

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<https://wrap.warwick.ac.uk/195620/>

How to cite:

The repository item page linked to above, will contain details on accessing citation guidance from the publisher.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>.



Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Natural variation modifies centromere-proximal meiotic crossover frequency and segregation distortion in *Arabidopsis thaliana*

Nicola Gorringer¹, Stephanie Topp¹, Robin Burns¹, Sota Yamaguchi², Fernando A. Rabanal³, Joiselle B. Fernandes¹, Detlef Weigel³, Tetsuji Kakutani², Matthew Naish^{1*} and Ian R. Henderson^{1*}

¹ Department of Plant Sciences, Downing Street, University of Cambridge, Cambridge, United Kingdom, CB2 3EA

² Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo 153-8902, Japan

³ Department of Molecular Biology, Max Planck Institute for Biology, Tübingen, D-72076, Tübingen, Germany

* Correspondence: mn359@cam.ac.uk and irh25@cam.ac.uk

Running title: Centromeric crossover and segregation.

Keywords: Centromere, meiosis, recombination, segregation distortion, Arabidopsis, natural variation, CENH3, DNA methylation, satellite repeats.

Abstract

Eukaryotic centromeres mediate chromosome segregation during cell division. Plant centromeres are loaded with CENH3-variant nucleosomes, which direct kinetochore formation and spindle-microtubule interaction. Centromeres are frequently composed of megabase-scale satellite repeat arrays, or retrotransposon nests. In monocentric genomes, such as the model plant *Arabidopsis thaliana*, pericentromeric heterochromatin surrounds the CENH3-occupied satellite arrays. A zone of suppressed meiotic crossover recombination contains the centromere and extends into the pericentromeres. Here, we explore how natural variation in Arabidopsis influences centromere-proximal crossover frequency, and segregation distortion, when centromeres are heterozygous. We used fluorescent crossover reporters to quantify the effect of genetic variation on centromere-proximal recombination in twelve F_1 hybrids between the reference strain Col-0 and non-reference accessions that captured Eurasian and relict diversity, and in total we measured 3,037,802 meioses. The majority of the F_1 hybrids (49 of 60) had significantly higher or lower centromere-proximal crossover frequency than inbreds. We relate hybrid crossover frequencies to patterns of nucleotide diversity and centromeric structural variation, and in a subset of seven accessions, to epigenetic patterns of CENH3 enrichment and DNA methylation. Using linear modelling, we observed that chromosome and accession, and their interaction, together explained 85% of variation in crossover frequency, consistent

1 with *cis*- and *trans*-acting modifying effects. The fluorescent reporters also allow
2 segregation distortion through meiosis to be quantified between hybrids and inbreds. We
3 observed a minority of hybrids (18 of 60) with distorted segregation through meiosis
4 compared to inbreds, that occurred with or without simultaneous change to centromere-
5 proximal crossover frequency. Linear modelling revealed that 56% of variation in
6 segregation distortion is explained by chromosome and accession, but with a stronger
7 effect of accession compared to crossover frequency. We discuss how Arabidopsis
8 centromeric structural heterozygosity may modify recombination, and cause segregation
9 distortion through meiosis.

11 Introduction

13 Meiosis is a specialised eukaryotic cell division, where a single round of DNA replication
14 is coupled to two rounds of chromosome segregation, which generates a tetrad of spores,
15 each with half the number of chromosomes (Villeneuve and Hillers 2001; Mercier *et al.*
16 2015). During meiosis, interhomolog recombination is initiated during prophase I, via the
17 formation of DNA double-strand breaks (DSBs) along the length of chromosomes by
18 SPO11 complexes (Arter and Keeney 2024). Meiotic DSB formation takes place in the
19 context of a proteinaceous axis - a multiprotein scaffold that organises chromosomes into
20 arrays of tethered chromatin loops (Kleckner 2006). As recombination proceeds, the axes
21 of homologous chromosomes pair, and are held in close physical proximity via the
22 synaptonemal complex (Kleckner 2006). This bivalent arrangement enables meiotic

1 recombination to complete using the homologous chromosome as a repair template,
2 which can result in reciprocal genetic exchange, termed crossover (Villeneuve and Hillers
3 2001; Kleckner 2006). Formation of interhomolog crossovers creates new combinations
4 of genetic variation along chromosomes (Villeneuve and Hillers 2001; Mercier *et al.*
5 2015), which are additionally required for balanced homolog segregation at the first
6 meiotic division (Grelon *et al.* 2001). Centromere behaviour is modified in meiosis, such
7 that sister chromatid centromeres are mono-orientated to the same cell pole during the
8 first meiotic division, whereas they are bi-orientated to opposite cell poles during the
9 second meiotic division (Koch and Marston 2025).

10
11 The frequency of meiotic crossovers varies widely along plant chromosomes, with the
12 presence of narrow 1-2 kilobase pair (kb) hotspots where recombination is elevated over
13 the genome average (Dooner and Martínez-Férez 1997; Choi *et al.* 2013; Drouaud *et al.*
14 2013; Rommel Fuentes *et al.* 2020; Szymanska-Lejman *et al.* 2023). Plant crossover
15 hotspots are influenced by both DNA sequence and chromatin, and are often found in
16 proximity to euchromatic gene promoters and terminators (Dooner and Martínez-Férez
17 1997; Choi *et al.* 2013; Drouaud *et al.* 2013; Rommel Fuentes *et al.* 2020; Szymanska-
18 Lejman *et al.* 2023). In the model plant *Arabidopsis thaliana* (hereafter *Arabidopsis*), the
19 total number of crossovers can be increased by changes in the activity of *trans* regulators,
20 for example via overexpression of the E3 ligase HEI10, or mutation of anti-crossover
21 factors (Séguéla-Arnaud *et al.* 2015; Ziolkowski *et al.* 2017). Crossovers are also subject
22 to ‘assurance’ and ‘interference’, which respectively describe that each chromosome pair
23 receives an obligate crossover, and that double crossovers are spaced further apart than

1 expected from uniformly random placement (Jones and Franklin 2006). HEI10 coarsening
2 via diffusion along the synaptonemal complex provides a model to understand crossover
3 obligation and interference (Morgan *et al.* 2021).

4
5 Many eukaryotic genomes also contain regions where crossovers are suppressed,
6 including heterochromatin, mating-type loci, sex chromosomes, and the centromeres
7 (Vincenten *et al.* 2015; Charlesworth 2017; Nambiar and Smith 2018; Rowan *et al.* 2019;
8 Fernandes *et al.* 2024). Crossover suppression can be caused both by the inhibitory
9 effects of heterochromatic marks on recombination, as well as sequence polymorphism
10 (heterozygosity) between homologs (Charlesworth 2017; Nambiar and Smith 2018;
11 Rowan *et al.* 2019; Fernandes *et al.* 2024). For example, acquisition of DNA methylation
12 and H3K9me2 methylation can inhibit meiotic DSBs and crossovers within Arabidopsis
13 recombination hotspots located close to genes (Yelina *et al.* 2015). The Arabidopsis
14 centromeric and pericentromeric regions are densely DNA methylated compared to the
15 chromosome arms, and are crossover suppressed (Naish *et al.* 2021; Fernandes *et al.*
16 2024; Rowan *et al.* 2019). However, loss of CG versus non-CG context DNA methylation
17 in Arabidopsis has opposite effects on centromere-proximal crossover frequency.
18 Specifically, *met1* and *ddm1* (CG) are observed to reduce proximal recombination,
19 whereas combinations of *drm1*, *drm2*, *cmt2*, and *cmt3* (non-CG) increase crossovers
20 (Colomé-Tatché *et al.* 2012; Mirouze *et al.* 2012; Melamed-Bessudo and Levy 2012;
21 Yelina *et al.* 2015, 2012; Underwood *et al.* 2018; Fernandes *et al.* 2024). The
22 heterochromatic histone variant H2A.W also contributes to centromere-proximal
23 crossover suppression in Arabidopsis (Son *et al.* 2025; Kuo *et al.* 2025). Heterozygous

structural variation further has a potent suppressive effect on crossovers (Rowan *et al.* 2019), and heterochromatic regions, including the centromeres, can be highly structurally polymorphic, even between closely related *Arabidopsis* accessions (Wlodzimierz *et al.* 2023b). Observations in budding yeast and mammals indicate that centromere-proximal crossovers can be associated with higher rates of gamete aneuploidy (Koehler *et al.* 1996; Lamb *et al.* 1996; Rockmill *et al.* 2006), providing a functional reason to suppress centromere-proximal crossovers. In addition, suppression of crossovers in heterochromatic regions may reduce the chance of illegitimate events, such as unequal crossover, when repetitive templates recombine (Smith 1976).

Centromeres are defined by the loading of variant nucleosomes that contain CENP-A (CENH3 in plants) instead of histone H3, which direct kinetochore formation and microtubule binding (Talbert *et al.* 2002; McKinley and Cheeseman 2016). In *Arabidopsis*, the centromeres consist of megabase arrays of 178 base pair *CEN178* satellite repeats, which are densely DNA methylated, within which a ~1-2 Mb sub-region is CENH3-occupied (Naish *et al.* 2021; Wlodzimierz *et al.* 2023b). The *CEN178* satellite arrays are highly structurally polymorphic between *Arabidopsis* accessions (Wlodzimierz *et al.* 2023b), reflecting the centromere paradox, where centromere sequences rapidly diversify, within and between species, despite playing a conserved and essential role during chromosome segregation (Henikoff *et al.* 2001). The DNA recombination and repair mechanisms that generate centromere satellite diversity are not well understood, and may include somatic and meiotic processes (Miga and Alexandrov 2021; Showman *et al.* 2024). Meiotic crossover maps generated from *Arabidopsis* Col-0×Ler-0 hybrids

1 showed that the centromeres are contained in megabase-scale non-recombining zones,
2 with proximal crossovers occurring in euchromatic gene islands embedded in the
3 pericentromeric heterochromatin (Fernandes *et al.* 2024).

4
5 An evolutionary explanation for rapid centromere evolution is provided by the drive
6 hypothesis (Fishman and Kelly 2015; Kursel and Malik 2018; Akeru *et al.* 2019). In
7 mammals and plants, female meiosis typically produces a single megaspore/egg that
8 forms the zygote upon fertilization, while the remaining three sister gametes/spores perish
9 (Fishman and Kelly 2015; Kursel and Malik 2018; Akeru *et al.* 2019). Several genetic
10 systems exist where centromere sequence variants bias their segregation into the
11 surviving female gamete, causing segregation distortion and non-Mendelian transmission
12 through meiosis (Fishman and Saunders 2008; Chmátal *et al.* 2014; Iwata-Otsubo *et al.*
13 2017; Akeru *et al.* 2017, 2019; Kumon *et al.* 2021). Segregation distortion can also occur
14 via 'gene drive' in the male germ line, where linked toxin-antidote gene pairs kill gametes
15 that do not inherit them (Nuckolls *et al.* 2017; Berube *et al.* 2024). However, the extent to
16 which intra-species centromere polymorphism causes variation in proximal crossover
17 frequency, and the extent to which this co-occurs with segregation distortion, is not well
18 explored.

19
20 *Arabidopsis* is a powerful model for studying centromere-proximal crossover
21 recombination and segregation distortion, as extensive natural genetic variation in
22 satellite array structures exists, with the centromeres of multiple accessions having been

1 fully assembled (Rabanal *et al.* 2022; Wlodzimierz *et al.* 2023b; Kang *et al.* 2023; Lian *et*
2 *al.* 2024). Furthermore, fluorescent reporter systems are available that allow sensitive
3 quantification of crossover frequency, and segregation distortion, across the centromeres
4 (Berchowitz and Copenhaver 2008; Wu *et al.* 2015; Fernandes *et al.* 2024). In this study,
5 we generated twelve *Arabidopsis* F₁ hybrids by crossing a panel of diverse accessions to
6 a common background (Col-0) that carries centromere-proximal fluorescent transgenic
7 reporters on each chromosome, to quantify meiotic crossover frequency and segregation
8 distortion, measuring 3,037,802 meioses in total. Compared to the Col-0 inbred parental
9 strains, the majority of hybrids showed either significantly higher or lower centromere-
10 proximal crossover frequency. We relate these crossover data to centromere genetic
11 diversity and structural sequence polymorphism between the recombining chromosomes,
12 and in a subset of seven accessions that represent Eurasian and relict backgrounds, to
13 epigenetic patterns of CENH3 enrichment and DNA methylation. In a minority of hybrids,
14 we detected significant segregation distortion of centromere haplotypes through meiosis,
15 compared to the Col-0 inbred parental strain, which occurred with or without associated
16 changes to centromere-proximal crossover frequency. Using linear modelling, we
17 observe that genetic variation between the centromere being measured, and the
18 accession crossed to, and their interaction, explains ~85% of the variation in hybrid
19 crossover frequency, and 56% of the variation in segregation distortion. We discuss
20 scenarios for how centromere structural polymorphism may modify recombination
21 frequency and cause segregation distortion in *Arabidopsis*.

Materials and Methods

Arabidopsis genetic stocks and growth conditions

All Col-0 FTL transgenic lines are TRAFFIC lines obtained from Prof. Scott Poethig (University of Pennsylvania, Philadelphia) (Wu *et al.* 2015; Poethig *et al.* 2022). The ANGE-B-2 and ANGE-B-10 accessions were provided by Fabrice Roux (CNRS, Toulouse), IP-Alo-19, IP-Cat-0, and IP-Med-0 were provided by Prof. Carlos Alonso-Blanco (CNB-CSIC, Madrid), Elk-3, Rab-1, and Tanz-1 were provided by Prof. Angela Hancock (Purdue University), and Jm-0, Ler-0, Cvi-0, and Etna-2 were obtained from the European Arabidopsis Stock Centre. All plants were grown in controlled environment chambers at 20°C, 60% humidity and under a 16-hour photoperiod.

Fluorescent seed scoring and calculation of crossover frequency and segregation distortion

Seed harvested from FTL hemizygous plants were cleaned to remove plant debris and images of a seed monolayer were captured using a Leica DFC310 FX dissecting microscope, using brightfield, or with ultraviolet filters. Images were analysed using an adapted CellProfiler pipeline (version 4.2.5) (Stirling *et al.* 2021), which identifies seeds as primary objects and measures fluorescence intensity for each object. Thresholds for fluorescence intensity were manually set for each group of seed images, following inspection of the seed fluorescence intensity histograms. FTL crossover frequency (centiMorgan, cM) was calculated using the formula:

$$cM = 100 \times (1 - [1 - 2(N_G + N_R)/N_T]^{1/2}),$$

where N_G is the number of green-alone fluorescent seeds, N_R is the number of red-alone fluorescent seeds, and N_T is the total number of seeds counted (Ziolkowski *et al.* 2015). To test for significant differences in crossover frequency between the Col-0 inbred controls and each F_1 hybrid, the N_R and N_G counts were summed across replicates for each genotype, and combined with the remainder of total counts to construct 2×2 contingency tables for Chi-square tests. To test for significant differences in segregation distortion between the Col-0 inbred controls and each F_1 hybrid, we compared the fluorescent and non-fluorescent seed counts to inbreds, using 2×2 contingency tables and Chi-square tests.

Jm-0 and Elk-3 long-read DNA sequencing, genome assembly and annotation

A single Jm-0 individual was grown at 23°C under 16 hours of light and harvested 26 days after germination. 300 mg of tissue powder was used for high-molecular weight genomic DNA extraction. The methods for DNA extraction, PacBio HiFi library preparation, genome assembly using hifiasm (Version 0.16.1-r375; (Cheng *et al.* 2021)), and quality assessment were as performed for the other ten HiFi-based assemblies used in this study (Włodzimierz *et al.* 2023b).

For Elk-3 genomic DNA extraction associated with ONT sequencing, seedlings were grown on ½ MS media and 1% sucrose. Three-week-old seedlings were moved to the dark for 48 hours prior to harvesting. Approximately 2 g of tissue was used per genomic

DNA extraction. Tissue was flash frozen and ground in liquid nitrogen, using a pestle and mortar for approximately 20 minutes. DNA was extracted using NucleoBond HMW DNA kit (Macherey-Nagel, Düren, Germany), as per the manufacturer's instructions. Library preparation followed the Nanopore SQK-LSK114 protocol and kit with the following modifications; approximately 1.2-1.5 µg of size-selected DNA in a volume of 48 µl was used for library preparation, the first bead clean-up after nic-repair and end-prep were replaced by Short Read Eliminator (SRE) size selection (PacBio), and libraries were sequenced in a R10.4.1 Promethion flowcell, with base-calling and DNA methylation quantification performed using dorado (v0.8.3.) with model dna_r10.4.1_e8.2_400bps_sup@v5.0.0 and --modified-bases 5mC_5hmC.

For genome assembly, reads were filtered for length and accuracy using Filtlong (Version 0.2.0) (--min_mean_q 90, --min_length 50,000), and assembled using Hifiasm (Version 0.23.0-r691), using --ont parameters (Cheng *et al.* 2021). Chromosomes were generated as single contigs and extracted and oriented using Ragtag (Version v2.1.0) scaffolding to the Col-CEN assembly (Naish *et al.* 2021; Alonge *et al.* 2022), and polished using medaka (version 2.0.1). For DNA methylation, reads were filtered for a minimum length of 30 kb and aggregated using Modkit pileup (Version 0.4.1), using the parameters --ignore h --mod-thresholds C:0.85.

Copies of the centromeric satellite repeat were identified using TRASH, providing a template of the *CEN178* consensus sequence (Włodzimierz *et al.* 2023a; b). To compare

centromere structural and repeat polymorphism, we used ReDotable to generate sequence identity dot-plots, as well as SyRI (Goel *et al.* 2019). To map the location of syntenic insertion sites for the FTL T-DNAs in the genome assemblies of the non-reference accessions, we extracted 20 kb of sequence upstream and downstream of the insertion site from the Col-CEN assembly. These sequences were aligned to each of the accession genomes using LASTZ to identify the syntenic position in each genome. Protein sequences of the CENH3 coding region were extracted from the 13 studied *A. thaliana* accessions, and *A. lyrata* (accession MN47) (Wlodzimierz *et al.* 2023b), which encodes two CENH3 paralogues (A1G11082 and A1G59290) (Kawabe *et al.* 2006). MAFFT (<https://github.com/GSLBiotech/mafft>) was used to align the protein sequences and snipit (<https://github.com/aineniarnh/snipit>) was used for data visualisation.

Calculation of pairwise genetic diversity (π) in FTL regions

Genomes from each accession used in the crosses were aligned to the Col-CEN assembly (Naish *et al.* 2021), using Winnowmap (Version 2.03) with a k-mer size of 19 (Jain *et al.* 2022). Sequence variant calling was performed using BCFtools (Version 1.21) (Danecek *et al.* 2021). Repetitive regions in the Col-CEN assembly were masked using RepeatMasker and TRASH, including centromeric satellites, transposable elements, organelle sequences, and rDNA (Wlodzimierz *et al.* 2023a). Single nucleotide polymorphisms (SNPs) were filtered to exclude those overlapping repeat-masked regions, to retain strictly biallelic variants, and to limit missing data to no more than 20% across the examined accessions. Pairwise nucleotide diversity (π) relative to Col-CEN

was calculated for each accession, within the positions of the FTL T-DNA markers on each chromosome.

Linear models of crossover frequency and segregation distortion

As each FTL interval varies in physical size, crossover frequency was normalised by interval length based on the Col-0 genome (centiMorgans per megabases (cM/Mb)). We modelled the variation in crossover frequency using the following formula:

$$\text{cM/Mb} \sim \text{chromosome} * \text{accession}$$

Where ‘chromosome’ denotes the specific FTL assayed (*CEN1-5*), and ‘accession’ is used as the natural accession crossed to Col-0. A linear mixed effect model using “replicate” as a random variable was used to control for experimental replication. The variance explained by each term, and their interactions, was calculated using Type III ANOVA, and the Sum of Squares used to estimate partial variance of each variable. Additional linear models were performed to test whether the distortion explained crossover frequency, in addition to the annotation features FTL interval size (for each accession), gene density, transposable element density, and *CEN178* satellite repeat density, which were added in a step-wise manner to the above model to determine if they would increase the proportion of variance of crossover frequency explained. Variation in

segregation distortion was modelled using the same structure, where the ratio of inheritance for the red or green fluorescent transgenes was used as the response variable. The variance in distortion explained by each term was determined as above and the same annotation features were included in models in a step-wise manner to test if they would increase the amount of variation explained.

CENH3 ChIP-seq

Approximately 4 g of 2-week-old seedlings were ground in liquid nitrogen. Nuclei were isolated in nuclei isolation buffer (1 M sucrose, 60 mM HEPES pH 8.0, 0.6% Triton X-100, 5 mM KCl, 5 mM MgCl₂, 5 mM EDTA, 0.4 mM PMSF, 1 mM pepstatin-A, 1×protease inhibitor cocktail) and crosslinked in 1% formaldehyde at room temperature for 25 minutes. The crosslinking reaction was quenched with 125 mM glycine and incubated at room temperature for a further 25 minutes. The nuclei were purified from cellular debris via two rounds of filtration through one layer of Miracloth and centrifuged at 2,500 g for 25 minutes at 4°C. The nuclei pellet was resuspended in EB2 buffer (0.25 M sucrose, 1% Triton X-100, 10 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 1 mM EDTA, 5 mM DTT, 0.1 mM PMSF, 1 mM pepstatin-A, 1×protease inhibitor cocktail) and centrifuged at 14,000 g for 10 minutes at 4°C. The pellet with the nuclei was resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 1% SDS, 10 mM EDTA, 0.1 mM PMSF, 1 mM pepstatin-A) and was sonicated using a Covaris E220 evolution with the following settings: power=150V, bursts per cycle=200, duty factor=20%, time=60 seconds.

Sonicated chromatin was centrifuged at 14,000 g and the supernatant was extracted and diluted with 1×volume of ChIP dilution buffer (1.1% Triton X-100, 20 mM Tris-HCl pH 8.0, 167 mM NaCl, 1.1 mM EDTA, 1 mM pepstatin-A, 1×protease inhibitor cocktail). The chromatin was incubated overnight at 4°C with 50 µl Protein A magnetic beads (Dynabeads, Thermo Fisher), pre-bound with 5 µl α-CENH3 raised to peptide acetyl-RTK HRV TRS QPR NQT DAC-amide (Eurogentec). The beads were collected on a magnetic rack and washed twice with low-salt wash buffer (150 mM NaCl, 0.1% SDS, 1% Triton X-100, 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.4 mM PMSF, 1 mM pepstatin-A, 1×protease inhibitor cocktail) and twice with high-salt wash buffer (500 mM NaCl, 0.1% SDS, 1% Triton X-100, 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.4 mM PMSF, 1 mM pepstatin-A, 1×protease inhibitor cocktail). Immunoprecipitated DNA–protein complexes were eluted from the beads (1% SDS, 0.1 M NaHCO₃) at 65°C for 15 minutes. Samples were reverse crosslinked by incubating with 0.24 M NaCl at 65°C overnight. Proteins and RNA were digested with Proteinase K treatment, and RNase A, and DNA was purified by phenol:chloroform:isoamyl alcohol (25:24:1) extraction and ethanol precipitation. Library preparation followed the Tecan Ovation Ultralow System v2 library protocol. ChIP samples were PCR amplified for 12 cycles and sequenced with 150-bp paired-end reads on the Illumina platform.

ChIP–seq data alignment, processing, and analysis

Deduplicated paired-end CENH3 ChIP–seq Illumina reads (2×150 bp) from Col-0, Cvi-0, Ler-0, Jm-0, Rab-1, Elk-3, and Tanz-1 were processed with Cutadapt (v.1.18) to remove adaptor sequences and low-quality bases (Phred+33-scaled quality <20). For each

accession, trimmed reads were aligned to the respective genome assembly using Bowtie2 (v.2.3.4.3), with the following settings: `--very-sensitive --no-mixed --no-discordant -k 200 --maxins 1000`. Read pairs with Bowtie2-assigned MAPQ of <10 were discarded using samtools (v.1.10). For retained read pairs that aligned to multiple locations, with varying alignment scores, the best alignment was selected. Alignments with more than two mismatches or consisting of only one read in a pair were discarded. For each dataset, bins per million mapped reads (BPM; equivalent to transcripts per million for RNA-seq data) coverage values were generated in bigWig and bedGraph formats with the bamCoverage tool from deepTools (v.3.5.0) and plotted using R. To define regions of CENH3 enrichment the start boundary was defined as the first incidence of three consecutive 10 kb windows which were greater than a value of $\log_2(\text{ChIP}/\text{Input})$ greater than 0.5, and the end boundary was defined by the first occurrence of three consecutive 10 kb windows were less than this value.

DNA methylation profiling and analysis

DNA base-calling, alignment, and methylation calling were performed using dorado (v0.8.3.), and models `dna_r10.4.1_e8.2_400bps_sup@v5.0.0` and `--modified-bases 5mC_5hmC`. The reads were filtered for a minimum length of 30 kb, and aggregated using Modkit pileup (Version 0.4.1), using parameters (`--ignore h --mod-thresholds C:0.85`). The average methylation for each DNA sequence context was calculated over 10 kb windows, using R.

Results

Quantifying centromere-proximal meiotic crossover frequency and segregation distortion using linked fluorescent reporter transgenes in Arabidopsis

In Arabidopsis, segregation of linked, hemizygous Fluorescent-Tagged Line (FTL) T-DNAs that express different colours of fluorescent protein in the pollen, or non-maternal seed tissues, can be used to quantify meiotic crossover frequency between the transgene insertions (Melamed-Bessudo *et al.* 2005; Wu *et al.* 2015; Poethig *et al.* 2022). Based on the locations of the centromeric *CEN178* satellite arrays in the Col-CEN genome assembly (Naish *et al.* 2021), we selected pairs of FTL T-DNAs generated in the Col-0 accession that define physical intervals that span each of the five centromeres (**Fig. 1A**, **Fig. S1** and **Table S1**) (Wu *et al.* 2015; Poethig *et al.* 2022). In the Col-CEN assembly, the FTL intervals range from 4.36 to 7.76 megabases (Mb) (**Fig. 1A**, **Fig. S1** and **Table S1**). In addition to the *CEN178* satellite arrays, the FTL *CEN* intervals contain varying amounts of pericentromeric heterochromatin, 5S rDNA copies, genes, and transposable elements (**Fig. 1A**, **Fig. S1** and **Table S1**) (Naish *et al.* 2021). When FTL T-DNAs are linked and hemizygous, and the plant is self-fertilized, the number of seeds with each fluorescent phenotype (red and green: red-alone: green-alone: no colour) can be used to quantify meiotic crossover frequency (**Fig. 1C-1D**) (Melamed-Bessudo *et al.* 2005; Francis *et al.* 2007; Wu *et al.* 2015; Poethig *et al.* 2022). If FTL hemizygous plants are self-fertilized, both male and female meiotic crossover are assayed (Melamed-Bessudo

1 *et al.* 2005; Francis *et al.* 2007; Wu *et al.* 2015; Poethig *et al.* 2022). In Col-0 inbreds, the
2 centromere-spanning FTL intervals have recombination rates in the range of 0.86-2.00
3 cM/Mb (**Table S1**), which is lower than the chromosome arm average of 3.80 cM/Mb
4 measured from Col-0×Ler-0 hybrids (Fernandes *et al.* 2024), consistent with centromere-
5 proximal suppression of crossover frequency.

6
7 FTL T-DNAs can also be used to quantify segregation distortion between generations.
8 When an FTL hemizygote undergoes meiosis, the Mendelian expectation is that exactly
9 half of the gametes will carry each T-DNA (**Fig. 1C** and **1E**). As our F₂ seed samples
10 represent self-fertilization of male and female meiotic products, the Mendelian
11 expectation is that two thirds of the seeds should show fluorescence for either T-DNA;
12 i.e., two thirds of the seeds should be red-fluorescent, and two thirds should be green-
13 fluorescent (**Fig. 1E**). In the case of segregation distortion, we would expect to observe
14 either an excess of seeds with red and green fluorescence if the Col-0 centromere interval
15 is over-transmitting, or fewer coloured seeds if it is under-transmitting (**Fig. 1E**). As the
16 centromere is in linkage with the markers, we would expect correlated inheritance
17 behaviour by both red and green FTL markers when centromere-linked distortion occurs.
18 In this way, we sought to use the FTL intervals to simultaneously quantify meiotic
19 crossover frequency, and segregation distortion, in centromere-proximal intervals of a
20 diverse panel of Arabidopsis F₁ hybrids, compared to Col-0 inbred lines.

To investigate the effect of natural variation on FTL centromere-proximal crossover frequency, we selected a diverse panel of twelve natural *Arabidopsis* accessions for analysis (**Fig. 1B** and **Tables S1-S2**). These accessions capture diversity from Eurasian (ANGE-B-2, ANGE-B-10, Ler-0, Col-0, and Jm-0), Iberian relict (IP-Cat-0, IP-Med-0, and IP-Alo-19), and non-Iberian relict (Etna-2, Tanz-1, Elk-3, Rab-1, and Cvi-0) genome-wide similarity groups (**Fig. 1B** and **Table S2**) (Naish *et al.* 2021; Wlodzimierz *et al.* 2023b). The Col-0 reference strain carrying the fluorescent reporters is a member of the Eurasian genetic group. Centromere-complete genome assemblies have been published for these accessions (Wlodzimierz *et al.* 2023b), except for Jm-0 and Elk-3. Therefore, Jm-0 was HiFi sequenced to 111× coverage, and Elk-3 was Oxford Nanopore (ONT) sequenced to 1,151× coverage with a subset of >50 kb reads used for genome assembly (248× coverage) using hifiasm (Cheng *et al.* 2021). For Elk-3, all chromosomes were assembled as single contigs, while for Jm-0 chromosomes there were nine gaps in total. For each genome assembly, we used the TRASH algorithm to identify the position and size of the centromeric *CEN178* arrays, as well as annotating genes using Helixer, and transposons using EDTA (**Table S1**) (Wlodzimierz, Hong, and Henderson 2023; Stiehler *et al.* 2021; Ou *et al.* 2019).

In our experiments, homozygous FTL lines in the Col-0 accession were crossed with the twelve inbred accessions to generate replicate F₁ hybrids. The F₁ hybrids were allowed to self-fertilise and their seeds analysed using fluorescence microscopy to quantify both crossover frequency, and segregation distortion (**Fig. 1C-1E** and **Fig. S2**). For each genotype, we scored an average of 12 F₁ replicates, with on average 2,005 seed analysed

per replicate, and 1,518,901 seed analysed in total, representing 3,037,802 meioses (Tables S3-S7). We first consider each centromere individually, and then together, and relate their genetic and epigenetic organisation to variation in centromere-proximal meiotic crossover frequency and segregation distortion.

Meiotic crossover frequency and segregation distortion within *CENTROMERE1*

We used the *CTL1.48* interval trafficline in the Col-0 background (Wu *et al.* 2015; Poethig *et al.* 2022), which contains a red-fluorescent protein encoding T-DNA *CR1111* at base pair position 13,216,744, and a green-fluorescent protein encoding T-DNA *CG30* at position 17,784,719 on chromosome 1 of the Col-CEN genome assembly (Fig. 2A, Fig. S1 and Table S1). For the purposes of this study, we renamed *CTL1.48* as *CENTROMERE1* (*CEN1*), which is 4.57 Mb long in the Col-CEN assembly. Relative to previous Col-0×Ler-0 genetic maps, 79% (3.60 Mb) of the *CEN1* interval corresponds to non-recombining zone (NRZ), with the remainder (0.96 Mb) corresponding to low-recombining zone (LRZ) (Fernandes *et al.* 2024). For each genotype, we quantified *CEN1* crossover frequency (centiMorgans, cM) from replicate individuals, using the counts of red alone (N_R) and green alone (N_G) seeds, compared to the total number (Table S3) (Melamed-Bessudo *et al.* 2005; Francis *et al.* 2007; Wu *et al.* 2015; Poethig *et al.* 2022). To test for significant differences in crossover frequency between the Col-0 inbred controls and each F_1 hybrid, the N_R and N_G counts were summed across replicates for

each genotype and combined with total counts to construct 2×2 contingency tables and Chi-square tests were performed.

Compared to Col-0 inbreds, we observed that ANGE-B-2×Col-0 and Ler-0×Col-0 F₁ hybrids had significantly lower crossover frequency, while recombination was significantly higher in eight hybrids (Chi-square tests, all $P < 2.05 \times 10^{-5}$) (**Fig. 2B** and **Table S3**). Interestingly, the two highest-recombining hybrids, Rab-1×Col-0 and Cvi-0×Col-0, are between diverged relict accessions and the Eurasian Col-0 parent, with larger *CEN178* centromere arrays that contain 8,523 and 7,717 additional copies, respectively (**Fig. 2B-2C** and **Table S1**). In contrast, the intra-Eurasian low-recombining ANGE-B-2×Col-0 and Ler-0×Col-0 hybrids are from accessions with *CEN178* arrays that are more similar in size and structure to Col-0 (**Fig. 2B-2C** and **Table S1**). However, overall there was no significant correlation between *CEN1* crossover frequency and the size difference of the hybrid *CEN178* arrays (defined as the difference in *CEN178* copies between the centromeres) ($r=0.023$, $P=0.120$) (**Fig. 2B-2C**, **Fig. S3** and **Table S1**). For example, the Tanz-1 accession, which has the largest *CEN1* *CEN178* arrays (34,131 copies), which are arranged in a series of inverted blocks, produced F₁ hybrids with a crossover frequency not significantly different to Col-0 inbreds (**Fig. 2B-2C**, **Fig. S4** and **Table S1**). Jm-0 was also notable as it bears a *CEN178* array on the opposite strand to Col-0 and yet produced hybrids with a significantly higher crossover frequency ($P=1.44 \times 10^{-18}$) (**Fig. 2B-2C**, **Fig. S4** and **Table S1**). These observations indicate that the size and structure of the *CEN178* arrays does not simply predict recombination frequency outcome.

In addition to assessing how satellite array polymorphism correlates with centromere-proximal recombination, we calculated pairwise nucleotide diversity (π) within the FTL regions, using high confidence biallelic SNPs, as an additional measure of the genetic diversity between the hybrid accessions and Col-0. As for *CEN178* copy number polymorphism, we did not observe a significant correlation between FTL interval π and crossover frequency (**Fig. S3** and **Table S1**). We conclude that although substantial variation in *CEN1* centromere-proximal crossover frequency exists between Arabidopsis hybrids, this is not significantly correlated with FTL nucleotide diversity, or structural variation in the *CEN178* arrays.

We next sought to use the FTL seed data to investigate segregation distortion. We calculated the ratio of red (N_R) to non-red fluorescent (N_{NR}), and green (N_G) to non-green fluorescent (N_{NG}) seeds. For each F_1 hybrid sample we compared the fluorescent and non-fluorescent seed counts to those from Col-0 inbreds, using 2×2 contingency tables, and performed Chi-square tests to assess significant differences (**Fig. 2B**). When both red and green inheritance ratios were significantly different from Col-0 inbreds, with the same direction of change, we considered this as evidence of centromere-linked segregation distortion. We observed that five F_1 hybrids, derived from the Ler-0, Tanz-1, IP-Med-0, IP-Alo-19, and Rab-1 accessions, showed significant over-transmission of the Col-0 centromere *CEN1* FTL interval (Chi square tests, all $P < 0.011$), whereas, three, with Elk-3, Jm-0, and ANGE-B-10, showed significant under-transmission (Chi square tests, all $P < 0.026$), compared to inbreds (**Fig. 2B** and **Table S1**). The direction and strength of distorted inheritance did not correlate with changes to meiotic crossover

frequency (**Fig. S2**). For example, the strongest distortion occurred against Tanz-1, where there was no significant change to *CEN1* crossover frequency, as well as in the highly recombining IP-Alo-19×Col-0 and Rab-1×Col-0 hybrids (**Fig. 2B** and **Table S1**). Equally, there was no correlation between the strength or direction of distorted inheritance, and how different the hybrid *CEN178* arrays were in size, or with FTL nucleotide diversity (π) (**Fig. 2B-2C, Fig. S3-S4** and **Table S1**). Notably, the Tanz-1 centromere, which has the largest *CEN178* arrays, showed the greatest under-transmission against Col-0 (**Fig. 2B-2C, Fig. S3-S4** and **Table S1**). Together, this reveals distorted *CEN1* inheritance in Arabidopsis hybrids, compared to inbreds, but which was not simply explained by *CEN178* structural variation, FTL nucleotide diversity, or centromere-proximal crossover frequency.

We next examined epigenetic organisation of the centromeres. In a subset of seven parental accessions; the Eurasians Col-0, Ler-0, and Jm-0, and the relicts Cvi-0, Tanz-1, Elk-3, and Rab-1, we compared ChIP-seq enrichment of the centromeric histone CENH3, together with DNA methylation in CG, CHG, and CHH sequence contexts across *CEN1* (Wlodzimierz *et al.* 2023b). We defined zones of CENH3 enrichment, and in all cases observed focused localisation within the *CEN178* satellite arrays (**Fig. 2D** and **Table S1**). We found CENH3 occupancy to be enriched in regions that were between 1.67 and 6.74 Mb wide, and typically with a single peak (**Fig. 2D**). Exceptions included the relict accessions Tanz-1 and Rab-1; in Tanz-1, CENH3 was enriched in a relatively broad 6.74 Mb region, although with strongest enrichment within a single *CEN178* inversion block, whereas in Rab-1, CENH3 was loaded asymmetrically within the largest satellite array

(**Fig. 2D**). In each accession, the centromeres are embedded in highly DNA methylated pericentromeric regions (**Fig. 2D**). The *CEN178* arrays themselves show similar CG context DNA methylation relative to the peri-centromeres, but with reduced CHG and CHH methylation (**Fig. 2D**). Taken together, this indicates stability of epigenetic organisation in the Arabidopsis centromeres, with respect to CENH3 loading, despite a high level of structural genetic variation in the underlying sequences. Overall, the size of the CENH3 zone of enrichment, and the pattern of DNA methylation, did not explain changes to crossover frequency, or the strength or direction of segregation distortion (**Fig. S5 and Table S1**).

Meiotic crossover frequency and segregation distortion within *CENTROMERE2*

For chromosome 2, we used the *CTL2.16* traffic line (Wu *et al.* 2015; Poethig *et al.* 2022), which consists of a red fluorescent protein encoding T-DNA *CR1014* located at base pair position 6,472,440 of chromosome 2 in the Col-CEN genome assembly, and a green-fluorescent protein encoding T-DNA *CG531* at position 2,108,979 bp. For the purposes of this study, we renamed *CTL2.16* as *CENTROMERE2* (*CEN2*), which is 4.36 Mb long in the Col-CEN assembly (**Fig. 3A, Fig. S1 and Table S1**). Relative to previous Col-0×Ler-0 genetic maps, 76% (3.30 Mb) of the *CEN2* interval corresponds to non-recombining zone (NRZ), with the remainder corresponding to low-recombining zone (LRZ) (17.2%, 0.75 Mb), and chromosome arm (7.1%, 0.31 Mb) sequences (Fernandes *et al.* 2024). In *CEN2*, we observed a single hybrid (with Ler-0) that had significantly lower

crossover frequency than Col-0 inbreds, whereas nine hybrids were significantly higher (Chi square tests, all $P \leq 0.023$) (**Fig. 3B**, **Table S1** and **Table S4**). As for *CEN1*, we did not observe a significant pairwise correlation between *CEN178* array size, structural polymorphism, or FTL interval nucleotide diversity, and *CEN2* crossover frequency in the hybrids (**Fig. 3B-3C**, **Fig. S3**, **S6** and **Table S1**). In hybrids with the divergent relict Tanz-1 and IP-Med-0 accessions, which have structurally polymorphic *CEN178* arrays relative to Col-0, including inversions, we observed significantly elevated hybrid crossovers, compared to Col-0 inbreds (Chi square tests, both $P \leq 3.45 \times 10^{-4}$) (**Fig. 3B-3C**, **Fig. S6** and **Table S1**), indicating again that satellite array genetic divergence is not a barrier to elevated centromere-proximal crossover frequency in Arabidopsis hybrids .

Compared to *CEN1*, fewer of the *CEN2* hybrids showed significant segregation distortion, with ANGE-B-10×Col-0, Tanz-1×Col-0 and Cvi-0×Col-0 hybrids showing over-transmission of the Col-0 centromere (Chi square tests, all $P \leq 0.034$) (**Fig. 3B** and **Table S1**). Similar to *CEN1*, the incidence of segregation distortion did not correlate either with FTL nucleotide diversity, *CEN178* array size differences, or centromere-proximal crossover frequency (**Fig. 3B-3C**, **Fig. S3** and **S6** and **Table S1**).

We compared the epigenetic landscape across *CEN2* and observed four accessions with zones of CENH3 enrichment 1.53-2.19 Mb, whereas the relicts Tanz-1, Rab-1 and Elk-3 showed larger *CEN178* arrays that contained broader 3.42-4.06 Mb CENH3-occupied regions (**Fig. 3D** and **Table S1**). The Rab-1 and Elk-3 hybrids both show significantly

higher *CEN3* crossover frequency, but no evidence of distortion (**Fig. 3B**). In each of the seven accessions, the region of CENH3 enrichment within *CEN2* coincides with high CG DNA methylation, and depleted CHG and CHH methylation in the *CEN178* arrays relative to the peri-centromeres (**Fig. 3D**). This reinforces the conclusion that although significant variation in centromere-proximal crossover frequency and segregation distortion exists between *Arabidopsis* hybrids, this is not easily explained by satellite array polymorphism, nucleotide diversity, or the zones of CENH3 enrichment and DNA methylation states.

Meiotic crossover frequency and segregation distortion within *CENTROMERE3*

To study chromosome 3, we combined the red-encoding FTL T-DNA *CR880*, located at base pair position 12,652,136 of chromosome 3 in the Col-CEN genome assembly, with a green-encoding T-DNA *CG811* at position 18,395,802, and named this interval *CENTROMERE3* (*CEN3*), which is 5.743 Mb long in Col-CEN (**Fig. 4A**, **Fig. S1** and **Table S1**) (Wu *et al.* 2015; Poethig *et al.* 2022). Relative to previous Col-0×Ler-0 genetic maps, 52% (3.0 Mb) of the *CEN3* interval corresponds to non-recombining zone (NRZ), with the remainder corresponding to low-recombining zone (LRZ) (34.2%, 1.1 Mb) and chromosome arm (13.6%, 0.78 Mb) sequences (Fernandes *et al.* 2024). In contrast to *CEN1* and *CEN2*, where the majority of hybrids showed significantly higher crossover frequency compared to inbreds, in *CEN3*, six hybrids were lower, and two were higher (Chi square tests, all $P < 2.73 \times 10^{-5}$) (**Fig. 4B** and **Table S5**). Interestingly, hybrids with the Cvi-0 and IP-Alo-19 relict accessions again showed elevated crossover frequency, despite structural

polymorphism in the *CEN178* arrays, compared to Col-0 inbreds (**Fig. 4B-4C**, **Fig. S3** and **Fig. S7** and **Table S1**). In Col-0, the *CEN3 CEN178* arrays have an inverted array structure (Naish *et al.* 2021), with inversions also present in the Elk-3, Ler-0, Rab-1, Cvi-0, and IP-Alo-19 satellite arrays, which together include the parents of both the highest (IP-Alo-19) and lowest (Elk-3) recombining hybrids, arguing that the *CEN178* inversion structure does not control crossover frequency (**Fig. 4B-4C**, **Fig. S7** and **Table S1**). Three hybrids with relict parents (Elk-3, Tanz-1, and Rab-1) showed significant over-transmission of the Col-0 centromere as monitored by FTL inheritance ratios (Chi square test, all $P=<0.0031$), whereas the Eurasian Jm-0×Col-0 hybrid showed significant Col-0 under-transmission (Chi square tests, both $P=<0.0011$) (**Fig. 4B** and **Table S1**). In the Rab-1×Col-0 hybrids, the *CEN3 CEN178* array is substantially larger than in Col-0 (2.14 vs 5.9 Mb), indicating again that larger array size does not associate with stronger distorted inheritance in Arabidopsis (**Fig. 4C** and **Table S1**). At the chromatin level, Col-0 shows a single 2.01 Mb region of CENH3 enrichment, whereas Rab-1 shows CENH3 enrichment on discrete large (4.9 Mb), and small (0.99 Mb) *CEN178* arrays (**Fig. 4D** and **Table S1**), which have the potential to contribute to distorted inheritance. With respect to CENH3 loading, Cvi-0, Ler-0, and Tanz-1 ChIP-seq enrichment was observed within *CEN178* sub-arrays that were mainly arranged on a single strand, whereas in Col-0 and Elk-3 CENH3 loading spans inverted satellite arrays (**Fig. 4D**).

Meiotic crossover frequency and segregation distortion within *CENTROMERE4*

For chromosome 4, we used the *CTL4.12* interval traffic line (Wu *et al.* 2015; Poethig *et al.* 2022), which consists of a red-encoding T-DNA *CR909* at base pair position 2,437,418 of chromosome 4 in the Col-CEN genome assembly, and a green-encoding T-DNA *CG350* at position 8,480,638. For the purposes of this study, we renamed *CTL4.12* as *CENTROMERE4* (*CEN4*), which is 6.04 Mb long in Col-CEN (**Fig. 5A**, **Fig. S1** and **Table S1**). Relative to previous Col-0×Ler-0 genetic maps, 67.8% (4.1 Mb) of the *CEN4* interval corresponds to non-recombining zone (NRZ), with the remainder corresponding to low-recombining zone (LRZ) (26.4%, 1.59 Mb), and chromosome arm (5.8%, 0.35 Mb) sequences (Fernandes *et al.* 2024). For *CEN4*, the majority of hybrids (11 of 12) showed significantly reduced crossover frequency compared to inbreds, with the only exception being the relict Cvi-0×Col-0 hybrid, which had significantly higher recombination (Chi square test, $P=1.31\times10^{-4}$) (**Fig. 5B** and **Table S6**). Col-0 is documented to have a large ‘knob’ inversion on the short arm of chromosome four that spans the *CR909* FTL T-DNA (Fransz *et al.* 2000, 2016), which is not shared with any of the twelve parental accessions crossed to (**Fig. S8**). This inversion almost certainly contributes substantially to the widespread suppressed crossover frequency in the *CEN4* hybrids tested here.

Two relict accessions, Tanz-1 and IP-Cat-0, produced hybrids with significant over-transmission of the Col-0 FTL centromere (Chi square tests, both $P=<1.80\times10^{-3}$) (**Fig. 5B** and **Table S1**). Both accessions share the knob inversion but are otherwise different from

one another in *CEN178* array size and structure (**Fig. 5B, S8 and Table S1**). Amongst the seven accessions profiled at the level of chromatin, the relicts Rab-1 and Tanz-1 showed the presence of multiple distinct *CEN178* arrays, one of which contained the majority of CENH3 in each case (**Fig. 5D**). In both Rab-1×Col-0 and Tanz-1×Col-0 hybrids, the smaller Col-0 *CEN178* array showed biased segregation distortion in its own favour (**Fig. 5B and Table S1**), further reinforcing that larger satellite arrays in Arabidopsis do not associate with distorted inheritance in their own favour.

Meiotic crossover frequency and segregation distortion within *CENTROMERE5*

We completed our study by crossing the red-encoding FTL T-DNA *CR1136* at base pair position 7,998,504 of chromosome 5 in the Col-CEN genome assembly, and a green-encoding FTL T-DNA *CG358* at position 15,759,366, to generate the *CENTROMERE5* (*CEN5*) interval, which is 7.76 Mb long (**Fig. 6A, Fig. S1 and Table S1**) (Wu *et al.* 2015; Poethig *et al.* 2022). Relative to previous Col-0×Ler-0 genetic maps, 48.9% (3.8 Mb) of the *CEN5* interval corresponds to non-recombining zone (NRZ), with the remainder corresponding to low-recombining zone (LRZ) (17.3%, 1.34 Mb), and chromosome arm (33.7%, 2.62 Mb) sequences (Fernandes *et al.* 2024). Compared to the other *CEN* FTL intervals, *CEN5* contains a higher proportion of pericentromeric and arm sequences, in addition to the *CEN178* satellite arrays (**Fig. S1**). In *CEN5*, seven hybrids showed significantly higher crossover frequency compared to inbreds, and two were significantly lower (Chi square tests, all $P < 4.87 \times 10^{-6}$) (**Fig. 6B and Table S7**), with the hybrids

1 derived from relicts IP-Alo-19 and IP-Cat-0 showing the highest crossover frequency.
 2 These Iberian relict accessions share a similar *CEN178* array, but which is orientated on
 3 the opposite strand to the Col-0 array (**Fig. 6C**). In contrast to the other centromeres, the
 4 Cvi-0×Col-0 hybrid had significantly lower crossover frequency (Chi square test,
 5 $P=1.77\times 10^{-46}$), suggestive of genetic or epigenetic features in this interval that suppress
 6 recombination (**Fig. 6B-6C** and **Table S1**). While the Cvi-0 *CEN178* array is substantially
 7 larger than Col-0, the Tanz-1 array is even larger, and Tanz-1×Col-0 hybrids had
 8 significantly higher crossover frequency than Col-0 inbreds (Chi square test, $P=5.81\times 10^{-10}$),
 9 which indicates independence of *CEN178* array structural variation on centromere-
 10 proximal recombination (**Fig. 6B-6C**, **Fig. S9** and **Table S1**). Five hybrids showed
 11 significant segregation distortion in favour of the Col-0 chromosome (Chi square tests, all
 12 $P<0.031$), whereas one hybrid, ANGE-B-2×Col-0, showed distortion against Col-0 (Chi
 13 square test, $P=5.53\times 10^{-5}$) (**Fig. 6B** and **Table S1**). As for the other centromeres, variation
 14 in centromere-proximal crossover frequency was uncorrelated with the incidence of
 15 segregation distortion (**Fig. 6B-6C**, **Fig. S2** and **Table S1**). With respect to CENH3
 16 loading, the Rab-1 and Tanz-1 relict accessions both displayed relatively large 5.7 and
 17 6.5 Mb *CEN178* arrays, respectively, that contained broad 5.57 and 7.45 Mb regions of
 18 CENH3 enrichment, although these was not associated with consistent effects on
 19 crossover frequency, or distorted inheritance (**Fig. 6B-6D** and **Table S1**).

Centromere-proximal crossover frequency and segregation distortion compared across all five chromosomes

In total, we quantified crossover frequency and segregation distortion in sixty hybrid combinations derived from twelve accessions that were analyzed across five centromeres each and compared them with inbred Col-0 FTL strains (**Figs. 2-7** and **Tables S1-S7**). In total, forty-nine of the hybrid centromere combinations had significantly higher ($n=27$), or lower ($n=22$), centromere-proximal crossover frequency, compared to Col-0 inbreds, consistent with widespread modification of centromere-proximal recombination by natural variation (**Figs. 2-7** and **Table S1**). The most consistently high-recombining hybrids were those derived from the relict accessions Cvi-0 and IP-Alo-19, where four of five *CEN* intervals showed significantly higher crossover than Col-0 inbreds, in each case (**Fig. 7A** and **Table S1**). In contrast, crossovers in Ler-0 x Col-0 hybrids were suppressed compared to inbreds, across four of the five centromeres (**Fig. 7A** and **Table S1**), which together may indicate the action of *trans*-modifying effects in these hybrids that coordinately influence recombination on multiple centromeres. Despite being more genetically diverged, the relict accessions produced proportionally more hybrids with significantly higher recombination (23 of 40, 65%), than the Eurasian accessions (4 of 20, 20%) (**Table S1**). This is also true when Iberian (9 of 15, 60%) and non-Iberian (14 of 25, 56%) relicts are considered separately (**Table S1**). Hybrid crossover frequency was most likely to be suppressed within *CEN4*, relative to inbreds, compared to the other centromeric intervals, with reduced crossovers in all but one of the twelve hybrids (**Fig. 7A** and **Table S1**). Suppression of *CEN4* crossover frequency is almost certainly due to

the well-documented knob-inversion in Col-0 that is not present in the other accessions in our study set (**Fig. S8**) (Fransz *et al.* 2000, 2016). Together, this indicates widespread modification of proximal crossover frequency, although with accession and centromere-specific effects.

Compared to crossover frequency, we observed fewer (18 of 60) instances of significant segregation distortion in the hybrids compared to inbreds (**Fig. 7B** and **Table S1**). There was no significant correlation between changes to crossover frequency and the incidence of segregation distortion (**Fig. S2** and **Table S1**), indicating that these phenomena are largely uncoupled. We observed eighteen instances of the Col-0 FTL chromosomes over-transmitting, and five cases where there was under-transmission, in hybrids compared to inbreds, with a greater incidence of distortion on the longer chromosomes one and five (13 of 18) (**Fig. 7B** and **Table S1**). Proportionally, there was a similar incidence of segregation distortion amongst hybrids with the diverged relict accessions (16 of 40, 40%), compared to hybrids between the Eurasians (7 of 20, 35%) (**Fig. 7** and **Table S1**). The Iberian (4 of 15, 27%) and non-Iberian (12 of 25, 48%) relict hybrids both showed distorted inheritance (**Table S1**). Interestingly, the Tanz-1 accession showed distortion in favour of Col-0 across four of five centromeres, potentially indicating a *trans* modifying effect acting across the centromeres in this accession (**Fig. 7B-7C** and **Table S1**).

To further consider our data together, we used linear modelling to investigate which factors explain variation in crossover frequency, and segregation distortion. First, as each

FTL interval varies in physical size, we used crossovers per megabase (cM/Mb) as the response variable. We considered the model:

$$\text{cM/Mb} \sim \text{chromosome} * \text{accession}$$

Where ‘chromosome’ denotes the CTL interval analysed (*CEN1* to *CEN5*), and ‘accession’ denotes the hybrid background that was measured. Our model explained 85.0% of the variation in crossover frequency, where the interaction between chromosome and accession explained the greatest amount (53.5%), compared to chromosome (13.1%), and accession (7.3%) alone (**Fig. 7B** and **Table S8**). This indicates that the influence of a specific centromere on crossover frequency is dependent on the accession crossed to. We tested additional models for crossover frequency that included the stepwise addition of the variables CTL interval length, gene density, transposon density, and *CEN178* density, but found these did not explain additional variation (**Table S8**). Including segregation distortion data in the model explained a small amount of crossover variation (0.2% for red inheritance and 0.6% for green inheritance), consistent with observations made for individual centromeres. Genotype replicates were included as a random variable to control for experimental replication and were not found to contribute to the variance explained.

We took a similar approach to model genetic control of segregation distortion. For each CTL interval, we previously observed that the inheritance ratios of red and green

fluorescence were highly correlated (**Fig. S2**), and similar results were obtained modelling the red or green seed fluorescence data (**Table S8**). We considered the model:

$$\text{segregation distortion} \sim \text{chromosome} * \text{accession}$$

The models explained 56.4% (green) and 53.2% (red) of the variation in distortion, where the interaction between chromosome and accession explained the greatest amount of variance (green=35.2%, red=30.2%), compared to chromosome (green=2.0%, red=4.0%) and accession (green=20.6%, red=20.0%) alone (**Fig 7D** and **Table S8**). This indicates that, like for crossover frequency, the influence of a specific centromere is dependent on the accession crossed to. In contrast to crossover frequency, accession had a stronger individual effect than chromosome (**Fig. 7D** and **Table S8**). Including cM/Mb data as a variable explained a small amount of variation in distortion (green=2.24%, red=1.63%) (**Table S8**), consistent with our previous observations that crossover frequency and distortion are not correlated. In contrast to cM/Mb, including the annotation features FTL interval size, transposon density, gene density and *CEN178* density explained a greater amount of variation in distortion (green=69.45%, red=66.67%) (**Table S8**), potentially indicating a stronger contribution of *cis*-acting genetic variation on distortion compared to crossover frequency.

Finally, as CENH3 is a key determinant of centromere identity (Talbert et al. 2002; McKinley and Cheeseman 2016), we investigated whether genetic polymorphism in

CENH3 might contribute to variation in centromere-proximal crossover frequency, or segregation distortion, observed between the hybrids studied here (**Fig. S10**). We observed few amino acid polymorphisms between Col-0 CENH3 and sequences in the accessions crossed to; with a G to A substitution in Elk-3 and IP-Cat-0, and a S to Y substitution in Cvi-0 (**Fig. S10**). In contrast, we observe a high number of amino acid variants, clustered in the N-terminal tail of CENH3, between *A. thaliana* and the sister species *A. lyrata* (**Fig. S10**), consistent with CENH3 diverging rapidly between species in the *Arabidopsis* genus (Kawabe, Nasuda, and Charlesworth 2006; Maheshwari et al. 2015). We conclude that genetic variation in CENH3 within *A. thaliana* is unlikely to explain the hybrid variation we observe in centromere-proximal crossover frequency and segregation distortion.

Discussion

We studied twelve *Arabidopsis* hybrids to survey the effects of naturally occurring variation on meiotic crossover frequency and segregation distortion across all five centromeres. We found widespread modification of centromere-proximal crossover frequency, with approximately equal cases where hybrid recombination was higher or lower than observed in inbreds. When considering genetic variation within the FTL intervals, including the size and structure of the *CEN178* satellite arrays and nucleotide diversity, there was no clear pairwise relationship between sequence divergence and recombination frequency. However, our linear modelling indicates a strong interacting

1 effect of *cis* and *trans* genetic variation on crossover frequency. Interestingly, the
2 accessions producing the highest-recombining hybrids, Cvi-0 and IP-Alo-19, are both
3 diverged relict accessions compared to the Eurasian Col-0 reference strain. This indicates
4 that an inhibitory effect of greater *cis* genetic divergence does not easily explain variation
5 in centromere-proximal meiotic recombination levels. These observations are also
6 consistent with previous work, where *Arabidopsis* hybrids showed higher recombination
7 than inbreds (Barth *et al.* 2001; Ziolkowski *et al.* 2015; Blackwell *et al.* 2020; Dluzewska
8 *et al.* 2023).

9
10 Larger inversions and structural variants in the chromosome arms are well known to
11 suppress crossover recombination in *cis* (Sturtevant 1917), and this is also the case for
12 the 1.2 Mb knob-inversion polymorphism within *CEN4* of Col-0 (Fransz *et al.* 2000, 2016).
13 As Cvi-0×Col-0 and IP-Alo-19×Col-0 hybrids had high recombination rates across four of
14 five centromeres, while recombination was low in four of the centromeres in Ler-0×Col-0
15 hybrids, we propose that Cvi-0, IP-Alo-19, and Ler-0 may also contribute *trans*-acting
16 recombination-modifying loci that exert a global effect on centromere-proximal crossovers
17 across multiple chromosomes. Further quantitative trait loci mapping experiments using
18 populations derived from these hybrids may reveal new genes that regulate the levels of
19 centromere-proximal crossover frequency, and the extent of *cis* versus *trans* acting
20 variation on recombination. We did not observe a strong effect of structural variation
21 between the centromere satellite arrays on crossover frequency. However, as
22 centromere-proximal crossovers have been fine-mapped to pericentromeric ‘low-
23 recombining zones’ (LRZ) that flank the central ‘non-recombining zone’ (NRZ) that

contains the satellite arrays (Fernandes *et al.* 2024), this may effectively shield recombination rates across these regions from the effect of structural variation in the core centromere regions.

Segregation distortion can be produced by a variety of causes during meiosis, or post-meiosis during germ line development. As our experiments involved self-fertilization of *Arabidopsis*, we are measuring both male and female meiotic transmission simultaneously. Meiotic drive of centromeres has been detected in female meiosis of mice, as well as monkeyflowers, where only one of the meiotic products survives to produce the egg cell (Finseth *et al.* 2022; Akera *et al.* 2019; Chmátal *et al.* 2014; Iwata-Otsubo *et al.* 2017; Akera *et al.* 2017; Kumon *et al.* 2021; Lampson and Black 2017; Kursel and Malik 2018; Malik 2009; Fishman and Willis 2005; Fishman and Saunders 2008). This pattern of development is similar in *Arabidopsis* ovules, where the megasporocyte gives rise to a tetrahedral or decussate tetrad, and where the most chalazal cell enlarges and the remaining three degenerate (Webb and Gunning 1990; Bowman 2012), meaning there is an opportunity for female meiotic drive to cause segregation distortion. Asymmetries in the spindle, or cell division, during female meiosis have the potential to be exploited by variant centromeres to cause segregation distortion into the surviving meiocyte (Malik 2009; Chmátal *et al.* 2014; Lampson and Black 2017; Iwata-Otsubo *et al.* 2017; Akera *et al.* 2017, 2019; Kursel and Malik 2018; Kumon *et al.* 2021). For example, in *Mimulus* interspecific hybrids, a 98:2 centromeric transmission advantage of the *D* haplotype has been observed through female meiosis, and 58:42 drive in con-specific crosses (Fishman and Saunders 2008; Fishman and Kelly 2015).

1 This correlates with a larger cytogenetic fluorescent *in situ* hybridization (FISH) signal for
2 the Cent728 satellite repeats in the driving *D* haplotype (Fishman and Saunders 2008).
3 Similarly, *Mus musculus* hybrids between laboratory strains and wild strains with smaller
4 centromeres (fewer minor satellite repeats and lower CENP-A loading) show biased
5 segregation of the lab-strain centromeres into the egg (Akeru *et al.* 2017, 2019). In our
6 study, although less frequent than centromere-proximal crossover modification, we
7 observed multiple instances of segregation distortion across the five centromeres.
8 Although, we did not observe a clear pairwise correlation between genetic divergence, or
9 the relative sizes of hybrid *CEN178* arrays, our linear modelling indicates a strong
10 combined effect of *cis* and *trans* genetic variation on distortion. We also note several
11 cases such as Tanz-1 *CEN1* and *CEN2* and Rab-1 *CEN4*, where the smaller Col-0
12 *CEN178* arrays show distorted inheritance at the expense of the larger relict
13 arrays. Therefore, larger satellite arrays may not universally cause stronger drive in their
14 own favour across species.

15
16 As we did not detect a strong correlation between segregation distortion and the size of
17 the *CEN178* satellite arrays, or the region of CENH3 loading, this may indicate that other
18 features of centromere organisation are important for distorted inheritance. For example,
19 the relative placement of CENH3 regions on the recombining homologous chromosomes
20 during meiosis may be more relevant for distortion, rather than satellite array sizes, or the
21 levels of CENH3 *per se*. During meiosis, formation of DSBs are more frequent in the
22 chromosome arms and are more likely to initiate interhomolog strand invasion and
23 recombination earlier in prophase-I, due to telomere-led synapsis (Choi *et al.* 2018;

Lambing *et al.* 2020). Since the Arabidopsis chromosome arms tend to be highly syntenic between accessions (Włodzimierz *et al.* 2023b; Lian *et al.* 2024), they will be more likely to synapse rapidly. We propose that the structurally polymorphic centromeres, which will form fewer DSBs, are likely to synapse in a delayed manner. In this scenario the flanking syntenic chromosome arms will have aligned and synapsed well, meaning there may be megabase regions in the centre that are not well aligned or synapsed, which could potentially manifest as a synapsis ‘buckle’ composed of the structurally polymorphic regions. Within these potentially buckled centromeric regions, although a similar area may be CENH3-occupied on both homologs, these regions could be positioned differently relative to the spindle microtubules, due to centromere structural polymorphism, in the context of the heterozygous bivalent structure (**Fig. S11**). We speculate that such offsets of CENH3 loading, and their differential interaction with putative spindle asymmetries during female meiosis, may contribute to centromeric segregation distortion that we observe in Arabidopsis hybrids. We also note that multiple instances of segregation distortion have been reported in Arabidopsis, suggesting such interactions could be widespread and result from interactions between multiple *cis* and *trans*-acting factors (Törjék *et al.* 2006; Salomé *et al.* 2012; Seymour *et al.* 2019).

Future studies that explore the cell biology of Arabidopsis meiosis, spindle formation, and how hybrid centromeres interact during meiotic pairing and recombination, may reveal mechanisms associated with segregation distortion. It will also be important to consider alternative models that may explain distortion, for example male-based gene drive via spore-killing, involving toxin-antidote systems that are linked to the centromere (Nuckolls

et al. 2017; Berube *et al.* 2024). In conclusion, our study reveals specific hybrid combinations where segregation distortion is evident across each of the Arabidopsis centromeres, as well as variation in meiotic crossover frequency. These hybrids will be useful study systems in which to investigate the molecular basis of segregation distortion and crossover variation over the centromeres.

Data availability

Long-read sequencing data and corresponding genome assemblies generated in this study for accessions Jm-0 and Elk-3 are available at the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena/browser/home>) under accession number PRJEB83900. Previously released long-read data and assemblies for the remaining accessions are available at ENA under accession number PRJEB55353 (Wlodzimierz *et al.* 2023b). ChIP-seq reads for Rab-1, Elk-3, and Jm-0 are available under ArrayExpress accession E-MTAB-15384.

Acknowledgements

We kindly thank Scott Poethig for providing the fluorescent TRAFFIC FTL reporter lines, and Fabrice Roux, Carlos Alonso-Blanco, and Angela Hancock for providing seed of Arabidopsis accessions.

Funding

We acknowledge support from UKRI/BBSRC grants BB/Y009487/1 and BB/V003984/1, ERC grant 101142254 EvoPanCen, a UKRI/BBSRC DTP studentship to NG, and a Broodbank Fellowship to MN.

Competing interests

DW also consults for KWS SE, a globally active plant breeder and seed producer. DW holds equity in Computomics, which advises plant breeders. All other authors declare no competing interests.

References

- Akera, Takashi, Lukáš Chmátal, Emily Trimm, Karren Yang, Chana Aonbangkhen, David M. Chenoweth, Carsten Janke, Richard M. Schultz, and Michael A. Lampson. 2017. "Spindle Asymmetry Drives Non-Mendelian Chromosome Segregation." *Science* 358 (6363): 668–72.
- Akera, Takashi, Emily Trimm, and Michael A. Lampson. 2019. "Molecular Strategies of Meiotic Cheating by Selfish Centromeres." *Cell* 178 (5): 1132–44.e10.
- Alonge, Michael, Ludivine Lebeigle, Melanie Kirsche, Katie Jenike, Shujun Ou, Sergey Aganezov, Xingang Wang, Zachary B. Lippman, Michael C. Schatz, and Sebastian Soyk. 2022. "Automated Assembly Scaffolding Using RagTag Elevates a New Tomato System for High-Throughput Genome Editing." *Genome Biology* 23 (1): 258.
- Arter, Meret, and Scott Keeney. 2024. "Divergence and Conservation of the Meiotic Recombination Machinery." *Nature Reviews. Genetics* 25 (5): 309–25.
- Barth, S., A. E. Melchinger, B. Devezi-Savula, and T. Lübberstedt. 2001. "Influence of Genetic Background and Heterozygosity on Meiotic Recombination in Arabidopsis Thaliana." *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada* 44 (6): 971–78.
- Berchowitz, Luke E., and Gregory P. Copenhaver. 2008. "Fluorescent Arabidopsis Tetrads: A Visual Assay for Quickly Developing Large Crossover and Crossover Interference Data Sets." *Nature Protocols* 3 (1): 41–50.

- 1 Berube, Benjamin, Evan Ernst, Jonathan Cahn, Benjamin Roche, Cristiane de Santis
2 Alves, Jason Lynn, Armin Scheben, et al. 2024. "Teosinte Pollen Drive Guides Maize
3 Diversification and Domestication by RNAi." *Nature* 633 (8029): 380–88.
4
- 5 Blackwell, Alexander R., Julia Dluzewska, Maja Szymanska-Lejman, Stuart Desjardins,
6 Andrew J. Tock, Nadia Kbiri, Christophe Lambing, et al. 2020. "MSH2 Shapes the
7 Meiotic Crossover Landscape in Relation to Interhomolog Polymorphism in
8 *Arabidopsis*." *The EMBO Journal* 39 (21): e104858.
9
- 10 Bowman, John. 2012. *Arabidopsis: An Atlas of Morphology and Development*. Springer
11 Science & Business Media.
12
- 13 Charlesworth, Deborah. 2017. "Evolution of Recombination Rates between Sex
14 Chromosomes." *Philosophical Transactions of the Royal Society of London. Series B,*
15 *Biological Sciences* 372 (1736). <https://doi.org/10.1098/rstb.2016.0456>.
16
- 17 Cheng, Haoyu, Gregory T. Concepcion, Xiaowen Feng, Haowen Zhang, and Heng Li.
18 2021. "Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with
19 Hifiasm." *Nature Methods* 18 (2): 170–75.
20
- 21 Chmátal, Lukáš, Sofia I. Gabriel, George P. Mitsainas, Jessica Martínez-Vargas, Jacint
22 Ventura, Jeremy B. Searle, Richard M. Schultz, and Michael A. Lampson. 2014.
23 "Centromere Strength Provides the Cell Biological Basis for Meiotic Drive and
24 Karyotype Evolution in Mice." *Current Biology: CB* 24 (19): 2295–2300.
25
- 26 Choi, Kyuha, Xiaohui Zhao, Krystyna A. Kelly, Oliver Venn, James D. Higgins, Nataliya
27 E. Yelina, Thomas J. Hardcastle, et al. 2013. "Arabidopsis Meiotic Crossover Hot Spots
28 Overlap with H2A.Z Nucleosomes at Gene Promoters." *Nature Genetics* 45 (11): 1327–
29 36.
30
- 31 Choi, Kyuha, Xiaohui Zhao, Andrew J. Tock, Christophe Lambing, Charles J.
32 Underwood, Thomas J. Hardcastle, Heïdi Serra, et al. 2018. "Nucleosomes and DNA
33 Methylation Shape Meiotic DSB Frequency in Arabidopsis Thaliana Transposons and
34 Gene Regulatory Regions." *Genome Research* 28 (4): 532–46.
35
- 36 Colomé-Tatché, Maria, Sandra Cortijo, René Wardenaar, Lionel Morgado, Benoit
37 Lahouze, Alexis Sarazin, Mathilde Etchevery, et al. 2012. "Features of the Arabidopsis
38 Recombination Landscape Resulting from the Combined Loss of Sequence Variation
39 and DNA Methylation." *Proceedings of the National Academy of Sciences of the United*
40 *States of America* 109 (40): 16240–45.
41
- 42 Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin
43 O. Pollard, Andrew Whitwham, et al. 2021. "Twelve Years of SAMtools and BCFtools."
44 *GigaScience* 10 (2). <https://doi.org/10.1093/gigascience/giab008>.
45
- 46 Dluzewska, Julia, Wojciech Dziegielelewski, Maja Szymanska-Lejman, Monika Gazecka,
Ian R. Henderson, James D. Higgins, and Piotr A. Ziolkowski. 2023. "MSH2 Stimulates

- Interfering and Inhibits Non-Interfering Crossovers in Response to Genetic Polymorphism." *Nature Communications* 14 (1): 6716.
- Dooner, H. K., and I. M. Martínez-Férez. 1997. "Recombination Occurs Uniformly within the Bronze Gene, a Meiotic Recombination Hotspot in the Maize Genome." *The Plant Cell* 9 (9): 1633–46.
- Drouaud, Jan, Hossein Khademian, Laurène Giraut, Vanessa Zanni, Sarah Bellalou, Ian R. Henderson, Matthieu Falque, and Christine Mézard. 2013. "Contrasted Patterns of Crossover and Non-Crossover at Arabidopsis Thaliana Meiotic Recombination Hotspots." *PLoS Genetics* 9 (11): e1003922.
- Fernandes, Joiselle B., Matthew Naish, Qichao Lian, Robin Burns, Andrew J. Tock, Fernando A. Rabanal, Piotr Wlodzimierz, et al. 2024. "Structural Variation and DNA Methylation Shape the Centromere-Proximal Meiotic Crossover Landscape in Arabidopsis." *Genome Biology* 25 (1): 30.
- Finseth, Findley, Keely Brown, Andrew Demaree, and Lila Fishman. 2022. "Supergene Potential of a Selfish Centromere." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 377 (1856): 20210208.
- Fishman, Lila, and John K. Kelly. 2015. "Centromere-Associated Meiotic Drive and Female Fitness Variation in *Mimulus*." *Evolution*. <https://doi.org/10.1111/evo.12661>.
- Fishman, Lila, and Arpiar Saunders. 2008. "Centromere-Associated Female Meiotic Drive Entails Male Fitness Costs in Monkeyflowers." *Science* 322 (5907): 1559–62.
- Fishman, Lila, and John H. Willis. 2005. "A Novel Meiotic Drive Locus Almost Completely Distorts Segregation in *Mimulus* (monkeyflower) Hybrids." *Genetics* 169 (1): 347–53.
- Francis, Kirk E., Sandy Y. Lam, Benjamin D. Harrison, Alexandra L. Bey, Luke E. Berchowitz, and Gregory P. Copenhaver. 2007. "Pollen Tetrad-Based Visual Assay for Meiotic Recombination in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* 104 (10): 3913–18.
- Fransz, Paul, Gabriella Linc, Cheng-Ruei Lee, Saulo Alves Aflitos, Jesse R. Lasky, Christopher Toomajian, Hoda Ali, et al. 2016. "Molecular, Genetic and Evolutionary Analysis of a Paracentric Inversion in Arabidopsis Thaliana." *The Plant Journal : For Cell and Molecular Biology* 88 (2): 159–78.
- Fransz, P. F., S. Armstrong, J. H. de Jong, L. D. Parnell, C. van Drunen, C. Dean, P. Zabel, T. Bisseling, and G. H. Jones. 2000. "Integrated Cytogenetic Map of Chromosome Arm 4S of A. Thaliana: Structural Organization of Heterochromatic Knob and Centromere Region." *Cell* 100 (3): 367–76.

- 1 Goel, Manish, Hequan Sun, Wen-Biao Jiao, and Korbinian Schneeberger. 2019. "SyRI:
2 Finding Genomic Rearrangements and Local Sequence Differences from Whole-
3 Genome Assemblies." *Genome Biology* 20 (1): 277.
- 4
- 5 Grelon, M., D. Vezon, G. Gendrot, and G. Pelletier. 2001. "AtSPO11-1 Is Necessary for
6 Efficient Meiotic Recombination in Plants." *The EMBO Journal* 20 (3): 589–600.
- 7
- 8 Henikoff, S., K. Ahmad, and H. S. Malik. 2001. "The Centromere Paradox: Stable
9 Inheritance with Rapidly Evolving DNA." *Science* 293 (5532): 1098–1102.
- 10
- 11 Iwata-Otsubo, Aiko, Jennine M. Dawicki-McKenna, Takashi Akera, Samantha J. Falk,
12 Lukáš Chmátal, Karren Yang, Beth A. Sullivan, Richard M. Schultz, Michael A.
13 Lampson, and Ben E. Black. 2017. "Expanded Satellite Repeats Amplify a Discrete
14 CENP-A Nucleosome Assembly Site on Chromosomes That Drive in Female Meiosis."
15 *Current Biology: CB* 27 (15): 2365–73.e8.
- 16
- 17 Jain, Chirag, Arang Rhie, Nancy F. Hansen, Sergey Koren, and Adam M. Phillippy.
18 2022. "Long-Read Mapping to Repetitive Reference Sequences Using Winnowmap2."
19 *Nature Methods* 19 (6): 705–10.
- 20
- 21 Jones, Gareth H., and F. Chris H. Franklin. 2006. "Meiotic Crossing-over: Obligation
22 and Interference." *Cell* 126 (2): 246–48.
- 23
- 24 Kang, Minghui, Haolin Wu, Huanhuan Liu, Wenyu Liu, Mingjia Zhu, Yu Han, Wei Liu, et
25 al. 2023. "The Pan-Genome and Local Adaptation of Arabidopsis Thaliana." *Nature*
26 *Communications* 14 (1): 6259.
- 27
- 28 Kawabe, Akira, Shuhei Nasuda, and Deborah Charlesworth. 2006. "Duplication of
29 Centromeric Histone H3 (HTR12) Gene in Arabidopsis Halleri and A. Lyrata, Plant
30 Species With Multiple Centromeric Satellite Sequences." *Genetics* 174 (4): 2021.
- 31
- 32 Kleckner, Nancy. 2006. "Chiasma Formation: Chromatin/axis Interplay and the Role(s)
33 of the Synaptonemal Complex." *Chromosoma* 115 (3): 175–94.
- 34
- 35 Koch, Lori B., and Adele L. Marston. 2025. "The Functional Organisation of the
36 Centromere and Kinetochore during Meiosis." *Current Opinion in Cell Biology* 94
37 (June):102486.
- 38
- 39 Koehler, K. E., C. L. Boulton, H. E. Collins, R. L. French, K. C. Herman, S. M. Lacefield,
40 L. D. Madden, C. D. Schuetz, and R. S. Hawley. 1996. "Spontaneous X Chromosome
41 MI and MII Nondisjunction Events in Drosophila Melanogaster Oocytes Have Different
42 Recombinational Histories." *Nature Genetics* 14 (4): 406–14.
- 43
- 44 Kumon, Tomohiro, Jun Ma, R. Brian Akins, Derek Stefanik, C. Erik Nordgren, Junhyong
45 Kim, Mia T. Levine, and Michael A. Lampson. 2021. "Parallel Pathways for Recruiting

Effector Proteins Determine Centromere Drive and Suppression.” *Cell* 184 (19): 4904–18.e11.

Kuo, Pallas, Andrew J. Tock, Xuexia Liu, Stephanie D. Topp, Zhenhui Zhong, Ian R. Henderson, and Christophe Lambing. 2025. “Histone Variant H2A.W7 Represses Meiotic Crossover Formation in Heterochromatin.” *Proceedings of the National Academy of Sciences of the United States of America* 122 (22): e2414166122.

Kursel, Lisa E., and Harmit S. Malik. 2018. “The Cellular Mechanisms and Consequences of Centromere Drive.” *Current Opinion in Cell Biology* 52 (June):58–65.

Lambing, Christophe, Pallas C. Kuo, Andrew J. Tock, Stephanie D. Topp, and Ian R. Henderson. 2020. “ASY1 Acts as a Dosage-Dependent Antagonist of Telomere-Led Recombination and Mediates Crossover Interference in Arabidopsis.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (24): 13647–58.

Lamb, N. E., S. B. Freeman, A. Savage-Austin, D. Pettay, L. Taft, J. Hersey, Y. Gu, et al. 1996. “Susceptible Chiasmate Configurations of Chromosome 21 Predispose to Non-Disjunction in Both Maternal Meiosis I and Meiosis II.” *Nature Genetics* 14 (4): 400–405.

Lampson, Michael A., and Ben E. Black. 2017. “Cellular and Molecular Mechanisms of Centromere Drive.” *Cold Spring Harbor Symposia on Quantitative Biology* 82:249–57.

Lian, Qichao, Bruno Huettel, Birgit Walkemeier, Baptiste Mayjonade, Céline Lopez-Roques, Lisa Gil, Fabrice Roux, Korbinian Schneeberger, and Raphael Mercier. 2024. “A Pan-Genome of 69 Arabidopsis Thaliana Accessions Reveals a Conserved Genome Structure throughout the Global Species Range.” *Nature Genetics* 56 (5): 982–91.

Maheshwari, Shamoni, Ek Han Tan, Allan West, F. Chris H. Franklin, Luca Comai, and Simon W. L. Chan. 2015. “Naturally Occurring Differences in CENH3 Affect Chromosome Segregation in Zygotic Mitosis of Hybrids.” *PLoS Genetics* 11 (1): e1004970.

Malik, Harmit S. 2009. “The Centromere-Drive Hypothesis: A Simple Basis for Centromere Complexity.” *Progress in Molecular and Subcellular Biology* 48:33–52.

McKinley, Kara L., and Iain M. Cheeseman. 2016. “The Molecular Basis for Centromere Identity and Function.” *Nature Reviews. Molecular Cell Biology* 17 (1): 16–29.

Melamed-Bessudo, Cathy, and Avraham A. Levy. 2012. “Deficiency in DNA Methylation Increases Meiotic Crossover Rates in Euchromatic but Not in Heterochromatic Regions in Arabidopsis.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (16): E981–88.

- 1 Melamed-Bessudo, Cathy, Elizabeth Yehuda, Antoine R. Stuitje, and Avraham A. Levy.
2 2005. "A New Seed-Based Assay for Meiotic Recombination in *Arabidopsis Thaliana*."
3 *The Plant Journal: For Cell and Molecular Biology* 43 (3): 458–66.
4
- 5 Mercier, Raphaël, Christine Mézard, Eric Jenczewski, Nicolas Macaisne, and Mathilde
6 Grelon. 2015. "The Molecular Biology of Meiosis in Plants." *Annual Review of Plant*
7 *Biology* 66:297–327.
8
- 9 Miga, Karen H., and Ivan A. Alexandrov. 2021. "Variation and Evolution of Human
10 Centromeres: A Field Guide and Perspective." *Annual Review of Genetics* 55
11 (November):583–602.
12
- 13 Mirouze, Marie, Michal Lieberman-Lazarovich, Riccardo Aversano, Etienne Bucher,
14 Joël Nicolet, Jon Reinders, and Jerzy Paszkowski. 2012. "Loss of DNA Methylation
15 Affects the Recombination Landscape in *Arabidopsis*." *Proceedings of the National*
16 *Academy of Sciences of the United States of America* 109 (15): 5880–85.
17
- 18 Morgan, Chris, John A. Fozard, Matthew Hartley, Ian R. Henderson, Kirsten Bomblies,
19 and Martin Howard. 2021. "Diffusion-Mediated HEI10 Coarsening Can Explain Meiotic
20 Crossover Positioning in *Arabidopsis*." *Nature Communications* 12 (1): 4674.
21
- 22 Naish, Matthew, Michael Alonge, Piotr Wlodzimierz, Andrew J. Tock, Bradley W.
23 Abramson, Anna Schmücker, Terezie Mandáková, et al. 2021. "The Genetic and
24 Epigenetic Landscape of the *Arabidopsis* Centromeres." *Science* 374 (6569): eabi7489.
25
- 26 Nambiar, Mridula, and Gerald R. Smith. 2018. "Pericentromere-Specific Cohesin
27 Complex Prevents Meiotic Pericentric DNA Double-Strand Breaks and Lethal
28 Crossovers." *Molecular Cell* 71 (4): 540–53.e4.
29
- 30 Nuckolls, Nicole L., María Angélica Bravo Núñez, Michael T. Eickbush, Janet M. Young,
31 Jeffrey J. Lange, Jonathan S. Yu, Gerald R. Smith, Sue L. Jaspersen, Harmit S. Malik,
32 and Sarah E. Zanders. 2017. "Genes Are Prolific Dual Poison-Antidote Meiotic Drivers."
33 *eLife* 6 (June). <https://doi.org/10.7554/eLife.26033>.
34
- 35 Ou, Shujun, Weija Su, Yi Liao, Kapeel Chougule, Jireh R. A. Agda, Adam J. Hellinga,
36 Carlos Santiago Blanco Lugo, et al. 2019. "Benchmarking Transposable Element
37 Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline." *Genome*
38 *Biology* 20 (1): 275.
39
- 40 Poethig, R. Scott, William L. Cullina, Erin Doody, Taré Floyd, Jim P. Fouracre, Tieqiang
41 Hu, Mingli Xu, and Jianfei Zhao. 2022. "Short-Interval Traffic Lines: Versatile Tools for
42 Genetic Analysis in *Arabidopsis Thaliana*." *G3* 12 (10).
43 <https://doi.org/10.1093/g3journal/jkac202>.
44
- 45 Rabanal, Fernando A., Maïke Gräff, Christa Lanz, Katrin Fritschi, Victor Llaca, Michelle
46 Lang, Pablo Carbonell-Bejerano, Ian Henderson, and Detlef Weigel. 2022. "Pushing the

- Limits of HiFi Assemblies Reveals Centromere Diversity between Two *Arabidopsis Thaliana* Genomes." *Nucleic Acids Research* 50 (21): 12309–27.
- Rockmill, Beth, Karen Voelkel-Meiman, and G. Shirleen Roeder. 2006. "Centromere-Proximal Crossovers Are Associated with Precocious Separation of Sister Chromatids during Meiosis in *Saccharomyces Cerevisiae*." *Genetics* 174 (4): 1745–54.
- Rommel Fuentes, Roven, Thamara Hesselink, Ronald Nieuwenhuis, Linda Bakker, Elio Schijlen, Willem van Dooijeweert, Sara Diaz Trivino, et al. 2020. "Meiotic Recombination Profiling of Interspecific Hybrid F1 Tomato Pollen by Linked Read Sequencing." *The Plant Journal: For Cell and Molecular Biology* 102 (3): 480–92.
- Rowan, Beth A., Darren Heavens, Tatiana R. Feuerborn, Andrew J. Tock, Ian R. Henderson, and Detlef Weigel. 2019. "An Ultra High-Density *Arabidopsis Thaliana* Crossover Map That Refines the Influences of Structural Variation and Epigenetic Features." *Genetics* 213 (3): 771–87.
- Salomé, P. A., K. Bomblies, J. Fitz, R. A. E. Laitinen, N. Warthmann, L. Yant, and D. Weigel. 2012. "The Recombination Landscape in *Arabidopsis Thaliana* F2 Populations." *Heredity* 108 (4): 447–55.
- Séguéla-Arnaud, Mathilde, Wayne Crismani, Cécile Larchevêque, Julien Mazel, Nicole Froger, Sandrine Choinard, Afef Lemhemdi, et al. 2015. "Multiple Mechanisms Limit Meiotic Crossovers: TOP3 α and Two BLM Homologs Antagonize Crossovers in Parallel to FANCM." *Proceedings of the National Academy of Sciences* 112 (15): 4713–18.
- Seymour, Danelle K., Eunyoung Chae, Burak I. Arioz, Daniel Koenig, and Detlef Weigel. 2019. "Transmission Ratio Distortion Is Frequent in *Arabidopsis Thaliana* Controlled Crosses." *Heredity* 122 (3): 294–304.
- Showman, Soyeon, Paul B. Talbert, Yiling Xu, Richard O. Adeyemi, and Steven Henikoff. 2024. "Expansion of Human Centromeric Arrays in Cells Undergoing Break-Induced Replication." *Cell Reports* 43 (3): 113851.
- Smith, George P. 1976. "Evolution of Repeated DNA Sequences by Unequal Crossover." *Science*. <https://doi.org/10.1126/science.1251186>.
- Son, Namil, Heejin Kim, Jaeil Kim, Jihye Park, Dohwan Byun, Sang-Jun Park, Hyein Kim, et al. 2025. "The Histone Variant H2A.W Restricts Heterochromatic Crossovers in." *Proceedings of the National Academy of Sciences of the United States of America* 122 (14): e2413698122.
- Stiehler, Felix, Marvin Steinborn, Stephan Scholz, Daniela Dey, Andreas P. M. Weber, and Alisandra K. Denton. 2021. "Helixer: Cross-Species Gene Annotation of Large Eukaryotic Genomes Using Deep Learning." *Bioinformatics (Oxford, England)* 36 (22-23): 5291–98.

- 1 Stirling, D. R., M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, and A.
2 Goodman. 2021. "CellProfiler 4: Improvements in Speed, Utility and Usability." *BMC*
3 *Bioinformatics* 22 (1). <https://doi.org/10.1186/s12859-021-04344-9>.
4
- 5 Sturtevant, A. H. 1917. "Genetic Factors Affecting the Strength of Linkage in
6 *Drosophila*." *Proceedings of the National Academy of Sciences of the United States of*
7 *America* 3 (9): 555–58.
8
- 9 Szymanska-Lejman, Maja, Wojciech Dziegielewski, Julia Dluzewska, Nadia Kbiri, Anna
10 Bieluszewska, R. Scott Poethig, and Piotr A. Ziolkowski. 2023. "The Effect of DNA
11 Polymorphisms and Natural Variation on Crossover Hotspot Activity in Arabidopsis
12 Hybrids." *Nature Communications* 14 (1): 33.
13
- 14 Talbert, Paul B., Ricardo Masuelli, Anand P. Tyagi, Luca Comai, and Steven Henikoff.
15 2002. "Centromeric Localization and Adaptive Evolution of an Arabidopsis Histone H3
16 Variant." *The Plant Cell*. <https://doi.org/10.1105/tpc.010425>.
17
- 18 Törjék, Ottó, Hanna Witucka-Wall, Rhonda C. Meyer, Maria von Korff, Barbara
19 Kusterer, Carsten Rautengarten, and Thomas Altmann. 2006. "Segregation Distortion in
20 Arabidopsis C24/Col-0 and Col-0/C24 Recombinant Inbred Line Populations Is due to
21 Reduced Fertility Caused by Epistatic Interaction of Two Loci." *TAG. Theoretical and*
22 *Applied Genetics. Theoretische Und Angewandte Genetik* 113 (8): 1551–61.
23
- 24 Underwood, Charles J., Kyuha Choi, Christophe Lambing, Xiaohui Zhao, Heïdi Serra,
25 Filipe Borges, Joe Simorowski, et al. 2018. "Epigenetic Activation of Meiotic
26 Recombination near Arabidopsis Thaliana Centromeres via Loss of H3K9me2 and Non -
27 CG DNA Methylation." *Genome Research* 28 (4): 519–31.
28
- 29 Villeneuve, A. M., and K. J. Hillers. 2001. "Whence Meiosis?" *Cell* 106 (6): 647–50.
30
- 31 Vincenten, Nadine, Lisa-Marie Kuhl, Isabel Lam, Ashwini Oke, Alastair R. W. Kerr,
32 Andreas Hochwagen, Jennifer Fung, Scott Keeney, Gerben Vader, and Adèle L.
33 Marston. 2015. "The Kinetochore Prevents Centromere-Proximal Crossover
34 Recombination during Meiosis." *eLife* 4 (December).
35 <https://doi.org/10.7554/eLife.10850>.
36
- 37 Webb, Mary C., and Brian E. S. Gunning. 1990. "Embryo Sac Development in
38 Arabidopsis Thaliana." *Sexual Plant Reproduction* 3 (4): 244–56.
39
- 40 Wlodzimierz, Piotr, Michael Hong, and Ian R. Henderson. 2023. "TRASH: Tandem
41 Repeat Annotation and Structural Hierarchy." *Bioinformatics (Oxford, England)* 39 (5).
42 <https://doi.org/10.1093/bioinformatics/btad308>.
43
- 44 Wlodzimierz, Piotr, Fernando A. Rabanal, Robin Burns, Matthew Naish, Elias Primetis,
45 Alison Scott, Terezie Mandáková, et al. 2023. "Cycles of Satellite and Transposon
46 Evolution in Arabidopsis Centromeres." *Nature* 618 (7965): 557–65.

Wu, Gang, Gabrielle Rossidivito, Tieqiang Hu, Yosef Berlyand, and R. Scott Poethig. 2015. "Traffic Lines: New Tools for Genetic Analysis in *Arabidopsis Thaliana*." *Genetics* 200 (1): 35–45.

Yelina, Nataliya E., Kyuha Choi, Liudmila Chelysheva, Malcolm Macaulay, Bastiaan de Snoo, Erik Wijnker, Nigel Miller, et al. 2012. "Epigenetic Remodeling of Meiotic Crossover Frequency in *Arabidopsis Thaliana* DNA Methyltransferase Mutants." *PLoS Genetics* 8 (8): e1002844.

Yelina, Nataliya E., Christophe Lambing, Thomas J. Hardcastle, Xiaohui Zhao, Bruno Santos, and Ian R. Henderson. 2015. "DNA Methylation Epigenetically Silences Crossover Hot Spots and Controls Chromosomal Domains of Meiotic Recombination in *Arabidopsis*." *Genes & Development* 29 (20): 2183–2202.

Ziolkowski, Piotr A., Luke E. Berchowitz, Christophe Lambing, Nataliya E. Yelina, Xiaohui Zhao, Krystyna A. Kelly, Kyuha Choi, et al. 2015. "Juxtaposition of Heterozygous and Homozygous Regions Causes Reciprocal Crossover Remodelling via Interference during *Arabidopsis* Meiosis." *eLife* 4 (March). <https://doi.org/10.7554/eLife.03708>.

Ziolkowski, Piotr A., Charles J. Underwood, Christophe Lambing, Marina Martinez-Garcia, Emma J. Lawrence, Liliana Ziolkowska, Catherine Griffin, et al. 2017. "Natural Variation and Dosage of the HEI10 Meiotic E3 Ligase Control Crossover Recombination." *Genes & Development* 31 (3): 306–17.

Figure 1. Quantifying centromere-proximal crossover frequency and segregation distortion using fluorescent reporters in *Arabidopsis* hybrids. **A.** Chromosome ideograms showing the location of FTL T-DNAs used in this study, and the *CEN178* satellite repeat arrays (dark grey), against the Col-CEN genome assembly (Naish *et al.* 2021). Red and green triangles indicate the location of FTL T-DNAs expressing dsRED and GFP, respectively (Wu *et al.* 2015; Poethig *et al.* 2022). **B.** Geographic map showing the origin of the *Arabidopsis* accessions studied. Accession locations are colour-coded by SNP PCA chromosome arm similarity groups (Włodzimierz *et al.* 2023b); Eurasian (purple), Iberian relict (pink), and non-Iberian relict (turquoise). **C.** Diagram showing genetic segregation of a hemizygous FTL chromosome through meiosis I (MI) and II (MII). Homologous chromosomes are shown in blue and orange, with the orange genotype

representing Col-0 FTL chromosomes, and FTL transgenes shown by red and green triangles. Two segregation scenarios are shown, either with or without a single crossover within the FTL interval. Alongside are representative micrographs of *FTL*/+ hemizygous seed imaged in brightfield, or visualising red or green fluorescence. Scale bar=3 mm. **D.** A representative histogram of red fluorescence (dsRED) of seed from a self-fertilized *CEN3*/+ hemizygous plant. The dotted line indicates the threshold that divides non-fluorescent (light grey) and red fluorescent (red) seed. To calculate crossover frequency (centiMorgan, cM) within the FTL genetic interval, the formula shown is applied to the counts of green (N_G) and red (N_R) fluorescent seed, and the total seed (N_T). **E.** Punnett square showing the expected male and female gametes produced by an FTL hemizygote, and the expected Mendelian inheritance ratio of three fluorescent to one non-fluorescent progeny, following self-fertilization. On the right, a diagram indicating that higher or lower ratios of fluorescent to non-fluorescent seed can indicate segregation distortion in progeny seed.

Figure 2. Meiotic crossover frequency, segregation distortion and genetic and epigenetic structure within *CENTROMERE1*. **A.** Chromosome ideogram showing the location of FTL T-DNAs used in this study, together with the position of the *CEN178* satellite repeat arrays (dark grey), shown against the Col-CEN genome assembly (Naish *et al.* 2021). Red and green triangles indicate the location of FTL T-DNAs expressing dsRED and GFP in seed from the *napA* promoter, respectively (Wu *et al.* 2015; Poethig *et al.* 2022). **B.** Crossover frequency (centiMorgan) values of individual replicates of F_1 hybrids between the accession listed on the x-axis, and the *CEN1* FTL reporters in a Col-

0 background. Mean values are shown by orange diamonds. The inbred Col-0×Col-0 FTL control is highlighted in green, via the horizontal dashed line. Stars at the top of the plot indicate hybrids with significantly different crossover frequencies to the inbred strain based on Chi-square tests ($P<0.05$). Beneath, the same data are analysed for the ratios of green (circles, short dashed line) and red fluorescent (triangles, long dashed line) seed to non-fluorescent seed and presented in the same way. **C.** Sequence similarity dot plots comparing the *CEN178* arrays from a subset of the accessions analyzed in B. Red and blue shading indicate sequence similarity on forward and reverse strands, using a 161 base pair window. Beneath the dotplot, the cM value for the corresponding *CEN1* hybrid from B. is shown, which is proportionally shaded. Additionally, beneath the dot plot are text indicators showing which accessions showed segregation distortion in B (Colour:Non-Colour, C:NonC). **D.** For the Col-0, Jm-0, and Ler-0 Eurasian accessions, and the Cvi-0, Elk-3, Rab-1, and Tanz-1 relict accessions, the CENH3 ChIP-seq enrichment (upper, purple) and DNA methylation (lower) in CG (dark pink), CHG (pink) and CHH (light pink) contexts across the *CEN1* region are shown. The position of *CEN178* arrays are indicated by ticks on the x-axis on the forward (red) or reverse (blue) strand.

Figure 3. Meiotic crossover frequency, segregation distortion and genetic and epigenetic structure within *CENTROMERE2*. **A.** Chromosome ideogram showing the location of FTL T-DNAs used in this study, together with the position of the *CEN178* satellite repeat arrays (dark grey), against the Col-CEN genome assembly (Naish *et al.* 2021). Red and green triangles indicate the location of FTL T-DNAs expressing dsRED and GFP in seed from the *napA* promoter, respectively (Wu *et al.* 2015; Poethig *et al.*

2022). **B.** Crossover frequency (centiMorgan) values of individual replicates of F_1 hybrids between the accession listed on the x-axis, and the *CEN2* FTL reporters in a Col-0 background. Mean values are shown by orange diamonds. The inbred Col-0×Col-0-FTL control is highlighted in green, via the horizontal dashed line. Stars at the top of the plot indicate hybrids with significantly different crossover frequencies to the inbred strain based on Chi-square tests ($P<0.05$). Beneath, the same data are analysed for the ratios of green (circles, short dashed line) and red (triangles, long dashed line) fluorescent seed to non-fluorescent seed and presented in the same way. **C.** Sequence similarity dot plots comparing the *CEN178* arrays from a subset of the accessions analyzed in B. Red and blue shading indicate sequence similarity on forward and reverse strands, using a 161 base pair window. Beneath the dotplot, the cM value for the corresponding *CEN2* hybrid from B. is shown, which is proportionally shaded. Additionally, beneath the dot plot are text indicators showing which accessions showed segregation distortion in B (Colour:Non-Colour, C:NonC). **D.** For the Col-0, Jm-0, and Ler-0 Eurasian accessions, and the Cvi-0, Elk-3, Rab-1, and Tanz-1 relict accessions, the CENH3 ChIP-seq enrichment (upper, purple) and DNA methylation (lower) in CG (dark pink), CHG (pink) and CHH (light pink) contexts across the *CEN1* region are shown. The position of *CEN178* arrays are indicated by ticks on the x-axis on the forward (red) or reverse (blue) strand.

Figure 4. Meiotic crossover frequency, segregation distortion and genetic and epigenetic structure within *CENTROMERE3*. **A.** Chromosome ideogram showing the location of FTL T-DNAs used in this study, together with the position of the *CEN178* satellite repeat arrays (dark grey), against the Col-CEN genome assembly (Naish *et al.*

2021). Red and green triangles indicate the location of FTL T-DNAs expressing dsRED and GFP in seed from the *napA* promoter, respectively (Wu *et al.* 2015; Poethig *et al.* 2022). **B.** Crossover frequency (centiMorgan) values of individual replicates of F₁ hybrids between the accession listed on the x-axis, and the *CEN3* FTL reporters in a Col-0 background. Mean values are shown by orange diamonds. The inbred Col-0×Col-0-FTL control is highlighted in green, via the horizontal dashed line. Stars at the top of the plot indicate hybrids with significantly different crossover frequencies to the inbred strain based on Chi-square tests ($P < 0.05$). Beneath, the same data are analysed for the ratios of green (circles, short dashed line) and red (triangles, long dashed line) fluorescent seed to non-fluorescent seed and presented in the same way. **C.** Sequence similarity dot plots comparing the *CEN178* arrays from a subset of the accessions analyzed in B. Red and blue shading indicate sequence similarity, on forward and reverse strands, using a 161 base pair window. Beneath the dotplot, the cM value for the corresponding *CEN3* hybrid from B. is shown, which is proportionally shaded. Additionally, beneath the dot plot are text indicators showing which accessions showed segregation distortion in B (Colour:Non-Colour, C:Non-C). **D.** For the Col-0, Jm-0, and Ler-0 Eurasian accessions, and the Cvi-0, Elk-3, Rab-1, and Tanz-1 relict accessions, the CENH3 ChIP-seq enrichment (upper, purple) and DNA methylation (lower) in CG (dark pink), CHG (pink) and CHH (light pink) contexts across the *CEN1* region are shown. The position of *CEN178* arrays are indicated by ticks on the x-axis on the forward (red) or reverse (blue) strand.

Figure 5. Meiotic crossover frequency, segregation distortion and genetic and epigenetic structure within *CENTROMERE4*. **A.** Chromosome ideogram showing the

location of FTL T-DNAs used in this study, together with the position of the *CEN178* satellite repeat arrays (dark grey), against the Col-CEN genome assembly (Naish *et al.* 2021). Red and green triangles indicate the location of FTL T-DNAs expressing dsRED and GFP in seed from the *napA* promoter, respectively (Wu *et al.* 2015; Poethig *et al.* 2022). **B.** Crossover frequency (centiMorgan) values of individual replicates of F₁ hybrids between the accession listed on the x-axis, and the *CEN4* FTL reporters in a Col-0 background. Mean values are shown by orange diamonds. The inbred Col-0×Col-0-FTL control is highlighted in green, via the horizontal dashed line. Stars at the top of the plot indicate hybrids with significantly different crossover frequencies to the inbred strain based on Chi-square tests ($P<0.05$). Beneath, the same data are analysed for the ratios of green (circles, short dashed line) and red (triangles, long dashed line) fluorescent seed to non-fluorescent seed and presented in the same way. **C.** Sequence similarity dot plots comparing the *CEN178* arrays from a subset of the accessions analyzed in B. Red and blue shading indicate sequence similarity on forward and reverse strands, using a 161 base pair window. Beneath the dotplot, the cM value for the corresponding *CEN4* hybrid from B. is shown, which is proportionally shaded. Additionally, beneath the dot plot are text indicators showing which accessions showed segregation distortion in B (Colour:Non-Colour, C:Non-C). **D.** For the Col-0, Jm-0, and Ler-0 Eurasian accessions, and the Cvi-0, Elk-3, Rab-1, and Tanz-1 relict accessions, the CENH3 ChIP-seq enrichment (upper, purple) and DNA methylation (lower) in CG (dark pink), CHG (pink) and CHH (light pink) contexts across the *CEN1* region are shown. The position of *CEN178* arrays are indicated by ticks on the x-axis on the forward (red) or reverse (blue) strand.

1 **Figure 6. Meiotic crossover frequency, segregation distortion and genetic and**
2 **epigenetic structure within *CENTROMERE5*. A.** Chromosome ideogram showing the
3 location of FTL T-DNAs used in this study, together with the position of the *CEN178*
4 satellite repeat arrays (dark grey), against the Col-CEN genome assembly (Naish *et al.*
5 2021). Red and green triangles indicate the location of FTL T-DNAs expressing dsRED
6 and GFP in seed from the *napA* promoter, respectively (Wu *et al.* 2015; Poethig *et al.*
7 2022). **B.** Crossover frequency (centiMorgan) values of individual replicates of F₁ hybrids
8 between the accession listed on the x-axis, and the *CEN5* FTL reporters in a Col-0
9 background. Mean values are shown by orange diamonds. The inbred Col-0×Col-0-FTL
10 control is highlighted in green, via the horizontal dashed line. Stars at the top of the plot
11 indicate hybrids with significantly different crossover frequencies to the inbred strain
12 based on Chi-square tests. Beneath, the same data are analysed for the ratios of green
13 (circles, short dashed line) and red (triangles, long dashed line) fluorescent seed to non-
14 fluorescent seed and presented in the same way. **C.** Sequence similarity dot plots
15 comparing the *CEN178* arrays from a subset of the accessions analyzed in B. Red and
16 blue shading indicate sequence similarity on forward and reverse strands, using a 161
17 base pair window. Beneath the dotplot, the cM value for the corresponding *CEN5* hybrid
18 from B. is shown, which is proportionally shaded. Additionally, beneath the dot plot are
19 text indicators showing which accessions showed segregation distortion in B. **D.** For the
20 Col-0, Jm-0, and Ler-0 Eurasian accessions, and the Cvi-0, Elk-3, Rab-1, and Tanz-1
21 relict accessions, the CENH3 ChIP-seq enrichment (upper, purple) and DNA methylation
22 (lower) in CG (dark pink), CHG (pink) and CHH (light pink) contexts across the *CEN1*

region are shown. The position of *CEN178* arrays are indicated by ticks on the x-axis on the forward (red) or reverse (blue) strand.

Figure 7. Crossover frequency and segregation distortion across twelve

Arabidopsis hybrids relative to Col-0 inbreds. A. Violin plots showing the variation in cM/Mb values for different *CEN* FTL intervals in different accession hybrids, and Col-0 inbreds. **B.** Linear modelling results showing the % variation in crossover frequency explained by accession (pink), chromosome (blue), the interaction between accession and chromosome (purple), and residuals (grey). **C.** As for A, but showing variation in red to non-red fluorescence ratios for each *CEN* FTL interval in different accession hybrids, and Col-0 inbreds. **D.** As for B, but modelling variation in red to non-red fluorescence ratios. **E.** For each chromosome, a sequence identity heat map is shown for Col-0 compared to Tanz-1 (where 4 of 5 centromeres showed significantly distorted inheritance against Tanz-1) and IP-Alo-19 (where 4 of 5 centromeres showed significantly higher crossover frequency compared to inbreds) centromeres. Text labels indicate whether the Tanz-1 and Alo-19 hybrids showed significantly higher (up arrow), lower (down arrow), or no significant (n.s) change, with the same signifiers used to indicate segregation distortion of the FTL T-DNAs, either in favour of Col-0 (up arrow), against (down arrow), or no significant change (n.s.). Beneath are histograms showing the number of estimates with specific % identity in the sequence identity heat maps for each accession and each centromere.

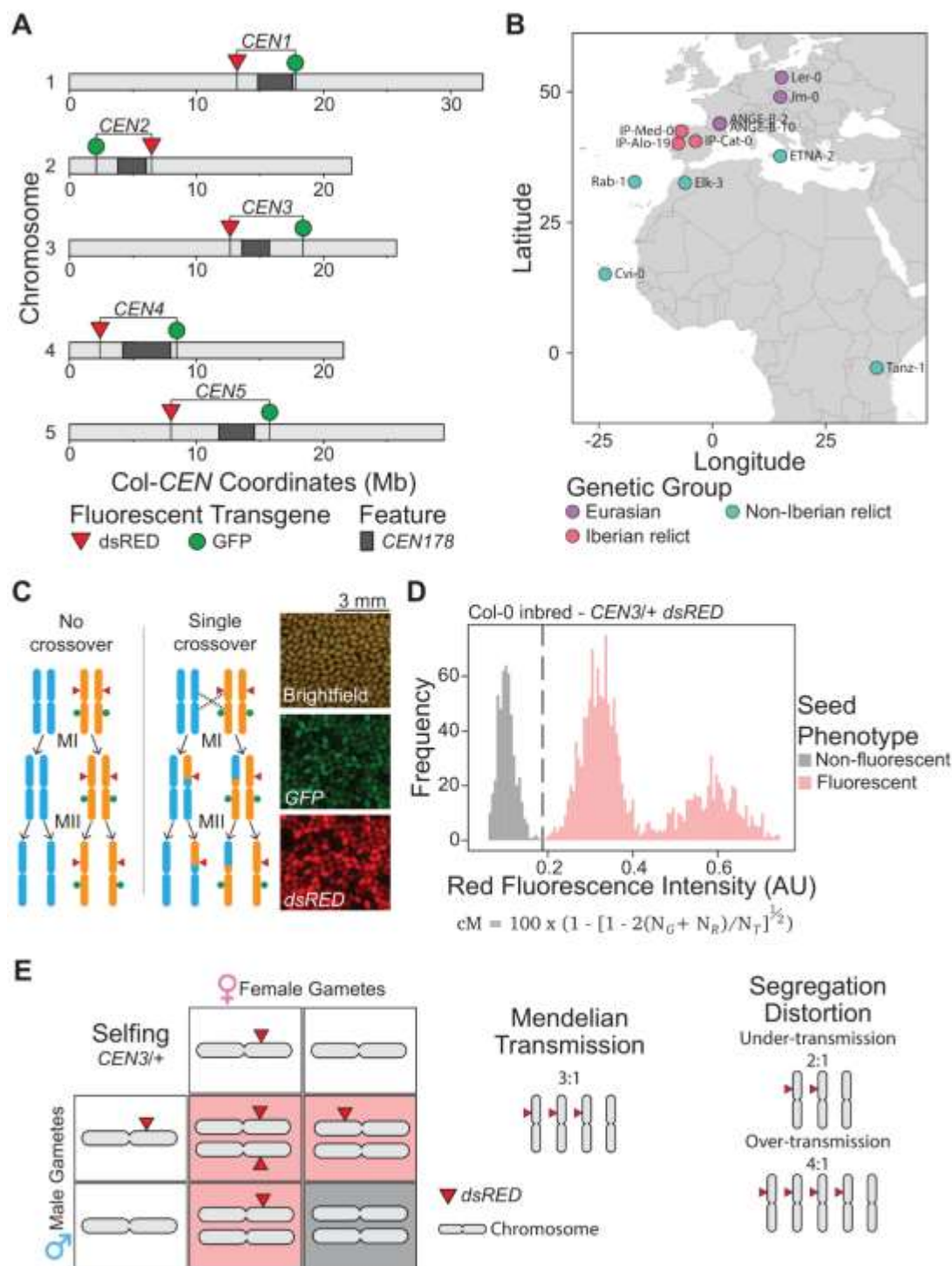


Figure 1
210x297 mm (x DPI)

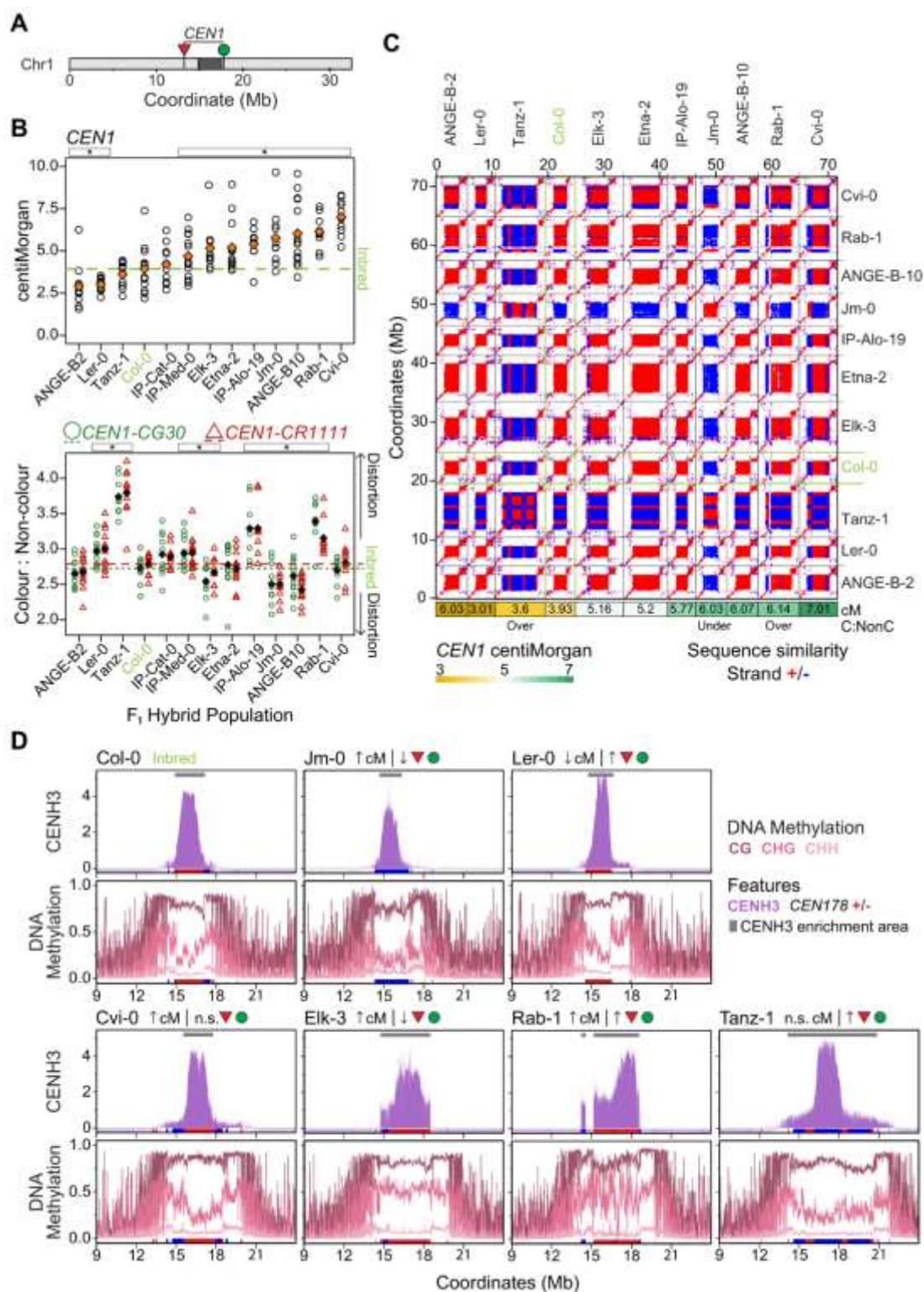
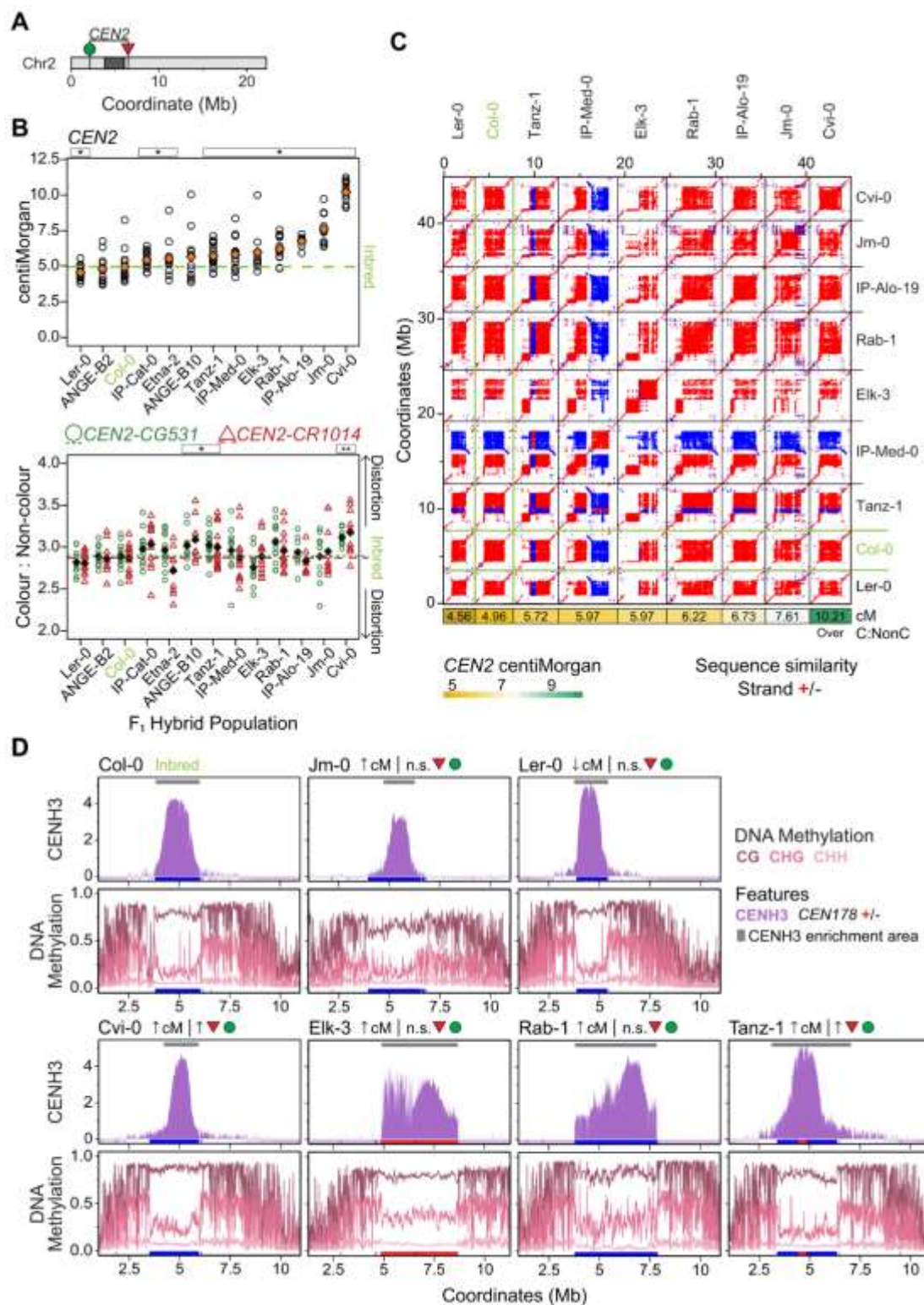


Figure 2
210x297 mm (x DPI)



1
2
3
4

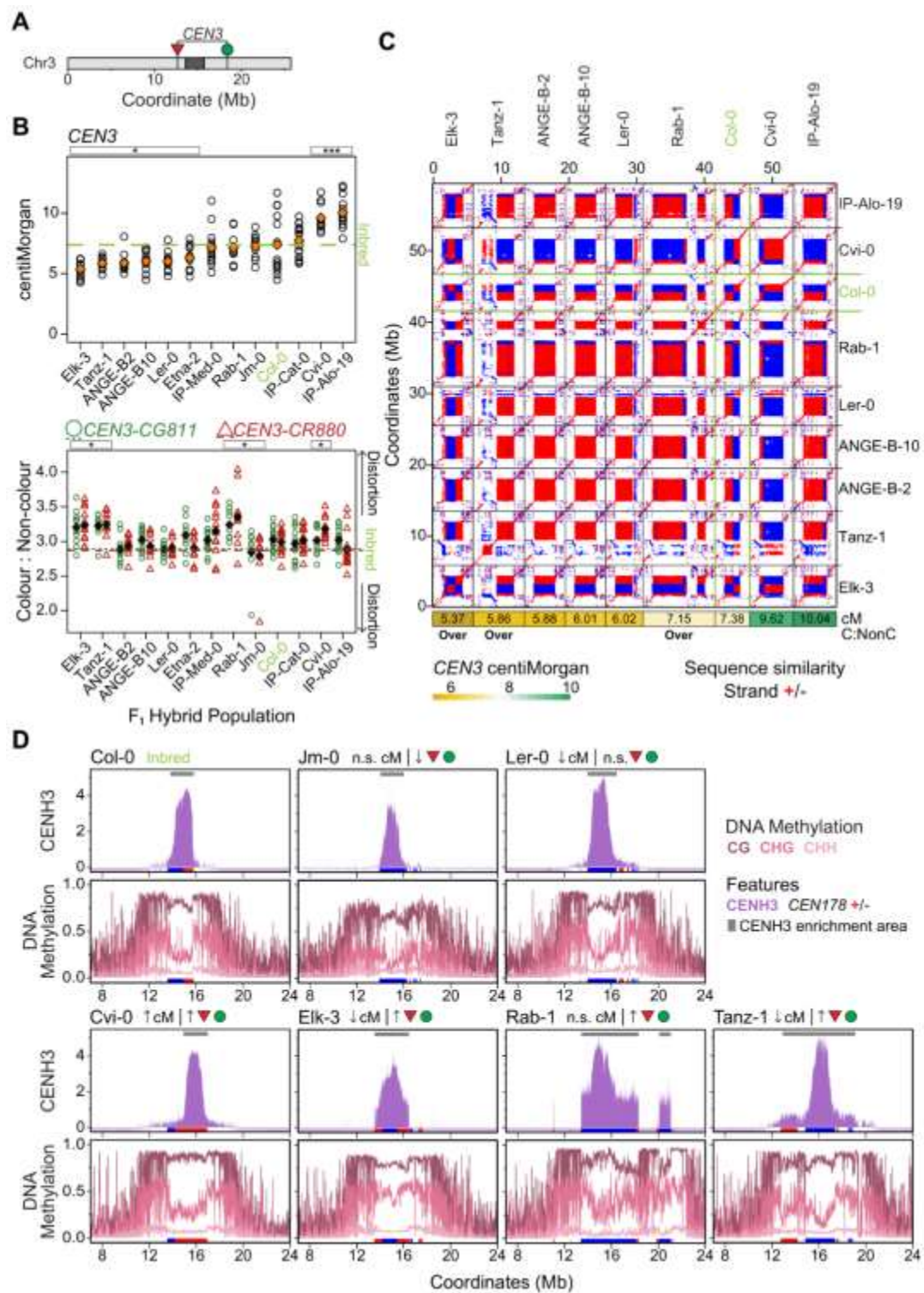


Figure 4
210x297 mm (x DPI)

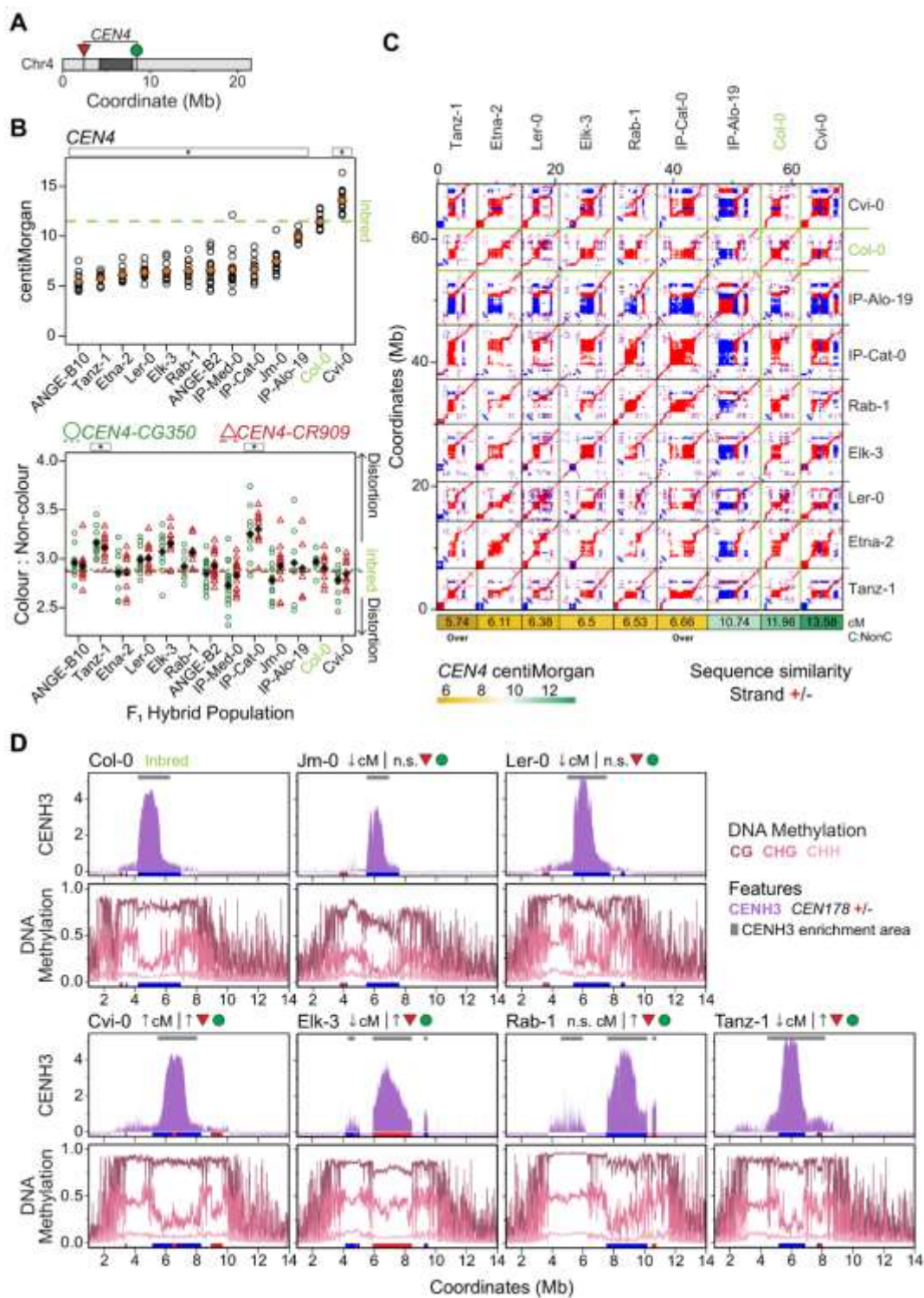


Figure 5
210x297 mm (x DPI)

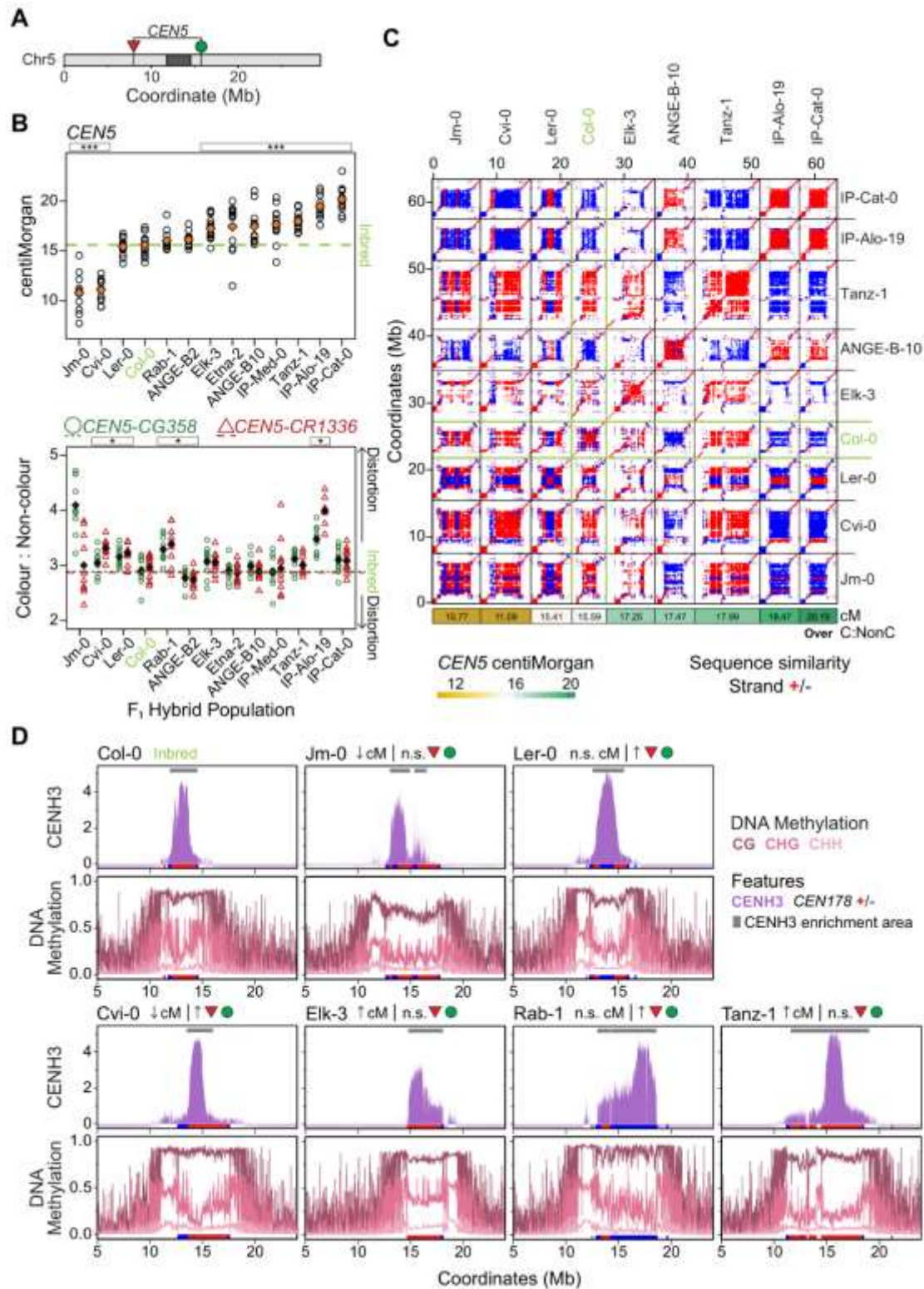


Figure 6
210x297 mm (x DPI)

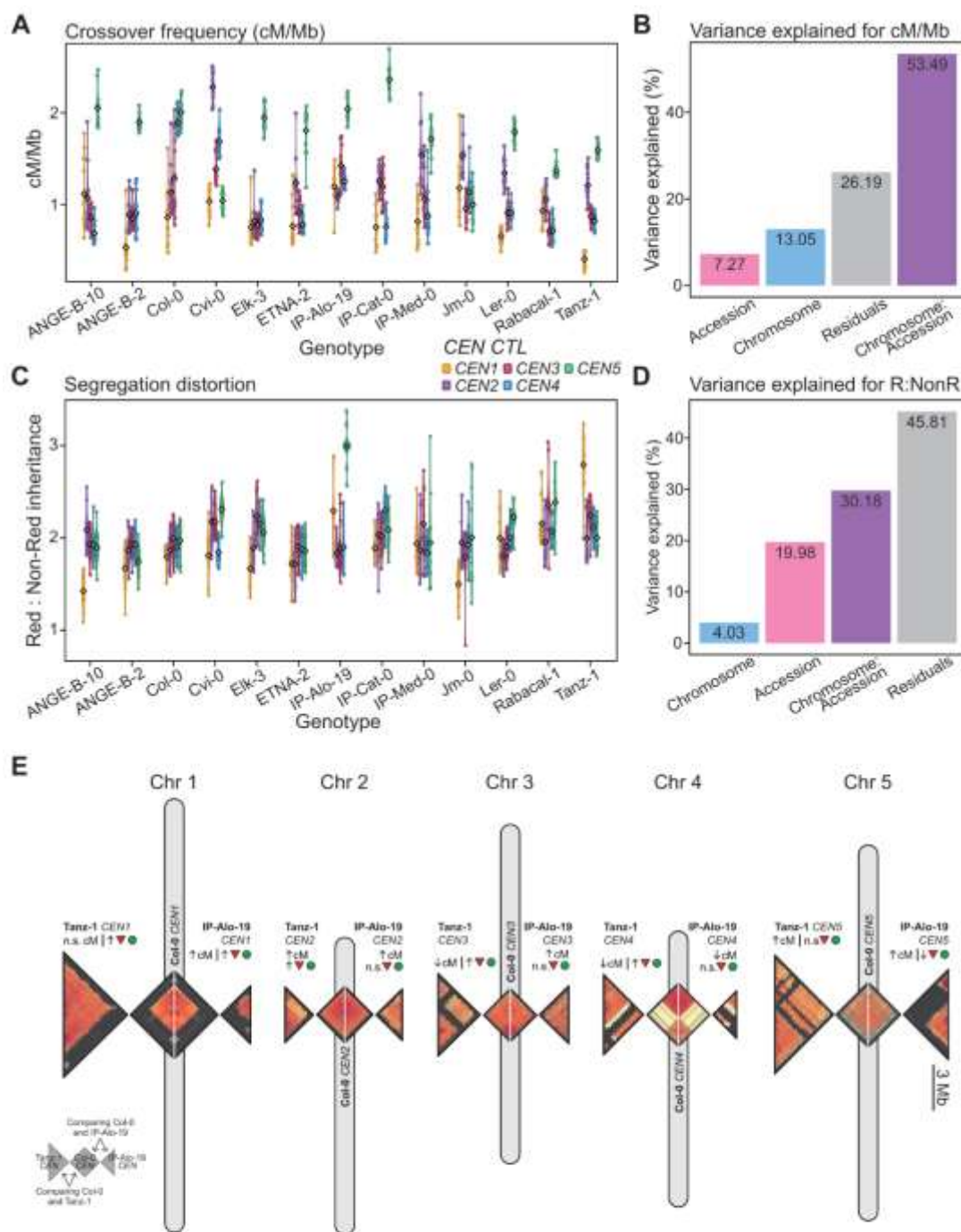


Figure 7
210x297 mm (x DPI)