Kinetochore and ionomic adaptation to whole genome duplication

Sian M. Bray[1], Tuomas Hämälä[1], Min Zhou[2], Silvia Busoms[3,4], Sina Fischer[1], Stuart D. Desjardins[5], Terezie Mandáková[6], Chris Moore[1], Thomas C. Mathers[7,8], Laura Cowan[1,9], Patrick Monnahan[3], Jordan Koch[3], Eva M. Wolf[10], Martin A. Lysak[6], Filip Kolar[11,12], James D. Higgins[5], Marcus A. Koch[10], and Levi Yant*[,1,11]

1. University of Nottingham, Nottingham NG7 2RD, United Kingdom
2. School of Life Sciences, Chongqing University, Chongqing 401331, China
3. Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park NR4 7UH, Norwich, UK
4. Department of Plant Physiology, Universitat Autònoma de Barcelona 08193 Barcelona, Spain
5. Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, UK
6. CEITEC–Central European Institute of Technology, Masaryk University, Brno 625 00, Czech Republic
7. Department of Crop Genetics, John Innes Centre, Norwich Research Park NR4 7UH, Norwich, UK
8. Current address: Tree of Life, Welcome Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK
9. Current address: Department of Plant Sciences, Cambridge University, Cambridge CB2 3EA, UK
10. Department of Biodiversity and Plant Systematics, Centre for Organismal Studies, Heidelberg University, 69120 Heidelberg, Germany
11. Department of Botany, Charles University, Benátská 2, CZ-12801 Prague, Czech Republic
12. Institute of Botany, The Czech Academy of Sciences, Zámek 1, CZ-252 43 Průhonice, Czech Republic

***Author for correspondence:** levi.yant@nottingham.ac.uk

Keywords: polyploidy; evolution; kinetochore; meiosis; ion homeostasis; population genomics

Abstract

Whole genome duplication (WGD) brings challenges to key processes like meiosis, but nevertheless is associated with diversification in all kingdoms. How is WGD tolerated, and what processes commonly evolve to stabilize the new polyploid lineage? Here we study this in *Cochlearia* spp., which have experienced multiple rounds of WGD in the last 300,000 years. We first generate a chromosome-scale genome and sequence 113 individuals from 33 diploid, tetraploid, hexaploid, and outgroup populations. We detect the clearest post-WGD selection signatures in functionally interacting kinetochore components and ion transporters. We structurally model these derived selected alleles, associating them with known WGD-relevant functional variation and compare these results to independent recent post-WGD selection in *Arabidopsis arenosa* and *Cardamine amara*. Some of the same biological processes evolve in all three WGDs, but specific genes recruited are flexible. This points to a polygenic basis for modifying systems that control the kinetochore, meiotic crossover number, DNA repair, ion homeostasis, and cell cycle. Given that DNA management (especially repair) is the most salient category with the strongest selection signal, we speculate that the generation rate of structural genomic variants may be altered by WGD in young polyploids, contributing to their occasionally spectacular adaptability observed across kingdoms.

45 Whole genome duplication (WGD, leading to polyploidy) is a dramatic mutation that disrupts
46 fundamental cellular processes. Yet, for those that can adapt to a transformed polyploid state,
47 WGD holds great promise[1-3]. Despite its importance to evolution, agriculture, and human health,
48 we do not yet know why some polyploids thrive, while others do not[1,4].
49 Immediately following WGD in autopolyploids (within-species WGD, not hybrid
50 allopolyploids), novel challenges arise. The most obvious concerns meiosis: the instant doubling
51 of homologs complicates chromosome pairing[5]. If a chromosome engages in crossing over with
52 more than one homolog, entanglements and breakage ensue at anaphase. WGD also disrupts
53 cellular equilibria, including ion homeostasis and cell size[6]. These challenges are insurmountable
54 for many nascent polyploids, although established autopolyploids can persist, indicating that they
55 can be overcome.
56 To date, work in two diploid-autotetraploid model systems has explicitly sought a basis of
57 adaptation to WGD in recent (< 300,000 year-old) autopolyploids. In *Arabidopsis arenosa,* a
58 handful of physically and functionally interacting meiosis proteins undergo adaptive evolution less
59 than 200,000 years post-WGD[7,8]. Derived alleles of these genes decrease chromosome
60 crossover rates, stabilizing meiosis[9-12]. Next, a pool-seq-based scan in *Cardamine*, a genus 17
61 million years diverged from *Arabidopsis,* showed only very modest convergence with *A. arenosa*
62 on the level of functional pathways for processes under selection shortly post-WGD[13]. Meiosis
63 showed little convergence and the signal of evolution of the meiosis genes obviously controlling
64 crossovers seen in *A. arenosa* was absent. These works gave insight but had important
65 limitations: there was very low sample number in *A. arenosa* (24 individuals[7]), and in *C. amara,*
66 pool-seq of only four populations using a highly fragmented reference[13]. More significantly, as
67 noted in the *C. amara* study, widespread vegetative reproduction in *C. amara* offers some escape
68 from selection for meiotic stability post-WGD, consistent with the results[13]. Thus, minimal
69 convergence between these systems leaves unresolved what salient processes stabilize recent
70 polyploids. This is important because the genomic changes that occur post-WGD may also help
71 explain why some polyploids are so successful and most are not.
72 Here we address this in a novel system that overcomes these limitations in a more
73 distantly related, independent, and successful set of WGD events within a single species flock (at
74 least 5 WGDs in *Cochlearia* the last 300,000 years)[14]. The *Cochlearia* species complex exhibits
75 diploid, autotetraploid, allohexaploid, octoploid and heptaploid cytotypes (Fig.1A)[14-17], with the
76 widespread autotetraploid cytotype[14], similar in age to *A. arenosa*[18,19]. *Cochlearia* is found across
77 Europe, from Spain to the Arctic, in a wide range of habitats including freshwater springs, coastal
78 cliffs, sand dunes, salt marshes, metal contaminated sites and roadside grit (Fig. 1A)[14,20-30]. A

79 broad habitat differentiation is evident by ploidy, with diploids typically found in upland freshwater
80 springs, autotetraploids overwhelmingly on coasts, often directly adjacent to seawater or
81 continuously submerged, and hexaploids in extreme salt marsh conditions. In fact, the hexaploid
82 *Cochlearia danica* is one of the most rapidly spreading invasive species in Europe, invading salted
83 roadways since the 1970's[23,31].
84 Here we first assess *Cochlearia* demography by individually resequencing 113 plants from
85 33 diploid, autotetraploid, hexaploid, and outgroup populations from across its ploidy-variable
86 range. We then focus our analysis on closely related diploids and autotetraploids in the UK and
87 scan for selective sweeps post-WGD. We dissect functional targets of adaptive evolution post-
88 WGD using protein modelling and identification of orthologous derived sites from functional
89 studies. Our results show convergence at the process level in very recent WGD adaptation events
90 in these three genera separated by greater spans of ~40 million years. This indicates that similar
91 processes adapt in response to WGD, but that specific genes recruited are far less constrained.
92 Surprisingly, also we also find strong signal of post-WGD evolution in several kinetochore
93 components, pointing to a novel mechanism of adaptation to polyploid meiosis and mitosis.
94
95 **Results and Discussion**
96 **Chromosome-level assembly.** To serve as a reference for our demographic analysis and
97 selection scans, we built a chromosome-level genome assembly of one diploid *Cochlearia*
98 *excelsa* individual (Styria, Gurktaler Alpen, Austria). We chose *C. excelsa* because it is an early
99 diverging diploid (2n = 12) and conveniently also a rare primarily selfing *Cochlearia* species. This
100 resulted in a highly contiguous primary assembly (contig N50=15 Mb; Fig.1B), generated from
101 Oxford Nanopore PromethION data (read N50=27 Kb). The primary assembly was performed
102 with Flye 2.9[32] and NECAT[33] with one round of polishing in Medaka[34] and Pilon[35] The assembled
103 contigs were then scaffolded to chromosome scale using Hi-C (Fig.1B) and a final cleanup
104 performed with Blobtools[36] (Fig. S1). Hi-C-guided chromosome arm orientations were confirmed
105 with concordant FISH and in silico mapping of telomeric and centromeric repeats (Fig.1B; Fig.
106 S2). This assembly consists of six primary scaffolds corresponding to the six *C. excelsa*
107 chromosomes with a scaffold N50 of 37 Mb and an overall genome size of 310 Mb (Fig.1B, C),
108 matching our estimated haploid genome size of 302 Mb (Fig. S3). Gene space representation
109 was very good, with 96.4% complete Brassicales BUSCOs[37] found in the assembly Fig.1C). We
110 performed an annotation incorporating RNA-seq data from the reference line (flower bud, leaf,
111 stem, and silique), protein homology information, and *ab initio* modelling with BRAKER2[38]. This
112 yielded 54,424 gene models across the six chromosomes.

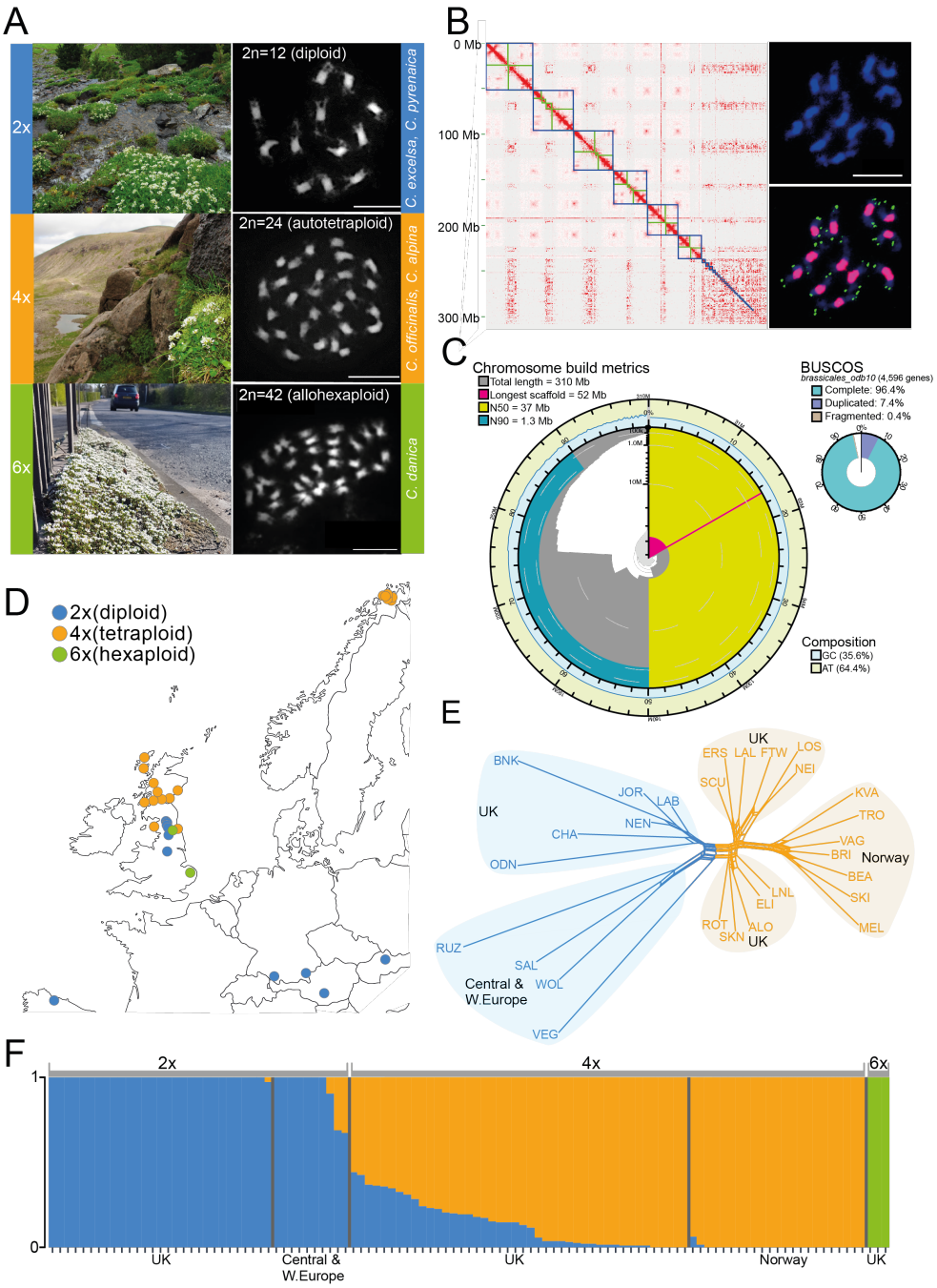

Figure 1 | Ploidy variation, genome assembly, sampling, and genetic structure. **(A)** Three *Cochlearia* species in this study (top: *C. pyrenaica* [2n=2x=12], middle: *C. officinalis* [2n=4x=24], bottom: *C. danica* [2n=6x=42]. Scale bars, 10 μm; **(B)** Hi-C contact map and cytology used to orient chromosome arms. Chromosomes are bounded in blue and centromeres indicated in green. Cytology shows DAPI-stained chromosomes with heterochromatic pericentromeres (top) and 102 bp satellite (pink) and Arabidopsis-type telomeric satellite (green) probes hybridizing to all (peri)centromeres and telomeres, respectively (bottom). Scale bar, 10 μm; **(C)** Chromosome-scale assembly of the diploid *C. excelsa* genome; **(D)** The 30 *Cochlearia* populations included in the short read sequencing (locations in Dataset S2); **(E)** Nei's genetic distances between the populations (hexaploids excluded) visualized in SplitsTree[39]; **(F)** fastSTRUCTURE[40] analysis (*k*=3, min alleles=8) of all *Cochlearia* individuals in the study, with regions and ploidies indicated. Blue=diploids; orange=tetraploids; green=hexaploids.

125 **Geographic distribution, ploidy variation, cohort construction and population sequencing.**
126 To determine optimal population contrasts for WGD-specific signatures of selection, we sampled
127 populations across the reported range of the *Cochlearia* species throughout Europe[14,17,22-25,27,28,30]
128 and conducted flow cytometry- and cytologically-based surveys of genome size and ploidy
129 variation (Fig.1A; Dataset S1 Ploidy Survey; Fig. S4). Measurements were normalised against
130 the diploid population with the most stable individual within-population genome size estimates,
131 WOL (Dataset S1). We focus our demographic analysis on the three most abundant ploidies
132 (Fig.1A): diploids = *Cochlearia pyrenaica* (inland UK and mainland Europe), autotetraploids =
133 *Cochlearia officinalis* (coastal UK and Norway), hexaploids = *Cochlearia danica* (inland UK).
134 Based on these ploidy surveys we chose 113 individuals from 33 populations for sequencing by
135 Illumina PE (average per-individual depth = 17x; minimum = 4x; Fig.1D; Dataset S2 Sample
136 Metrics). The final dataset consisted of 18,307,309 SNPs, on average one variant every 17 bp
137 (quality and depth filtered; Methods).
138
139 **Genetic structure.** To assess structure in our dataset, we first performed fastSTRUCTURE[40]
140 (Fig. 1F) on our 109 *Cochlearia* individuals, excluding outgroup sister genus Ionopsidium (23,733
141 LD-pruned, biallelic 4-fold-degenerate SNPs; max 20% missing data; min minor allele frequency
142 = 0.02). *K*=3 maximized marginal likelihood and grouped samples by ploidy. Focusing on diploids
143 vs tetraploids, PCA confirmed that ploidy dominates over geography (PC1 [ploidy] = 27% of
144 variance explained; PC2 [UK vs Norwegian tetraploids] = 11% of variance explained; Fig. S5).
145 This is reflected also in clean geographical groupings by SplitsTree[39] analyses, which visualize
146 simple genetic distances (Fig. 1E). This is consistent with a rapid inter- or peri-glacial radiation
147 and postglacial migration such as found in other *Cochlearia* species[14].
148
149 **Meiotic and ion homeostasis-related phenotypic shifts upon WGD.** In a young autopolyploid,
150 initial maintenance of diploid-like crossover frequencies can lead to chromosome entanglements
151 and breakage at meiosis. We therefore confirmed the establishment of meiotic stability of these
152 autotetraploids (Fig. 2). Similar to *A. arenosa* autotetraploids[7], we found a significant per-
153 chromosome reduction in class I mature crossovers, evidenced by HEI10 foci ($p < 0.00001$; Mann-
154 Whitney). This translated to a degree of meiotic stability in the *Cochlearia* tetraploids, although
155 we observed considerably greater variation in meiotic stability in autotetraploids, which varied
156 dramatically both within and between families for multivalent production, as seen in *A. arenosa*.
157 This high variability suggests segregating variation for factors promoting stability (Fig. 2C, D).

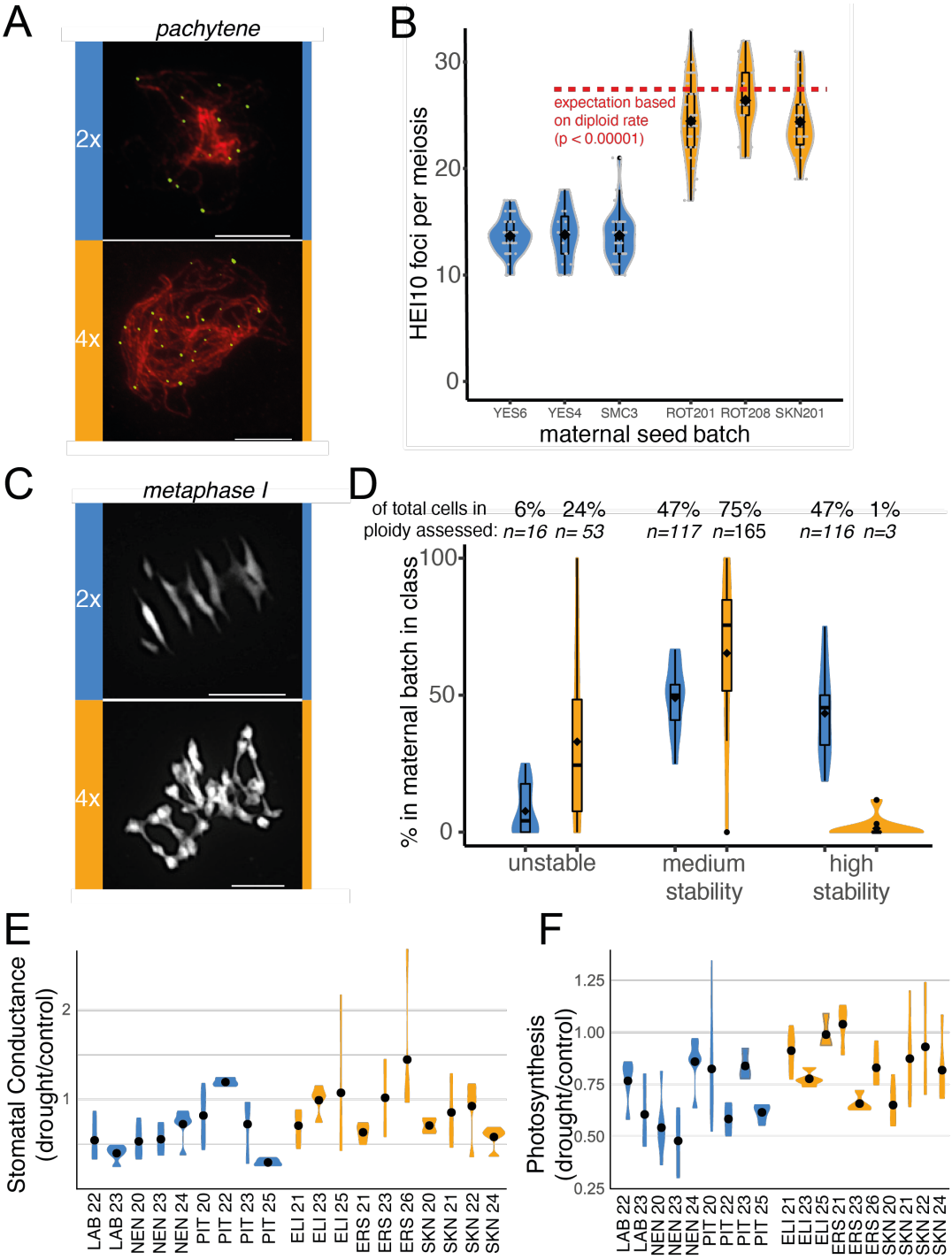

Figure 2 | Reduction in crossover number, improved drought response, and high inter-lineage phenotypic variability upon WGD. (A and B) Quantification of mature meiotic crossover events by HEI10 staining (green) on pachytene chromosomes stained by ZIP1 (red), showing a significant per-chromosome downregulation of crossovers in *Cochlearia* autotetraploids (p <0.00001; Mann-Whitney); **(C)** Metaphase I chromosome spreads and **(D)** quantification of meiotic stability, showing decreased, but highly variable stability in tetraploids relative to diploids. Increased stomatal conductance **(E)** and photosynthetic **(F)** rates under dehydration stress conditions in tetraploids. Plots show median and variation of drought stressed plants in comparison to well-watered plants. (Sign differences between diploids and autotetraploids as seen in one-way ANOVA with posthoc Tukey; Table S4) Blue=diploids; orange=tetraploids. Scale bars, 10 μm.

168 *Cochlearia* in the UK exhibits a broadly disjunct geographic distribution by ploidy, with
169 diploids (n=6) deeply inland and autotetraploids inhabiting coastal regions of the highest salinity,
170 including full seawater submergence (Fig.1A; Fig. S4). A direct mechanistic link between WGD
171 and salinity tolerance was established in *Arabidopsis thaliana*, where first-generation
172 neoautopolyploids (otherwise isogenic with diploid siblings) show elevated salinity tolerance and
173 intracellular potassium[41]. We therefore tested for ploidy-related differences in salinity tolerance
174 and dehydration stress tolerance in wild *Cochlearia*. Interestingly, in terms of overall plant survival,
175 we found extreme salinity tolerance in all ecotypes tested, with even diploids tolerating up to
176 600mM NaCl (salinity level of seawater), along with all higher ploidies (Table S1). Tetraploids
177 showed signal of increased drought tolerance, with both elevated stomatal conductance and net
178 photosynthetic rates under drought, relative to diploids (Fig. 2E and F; Tables S2-S4). This
179 suggests an adaptive benefit specific to higher ploidies in response to drought. A benefit under
180 salinity stress may be tempered by preadaptation to a stringent ionic challenge even in the diploids
181 of this species flock, consistent with their halophyte and cold-loving nature[42,43].
182
183 **Selective sweeps associated with WGD.** To identify candidate genes and processes mediating
184 adaptation to WGD, we next focused on the 18 geographically proximal populations from the UK
185 (44 autotetraploid individuals and 29 diploid individuals) with good sequencing coverage.
186 Concentrating the selection scan on UK diploids and autotetraploids minimizes the effect of
187 genetic structure that would be introduced if we were to use the mainland European samples as
188 well. To guide our selection scan window size choice, we calculated pairwise linkage decay, which
189 was rapid in both diploids and tetraploids, with near complete lack of genotypic correlations within
190 2 kb to very low background levels (Fig. 3A). We thus calculated in 1kb windows a battery of
191 differentiation metrics (Dxy[44], Rho[45], Hudson's F_{ST}[46], Nei's F_{ST}[47], Weir-Cochran's F_{ST}[48], and
192 groupwise allele frequency difference [AFD]) genome-wide (minimum = 15; mean = 101 SNPs
193 per window). After filtering, these scans overlapped 40,245 of 54,424 total predicted genes, or
194 74% of gene coding loci with sufficient coverage for assessment.
195 To empirically determine which metric most reliably identified genomic regions that exhibit
196 localized peaks in AFD indicative of specific sweeps upon WGD, we performed an inspection of
197 all AFD plots in all outlier tails (see Methods). From this, we identified Hudson's F_{ST} as most
198 reliably identifying regions of localized AFD peaks (and not e.g., low diversity[44]). This formulation
199 of F_{ST} brings the added benefit of robustness for unequal population sizes and presence of rare
200 variants[49] and direct comparisons to the *C. amara* study, which used this same metric[13]. We
201 therefore extracted windows in the top 1% of this distribution as candidate outliers, consisting of

1,823 1kb windows, overlapping 753 gene-coding loci for which we could obtain functional descriptions primarily from orthology (or, lacking this, close homology) to *Arabidopsis thaliana* (Dataset S3 Selective Sweep Candidates). We focus on the most extreme 25 of these (to the right of the dashed line in Fig 3B; Table 1), which we confirm exhibit elevated Dxy values (outlier F_{ST} peak Dxy=0.17; mean outside peak Dxy=0.06; Mann Whitney U test: $P < 2 \times 10^{-16}$; Fig. 3C).

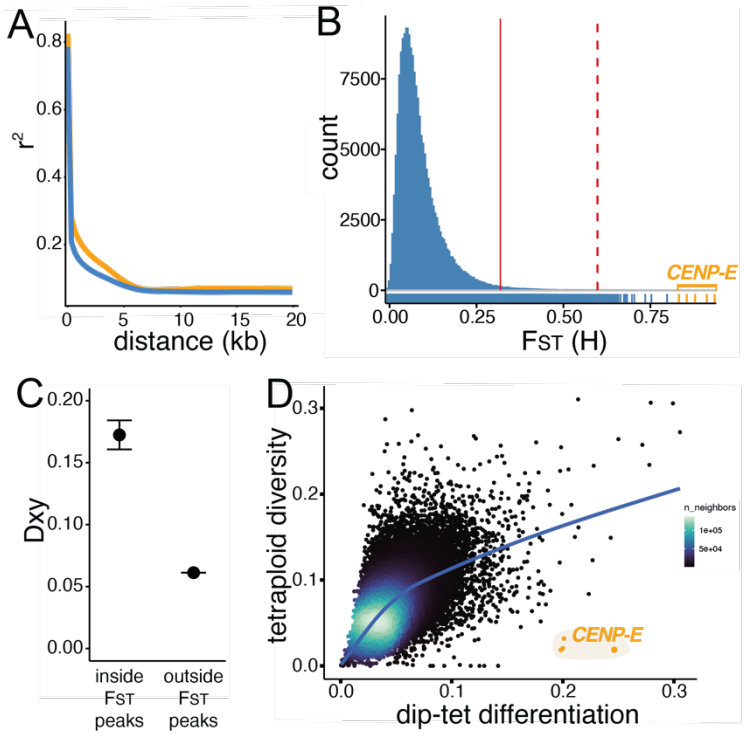

Figure 3 | Rapid linkage decay and empirical outlier analysis. (A) Immediate decay of genotypic correlations (r^2) observed in both diploid (blue) and tetraploid (orange) *Cochlearia*; **(B)** Distribution of genome wide F_{ST} values for 182,327 1kb windows. The dashed red line gives the extreme stringency F_{ST} cutoff of the top 25 genome-wide outlier genes; the solid red line gives the 1% cutoff; windows inside the *CENP-E* gene coding region are highlighted in orange; **(C)** Dxy values are significantly elevated inside F_{ST} peaks (Mann Whitney U test: $P < 2 \times 10^{-16}$); **(D)** *CENP-E*, the #1 genome-wide F_{ST} outlier, also exhibits greatly excess differentiation for its level of diversity in the tetraploids, a classical signal of selective sweep.

Complementing this approach, we also focused our 1% outlier list on gene coding regions with a fineMAV approach[50]. Using Grantham scores to estimate functional impact of each non-synonymous amino acid change encoded by a given SNP, this approach scales the severity of predicted amino acid change by the AFD between groups. Of the 107,055 non-synonymous-encoding SNPs assigned a MAV score, the top 1% outliers from the empirical distribution were intersected with our F_{ST} outliers, yielding a protein-evolution-oriented list of 159 gene coding loci, harboring 290 MAV SNPs (bold in Dataset S3 Selective Sweep Candidates; 1% F_{ST} outliers with 10 or more 1% extreme MAV outliers are given in Table 1). By these approaches, we could resolve clear gene-specific peaks of F_{ST} (Fig. 4A) and candidate selective sweep alleles in our

225 top 25 genome-wide outliers (Fig. 4 B-G).

F_{ST} rank	*Cochlearia* ID	*A.thaliana* ID	Name	Description
1*	g7445	AT3G10180	CENP-E	**CENTROMERE PROTEIN E.** Kinetochore protein that moves mono-oriented chromosomes to the spindle equator[51]. Cooperates with chromokinesins and dynein to mediate chromosome congression. Activity is regulated by post-translational modifications, protein interactions and autoinhibition Tetraploid cancers are far more susceptible to CENP-E inhibitors than diploids[52-54].
2	g31016	AT1G06670	NIH	NUCLEAR DEIH-BOX HELICASE. Binds DNA without clear specificity[55].
3	g7446	AT3G10070	TAF12	TATA-ASSOCIATED II 58. Controls stress responsive root growth[56]
4	g40185	AT2G26690	NPF6.2	NITRATE TRANSPORTER 6.2. Mediates drought stress response[57]
5	g25338	AT4G00060	MEE44	MATERNAL EFFECT EMBRO ARREST 44: Part of the RNA TRAMP complex
6	g49945	AT1G15660	CENP-C	**CENTROMERE PROTEIN C**. Kinetochore protein that is critical for centromere identity in both mitosis and meiosis. Loss of CENP-C results in aneuploidy and cell death[60]. Necessary in mitotic cells for kinetochore assembly centromere establishment. Mutants also fail to retain centromeric SC in late pachytene[58,59].
7	g25334	AT1G05940	CAT9	CATIONIC AMINO ACID TRANSPORTER 9
8	g25335	AT4G00026	SD3	**SEGREGATION DISTORTION 3**. Mutants have lower ploidy[60].
9	g6996	AT3G16730	CAP-H2 / HEB2	**CONDENSIN-II COMPLEX SUBUNIT H2**. Functions in DSB repair and arranges interphase chromatin[61,62]. Plays a role in alleviating DNA damage and genome integrity in *A.thaliana*[62]. Condensin II-depleted human cells have a defect in homologous recombination-mediated repair[63].
10	g24649	AT5G40595		Unknown protein
11	g24648	AT4G02000		Ta11-like non-LTR retrotransposon
12	g33330	AT1G54310		S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
13	g40302	AT4G10310	HKT1	**HIGH-AFFINITY K+ TRANSPORTER 1**. Sodium transporter. Mediates salinity tolerance in wild *A. thaliana* populations[64]. Under selection post-WGD in *A. arenosa*[7].
14	g33311	AT5G61390	NEN2	NAC45/86-DEPENDENT EXONUCLEASE-DOMAIN PROTEIN 2
15	g10739	AT2G01980	SOS1	**SALT OVERLY SENSITIVE 1.** Plasma membrane-localized Na+/H+ antiporter that extrudes Na+ from cells. Is essential for plant salt and stress tolerance.
16	g46402	AT1G77990	SULTR2	SULPHATE TRANSPORTER 2;2. A low-affinity sulfate transporter
17	g4631	AT5G19270		Reverse transcriptase-like protein
18	g4632	AT2G07200		Cysteine proteinases superfamily protein
19	g25121	AT5G35970		P-loop containing nucleoside triphosphate hydrolases superfamily protein
20	g45503	AT5G08620	STRS2	STRESS RESPONSE SUPPRESSOR 2. A DEA(D/H)-box helicase involved in drought, salt and cold stress responses.
21	g25112	AT5G35980	YAK1	YAK1-related. Controls cell cycle and regulates drought tolerance[65,66].
22	g40349	AT4G11110	SPA2	SPA1-related 2. Convergent with WGD adaptation in *A. arenosa*[7].
23	g54387	AT3G01420	DOX1	An alpha-dioxygenase involved in protection against oxidative stress.
24	g25300	AT5G34940	GUS3	GLUCURONIDASE 3
25	g45090	AT1G65320	CBSX6	Cystathionine beta-synthase family protein
15 mav	g39361	AT5G55820	INCEP / WYRD	**INNER CENTROMERE PROTEIN**. The largest subunit of the Chromosome Passenger Complex (CPC), and directly binds to all other subunits in animals and yeast[67]. The CPC ensures that all kinetochores are attached to microtubules emanating from opposing poles. INCENP is necessary for normal mitotic divisions[68]. At mitosis and meiosis localises to kinetochores, and later, the phragmoplast[67].
12 mav	g25323	AT4G29090		Ribonuclease H-like superfamily protein
11 mav	g23778	AT4G25290		DNA photolyase
11 mav	g54385	AT1G56145	CORK1	A LRR receptor kinase required for cellooligomer-induced responses.
10 mav	g33400	AT1G50240	FU	**FUSED**. An ARM repeat domain-containing protein kinase involved in male meiosis. Tightly localized to nascent phragmoplast and with the expanding phragmoplast ring.

226 **Table 1 | Top selective sweep candidates following WGD in *Cochlearia*.** Top 25 of 40,245 genes
227 assessed, with 5 additional genes (bottom) with 10 or more MAV SNPs and in top 1% F_{ST} tail. * = contains
228 37/39 whole-genome fixed differences between diploids and tetraploids and is also the only gene with three
229 of the top 25 Dxy windows. Genome-wide F_{ST} rank is given in Column 1.
230

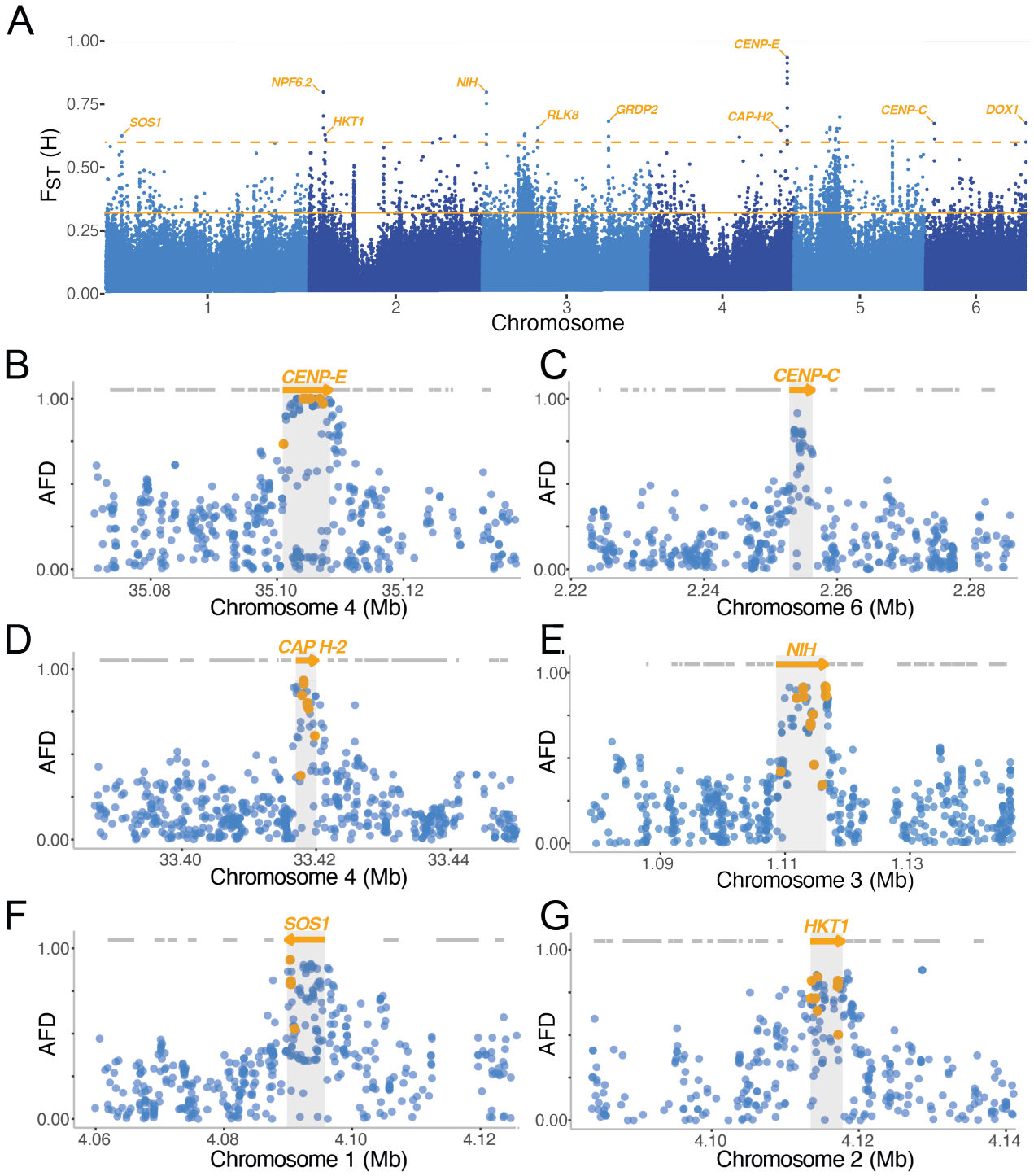

Figure 4 | Selective sweep signatures of DNA management and ion homeostasis alleles. (A) Ploidy-specific differentiation across the *Cochlearia* genome. The dashed orange line gives the extreme stringency F_{ST} cutoff of the top 25 outlier genes (of 40,245 assessed); the solid line gives a 1% F_{ST} cutoff; **(B-G)** Examples of selective sweep signal among 6 top outlier genes; **(B-E)** Represent kinetochore or DNA management; **(F-G)** ion homeostasis functional categories. The X axis gives genome position in megabases (Mb). The Y axis gives AFD values at single SNPs (dots) between diploid and tetraploid *Cochlearia*. Orange arrows indicate genes overlapping the top F_{ST} gene outliers, and grey lines indicate neighboring genes. Orange dots indicate MAV outliers.

240 **Functional processes under selection post-WGD in *Cochlearia*.** While we focus our
241 discussion below to the top 25 genome-wide outliers, a broader, 1% F_{ST} list of 753 selective sweep
242 candidates (Dataset S3) yield particularly informative gene ontology (GO) enrichments, with 181
243 significantly enriched categories (using a conservative 'elim' Exact Test; Dataset S4 Gene
244 Ontology Enrichment Cochlearia). Many of the enriched categories can be grouped into three
245 classes congruent with WGD-associated changes[1,2,4-7,69]: 1) DNA management: 29 categories
246 relate to DNA integration, cell division, meiotic chromosome segregation, mitosis, DNA repair,
247 and recombination; 2) Ion homeostasis: 26 categories relate to ion transport (principally
248 extrusion), cation homeostasis, salt stress, and stomata; and 3) Cell stoichiometry: 7 categories
249 relate to global gene expression, cell wall, and biosynthetics, pointing to both global gene
250 expression and 'nucleotypic'[6] changes upon WGD. These categories often overlap, for example
251 'cell stoichiometry' and 'DNA management' in terms of global RNA transcriptional changes post-
252 WGD, which should alter due to doubled DNA template.
253
254 **Kinetochore evolution upon WGD.** Genome-wide, the most dramatic selection signature is
255 directly over the coding region of the *CENTROMERE PROTEIN E* ortholog (*CENP-E*; Fig. 4A,
256 B)[70]. This gene overlaps the 5 top F_{ST} outlier windows (mean F_{ST} = 0.88) and includes 37 of the
257 39 genome-wide fixed differences between ploidies. An essential kinetochore protein, *CENP-E*
258 moves mono-oriented chromosomes to the spindle equator, mediating congression[51,71,72].
259 Strikingly, tetraploid cancers are far more susceptible to CENP-E inhibitors than diploids[52-54,73,74].
260 The *Cochlearia* CENP-E coding region contains 26 SNPs (synonymous or non-
261 synonymous) that are highly differentiated between diploids and tetraploids (>50% AFD). Six of
262 these are unique to *Cochlearia* tetraploids, at highly conserved sites across angiosperms (Fig.
263 5A). None are in characterized conserved functional regions of the kinase domain[75], meaning
264 motor activity is likely intact. Most of the tetraploid-specific changes are in the coiled coil regions,
265 which in animals are important for regulation of cell division via phosphorylation and protein-
266 protein interactions[51]. For example, point mutations in the coiled coils are associated with human
267 disease (e.g., microcephaly[76,77]). In humans and Xenopus, these regions are known to be
268 extensively phosphorylated during the cell cycle and may be involved in the autoinhibition of
269 CENP-E[78-81]. Indeed, we see four tetraploid-specific changes that may affect regulation via the
270 loss of phosphorylation (S717A, S821A, S1059L and S1169). Three additional changes toward
271 the C-terminus are in a cargo (chromosomes) binding region. Four of the tetraploid-specific
272 changes show remarkable conservation across plants, being otherwise absolutely conserved
273 across CENP-E-like kinesins (Fig. 5A; A607G, Q613D, R899Q and Q1024E). Taken together,

these data suggest changed regulation of CENP-E at mitosis and/or meiosis. This is consistent with functional evidence from *A. thaliana* showing that *CENP-E* mutations extend the cell cycle[81].

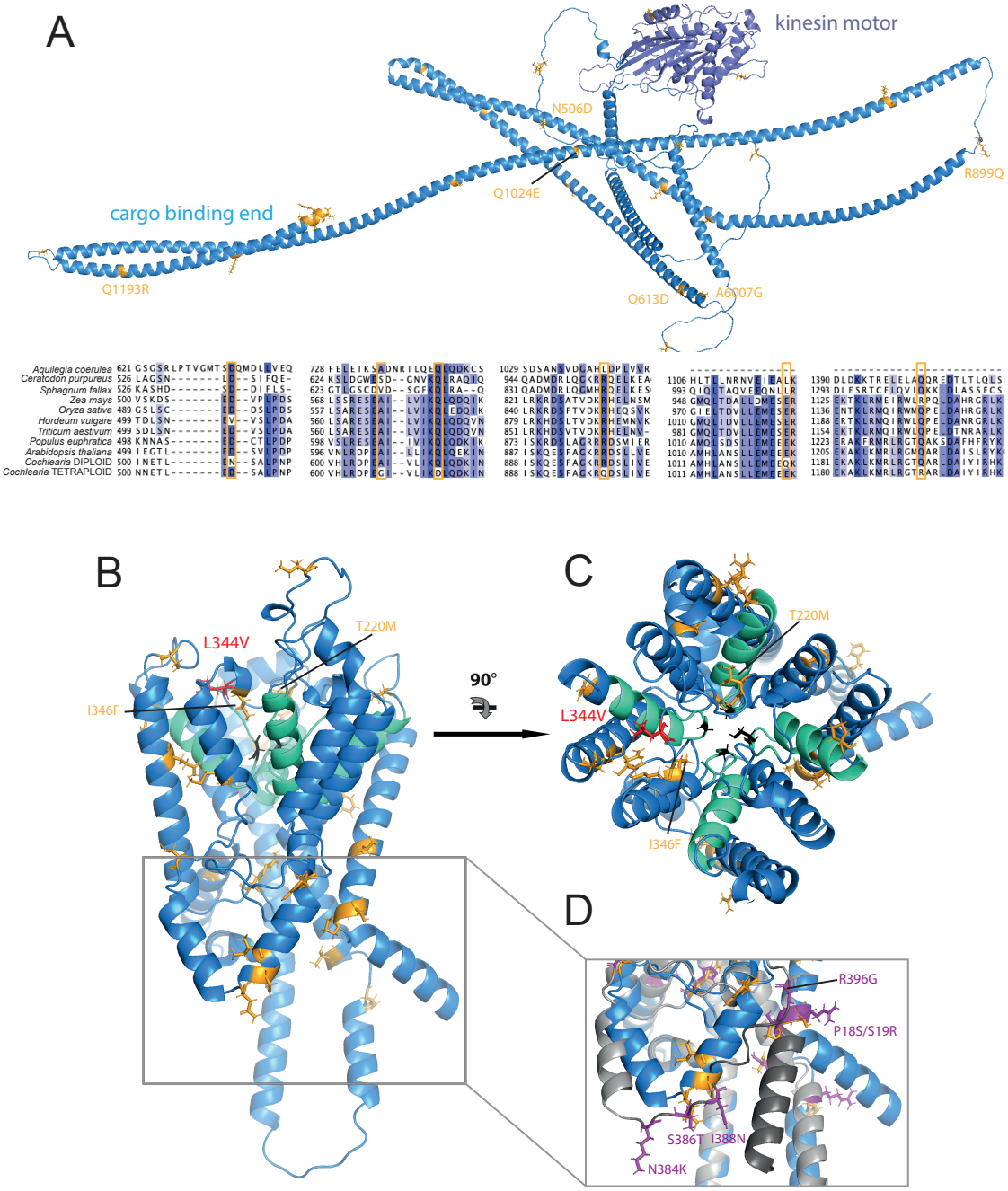

Figure 5 | Tetraploid-derived protein structure changes in CENP-E and HKT1. (A) Kinetochore subunit CENP-E structural prediction and pan-plant diversity alignment. Structure: diploid structure (blue), consensus (>50% allele frequency) derived polymorphisms in tetraploids (orange), kinesin motor domain (slate). Alignment: colour by percentage identity; consensus mutations at highly conserved sites (orange boxes); **(B-D)** Ion channel HKT1: diploid structure (blue), mutation sites (orange; diploids, magenta: tetraploids), pore domains (teal); **(C)** HKT1 extracellular surface. Salinity tolerance-mediating change L344V in red and selectivity filter residues S/G1, G2, G3 and G4 (black); **(D)** superimposition of diploid (blue) and tetraploid (grey) intracellular domains. Structures that are predicted to have rotated are highlighted on the tetraploid structure (dark grey). Tetraploid consensus mutations (magenta).

286 Our top outlier list (Table 1) contains two additional orthologs of *A. thaliana* kinetochore
287 components: *CENP-C*, an essential kinetochore component in both mitosis and meiosis, needed
288 for centromere identity in plants, yeast, *Drosophila*, and humans[58,59,82] (Fig. 4C) and *INNER*
289 *CENTROMERE PROTEIN (INCENP)*, which controls mitotic and meiotic chromosome
290 segregation and cytokinesis in plants, yeast and animals[68,83]. At mitosis and meiosis *INCENP*
291 localizes to kinetochores, and later, the phragmoplast as the main subunit of the Chromosome
292 Passenger Complex[67]. Both INCENP and CENP-C contain 1% outlier-MAV SNPs, with *INCENP*
293 harboring a remarkable 15 MAV outlier SNPs, the greatest number of any gene in the genome.
294
295 **Evolution of DNA repair and transcription.** Several of the top signals of selective sweeps are
296 in DNA repair-related genes, for example in *CONDENSIN-II COMPLEX SUBUNIT*
297 *H2/HYPERSENSITIVITY TO EXCESS BORON 2* (Fig. 4D), which functions directly in double-
298 strand break (DSB) repair[62,84] and chromatin management in plants, mouse and *Drosophila*[61,85].
299 *Condensin II* is required for proper DNA DSB repair by homologous recombination (HR) repair in
300 *A. thaliana* and humans[62,63], and has been implicated association and dissociation of
301 centromeres[85]. We find in our 1% F_{ST} outliers an outlier also at a homolog of *DAYSLEEPER*, an
302 essential domesticated transposase[86]. *DAYSLEEPER* binds the Kubox1 motif upstream of the
303 DNA repair gene *Ku70* to regulate non-homologous end joining double-strand break repair, the
304 only alternative to HR[87,88].
305 In a young polyploid, the total DNA content doubles but the protein content and cell size
306 does not scale accordingly[6], so we predicted that the control of global gene expression, like
307 meiosis, should undergo adaptive evolution post-WGD. Here we see signal of this, with a suite of
308 DNA or RNA polymerase-associated genes among our selective sweep outliers. In our 1% F_{ST}
309 outliers this includes *NRPB9*, an RNA polymerase subunit that is implicated in transcription
310 initiation, processivity, fidelity, proofreading and DNA repair[89-93] as well as the ortholog of *MED13*,
311 of the mediator complex, which is essential for the production of nearly all cellular transcripts[94]
312 (Dataset S3).
313
314 **Evolution of ion homeostasis, transport and stress signaling.** The ionomic equilibrium of the
315 cell is immediately disrupted at WGD[41]; in particular K^+ concentrations are increased instantly,
316 consistent with increases in salinity tolerance in synthetic *A. thaliana* autotetraploids[41]. In our
317 young tetraploids, among the top selective sweeps are ion channels that function explicitly to
318 remove K^+, Na^+ and other cations from the cell. At F_{ST} rank 13 genome-wide, we see the ortholog
319 of *HIGH-AFFINITY K^+ TRANSPORTER*[64,95] *(HKT1;* Fig 4G, 5B*),* and at rank 15, *SALT OVERLY*

320 *SENSITIVE 1 (SOS1)*, a membrane Na^+/H^+ transporter that removes excessive Na^+ from the cell
321 and is central to salt tolerance[96,97] (Fig 4F). Several studies have demonstrated adaptive natural
322 variation in response to salinity in *HKT1*[64,95], but among natural systems there have been to date
323 no works until now that implicate *SOS1,* although this gene is central to salt tolerance pathways
324 (discussed in [98]).
325 Our modeling confirms that the *Cochlearia HKT* under selection is class 1 from its
326 selectivity filter residue configuration (S-G-G-G), indicating it is likely Na^+ selective (and K^+ non-
327 selective)[99]. There is a L344V mutation in the tetraploid relative to the diploid; remarkably, this is
328 the identical site and amino acid change that is associated with salt tolerance in rice OsHKT1;5
329 (Fig 5B)[100,101]: functional confirmation in rice shows the orthologous site substitution to valine in
330 our tetraploid to be associated with salt tolerance (including faster Na^+ transport), while the diploid
331 leucine is associated with salt sensitivity (including slower Na^+ transport)[100]. Given that the
332 tetraploids live overwhelmingly in coastal regions where they are exposed to extreme Na levels,
333 while the diploid lives in low Na freshwater streams, this makes biological sense. While the close
334 proximity of L344 alone is likely enough to disrupt pore rigidity via its larger side chain relative to
335 V, in *Cochlearia* HKT1 we also see T220M (a highly conserved residue) and I346F mutations
336 which are even closer to the P1-4 pore domains that hold S-G2-G3-G4 together to create the
337 selectivity control. These mutations all introduce large side chains that likely affect structural
338 dynamics at the P1-4 pore[100,101]. In addition, mutations F326I, R200P L180F, Q303H, M360I are
339 all in (R200P and F326I) or in contact with the four alpha helices that stabilize the SGGG
340 selectivity filter. While this may seem an excessive quantity of mutations, suggesting gene
341 inactivation and relaxation of selection, all the sites except M360I, T220M and L344V are loosly
342 conserved, suggesting flexibility in these regions.
343 There are also mutations on the cytosolic side of the protein, a few poorly conserved
344 residues appear to induce small structural changes (Fig 5D). This includes a cluster of changed
345 residues (I388N, S386T, N384K, R396G) which are predicted to break up an alpha helicase in
346 the tetraploid, and P18S and S19R, which appear to induce a break in the first alpha helicase of
347 the protein. To our knowledge this domain is not functionally characterized, though given this
348 positioning, this could represent a change in signaling or regulation.
349 Congruent with its conserved central role in ion homeostasis, *SOS1* is highly conserved
350 and tetraploid-specific changes are not near active transport (nucleotide or ion) binding sites, nor
351 dimerization domains. Instead, a tight cluster of 3 mutations marks the boundary between the β-
352 sheet-rich cytoplasmic domain (β-CTD) and the C-terminal autoinhibition (CTA) domain which
353 contains a further 3 mutations (K1014E, T1075S and R1101Q). The CTA is unstructured;

354 however, it has been experimentally shown that truncation, or just two point mutations, can
355 radically change the behavior of the channel, presumably by releasing autoinhibition increasing
356 Na^+ transport and therefore increasing salt tolerance[102]. Notably, this includes one T1075S
357 substitution: S and T are two of the three amino acids that can be phosphorylated and it has been
358 shown that this exact residue behaves differently when phosphorylated, which suggests that the
359 choice of S or T may have evolutionary consequences[103].
360 Ion homeostasis shifts should be associated with changes in responses to salt, osmotic,
361 and cold stress, as all these stressors have a common osmotic basis. Such a link between
362 immediate ionomic changes in the polyploid cell may be a key functional basis for the observed
363 ecotypic differentiation of young polyploids, especially as observed in arctic and alpine conditions.
364 We accordingly see in our top outliers categories of relevant genes, for example among the top
365 candidates the ortholog of *DEAD-BOX RNA HELICASE 25 (STRS2)*, identified in *A. thaliana* as
366 a repressor of stress signaling for salt, osmotic, and cold stress[104,105]. This gene also controls
367 freezing tolerance[106], highly relevant to the cold-loving arctic and alpine history of *Cochlearia*[14].
368 To assess if perhaps these sweeps were better associated with ecotype differences
369 between ploidies, we performed a salt tolerance experiment on diploid and tetraploid plants.
370 Interestingly, given their divergent ecotype preferences (Fig. S4), with tetraploids found in more
371 saline conditions, we found that the diploid *Cochlearia* are in fact more salt tolerant than the
372 tetraploids ($p = 2.178 \times 10^{-05}$; See Supplementary Text 1 and Table S5.). This finding also
373 contrasts with observations of increased salinity tolerance in neotetraploid *Arabidopsis thaliana*[41].
374 Again, however, this may be a signal of preadaptation to osmotic challenge (common to freezing,
375 salinity, and dehydration) across the halophyte *Cochlearia*[42].
376 Relevant also to these phenotypes, genes involved in stomatal function were outliers post-
377 WGD, such as the ortholog of *OPEN STOMATA2*, a target of ABA stress signaling to close the
378 stomata during drought response[107]. This gene is an ATPase in the plasma membrane that drives
379 hyperpolarization and initiates stomatal opening[105]. We also see *TOO MANY MOUTHS*, where a
380 mutation leads to disruption of asymmetric cell division during stomata development[108]. Finally,
381 we see selection signal in *STOMAGEN*, which acts on the epidermis to increase stomatal
382 formation[109]. Sweeps in these loci are consistent with the phenotypic shifts we observe of
383 increased stomatal conductance and net photosynthetic rate under drought conditions in
384 tetraploid *Cochlearia* populations relative to diploids (Fig. 2E; Supplementary Text 2).
385
386 **Gene-level convergence.** To test for convergence at the ortholog level, we first determined
387 orthogroups[110] between *Cochlearia, A. arenosa,* and *C. amara* (Methods). Top 1% F_{ST} outliers for

388 *Cochlearia* (n=753; Dataset S3), *A. arenosa* (n=452; Dataset S5), and *C. amara* (n=229; Dataset
389 S6) were considered orthologues if they were part of the same orthogroup. By this criterion not a
390 single ortholog was under selection in all three species (Fig. S6A; Dataset S7 Gene Ortholog
391 Convergence). This approach depends on strict 3-way orthogroup assignment, so we then
392 searched for convergence by assigning all genes in the outlier lists to a nearest *A. thaliana*
393 homolog. By this 'nearest homolog' criterion, only one gene was selected in all three WGDs:
394 *DAYSLEEPER*, an essential domesticated transposase[86] with a role in regulating non-
395 homologous end joining double-strand break repair (Fig. S6B; Dataset S8).
396 Interestingly, by both homolog assignment methods, several of the best *Cochlearia* WGD
397 adaptation candidates are candidates also in *A. arenosa*: *ASY3,* functionally validated[9,12] to have
398 an primary role in stabilizing autotetraploid meiosis in *A. arenosa* is also in our 1% F_{ST} outlier list
399 in *Cochlearia*[9]. This gene, along with *ASY1*, is critical for formation of meiotic chromosome axes,
400 and tetraploid alleles of both genes result in fewer deleterious multichromosome associations and
401 more rod-shaped bivalents in metaphase I[12]. We also see *CYCD5;1*, which is a QTL for
402 endoreduplication[111]. Additionally, the salinity and osmotic genes *HKT1* and *OST2* in both top
403 candidate lists (Fig. S6A; Dataset S7 Gene Ortholog Convergence). All of these genes are
404 involved in processes that have been implicated in adaptation to WGD[1,2,4-7] and therefore stand
405 as good candidates in salient challenges to nascent polyploids. We note that overlap between
406 *Cochlearia* and both *A. arenosa* and *C. amara* candidates was greater than expected by chance
407 (SuperExactTest p=0.0024 and p=0.0047 respectively), but only marginally for *C. amara* and *A.*
408 *arenosa* (SuperExactTest p=0.014).
409
410 **Process-level convergence.** We reasoned that there may be similarities in processes under
411 selection between the three independent WGDs, despite modest gene-level convergence. To
412 estimate this, we first compared our GO results from those published in *A. arenosa* and *C. amara*.
413 The much greater signal of overlap (process-level convergence; Fig. S6C) was between
414 *Cochlearia* and *A. arenosa*: of the 113 GO biological process terms significantly enriched in
415 *Cochlearia*, 17 were among the 73 GO terms enriched in *A. arenosa* (Dataset S4). These were
416 high-level GO terms including representatives of ploidy-relevant categories, e.g. 'cell division',
417 'transmembrane transport, and 'regulation of RNA metabolic processes'.
418 Despite this evident convergence, in *Cochlearia* an array of DNA repair and kinetochore
419 genes were among top candidates, signaling a shift relative to *A. arenosa*, where a more focused
420 prophase I-oriented signal emerged primarily around Synaptonemal Complex (SC)-associated
421 proteins mediating lower crossover rates[7]. However, given that *CENP-C* mutants fail to retain an

422 SC at the centromere and *CENP-C* appears to have functions in synapsis, cohesion, and
423 centromere clustering[112], there may be a closer parallel between the obviously 'SC- focused'
424 adaptive response in *A. arenosa* and that in *Cochlearia*. Indeed, aside from *ASY3*, in the
425 *Cochlearia* 1% F_{ST} outlier list we also see *ATAXIA-TELANGIECTASIA MUTATED (ATM;* also with
426 6 MAV SNPs; Dataset S3). This gene controls meiotic DNA double-strand break formation and
427 recombination and affects synaptonemal complex organization[113]. We also see a *PDS5* cohesion
428 cofactor ortholog and *REC Q MEDIATED INSTABILITY 1* (*RMI1*), which suppresses somatic
429 crossovers and is essential for resolution of meiotic recombination intermediates[114]. Notable also
430 are *BRCA2-like B* (which is essential at meiosis and interacts with *Rad51, Dss1,* and *Dmc1*) and
431 *SHUGOSHIN C* (which protects meiotic centromere cohesion).

432 Further evidence of functional association outliers found in *Cochlearia* with those in *A.*
433 *arenosa* can be observed in protein interaction information from the STRING database, which
434 provides an estimate of proteins' joint contributions to a shared function[115]. Using comparable 1%
435 F_{ST} outlier lists from the two species, we see many connections between these independent
436 WGDs. Here we see particularly large clusters (Fig. 6A,B) in center of the overall network (Fig.
437 6C). In particular, the endopolyploidy gene[111] *CYCD5;1,* and *DNA Pol V,* a shared outlier in both
438 species, interact with a broad array of other outliers in each selection scan (Fig. 6A). This analysis
439 reveals, for example, that *DNA Pol V* shares as a top candidate either *NRPB9A* (in *A. arenosa*)
440 or *NRPB9B* (in *Cochlearia*). These subunits are partly redundant interactors with *Pol II, IV*, and *V*
441 and the double mutant is fatal in *A. thaliana*[91].

442 Taken together, these results indicate that adaptive evolution in response to WGD is
443 focused on particular functions instead of specific genes. These functions involve DNA
444 management and ion homeostasis. However, it is unclear why particular solutions are favored in
445 one species relative to another. A degree of stochasticity depending on available standing
446 variation can be expected, but species histories likely play a role, offering preadaptations that may
447 'nudge' evolution. For example, our analysis of salinity tolerance in *Cochlearia* gave the surprising
448 result that the diploid was at least as tolerant to extreme salt concentrations as the tetraploid,
449 although the diploid is found predominantly inland, except for rare diploid coastal populations from
450 Spain. A postglacial and boreal spread of the diploid towards the UK may have brought salinity
451 and cold tolerance along the way[116], altering the genomic substrate upon which selection acted
452 in response to WGD-associated ionomic challenge.

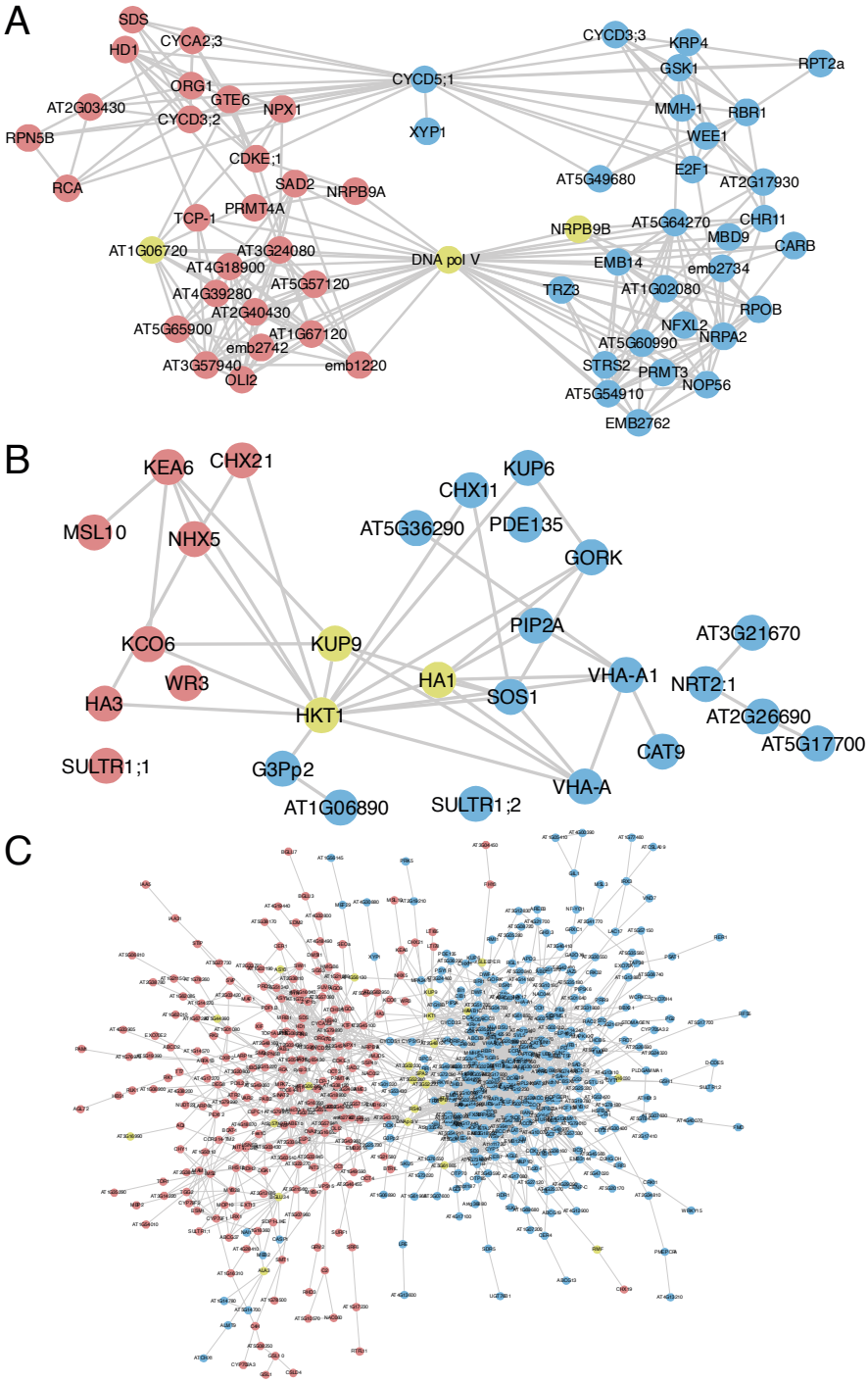

Figure 6 | Evidence for functional convergence between *Cochlearia* and *Arabidopsis arenosa* following independent WGDs. STRING[115] plots show *Cochlearia* candidate genes (blue) and *A. arenosa* candidate genes (red). Convergent genes that are present in both species' outlier lists as selection candidates are in yellow. **(A)** A large shared cluster surrounding the endopolyploidy gene *CYCD5;1*, which has many connections to large cluster centering on *DNA pol V*, which is an outlier in both datasets; **(B)** ion transport-related genes with a highly interconnected cluster of top outliers in both genome scans, *HKT1, KUP9,* and *HA1;* **(C)** The entire set of candidates in both genome scans for which the STRING database has information.

462 **Conclusion**

463 Following whole genome duplication (WGD), the newly polyploid cell requires modifications to
464 chromosome segregation, ion homeostasis, stomatal function, and diverse other processes[1,2,4,6].
465 Here we investigated the signals of adaptive evolution post-WGD in a successful novel polyploid
466 system, *Cochlearia*. We discovered striking signals of selection in core kinetochore components,
467 *CENP-E, CENP-C, INCEP, CAP-H2,* and others, as well as well-studied ion homeostasis loci,
468 *HKT1* and *SOS1.* We detail specific changes in these proteins upon WGD which are known in
469 model systems to directly modulate WGD-relevant function. We also compare our results to
470 independent WGD adaptation events, finding convergence of these processes, but not genes,
471 indicating a highly flexible array of adaptive mechanisms.

472 Our results also suggest a hypothesis for the occasionally spectacular adaptability of
473 polyploids. We observed a broad array of DNA management and repair processes under selection
474 in all species, and especially *Cochlearia*. This may signal a temporarily increased post-WGD
475 susceptibility to DNA damage, due to suboptimal function of DNA repair genes during the process
476 of adaptation to the WGD state. This may result in a relative 'mutator phenotype' in neopolyploids.
477 Such a mutator phenotype has been plainly observed in polyploid metastatic human cancers,
478 which not only exhibit SNP-level hypermutator phenotypes, but also dramatic structural variation
479 in malignant aneuploid swarms that are associated with progression[3]. We speculate that a parallel
480 to this may exists following other WGDs. Whether this hypothesis is further supported by future
481 discoveries, the centrality of WGD to evolution, ecology and agriculture, underscores the
482 importance of understanding the processes mediating adaptation to—and perhaps also by—
483 WGD.

Methods

Plant material. We first located 89 *Cochlearia* populations throughout Europe and collected population samplings of plants from each, aiming for at least 10 plants per population, with each sampled plant a minimum of 2 meters from any other. Of these, we selected 33 geographically and ecotypically diverse representative populations for population resequencing, including the outgroup *Ionopsidium* (Dataset S2 Sample Metrics). An average of 4 individuals per population were sequenced. A total of 149 individuals were initially sequenced, which was narrowed down by a cutoff of minimum average depth of 4x, leaving 113 individuals from 33 populations in the final analysed dataset, including the outgroup *Ionopsidium*.

Ploidy Determination by flow cytometry. DNA content and ploidy were inferred for populations using flow cytometry (Dataset S1). Approximately 1 square cm of leaf material was diced alongside an internal reference using razor blades in 1 ml ice cold extraction buffer (either 45 mM $MgCl_2$, 30 mM sodium citrate, 20mM MOPS, 1% Triton-100, pH 7 with NaOH for relative staining or 0.1 M citric acid, 0.5% Tween 20 for absolute measurements). The resultant slurry was then filtered through a 40-µm nylon mesh before the nuclei were stained with the addition of 1 ml staining buffer (either CyStain UV precise P [Sysmex, Fluorescence emission: 435nm to 500nm] for relative ploidy, or Otto 2 buffer [0.4 M $Na_2HPO_4 \cdot 12H_2O$, Propidium iodide 50 µgmL−1, RNase 50 µgmL−1], for absolute DNA content). After 1 minute of incubation at room temperature the sample was run for 5,000 particles on either a Partec PA II flow cytometer or a BD FACS Melody. Histograms were evaluated using FlowJo software version 10.6.1.

HEI10 immunostaining. Pachytene chromosome spreads were prepared from fixed anthers, as described previously[117]. Immunostaining was conducted using two primary antibodies: anti-AtZYP1C rat, 1:500[118] and anti-HvHEI10 rabbit, 1:250[119], followed by two secondary antibodies: goat anti-rat Alexa Fluor® 594 (Invitrogen) and goat anti-rabbit Alexa Fluor® 488 (Invitrogen). A Nikon Eclipse Ci fluorescence microscope equipped with NIS elements software was used to capture and quantify images. HEI10 foci were counted at pachytene using NIS software and significance established using the Mann-Whitney U test (Minitab v 18.1.0.0).

Fluorescence *in situ* hybridization. Mitotic chromosome spreads from fixed root tips were prepared as described previously[120]. The *Arabidopsis*-type telomere repeat $(TTTAGGG)_n$ was prepared according to [121]. The *Cochlearia*-specific 102-bp (GTTAGATGTTTCATAAGTTCGTCAA ACTTGTACAAAGCTCATTGAGACACTTATAAGCACTCATGTTGCATGAAACTTGGTTTAGAG TCCTAGAAACGCGTT) tandem repeat was designed and prepared based on Mandáková et al. (2013) and used for identification of centromeres. The DNA probes were labeled by nick translation with biotin-dUTP and digoxigenin-dUTP according to [122], pooled and precipitated by adding 1/10 volume of 3 M sodium acetate, pH 5.2, and 2.5 volumes of ice-cold 96% ethanol and kept at -20°C for 30 min. The pellet was then centrifuged at 13,000 g at 4°C for 30 min. The pellet was resuspended in 20 µl of the hybridization mix (50% formamide and 10% dextran sulfate in 2×SSC) per slide. 20 µl of the probe was pipetted onto a chromosome-containing slide. The cover slips were framed with rubber cement. The probe and chromosomes were denatured together on a hot plate at 80°C for 2 min and incubated in a moist chamber at 37°C overnight. Post-hybridization washing was performed according to [122]. After immunodetection, chromosomes

528 were counterstained with 4', 6-diamidino-2-phenylindole (DAPI, 2 µg/ml) in Vectashield (Vector
529 Laboratories). The preparations were photographed using a Zeiss Axioimager Z2 epifluorescence
530 microscope with a CoolCube camera (MetaSystems). The three monochromatic images were
531 pseudocolored, merged and cropped using Photoshop CS (Adobe Systems) and Image J
532 (National Institutes of Health) softwares.
533
534 **Reference Genome Assembly and Alignment.** We generated a long read-based *de*
535 *novo* genome assembly using Oxford Nanopore and Hi-C approaches, below.
536
537 **High Molecular Weight DNA isolation and Oxford Nanopore sequencing**. A total of 0.4 g
538 *Cochlearia excelsa* leaf material from one individual plant was ground using liquid nitrogen before
539 the addition of 10 ml of CTAB DNA extraction buffer (100 mM Tris-HCl, 2% CTAB, 1.4 M NaCl,
540 20 mM EDTA, and 0.004 mg/ml Proteinase K). The mixture was incubated at 55°C for 1 hour then
541 cooled on ice before the addition of 5 ml Chloroform. This was then centrifuged at 3000 rpm for
542 30 minutes and the upper phase taken, this was added to 1X volume of
543 phenol:chloroform:isoamyl-alcohol and spun for 30 minutes at 3000 rpm. Again, the upper phase
544 was taken and mixed with a 10% volume of 3M NaOAc and 2.5X volume of 100% ethanol at 4
545 °C. This was incubated on ice for 30 minutes before being centrifuged for 30 minutes at 3000 rpm
546 and 4 °C. Three times the pellet was washed in 4ml 70% ethanol at 4 °C before being centrifuged
547 again for 10 minutes at 3000 rpm and 4°C. The pellet was then air dried and resuspend in 300 µl
548 nuclease-free water containing 0.0036 mg/ml RNase A. The quantity and quality of high molecular
549 weight DNA was checked on a Qubit Fluorometer 2.0 (Invitrogen) using the Qubit dsDNA HS
550 Assay kit. Fragment sizes were assessed using a Q-card (OpGen Argus) and the Genomic DNA
551 Tapestation assay (Agilent). Removal of short DNA fragments and final purification to HMW DNA
552 was performed with the Circulomics Short Read Eliminator XS kit.
553
554 Long read libraries were prepared using the Genomic DNA by Ligation kit (SQK-LSK109; Oxford
555 Nanopore Technologies) following the manufacturer's procedure. Libraries were then loaded onto
556 a R9.4.1 PromethION Flow Cell (Oxford Nanopore Technologies) and run on a PromethION Beta
557 sequencer. Due to the rapid accumulation of blocked flow cell pores or due to apparent read
558 length anomalies on some *Cochlearia* runs, flow cells used in runs were treated with a nuclease
559 flush to digest blocking DNA fragments before loading with fresh library according to the Oxford
560 Nanopore Technologies Nuclease Flush protocol, version NFL_9076_v109_revD_08Oct2018.
561
562 **Genome size estimation and computational ploidy inference.** We used KMC[123] to create a k-
563 mer frequency spectrum (Kmer length=21) of trimmed Illumina reads. We then used
564 GenomeScope 2.0 (parameters: -k 21 -m 61) and Smudgeplot[124] to estimate genome size and
565 heterozygosity from k-mer spectra.
566
567 **Data processing and assembly**. Fast5 sequences produced by PromethION sequencing were
568 base called using the Guppy 6 high accuracy base calling model (dna_r9.4.1_450bps_hac.cfg)
569 and the resulting fastq files were quality filtered by the base caller. A total of 17.2 GB base called
570 data were generated for the primary assembly, resulting in 60x expected coverage. Primary

571 assembly was performed in Flye[32] and Necat[33]. The contigs were polished to improve the single-
572 base accuracy in a single round of polishing in Medaka[34] and Pilon[35].
573
574 **Pseudomolecule construction by Hi-C, assembly cleanup, and polishing.** To scaffold the
575 assembled contigs into pseudomolecules, we performed chromosome conformation capture
576 using HiC. Leaves from a single plant were snap-frozen in liquid N and ground to a fine powder
577 using mortar and pestle. The sample was then homogenised, cross-linked and shipped to Phase
578 Genomics (Seattle, USA), who prepared and sequenced an in vivo Hi-C library. After filtering low-
579 quality reads with Trimmomatic[125], we aligned the Hi-C reads against the contig-level assembly
580 using bwa-mem[126] (settings -5 -S -P) and removed PCR duplicated using Picard Tools
581 (https://broadinstitute.github.io/picard/). We used 3D-DNA[127] to conduct the initial scaffolding,
582 followed by a manual curation in Juicebox[128]. After manually assigning chromosome boundaries,
583 we searched for centromeric and telomeric repeats to orient the chromosome arms and to assess
584 the completeness of the assembled pseudomolecules. To identify the centromeric repeat motif in
585 *C. excelsa*, we used the RepeatExplorer[129] pipeline to search for repetitive elements from short-
586 read sequence data originating from the reference individual. RepeatExplorer discovered a highly
587 abundant 102 nucleotide repeat element (comprising 21% of the short-read sequence), which we
588 confirmed as the centromeric repeat motif by fluorescence in situ hybridisation. Using BLAST, we
589 localised the centromeric and telomeric (TTTAGGG) repeats and used them to orient the
590 chromosome arms. We performed a final assembly cleanup in Blobtools[36] (Fig.S1). Gene space
591 completeness was assessed using BUSCO version 3.0.2)[37].
592
593 **Assembly annotation and RNA-seq**. Prior to gene annotation, we identified and masked
594 transposable element (TE) sequences from the genome assembly. To do so, we used the EDTA
595 pipeline[130], which combines multiple methods to comprehensively identify both retrotransposons
596 and DNA transposons. After running EDTA on our chromosome-level genome assembly, we
597 performed BLAST queries against a curated protein database from Swiss-Prot to remove putative
598 gene sequences from the TE library and masked the remaining sequences from the assembly
599 using RepeatMasker (https://www.repeatmasker.org).
600
601 We then used the BRAKER2[38] pipeline to conduct gene annotatation on the TE-masked genome
602 assembly. Evidence types included RNAseq data from the identical *C. excelsa* line and protein
603 data from related species. RNA-seq was generated from bud, stem and leaf tissue. Total RNA
604 was extracted from each tissue using the Qiagen RNeasy Extraction Kit. Stranded RNA libraries
605 with polyA were constructed Using NEB Next Ultra II Directional RNA Library Prep Kit for Illumina
606 and then evaluated by qPCR, TapeStation and Qubit at the DeepSeq facility (Nottingham, UK)
607 before being sequenced at PE 150 at Novogene, inc (Cambridge, UK). We mapped the RNA-seq
608 reads of each tissue to our reference genome using STAR[131] with default parameters (-
609 twopassMode Basic) before running BRAKER2. Running BRAKER2 without UTR prediction
610 generated more gene models and much better BUSCO metrics than with UTR prediction (97.8%
611 [raw, pre-Blobtools trimmed] complete BUSCOs without UTR prediction vs 91.7% with UTR
612 prediction), so for the final annotation we used the more complete set and ran BRAKER2 without
613 UTR prediction.
614

615 **Population Resequencing and Analysis.**
616 **Library preparation and sequencing.** DNA was prepared using the commercially available
617 DNeasy Plant Mini Kit from Qiagen. DNA libraries were made using TruSeq DNA PCR-free Library
618 kit from Illumina as per the manufacturer's instructions and were multiplexed based on
619 concentrations measured with a Qubit Flourometer 2.0 (Invitrogen) using the Qubit dsDNA HS
620 Assay kit. Sequencing was carried out on either NextSeq 550 (Illumina) in house (4 runs) or sent
621 to Novogene for Illumina Hiseq X, PE150 sequencing (2 runs).
622
623 **Data preparation, alignment, and genotyping.** Reads were quality trimmed with Trimmomatic
624 0.39[125] (PE -phred33 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:50) and then
625 aligned to the *C. excelsa* reference using bwa-mem[132] and further processed with samtools[133].
626 Duplicate reads were removed and read group IDs added to the bam files using Picard (version
627 1.134). Indels were realigned with GATK (version 4.2.3.0)[134]. Samples were first genotyped
628 individually with "HaplotypeCaller" (--emit-ref-confidence BP_RESOLUTION --min-base-quality-
629 score 25 --minimum-mapping-quality 25) and were then genotyped jointly using
630 "GenotypeGVCFs" in GATK (version 4.2.3.0). The resulting VCF files were then filtered for
631 biallelic sites and mapping quality (QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5,
632 ReadPosRankSum < -8.0, HaplotypeScore < 13.0). The VCF was then filtered by depth. Based
633 on this distribution a depth cutoff of 4,322 was applied to the VCF containing the dataset and this
634 was then used as a mask for the final VCF containing all individuals.
635
636 **Demographic analysis**. We first inferred genetic relationships between individuals using
637 principal component analysis (PCA). Following [135], we estimated a matrix of genetic covariances
638 between each pair of individuals. For two individuals, i and j, covariance (C) was calculated as:
639

640 $$C_{ij} = \frac{1}{m}\sum_{s=1}^{m}\frac{(g_{is} - x_{ij}p_s)(g_{js} - x_{ij}p_s)}{x_{ij}p_s(1-p_s)},$$

641
642 where m is the number of variable sites, g_{is} is the genotype of individual i in site s, x is the average
643 ploidy level of the two individuals, and p is the alternate allele frequency. PCA was performed on
644 the matrix using the R function *prcomp*, setting scaling to TRUE, and first two axes of the rotated
645 data extracted for plotting. For fastSTRUCTURE we followed [136] by randomly subsampling two
646 alleles from tetraploid and hexaploid populations using a custom script. We have previously
647 demonstrated that results generated in this way are directly comparable to results generated with
648 the full dataset in STRUCTURE[136]. We calculated Nei's distances among all individuals in
649 StAMPP[137] and visualized these using SplitsTree[39]. Linkage disequilibrium was estimated using
650 ldsep[138]. To avoid biasing the estimates with unequal sample sizes, we chose 39 diploids and
651 tetraploids for the analysis. To reduce computation time, the analysis was performed on 4-fold
652 sites from a single chromosome (chromosome 1). To visualize the decay of LD as a function of
653 physical distance, we calculated average r^2 in 10 bp non-overlapping windows and fit a loess
654 curve on the binned data.
655
656 **Window-based scan for selective sweep signatures**. We performed a window-based
657 divergence scan for selection consisting of 1 kb windows that contained at least 15 SNPs. The

658 data were filtered as described above and in addition was filtered for no more than 20% missing
659 data and a depth of >= 8x. We calculated the following metrics: Rho[45], Nei's F_{ST} [47], Weir-Cochran's
660 F_{ST}[48], Hudson's F_{ST}[46], Dxy[44], number of fixed differences and average groupwise allele frequency
661 difference (AFD). To determine the best metric to detect localised peaks of divergence we
662 performed a quantitative analysis of AFD plot quality for all 1% outliers of each metric. Each
663 window was given a score of 0-4, with 0 being the lowest quality and 4 the highest. Scores were
664 based on two qualities: peak height and peak specificity. For peak height one point was awarded
665 if the window contained one SNP of AFD between 0.5 and 0.7, and two points were awarded for
666 any SNP of AFD > 0.7. Likewise, for peak specificity two points were awarded for an AFD peak
667 that was restricted to a single gene and one point was awarded for a peak that was restricted to
668 2-3 genes. Compared to all other single 1% outlier lists and all permutations of overlapped 1%
669 outlier lists, the top 1% outliers from Hudson's F_{ST} performed most favorably as it maximized the
670 number of '4' and '3' scores while minimizing the number of '1' and '0' scores. Finally, we masked
671 from downstream analysis a region of uniformly high differentiation marking a suspected inversion
672 at scaffold 6 (between 5,890,246 and 6,137,362 bp).
673
674 **MAV analysis.** Following [13], we performed a FineMAV[50]-like analysis on all biallelic, non-
675 synonymous SNPs passing the same filters as the window-based selection scan. SNPs were
676 assigned a Grantham score according to the amino acid change and this was scaled by the AFD
677 between ploidies. SNPs were first filtered for a minimum AFD of 0.25. The top 1% outliers of all
678 these MAV-SNPs were then overlapped with the genes in our 1% F_{ST} outlier windows to give a
679 refined list of candidate genes that contain potentially functionally significant non-synonymous
680 mutations at high AFD between cytotypes. The code outlining this can be found at
681 https://github.com/paajanen/meiosis_protein_evolution/tree/master/FAAD.
682
683 **Orthogrouping and Reciprocal Best Blast Hits.** We performed an orthogroup analysis using
684 Orthofinder version 2.5.5[110]. to infer orthologous groups (OGs) from four species (*C. amara, A.*
685 *lyrata, A. thaliana, C. excelsa*). A total of 25,199 OGs were found. Best reciprocal blast hits (RBHs)
686 for *Cochlearia* and *A. thaliana* genes were found using BLAST version 2.9.0. *Cochlearia* genes
687 were then assigned an *A. thaliana* gene ID for GO enrichment analysis in one of five ways. First
688 if the genes' OG contained only one *A. thaliana* gene ID, that gene ID was used. If the OG
689 contained more than one *A. thaliana* gene ID then the RBH was taken. If there was no RBH then
690 the OG gene with the lowest E-value in a BLAST versus the TAIR10 database was taken. If no
691 OG contained the *Cochlearia* gene then the RBH was taken. Finally, if there was no OG or RBH
692 then the gene with the lowest E-value in a BLAST versus the TAIR10 database was taken.
693 BLASTs were performed using the TAIR10.1 genome with data generated on 2023-01-02.
694
695 **GO Enrichment Analysis.** To infer functions significantly associated with directional selection
696 following WGD, we performed gene ontology enrichment of candidate genes in the R package
697 TopGO v.2.52[139], using our *Cochlearia* universe set. We tested for overrepresented Gene
698 Ontology (GO) terms within the three domains Biological Process (BP), Cellular Component (CC)
699 and Molecular Function (MF) using Fisher's exact test with conservative 'elim' method, which
700 tests for enrichment of terms from the bottom of the hierarchy to the top and discards any genes
701 that are significantly enriched in a descendant GO term. We used a significance cut-off of 0.05.

Generation of Consensus Sequences. Consensus sequences were generated for proteins of interest so that they could be closely inspected via MSAs and 3D protein structure prediction. Genomic regions were selected for either all diploids or all tetraploids present in the selection scan with GATK SelectVariants, while simultaneously being filtered for biallelic SNPs, "--max-nocall-fraction 0.2" and "-select 'AF > 0.5'". A consensus sequence was generated for exons by combining samtools faidx and bcftools consensus. Finally, a VCF containing only non-biallelic variation was manually inspected and any multiallelic variants at AF>0.5 and max-nocall-fraction<0.2 manually incorporated into the consensus.

Multiple Sequence Alignments (MSAs). We generated multiple sequence alignments using Clustal-Omega[140] in combination with amino acid sequences from the GenBank database. Sequences were selected either because the genes/proteins were well studied in other organisms or to give a phylogenetically broad coverage. Alignments were manually refined and visualised in JalView[141].

Protein modeling. Protein homology models were created using AlphaFold[142] version 2.1 on the Czech national HPC MetaCentrum. The full database was uses with a model preset of monomer and a maximum template data of 2020-05-14. Structures were visualised and images generated in the PyMOL Molecular Graphics System (Version 2.0 Schrödinger, LLC).

Data Availability

Sequence data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI (https://www.ebi.ac.uk/ena/browser/view/PRJEB66308) under accession number PRJEB66308.

Acknowledgements

We thank Kirsten Bomblies, Lara Hebberecht-Lopez, Mary Bray and Nigel Bray for assistance collecting plant material. We further acknowledge wet lab support from Anna Loreth (Heidelberg). This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program [grant number ERC-StG 679056 HOTSPOT], via a grant to LY and under the Marie Skłodowska-Curie grant agreement No. 101022295 to TH. Nanopore sequencing was performed in the University of Nottingham DeepSeq sequencing facility. We are grateful for access to the University of Nottingham's Augusta HPC service and the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme 'Projects of Large Research, Development, and Innovations Infrastructures'. We thank local authorities for issuing permission to collect samples (Slovak Ministry for Environment, permission No 062-219/18).

Author Contributions: LY and SMB conceived the study. LY, SMB, MK, PM, and JK performed collections. SMB, SB, SDD, T.Mandáková, CM, LC, MZ, and SF performed laboratory experiments. SMB, TH, SDD, CM TM, T.Mathers, T.Mandáková, EW, MB, SB, SF, PP, MK and LY performed analyses. LY and SMB wrote the manuscript. All authors approved the final manuscript.

Competing Interests statement: The authors declare no competing interests.

References

1 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. *Frontiers in Ecology and Evolution* **6** (2018). https://doi.org:10.3389/fevo.2018.00117

2 Yant, L. & Bomblies, K. Genome management and mismanagement--cell-level opportunities and challenges of whole-genome duplication. *Genes & Development* **29**, 2405-2419 (2015). https://doi.org:10.1101/gad.271072.115

3 Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat Rev Genet* **18**, 411-424 (2017). https://doi.org:10.1038/nrg.2017.26

4 Bomblies, K. When everything changes at once: finding a new normal after genome duplication. *Proc Biol Sci* **287**, 20202154 (2020). https://doi.org:10.1098/rspb.2020.2154

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

6 Doyle, J. J. & Coate, J. E. Polyploidy, the Nucleotype, and Novelty: The Impact of Genome Doubling on the Biology of the Cell. *International Journal of Plant Sciences* **180**, 1-52 (2019). https://doi.org:10.1086/700636

7 Yant, L. *et al.* Meiotic adaptation to genome duplication in Arabidopsis arenosa. *Current Biology* **23**, 2151-2156 (2013). https://doi.org:10.1016/j.cub.2013.08.059

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

9 Seear, P. J. *et al.* A novel allele of ASY3 is associated with greater meiotic stability in autotetraploid Arabidopsis lyrata. *PLoS Genetics* **16**, e1008900 (2020). https://doi.org:10.1371/journal.pgen.1008900

10 Morgan, C., Knight, E. & Bomblies, K. The meiotic cohesin subunit REC8 contributes to multigenic adaptive evolution of autopolyploid meiosis in Arabidopsis arenosa. *PLoS Genetics* **18**, e1010304 (2022).

11 Morgan, C. *et al.* Evolution of crossover interference enables stable autopolyploidy by ensuring pairwise partner connections in Arabidopsis arenosa. *Current Biology* **31**, 4713-4726 e4714 (2021). https://doi.org:10.1016/j.cub.2021.08.028

12 Morgan, C., Zhang, H., Henry, C. E., Franklin, F. C. H. & Bomblies, K. Derived alleles of two axis proteins affect meiotic traits in autotetraploid Arabidopsis arenosa. *Proceedings of the National Academy of Sciences* **117**, 8980-8988 (2020).

13 Bohutinska, M. *et al.* Novelty and Convergence in Adaptation to Whole Genome Duplication. *Molecular Biology and Evolution* **38**, 3910-3924 (2021). https://doi.org:10.1093/molbev/msab096

14 Wolf, E., Gaquerel, E., Scharmann, M., Yant, L. & Koch, M. A. Evolutionary footprints of a cold relic in a rapidly warming world. *Elife* **10**, e71572 (2021).

15 Gill JJB, M. H., Fearn, GM Cytotaxonomic studies on the Cochlearia officinalis L. group from inland stations in Britain. *Watsonia* **12**, 15-21 (1978).

16 Gill, J. Cytogenetic studies in Cochlearia L.(Cruciferae). The origins of C. officinalis L. and C. micacea Marshall. *Genetica* **44**, 217-234 (1973).

17 Gill, J. Cytogenetic studies in Cochlearia L. The chromosomal homogeneity within both the 2 n= 12 diploids and the 2 n= 14 diploids and the cytogenetic relationship between the two chromosome levels. *Annals of Botany* **35**, 947-956 (1971).

18 Marburger, S. *et al.* Interspecific introgression mediates adaptation to whole genome duplication. *Nature Communications* **10**, 5218 (2019). https://doi.org:10.1038/s41467-019-13159-5

19 Arnold, B., Kim, S.-T. & Bomblies, K. Single geographic origin of a widespread autotetraploid Arabidopsis arenosa lineage followed by interploidy admixture. *Molecular Biology and Evolution* **32**, 1382-1395 (2015).

20 Fearn, G. A morphological and cytological investigation of Cochlearia populations on the Gower Peninsula, Glamorgan. *New Phytologist* **79**, 455-458 (1977).
21 Stace, C. A. Hybridization and the Flora of the British Isles. *Botanical Society of the British Isles* (1975).
22 Koch, M. Genetic differentiation and speciation in prealpine Cochlearia: Allohexaploid Cochlearia bavarica Vogt (Brassicaceae) compared to its diploid ancestor Cochlearia pyrenaica DC. in Germany and Austria. *Plant Systematics and Evolution* **232**, 35-49 (2002).
23 Koch, M. Zur Ausbreitung des Dänischen Löffelkrautes (Cochlearia danica L.) als Küstensippe in das niedersächsische Binnenland. *Florist. Rundbr* **30**, 20-23 (1996).
24 Koch, M. Kurznotiz zur südlichen Ausbreitung des Dänischen Löffelkrauts (Cochlearia danica L.) in Nordrhein-Westfalen. *Flor. Rundbr* **30**, 136-138 (1997).
25 Koch, M., Dobeš, C., Bernhardt, K.-G. & Kochjarová, J. Cochlearia macrorrhiza (Brassicaceae): A bridging species between Cochlearia taxa from the Eastern Alps and the Carpathians? *Plant Systematics and Evolution* **242**, 137-147 (2003).
26 Koch, M., Hurka, H. & Mummenhoff, K. Chloroplast DNA restriction site variation and RAPD-analyses in Cochlearia (Brassicaceae): Biosystematics and speciation. *Nordic Journal of Botany* **16**, 585-603 (1996).
27 Koch, M., Huthmann, M. & Hurka, H. Isozymes, speciation and evolution in the polyploid complex Cochlearia L.(Brassicaceae). *Botanica Acta* **111**, 411-425 (1998).
28 Koch, M., Mummenhoff, K. & Hurka, H. Molecular phylogenetics of Cochlearia (Brassicaceae) and allied genera based on nuclear ribosomal ITS DNA sequence analysis contradict traditional concepts of their evolutionary relationship. *Plant Systematics and Evolution* **216**, 207-230 (1999).
29 Koch, M. A. Mid-Miocene divergence of Ionopsidium and Cochlearia and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). *Taxon* **61**, 76-92 (2012).
30 Koch, M. A. *et al.* The Quaternary evolutionary history of Bristol rock cress (Arabis scabra, Brassicaceae), a Mediterranean element with an outpost in the north-western Atlantic region. *Annals of Botany* **126**, 103-118 (2020).
31 Fekete, R., Mesterhazy, A., Valko, O. & Molnár V, A. A hitchhiker from the beach: the spread of the maritime halophyte Cochlearia danica along salted continental roads. *Preslia* **90**, 23-37 (2018).
32 Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540-546 (2019).
33 Chen, Y. *et al.* Fast and accurate assembly of Nanopore reads via progressive error correction and adaptive read selection. *BioRxiv*, 2020.2002. 2001.930107 (2020).
34 Medaka (Available from: https://github.com/nanoporetech/medaka).
35 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* **9**, e112963 (2014).
36 Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. *F1000Research* **6**, 1287 (2017).
37 Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**, 4647-4654 (2021). https://doi.org:10.1093/molbev/msab199
38 Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics* **3**, lqaa108 (2021).
39 Huson, D. H. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68-73 (1998).

40 Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197**, 573-589 (2014).
41 Chao, D. Y. *et al.* Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis. *Science* **341**, 658-659 (2013). https://doi.org:10.1126/science.1240561
42 Nawaz, I., Iqbal, M., Bliek, M. & Schat, H. 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 (2017). https://doi.org:10.1016/j.scitotenv.2017.01.111
43 Pegtel, D. M. Effect of ploidy level on fruit morphology, seed germination and juvenile growth in scurvy grass (Cochlearia officinalis L. sl, Brassicaceae). *Plant Species Biology* **14**, 201-215 (1999).
44 Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology* **23**, 3133-3157 (2014).
45 Ronfort, J., Jenczewski, E., Bataillon, T. & Rousset, F. Analysis of population structure in autotetraploid species. *Genetics* **150**, 921-930 (1998).
46 Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583-589 (1992).
47 Nei, M. Genetic distance between populations. *The American Naturalist* **106**, 283-292 (1972).
48 Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370 (1984).
49 Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome research* **23**, 1514-1521 (2013).
50 Szpak, M. *et al.* FineMAV: prioritizing candidate genetic variants driving local adaptations in human populations. *Genome biology* **19**, 1-18 (2018).
51 Craske, B. & Welburn, J. P. Leaving no-one behind: how CENP-E facilitates chromosome alignment. *Essays in biochemistry* **64**, 313-324 (2020).
52 Yoshizawa, K. *et al.* Tetraploidy-linked sensitization to CENP-E inhibition in human cells. *Molecular oncology* (2023).
53 Pisa, R., Phua, D. Y. Z. & Kapoor, T. M. Distinct Mechanisms of Resistance to a CENP-E Inhibitor Emerge in Near-Haploid and Diploid Cancer Cells. *Cell Chemical Biology* **27**, 850-857 e856 (2020). https://doi.org:10.1016/j.chembiol.2020.05.003
54 Real, A. M., Marsiglia, W. M. & Dar, A. C. Ploidy Leads a Molecular Motor to Walk Different Paths to Drug Resistance. *Cell Chemical Biology* **27**, 770-772 (2020). https://doi.org:10.1016/j.chembiol.2020.06.019
55 Isono, K., Yamamoto, H., Satoh, K. & Kobayashi, H. An Arabidopsis cDNA encoding a DNA-binding protein that is highly similar to the DEAH family of RNA/DNA helicase genes. *Nucleic acids research* **27**, 3728-3735 (1999).
56 Kim, J. S. *et al.* Arabidopsis TBP-ASSOCIATED FACTOR 12 ortholog NOBIRO6 controls root elongation with unfolded protein response cofactor activity. *Proceedings of the National Academy of Sciences* **119** (2022). https://doi.org:10.1073/pnas.2120219119
57 Rolly, N. K. & Yun, B. W. Regulation of Nitrate (NO(3)) Transporters and Glutamate Synthase-Encoding Genes under Drought Stress in Arabidopsis: The Regulatory Role of AtbZIP62 Transcription Factor. *Plants* **10** (2021). https://doi.org:10.3390/plants10102149
58 Fellmeth, J. E. & McKim, K. S. Meiotic CENP-C is a shepherd: bridging the space between the centromere and the kinetochore in time and space. *Essays Biochem* **64**, 251-261 (2020). https://doi.org:10.1042/EBC20190080
59 Ogura, Y., Shibata, F., Sato, H. & Murata, M. Characterization of a CENP-C homolog in Arabidopsis thaliana. *Genes & genetic systems* **79**, 139-144 (2004).

60 Hamasaki, H. *et al.* SD3, an Arabidopsis thaliana homolog of TIM21, affects intracellular ATP levels and seedling development. *Molecular Plant* **5**, 461-471 (2012). https://doi.org:10.1093/mp/ssr088

61 Municio, C. *et al.* The Arabidopsis condensin CAP-D subunits arrange interphase chromatin. *New Phytologist* **230**, 972-987 (2021). https://doi.org:10.1111/nph.17221

62 Sakamoto, T. *et al.* Condensin II alleviates DNA damage and is essential for tolerance of boron overload stress in Arabidopsis. *Plant Cell* **23**, 3533-3546 (2011). https://doi.org:10.1105/tpc.111.086314

63 Wood, J. L., Liang, Y., Li, K. & Chen, J. Microcephalin/MCPH1 associates with the condensin II complex to function in homologous recombination repair∗. *Journal of Biological Chemistry* **283**, 29586-29592 (2008).

64 Busoms, S. *et al.* Fluctuating selection on migrant adaptive sodium transporter alleles in coastal Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **115**, E12443-E12452 (2018). https://doi.org:10.1073/pnas.1816964115

65 Barrada, A. *et al.* A TOR-YAK1 signaling axis controls cell cycle, meristem activity and plant growth in Arabidopsis. *Development* **146**, dev171298 (2019).

66 Kim, D., Ntui, V. O. & Xiong, L. Arabidopsis YAK 1 regulates abscisic acid response and drought resistance. *FEBS letters* **590**, 2201-2209 (2016).

67 Komaki, S. *et al.* Functional Analysis of the Plant Chromosomal Passenger Complex. *Plant Physiol* **183**, 1586-1599 (2020). https://doi.org:10.1104/pp.20.00344

68 Kirioukhova, O. *et al.* Female gametophytic cell specification and seed development require the function of the putative Arabidopsis INCENP ortholog WYRD. *Development* **138**, 3409-3420 (2011). https://doi.org:10.1242/dev.060384

69 Bomblies, K., Jones, G., Franklin, C., Zickler, D. & Kleckner, N. The challenge of evolving stable polyploidy: could an increase in “crossover interference distance” play a central role? *Chromosoma* **125**, 287-300 (2016).

70 Diaz, K. J. *A novel kinesin, Arabidopsis thaliana Centromere associated protein E2, is necessary for the progression of mitosis and meiosis*. (University of Massachusetts Boston, 2014).

71 Iemura, K. & Tanaka, K. Chromokinesin Kid and kinetochore kinesin CENP-E differentially support chromosome congression without end-on attachment to microtubules. *Nature Communications* **6**, 6447 (2015). https://doi.org:10.1038/ncomms7447

72 She, Z. Y. *et al.* Kinesin-7 CENP-E regulates chromosome alignment and genome stability of spermatogenic cells. *Cell Death Discov* **6**, 25 (2020). https://doi.org:10.1038/s41420-020-0261-8

73 Wood, K. W. *et al.* Antitumor activity of an allosteric inhibitor of centromere-associated protein-E. *Proceedings of the National Academy of Sciences* **107**, 5839-5844 (2010). https://doi.org:10.1073/pnas.0915068107

74 Matsuo, K. & Tamaoki, N. Rational design and development of a lit-active photoswitchable inhibitor targeting CENP-E. *Org Biomol Chem* **19**, 6979-6984 (2021). https://doi.org:10.1039/d1ob01332g

75 Marx, A., Hoenger, A. & Mandelkow, E. Structures of kinesin motor proteins. *Cell motility and the cytoskeleton* **66**, 958-966 (2009).

76 Mirzaa, G. M. *et al.* Mutations in CENPE define a novel kinetochore-centromeric mechanism for microcephalic primordial dwarfism. *Human genetics* **133**, 1023-1039 (2014).

77 Yen, T. J., Li, G., Schaar, B. T., Szilak, I. & Cleveland, D. W. CENP-E is a putative kinetochore motor that accumulates just before mitosis. *Nature* **359**, 536-539 (1992).

78 Huang, Y. *et al.* BubR1 phosphorylates CENP-E as a switch enabling the transition from lateral association to end-on capture of spindle microtubules. *Cell Research* **29**, 562-578 (2019).

79 Kim, Y., Holland, A. J., Lan, W. & Cleveland, D. W. Aurora kinases and protein phosphatase 1 mediate chromosome congression through regulation of CENP-E. *Cell* **142**, 444-455 (2010).
80 Nousiainen, M., Silljé, H. H., Sauer, G., Nigg, E. A. & Körner, R. Phosphoproteome analysis of the human mitotic spindle. *Proceedings of the National Academy of Sciences* **103**, 5391-5396 (2006).
81 Espeut, J. *et al.* Phosphorylation relieves autoinhibition of the kinetochore motor CENP-E. *Molecular cell* **29**, 637-643 (2008).
82 Killinger, K. *et al.* Auto-inhibition of Mif2/CENP-C ensures centromere-dependent kinetochore assembly in budding yeast. *The EMBO journal* **39**, e102938 (2020). https://doi.org:10.15252/embj.2019102938
83 Vagnarelli, P. & Earnshaw, W. C. Chromosomal passengers: the four-dimensional regulation of mitotic events. *Chromosoma* **113**, 211-222 (2004).
84 Smith, S. J., Osman, K. & Franklin, F. C. H. The condensin complexes play distinct roles to ensure normal chromosome morphogenesis during meiotic division in A rabidopsis. *The Plant Journal* **80**, 255-268 (2014).
85 Sakamoto, T., Sugiyama, T., Yamashita, T. & Matsunaga, S. Plant condensin II is required for the correct spatial relationship between centromeres and rDNA arrays. *Nucleus* **10**, 116-125 (2019).
86 Bundock, P. & Hooykaas, P. An Arabidopsis hAT-like transposase is essential for plant development. *Nature* **436**, 282-284 (2005).
87 Chang, W.-C. *et al.* Regulation of Ku gene promoters in Arabidopsis by hormones and stress. *Functional Plant Biology* **35**, 265-280 (2008).
88 Tamura, K., Adachi, Y., Chiba, K., Oguchi, K. & Takahashi, H. Identification of Ku70 and Ku80 homologues in Arabidopsis thaliana: evidence for a role in the repair of DNA double-strand breaks. *The Plant Journal* **29**, 771-781 (2002).
89 Li, S., Ding, B., Chen, R., Ruggiero, C. & Chen, X. Evidence that the transcription elongation function of Rpb9 is involved in transcription-coupled DNA repair in Saccharomyces cerevisiae. *Molecular and cellular biology* **26**, 9430-9441 (2006).
90 Hull, M. W., McKune, K. & Woychik, N. A. RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes & Development* **9**, 481-490 (1995).
91 Tan, E. H., Blevins, T., Ream, T. S. & Pikaard, C. S. Functional consequences of subunit diversity in RNA polymerases II and V. *Cell reports* **1**, 208-214 (2012).
92 Hemming, S. A. *et al.* RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *Journal of Biological Chemistry* **275**, 35506-35511 (2000).
93 Nesser, N. K., Peterson, D. O. & Hawley, D. K. RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proceedings of the National Academy of Sciences* **103**, 3268-3273 (2006).
94 Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nature reviews Molecular cell biology* **16**, 155-166 (2015).
95 Baxter, I. *et al.* A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1; 1. *PLoS Genetics* **6**, e1001193 (2010).
96 Shi, H., Ishitani, M., Kim, C. & Zhu, J.-K. The Arabidopsis thaliana salt tolerance gene SOS1 encodes a putative Na+/H+ antiporter. *Proceedings of the National Academy of Sciences* **97**, 6896-6901 (2000).
97 Zhou, M. & Wang, W. SOS1 safeguards plant circadian rhythm against daily salt fluctuations. *Proceedings of the National Academy of Sciences* **119**, e2212950119 (2022). https://doi.org:10.1073/pnas.2212950119
98 Busoms, S., Fischer, S. & Yant, L. Chasing the mechanisms of ecologically adaptive salinity tolerance. *Plant Communications* (2023). https://doi.org:https://doi.org/10.1016/j.xplc.2023.100571

99 Furini, S. & Domene, C. K+ and Na+ conduction in selective and nonselective ion channels via molecular dynamics simulations. *Biophysical journal* **105**, 1737-1745 (2013).
100 Cotsaftis, O., Plett, D., Shirley, N., Tester, M. & Hrmova, M. A two-staged model of Na+ exclusion in rice explained by 3D modeling of HKT transporters and alternative splicing. *PloS one* **7**, e39865 (2012).
101 Almeida, P., Katschnig, D. & De Boer, A. H. HKT transporters—state of the art. *International Journal of Molecular Sciences* **14**, 20359-20385 (2013).
102 Wang, Y. *et al.* Architecture and autoinhibitory mechanism of the plasma membrane Na+/H+ antiporter SOS1 in Arabidopsis. *Nature Communications* **14**, 4487 (2023).
103 Pandey, A. K. *et al.* An Inherent Difference between Serine and Threonine Phosphorylation: Phosphothreonine Strongly Prefers a Highly Ordered, Compact, Cyclic Conformation. *ACS Chemical Biology* (2023).
104 Kant, P., Kant, S., Gordon, M., Shaked, R. & Barak, S. STRESS RESPONSE SUPPRESSOR1 and STRESS RESPONSE SUPPRESSOR2, two DEAD-box RNA helicases that attenuate Arabidopsis responses to multiple abiotic stresses. *Plant physiology* **145**, 814-830 (2007).
105 Khan, A. *et al.* The A rabidopsis STRESS RESPONSE SUPPRESSOR DEAD-box RNA helicases are nucleolar-and chromocenter-localized proteins that undergo stress-mediated relocalization and are involved in epigenetic gene silencing. *The Plant Journal* **79**, 28-43 (2014).
106 Kim, J. S., Kim, K. A., Oh, T. R., Park, C. M. & Kang, H. Functional characterization of DEAD-box RNA helicases in Arabidopsis thaliana under abiotic stress conditions. *Plant and cell physiology* **49**, 1563-1571 (2008).
107 Merlot, S. *et al.* Constitutive activation of a plasma membrane H+-ATPase prevents abscisic acid-mediated stomatal closure. *The EMBO journal* **26**, 3216-3226 (2007).
108 Nadeau, J. A. & Sack, F. D. Control of stomatal distribution on the Arabidopsis leaf surface. *Science* **296**, 1697-1700 (2002).
109 Sugano, S. S. *et al.* Stomagen positively regulates stomatal density in Arabidopsis. *Nature* **463**, 241-244 (2010).
110 Emms, D. M. & Kelly, S. OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *BioRxiv* **466201** (2018).
111 Sterken, R. *et al.* Combined linkage and association mapping reveals CYCD5; 1 as a quantitative trait gene for endoreduplication in Arabidopsis. *Proceedings of the National Academy of Sciences* **109**, 4678-4683 (2012).
112 Unhavaithaya, Y. & Orr-Weaver, T. L. Centromere proteins CENP-C and CAL1 functionally interact in meiosis for centromere clustering, pairing, and chromosome segregation. *Proceedings of the National Academy of Sciences* **110**, 19878-19883 (2013).
113 Kurzbauer, M.-T. *et al.* ATM controls meiotic DNA double-strand break formation and recombination and affects synaptonemal complex organization in plants. *The Plant Cell* **33**, 1633-1656 (2021).
114 Bonnet, S., Knoll, A., Hartung, F. & Puchta, H. Different functions for the domains of the Arabidopsis thaliana RMI1 protein in DNA cross-link repair, somatic and meiotic recombination. *Nucleic acids research* **41**, 9349-9360 (2013).
115 Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47**, D607-D613 (2019).
116 Eisenschmid, K., Jabbusch, S. & Koch, M. A. Evolutionary footprints of cold adaptation in arctic-alpine Cochlearia (Brassicaceae)–Evidence from freezing experiments and electrolyte leakage. *Perspectives in Plant Ecology, Evolution and Systematics* **59**, 125728 (2023).

1052 117 Desjardins, S., Kanyuka, K. & Higgins, J. D. A cytological analysis of wheat meiosis
1053 targeted by virus-induced gene silencing (VIGS). *Plant Meiosis: Methods and Protocols*,
1054 319-330 (2020).
1055 118 Higgins, J. D., Sanchez-Moran, E., Armstrong, S. J., Jones, G. H. & Franklin, F. C. H. The
1056 Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis
1057 and normal fidelity of crossing over. *Genes & Development* **19**, 2488-2500 (2005).
1058 119 Desjardins, S. D. *et al.* MutS homologue 4 and MutS homologue 5 maintain the obligate
1059 crossover in wheat despite stepwise gene loss following polyploidization. *Plant Physiology*
1060 **183**, 1545-1558 (2020).
1061 120 Mandáková, T. & Lysak, M. A. Chromosome preparation for cytogenetic analyses in
1062 Arabidopsis. *Current protocols in plant biology* **1**, 43-51 (2016).
1063 121 Ijdo, J., Wells, R., Baldini, A. & Reeders, S. Improved telomere detection using a telomere
1064 repeat probe (TTAGGG) n generated by PCR. *Nucleic acids research* **19**, 4780 (1991).
1065 122 Mandáková, T. & Lysak, M. A. Painting of Arabidopsis chromosomes with chromosome-
1066 specific BAC clones. *Current protocols in plant biology* **1**, 359-371 (2016).
1067 123 Deorowicz, S., Debudaj-Grabysz, A. & Grabowski, S. Disk-based k-mer counting on a PC.
1068 *BMC bioinformatics* **14**, 1-12 (2013).
1069 124 Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and
1070 Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* **11**,
1071 1432 (2020). https://doi.org:10.1038/s41467-020-14998-3
1072 125 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
1073 sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
1074 126 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1075 *arXiv* (2013).
1076 127 Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields
1077 chromosome-length scaffolds. *Science* **356**, 92-95 (2017).
1078 128 Durand, N. C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with
1079 unlimited zoom. *Cell systems* **3**, 99-101 (2016).
1080 129 Novák, P., Neumann, P., Pech, J., Steinhaisl, J. & Macas, J. RepeatExplorer: a Galaxy-
1081 based web server for genome-wide characterization of eukaryotic repetitive elements from
1082 next-generation sequence reads. *Bioinformatics* **29**, 792-793 (2013).
1083 130 Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a
1084 streamlined. *BioRxiv* **657890** (2019).
1085 131 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1086 (2013).
1087 132 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler
1088 transform. *Bioinformatics* **25**, 1754-1760 (2009).
1089 133 Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-
1090 2079 (2009).
1091 134 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
1092 next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).
1093 135 Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS*
1094 *genetics* **2**, e190 (2006).
1095 136 Monnahan, P. *et al.* Pervasive population genomic consequences of genome duplication
1096 in Arabidopsis arenosa. *Nature Ecology and Evolution* **3**, 457-468 (2019).
1097 https://doi.org:10.1038/s41559-019-0807-4
1098 137 Pembleton, L. W., Cogan, N. O. & Forster, J. W. St AMPP: An R package for calculation
1099 of genetic differentiation and structure of mixed-ploidy level populations. *Molecular*
1100 *Ecology Resources* **13**, 946-952 (2013).
1101 138 Gerard, D. Pairwise linkage disequilibrium estimation for polyploids. *Molecular Ecology*
1102 *Resources* **21**, 1230-1242 (2021).

1103 139 Alexa, A. & Rahnenführer, J. Gene set enrichment analysis with topGO. *Bioconductor*
1104 *Improv* **27**, 1-26 (2009).
1105 140 Sievers, F. & Higgins, D. G. Clustal omega. *Current protocols in bioinformatics* **48**, 3.13.
1106 11-13.13. 16 (2014).
1107 141 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version
1108 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**,
1109 1189-1191 (2009).
1110 142 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,
1111 583-589 (2021).
1112