestone	This study	cultiv.
STE	8	4x	RUDERAL	80	52.28028	16.70944	Pol	Stęszew, railway, secondary	This study	cultiv.
SWA	10	4x	C.EUROPE	700	48.44784	9.422422	D	Grindel Steige + Upfinger Steige, shady rocks in beech forest, NE-facing, calcareous	Hollister et al. 2012 (US)	cultiv.
SZI	5	2x	PANN.	130	46.80667	17.43444	Hun	Szigligeti vár, vulcanite	This study	cultiv.

TBG	5	4x	RUDERAL	640	48.13972	8.23667	D	Triberg, railway, secondary	Hollister et al. 2012 (TBG)	cultiv.
TKO	8	4x	W.CARP.-4x	1783	49.20451	19.7352	Sk	Tri Kopy, rocks, screes, neutral - lime-enriched silicate (mylonite)	This study	wild
TRD	6	2x	W.CARP.-2x	1380	49.251596	20.206285	Sk	Tristárska dolina, alpine scree, limestone	This study	wild
TRE	8	4x	W.CARP.-4x	280	48.89417	18.04472	Sk	Trenčín, rocks at the castle ruin, calcareous	This study	cultiv.
TRT	8	4x	W.CARP.-4x	1700	49.24932	20.20498	Sk	Tristárska dolina, alpine scree, limestone	This study	wild
TZI	9	4x	S.CARP.-4x	511	46.56667	23.67417	Rom	Cheile Turzii, limestone gorge	This study	cultiv.
VEL	8	2x	W.CARP.-2x	1823	49.162	20.15419	Sk	Velická dolina, alluvial gravel, wet rocks, alpine, neutral - lime-enriched silicate (mylonite)	This study	wild
VID	8	2x	S.CARP.-2x	900	45.36392	24.63756	Rom	Vidraru, disturbed site in beech forest above road, silicate	This study	wild
WEK	8	4x	C.EUROPE	359	48.405022	15.472906	At	Weißenkirchen, oak-pine forest and former vineyard, silicate	This study	wild
ZAP	5	4x	W.CARP.-4x	915	49.278343	19.96706	Pol	Zakopane, limestone	This study	wild

Table S12. Number of sites per populations used in the genomic load analysis.

	Population	lineage	Total sites
diploids	FOJ	*Dinaric-2x*	12,400,402
	BIH	*Dinaric-2x*	12,482,908
	VEL	*W. Carp.-2x*	12,928,709
	GOR	*S. Carp.-2x*	12,804,746
	HNE	*Pann.-2x*	10,053,779
	RZA	*S. Carp.-2x*	11,284,904
	SNO	*W. Carp.-2x*	9,450,560
	MIE	*Baltic-2x*	12,843,047
	PRE	*Baltic-2x*	13,105,828
tetraploids	TKO	*W. Carp.-4x*	12,930,327
	SCH	*C. Europe-4x*	10,346,987
	KOW	*Ruderal-4x*	12,571,603
	DRA	*S. Carp -4x*	11,510,853
	HOC	*C. Europe-4x*	9,040,625
	KAS	*C. Europe-4x*	11,634,487
	STE	*Ruderal-4x*	11,066,987
	TZI	*S. Carp.-4x*	10,777,235
	SPI	*W. Carp.-4x*	13,663,634

Table S13 Average genome-wide frequencies of alleles diagnostic to each diploid lineage of *A. arenosa* (in columns) and *A. lyrata*-specific alleles (reference alleles with max frequency 0.07 in total *A. arenosa* dataset) in *A. arenosa* populations and lineages (in rows). Values above 0.05 (*A. arenosa* diploid alleles) and 0.02 (*A. lyrata* alleles) are in bold.

		alleles specific to diploid lineage					
tetraploid population	tetraploid lineage	Pannonian-2x	Dinaric-2x	Baltic-2x	S. Carp.-2x	W. Carp.-2x	*A. lyrata*
BGS	*C. Europe*	0.008	0.012	0.032	0.032	**0.161**	**0.022**
BRD	*C. Europe*	0.01	0.013	0.01	0.02	**0.144**	**0.029**
CHO	*C. Europe*	0.005	0.008	0.006	0.01	**0.065**	0.018
DRA	*S. Carp.*	0.005	0.007	0.007	**0.138**	**0.064**	0.011
GUL	*C. Europe*	0.007	0.013	0.009	0.018	**0.134**	0.017
HOC	*C. Europe*	0.009	0.015	0.009	0.023	**0.139**	0.015
KAS	*C. Europe*	0.008	0.016	0.012	0.024	**0.163**	0.015
KOS	*C. Europe*	0.007	0.012	0.011	0.017	**0.192**	**0.039**
KOW	*Ruderal*	0.006	0.009	**0.082**	**0.063**	**0.107**	**0.031**
LAC	*S. Carp.*	0.007	0.01	0.01	**0.076**	**0.178**	0.008
RFT	*C. Europe*	0.009	0.014	0.009	0.021	**0.112**	0.011
SCH	*C. Europe*	0.004	0.007	0.006	0.012	**0.145**	**0.023**
SPI	*W. Carp*	0.005	0.008	0.011	0.012	**0.253**	0.005
STE	*Ruderal*	0.006	0.01	**0.082**	**0.06**	**0.113**	**0.039**
SWA	*C. Europe*	0.011	0.016	0.01	0.024	**0.125**	0.01
TBG	*Ruderal*	0.007	0.009	**0.068**	**0.054**	**0.119**	**0.033**
TKO	*W. Carp*	0.004	0.007	0.008	0.009	**0.37**	0.005
TRE	*W. Carp*	0.014	0.01	0.011	0.014	**0.184**	0.009
TRT	*W. Carp*	0.003	0.006	0.008	0.007	**0.316**	0.007
TZI	*S. Carp.*	0.006	0.009	0.011	**0.079**	**0.077**	0.011
WEK	*C. Europe*	0.019	0.004	0.005	0.009	**0.071**	**0.046**
ZAP	*W. Carp.*	0.004	0.006	0.009	0.008	**0.262**	0.016
C. Europe-4x		0.009	0.012	0.011	0.019	0.132	**0.022**
Ruderal-4x		0.006	0.009	0.077	0.059	0.113	**0.034**
S. Carp.-4x		0.006	0.009	0.009	0.098	0.106	0.01
W. Carp.-4x		0.006	0.007	0.009	0.01	0.277	0.008
Diploids		-	-	-	-	-	0.009
N sites		81744	65201	35320	25378	28630	27637
Differences among groups (ANOVA test)		$F(3,18) = 0.6$, n.s.	$F(3,18) = 0.34$, n.s.	$F(3,18) = 52.3$, $p < 0.001$	$F(3,18) = 33.4$, $p < 0.001$	$F(3,18) = 14.5$, $p < 0.001$	$F(3,18) = 6.77$, $p = 0.003$

Table S14 Parameter estimates from the DFE-alpha analysis based on the folded sites frequency spectrum for all diploid and tetraploid populations. β and E*s* are the shape and mean selection strength of selection of the fitted gamma distribution. The fraction of considered neutral and the proportion of sites in increasing bins of strength of purifying selection is given. α and ωα are the proportion of adaptive substitution and proportion of adaptive substitution relative to neutral. Number in bold are the median value of parameter estimates for diploid and tetraploid populations

Pop	Ploidy	β	Es	Fraction neutral	Nes : 0 - 1	Nes : 1 - 10	Nes : 10 - >100	se_X10.100	α	ωα
BEL	2	0.23 (2.01E-03)	-4.3 (2.71)	0.2	0.2 (4.62E-04)	0.14 (1.36E-03)	0.67 (1.36E-03)	0.42	0.42 (1.65E-03)	0.13 (5.30E-04)
BIH	2	0.196 (6.64E-04)	-26.02 (0.58)	0.2	0.2 (2.87E-04)	0.12 (3.70E-04)	0.68 (3.08E-04)	0.39	0.39 (1.12E-03)	0.12 (4.07E-04)
FOJ	2	0.168 (1.20E-03)	-91.56 (3.44)	0.19	0.19 (4.35E-04)	0.09 (9.78E-04)	0.72 (1.31E-03)	0.41	0.41 (1.24E-03)	0.12 (4.05E-04)
GOR	2	0.172 (1.28E-03)	-25.62 (18.56)	0.21	0.21 (4.84E-04)	0.1 (9.61E-04)	0.69 (1.28E-03)	0.37	0.37 (1.46E-03)	0.11 (4.70E-04)
HNE	2	0.193 (6.31E-04)	-110.63 (2.95)	0.18	0.18 (2.82E-04)	0.1 (3.26E-04)	0.72 (3.42E-04)	0.44	0.44 (1.17E-03)	0.13 (4.22E-04)
KZL	2	0.215 (9.74E-04)	-7.1 (0.4)	0.19	0.19 (3.15E-04)	0.12 (5.64E-04)	0.69 (4.43E-04)	0.41	0.41 (1.32E-03)	0.12 (4.57E-04)
MIE	2	0.141 (1.00E-03)	-146.86 (133.78)	0.21	0.21 (3.79E-04)	0.08 (7.10E-04)	0.71 (9.58E-04)	0.35	0.35 (1.16E-03)	0.11 (3.81E-04)
PRE	2	0.138 (4.74E-04)	-550.06 (27.29)	0.2	0.2 (2.56E-04)	0.08 (2.44E-04)	0.72 (2.53E-04)	0.39	0.39 (9.86E-04)	0.12 (3.48E-04)
RZA	2	0.241 (1.82E-03)	-3.73 (1.16)	0.19	0.19 (5.97E-04)	0.14 (9.70E-04)	0.67 (5.10E-04)	0.44	0.44 (2.33E-03)	0.13 (7.53E-04)
SNO	2	0.182 (8.12E-04)	-6.8 (0.18)	0.24	0.24 (3.03E-04)	0.12 (5.88E-04)	0.64 (5.27E-04)	0.36	0.36 (1.14E-03)	0.12 (4.54E-04)
SZI	2	0.064 (2.75E+00)	-2.29E+07 (9.08E+08)	0.22	0.22 (5.97E-03)	0.04 (2.30E-02)	0.74 (1.82E-02)	0.14	0.14 (2.34E-02)	0.03 (5.79E-03)
TRD	2	0.217 (5.16E-04)	-5.22 (0.08)	0.19	0.19 (1.95E-04)	0.13 (3.18E-04)	0.68 (3.16E-04)	0.45	0.45 (6.58E-04)	0.14 (2.57E-04)
VEL	2	0.213 (4.68E-04)	-13.63 (0.19)	0.2	0.2 (2.16E-04)	0.12 (2.58E-04)	0.68 (2.54E-04)	0.44	0.44 (8.19E-04)	0.14 (3.18E-04)
VID	2	0.22 (5.21E-04)	-5.29 (0.21)	0.2	0.2 (2.28E-04)	0.13 (3.08E-04)	0.67 (3.05E-04)	0.43	0.43 (8.73E-04)	0.13 (3.37E-04)
		0.19	**-19.62**	**0.9**	**0.2**	**0.12**	**0**	**0.68**	**0.41**	**0.12**
BGS	4	0.241 (8.30E-04)	-11.28 (0.08)	0.19	0.19 (2.84E-04)	0.14 (4.34E-04)	0.67 (2.89E-04)	0.54	0.54 (9.16E-04)	0.2 (4.79E-04)
BRD	4	0.285 (8.14E-04)	-11.7 (0.12)	0.18	0.18 (2.59E-04)	0.16 (4.42E-04)	0.66 (2.95E-04)	0.55	0.55 (9.49E-04)	0.19 (4.79E-04)
CHO	4	0.297 (1.37E-03)	-8.91 (0.17)	0.18	0.18 (4.37E-04)	0.17 (7.16E-04)	0.65 (4.06E-04)	0.58	0.58 (1.47E-03)	0.21 (7.45E-04)
DRA	4	0.159 (2.92E-03)	-33.94 (1.07)	0.21	0.21 (9.28E-04)	0.09 (1.54E-03)	0.69 (6.58E-04)	0.43	0.43 (3.11E-03)	0.15 (1.12E-03)
GUL	4	0.249 (1.16E-03)	-12.06 (0.11)	0.18	0.18 (3.56E-04)	0.14 (5.76E-04)	0.68 (3.09E-04)	0.55	0.55 (1.19E-03)	0.19 (6.09E-04)
HOC	4	0.282 (7.78E-04)	-7.15 (0.04)	0.18	0.18 (2.85E-04)	0.16 (4.02E-04)	0.66 (2.98E-04)	0.55	0.55 (9.64E-04)	0.19 (4.73E-04)
KAS	4	0.168 (3.08E-04)	-142.87 (1.73)	0.2	0.2 (2.03E-04)	0.09 (1.61E-04)	0.71 (2.23E-04)	0.45	0.45 (8.71E-04)	0.15 (4.16E-04)
KOS	4	0.235 (8.44E-04)	-25.54 (0.23)	0.19	0.19 (2.87E-04)	0.14 (4.64E-04)	0.67 (3.30E-04)	0.52	0.52 (9.62E-04)	0.19 (4.88E-04)
KOW	4	0.181 (2.60E-04)	-200.01 (2.25)	0.2	0.2 (1.51E-04)	0.1 (1.67E-04)	0.7 (2.33E-04)	0.49	0.49 (7.19E-04)	0.17 (4.35E-04)
LAC	4	0.173 (1.45E-03)	-24.99 (41.33)	0.21	0.21 (2.32E-04)	0.1 (9.71E-04)	0.69 (9.88E-04)	0.39	0.39 (1.00E-03)	0.12 (3.76E-04)
RFT	4	0.243 (7.55E-04)	-14.1 (0.1)	0.19	0.19 (2.76E-04)	0.14 (3.86E-04)	0.67 (2.78E-04)	0.5	0.5 (9.98E-04)	0.16 (4.59E-04)
SCH	4	0.2 (1.29E-03)	-123.78 (2.24)	0.19	0.19 (3.70E-04)	0.11 (6.98E-04)	0.69 (4.37E-04)	0.47	0.47 (1.32E-03)	0.16 (5.40E-04)
SPI	4	0.221 (6.38E-04)	-24.34 (0.28)	0.19	0.19 (2.52E-04)	0.12 (3.16E-04)	0.69 (2.25E-04)	0.51	0.51 (9.41E-04)	0.17 (4.32E-04)
STE	4	0.219 (7.72E-04)	-28.16 (0.33)	0.19	0.19 (2.64E-04)	0.12 (4.16E-04)	0.69 (3.00E-04)	0.56	0.56 (9.04E-04)	0.22 (5.68E-04)
SWA	4	0.241 (5.88E-04)	-16.78 (0.13)	0.19	0.19 (2.23E-04)	0.14 (3.35E-04)	0.66 (2.72E-04)	0.48	0.48 (8.29E-04)	0.16 (3.74E-04)
TBG	4	0.228 (7.88E-04)	-22.51 (0.21)	0.2	0.2 (2.60E-04)	0.14 (4.53E-04)	0.67 (3.35E-04)	0.51	0.51 (9.43E-04)	0.18 (4.93E-04)
TKO	4	0.231 (7.57E-04)	-7.47 (0.09)	0.19	0.19 (2.90E-04)	0.13 (4.06E-04)	0.68 (2.76E-04)	0.5	0.5 (1.02E-03)	0.17 (4.19E-04)
TRE	4	0.198 (5.82E-04)	-48.73 (0.51)	0.2	0.2 (2.17E-04)	0.11 (3.25E-04)	0.69 (2.67E-04)	0.48	0.48 (8.06E-04)	0.16 (3.69E-04)
TRT	4	0.187 (1.05E-03)	-13.78 (0.18)	0.2	0.2 (3.32E-04)	0.11 (5.80E-04)	0.69 (3.68E-04)	0.45	0.45 (1.25E-03)	0.15 (4.74E-04)
TZI	4	0.168 (7.58E-04)	-17.66 (0.32)	0.22	0.22 (3.29E-04)	0.1 (4.07E-04)	0.68 (2.87E-04)	0.44	0.44 (1.37E-03)	0.15 (6.65E-04)
WEK	4	0.113 (5.63E-04)	-379.95 (43.38)	0.22	0.22 (2.64E-04)	0.06 (3.13E-04)	0.72 (2.67E-04)	0.53	0.53 (1.05E-03)	0.22 (8.13E-04)
ZAP	4	0.165 (3.49E-04)	-28.45 (0.44)	0.2	0.2 (1.79E - 04)	0.09 (1.94E-04)	0.71 (2.17E-04)	0.46	0.46 (7.28E-04)	0.15 (3.58E-04)
		0.22	**-23.42**	**0.88**	**0.19**	**0.12**	**0.68**	**0**	**0.5**	**0.17**

List of supplementary files

File_S1_popstats.xlsx Measures of genome-wide diversity within the 36 populations of *A. arenosa* with ≥ 5 individuals sequenced and details on inclusion of the populations in the downstream analyses.

File_S2_NGS_Analysis_Steps_300.xlsx. Steps used for processing, mapping, and variant calling.

File_S3 _Summary_QC.xlsx. Sequence processing quality assessment of each sequenced individual. PE – Paired End reads, SE – Single End reads, TSPf – TruSeq PCR free, DOC – Depth Of Coverage, DOCq – DOC considering quality cutoffs of minimum base quality of 25 and minimum mapping quality of 25, DOCq bases_above_4 – percentage of bases that had a DOCq of 4x or more, mapping efficiency = mapped reads/total reads. The second sheet shows the number of SNPs per scaffold/chromosome that either passed or failed all filters and final N of SNPs per scaffold after filtering.

File_S4_AFS.pdf Unfolded allele frequency spectra of the 36 A. *arenosa* populations with ≥5 individuals. For populations with >5 individuals, we randomly downsampled to 5 individuals at each site.

File_S5_plastome_sequences.fas Fasta of 291 plastome sequences from our study used for construction of the plastome phylogeny tree (inverted repeats and regions around complex indels were later excluded).

File_S6_plastome_tree.pdf Maximum likelihood phylogeny of *Arabidopsis* plastomes from our study and of Novikova et al. (2016).

File_S7_fsc2.docx Example parameter files used in *fastsimcoal2.*

File_S8_AFS.txt Data used to generate File_S4_AFS.pdf. To calculate the scaled relative frequency in each population, we first calculated the relative frequency in each bin for each data type (All sites, 4dg sites, and nonsynonymous sites). Then, we scaled each of these relative frequencies with respect to the maximum relative frequency observed across data types for that population. We then divided each relative frequency by this maximum to scale all observations relative to this max value, whose scaled value now becomes one. This scaling preserves the relative proportions within a data type, but also allows us to compare relative frequencies across data types by scaling to this common value.