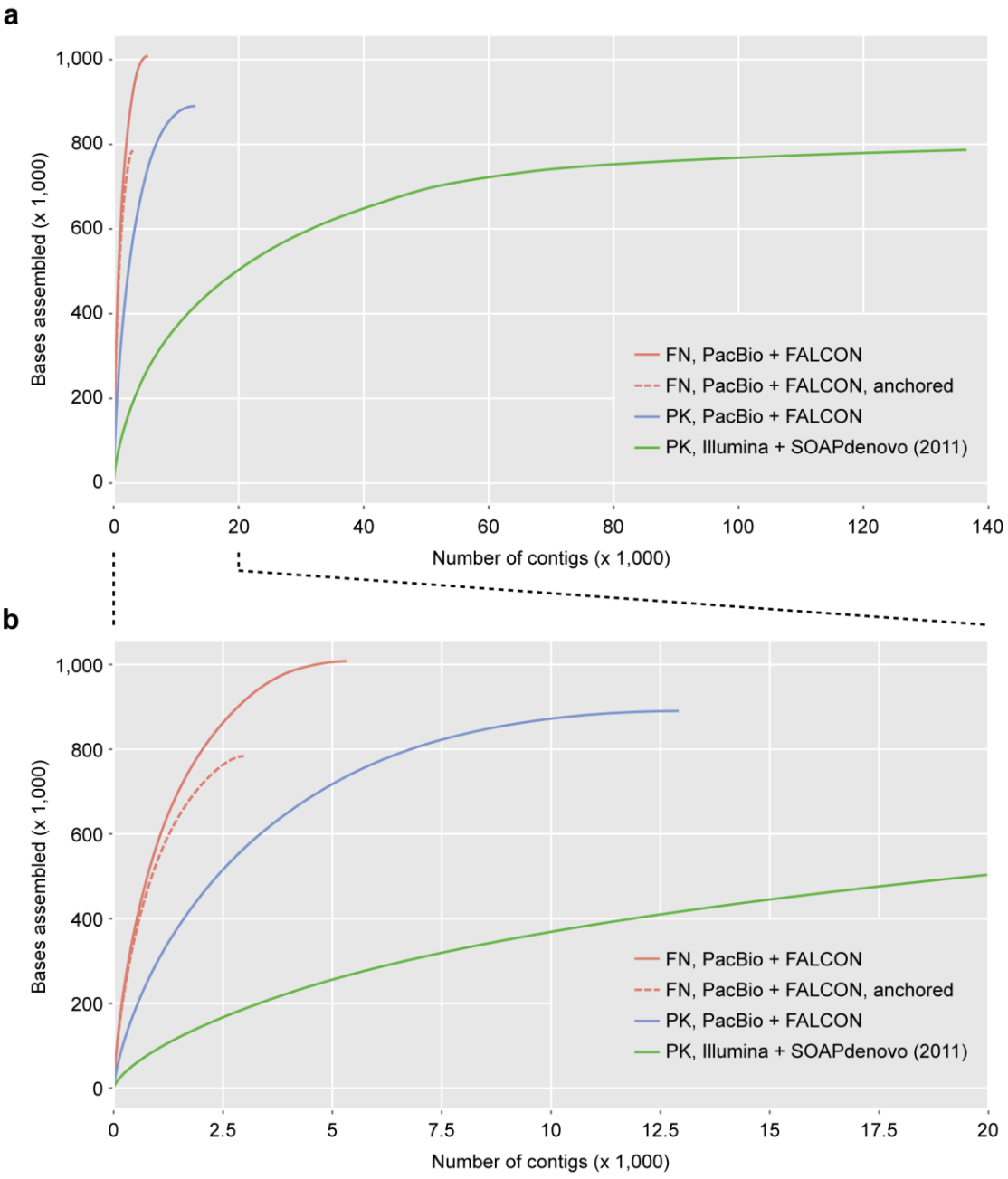
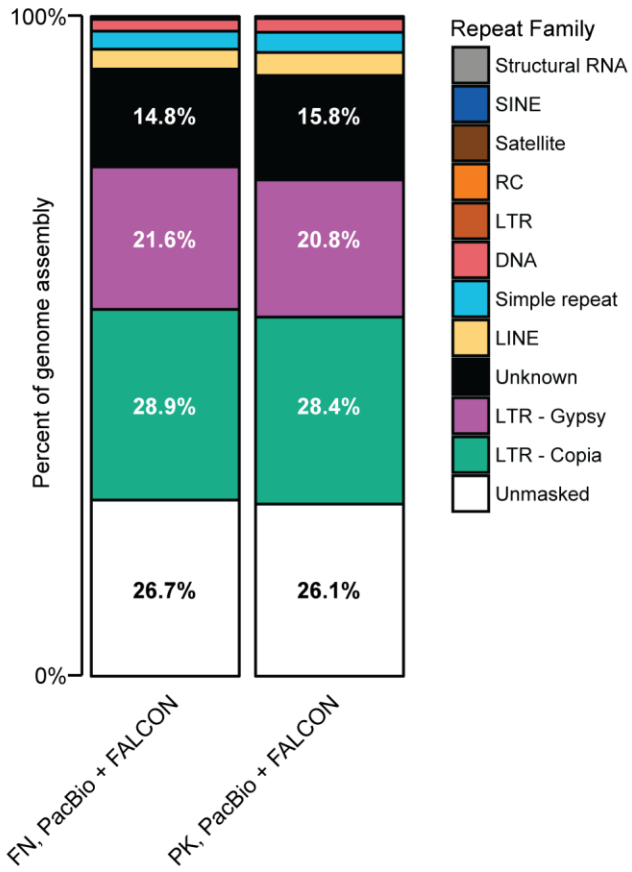
Supplemental figures and table for:
A physical and genetic map of *Cannabis sativa* identifies extensive rearrangement at the THC/CBD acid synthase locus

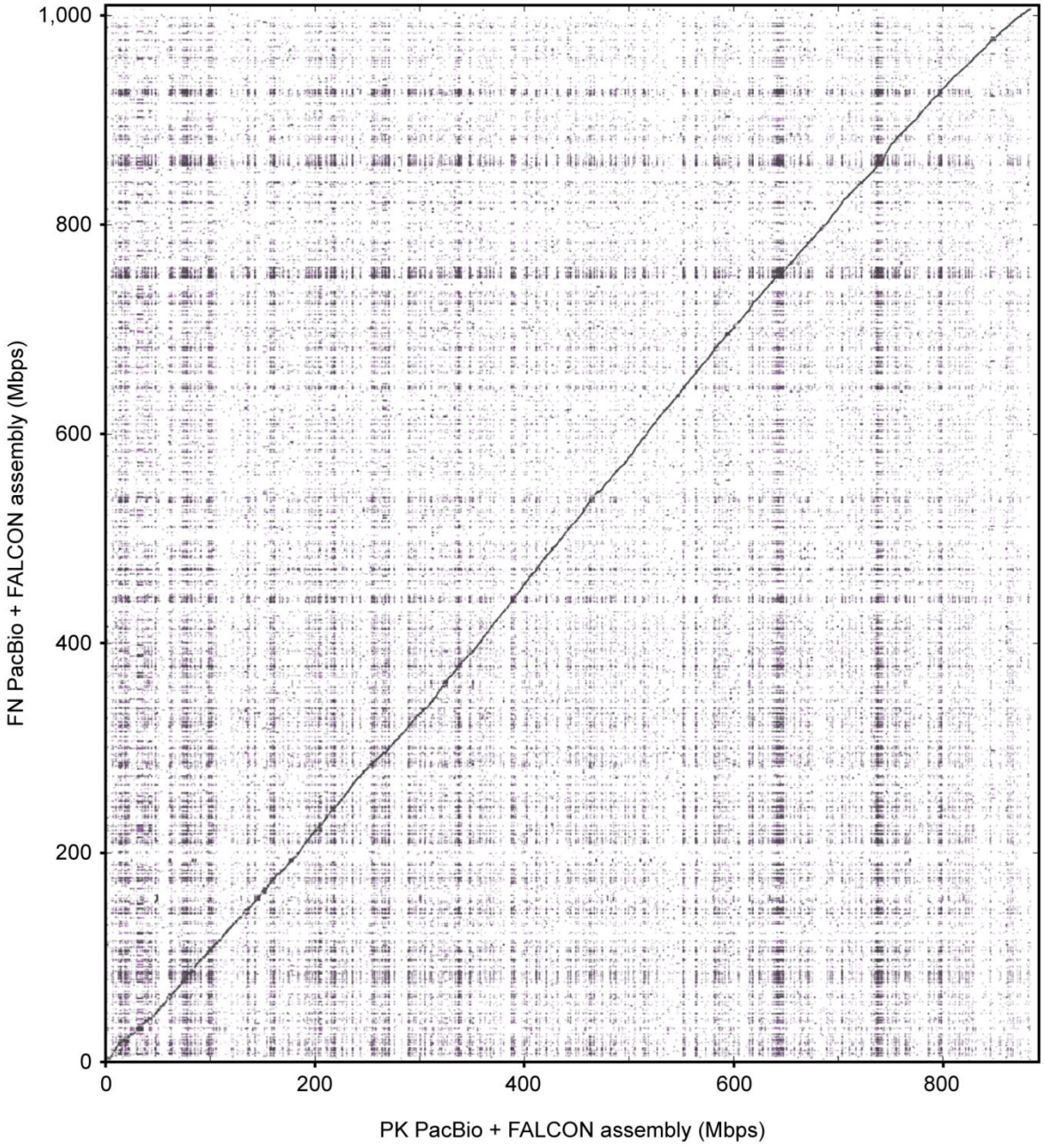
Kaitlin U. Laverty, Jake M. Stout, Mitchell J. Sullivan, Hardik Shah, Navdeep Gill, Larry Holbrook, Gintaras Deikus, Robert Sebra, Timothy R. Hughes, Jonathan E. Page, and Harm van Bakel



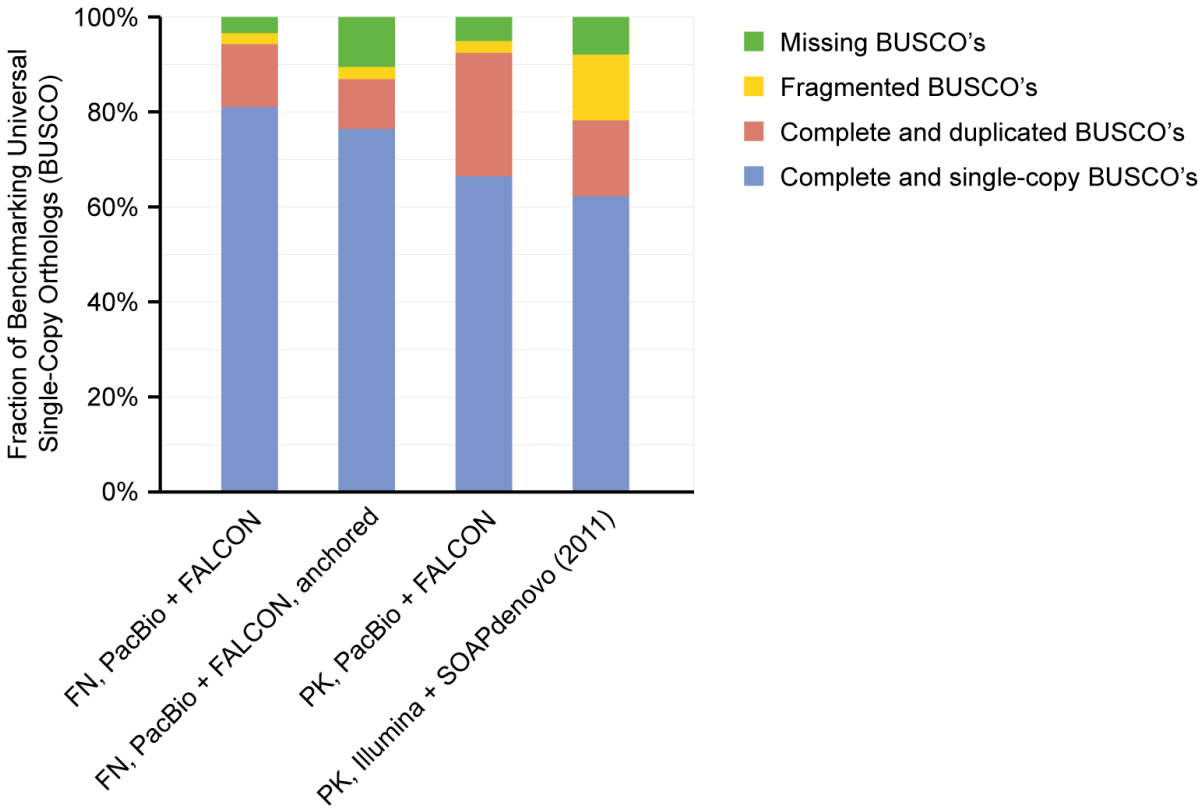
Supplemental Figure 1. Cumulative contig sizes for PK and FN assemblies.
(a) The plot compares the cumulative contig size distribution of the PK and FN PacBio Falcon assemblies to the SOAPdenovo assembly of illumina data in the 2011 draft genome publication. The anchored FN assembly consists of the subset of scaffolds that could be placed in the genetic map. **(b)** Same as (a) but for the first 20,000 contigs.



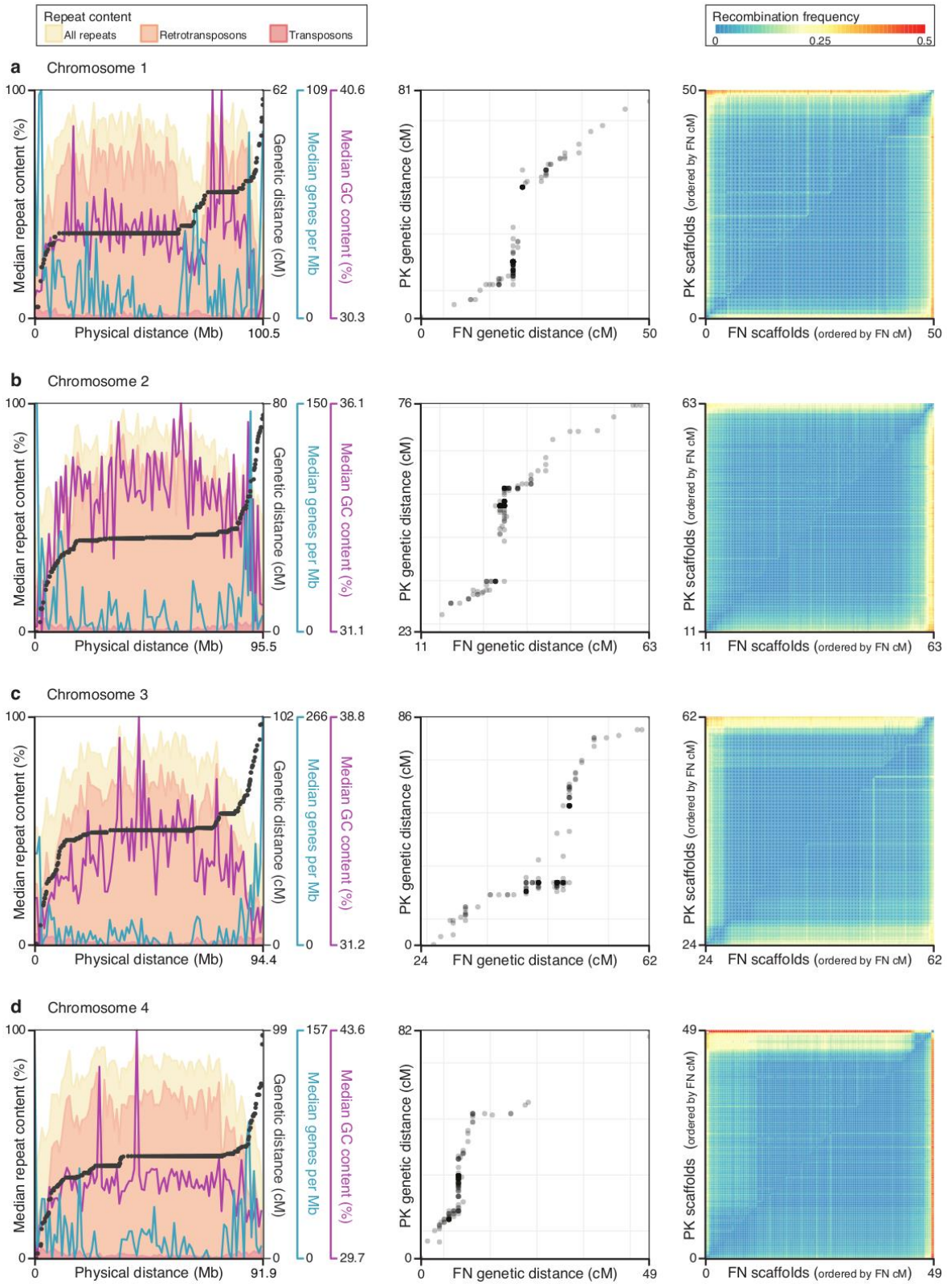
Supplemental Figure 2. Results of repeatmasker analysis.
Breakdown of major repeat families found in the **(a)** FN and **(b)** PK PacBio + FALCON assemblies.



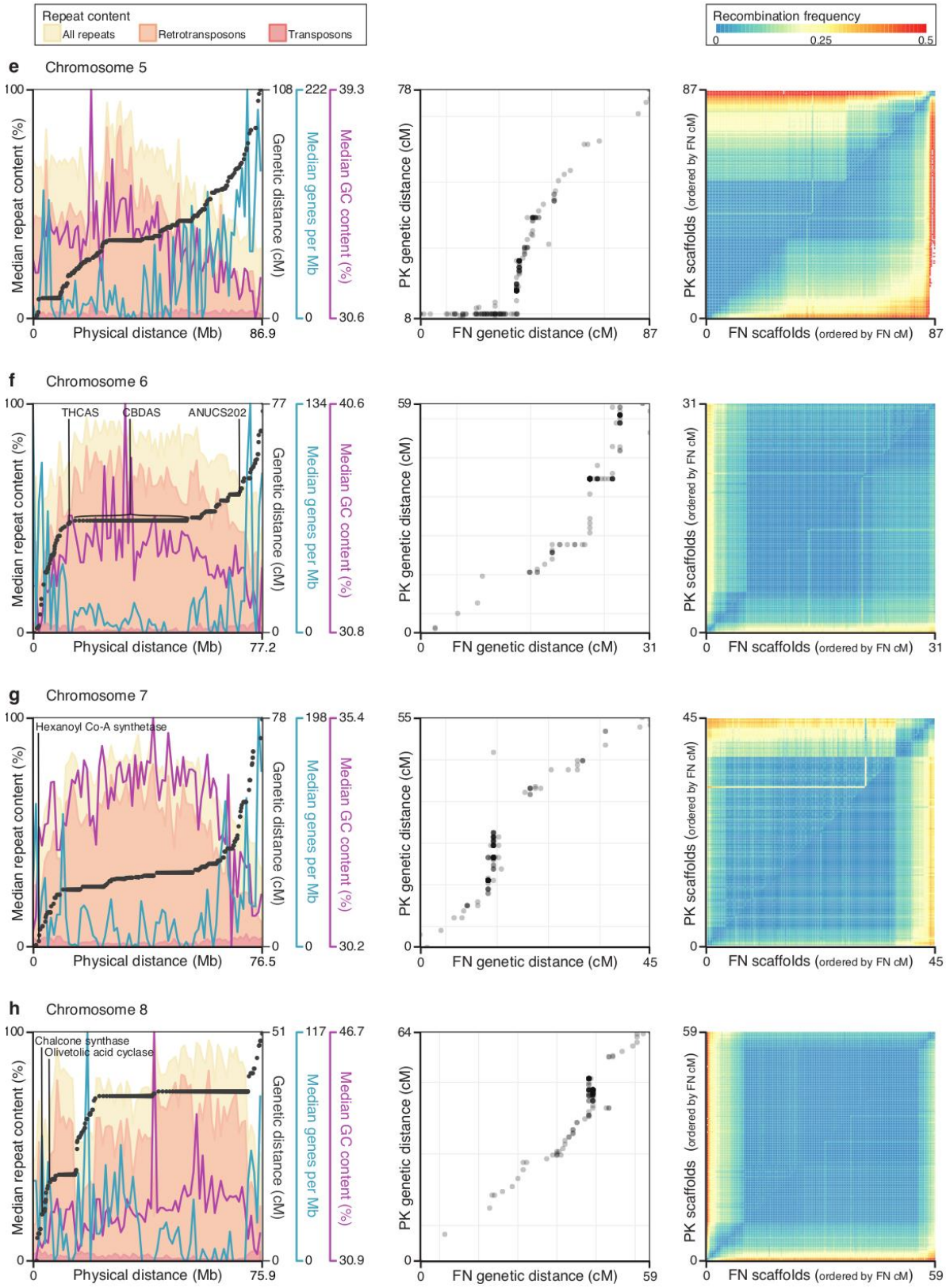
Supplemental Figure 3. Dot-plot comparing the PK and FN assemblies.
For each alignment between PK and FN with >95% identity and a length >2,500bp, a dot was placed at 1,000bp increments along the alignment. Grey dots indicate direct alignments and purple dots correspond to inverted alignments.



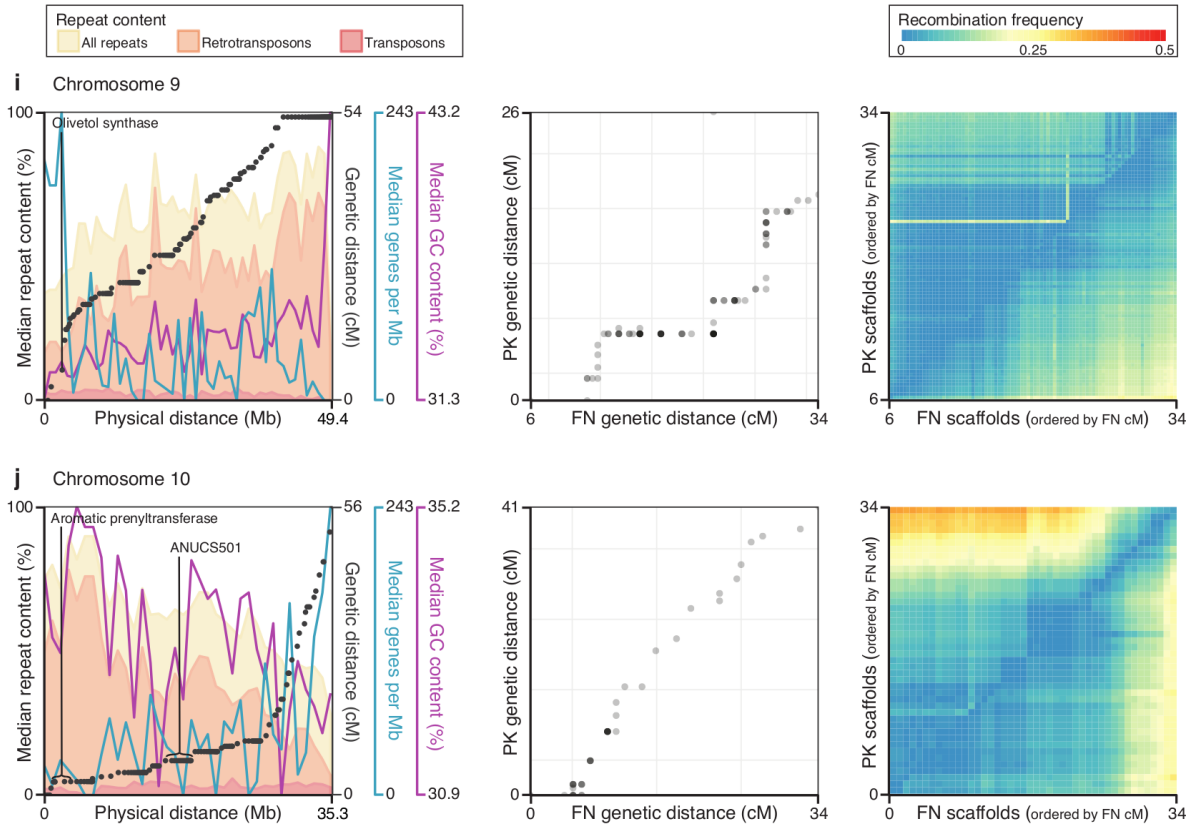
Supplemental Figure 4. BUSCO assessment of genome assembly completeness. Estimate of assembly completeness based on the presence of eudicotyledons universal single-copy orthologs selected from OrthoDB release 10. Gene presence was estimated using BUSCO v3.



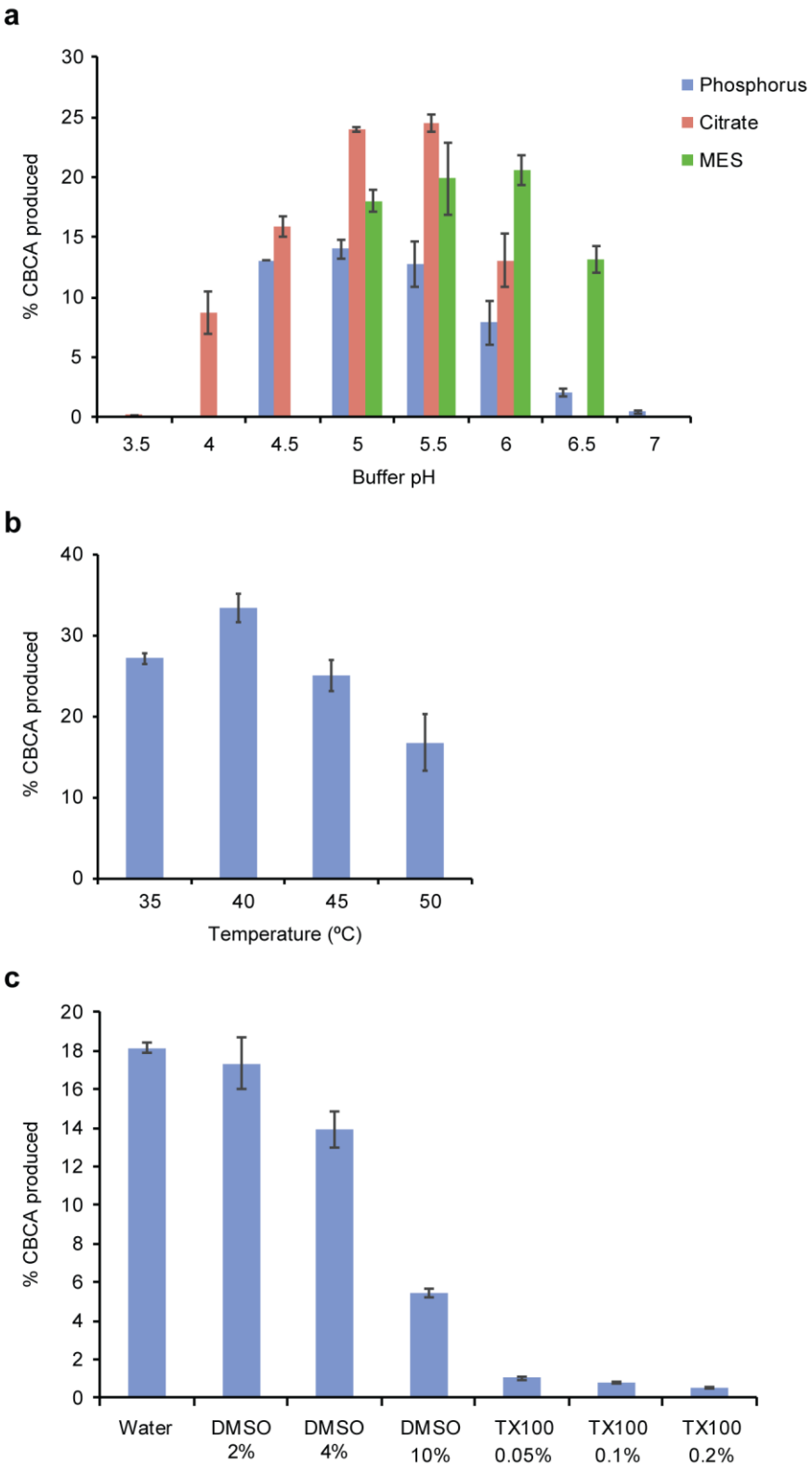
Supplemental Figure 5. Comparison of the genetic maps for PK and FN.
(Continued on next page)



Supplemental Figure 5. Comparison of the genetic maps for PK and FN.
(Continued on next page)



Supplemental Figure 5. Comparison of the genetic maps for PK and FN.
Three panels are shown for each chromosome. *Left*, a map as described in Figure 1, for the merged genetic map. *Middle*, scatter plot showing genetic positions (in cM) as determined by MSTmap for the same scaffolds in FN (horizontal axis) and PK (vertical axis). Regions that recombine in one parent and not the other may contain inversions or other rearrangements that differ between the parents. *Right*, heatmap showing recombination frequencies for FN (lower right) and PK (upper left), with scaffold order determined by the order of scaffolds in FN. Within regions of no recombination, scaffolds are ordered by size..



Supplemental Figure 6. Optimization of *CBCAS* reaction conditions.
Values are expressed as a fraction of CBCA produced (peak area of CBCA A265 nm / (peak area of CBGA A265 nm + peak area of CBCA A265 nm) * 100). **(a)** Effect of buffer and buffer pH on CBCA production. **(b)** Effect of incubation temperature on CBCA production. Reactions were performed in pH 5.5 citrate buffer. **(c)** Effect of additives on CBCA production. Reactions were performed in pH 5.5 citrate buffer at 35 °C.

Marker	GenBank Accession	Locations in FN FALCON assembly	Identity
MADC2	JF298280.1	000213F:489286-489570	100.00%
MADC3	AB021658.1	No alignment with > 98% ID	-
MADC4	AB021659.1	001104F:32868-33441	98.26
MADC5	AF364954.1	002457F:85456-86412	98.86%
MADC6	AF364955.1	004222F:33414-33562 004078F:34784-34932 003160F:27052-27200 003035F:54853-55001 001485F:27128-27276 004013F:17944-18092	100.00% 100.00% 100.00% 100.00% 100.00% 99.33%

Supplemental Table 1. Locations of male-specific RAPD markers in the FN assembly. Five Male-Associated DNA from Cannabis sativa (MADC) markers were aligned to the FN assembly using BLASTN. Locations and identities are shown for all alignments with > 98% identity. None of the scaffolds that contained a high identity alignment were placed in the anchored FN map.