Impact of whole-genome duplications on structural variant evolution in the plant genus *Cochlearia*

Tuomas Hämälä[1,*], Christopher Moore[1], Laura Cowan[1], Matthew Carlile[1], David Gopaulchan[2], Marie K. Brandrud[3], Siri Birkeland[3,4], Matthew Loose[1], Filip Kolář[5,6], Marcus A. Koch[7], and Levi Yant[1,*]

[1]School of Life Sciences, University of Nottingham, Nottingham, UK

[2]School of Biosciences, University of Nottingham, Nottingham, UK

[3]Natural History Museum, University of Oslo, Oslo, Norway

[4]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

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

[6]Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic

[7]Centre for Organismal Studies, University of Heidelberg, Heidelberg, Germany

*tuomas.hamala@gmail.com, *levi.yant@nottingham.ac.uk

Abstract

Polyploidy, the result of whole-genome duplication (WGD), is a major driver of eukaryote evolution. All angiosperms have undergone ancient WGDs, and today stable polyploids account for a substantial portion of both wild and domesticated plant species. Despite its centrality in evolution, WGD is a hugely disruptive mutation, and we still lack a clear understanding of its fitness consequences. Here, we study whether WGD results in greater diversity of genomic structural variants (SVs) and how this influences evolutionary dynamics in nature. Using a set of long-read sequenced samples from the plant genus *Cochlearia* (Brassicaceae), which contains diploids and a recent ploidy series up to octoploid, we show that masking of recessive mutations due to WGD has led to a substantial and progressive accumulation of genic SVs across four ploidy levels. Such SVs likely constitute a genetic load and thus reduce the adaptive potential of polyploid populations. However, this SV accumulation also provides a rich pool of standing genetic variation upon which selection may act in novel environments. By constructing a graph-based pangenome for *Cochlearia*, we identify SVs in hundreds of samples and study the genomic basis of environmental adaptation. We find putatively beneficial SVs involved in pathogen resistance, root development, and salt tolerance, many of which are unique to polyploids. Finally, we explore the adaptive landscapes of SVs and SNPs, identify geographical regions where SVs make novel contributions to adaptive variation, and predict that the role of SVs in environmental adaptation increases due to rapid climate change.

Keywords: autopolyploidy, structural variants, genetic load, pangenome, climate change

Introduction

Whole-genome duplication (WGD) is a dramatic mutation that directly challenges the stability of meiosis and DNA management (Bomblies, 2023; Yant and Bomblies, 2015). As such, WGDs are often fatal, but the resulting polyploids that survive these initial obstacles may ultimately thrive (Comai, 2005). Indeed, WGDs have likely contributed to the emergence of all major eukaryotic lineages (Van de Peer et al., 2017), with particular importance in the evolution of plants (Wood et al., 2009). WGDs also have a direct economic impact, as the majority of our most important crop species are polyploid (Salman-Minkov et al., 2016). Understanding how evolutionary dynamics are altered by WGDs is therefore a fundamental goal in evolutionary biology, with applications reaching into agriculture. However, much of the genomic work related to WGDs is conducted on allopolyploids (polyploids resulting from the joining of two lineages), in which the effects of WGDs are confounded by hybridisation. Polyploids resulting from within-species WGDs (autopolyploids), by contrast, provide feasible systems to directly assess how evolutionary processes are shaped by WGDs.

Autopolyploidy is typically characterised by random pairing of chromosomes in meiosis (in allopolyploids chromosome pairing typically happens within subgenomes), resulting in predictable changes in population genetic processes (Haldane, 1930; Wright, 1938). All else being equal, doubling of the genome increases the mutational input, number of recombination events per individual, and the effective population size, leading to an increase in genetic diversity and a decrease in effective linkage (Caballero, 1994; Otto and Gerstein, 2008). Dominance structure is also transformed by WGDs, leading to more efficient masking of recessive mutations (Haldane, 1932; Otto and Whitton, 2000). Thus, increased diversity combined with masking of deleterious mutations may initially raise the adaptive potential of nascent polyploids (Otto, 2007). In the long-term, however, the hidden deleterious mutations might prove difficult to purge and allelic masking not only increases genetic load but also reduces the efficacy of positive selection (Haldane, 1932; Hill, 1970; Otto and Whitton, 2000; Ronfort, 1999), resulting in negative fitness consequences for aging polyploids (Baduel et al., 2018). We can therefore expect both beneficial and detrimental effects arising from WGDs, with empirical support found for some of the theoretical predictions (Baduel et al., 2019; Fisher et al., 2018; Monnahan et al., 2019; Selmecki et al., 2015).

Despite decades of work on polyploid genetics, an aspect that has remained largely unexplored is genomic structural variants (SVs). SVs encompass variants that influence the presence, abundance, location, and/or the orientation of the nucleotide sequence, typically defined as being greater than 50 bp in length. Studies of diploid organisms have established that SVs generally cover much more of the genome than point mutations (Hufford et al., 2021; Liao et al., 2023; Liu et al., 2020; Qin et al., 2021), suggesting that they can have a major

impact on the adaptive potential of populations and species. Given their disruptive effects on chromosomal structure, newly emerged SVs tend to be strongly deleterious and thus reduce the fitness of the host (Hämälä et al., 2021; Zhou et al., 2019). Yet SVs have also been associated with adaptive phenotypes in multiple species (Hof et al., 2016; Küpper et al., 2015; Studer et al., 2011; Sutton et al., 2007; Todesco et al., 2020), demonstrating that individual SVs can have beneficial fitness effects. In polyploids, however, the trajectory of SV evolution remains poorly understood, with existing knowledge primarily coming from allopolyploid crop genomes (He et al., 2021; Lovell et al., 2021; Walkowiak et al., 2020). In turn, we are missing an assessment of SV diversity in wild autopolyploid systems, leaving unknown the impact of WGDs on SV evolution in natural contexts.

Here, we study the impact of WGDs on SV landscapes as a function of ploidy, with a focus on understanding the influence of SVs on the adaptive potential of wild autopolyploid populations and species. Given the increased mutational input, combined with more complicated recombination and DNA repair machinery (Bomblies, 2023), we may expect SV emergence to increase in polyploids compared to diploids. We therefore ask how SVs influence the genetic load of polyploid populations, but also explore whether SVs provide unique benefits to polyploids. By analysing hundreds of genomes from the plant genus *Cochlearia*, we find both negative and positive interactions between WGDs and SVs. Masking of recessive mutations has increased the accumulation of deleterious SVs in polyploids, likely reducing the adaptive potential of these populations. However, we also discover apparent benefits resulting from the accumulation of SVs, as many more ploidy-specific SVs contribute to local adaptation in polyploids than in diploids. Finally, we propose that range-edge populations can especially benefit from large-effect SVs, and that SV-mediated adaptation could become more prominent in the future due to rapid climate change. Overall, our results provide important insights into the evolutionary relationship between WGDs and SVs – an aspect that likely has a major impact on the adaptive potential of polyploid organisms.

Results

Genetic composition of the *Cochlearia* genus

To study the impact of WGDs on SVs evolution in wild species, we conducted extensive long- and short-read sequencing on the plant genus *Cochlearia* (Brassicaceae). As an evolutionarily dynamic genus, *Cochlearia* exhibits highly labile genome structure, with two base chromosome numbers (n = 6 and n = 7) and multiple ploidy levels (from diploids to dodecaploids) found among the species (Crane and Gairdner, 1923; Gill, 1971; Koch et al., 1998, 1996; Koch, 2012; Saunte, 1955). Here, we focus on populations from the diploid n = 6 species *C. pyrenaica*, *C. excelsa, C. aestuaria*, and *C. islandica*; diploid n = 7 species *C.*

groenlandica and *C. triactylites*; tetraploid *n* = 6 species *C. officinalis, C. micacea*, and *C. alpina*; hexaploid species *C. tatrae, C. bavarica*, and *C. polonica*; and octoploid species *C. anglica*. The tetraploids likely resulted from within-species WGDs (autopolyploids), as evidenced by widespread multivalent formation at meiosis (Bray et al., 2023), whereas the evolutionary history of the higher ploidies is more complex, potentially involving both auto- and allopolyploidisation events. See Koch (2012) and Wolf et al. (2021) for comprehensive discussions of the evolutionary history of the *Cochlearia* species.

In total, our dataset comprised 23 samples sequenced with Oxford Nanopore (ONT) or Pacific Biosciences (PacBio) long-read technologies and 351 samples sequenced with Illumina short-read technology. The individuals represent 76 populations, covering the primary range of *Cochlearia* throughout Europe (Fig. 1A), along with locations in Svalbard and North America (Dataset S1). We first used SNP data derived from short-read sequencing to examine patterns of genetic diversity and differentiation among the *Cochlearia* populations. Compared to the diploid populations, polyploids exhibited lower levels of genetic diversity (Fig. 1B) and more negative Tajima's *D* (Fig. 1C), potentially reflecting bottlenecks and subsequent expansions resulting from the recent establishment of these populations (Wolf et al., 2021). A principal component analysis (PCA) indicated genetic clustering both due to ploidy and geographical location: the first principal component (PC) primarily corresponded to a difference between diploids and polyploids, whereas the second PC revealed clustering among multiple locations (Fig. 1D). The geographical clustering was also evident in within-ploidy PCAs, while little separation was found based on species assignments (Fig. S1). We further discovered a signal of isolation-by-distance, with between population F_{ST} estimates increasing with geographical distance, especially among the diploids (Fig. 1E). However, by using redundancy analysis (RDA) to model the role of geography, climate, and ploidy in explaining differentiation among the populations, we found climatic conditions to be a better predictor of genetic differentiation than either geographical distance or ploidy (Fig. 1F).
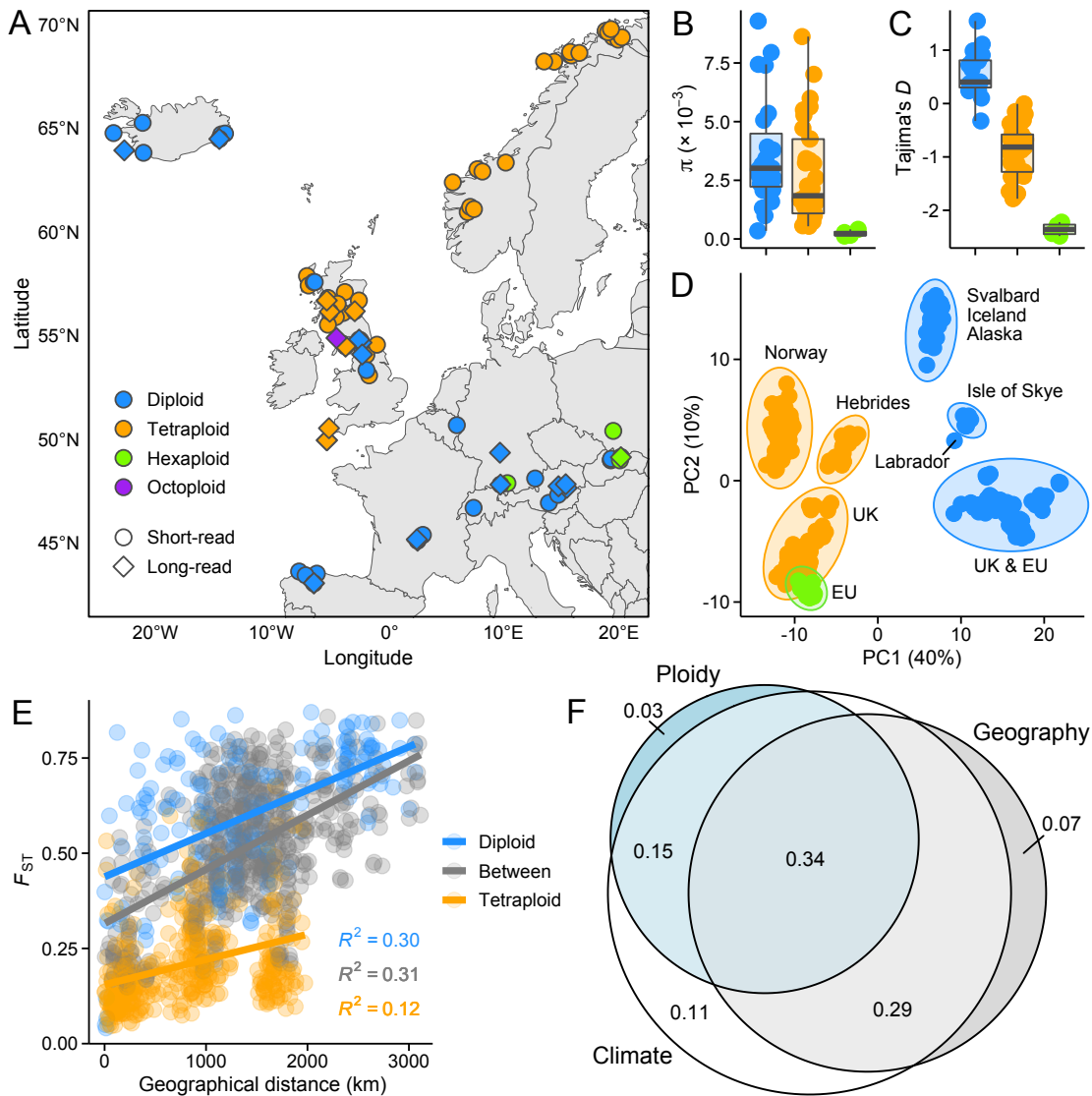
Figure 1. Locations and genetic variation among *Cochlearia* populations in this study. **A**: Map depicting European sampling locations. Shown are short-read sequenced populations (circles) and long-read sequenced individuals (diamonds). **B**: Pairwise nucleotide diversity (π) estimates for the short-read sequenced populations. **C**: Tajima's *D* estimates for the short-read sequenced populations. **D**: First two axes of a PCA. The proportion of variance explained by the PCs is shown in parentheses. **E**: Relationship between F_{ST} and geographical distance among diploid and tetraploid populations (between = diploid vs. tetraploid). **F**: The role of geography, climate, and ploidy in explaining genetic differentiation among these *Cochlearia* populations. Adjusted R^2 values from partial RDA models are shown in the circles. Note that the same colour legend applies to panels A–E.

SV identification and methylation assessment using long-read sequencing

Based on the analysis of SNP data, we found evidence that polyploidy influences the genetic composition of the *Cochlearia* genus. To explore whether WGDs also have an impact on SV landscapes, we performed long-read sequencing to identify SVs in 23 samples chosen to represent diverse lineages and ploidies. However, due to low sequencing depth, we excluded four diploids from our main analyses (Table S1), resulting in a set of 10 diploids, seven

6

tetraploids, one hexaploid, and one octoploid. After aligning reads against the chromosome build *C. excelsa* reference genome (Bray et al., 2023), we used Sniffles2 (Smolka et al., 2022) to identify SVs from the alignments. First, as Sniffles2 was developed primarily for diploid organisms, we used simulated data to confirm that it likely has a good power to detect SVs in our high-depth (mean depth = 68) autotetraploid samples (Fig. 2A and Table S2). We focused our analyses on insertions and deletions between 50 and 100,000 bp in size and filtered them for variant quality, missing data, and sequencing depth. After filtering, we retained 33,376 SVs in diploids and 47,755 in tetraploids. As both sequencing depth and read length can influence the power to detect SVs, we confirmed that tetraploids also carried more SVs after downsampling the alignments to equal number of base pairs (Fig. S2; *P* = 0.001, Wilcoxon rank-sum test). By comparing the SV sequence against our transposable element (TE) library, we found that in both diploids and tetraploids ~45% of the SVs contained TE sequence (Fig. S3), suggesting that many of the SVs are likely the result of TE mobilisation. To examine whether TE activity, and thus the potential of TEs to generate new SVs, differs between the ploidies, we quantified TE methylation using our ONT sequenced samples. Although we observed higher methylation levels in tetraploids than in diploids (Fig. S4), the pattern was not unique to TEs, and once we controlled for the genome-wide difference in methylation levels, we saw no evidence that TE families are systematically hyper- or hypomethylated in tetraploids (Fig. S5). Indeed, by estimating putative insertion times for TEs, we found no significant differences between the ploidies (Fig. S6), indicating that differential silencing of TEs is not a major factor shaping SV landscapes in diploids and tetraploids.

Masking progressively increases SV load in polyploids

To gain insight into the fitness effects of SVs, we estimated allele frequency spectra (AFS) for SVs found at genic (exons) and intergenic (> 5 kb away from genes) regions and compared these to nonsynonymous (0-fold) and synonymous (4-fold) SNPs. Overall, we found the AFS of diploids and tetraploids to be similar, except for an excess of rare exonic SVs in tetraploids (Fig. 2B). In the absence of mutation rate difference, such excess could either indicate stronger purifying selection in tetraploids or that recently emerged SVs are tolerated at functional regions because their effects are being masked by the additional allelic copies. To answer this question, we summarised the AFS using Tajima's *D* and compared two sets of estimates: ones estimated from the whole AFS and ones with singletons (i.e., variants with only single allele present) excluded. We found that the exclusion of singletons removed the excess of rare exonic SVs (Fig. 2C), supporting the idea that such SVs are being retained due to more efficient masking in tetraploids (as stronger purifying selection would skew the whole AFS towards rare variants). Therefore, our AFS-based analyses suggest that masking of recessive mutations allows SVs to accumulate in tetraploids that would have been purged by

7

purifying selection in diploids. We acknowledge, however, that these analyses rely on the correct identification of the SV genotypes, which can be challenging in polyploids, despite our validation (Fig. 2A). Thus, as an alternative approach, we examined the genomic locations of the SVs (regardless of their genotypes) and compared the observed numbers of SVs found overlapping different genomic features to random expectations. Given that this analysis does not require population-level data, we also included the hexaploid and octoploid samples. As expected, we discovered an overall deficit of exonic SVs and an excess of intergenic SVs (Fig. 2D). However, the deficit was greater in diploids than in polyploids, with the amount decreasing progressively with increasing ploidy (Fig. 2D). Furthermore, by examining the level of coding sequence conservation at genes affected by the SVs, we found a similar cline between all ploidies (Fig. 2E), indicating that SVs are being retained in genes under stronger selective constraint in polyploids than in diploids. Both results further support our conclusion that masking allows recessive SVs to accumulate in polyploids, likely progressively increasing the genetic load of higher ploidy populations.
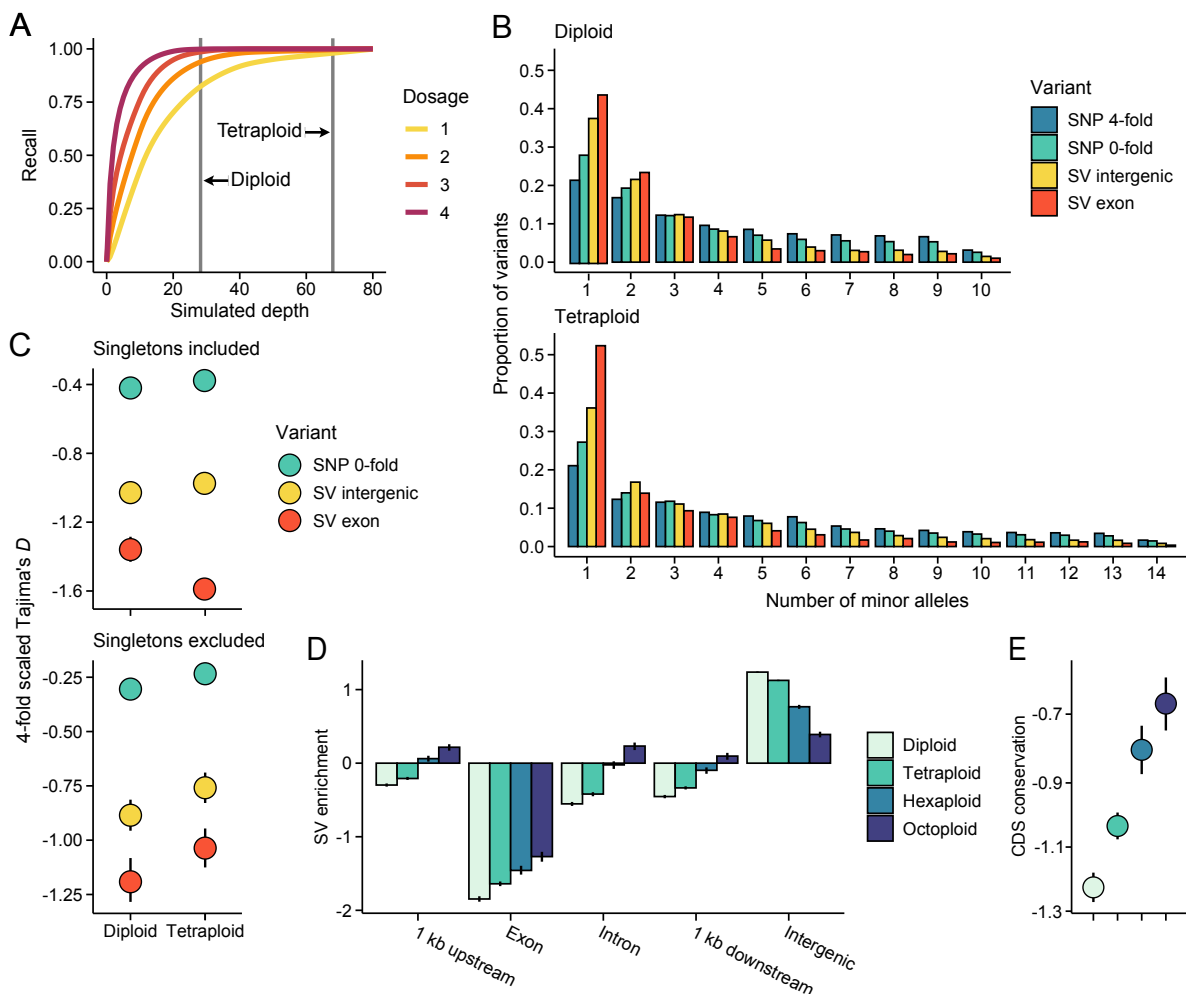

Figure 2. Fitness effects of structural variants (SVs). **A**: Validation of the SV caller. Shown are recall estimates for different levels of read depth over simulated insertions and deletions. For

each of the simulated scenarios, precision estimates were > 0.95. Grey horizontal lines mark the mean sequencing depth of our diploid and tetraploid samples. **B**: Folded allele frequency spectra (AFS) for different variant types in diploids and tetraploids. **C**: Tajima's *D* estimated from the whole AFS (singletons included) and AFS with singletons excluded. Values for nonsynonymous (0-fold) SNPs and SVs were scaled in relation to synonymous (4-fold) sites. **D**: SV enrichment across different genomic features in diploids and polyploids. Shown are $\log_2$-transformed ratios of observed to expected numbers of SVs. **E**: Coding sequence (CDS) conservation of genes affected by SVs (exon in panel D). Shown are median standardised GERP scores estimated among 30 eudicot species (lower values indicate weaker conservation). Note that the same colour legend applies to panels D and E. In panels C, D, and E, error bars indicate 95% bootstrap-based CIs.

Cochlearia pangenome reveals the influence of SVs in environmental adaptation

Although our analyses of the long-read data suggest that masking has increased the accumulation of deleterious SVs in polyploids, we might expect that some SVs provide selective benefits for the *Cochlearia* populations. Therefore, to examine the role of SVs in environmental adaptation, we constructed a graph-based pangenome for *Cochlearia* and used it to genotype 135,574 SVs in 351 short-read sequenced samples. After filtering the SVs for variant quality, missing data, and minor allele frequency (MAF), we used 11,393 (diploids), 18,313 (tetraploids), and 15,959 (all: diploids, tetraploids, and hexaploids) SVs to conduct genotype-environment association (GEA) analyses. Our analysis identified 49 SVs strongly associated with climatic variables in diploids, 159 in tetraploids, and 122 when considering all ploidy levels (Fig. 3A). Among the top outliers, we discovered SVs closely adjacent (< 1 kb) to genes *ACT7* (diploids), involved in root development and germination (Gilliland et al., 2003); *RIN2* (tetraploids), involved in pathogen resistance (Kawasaki et al., 2005); and *SIGA* (all), involved in control of plastid gene expression (Onda et al., 2008). Gene ontology (GO) terms related to root development and pathogen resistance (hypersensitive response) were also highly enriched among all outliers in diploids and tetraploids, respectively (Fig. 3B). By combining the three outlier sets, we discovered enrichment for additional GO terms, such as cellular response to salt (Fig. 3B). The candidate genes and biological processes associated with outlier SVs were largely independent of those identified with SNPs, as fewer than expected genes (odds ratio = 0.02; *P* < 2 × 10^{-16}, Fisher's exact test) and none of the GO terms were shared between the SV- and SNP-based outlier sets.

Given the greater accumulation of SVs in polyploids, we might assume that ploidy-specific SVs are more likely to contribute to adaptation in tetraploids than in diploids. Consistent with this expectation, we discovered that larger proportion of common SVs (MAF > 0.05) were ploidy-specific in tetraploids than in diploids (Fig. 3C; *P* < 2 × 10^{-16}, Fisher's exact test), including GEA outlier SVs (Fig. 3C; *P* = 0.01, Fisher's exact test). For example, two top outlier SVs in our GEA, closely adjacent to gene *RIN2*, were only polymorphic among the tetraploid populations (Fig. 3D). We further note that the proportion of ploidy-specific SVs in

tetraploids is likely underestimated, as our long-read data covers only part of the tetraploid range (the large cluster of diversity in Norway is missing, Fig. 1D), whereas among diploids there was a good correspondence between long- and short-read sequencing (Fig. 1A).
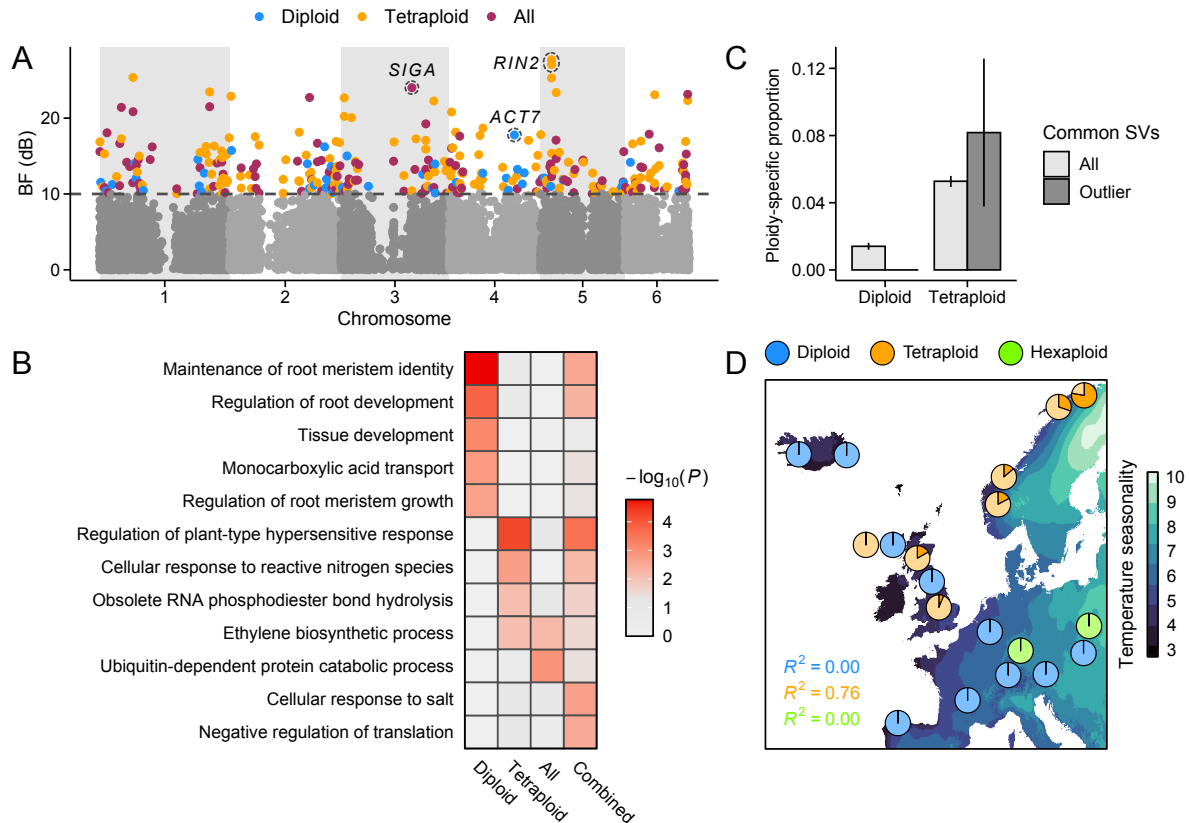

Figure 3. Footprints of environmental adaptation at SVs. **A**: Bayes Factor (BF) estimates from genotype-environment association (GEA) analyses conducted among diploid, tetraploid, and all (diploids, tetraploids, and hexaploids) populations. SVs with BF (dB) ≥ 10 were considered putatively adaptive. Top candidate gene (< 1 kb of the top outlier SV) from each analysis is highlighted. **B**: Results from a gene ontology (GO) enrichment analysis conducted on candidate genes. Included are GO terms with $P < 0.01$ from each of the three GEA sets as well as from a set with all outliers combined. **C**: The proportion of common SVs (MAF > 0.05) found only among diploid or tetraploid populations (ploidy-specific). Error bars show 95% bootstrap-based CIs. **D**: Example of a tetraploid-specific outlier SV, found 107 bp upstream of a gene *RIN2.* Pie charts show the frequencies of reference (lighter colour) and alternative (darker colour) alleles in closely adjacent populations.

The role of SVs in environmental adaptation increases due to climate change

Our results suggest that SVs make novel contributions to environmental adaptation in *Cochlearia*. To gain more insight into the geographical distribution of this adaptive variation, we predicted adaptive landscapes across the European range of the genus (Fig. S7). By leveraging the associations between genetic and environmental variables, adaptive landscapes can be used to project adaptive variation to unsampled locations (Fitzpatrick and Keller, 2015) and to model population vulnerability under climate change (Rellstab et al.,

10

2021). Here, however, we extend this approach to identify geographical regions where SVs potentially make unique contributions to adaptive variation by visualising the difference between SV- and SNP-based landscapes. Our analysis identified the northern (Norway and Iceland) and southern (Spain and France) edges of the *Cochlearia* range as locations where the adaptive landscapes are most strongly diverged (Fig. 4). Furthermore, by conducting a prediction using environmental variables projected for years 2061 – 2080, we discovered that the disparity between SV- and SNP-based adaptive landscapes may increase due to climate change (assuming that populations mainly track climate change through existing variation), especially in populations at the southern edge of the *Cochlearia* range (Fig. 4). We note that the same populations are not, according to our analysis, the ones most vulnerable to climate change, which are primarily found at the eastern edge of the *Cochlearia* range (Fig. S8).
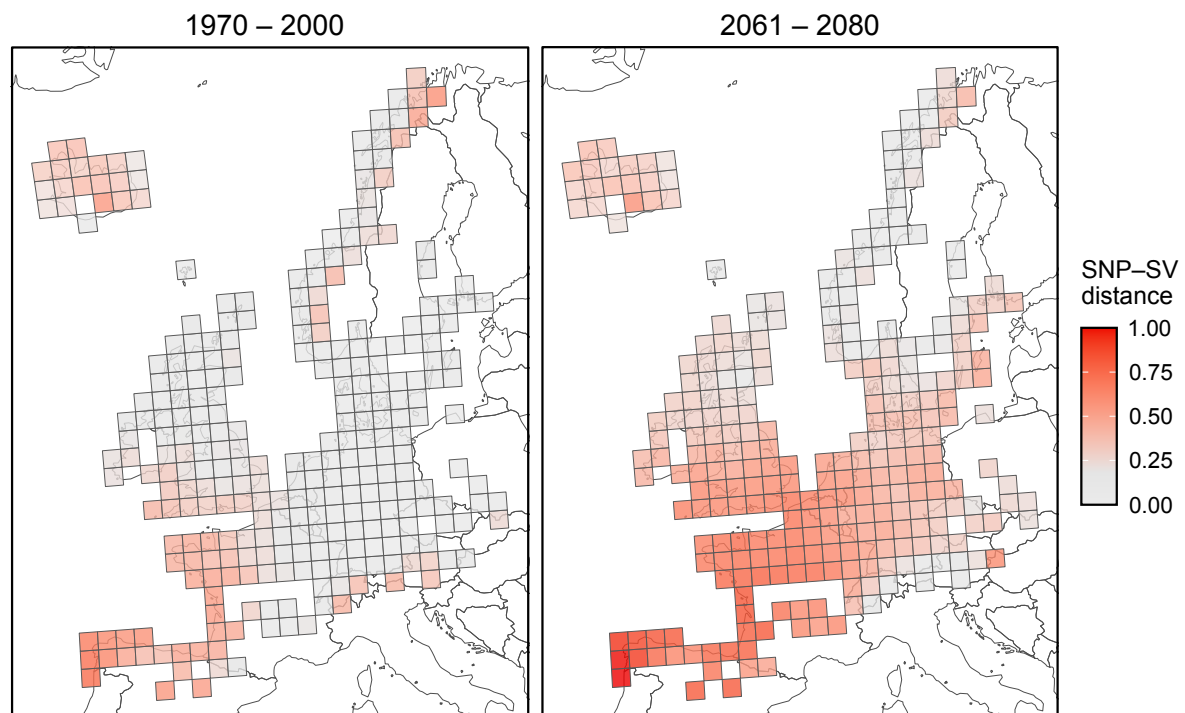

Figure 4. Adaptive distance between SNPs and SVs across the European range of *Cochlearia*. Colour scale indicates the level of unique contributions that SVs are predicted to make to adaptive variation (Euclidean distance between the adaptive indices). First panel shows past adaptation, based on 11 bioclimatic variables collected between 1970 and 2000, and the second panel shows a projection for years 2061 – 2080.

Discussion

How WGDs influence adaptive evolution is a long-standing question in evolutionary biology. Based on both theoretical (Haldane, 1930; Hill, 1970; Otto and Whitton, 2000; Wright, 1938) and empirical (Baduel et al., 2019; Fisher et al., 2018; Monnahan et al., 2019; Selmecki et al., 2015) work, we can expect pervasive fitness consequences arising from WGDs. However, the evolutionary relationship between WGDs and SVs is poorly understood, partly because SV identification has been challenging using short-read sequencing technologies (Mahmoud et al., 2019).

Here, we used long-read sequencing and pangenomics to study the impact of WGDs on SV landscapes in the plant genus *Cochlearia*. We discovered a substantial accumulation of genic SVs in polyploids that likely would have been purged by purifying selection in diploids. Theory suggests that such hidden load can have a major impact on the long-term fate of polyploid populations (Haldane, 1932; Otto and Whitton, 2000; Ronfort, 1999), contributing to eventual extinction or rediploidisation (Baduel et al., 2018). Although previous studies have discovered an excess of nonsynonymous SNPs (Monnahan et al., 2019) and TEs (Baduel et al., 2019) in recently founded autotetraploids, we may expect that SV accumulation has a particularly strong effect on the genetic load of polyploid populations. SVs are not only more deleterious than point mutations (Hämälä et al., 2021; Zhou et al., 2019) but also could be more frequently generated in polyploids due more complicated recombination and DNA repair machinery (Bomblies, 2023), as experimentally shown in yeast (Selmecki et al., 2015). Indeed, we previously showed that genes involved in DNA repair have evolved rapidly in the tetraploid *C. officinalis* since its divergence from the diploid *C. pyrenaica* (Bray et al., 2023), suggesting that WGD in *Cochlearia* has resulted in a shift in the (internal) selective environment due to extra challenges in DNA management.

Assuming that most deleterious mutations are partially recessive (Agrawal and Whitlock, 2011), SVs could have two major consequences for the fate of autopolyploid populations: 1) The point at which the genetic load of a newly formed autotetraploid population exceeds that of its diploid progenitor is reached faster with stronger selection coefficients (Otto and Whitton, 2000), meaning that SVs (compared to point mutations) could shorten the period of beneficial fitness effects arising from WGDs. 2) Once population reaches equilibrium, the fitness reduction due to deleterious mutations is roughly equal to the product of ploidy level and mutation rate (Otto and Whitton, 2000; Ronfort, 1999), indicating that higher rate of SV emergence in polyploids would increase the genetic load beyond that predicted from ploidy alone. Therefore, the accumulating SV load is likely an important factor limiting the adaptive potential of polyploid organisms, especially among the higher ploidies. We acknowledge, however, that in our hexaploid and octoploid samples the load inference could be influenced

by their mixed auto- and allopolyploid histories (Wolf et al., 2021), as subgenome dominance and lack of homoeologous recombination may increase the accumulation of deleterious mutations in allopolyploids compared to autopolyploids (Conover and Wendel, 2022). Nevertheless, the progressive accumulation of genic SVs across the four ploidy levels supports the idea that increasing ploidy leads to more efficient masking of recessive mutations, thus reducing the efficacy of purifying selection.

Despite the increased SV loads in polyploids, we also discovered apparent benefits resulting from the SV accumulation. Ploidy-specific SVs were more likely to contribute to local adaptation in tetraploids than in diploids, suggesting that the greater SV diversity in polyploids occasionally gets harnessed by positive selection. Furthermore, as interploidy gene flow is almost exclusively unidirectional from diploids to tetraploids (Baduel et al., 2018; Morgan et al., 2021), tetraploids are more likely to benefit from adaptive SVs originating in diploids than *vice versa*. Therefore, our results suggest that SVs contribute to the greater diversity of adaptive alleles available for polyploids (Bohutínská et al., 2023), compensating for some of the detrimental effects arising from the increased SV load. By analysing genes closely adjacent to the outlier SVs, we discovered enrichment of genes involved in different biological processes. The most prominent were related to root development in diploids, pathogen resistance in tetraploids, and salt tolerance in the combined dataset. The fast evolving resistance genes have been associated with SVs in multiple plant species (Bayer et al., 2020), highlighting the importance of structural variation in the ongoing arms race between plants and their pathogens (Anderson et al., 2010). Additionally, SVs have been linked to salinity adaptation in *Arabidopsis* (Busoms et al., 2018), making SV-mediated adaptation a potential contributor to the high salt tolerance exhibited by these *Cochlearia* species (Bray et al., 2023; Pegtel, 1999). The *Cochlearia* species (particularly the diploid *C. pyrenaica*) are also highly metallicolous (Nawaz et al., 2017), which could be associated with root development (Dijk et al., 2022) and thus SV-mediated adaptation. Importantly, these candidate genes and biological processes were not detected using SNPs, demonstrating that SVs need to be considered for a comprehensive view of adaptive processes.

To gain more insight into the unique roles of adaptive SVs, we searched for differences between SV- and SNP-based adaptive landscapes (Capblancq and Forester, 2021). The northern and southern range edges were highlighted as regions where the adaptive landscapes are most strongly diverged, potentially indicating novel contributions made by SVs. Indeed, we might expect to find more adaptive SVs in range-edge populations, as large-effect mutations tend to be favoured in populations that are far from their selective optima (Hämälä et al., 2020; Orr, 1998). By conducting a prediction using future climate projections, our modelling further suggests that SVs currently conferring adaptation to the southern environment could increase in frequency and spread northward due to climate change,

resulting in greater divergence between the SV- and SNP-based adaptive landscapes. Furthermore, this analysis expects that populations track shifting fitness optima through existing variation, but SV emergence could also increase due to climate change, as environmental stress is known to induce TE mobilisation (Hämälä et al., 2022; Wos et al., 2021), potentially providing more opportunities for SV-mediated adaptation.

Conclusions

By conducting extensive long- and short-read sequencing on the plant genus *Cochlearia*, we have gained novel insights into the evolutionary relationship between WGDs and SVs. We discovered a progressive accumulation of genic SVs across four ploidy levels, indicating increased SV loads in polyploids compared to diploids. Given the strongly negative fitness effects of SVs, we expect such SV loads to limit the long-term adaptability of polyploid populations and species. However, by constructing a graph-based pangenome for *Cochlearia*, we also found putative benefits arising from the SV accumulation, as ploidy-specific SVs were more likely to contribute to local adaptation in tetraploids than in diploids. Finally, our modelling work highlighted the unique roles of SVs in adaptation to current and future climates. Overall, our comprehensive analysis of SVs in *Cochlearia* sheds light on important and largely unexplored aspect of polyploid genomes, broadening the perspective of polyploid evolution as well as the evolution of structural variation in wild populations and species.

Material and methods

High molecular weight DNA isolation, Oxford Nanopore, and PacBio HiFi sequencing

To study the evolutionary role of SVs in *Cochlearia*, we collected samples from 23 individuals to be used in long-read sequencing. The set included 14 diploids, seven tetraploids, one hexaploid, and one octoploid (Table S1). Before starting DNA isolation, Carlson lysis buffer (100 mM Tris-HCl, 2% CTAB, 1.4 M NaCl, 20 mM EDTA, 1% PEG 8000) was mixed with 0.3 g PVPP and 50 ul B-mercaptoethanol and preheated to 65 °C. Leaf material from individual plants was ground into the heated solution and incubated for an hour at 65 °C. 20 ml chloroform was then added and mixed by inverting. The mixture was centrifuged at 3500 rpm (4 °C) for 15 minutes, the top layer of the lysate added to 1 × volume isopropanol, inverted to mix, and incubated at –80 °C for 15 minutes before being centrifuged at 3500 rpm (4°C) for 45 minutes. The supernatant was removed, the pellet air dried (sterile wipes were also used to dry the side walls of the tube) and resuspended in 500 ul nuclease-free water containing 2 ul of RNase A before being left to incubate at 37°C for 45 minutes. Samples were column purified with a Qiagen Blood and Cell Culture DNA Maxi Kit clean up. The DNA concentration was checked on a Qubit Fluorometer 2.0 (Invitrogen) using the Qubit dsDNA HS Assay kit.

Fragment sizes were assessed using the Genomic DNA Tapestation assay (Agilent). Removal of short DNA fragments and final purification to high molecular weight DNA was performed with the Circulomics Short Read Eliminator XS kit. After DNA isolation, two samples were used for Pacific Biosciences (PacBio) HiFi sequencing and 21 samples for Oxford Nanopore Technologies (ONT) sequencing.

ONT libraries were prepared using the Genomic DNA Ligation kit SQK-LSK109 following manufacturer's procedure. Libraries were loaded onto R9.4.1 PromethION Flow Cells and run on a PromethION Beta sequencer. Due to the rapid accumulation of blocked flow cell pores or due to apparent read length anomalies on some *Cochlearia* runs, flow cells used in the runs were treated with a nuclease flush to digest blocking DNA fragments before loading with fresh libraries according to the ONT Nuclease Flush protocol (version NFL_9076_v109_revD_08Oct2018). FAST5 sequences produced by PromethION sequencer were basecalled using the Guppy6 (https://community.nanoporetech.com) high accuracy basecalling model (dna_r9.4.1_450bps_hac.cfg) and the resulting FASTQ files quality filtered by the basecaller. PacBio sequencing was performed on a Sequel IIe at Novogene Europe (Cambridge, UK) in CCS mode.

Short-read library preparation and sequencing

We also used a set of 109 short-read sequenced *Cochlearia* individuals from Bray et al. (2023), which includes 39 diploids and 70 tetraploids. Although this sampling covers several locations across the Western and Northern Europe, it is mainly focused on the UK. To expand our sampling to more varied environments, we additionally collected 242 *Cochlearia* individuals across Europe and North America, leading to a final set of 351 individuals from 76 populations used for short-read sequencing (Dataset S1).

DNA was prepared using the commercially available DNeasy Plant Mini Kit from Qiagen (Qiagen: 69204). Illumina libraries were constructed from genomic DNA using Illumina DNA Prep library kit and IDT for Illumina DNA/RNA Unique Dual Index sets. Library preparation was performed using a Mosquito HV (SPT Labtech) liquid handling robot. The standard protocol timings and reagents were used, but with 1/10th reagent volumes at all steps. A total of 9 – 48 ng of DNA was used as library input and 5 cycles of PCR were used for the library amplification step. Individual libraries were pooled together and size selected using 0.65 $\times$ AMPure XP beads to minimise library fragments < 300 bp. Library pools were sequenced on a Novaseq 6000 using 2 $\times$ 150 bp paired-end reads at Novogene Europe (Cambridge, UK).

Transposable element annotation

We previously identified TEs from the *C. excelsa* reference genome (Bray et al., 2023). However, as the reference originated from a selfing diploid, we additionally assembled the genomes of one outcrossing diploid (*C. pyrenaica*) and one outcrossing tetraploid (*C. officinalis*) to expand our library of *Cochlearia* TEs. To do so, the individuals were sequenced using PacBio HiFi reads to an estimated depth of ~20 (diploid) and ~40 (tetraploid) $\times$ the haploid genome size. The reads were then *de novo* assembled using hifiasm (Cheng et al., 2021) and haplotigs purged from the primarily assemblies using purge_dups (Guan et al., 2020). The resulting assemblies had a total size of 359 (diploid) and 315 (tetraploid) mb, with contig N50 of 2.6 mb (diploid) and 630 kb (tetraploid). BUSCO (Simão et al., 2015) analysis indicated high completeness of the gene space, with 96% of the single-copy Brassicales genes found in both assemblies (Fig. S9). As with the *C. excelsa* reference genome, we annotated the assemblies using the EDTA pipeline (Ou et al., 2019), which includes multiple methods to comprehensively identify both retrotransposons and DNA transposons. To generate a single TE library across the three species, we used the cleanup_nested.pl script from EDTA to remove redundant (> 95% identical) consensus sequences from the combined library. We last conducted BLAST queries against a curated plant protein database from Swiss-Prot to remove likely gene sequences from the TE library. See Fig. S10 for outline of the annotated TE superfamilies.

Short-read processing and variant calling

Low quality reads and sequencing adapters were removed using Trimmomatic (Bolger et al., 2014) and the surviving reads aligned to the *C. excelsa* refence genome using bwa-mem2 (Vasimuddin et al., 2019). Although we aligned reads from multiple species against a single reference, alignment proportions were high for all samples (between 80% and 99%), likely reflecting the shallow divergence between the *Cochlearia* species (Wolf et al., 2021). We removed duplicated reads using Picard tools (https://broadinstitute.github.io/picard/) and identified SNPs using GATK4 (Van der Auwera and O'Connor, 2020). Filtering of the variant calls was based on the GATK's best practices protocol, and we included filters for mapping quality (MQ $\geq$ 40 and MQRankSum $\geq$ –12.5), variant confidence (QD $\geq$ 2), strand bias (FS < 60), read position bias (ReadPosRankSum $\geq$ –8), and genotype quality (GQ $\geq$ 15). We further removed SNPs with per-sample sequencing depth $\geq$ 1.6 $\times$ the mean depth to avoid issues caused by paralogous mapping.

Analyses of genetic variation

We used the short-read-based SNP calls to infer genetic relationships among our diploid and polyploid *Cochlearia* populations. First, we estimated pairwise nucleotide diversity (π) and Tajima's *D* for each population using both mono- and biallelic sites. We then conducted a principal components analysis (PCA) using linkage-pruned ($r^2 \leq 0.1$ within 100 SNPs, minor allele frequency [MAF] > 0.05) SNPs found at synonymous (4-fold) sites. Following Patterson et al., (2006), we estimated a covariance matrix representing the genetic relationships among each pair of individuals. For two individuals, i and j, covariance (C) was calculated as:

$$C_{ij} = \frac{1}{m} \sum_{s=1}^{m} \frac{(g_{is} - x_{ij}p_s)(g_{js} - x_{ij}p_s)}{x_{ij}p_s(1 - p_s)},$$

where m is the number of variable sites, g_{is} is the genotype of individual i in site s, x is the average ploidy level of the two individuals, and p is the alternate allele frequency. We then conducted PCA on the matrix using the R function prcomp and extracted the first two axes of the rotated data for plotting. We also estimated genetic differentiation between populations using F_{ST}. Here, we employed the F_{ST} measure by Hudson et al. (1992), as recommended by Bhatia et al. (2013).

To disentangle drivers of genetic differentiation among the *Cochlearia* populations, we tested for a pattern of isolation-by-distance and isolation-by-environment. Following Capblancq and Forester (2021), we performed redundancy analyses (RDA) using the R package vegan (Oksanen et al., 2022). We first estimated allele frequencies for the populations using linkage-pruned SNPs with MAF > 0.05 and $\leq$ 20% missing data. Missing population frequencies were imputed by randomly drawing them from a beta distribution with scale parameters calculated from the mean and variance of the non-missing values. We then extracted all 19 bioclimatic variables from WorldClim (Fick and Hijmans, 2017) and conducted forward model selection using RDA to identify variables explaining a significant proportion of genetic variation. Based on 1000 permutations, we kept 14 variables with *P* < 0.01. To transform the spatial structure of our data into a format usable in RDA, we conducted a principal coordinates analysis on a geographical distance matrix, retaining ten principal coordinates after forward model selection. Last, using partial RDA, we decomposed the effects of climate, geography, and ploidy in explaining genetic variation among the *Cochlearia* populations.

Validation of the SV caller

We aligned both ONT and PacBio long-reads against the *C. excelsa* reference genome using minimap2 (Li, 2018) and identified SVs from the alignments using Sniffles2 (Smolka et al., 2022). Our main analyses were based on 10 diploids (we excluded four diploids due to low

sequencing depth, Table S1) and seven tetraploids. As Sniffles2 expects the reads to originate from diploid organisms, we first used simulated data to evaluate its performance in autotetraploids. To estimate parameter values for the simulations, we used NanoPlot (De Coster et al., 2018) to calculate the mean and SD of read lengths across all samples. By aligning reads from the *C. excelsa* reference individual against the reference genome, we estimated an empirical error rate of 4% for the ONT reads. We note, however, this is likely a conservative estimate, as we assumed that all differences between the assembly and sequencing reads were due to sequencing errors, whereas such differences may also result from erroneous alignments and heterozygous SNPs (although heterozygous SNPs should be relatively rare in the selfing reference individual). We then randomly chose a single 10 kb region from the reference genome and introduced a 1 kb insertion or deletion into it. Using PBSIM2 (Ono et al., 2021), we generated simulated ONT reads from the modified and unmodified FASTA files, and combined them assuming average read proportions for simplex (1/4), duplex (2/4), triplex (3/4), and quadruplex (4/4) mutations. The total read depth was either 5, 10, 20, 40, or 80 × the simulated region. As Sniffles2 includes a preset for detecting rare SVs (i.e., SVs that are supported by only few reads), we assessed the performance of both the standard and rare SV modes. Based on the results (Table S2), we chose the rare SV mode for all subsequent Sniffles2 runs.

Long-read-based variant calling and genotyping

We called SV candidates individually for each sample and joined them into multi-sample VCF files using the population calling algorithm in Sniffles2 (Smolka et al., 2022). To reduce false positives caused by erroneous alignments, we included the reference (highly homozygous) individual in all multi-sample VCF files and excluded SVs that were called as heterozygotes or alternate homozygotes in the refence sample. We further focused our analyses on insertions and deletions (variant quality $\geq$ 20) between 50 bp and 100 kb, as methods based on read alignments are generally less accurate at detecting other types of SVs (e.g., tandem duplications and inversions) as well as very large SVs (Mahmoud et al., 2019). We also identified SNPs from the long-read alignments using Clair3 (Zheng et al., 2022). Running Clair3 on a model trained with R9.4 and Guppy6 reads (or alternatively with HiFi for the two PacBio sequenced individuals), we generated gVCF files for each individual and combined them into multi-sample VCF files using GATK4 (Van der Auwera and O'Connor, 2020).

Although our simulations suggest that Sniffles2 has good power to detect insertions and deletions in autotetraploids (Table S2), the genotype calls are incorrect due to diploid-specific genotyping model. Therefore, we collected allele count data (i.e., the number of reads supporting the reference and alternate alleles in each variant) for SVs and SNPs, and used

18

the R package Updog (Gerard et al., 2018) to estimate genotype likelihoods and probabilities. We required that variants used in Updog had $\leq 20\%$ missing data and were covered by ≥ 10 reads in diploids and ≥ 20 reads in tetraploids. To include genotype uncertainty directly into our analyses, we estimated the allelic dosage, or the expected genotype, from the genotype probabilities as $\mathrm{E}[G] = \sum_{g=0}^{4} g P(G = g)$, where G is the genotype. We then repeated this dosage estimation for the diploids to make the ploidy comparison equal.

Differential methylation analysis

To quantify DNA methylation, we used Deepsignal-plant (Ni et al., 2021) to detect 5-Methylcytosine (5mC) states of individual bases in the ONT data. To do so, we first used Tombo (Stoiber et al., 2017) to assign basecalls and genomic locations to raw signal reads. Then, using a model trained on *Arabidopsis thaliana* and *Oryza sativa* R9.4 reads, we had Deepsignal-plant estimate methylation frequencies in three sequence contexts, CG, CHG, and CHH (where H is A, T, or C), for each ONT sequenced individual with a mean depth ≥ 10. We last used cytosines covered by ≥ 6 reads to calculate methylation levels across genes and TEs.

We searched for differentially methylated TE families between diploids and tetraploids to assess TE activity. To do so, we used logistic regression and likelihood-ratio tests (LRTs) to identify associations between methylation levels and the ploidy. To control for the effects of population structure on methylation patterns, we conducted PCA on genome-wide methylation levels, used scree plots to define the number of main components in the data, and included the extracted PCs as cofactors in the models. *P*-values from the LRTs were transformed to false discovery rate-based *Q*-values (Storey, 2002) to account for multiple testing. We considered TE families with *Q* < 0.05 as differentially methylated between the ploidies.

Estimation of TE insertion times

We identified non-refence TE insertions from the ONT alignments using TELR, which has shown good performance in highly heterozygous, polyploid samples (Han et al., 2022). TELR combines Sniffles (we modified the pipeline to use Sniffles2 in the rare SV mode) and RepeatMasker (http://www.repeatmasker.org) to first identify TE insertions and then performs local assembly of the inserted sequences using wtdbg2 (Ruan and Li, 2020). After running TELR on each sample, we aligned the inserted sequences against the consensus TEs using MAFFT (Nakamura et al., 2018) and calculated sequence divergence (K) using the F81 substitution model (Felsenstein, 1981) implemented in the R package phangorn (Schliep, 2011). Last, we estimated insertion times as $T = \frac{K}{2}/\mu$, where μ is the per year substitution

rate, here assumed to be equal to the per generation mutation rate estimated for *Arabidopsis thaliana* (6.95×10^{-9} per base pair, Weng et al., 2019).

Fitness effects of SVs

We assessed the fitness effects of SVs by first analysing their allele frequency spectra (AFS). Using the estimates of allelic dosage, we calculated folded AFS for SVs found in genic (exons) and intergenic (> 5 kb away from genes) regions and compared them to synonymous (4-fold) and nonsynonymous (0-fold) SNPs. In the case of missing data (max 20%), we imputed the missing alleles by drawing them from a Bernoulli distribution. We further evaluated the selective removal of SVs by calculating the ratio of observed to expected numbers of SVs found overlapping different genomic features (1 kb up- and downstream of genes, exons, introns, intergenic). The expected numbers were estimated by defining the proportion of the genome that is covered by each feature (i.e., under random expectations, SVs would be distributed according to those proportions). We note, however, that these expectations could be affected by variation in mutations rates and insertion preference of TEs, but here were assumed that such biases are, on average, equal between the ploidies (this was confirmed for TEs, Figs. S4 – S6).

To determine the level of selective constraint on genes affected by the SVs, we estimated coding sequence conservation using GERP++ (Davydov et al., 2010). We first selected 29 eudicot species from the clade Superrosidae (Table S3), whose divergence times ranged from 20 million years (*Lobularia maritima*) to 123 million years (*Vitis vinifera*) in relation to *C. excelsa* (Hohmann et al., 2015). To identify sequence homologs, we conducted BLAST searches against species-specific protein databases, selecting only the best match with an *e*-value < 1×10^{-5} for each gene. We aligned the coding sequences using MAFFT (Nakamura et al., 2018), keeping only homolog sets with 15 or more species. We then chose 1000 random genes with no missing species, extracted synonymous sites based on the *C. excelsa* sequence, and estimated a maximum likelihood tree using the R package phangorn (Schliep, 2011). Based on the species tree and multiple alignments, we used GERP++ to estimate the rejected substitutions score for sites in the *C. excelsa* coding sequence, indicating the degree of nucleotide conservation relative to the synonymous substitution rate. Finally, we normalised the GERP scores using the range of possible values (as the range depends on the sample size of a particular site), calculated a median for each gene, and standardised the gene-specific estimates to a median of zero and MAD (median absolute deviation) of one.

Pangenome construction and SV genotyping

To more broadly study the evolutionary impact of SVs in *Cochlearia*, we genotyped our long-read based SVs in a set of 351 short-read sequenced individuals (Dataset S1). First, we

identified SVs from all 23 long-read sequenced individuals, including four diploids previously excluded due to low sequencing depth, one hexaploid, and one octoploid, to construct a pangenome graph to serve as a reference for the short-real alignments. We kept all insertions and deletions filling the following requirements: not identified in the reference individual, length between 50 bp and 100 kb, variant quality $\geq$ 20, supported by $\geq$ 4 reads, and the proportion of supporting reads $\geq$ 0.1 of all reads. We then used vg (Garrison et al., 2018) to construct a pan-genome graph based on the chromosome build *C. excelsa* reference genome and the resulting 135,574 SVs. The short-read data were aligned to the pangenome graph using vg map (Garrison et al., 2018) and SVs genotyped using vg call (Hickey et al., 2020). We last combined the individual-based SV calls into multi-sample VCF files using BCFtools (Li, 2011) and estimated genotype probabilities and allelic dosage using Updog (Gerard et al., 2018).

Genotype-environment association analyses

We tested for an association between genetic and environmental variables to identify loci involved in local adaptation. To do so, we characterised the growing environment of 70 European *Cochlearia* populations (we excluded three populations from North America and three populations from Svalbard, as they represented clear climatic outliers) using 11 bioclimatic variables (Fig. S11) identified with RDA (see "Analyses of genetic variation" for more details) and conducted genotype environment association (GEA) analyses using BayPass (Gautier, 2015). BayPass was ran on SV and SNP data compiled for three sets of samples: diploids, tetraploids, and all (diploids, tetraploids, and hexaploids combined). Note that BayPass works on population-specific allele frequencies (and not individual genotypes), making it suitable for polyploids. We required the variants to have MAF > 0.05 and $\leq$ 20% missing data to be included in the analyses. To control for the confounding effects of population structure, we included covariance matrices estimated using synonymous, linkage-pruned SNPs into all BayPass runs. Following the recommendation of Gautier (2015), we repeated each run ten times with different seed numbers (settings for the priors and the MCMC sampling were left default) and calculated a median Bayes Factor (BF) for the variants. Variants with median deciban (dB) BF $\geq$ 10 were considered putatively adaptive (corresponding to strong evidence for an association between genetic and environmental variables).

To better understand the functional importance of the outlier SVs, we conducted gene ontology (GO) enrichment analyses using the R package topGO (Alexa and Rahnenfuhrer, 2023). For each outlier SV and SNP, we included the closest gene within 1 kb and ran GO enrichment analyses using the weight01 algorithm and Fisher's exact test. We defined the background distribution of GO terms using only genes $\leq$ 1 kb of SVs and SNPs. Following the

recommendation of Alexa and Rahnenfuhrer (2023), we considered GO terms with $P < 0.01$ as significantly enriched among the candidate gene sets.

Adaptive landscapes and climate change

We used RDA (Oksanen et al., 2022) to explore the adaptive landscapes of SVs and SNPs. First, we estimated population allele frequencies for each outlier loci identified with BayPass (loci combined from diploid, tetraploid, and all runs) and imputed missing frequencies by drawing them from a beta distribution. We then used RDA to search for multivariate association between allele frequencies and the 11 bioclimatic variables used in the GEA. Following Capblancq and Forester (2021), we used the loadings of the first two RDA axes to predict the adaptive index for each environmental pixel across Europe (see Fig. S12 for a biplot of the loadings). We then acquired occurrence data for the *Cochlearia* genus from Global Biodiversity Information Facility (GBIF.org, 2023), cleaned up the records using both an automated tool (Zizka et al., 2019) and manual curation, and summarised the results into 100 × 100 km grid points comprising the European range of *Cochlearia*. We did this prediction using both outlier SNPs and SVs, and plotted the adaptive distance (Euclidean distance between the adaptive indices) between the two variant types to identify geographical regions where SVs potentially make unique contributions to the adaptive landscape. To explore possible effects of climate change on the adaptive landscapes, we used a Shared Socioeconomic Pathways (SSP) scenario SSP3-7.0 (Yukimoto et al., 2019) to model the increase in greenhouse gas concentrations by years 2061 – 2080.

Data availability

Sequence data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB66308. Scripts for conducting the analyses are available at https://github.com/thamala/polySV.

Author contributions

L.Y., T.H, and M.A.K. conceived the study. T.H. performed analyses. C.M., L.C., M.C., D.G., and M.L. performed laboratory work. M.A.K., M.K.B., S.B., and F.K. provided materials. T.H. wrote the manuscript with input from other authors.

Acknowledgements

References

Agrawal AF, Whitlock MC. 2011. Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* **187**:553–566. doi:10.1534/genetics.110.124560

Alexa A, Rahnenfuhrer J. 2023. topGO: enrichment analysis for Gene Ontology. *R Package Version 2520*.

Anderson JP, Gleason CA, Foley RC, Thrall PH, Burdon JB, Singh KB. 2010. Plants versus pathogens: an evolutionary arms race. *Funct Plant Biol FPB* **37**:499–512. doi:10.1071/FP09304

Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. 2018. The "Polyploid Hop": Shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front Ecol Evol* **6**:1–19. doi:10.3389/fevo.2018.00117

Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V. 2019. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat Commun* **10**:1–10. doi:10.1038/s41467-019-13730-0

Bayer PE, Golicz AA, Scheben A, Batley J, Edwards D. 2020. Plant pan-genomes are the new reference. *Nat Plants* **6**:914–920. doi:10.1038/s41477-020-0733-0

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting F_{ST} The impact of rare variants. *Genome Res* **23**:1514–21. doi:10.1101/gr.154831.113

Bohutínská M, Petříková E, Booker TR, Cobo CV, Vlček J, Šrámková G, Poštulková A, Hojka J, Marhold K, Yant L, Kolář F, Schmickl R. 2023. Mosaic haplotypes underlie repeated adaptation to whole genome duplication in *Arabidopsis lyrata* and *Arabidopsis arenosa*. doi:10.1101/2023.01.11.523565

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170

Bomblies K. 2023. Learning to tango with four (or more): the molecular basis of adaptation to polyploid meiosis. *Plant Reprod* **36**:107–124. doi:10.1007/s00497-022-00448-1

Bray SM, Hämälä T, Zhou M, Busoms S, Fischer S, Desjardins SD, Mandakova T, Moore C, Mathers TC, Cowan L, Monnahan P, Koch J, Wolf EM, Lysak MA, Kolář F, Higgins JD, Koch MA, Yant L. 2023. Kinetochore and ionomic adaptation to whole genome duplication. doi:10.1101/2023.09.27.559727

Busoms S, Paajanen P, Marburger S, Bray S, Huang XY, Poschenrieder C, Yant L, Salt DE. 2018. Fluctuating selection on migrant adaptive sodium transporter alleles in coastal *Arabidopsis thaliana*. *Proc Natl Acad Sci* **115**:E12443–E12452. doi:10.1073/pnas.1816964115

Caballero A. 1994. Developments in the prediction of effective population size. *Heredity* **73**:657–679. doi:10.1038/hdy.1994.174

Capblancq T, Forester BR. 2021. Redundancy analysis: A Swiss army knife for landscape genomics. *Methods Ecol Evol* **12**:2298–2309. doi:10.1111/2041-210X.13722

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**:170–175. doi:10.1038/s41592-020-01056-5

Comai L. 2005. The advantages and disadvantages of being polyploid. *Nat Rev Genet* **6**:836–846. doi:10.1038/nrg1711

Conover JL, Wendel JF. 2022. Deleterious mutations accumulate faster in allopolyploid than diploid cotton (*Gossypium*) and unequally between subgenomes. *Mol Biol Evol* **39**:msac024. doi:10.1093/molbev/msac024

Crane MB, Gairdner AE. 1923. Species-crosses in *Cochlearia*, with a preliminary account of their cytology. *J Genet* **13**:187–200.

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Comput Biol* **6**:e1001025. doi:10.1371/journal.pcbi.1001025

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**:2666–2669. doi:10.1093/bioinformatics/bty149

Dijk JR van, Kranchev M, Blust R, Cuypers A, Vissenberg K. 2022. Arabidopsis root growth and development under metal exposure presented in an adverse outcome pathway framework. *Plant Cell Environ* **45**:737. doi:10.1111/pce.14147

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**:368–376. doi:10.1007/BF01734359

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* **37**:4302–4315. doi:10.1002/joc.5086

Fisher KJ, Buskirk SW, Vignogna RC, Marad DA, Lang GI. 2018. Adaptive genome duplication affects patterns of molecular evolution in *Saccharomyces cerevisiae*. *PLOS Genet* **14**:e1007396. doi:10.1371/journal.pgen.1007396

Fitzpatrick MC, Keller SR. 2015. Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett* **18**:1–16. doi:10.1111/ele.12376

Garrison E, Sirén J, Novak AM, Hickey G, Eizenga JM, Dawson ET, Jones W, Garg S, Markello C, Lin MF, Paten B, Durbin R. 2018. Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat Biotechnol* **36**:875–879. doi:10.1038/nbt.4227

Gautier M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**:1555–1579. doi:10.1534/genetics.115.181453

GBIF.org. 2023. GBIF occurrence download. doi:10.15468/dl.z297uw

Gerard D, Luis Felipe Ventorim Ferrão, Garcia AAF, Stephens M. 2018. Genotyping polyploids from messy sequencing data. *Genetics* **210**:789–807. doi:10.1534/genetics.118.301468

Gill JJB. 1971. Cytogenetic studies in *Cochlearia* L. The chromosomal homogeneity within both the 2n = 12 diploids and the 2n = 14 diploids and the cytogenetic relationship between the two chromosome levels. *Ann Bot* **35**:947–956.

Gilliland LU, Pawloski LC, Kandasamy MK, Meagher RB. 2003. Arabidopsis actin gene ACT7 plays an essential role in germination and root growth. *Plant J* **33**:319–328. doi:10.1046/j.1365-313x.2003.01626.x

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**:2896–2898. doi:10.1093/bioinformatics/btaa025

Haldane JBS. 1932. The causes of evolution. New York: Longmans, Green and Co.

Haldane JBS. 1930. Theoretical genetics of autopolyploids. *J Genet* **22**:359–372. doi:10.1007/BF02984197

Hämälä T, Gorton AJ, Moeller DA, Tiffin P. 2020. Pleiotropy facilitates local adaptation to distant optima in common ragweed (*Ambrosia artemisiifolia*). *PLOS Genet* **16**:e1008707. doi:10.1371/journal.pgen.1008707

Hämälä T, Wafula EK, Guiltinan MJ, Ralph PE, DePamphilis CW, Tiffin P. 2021. Genomic structural variants constrain and facilitate adaptation in natural populations of *Theobroma cacao*, the chocolate tree. *Proc Natl Acad Sci* **118**:e2102914118. doi:10.1073/pnas.2102914118

Hämälä T, Weixuan N, Kuittinen H, Aryamanesh N, Savolainen O. 2022. Environmental response in gene expression and DNA methylation reveals factors influencing the adaptive potential of *Arabidopsis lyrata*. *eLife* **11**:e83115. doi:10.1101/2022.03.24.485516

24

Han S, Dias GB, Basting PJ, Viswanatha R, Perrimon N, Bergman CM. 2022. Local assembly of long reads enables phylogenomics of transposable elements in a polyploid cell line. *Nucleic Acids Res* gkac794. doi:10.1093/nar/gkac794

He Z, Ji R, Havlickova L, Wang L, Li Y, Lee HT, Song J, Koh C, Yang J, Zhang M, Parkin IAP, Wang X, Edwards D, King GJ, Zou J, Liu K, Snowdon RJ, Banga SS, Machackova I, Bancroft I. 2021. Genome structural evolution in *Brassica* crops. *Nat Plants* **7**:757–765. doi:10.1038/s41477-021-00928-8

Hickey G, Heller D, Monlong J, Sibbesen JA, Sirén J, Eizenga J, Dawson ET, Garrison E, Novak AM, Paten B. 2020. Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biol* **21**:35. doi:10.1186/s13059-020-1941-7

Hill RR. 1970. Selection in autotetraploids. *Theor Appl Genet* **41**:181–186. doi:10.1007/BF00277621

Hof AE van't, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**:102–105. doi:10.1038/nature17951

Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* **27**:2770–2784. doi:10.1105/tpc.15.00482

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**:583–589.

Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y, Coletta RD, Tittes S, Hudson AI, Marand AP, Wei S, Lu Z, Wang B, Tello-Ruiz MK, Piri RD, Wang N, Kim D won, Zeng Y, O'Connor CH, Li X, Gilbert AM, Baggs E, Krasileva KV, Portwood JL, Cannon EKS, Andorf CM, Manchanda N, Snodgrass SJ, Hufnagel DE, Jiang Q, Pedersen S, Syring ML, Kudrna DA, Llaca V, Fengler K, Schmitz RJ, Ross-Ibarra J, Yu J, Gent JI, Hirsch CN, Ware D, Dawe RK. 2021. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**:655–662.

Kawasaki T, Nam J, Boyes DC, Holt BF, Hubert DA, Wiig A, Dangl JL. 2005. A duplicated pair of Arabidopsis RING-finger E3 ligases contribute to the RPM1- and RPS2-mediated hypersensitive response. *Plant J* **44**:258–270. doi:10.1111/j.1365-313X.2005.02525.x

Koch M, Hurka H, Mummenhoff K. 1996. Chloroplast DNA restriction site variation and RAPD-analyses in *Cochlearia* (Brassicaceae): Biosystematics and speciation. *Nord J Bot* **16**:585–603. doi:10.1111/j.1756-1051.1996.tb00276.x

Koch M, Huthmann M, Hurka H. 1998. Isozymes, speciation and evolution in the polyploid complex *Cochlearia* L. (Brassicaceae). *Bot Acta* **111**:411–425. doi:10.1111/j.1438-8677.1998.tb00727.x

Koch MA. 2012. Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe *Cochlearieae* (Brassicaceae). *TAXON* **61**:76–92. doi:10.1002/tax.611006

Küpper C, Stocks M, Risse JE, Remedios ND, Farrell LL, McRae SB, Morgan TC, Karlionova N, Pinchuk P, Verkuil YI, Kitaysky AS, Wingfield JC, Piersma T, Zeng K, Slate J, Blaxter M, Lank DB, Burke T. 2015. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat Genet* **48**:79–83. doi:10.1038/ng.3443

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100. doi:10.1093/bioinformatics/bty191

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–93. doi:10.1093/bioinformatics/btr509

Liao W-W, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, Buonaiuto S, Chang XH, Cheng H, Chu J, Colonna V, Eizenga JM, Feng X, Fischer C, Fulton RS, Garg S, Groza C, Guarracino A, Harvey WT, Heumos S, Howe K, Jain M, Lu T-Y, Markello C, Martin FJ, Mitchell MW, Munson KM, Mwaniki MN,

Novak AM, Olsen HE, Pesout T, Porubsky D, Prins P, Sibbesen JA, Sirén J, Tomlinson C, Villani F, Vollger MR, Antonacci-Fulton LL, Baid G, Baker CA, Belyaeva A, Billis K, Carroll A, Chang P-C, Cody S, Cook DE, Cook-Deegan RM, Cornejo OE, Diekhans M, Ebert P, Fairley S, Fedrigo O, Felsenfeld AL, Formenti G, Frankish A, Gao Y, Garrison NA, Giron CG, Green RE, Haggerty L, Hoekzema K, Hourlier T, Ji HP, Kenny EE, Koenig BA, Kolesnikov A, Korbel JO, Kordosky J, Koren S, Lee H, Lewis AP, Magalhães H, Marco-Sola S, Marijon P, McCartney A, McDaniel J, Mountcastle J, Nattestad M, Nurk S, Olson ND, Popejoy AB, Puiu D, Rautiainen M, Regier AA, Rhie A, Sacco S, Sanders AD, Schneider VA, Schultz BI, Shafin K, Smith MW, Sofia HJ, Abou Tayoun AN, Thibaud-Nissen F, Tricomi FF, Wagner J, Walenz B, Wood JMD, Zimin AV, Bourque G, Chaisson MJP, Flicek P, Phillippy AM, Zook JM, Eichler EE, Haussler D, Wang T, Jarvis ED, Miga KH, Garrison E, Marschall T, Hall IM, Li H, Paten B. 2023. A draft human pangenome reference. *Nature* **617**:312–324. doi:10.1038/s41586-023-05896-x

Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou GA, Zhang H, Liu Z, Shi M, Huang X, Li Y, Zhang M, Wang Z, Zhu B, Han B, Liang C, Tian Z. 2020. Pan-genome of wild and cultivated soybeans. *Cell* **182**:162–176. doi:10.1016/j.cell.2020.05.023

Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S, Barry K, Bonos S, Boston L, Daum C, Deshpande S, Ewing A, Grabowski PP, Haque T, Harrison M, Jiang J, Kudrna D, Lipzen A, Pendergast TH, Plott C, Qi P, Saski CA, Shakirov EV, Sims D, Sharma M, Sharma R, Stewart A, Singan VR, Tang Y, Thibivillier S, Webber J, Weng X, Williams M, Wu GA, Yoshinaga Y, Zane M, Zhang L, Zhang J, Behrman KD, Boe AR, Fay PA, Fritschi FB, Jastrow JD, Lloyd-Reilley J, Martínez-Reyna JM, Matamala R, Mitchell RB, Rouquette FM, Ronald P, Saha M, Tobias CM, Udvardi M, Wing RA, Wu Y, Bartley LE, Casler M, Devos KM, Lowry DB, Rokhsar DS, Grimwood J, Juenger TE, Schmutz J. 2021. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* **590**:438–444. doi:10.1038/s41586-020-03127-1

Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. 2019. Structural variant calling: The long and the short of it. *Genome Biol* **20**:1–14. doi:10.1186/s13059-019-1828-7

Monnahan P, Kolář F, Baduel P, Sailer C, Koch J, Horvath R, Laenen B, Schmickl R, Paajanen P, Šrámková G, Bohutínská M, Arnold B, Weisman CM, Marhold K, Slotte T, Bomblies K, Yant L. 2019. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat Ecol Evol* **3**:457–468. doi:10.1038/s41559-019-0807-4

Morgan EJ, Čertner M, Lučanová M, Deniz U, Kubíková K, Venon A, Kovářík O, Lafon Placette C, Kolář F. 2021. Disentangling the components of triploid block and its fitness consequences in natural diploid–tetraploid contact zones of *Arabidopsis arenosa*. *New Phytol* **232**:1449–1462. doi:10.1111/nph.17357

Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**:2490–2492. doi:10.1093/bioinformatics/bty121

Nawaz I, Iqbal M, Bliek M, Schat H. 2017. Salt and heavy metal tolerance and expression levels of candidate tolerance genes among four extremophile *Cochlearia* species with contrasting habitat preferences. *Sci Total Environ* **584–585**:731–741. doi:10.1016/j.scitotenv.2017.01.111

Ni P, Huang N, Nie F, Zhang J, Zhang Z, Wu B, Bai L, Liu W, Xiao C-L, Luo F, Wang J. 2021. Genome-wide detection of cytosine methylations in plant from Nanopore data using deep learning. *Nat Commun* **12**:5976. doi:10.1038/s41467-021-26278-9

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2022. vegan: Community ecology package. *R Package Version 26-4*.

Onda Y, Yagi Y, Saito Y, Takenaka N, Toyoshima Y. 2008. Light induction of Arabidopsis SIG1 and SIG5 transcripts in mature leaves: differential roles of cryptochrome 1 and

cryptochrome 2 and dual function of SIG5 in the recognition of plastid promoters. *Plant J* **55**:968–978. doi:10.1111/j.1365-313X.2008.03567.x

Ono Y, Asai K, Hamada M. 2021. PBSIM2: a simulator for long-read sequencers with a novel generative model of quality scores. *Bioinformatics* **37**:589–595. doi:10.1093/bioinformatics/btaa835

Orr A. 1998. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* **52**:935–949. doi:10.1111/j.1558-5646.1998.tb01823.x

Otto SP. 2007. The evolutionary consequences of polyploidy. *Cell* **131**:452–462. doi:10.1016/j.cell.2007.10.022

Otto SP, Gerstein AC. 2008. The evolution of haploidy and diploidy. *Curr Biol* **18**:R1121–R1124. doi:10.1016/j.cub.2008.09.039

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet* **34**:401–437.

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, Jiang N, Hirsch CN, Hufford MB. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**:275. doi:10.1186/s13059-019-1905-y

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet* **2**:2074–2093. doi:10.1371/journal.pgen.0020190

Pegtel DM. 1999. Effect of ploidy level on fruit morphology, seed germination and juvenile growth in scurvy grass (*Cochlearia officinalis* L. sl, Brassicaceae). *Plant Species Biol* **14**:201–215.

Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, He Q, Ou S, Zhang H, Li Xuanzhao, Li Xiuxiu, Li Y, Liao Y, Gao Q, Tu B, Yuan H, Ma B, Wang Y, Qian Y, Fan S, Li W, Wang J, He M, Yin J, Li T, Jiang N, Chen X, Liang C, Li S. 2021. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**:3542-3558.e16. doi:10.1016/j.cell.2021.04.046

Rellstab C, Dauphin B, Exposito-Alonso M. 2021. Prospects and limitations of genomic offset in conservation management. *Evol Appl* **14**:1202–1212. doi:10.1111/eva.13205

Ronfort J. 1999. The mutation load under tetrasomic inheritance and its consequences for the evolution of the selfing rate in autotetraploid species. *Genet Res* **74**:31–42. doi:10.1017/S0016672399003845

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17**:155–158. doi:10.1038/s41592-019-0669-3

Salman-Minkov A, Sabath N, Mayrose I. 2016. Whole-genome duplication as a key factor in crop domestication. *Nat Plants* **2**:1–4. doi:10.1038/NPLANTS.2016.115

Saunte LH. 1955. Cyto-genetical studies in the *Cochlearia officinalis* complex. *Hereditas* **41**:499–515.

Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**:592–593. doi:10.1093/bioinformatics/btq706

Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, Sorenson AL, De S, Kishony R, Michor F, Dowell R, Pellman D. 2015. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**:349–351. doi:10.1038/nature14187

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212. doi:10.1093/bioinformatics/btv351

Smolka M, Paulin LF, Grochowski CM, Mahmoud M, Behera S, Gandhi M, Hong K, Pehlivan D, Scholz SW, Carvalho CMB, Proukakis C, Sedlazeck FJ. 2022. Comprehensive structural variant detection: from mosaic to population-level. doi:10.1101/2022.04.04.487055

Stoiber M, Quick J, Egan R, Lee JE, Celniker S, Neely RK, Loman N, Pennacchio LA, Brown J. 2017. De novo Identification of DNA modifications enabled by genome-guided Nanopore signal processing. doi:10.1101/094672

Storey JD. 2002. A direct approach to false discovery rates. *J R Stat Soc Ser B* **64**:479–498.

Studer A, Zhao Q, Ross-Ibarra J, Doebley J. 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet* **43**:1160–1163. doi:10.1038/ng.942

Sutton T, Baumann U, Hayes J, Collins NC, Shi BJ, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M, Langridge P. 2007. Boron-toxicity tolerance in barley arising from efflux transporter amplification. *Science* **318**:1446–1449. doi:10.1126/science.1146853

Todesco M, Owens G, Bercovich N, Légaré J-S, Soudi S, Burge D, Huang K, Ostevik K, Drummond E, Imerovski I, Lande K, Pascual M, Cheung W, Staton SE, Muños S, Nielsen R, Donovan L, Burke J, Yeaman S, Rieseberg L. 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**:602–607. doi:10.1101/790279

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nat Rev Genet* **18**:411–424. doi:10.1038/nrg.2017.26

Van der Auwera GA, O'Connor DB. 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra. O'Reilly Media.

Vasimuddin Md, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Presented at the 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). pp. 314–324. doi:10.1109/IPDPS.2019.00041

Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D, Klymiuk V, Byrns B, Gundlach H, Bandi V, Siri JN, Nilsen K, Aquino C, Himmelbach A, Copetti D, Ban T, Venturini L, Bevan M, Clavijo B, Koo D-H, Ens J, Wiebe K, N'Diaye A, Fritz AK, Gutwin C, Fiebig A, Fosker C, Fu BX, Accinelli GG, Gardner KA, Fradgley N, Gutierrez-Gonzalez J, Halstead-Nussloch G, Hatakeyama M, Koh CS, Deek J, Costamagna AC, Fobert P, Heavens D, Kanamori H, Kawaura K, Kobayashi F, Krasileva K, Kuo T, McKenzie N, Murata K, Nabeka Y, Paape T, Padmarasu S, Percival-Alwyn L, Kagale S, Scholz U, Sese J, Juliana P, Singh R, Shimizu-Inatsugi R, Swarbreck D, Cockram J, Budak H, Tameshige T, Tanaka T, Tsuji H, Wright J, Wu J, Steuernagel B, Small I, Cloutier S, Keeble-Gagnère G, Muehlbauer G, Tibbets J, Nasuda S, Melonek J, Hucl PJ, Sharpe AG, Clark M, Legg E, Bharti A, Langridge P, Hall A, Uauy C, Mascher M, Krattinger SG, Handa H, Shimizu KK, Distelfeld A, Chalmers K, Keller B, Mayer KFX, Poland J, Stein N, McCartney CA, Spannagl M, Wicker T, Pozniak CJ. 2020. Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**:277–283. doi:10.1038/s41586-020-2961-x

Weng ML, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG, Weigel D, Fenster CB. 2019. Fine-grained analysis of spontaneous mutation spectrum and frequency in *Arabidopsis thaliana*. *Genetics* **211**:703–714. doi:10.1534/genetics.118.301721

Wolf E, Gaquerel E, Scharmann M, Yant L, Koch MA. 2021. Evolutionary footprints of a cold relic in a rapidly warming world. *eLife* **10**:e71572. doi:10.7554/eLife.71572

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci* **106**:13875–13879. doi:10.1073/pnas.0811575106

Wos G, Choudhury RR, Kolář F, Parisod C. 2021. Transcriptional activity of transposable elements along an elevational gradient in *Arabidopsis arenosa*. *Mob DNA* **12**:1–12. doi:10.1186/s13100-021-00236-0

Wright S. 1938. The distribution of gene frequencies in populations of polyploids. *Proc Natl Acad Sci* **24**:372–377. doi:10.1073/pnas.24.9.372

Yant L, Bomblies K. 2015. Genome management and mismanagement – cell-level opportunities and challenges of whole-genome duplication. *Genes Dev* **29**:2405–2419. doi:10.1101/gad.271072.115

Yukimoto S, Koshiro T, Kawai H, Oshima N, Yoshida K, Urakawa S, Tsujino H, Deushi M, Tanaka T, Hosaka M, Yoshimura H, Shindo E, Mizuta R, Ishii M, Obata A, Adachi Y.

2019. MRI MRI-ESM2.0 model output prepared for CMIP6 AerChemMIP. doi:10.22033/ESGF/CMIP6.633

Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, Luo R. 2022. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci* **2**:797–803. doi:10.1038/s43588-022-00387-x

Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T, Cantu D, Gaut BS. 2019. The population genetics of structural variants in grapevine domestication. *Nat Plants* **5**:965–979. doi:10.1038/s41477-019-0507-8

Zizka A, Silvestro D, Andermann T, Azevedo J, Duarte Ritter C, Edler D, Farooq H, Herdean A, Ariza M, Scharn R, Svantesson S, Wengström N, Zizka V, Antonelli A. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol Evol* **10**:744–751. doi:10.1111/2041-210X.13152