











# Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles

Veronika Konečná <sup>1,2</sup>, Sian Bray<sup>3</sup>, Jakub Vlček <sup>1,4,5</sup>, Magdalena Bohutínská<sup>1,2</sup>, Doubravka Požárová<sup>1</sup>, Rimjhim Roy Choudhury<sup>6,7</sup>, Anita Bollmann-Giolai <sup>8</sup>, Paulina Flis <sup>3</sup>, David E. Salt <sup>3</sup>, Christian Parisod <sup>6</sup>, Levi Yant <sup>9,10</sup>  & Filip Kolář <sup>1,2,10</sup> 

Relative contributions of pre-existing vs de novo genomic variation to adaptation are poorly understood, especially in polyploid organisms. We assess this in high resolution using autotetraploid *Arabidopsis arenosa*, which repeatedly adapted to toxic serpentine soils that exhibit skewed elemental profiles. Leveraging a fivefold replicated serpentine invasion, we assess selection on SNPs and structural variants (TEs) in 78 resequenced individuals and discover significant parallelism in candidate genes involved in ion homeostasis. We further model parallel selection and infer repeated sweeps on a shared pool of variants in nearly all these loci, supporting theoretical expectations. A single striking exception is represented by *TWO PORE CHANNEL 1*, which exhibits convergent evolution from independent de novo mutations at an identical, otherwise conserved site at the calcium channel selectivity gate. Taken together, this suggests that polyploid populations can rapidly adapt to environmental extremes, calling on both pre-existing variation and novel polymorphisms.

<sup>1</sup>Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic. <sup>2</sup>The Czech Academy of Sciences, Institute of Botany, Průhonice, Czech Republic. <sup>3</sup>Future Food Beacon and School of Biosciences, University of Nottingham, Nottingham, UK. <sup>4</sup>Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic. <sup>5</sup>Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. <sup>6</sup>Institute of Plant Sciences, University of Berne, Bern, Switzerland. <sup>7</sup>Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland. <sup>8</sup>John Innes Centre (JIC), Norwich Research Park, Norwich, UK. <sup>9</sup>Future Food Beacon and School of Life Sciences, University of Nottingham, Nottingham, UK. <sup>10</sup>These authors jointly supervised this work: Levi Yant, Filip Kolář. ✉email: [levi.yant@gmail.com](mailto:levi.yant@gmail.com); [filip.kolar@natur.cuni.cz](mailto:filip.kolar@natur.cuni.cz)

Rapid adaptation to novel environments is thought to be enhanced by the availability of genetic variation; however, the relative contribution of standing variation versus the role of novel mutation is a matter of debate<sup>1,2</sup>, especially in higher ploidy organisms. Whole-genome duplication (WGD; leading to polyploidisation) is a major force underlying diversification across eukaryotic kingdoms, seen most clearly in plants<sup>3–5</sup> with various effects on genetic variation<sup>6–8</sup>. While WGD is clearly associated with environmental change or stress<sup>5,9</sup>, the precise impact of WGD on adaptability is largely unknown in multicellular organisms, and there is virtually no work assessing the evolutionary sources of adaptive genetic variation in young polyploids. Work in autopolyploids, which clearly isolate effects of WGD from hybridisation (which is confounded in allopolyploids), indicates that subtle genomic changes may follow WGD alone<sup>8,10,11</sup>, which raises the question of when their adaptive value may originate.

Autopolyploidy is expected to alter selective and adaptive process in many ways, but a dearth of empirical data prevents synthetic evaluation. Besides immediate phenotypic<sup>12–14</sup> and genomic<sup>10,11</sup> changes following WGD, theory is unsettled regarding how adaptation proceeds as the autopolyploid lineage diversifies and adapts to novel challenges. On the one hand, autopolyploids can mask deleterious alleles and accumulate cryptic allelic diversity<sup>7</sup>. In addition, the number of mutational targets is multiplied in autopolyploids, meaning that new alleles are introduced more quickly<sup>6,15,16</sup>. This could promote adaptation<sup>6,17</sup>. On the other hand, reduced rates of allele frequency changes may retard adaptation<sup>6,18</sup>, particularly for de novo mutations, which emerge in a population at initially low frequencies<sup>19</sup>. Recent advances in theory and simulations suggest potential solutions to this controversy. Polyploidy may promote adaptation under scenarios of rapid environmental change (e.g. colonisation of challenging habitats) when selection is strong and originally neutral or mildly deleterious alleles standing in polyploid populations may become beneficial<sup>5,20</sup>. However, empirical evidence supporting this scenario is fragmentary. There is broad correlative evidence that polyploids are good colonisers of areas experiencing environmental flux (e.g. the Arctic<sup>21,22</sup>, stressful habitats<sup>8,23,24</sup>, and heterogeneous environments<sup>25</sup>). However, the genomic basis of such polyploid adaptability—and whether their primary source of adaptive alleles is high diversity (standing variation) or large mutational target size (de novo mutations)—remains unknown.

We focus on natural autotetraploid *Arabidopsis arenosa* populations repeatedly facing one of the greatest environmental challenges for plant life—naturally toxic serpentine soils. Serpentine soils occur as islands in the landscape with no intermediate habitats and are defined by peculiar elemental contents (highly skewed Ca/Mg ratio and elevated heavy metals such as Cr, Co, and Ni), that may be further combined with low nutrient availability and propensity for drought<sup>26</sup>. *Arabidopsis arenosa* is a well-characterised, natural diploid-autotetraploid species with large and genetically diverse outcrossing populations<sup>27</sup>. The widespread autotetraploids, which originated from a single diploid lineage ~19–31k generations ago<sup>8</sup>, harbour increased adaptive diversity genome wide<sup>8</sup> and currently occupy a broader ecological niche than their diploid sisters<sup>28</sup>, including serpentine outcrops<sup>29</sup>, railway lines<sup>30,31</sup>, and contaminated mine tailings<sup>32,33</sup>. This makes *A. arenosa* a promising model for empirical inquiries of adaptation in autopolyploids<sup>8,27</sup>. As a proof of concept, selective ion uptake phenotypes and a polygenic basis for serpentine adaptation have been suggested from a single *A. arenosa* serpentine population<sup>29</sup>. However, limited sampling left unknown whether the same genes are generally (re)used and what is the evolutionary source of the selected alleles, i.e. leaving

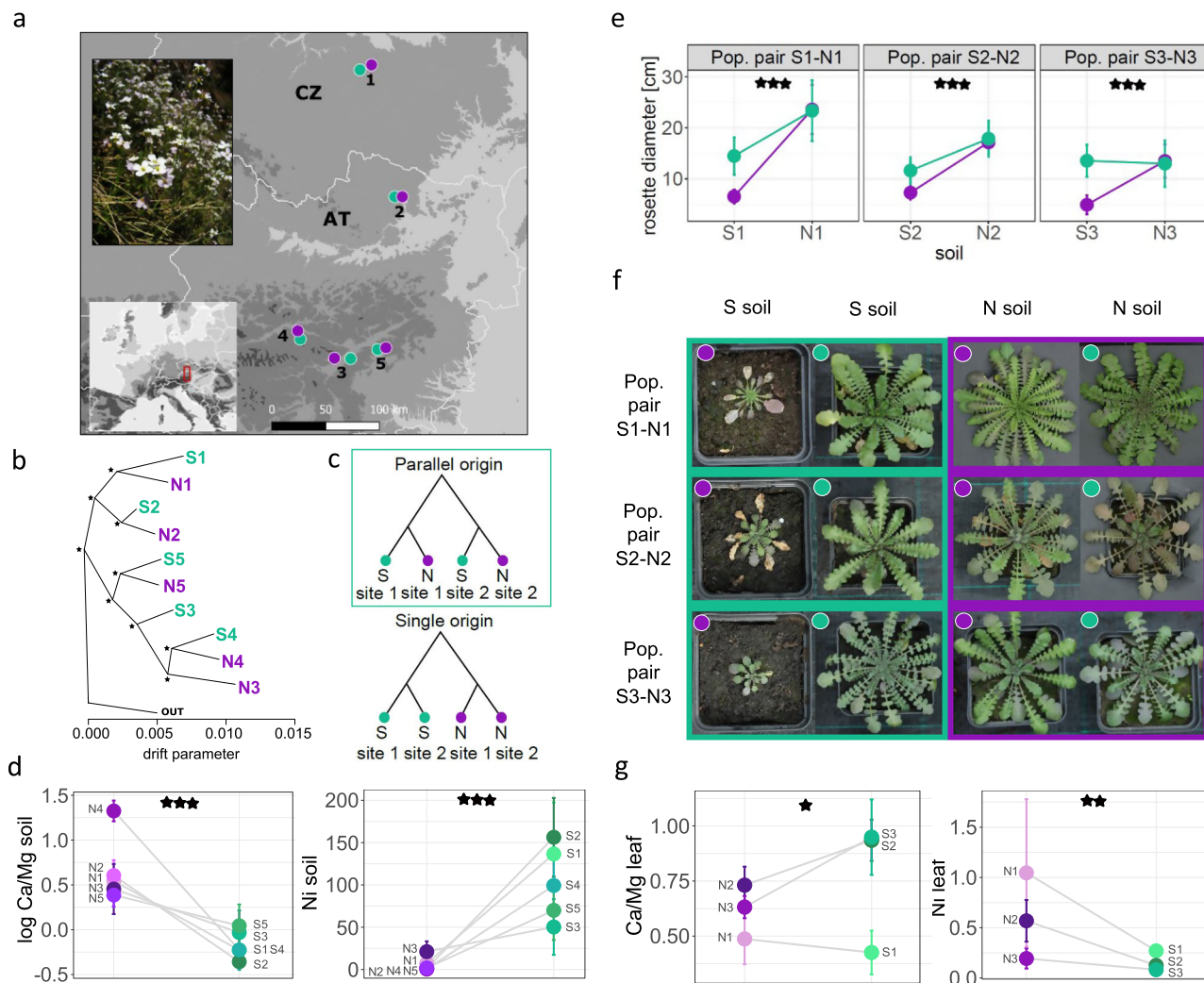
unresolved the evolutionary dynamics and mechanism underlying these striking adaptations. We ask specifically: (1) Does gene-level parallelism in autotetraploid *A. arenosa* dominantly reflect repeated sampling from the large pool of shared variation that is expected to be maintained in autopolyploids? and (2) Is repeated adaptation from novel mutations feasible in autotetraploid populations?

In this work, we deconstruct the sources of parallel adaptive variation in *A. arenosa*. First, we sample five serpentine/non-serpentine population pairs of autotetraploid *A. arenosa* and demonstrate rapid parallel adaptation by combining demographic analysis and reciprocal transplant experiments. Taking advantage of the power of this fivefold replicated natural selection experiment, we identify candidate adaptive loci from population resequencing data and find significant parallelism underlying serpentine adaptation. We then model parallel selection using a designated framework and statistically infer the evolutionary sources of parallel adaptive variation for all candidate loci. In line with theory, we find that shared variation is the vastly prevalent source of parallel adaptive variants in serpentine *A. arenosa*. However, we also discover an exceptional locus exhibiting footprints of selection on alleles originating from two distinct de novo mutations. In line with the latter hypothesis, this demonstrates that the rapid selection of novel alleles is still feasible in autopolyploids, indicating broad evolutionary flexibility of lineages with doubled genomes.

## Results

**Parallel serpentine adaptation.** First, we inferred independent colonisation of each serpentine site by different local *A. arenosa* populations. To do this, we resequenced five pairs of geographically proximate serpentine (S) and non-serpentine (N) populations covering all known serpentine sites occupied by the species to date (8 individuals per population on average, mean sequencing depth 21×; Fig. 1a, Supplementary Fig. 1, Supplementary Data 1, and Supplementary Tables 1–3). Phylogenetic, ordination, and Bayesian analyses based on nearly neutral fourfold-degenerate (4dg) sites demonstrated overall grouping of populations by spatial proximity, not by substrate. In all but one case, the adjacent S and N populations occupied sister position in the population tree and belonged to the same Bayesian cluster; only the population S3 occupied somewhat isolated position yet still within the lineage of Eastern Alpine populations (Fig. 1b and Supplementary Fig. 1). We thus further tested the independent colonisation of each serpentine site by coalescent simulations. Consistently over all possible pairwise iterations of S–N population pairs ( $n = 10$ ), the scenario of independent colonisation of each serpentine site was more likely than any scenario assuming sister position of two S populations (Fig. 1c). Note that subsequent gene flow between substrate types within each S–N population pair was unlikely as the assumption of migration within each population pair had not significantly improved the model fit (Supplementary Fig. 2 and Supplementary Data 2). Reflecting the independent origin of the five S populations, we analysed each serpentine colonisation event separately in the following analyses to take into account neutral population structure in the data, using the spatially closest N population as a contrast where needed. The very low differentiation between S and proximate N populations and consistently low population split times (Table 1) indicate very recent, postglacial serpentine invasions. There is no evidence of bottleneck associated with colonisation, as S and N populations exhibited similar nucleotide diversity and Tajima's  $D$  values (Table 1).

To assess whether the colonisation of serpentine soils was accompanied by substrate adaptation, we combined ionomics



**Fig. 1 Parallel adaptation of *Arabidopsis arenosa* to challenging serpentine soils.** **a** Locations of the investigated serpentine (S, green) and non-serpentine (N, violet) populations sampled as spatially proximate pairs (numbers) in Central Europe with an illustrative photo of an S population (photo was taken by F. Kolár). **b** Allele frequency covariance graph of populations based on ~870,000 fourfold-degenerate SNPs; asterisks show the 100 bootstrap branch support. The outgroup (OUT) is represented by a tetraploid population from Western Carpathians, the ancestral area of tetraploid *A. arenosa*<sup>93</sup>. **c** Two contrasting evolutionary scenarios of serpentine colonisation compared in coalescent simulations; the topology assuming independent serpentine colonisations (framed in green) received the highest support consistently across all 10 pairwise combinations of S–N population pairs. **d** Differences in Ca/Mg ratio and in Ni concentrations [µg/g] in S and N soils from the original sampling sites ( $n = 78$  individual samples, one-way ANOVAs:  $F_{1,77} = 26.5$ ,  $p = 1.94 \times 10^{-6}$  and  $F_{1,77} = 117.4$ ,  $p = 2.01 \times 10^{-16}$  for Ca/Mg and Ni, respectively). **e** Differences in maximum rosette size of three population pairs attained after 3 months of cultivation in local serpentine and non-serpentine substrates (significance of the soil treatment  $\times$  soil origin interaction in a two-way ANOVA is indicated:  $F_{1,90} = 21.6$ ,  $p = 1.17 \times 10^{-5}$ ,  $F_{1,96} = 12.3$ ,  $p = 6.88 \times 10^{-4}$ , and  $F_{1,85} = 42$ ,  $p = 5.68 \times 10^{-9}$  for population pairs 1, 2 and 3, respectively). **f** Example photos illustrating parallel growth response in the three population pairs to serpentine soils (green frame) depending on the soil of origin (dot colour) (photo was taken by V. Konečná). **g** Differences in ion uptake between originally S and N individuals when cultivated in serpentine soils; Ni concentrations were standardised by corresponding soil Ni values ( $n = 28$  individual samples, one-way ANOVAs:  $F_{1,27} = 6.2$ ,  $p = 0.019$  and  $F_{1,27} = 13.5$ ,  $p = 0.001$  for Ca/Mg ratio and Ni, respectively). Points denote mean, error bars depict standard error of mean in charts **e**, **d**, **g**. Source data underlying Fig. 1d, g are provided as a Source data file.

with a reciprocal transplant experiment. First, using ionic profiling of native soil associated with each sequenced individual, we characterised major chemical parameters differentiating on both substrates (Fig. 1d and Supplementary Figs. 3 and 4). Among the 20 elements investigated (Supplementary Fig. 3a), only the bioavailable concentration of Mg, Ni, Co, and Ca/Mg ratio consistently differentiated both soil types (Bonferroni-corrected one-way analysis of variance (ANOVA) taking population pair as a random variable). Serpentine sites were not macronutrient poor (Supplementary Table 4) and were not differentiated from non-serpentines by bioclimatic parameters (annual temperature, precipitation, and elevation; Supplementary

Fig. 3), indicating that skewed Ca/Mg ratios and elevated heavy metal content are likely the primary selective agents on the sampled serpentine sites<sup>34–36</sup>.

We then tested for differential fitness response towards serpentine soil between populations of S versus N origin using reciprocal transplant experiments. We cultivated plants from three population pairs (S1–N1, S2–N2, and S3–N3) on both native soil types within each pair for three months (until attaining maximum rosette size), observing significantly better germination and growth of the S plants in their native serpentine substrate as compared to their closest N relatives. First, we found a significant interaction between soil type and soil of origin at germination

**Table 1** Between-population divergence and within-population diversity of the five investigated serpentine/non-serpentine population pairs inferred from genome-wide fourfold-degenerate single nucleotide polymorphisms.

Population pair	Divergence (generations) <sup>a</sup>	Pairwise $F_{ST}$	Nucleotide diversity <sup>b</sup>	Tajima's $D^b$
S1-N1	(774) 4317 (6284)	0.069	0.0292/0.0276	0.266/0.485
S2-N2	(500) 2690 (3826)	0.029	0.0306/0.0285	0.131/0.372
S3-N3	(794) 3912 (6316)	0.085	0.0307/0.0287	-0.046/0.350
S4-N4	(812) 2918 (3542)	0.057	0.0304/0.0297	0.132/0.177
S5-N5	(546) 3539 (4869)	0.047	0.0292/0.0296	0.267/0.089

<sup>a</sup>Divergence between proximal S-N populations. Mean and 95% confidence intervals inferred by bootstrapping, estimated by coalescent simulations. Assuming 2-year generation time<sup>93</sup>, all estimates indicate recent postglacial divergence.

<sup>b</sup>Genome-wide nucleotide diversity ( $\pi$ ) and Tajima's D of each S-N population.

(generalised linear model (GLM) with binomial errors taking population pair as a random variable,  $\chi^2 = 22.436$ ,  $p < 0.001$ ), although the fitness disadvantage of N plants in serpentine soil varied across population pairs (Supplementary Fig. 5). During subsequent cultivation, we recorded zero mortality but found a significant interaction effect between soil treatment and soil of origin on growth, as approximated by maximum rosette sizes (two-way ANOVA taking population pair as a random variable,  $F_{1,277} = 55.5$ ,  $p < 0.001$ , Fig. 1e; see Supplementary Fig. 6 for rosette size temporal development). Once again, the S plants consistently produced significantly larger rosettes (by 47% on average) than their N counterparts when grown in serpentine soil, indicating consistent substrate adaptation (Fig. 1e, f). Finally, we evaluated differences in Ni and Ca/Mg accumulation in leaves harvested on plants cultivated in serpentine soils. Consistent with adaptive responses to soil chemistry, we found higher Ca/Mg ratio and reduced uptake of Ni (lower leaf/soil ratio) in tissue of serpentine plants relative to their non-serpentine counterparts (Fig. 1g). Taken together, our demographic analysis complemented by transplant experiments support recent parallel serpentine adaptation of autotetraploid *A. arenosa* at five distinct sites, exhaustively covering all known serpentine populations of the species.

**Parallel genomic footprints of selection on serpentine at single-nucleotide polymorphisms (SNPs) and transposable elements (TEs).** Using these five natural replicates of serpentine adaptation, we sought the genomic basis and evolutionary source of the parallel adaptations. To do this, we combined divergence scans and environmental association analysis to refine the list of loci for parallel selection modelling only to the candidates that repeatedly differentiated across multiple population pairs and were significantly associated with the selective soil environment. First, we identified initial inclusive lists of gene-coding loci exhibiting excessive differentiation between paired populations using 1% outlier  $F_{ST}$  window-based scans (490–525 candidate genes per pair; details in ‘Methods’; Supplementary Data 3). These most inclusive lists must be interpreted with caution, as they are based on a simple assumption that the most differentiated regions are under directional selection<sup>37</sup>. However, in support of their relevance a gene ontology (GO) analysis of the 2245 candidates from all five population pairs shows significant enrichment (Fisher’s exact test;  $p < 0.05$ ) of ‘biological processes’, ‘molecular functions’, and ‘cellular components’ considered relevant to serpentine adaptation<sup>26,34</sup>, such as inorganic anion transport, ion homeostasis, post-embryonic development, and calcium transmembrane transporter activity (Fig. 2b and Supplementary Data 4).

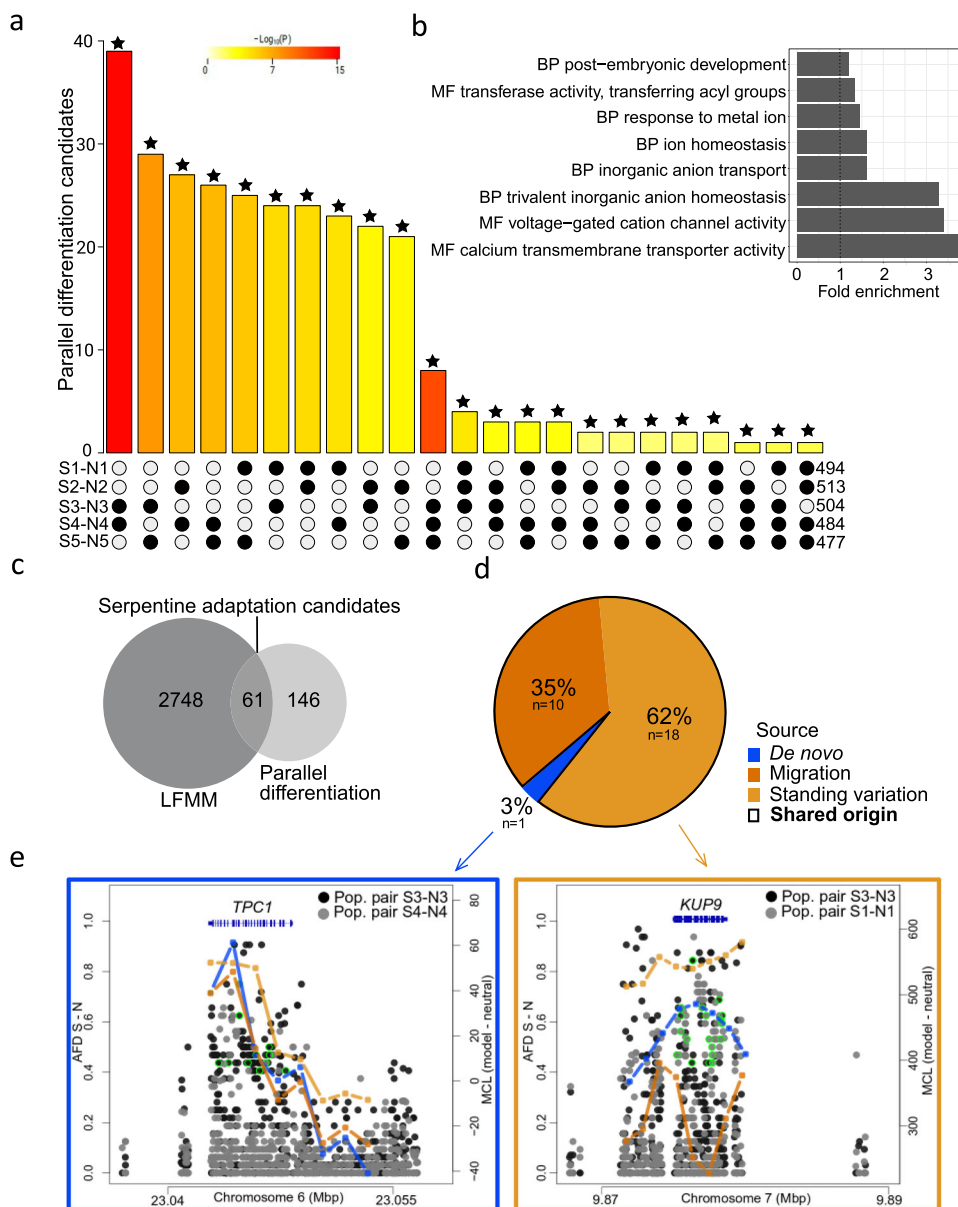
To refine this broad list and pinpoint parallel evolution candidate genes, we overlapped these candidate gene lists across population pairs, identifying 207 ‘parallel differentiation candidates’ that represent divergence outliers in at least two S–N population pairs. The level of parallelism was greater than

expected by chance for all pairs of S–N contrasts (Fisher’s exact test; Fig. 2a and Supplementary Data 4) and we hereafter refer to this as ‘significant parallelism’. Such a fraction of parallel gene candidates (0.02–0.04 out of all candidates from that particular population pair) is in line with other naturally adapting systems of comparable divergence<sup>33,38–40</sup>. The parallel differentiation candidates were significantly enriched ( $p < 0.05$ ) for GO terms, such as regulation of ion transmembrane transport, voltage-gated calcium and potassium channel activity or plasma membrane (Supplementary Data 4). The absence of common candidates across all five population pairs may reflect a complex genetic basis of the traits allowing for the modulation of the same pathway by different genes in some populations. This is supported by significant functional parallelism, i.e. higher than random number of overlapping GO terms that were repeatedly identified by separate enrichment analyses of outlier gene list from each population pair (Supplementary Fig. 7 and Supplementary Data 4). Additionally, adaptation via partial (soft) sweeps, which are likely to occur in autotetraploids<sup>19</sup>, might have further limited the power of our divergence scans in some loci and populations.

As a complementary approach, we inferred candidates directly associated with the distinctive chemical characteristics of serpentine soil by performing environmental association analysis using latent factor mixed models (LFMM)<sup>41</sup>. This analysis quantitatively determines the association between each soil elemental concentration and SNPs across the genome in both S and N populations at the level of individual plants (in total 78). We identified 2,809 genes (LFMM candidates) harbouring  $\geq 1$  SNP significantly associated with at least one distinctive serpentine soil parameter previously identified by ionic analysis (Ca/Mg ratio, high Mg, Ni, and Co; Supplementary Data 5). Finally, we overlapped the LFMM candidates with the parallel differentiation candidates to produce a final refined list of 61 ‘serpentine adaptation candidates’ (Fig. 2c, Supplementary Fig. 8, and Supplementary Data 6 and 7). This conservative approach aims to identify the strongest candidates underlying serpentine adaptation for further model-based inference of the sources of variation in the next section. We note that this approach discards population-specific (private) candidates and cases of distinct genetic architecture of a trait (e.g. distinct genes affecting the same pathway) and thus cannot quantify the overall genome proportion that evolves in parallel. Importantly, however, it also minimises false positives from population-specific selection and genetic drift.

These 61 serpentine adaptation candidates were significantly enriched ( $p < 0.05$ ) for categories related, for example, to regulation of ion transmembrane transport, and specifically, voltage-gated calcium and potassium channel activity (Supplementary Data 7). Candidates included the *NRT2.1* and *NRT2.2* high-affinity nitrate transporters, which act as repressors of lateral root initiation<sup>42,43</sup>; *RHF1A*, which is involved in gametogenesis and transferase activities<sup>44,45</sup>; *TPC1*, a central calcium channel





**Fig. 2 Parallel serpentine adaptation candidates and the sources of parallel variants in *A. arenosa*.** **a** Intersection of candidates from each population pair (S1-N1 to S5-N5) demonstrating more genes repeatedly found as candidates across two, three, and four population pairs than expected by chance alone (all intersections were significant at  $p < 0.01$  (highlighted by asterisks), one-sided Fisher’s exact test); note: the colour intensity of the bars represents the  $p$  value significance of the intersections. **b** Gene ontology (GO) enrichment of the candidates (across all population pairs); GO categories: biological process (BP) and molecular function (MF); for complete list of GO terms, see Supplementary Data 4. **c** Overlap between parallel differentiation candidates and latent factor mixed model (LFMM) candidates resulting in 61 serpentine adaptation candidate genes. **d** Proportions of serpentine adaptation candidates originating from de novo mutations or being of shared origin out of the total of 29 cases of non-neutral parallelism as inferred by the Distinguishing among Modes of Convergence approach (DMC; see text for details). **e** Two examples of parallel candidate loci, illustrating nucleotide divergence and maximum composite log-likelihood (MCL) estimation of the source of the selected alleles in these particular loci inferred in DMC. Allele frequency difference (AFD) for locus with independent de novo mutations (left) and with parallel recruitment of shared ancestral standing variation (right). Left y-axis: AFD between S and N populations. Dots: AFD values of individual SNPs; bright green circles: non-synonymous SNPs with  $AFD \geq 0.4$ ; lines (right y-axis): MCL difference between neutral versus parallel selection scenario following colour scheme in **d**; gene models are in blue.

that mediates plant-wide stress signalling and tolerance<sup>46</sup>, and potassium transporters *AKT5* and *KUP9*. Furthermore, when we compared our serpentine adaptation candidates ( $n = 61$ ) to candidate loci for parallel serpentine adaptation in *Arabidopsis lyrata* ( $n = 62$ ) from a previous study<sup>47</sup>, we found two loci in common (significant overlap;  $p < 0.007$ ), *KUP9* and *TPC1*, further supporting important roles of these two ion transporters in repeated adaptation to serpentine soil. In addition, when overlapping the candidate genes detected at least in one of our

five population pairs with serpentine *A. lyrata* study we revealed additional convergent loci involved in ion homeostasis, calcium, nickel, and potassium transmembrane transport (Supplementary Table 5), suggesting existence of ‘hotspot’ regions in *Arabidopsis* genome in response to serpentine stress. An additional candidate gene (*FPN2 = IREG2*) investigated in *Alyssum* (Brassicaceae; Sobczyk et al.<sup>48</sup>) has been found to be shared between three population pairs. Finally, when comparing to the only genomically investigated serpentine system outside Brassicaceae

(*Mimulus*, Phrymaceae; Selby<sup>49</sup>), there was only limited overlap in two loci with one of our population pair. On the other hand, similar functions were enriched altogether suggesting parallel adaptation through similar pathways in very divergent (~140 myr) species.

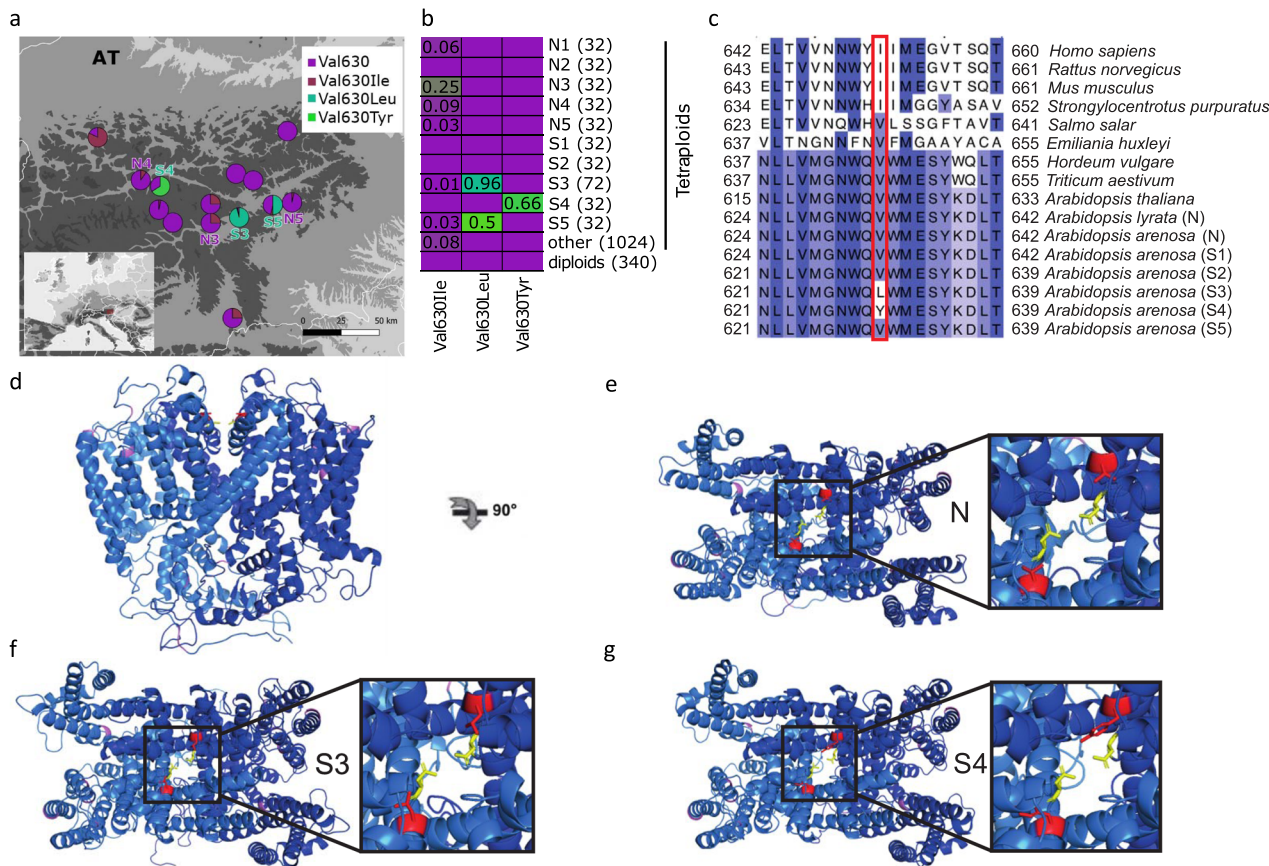
SNP data only present part of the picture and, despite linkage, do not capture structural variation. Specific TE families can be activated by abiotic stresses and possibly contribute to adaptation to challenging environments<sup>11,50–52</sup>. Thus, we also investigated divergence at TEs in population pairs 1 to 4 (relatively lower coverage of the N5 population did not permit this analysis) based on 21,690 TE variants called using the TEPID approach that is specifically designed for population TE variation studies<sup>53</sup>. Assuming linkage between each TE variant and surrounding SNPs (in the proximity of  $\pm 100$  bp), we applied a similar differentiation outlier window-based workflow as specified above and identified 92–115 TE-associated candidate genes per S–N contrasts (Supplementary Data 8). In comparison with the list of candidates based on SNPs (for the same four population pairs,  $n = 1,853$ ), we observed the overlap of 46 genes. The GO enrichment of TE-associated candidates from all four population pairs resulted in significant enrichment ( $p < 0.05$ ) of functions such as transmembrane transport, water channel activity and symporter activity (Supplementary Data 9). By overlapping the lists of TE-associated candidates across S–N pairs, we identified 13 parallel TE-associated differentiation candidates (Supplementary Fig. 9 and Supplementary Data 10; significant overlap,  $p < 0.05$ ). These loci included the plasma membrane protein *PIP2*, the putative apoplasmic peroxidase *PRX37*, and *RALF-LIKE 28*, which is involved in calcium signalling. This suggests a potential impact of TEs on serpentine adaptation and gives discrete candidates for future study.

**Sources of adaptive variation.** Next, we tested whether variants in each serpentine adaptation candidate have arisen by parallel de novo mutations or instead came from pre-existing variation shared across populations. To do so, we modelled allele frequency covariance around repeatedly selected sites for each locus and identified the most likely of the four possible evolutionary scenarios using a designated ‘Distinguishing among Modes of Convergence’ (DMC) approach<sup>54</sup>: (i) a null-model assuming no selection (neutral model), (ii) independent de novo mutations at the same locus, (iii) repeated sampling of shared ancestral variation, and (iv) sharing of adaptive variants via migration between adapted populations. For simplicity, we considered scenarios (iii) and (iv) jointly as ‘shared origin’ because both processes operate on alleles of a single mutational origin, in contrast to scenario (ii). To choose the best fitting scenario for each of the 61 candidate genes, we compared the maximum composite log-likelihoods (MCLs) between the four scenarios (see ‘Methods’; Supplementary Table 6). This analysis indicated that parallel selection exceeded the neutral model for 62 out of the total 84 candidate cases of parallelism (i.e. cases when two population pairs shared one of the 61 serpentine adaptation candidates). To focus only on well-justified candidates of adaptation *within* the serpentine populations, we excluded an additional 33 cases where the scenario of parallel selection with the highest MCL estimate in serpentine populations was not considerably higher (>10%) than this estimate in non-serpentine populations, which resulted in 29 candidate cases of serpentine adaptation parallelism. Shared origins dominated these results, representing 97% of the cases (28/29; Fig. 2d and Supplementary Data 7). The alternate non-neutral scenario, parallel de novo origin, was supported only for a single locus, *TWO PORE CHANNEL 1* (*TPC1*) in one case (S3–N3 and S4–N4; Fig. 2e). Using a more permissive threshold for

identifying differentiation candidates (3% outliers, leading to a fivefold increase in parallel candidates) resulted in a similar DMC estimate of the proportion of the shared variation scenario (103/114 cases, i.e. 90%; Supplementary Fig. 10 and Supplementary Data 11), indicating that our inference of the dominant role of shared variation in genic parallelism is not dependent on a particular stringent outlier threshold. Finally, we applied a similar approach to parallel TE-associated differentiation candidates ( $n = 13$ ) assuming selection on TE variants left a footprint in surrounding SNP–allele frequency covariance. We found a single non-neutral candidate, *ATPUX7*, for which parallel selection on standing variation was inferred (Supplementary Data 10). In summary, by a combination of genome-wide scanning with a designated modelling approach, we find that a non-random fraction of loci is likely reused by selection on serpentine, sourcing almost exclusively from a pool of alleles shared across the variable autotetraploid populations. Note that our conservative approach, focussed on identifying regions of repeated excessive differentiation and significant soil-related allele frequency differences, is not designed to cover the entire range of adaptive loci. Further research is thus needed to comprehensively cover the complete landscape of adaptation in autotetraploid *A. arenosa*.

**Rapid recruitment of convergent de novo mutations at the calcium channel *TPC1*.** One advantage of the DMC approach is an objective model selection procedure. However, it does not give fine scale information about the distribution of sequence variation at particular alleles. Therefore, we further investigated candidate alleles of the *TPC1* gene, for which DMC results suggested the sweep of different de novo mutations in independent serpentine populations. Upon closer inspection of all short-read sequences complemented by Sanger sequencing of additional 40 individuals from the three serpentine populations, a remarkably specific selection signal emerged. We found two absolutely serpentine-specific, high-frequency, non-synonymous mutations only at residue 630, overlapping the region of the highest MCL estimate for the de novo scenario in DMC, and directly adjacent to the selectivity gate of the protein in structural homology models (Fig. 3 and Supplementary Table 7). Of the two, the polymorphism Val630Leu is nearly fixed in the S3 population (25 homozygous Leu630 individuals and four heterozygous Leu630/Val630 individuals out of 29 individuals) and is at a high frequency in the S5 population (one homozygous Leu630 individual and 17 heterozygous out of 20 individuals); the second convergent Val630Tyr mutation is at high frequency in the S4 population (four homozygous Tyr630 individuals and 18 heterozygous Tyr630/Val630 out of 25 individuals; Fig. 3a, b and Supplementary Fig. 11). Strikingly, Val630Tyr requires a three-nucleotide mutation covering the entire codon (GTA to TAT). Neither of these variants were found in any other *A. arenosa* population in a range-wide catalogue<sup>8</sup> encompassing 1724 *TPC1* alleles (including 368 alleles from the focal area of Eastern Alps; Fig. 3a) nor in the available short-read data of the other two *Arabidopsis* outcrossing species (224 *A. lyrata* and 178 *Arabidopsis halleri* alleles, respectively, Supplementary Fig. 12), indicating that both are private to serpentine populations. Altogether, the absolute lack of either *A. arenosa* serpentine-specific variant in non-serpentine sampling across the genus strongly supports the conclusions of the DMC modelling of their independent de novo mutation origin.

To investigate the potential functional impact of these high-frequency, convergent amino acid changes, we first generated an alignment for *TPC1* homologues across the plant and animal kingdoms. Residue 630 (634/633 in *Arabidopsis thaliana*/*A. lyrata*) is conserved as either a Val or Ile across kingdoms,



**Fig. 3 Serpentine-private, convergent de novo high-impact protein changes in the TWO PORE CHANNEL 1 (TPC1) locus.** **a** All populations with serpentine-specific variants and all other resequenced *A. arenosa* populations in the focal area of Eastern Alps, showing frequencies of amino acid substitutions at residue 630 as pie charts (map drawn by V. Konečná). **b** Population frequencies of substitutions in the residue 630 among 1,724 alleles from range-wide *A. arenosa* resequenced samples. Colours denote frequencies from ancestral non-serpentine (violet) to serpentine-specific alleles (green) and number in brackets denotes total *N* of alleles screened. **c** Cross-kingdom conservation of the site shown by multiple sequence alignment of surrounding exon, including consensus sequences (AF > 0.5) from all serpentine *A. arenosa* populations (in S5 population, the frequency of Val630Leu is 0.5). Residues are coloured according to the percentage that matches the consensus sequence from 100% (dark blue) to 0% (white), the position of the serpentine-specific high-frequency non-synonymous polymorphism is highlighted in red. **d–g** Structural homology models of *A. arenosa* TPC1 alleles. Dimeric subunits are coloured blue or marine. Non-synonymous variation that is not linked to serpentine soil is coloured deep purple. Residue 630 is coloured red and drawn as sticks. The adjacent residue, 627 (631 in *A. thaliana*), which has an experimentally demonstrated key role in selectivity control, is yellow and drawn as sticks. **d** Side view of the non-serpentine allele. **e** Top view of the non-serpentine allele with the detail of the pore opening depicted in the inset. **f** Top view of the Leu630 allele private for S3 and S5 populations. **g** Top view of the Tyr630 allele private for S4 population. Source data underlying Fig. 3a, b are provided as a Source data file.

except for the serpentine *Arabidopsis* populations (Fig. 3b, c and Supplementary Fig. 13). Val/Ile and Tyr are disparate amino acids in both size and chemical properties, so this substitution has high potential for a functional effect. Although the difference between Val/Ile and Leu is not as radical, we predicted that the physical difference between the side chains of Val/Ile versus Leu (the second terminal methyl group on the amino acid chain) may also have a functional effect by making new contacts in the tertiary structure. To test this, we performed structural homology modelling of all alleles found in *A. arenosa* (Fig. 3d–g), using two crystallographically determined structures as a template (PDB codes 5DQQ and 5E1)<sup>55,56</sup>. In the tertiary structure, residue 630 sits adjacent to the Asn residue (Asn627 in *A. arenosa*), which forms the pore’s constriction point and has been shown to control ion selectivity in *A. thaliana*<sup>55–58</sup>. In *A. thaliana*, this Asn627 residue, when substituted by site-directed mutagenesis to the human homologue state, can cause Na<sup>+</sup> non-selective *A. thaliana* TPC1 to adopt the Na<sup>+</sup> selectivity of human TPC1<sup>56</sup>. Depending on the rotameric conformation adopted, the Leu630 allele forms contacts with the selectivity-

determining Asn627 residue, while the non-serpentine Val630 does not (Fig. 3e).

Our modelling suggests that the Tyr630 residue is even more disruptive: the Tyr side chain can adopt one of the two broad conformations, either sticking into the channel, where it occludes the opening, or sticking away from the channel and directly into surrounding residues, which have been shown to be important for stabilising Asn627 in *A. thaliana*. Both of the conformations seen, shown in heterodimer form (Fig. 3g), are highly likely to disrupt the stability of Asn627, thereby modifying the selectivity of the channel. Finally, to determine likely rotameric conformations, we generated 100 models of the S4 homodimer (two Tyr630 alleles). In this case, the two residues were significantly less likely to both point into the channel (Supplementary Table 8), suggesting that this is difficult for the structure to accommodate. Because the S4 Tyr630 allele is predominantly in a heterozygous state with the Val630 allele in nature, we also generated 100 models of the S4 heterodimer (one Tyr630 and one Val630 allele). In the heterodimer, there was no significant difference between the occurrence of the two side-chain conformations (Supplementary



Table 8), suggesting that either disruptive conformation, sticking into the pore or affecting nearby functional residues, is possible. Taken together, these results suggest that even the presence of a single Tyr630 variant within a dimer will have a substantial impact on *TPC1* function and that the Tyr630 allele is likely to be dominant or partially dominant. This prediction is consistent with its predominant occurrence in a heterozygous state and a ~50% allele frequency in S4 (Supplementary Fig. 11). We also speculate that if homodimers of the Tyr allele are too disruptive to *TPC1* function this may result in a heterozygote advantage maintained by balancing selection. In conclusion, our results suggest that the serpentine-linked allelic variation at residue 630 impacts the selectivity of *TPC1*, with functional implications, and that independent convergent de novo mutations have been repeatedly selected upon during adaptation to serpentine soils. Dedicated, electrophysiological single vacuole conductance experiments are required to explore functional changes in detail at *TPC1*. However, the exceptionally suggestive convergent changes we discovered and modelled to structures at the pore selectivity gate force the speculation that they mediate change in the relative conductance of the divalent cations  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ , the highly skewed ratios of which stand as hallmarks of serpentine soils<sup>26</sup>.

## Discussion

Here we inferred the evolutionary sources of adaptive variation in autotetraploid populations by taking advantage of naturally replicated adaptation to toxic serpentine soil in wild *A. arenosa*. Using a designated statistical approach leveraging parallelism, we inferred that nearly all parallel serpentine adaptation candidates were sourced from a common pool of alleles that was shared across populations. However, for one exceptional candidate—a central calcium channel shown to mediate stress signalling<sup>46</sup>—we identified independent de novo mutations at the same otherwise highly conserved site with likely functional consequences. Our approach informs on natural sources of parallel adaptive variation in autotetraploid populations, which has not yet been investigated genome-wide in a natural polyploid system. Yet we refrain from direct comparison with diploid ancestors because diploid serpentine populations are not known in *A. arenosa*, leaving a space for further study investigating other species encompassing multiple ploidies facing the same environmental challenge.

The potential of polyploidy to enhance adaptation is a matter of ongoing debate that is mainly fuelled by theory-based controversies revolving around the efficiency of selection<sup>6,19,59,60</sup> and observations of the frequencies of WGD events in space and time<sup>5,22</sup>. In contrast, empirical population-level investigations that unravel evolutionary mechanisms operating in natural autopolyploids are very scarce. It has been shown that autopolyploids may rapidly react to new challenges by landscape genetic<sup>23</sup> and experimental studies<sup>24,61</sup>. Our transplant experiments coupled with demographic investigations support this and further demonstrate that such rapid adaptation may be repeated many times within a species, partially drawing on the same variants. Furthermore, a previous study in *A. arenosa* showed that the genome-wide proportion of non-synonymous polymorphisms fixed by directional selection was higher in tetraploids than in diploids, suggesting increased adaptive variation in natural autopolyploids<sup>8</sup>. However, it remains unclear whether such variation reflects increased input of novel mutations (as observed, e.g. in experimental yeast populations<sup>17</sup>) or sampling from increased standing variation (as predicted by theory<sup>5,6</sup>). Here we find nearly exclusive repeated sampling from a shared pool of variants. Such a dominant role of pre-existing variation is in line with the studies of parallel adaptation in diploid systems such as

*Littorina* snails<sup>62,63</sup>, stickleback fishes<sup>64,65</sup>, *Heliconius* butterflies<sup>66</sup>, *Sinosuthora webbiana* vinous-throated parrotbill birds<sup>39</sup>, *Ipomoea purpurea* morning glories<sup>67</sup>, *Apis cerana* Asian honeybees<sup>68</sup>, or *Coilia nasus* fishes<sup>69</sup>. Yet examples are lacking from autopolyploid systems, where a large pool of standing variation is expected by theory due to larger effective population size and polysomic masking of allelic variation<sup>5,70</sup>. Sharing alleles that have persisted in a specific genomic environment already for some time may be particularly beneficial under intense selection when rapid adaptive responses are needed<sup>67,71–73</sup>. In addition, standing genetic variation likely minimises negative pleiotropic effects of linked variants<sup>67,74,75</sup>. It should be noted, however, that our estimates may be biased upward for shared alleles by focussing only on cases of parallelism, which provided a testable framework for our inference of the sources of variants. Larger fractions of novel mutations may be represented among the non-parallel adaptive variation, which is, however, harder to identify.

In contrast to shared variants, empirical evidence for parallel de novo mutations within species is rare even in diploids<sup>73,76–78</sup> and we lack any example from polyploids. Theory suggests that such a scenario is unlikely for autopolyploids, as reduced efficacy of selection on a novel, initially low-frequency variants is predicted for most dominance states in autopolyploids<sup>6,18,19</sup>. On the other hand, beneficial alleles are introduced at increased rates in doubled genomes<sup>6,17</sup> and additional variation may accumulate due to polysomic masking<sup>6</sup>. Here we provide an example of parallel recruitment of two distinct de novo mutations with likely phenotypic effect in separate polyploid populations within one species, demonstrating that adaptive sourcing from novel polymorphisms is in fact feasible even in autopolyploids. Interestingly, high frequencies of homozygous individuals in one population demonstrates that such novel variants may approach fixation, in stark contrast to theory, which predicts incomplete sweeps of dominant mutations to be prevalent in autotetraploids<sup>6,19</sup>. On the other hand, the prevalence of heterozygotes in the other serpentine population together with results of structural modelling suggest (at least partial) dominance of the serpentine allele.<sup>6</sup>

Overall, our study demonstrates that rapid environmental adaptation may repeatedly occur in established autopolyploid populations. Footprints of selection at similar genomic positions mostly occur because of the repeated recruitment from a large pool of pre-existing variation, yet exceptionally also from recurrent de novo mutations. Thus, these results support the emerging view of autopolyploids as diverse evolutionary amalgamates, capable of flexible adaptation in response to environmental challenge.

## Methods

**Field sampling.** Serpentes occur in Central Europe as scattered edaphic ‘islands’ surrounded by open rocky habitats on other substrates in which autotetraploid *A. arenosa* frequently occur. In contrast, *A. arenosa* colonised only some serpentine sites in this area<sup>79,80</sup> indirectly suggesting that colonisation of serpentine sites by surrounding non-serpentine populations happened in parallel and was probably linked with local substrate adaptation. To test this hypothesis, we sampled all five serpentine (S) populations of *A. arenosa* known to date and complemented each by a proximal (<19 km distant) non-serpentine (N) population. All N populations grew in similar vegetation (rocky outcrops in open forests or grasslands) and soil type (siliceous to neutral rocks; Supplementary Table 1). Although we observed a considerable variation in the overall soil chemistry in our samples (Supplementary Fig. 14), the principal soil factors differentiating between S and N populations were always the same—higher Mg, Ni, and Co and lower Ca/Mg in S populations. Diploid serpentine *A. arenosa* is not known, even though serpentine barrens are frequent in some diploid-dominated areas, such as the Balkan peninsula. The sampled populations covered considerable elevational gradient (414–1750 m a.s.l.), but the differences in elevation within the pairs were small except for one pair where no nearby subalpine non-serpentine population exists (population pair S4–N4, difference 740 m).

We sampled eight individuals per every population for genomic analysis and confirmed their tetraploid level by flow cytometry. For each individual, we also sampled soil from very close proximity to the roots (~10–20 cm below ground),



except for N5 population for which genotyped data were already taken from the previous study<sup>8</sup>. There we collected an additional eight soil samples and use their average in the following environmental association analysis. For the transplant experiment, we also collected seeds (~20–30 maternal plants/population) from three population pairs (S1–N1, S2–N2, S3–N3) and bulks of soil ~80 l (sieved afterwards) from the natural sites occupied by these six populations.

**DNA extraction, library preparation, sequencing, raw data processing, and filtration.** We stored all samples for this study in RNAlater (R0901-500ML, SIGMA-ALDRICH CO LTD) to avoid genomic DNA degradation and we further prepared the leaf material as described in refs. 81–83. We extracted DNA as described in Supplementary Method 1. Genomic libraries for sequencing were prepared using the Illumina TRUSeq PCR-free library. Libraries were sequenced as 150 bp paired-end reads on a HiSeq 4000 (3 lanes in total) by Norwegian Sequencing Centre, University of Oslo.

We used trimmomatic-0.36<sup>84</sup> to remove adaptor sequences and low-quality base pairs (<15 PHRED quality score). Trimmed reads >100 bp were mapped to reference genome of North American *A. lyrata*<sup>85</sup> by bwa-0.7.15<sup>86</sup> (<https://rcc.uchicago.edu/docs/software/modules/bwa/midway2/0.7.15.html>) with default setting. Duplicated reads were identified by picard-2.8 (<https://github.com/broadinstitute/picard>) and discarded together with reads that showed low mapping quality (<25). Afterwards we used GATK v.3.7 to call and filter reliable variants and invariant sites according to best practices<sup>87</sup> (complete variant calling pipeline available at <https://github.com/vlkojly/Fastq-to-vcf>). Namely, we used the HaplotypeCaller module to call variants per individual using the ploidy = 4 option, which enables calling full tetraploid genotypes. Then we aggregated variants across all individuals by module GenotypeGVCFs. We selected only biallelic SNPs and removed those that matched the following criteria: Quality by Depth (QD) < 2.0, FisherStrand (FS) > 60.0, RMSMappingQuality (MQ) < 40.0, MappingQualityRankSumTest (MQRS) < -12.5, ReadPosRankSum < -8.0, StrandOddsRatio (SOR) > 3.0. We called invariant sites also with the GATK pipeline similarly to variants, and we removed sites where QUAL was <15. Both variants and invariants were masked for sites with average read depth (RD) > 2 times standard deviation as these sites were most likely located in duplicated regions and we also masked regions with excessive heterozygosity, representing likely paralogous mis-assembled regions, following Monahan et al.<sup>8</sup>. One individual per each S2 and N5 populations was excluded due to exceptionally bad data quality (low percentage of mapped reads and low RD, <10 on average), leaving us with a final data set of 78 individuals that were used in genomic analyses. This pre-filtered data set contained 110,358,565 sites (of which 11,744,200 were SNPs) with average depth of coverage 21× (Supplementary Data 1). This way of genotyping leads to allele frequency estimates that are well comparable with previous estimates<sup>88</sup> including a broad range-wide *A. arenosa* sampling<sup>8</sup> (the site frequency spectra (SFS) are presented in Supplementary Fig. 15). Our site frequency estimates, which were constructed by program *est-sfs*<sup>89</sup>, are likely not biased by the number of individuals as was demonstrated by consistent site frequency estimates when subsampling one more deeply sampled population to the most common number of 32 chromosomes (S3; when including additional nine individuals from Arnold et al.<sup>29</sup>; Supplementary Fig. 16).

**Reconstruction of population genetic structure.** We inferred the population genetic structure, diversity, and relationships among individuals from putatively neutral 4dG SNPs filtered for DP > 8 per individual and maximum fraction of filtered genotypes (MFFG) of 0.2, i.e. allowing max. 20% missing calls per site (1,042,793 SNPs with a total of 0.49% missing data; see Supplementary Tables 1–3 and Supplementary Data 1 for description of data sets and filtration criteria). We used several complementary approaches. First, we ran principal component analysis (PCA) on individual genotypes using *gPCA* function in *adegenet* v.2.1.1 replacing the missing values by average allele frequency for that locus. Second, we applied model-based clustering with accelerated variational inference in *fastStructure* v.1.0<sup>90</sup>. To remove the effect of linkage, we randomly selected one SNP per a 1 kb window, keeping 10 kb distance between the windows and, additionally, filtered for minimum minor allele frequency (MAF) = 0.05 resulting in a data set of 9,923 SNPs. As *fastStructure* does not handle the polyploid genotypes directly, we randomly subsampled two alleles per each tetraploid site using a custom script. This approach has been demonstrated to provide unbiased clustering in auto-tetraploid samples in general<sup>91</sup> and *Arabidopsis* in particular<sup>8</sup>. We ran *fastStructure* with 10 replicates under  $K = 5$  (corresponding to the number of population pairs) with default settings. Third, we inferred relationships among populations using allele frequency covariance graphs implemented in *TreeMix* v.1.13<sup>92</sup>. We used custom python3 scripts (available at [https://github.com/mbohutinska/TreeMix\\_input](https://github.com/mbohutinska/TreeMix_input)) to create the input files. We ran *TreeMix* analysis rooted with an outgroup population (tetraploid *A. arenosa* population ‘Hranovnica’ from the area of origin of the autotetraploid cytotype in Western Carpathians<sup>93</sup>). We repeated the analysis over the range of 0–6 migration edges to investigate the change in the explanatory power of the model when assuming migration event(s) and found that adding migration to the model did not lead to large improvement (Supplementary Fig. 17). We bootstrapped the scenario without migration (the topology did not change with adding the migrations) choosing bootstrap block size 1 kb (the same window size also for the divergence scan, see below) and 100 replicates and

summarised the results using *SumTrees.py* function in *DendroPy*<sup>78</sup>. Finally, we calculated nucleotide diversity ( $\pi$ ) and Tajima’s  $D$  for each population pair and pairwise differentiation ( $F_{ST}$ ) (Supplementary Table 9) for each population pair using custom python3 scripts (available at [https://github.com/mbohutinska/ScanTools\\_ProtEvol](https://github.com/mbohutinska/ScanTools_ProtEvol); see Supplementary Table 3 for the number of sites per each population). For nucleotide diversity calculation, we down-sampled each population to six individuals on a per-site basis to keep equal sample per each population while also keeping the maximum number of sites with zero missingness.

**Demographic inference.** We performed demographic analyses in *fastsimcoal* v.2.6<sup>94</sup> to specifically test for parallel origin of serpentine populations and to estimate divergence time between serpentine and proximal non-serpentine populations. We constructed unfolded multidimensional SFS from the variant and invariant 4dG sites (filtered in the same ways as above, Supplementary Table 2) using custom python scripts published in our earlier study (*FSC2input.py* at <https://github.com/pmonahan/ScanTools/>)<sup>8</sup>. We repolarized a subset of sites using genotyped individuals across closely related diploid *Arabidopsis* species to avoid erroneous inference of ancestral state based on a single reference *A. lyrata* individual following Monahan<sup>8</sup>.

First, we tested for parallel origin of serpentine populations using population quartets (two pairs of geographically proximal serpentine and non-serpentine populations) and iterated such pairs across all combinations of regions (10 pairwise combinations among the five regions in total). For each quartet, we created four-dimensional SFS and compared following the four evolutionary scenarios (Supplementary Fig. 2): (i) parallel origin of serpentine ecotype—sister position of serpentine and non-serpentine populations within the same region, (ii) parallel origin with migration—the same topology with additional gene flow between serpentine and the proximal non-serpentine population, (iii) single origin of serpentine ecotype—sister position of serpentine populations and of non-serpentine populations, respectively, and (iv) single origin with migration—the same topology with additional gene flow between serpentine and the proximal non-serpentine populations. For each scenario and population quartet, 50 *fastsimcoal* runs were performed. For each run, we allowed for 40 ECM optimisation cycles to estimate the parameters and 100,000 simulations in each step to estimate the expected SFS. We used wide range of initial parameters (effective population size, divergence times, migration rates; see the example \*.est and \*.tpl files provided for each model tested in the Supplementary Data 12) and assumed mutation rate of  $4.3 \times 10^{-8}$  inferred for *A. arenosa* previously<sup>32</sup>. Further, we extracted the best likelihood partition for each *fastsimcoal* run, calculated Akaike information criterion (AIC), and summarised the AIC values across the 50 *fastsimcoal* runs. The scenario with lowest median AIC values within each particular population quartet was preferred (Supplementary Fig. 2 and Supplementary Data 2).

Second, we estimated divergence time between S and N populations from each population pair (i.e. S1–N1, S2–N2, S3–N3, S4–N4, and S5–N5) based on two-dimensional SFS using the same *fastsimcoal* settings as above. We simulated according to models of two-population split, not assuming migration because the model with migration did not significantly increase the model fit across the quartets of populations (Supplementary Fig. 2 and Supplementary Data 2). To calculate 95% confidence intervals for parameter estimates (Table 1), we sampled with replacement the original SNP matrices to create 100 bootstrap replicates of the two-dimensional SFS per each of the five population pairs.

**Window-based scans for directional selection.** We leveraged the fivefold-replicated natural set-up to identify candidate genes that show repeated footprints of selection across multiple events of serpentine colonisation. First, we identified genes of excessive divergence for each pair of proximal serpentine (S)–non-serpentine (N) populations (five pairs in total). Reflecting hierarchical structure in the data, we avoided merging multiple populations into larger units for the estimation of  $F_{ST}$  and strictly worked in a pairwise design. We admit that population S3 occupies a somewhat separate position and its ancestral non-serpentine population might have thus remained unsampled (or got extinct)—we therefore used the spatially closest population N3 as the most representative paired population available in our sampling. We calculated pairwise  $F_{ST}$ <sup>95</sup> for non-overlapping 1 kbp windows along the genome with the minimum of 10 SNPs per window to exclude potential biases in  $F_{ST}$  estimation caused by low-informative windows<sup>96</sup>. We used the custom script ([https://github.com/mbohutinska/ScanTools\\_ProtEvol](https://github.com/mbohutinska/ScanTools_ProtEvol)) based on *ScanTools* pipeline that has been successfully applied in our previous analyses of autotetraploid *A. arenosa*<sup>8</sup>. The window size of 1 kbp was selected to properly account for the average genome-wide linkage disequilibrium (LD) decay of genotypic correlations (150–800 bp) previously estimated in autotetraploid *A. arenosa*<sup>40</sup>. We used windows of fixed length and thus with homogeneous position on genome across all population pairs (in contrast to windows defined by number of SNPs and thus varying in exact position and length) to facilitate comparisons of selection candidates across distinct population pairs. Our  $F_{ST}$  estimates are unlikely to be strongly affected by varying numbers of SNPs per window, as the correlation between  $F_{ST}$  and number of SNPs per window was very weak (Spearman’s rank correlation coefficient varied from 0.02 to 0.04 across population pairs; Supplementary Fig. 18). We identified the upper 99% quantile of all 1 kbp windows in the empirical distribution of  $F_{ST}$  metric per each population pair. Then we identified

initial lists of genes with excessive differentiation for each S–N population pair as genes overlapping with the 1% outlier windows using *A. lyrata* gene annotation<sup>97</sup>, where the gene includes 5' untranslated regions (UTRs), start codons, exons, introns, stop codons, and 3' UTRs. Then we refined this inclusive list and identified parallel differentiation candidates as genes identified as candidates in at least two population pairs. We tested whether such overlap is higher than a random number of overlapping items given the sample size using Fisher's exact test in Super-ExactTest R package<sup>98</sup>. Our  $F_{ST}$ -based detection of outlier windows was not largely biased towards regions with low recombination rate (based on the available *A. lyrata* recombination map<sup>99</sup>; Supplementary Fig. 19).

**Environmental association analysis.** To further refine the candidate list to genes associated with the discriminative serpentine soil parameters, we performed environmental association analysis using LFMMs—LFMM 2 (<https://bcm-uga.github.io/lfmm/>)<sup>41</sup>. We tested the association of allele frequencies at each SNP for each individual with associated soil concentration of the key elements differentiating serpentine and non-serpentine soils: Ca/Mg ratio and bioavailable soil concentrations of Co, Mg, and Ni. Only those elements were significant in one-way ANOVAs (Bonferroni corrected) testing differences in elemental soil concentration between S and N population, taking population pair as a random variable. We retained 1,783,055 SNPs without missing data and MAF > 0.05 as an input for the LFMM analysis. LFMM accounts for a discrete number of ancestral population groups as latent factors. We used five latent factors reflecting the number of population pairs. Due to hierarchical structure in the data (PCA based on ~1 M 4dg SNPs indicated five main components, yet the first three axes alone also explained considerable variation; Supplementary Fig. 20), we also performed the additional analysis assuming three latent factors. As such analysis had only a minor effect on the total number of serpentine adaptation candidates (reducing their number by only two), we further used a candidate list based on five latent factors that corresponds to the total number of population pairs, thus it is also better comparable with the parallel differentiation candidates. To identify SNPs significantly associated with soil variables, we transformed  $p$  values to false discovery rate (<0.05) based on  $q$  values using the  $q$ value R package v.2.20<sup>100</sup>. Finally, we annotated the candidate SNPs to genes, termed 'LFMM candidates' (at least one significantly associated SNP per candidate gene).

We made a final shortlist of serpentine adaptation candidates by overlapping the LFMM candidates, reflecting significant association with important soil elements, with the previously identified parallel differentiation candidates, mirroring regions of excessive differentiation repeatedly found across parallel population pairs. For visualisation purposes (Fig. 2e), we annotated SNPs in the serpentine adaptation candidates using SnpEff v. 4.3<sup>101</sup> following *A. lyrata* version 2 genome annotation<sup>97</sup>.

**TE variant calling and analysis.** TE variants (insertions or deletions) among sequenced individuals were identified and genotyped in population pairs 1–4 using TEPID v.0.8<sup>53</sup> following the approach described in Rogivue et al.<sup>50</sup> (relatively lower coverage of the N5 population did not permit this analysis in the last pair). We annotated TEs based on available *A. lyrata* TE reference<sup>102</sup>. TEPID is based on split and discordant read mapping information and employs read mapping quality, sequencing breakpoints, and local variation in sequencing coverage to call the absence of reference TEs as well as the presence of non-reference TE copies. This method is specifically suited for studies at the population level as it takes intra-population polymorphism into account to refine TE calls in focal samples by supporting reliable call of non-reference alleles under lower thresholds when found in other individuals of the population. We filtered the data set by excluding variants with MFFG > 0.2 and DP < 8, which resulted in 21,690 TE variants (13,542 deletions and 8,148 insertions as compared to the reference).

Assuming linkage between TE variants and nearby SNPs<sup>53</sup>, we calculated pairwise  $F_{ST}$ <sup>95</sup> using SNP frequencies (in the same way as specified above) for non-overlapping 1 kbp windows containing TE variant(s) for each population pair. The candidate windows for directional selection were identified as the upper 99% quantile of all windows containing a TE variant in the empirical distribution of  $F_{ST}$  metric per each population pair. Further, we identified candidate genes (TE-associated candidates) as those present up to +/- 2 kbp upstream and downstream from the candidate TE variant (assuming functional impact of TE variant until such distance, following Hollister et al.<sup>103</sup>). Finally, we identified parallel TE-associated differentiation candidates as those loci that appeared as candidates in at least two population pairs.

**GO enrichment analysis.** We inferred potential functional consequences of the candidate gene lists using GO enrichment tests within biological processes, molecular functions, and cellular components domains. We applied Fisher's exact test ( $p < 0.05$ ) with the 'elim' algorithm implemented in topGO v.2.42 R package<sup>104</sup>. We worked with *A. thaliana* orthologues of *A. lyrata* genes obtained using biomaRt<sup>105</sup> and *A. thaliana* was also used as the background gene universe in all gene set enrichment analyses. The used 'elim' algorithm traverses the GO hierarchy from the bottom to the top, discarding genes that have already been mapped to significant child terms while accounting for the total number of genes annotated in the GO term<sup>104,106</sup>.

**Modelling the sources of adaptive variation.** For each serpentine adaptation candidate ( $n = 61$  and  $n = 13$  identified using SNPs and TE-associated variants, respectively), we modelled whether it exhibits patterns of parallel selection that is beyond neutrality and if so whether the parallel selection operated on de novo mutations or rather called on pre-existing variation shared across populations. We used model-based likelihood approach that is specifically designed to identify loci involved in parallel evolution and to distinguish among their evolutionary sources (DMC<sup>54</sup>). Convergent is analogous to parallel in this case as the entire approach is designed for closely related populations.

We considered the following four evolutionary scenarios assuming distinct variation sources: (i) no selection (neutral model), (ii) independent de novo mutations at the same locus, (iii) repeated sampling of ancestral variation that was standing in the non-adapted populations prior the onset of selection, and (iv) transfer of adaptive variants via gene flow (migration) from another adapted population. We interpreted the last two scenarios together as a variation that is shared across populations as both operate on the same allele(s) that do not reflect independent mutations.

We estimated composite log-likelihoods for each gene and under each scenario using a broad range of realistic parameters taking into account demographic history of our populations inferred previously and following recommendations in Lee and Coop<sup>54</sup> (positions of selected sites, selection coefficients, migration rates, times for which allele was standing in the populations prior to onset of selection, and initial allele frequencies prior to selection; see Supplementary Table 6 for the summary of all parameters and their ranges). We chose to place selected sites at eight locations at equal distance (default value recommended by authors <https://github.com/kristinmlee/dmc>) from each other along the particular gene. Such a density (one site per ~500 bp on average, as the mean length of serpentine adaptation candidate is ~4,000 bp) is in fact well within the range of the LD decay of 150–800 bp, which was estimated in *A. arenosa*<sup>40</sup>. For the calculation of co-ancestry decay, we also considered a 25 kbp upstream and downstream region from each gene. To choose the best fitting scenario for each candidate, we first estimated the MCL over the parameters for each of the three parallel selection scenarios and a neutral scenario. We selected among the parallel selection models by choosing the model with the highest MCL, following the approach of ref. <sup>49</sup>. Further, we considered the case significantly non-neutral only if the MCL difference between the selected parallel model and the corresponding neutral model was higher than the maximum of the distribution of the differences from the simulated neutral data in *A. arenosa* inferred in Bohutínská et al.<sup>40</sup> (i.e. MCL difference > 21, a conservative estimate). Further, to focus only on divergence caused by selection in serpentine populations (i.e. eliminating divergence signals caused by selection in N populations), only cases of selection in the serpentine populations in which the scenario of parallel selection had a considerably higher MCL estimate (>10%) than selection in non-serpentine populations were taken into account.

To ensure that our inference on the relative importance of shared versus de novo variation is not biased by arbitrary outlier threshold selection, we re-analysed the SNP data set using a 3% outlier  $F_{ST}$  threshold for identifying differentiation candidates. We overlapped the resulting 1,179 parallel differentiation candidates with the LFMM candidates and subjected the resulting 246 serpentine adaptation candidates to DMC modelling in the same way as described above (Supplementary Data 11). For each of the 246 genes and parallel quartets of populations (in total 420 cases of parallelism) we again compared the four scenarios as described above (Supplementary Fig. 10 and Supplementary Data 11).

**Reciprocal transplant experiment.** To test for local substrate adaptation in three serpentine populations, we compared plant fitness in the native versus foreign soil in a reciprocal transplant experiment. We reciprocally transplanted plants of serpentine and non-serpentine origin from three population pairs (S1–N1, S2–N2, S3–N3) that served as representatives of independent colonisation in each broader geographic region (Bohemian Massif, lower Austria, and Eastern Alps, respectively). As the most stressful factor for *A. arenosa* populations growing on serpentine sites is the substrate (Arnold et al.<sup>29</sup> and Supplementary Fig. 3a, b), we isolated the soil effect by cultivating plants in similar climatic conditions in the greenhouse.

For each pair, we cultivated the plants in serpentine and non-serpentine soil originating from their original sites (i.e. S1 plant cultivated in S1 and N1 soil and vice versa) and tested for the interaction between the soil treatment and soil of origin in selected fitness indicators (germination and rosette diameter sizes). We germinated seeds from 12 maternal plants (each representing a seed family of a mixture of full- and half-sibs) from each population in Ppetri dishes filled by either type of soil (15 seeds/family/treatment). Seeds germinated in the growth chamber (Conviron) under conditions approximating spring season at the original sites: 12 h dark at 10 °C and 12 h light at 20 °C. We recorded the germination date as the appearance of cotyledon leaves for the period of 20 days after which there were no new seedlings emerging. We tested for the effect of substrate of origin (serpentine versus non-serpentine), soil treatment (serpentine versus non-serpentine), and their interaction on germination proportion using GLM with binomial errors. To account for lineage-specific differences between population pairs, which are uninformative for the overall assessment of the fitness response towards serpentine, we treated population pairs as a random variable.

Due to zero germination of N1 seeds in S1 soil, we measured differential growth response on plants that were germinated in the non-serpentine soils and were subjected to the differential soil treatment later, in a seedling stage. We chose 44–50 seedlings equally representing progeny of 11 maternal plants per each population (in total 284 seedlings), transferred each plant to a separate pot filled either with ~1 L of the original or the alternative paired soil (i.e. S1 soil for N1 population and vice versa). We randomly swapped the position of each pot twice a week and watered them with tap water when needed. We measured the rosette diameter and counted the number of leaves (which correlated with the rosette diameter,  $R = 0.85$ ,  $p < 0.001$ ) twice a week for five weeks until rosette growth reached a plateau (Supplementary Fig. 6). By that time, we observed zero mortality and only negligible flowering (1%, four of the 284 plants). We tested whether soil treatment (serpentine versus non-serpentine) with the interaction of soil of origin (serpentine versus non-serpentine) had a significant effect on rosette diameter sizes (the maximum rosette diameter sizes from the last tenth measurement as a dependent variable), using two-way ANOVA taking population pair (1–3) as a random factor.

**Elemental analysis of soil and leaf samples.** We quantified the soil elemental composition by inductively coupled plasma mass spectrometry (ICP-MS; PerkinElmer NexION 2000, University of Nottingham). We monitored 21 elements (Na, Mg, P, S, K, Ca, Ti, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo, Cd, and Pb) in the soil extract samples. Individual soil samples of genotyped individuals (80 samples in total) were dried at 60 °C. Soil samples were sieved afterwards. Samples were prepared according to a protocol summarised in Supplementary Method 2. We quantified the elemental soil and leaf composition of the elements in samples from reciprocal transplant experiment by ICP OES spectrometer INTEGRA 6000 (GBC, Dandenong Australia). We monitored three elements that were identified as key elements differentiating S and N soils of the natural populations (Ca, Mg, and Ni) and decomposed samples prior the analysis. For details, see Supplementary Method 3.

**Screening natural variation in the *TPC1* locus.** To screen a broader set of *TPC1* genotypes in the relevant serpentine populations S3–S5, we Sanger-sequenced additional individuals sampled at the original sites of the focal S3, S4, and S5 populations (11, 17, and 12 individuals, respectively) exhibiting non-synonymous variation in the *TPC1* locus. For the amplification of the exon around the candidate site, we used specifically designed primers (Supplementary Table 10). The mix for PCR contained 0.3 µL of forward and reverse primer each, 14.2 µL of ddH<sub>2</sub>O, 0.2 µL of MyTag DNA polymerase, and 4 µL of reaction buffer MyTag, and we added 1 µL (10 ng) of DNA. The PCR amplification was conducted in a thermocycler (Eppendorf Mastercycler Pro) under the following conditions: 1 min of denaturation at 95 °C, followed by 35 cycles: 20 s at 95 °C, 25 s at 60 °C, 45 s at 72 °C, and a final extension for 5 min at 72 °C. Amplification products of high purity were sequenced at 3130xl Genetic Analyser (DNA laboratory of Faculty of Science, Charles University, Prague).

Then we checked whether the candidate alleles, inferred as serpentine specific in our data set, are also absent in a published broad non-serpentine sampling among outcrossing *Arabidopsis* species<sup>8,33,107–112</sup>. We downloaded all the available short-read genomic sequences published with the referred studies, called variants using the same approach as described above, and checked the genotypes at the candidate site (residue 630 in *A. arenosa* and 633 in *A. lyrata* and *A. halleri*). In total, we screened 1724 alleles of *A. arenosa*, 178 alleles of *A. halleri*, and 224 alleles of *A. lyrata*.

To visually compare variation at the entire *TPC1* locus, we generated consensus sequences for the group of all five non-serpentine *A. arenosa* populations and for each separate serpentine population using the Variant Call Format (VCF) with all variants in the region of scaffold\_6:23,042,733–23,048,601 using bcftools (Supplementary Fig. 21). Sites were included in the consensus sequence if they had AF > 50%. As the original VCF contained only biallelic sites, an additional multiallelic VCF was also created using GATK and any variants with AF > 50% were manually added to the biallelic consensus sequence. The *A. lyrata* and *A. thaliana* non-serpentine sequences were assumed to match the corresponding reference for each species.

Finally, we screened the variation in the *TPC1* locus at deep phylogenetic scales. We generated multiple sequence alignments using Clustal-Omega<sup>113</sup> from the available mRNA sequences from GenBank (*Emiliana huxleyi*, *Hordeum vulgare*, *A. thaliana*, *Triticum aestivum*, *Salmo salar*, *Strongylocentrotus purpuratus*, *Homo sapiens*, *Rattus norvegicus*, and *Mus musculus*) that were complemented by the consensus *Arabidopsis* sequences described above. Alignments were manually refined and visualised in JalView<sup>114</sup>.

**Structural homology models.** Structural homology models of dimeric *TPC1* were generated with Modeller v. 9.24<sup>115</sup> using two *A. thaliana* crystal structures (5E1J and 5DQQ<sup>55,57</sup>) as templates. The final model was determined by its discrete optimised protein energy score.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A Reporting Summary for this Article is available as a Supplementary Information file. Sequence data generated in this study have been deposited in the GenBank SRA database as a BioProject PRJNA667586 (populations S1–S5, N1–N4). Additional sequence data used in this study are deposited in the GenBank SRA database within a BioProject PRJNA325082 (population N5). Source data are provided with this paper.

## Code availability

Newly developed scripts are available at GitHub [<https://github.com/vlkofly/Fastq-to-vcf>] and [[https://github.com/mbohutinska/ScanTools\\_ProcEvolve](https://github.com/mbohutinska/ScanTools_ProcEvolve)].

Received: 25 January 2021; Accepted: 21 July 2021;

Published online: 17 August 2021

## References

- Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
- Barrett, R. D. H. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
- Wood, T. E. et al. The frequency of polyploid speciation in vascular plants. *Proc. Natl Acad. Sci. USA* **106**, 13875–13879 (2009).
- Soltis, D. E., Visger, C. J. & Soltis, P. S. The polyploidy revolution then...and now: Stebbins revisited. *Am. J. Bot.* **101**, 1057–1078 (2014).
- Van De Peer, Y., Mizrahi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
- Otto, S. P. & Whitton, J. Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437 (2000).
- Otto, S. P. The evolutionary consequences of polyploidy. *Cell* **131**, 452–462 (2007).
- Monnahan, P. et al. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* **3**, 457 (2019).
- Van de Peer, Y., Ashman, T. L., Soltis, P. S. & Soltis, D. E. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**, 11–26 (2020).
- Bardil, A., Tayalé, A. & Parisod, C. Evolutionary dynamics of retrotransposons following autopolyploidy in the Buckler Mustard species complex. *Plant J.* **82**, 621–631 (2015).
- Baduel, P., Quadrona, L., Hunter, B., Bomblies, K. & Colot, V. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat. Commun.* **10**, 5818 (2019).
- Ramsey, J. Polyploidy and ecological adaptation in wild yarrow. *Proc. Natl Acad. Sci. USA* **108**, 7096–7101 (2011).
- Chao, D. et al. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* **341**, 658–659 (2013).
- Bomblies, K. When everything changes at once: finding a new normal after genome duplication. *Proc. R. Soc. B Biol. Sci.* **287**, 20202154 (2020).
- Soltis, P. S. & Soltis, D. E. The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl Acad. Sci. USA* **97**, 7051–7057 (2000).
- Haldane, J. B. S. *The Causes of Evolution* (Princeton University Press, 1932).
- Selmecki, A. M. et al. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–351 (2015).
- Gerstein, A. C. & Otto, S. P. Ploidy and the causes of genomic evolution. *J. Hered.* **100**, 571–581 (2009).
- Monnahan, P. & Brandvain, Y. The effect of autopolyploidy on population genetic signals of hard sweeps. *Biol. Lett.* **16**, 20190796 (2020).
- Yao, Y., Carretero-Paulet, L. & Van de Peer, Y. Using digital organisms to study the evolutionary consequences of whole genome duplication and polyploidy. *PLoS ONE* **14**, e0220257 (2019).
- Brochmann, C. et al. Polyploidy in arctic plants. *Biol. J. Linn. Soc.* **82**, 521–536 (2004).
- Rice, A. et al. The global biogeography of polyploid plants. *Nat. Ecol. Evol.* **3**, 265–273 (2019).
- Parisod, C. & Besnard, G. Glacial in situ survival in the Western Alps and polytypic autopolyploidy in *Biscutella laevigata* L. (Brassicaceae). *Mol. Ecol.* **16**, 2755–2767 (2007).
- Martin, S. L. & Husband, B. C. Adaptation of diploid and tetraploid *Chamerion angustifolium* to elevation but not local environment. *Evolution* **67**, 1780–1791 (2013).



25. Wei, N., Cronn, R., Liston, A. & Ashman, T. L. Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments. *N. Phytol.* **221**, 2286–2297 (2019).
26. O'Dell, R. E. & Rajakaruna, N. in *Serpentine: Evolution and Ecology in a Model System* (eds Harrison, S. & Rajakaruna, N.) 97–137 (University of California Press, 2011).
27. Yant, L. & Bomblies, K. Genomic studies of adaptive evolution in outcrossing Arabidopsis species. *Curr. Opin. Plant Biol.* **36**, 9–14 (2017).
28. Molina-Henao, Y. F. & Hopkins, R. Autopolyploid lineage shows climatic niche expansion but not divergence in *Arabidopsis arenosa*. *Am. J. Bot.* **106**, 61–70 (2019).
29. Arnold, B. J. et al. Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl Acad. Sci. USA* **113**, 8320–8325 (2016).
30. Baduel, P., Hunter, B., Yeola, S. & Bomblies, K. Genetic basis and evolution of rapid cycling in railway populations of tetraploid *Arabidopsis arenosa*. *PLoS Genet.* **14**, 1–26 (2018).
31. Baduel, P., Arnold, B., Weisman, C. M., Hunter, B. & Bomblies, K. Habitat-associated life history and stress-tolerance variation in *Arabidopsis arenosa*. *Plant Physiol.* **171**, 437–451 (2016).
32. Przedpelska, E. & Wierzbička, M. *Arabidopsis arenosa* (Brassicaceae) from a lead-zinc waste heap in southern Poland - a plant with high tolerance to heavy metals. *Plant Soil* **299**, 43–53 (2007).
33. Preite, V. et al. Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. *Philos. Trans. R. Soc. B* **374**, 20180243 (2019).
34. Brady, K. U., Kruckeberg, A. R. & Bradshaw, H. D. Jr Evolutionary ecology of plant adaptation to serpentine soils. *Annu. Rev. Ecol. Evol. Syst.* **36**, 243–266 (2005).
35. Kazakou, E., Dimitrakopoulos, P. G., Baker, A. J. M., Reeves, R. D. & Troumbis, A. Y. Hypotheses, mechanisms and trade-offs of tolerance and adaptation to serpentine soils: from species to ecosystem level. *Biol. Rev.* **83**, 495–508 (2008).
36. Konečná, V., Yant, L. & Kolář, F. The evolutionary genomics of serpentine adaptation. *Front. Plant Sci.* **11**, 574616 (2020).
37. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat. Rev. Genet.* **10**, 639–650 (2009).
38. Takuno, S. et al. Independent molecular basis of convergent highland adaptation in maize. *Genetics* **200**, 1297–1312 (2015).
39. Lai, Y. T. et al. Standing genetic variation as the predominant source for adaptation of a songbird. *Proc. Natl Acad. Sci. USA* **116**, 2152–2157 (2019).
40. Bohutinská, M. et al. Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *Proc. Natl Acad. Sci. USA* **118**, e2022713118 (2021).
41. Caye, K., Jumentier, B., Lepeule, J. & François, O. LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Mol. Biol. Evol.* **36**, 852–860 (2019).
42. Remans, T. et al. A central role for the nitrate transporter NRT2.1 in the integrated morphological and physiological responses of the root system to nitrogen limitation in *Arabidopsis*. *Plant Physiol.* **140**, 909–921 (2006).
43. Little, D. Y. et al. The putative high-affinity nitrate transporter NRT2.1 represses lateral root initiation in response to nutritional cues. *Proc. Natl Acad. Sci. USA* **102**, 13693–13698 (2005).
44. Liu, J. et al. Targeted degradation of the cyclin-dependent kinase inhibitor ICK4/KRP6 by RING-type E3 ligases is essential for mitotic cell cycle progression during *Arabidopsis* gametogenesis. *Plant Cell* **20**, 1538–1554 (2008).
45. Stone, S. L. et al. Functional analysis of the RING-type ubiquitin ligase family of *Arabidopsis*. *Plant Physiol.* **137**, 13–30 (2005).
46. Choi, W. G., Toyota, M., Kim, S. H., Hilleary, R. & Gilroy, S. Salt stress-induced Ca<sup>2+</sup> waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proc. Natl Acad. Sci. USA* **111**, 6497–6502 (2014).
47. Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T. & Nuzhdin, S. V. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* **42**, 260–263 (2010).
48. Sobczyk, M. K., Smith, J. A. C., Pollard, A. J. & Filatov, D. A. Evolution of nickel hyperaccumulation and serpentine adaptation in the *Alyssum serpyllifolium* species complex. *Heredity* **118**, 31–41 (2017).
49. Selby, J. P. *The Genetic Basis of Local Adaptation to Serpentine Soils in Mimulus guttatus*. Doctoral dissertation, Duke University (2014).
50. Rogivue, A. et al. Genome-wide variation in nucleotides and retrotransposons in alpine populations of *Arabis alpina* (Brassicaceae). *Mol. Ecol. Resour.* **19**, 773–787 (2019).
51. Wos, G., Choudhury, R. R., Kolář, F. & Parisod, C. Transcriptional activity of transposable elements along an elevational gradient in *Arabidopsis arenosa*. *Mob. DNA* **12**, 1–13 (2021).
52. Grandbastien, M.-A. et al. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **110**, 229–241 (2005).
53. Stuart, T. et al. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *Elife* **5**, 1–27 (2016).
54. Lee, K. M. & Coop, G. Distinguishing among modes of convergent adaptation using population genomic data. *Genetics* **207**, 1591–1619 (2017).
55. Kintzer, A. F. & Stroud, R. M. Structure, inhibition and regulation of two-pore channel TPC1 from *Arabidopsis thaliana*. *Nature* **531**, 258–262 (2016).
56. Guo, J., Zeng, W. & Jiang, Y. Tuning the ion selectivity of two-pore channels. *Proc. Natl Acad. Sci. USA* **114**, 1009–1014 (2017).
57. Guo, J. et al. Structure of the voltage-gated two-pore channel TPC1 from *Arabidopsis thaliana*. *Nature* **531**, 196–201 (2016).
58. Kintzer, A. F. et al. Structural basis for activation of voltage sensor domains in an ion channel TPC1. *Proc. Natl Acad. Sci. USA* **115**, 9095–9104 (2018).
59. Griswold, C. K. & Williamson, M. W. A two-locus model of selection in autotetraploids: chromosomal gametic disequilibrium and selection for an adaptive epistatic gene combination. *Heredity* **119**, 314–327 (2017).
60. Mostafaei, N. & Griswold, C. K. Two-locus local adaptation by additive or epistatic gene combinations in autotetraploids versus diploids. *J. Hered.* **110**, 866–879 (2019).
61. Burgess, K. S., Etterson, J. R. & Galloway, L. F. Artificial selection shifts flowering phenology and other correlated traits in an autotetraploid herb. *Heredity* **99**, 641–648 (2007).
62. Morales, H. E. et al. Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. *Sci. Adv.* **5**, eaav9963 (2019).
63. Ravinet, M. et al. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol. Ecol.* **25**, 287–305 (2016).
64. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
65. Colosimo, P. F. et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
66. Pardo-Diaz, C. et al. Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* **8**, e1002752 (2012).
67. Van Etten, M., Lee, K. M., Chang, S. M. & Baucom, R. S. Parallel and nonparallel genomic responses contribute to herbicide resistance in *Ipomoea purpurea*, a common agricultural weed. *PLoS Genet.* **16**, e1008593 (2020).
68. Ji, Y. et al. Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Sci. Adv.* **6**, eabd3590 (2020).
69. Zong, S.-B., Li, Y.-L. & Liu, J.-X. Genomic architecture of rapid parallel adaptation to fresh water in a wild fish. *Mol. Biol. Evol.* **38**, 1317–1329 (2020).
70. Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F. & Yant, L. The 'Polyploid Hop': shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* **6**, 1–19 (2018).
71. Oziolor, E. M. et al. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* **364**, 455–457 (2019).
72. Reid, N. M. et al. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* **354**, 1305–1308 (2016).
73. Kreiner, J. M. et al. Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. *Proc. Natl Acad. Sci. USA* **116**, 21076–21084 (2019).
74. Przeworski, M., Coop, G. & Wall, J. D. The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312 (2005).
75. Hermisson, J. & Pennings, P. S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
76. Chan, Y. F. et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a pitxl enhancer. *Science* **327**, 302–305 (2010).
77. Tishkoff, S. A. et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).
78. Xie, K. T. et al. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* **84**, 81–84 (2019).
79. Justin, C. Über bemerkenswerte vorkommen ausgewählter pflanzensippen auf serpentinstandorten Österreichs, Sloweniens sowie der Tschechischen Republik. *Linzer Biol. Beiträge* **25**, 1033–1091 (1993).
80. Punz, W., Aigner, B., Sieghardt, H., Justin, C. & Zechmeister, H. G. Serpentinophyten im Burgenland. *Verhandlungen Zool. Ges. Österreich* **147**, 83–92 (2010).
81. Bodenhausen, N., Horton, M. W. & Bergelson, J. Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS ONE* **8**, e56329 (2013).
82. Horton, M. W. et al. Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nat. Commun.* **5**, 1–7 (2014).
83. Qvit-Raz, N., Jurkevitch, E. & Belkin, S. Drop-size soda lakes: transient microbial habitats on a salt-secreting desert tree. *Genetics* **178**, 1615–1622 (2008).
84. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
85. Hu, T. T. et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476–481 (2011).
86. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

87. Mckenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
88. Hollister, J. D. et al. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet.* **8**, e1003093 (2012).
89. Keightley, P. D. & Jackson, B. C. Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics* **209**, 897–906 (2018).
90. Raj, A., Stephens, M. & Pritchard, J. K. FastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589 (2014).
91. Stift, M., Kolář, F. & Meirmans, P. G. Structure is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity* **123**, 429–441 (2019).
92. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e100296 (2012).
93. 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).
94. Excoffier, L. & Foll, M. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332–1334 (2011).
95. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
96. Beissinger, T. M., Rosa, G. J., Kaeppler, S. M., Gianola, D. & De Leon, N. Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genet. Sel. Evol.* **47**, 1–9 (2015).
97. Rawat, V. et al. Improving the annotation of *Arabidopsis lyrata* using RNA-Seq data. *PLoS ONE* **10**, 1–12 (2015).
98. Wang, M., Zhao, Y. & Zhang, B. Efficient test and visualization of multi-set intersections. *Sci. Rep.* **5**, 1–12 (2015).
99. Hämälä, T. & Savolainen, O. Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*. *Mol. Biol. Evol.* **36**, 2557–2571 (2019).
100. Storey, J., Bass, A., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. R package version 2.20.0 (2020).
101. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
102. Legrand, S. et al. Differential retention of transposable element-derived sequences in outcrossing *Arabidopsis* genomes. *Mob. DNA* **10**, 1–17 (2019).
103. Hollister, J. D. et al. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl Acad. Sci. USA* **108**, 2322–2327 (2011).
104. Alexa, A. Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology. R package version 2.44.0 (2021).
105. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184 (2009).
106. Grossmann, S., Bauer, S., Robinson, P. N. & Vingron, M. Improved detection of overrepresentation of Gene-Ontology annotations with parent-child analysis. *Bioinformatics* **23**, 3024–3031 (2007).
107. 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).
108. Hämälä, T., Mattila, T. M., Leinonen, P. H., Kuitinen, H. & Savolainen, O. Role of seed germination in adaptation and reproductive isolation in *Arabidopsis lyrata*. *Mol. Ecol.* **26**, 3484–3496 (2017).
109. Mattila, T. M., Tyrmi, J., Pyhäjärvi, T. & Savolainen, O. Genome-wide analysis of colonization history and concomitant selection in *Arabidopsis lyrata*. *Mol. Biol. Evol.* **34**, 2665–2677 (2017).
110. Guggisberg, A. et al. The genomic basis of adaptation to calcareous and siliceous soils in *Arabidopsis lyrata*. *Mol. Ecol.* **27**, 5088–5103 (2018).
111. Hämälä, T., Mattila, T. M. & Savolainen, O. Local adaptation and ecological differentiation under selection, migration, and drift in *Arabidopsis lyrata*. *Evolution* **72**, 1373–1386 (2018).
112. Marburger, S. et al. Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* **10**, 1–11 (2019).
113. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
114. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
115. Šali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins Struct. Funct. Bioinformatics* **23**, 318–326 (1995).

## Acknowledgements

This work was supported by the Czech Science Foundation (project 20-22783S to F.K.), Charles University (project Primus/SCI/35 to F.K. and Charles University Grant Agency grant No. 410120 to V.K.), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [grant number ERC-StG 679056 HOTSPOT to L.Y. and ERC-StG 850852 DOUBLE ADAPT to F.K.]. Additional support was provided by the long-term research development project No. RVO 67985939 of the Czech Academy of Sciences. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085. The authors thank Lenka Flašková, Gabriela Šrámková, Anna Krejčová, and Melliha Allen for help with laboratory work and result interpretation.

## Author contributions

F.K., L.Y., and V.K. conceived the study. V.K., A.B.-G., L.Y., F.K., and M.B. performed field collections. A.B.-G., P.F., and V.K. did laboratory work. V.K., S.B., J.V., R.R.C., and M.B. performed analyses. V.K. and D.P. performed experiments. V.K., L.Y., and F.K. wrote the manuscript with input from all authors. All authors approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-25256-5>.

**Correspondence** and requests for materials should be addressed to L.Y. or F.K.

**Peer review information** *Nature Communications* thanks John Willis and other anonymous reviewers for their contributions to the peer review of this work. Peer review reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021