Supplementary Information for

The recombination landscape in *Arabidopsis thaliana* F_2 populations

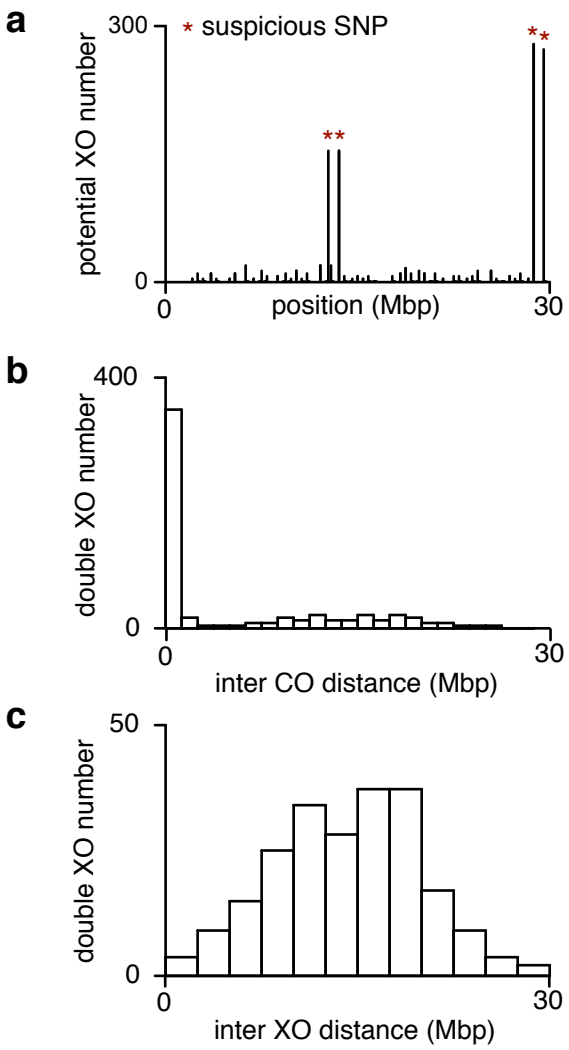
Patrice A. Salomé, Kirsten Bomblies, Joffrey Fitz, Roosa A. E. Laitinen, Norman Warthmann, Levi Yant, and Detlef Weigel

11 supplementary figures:

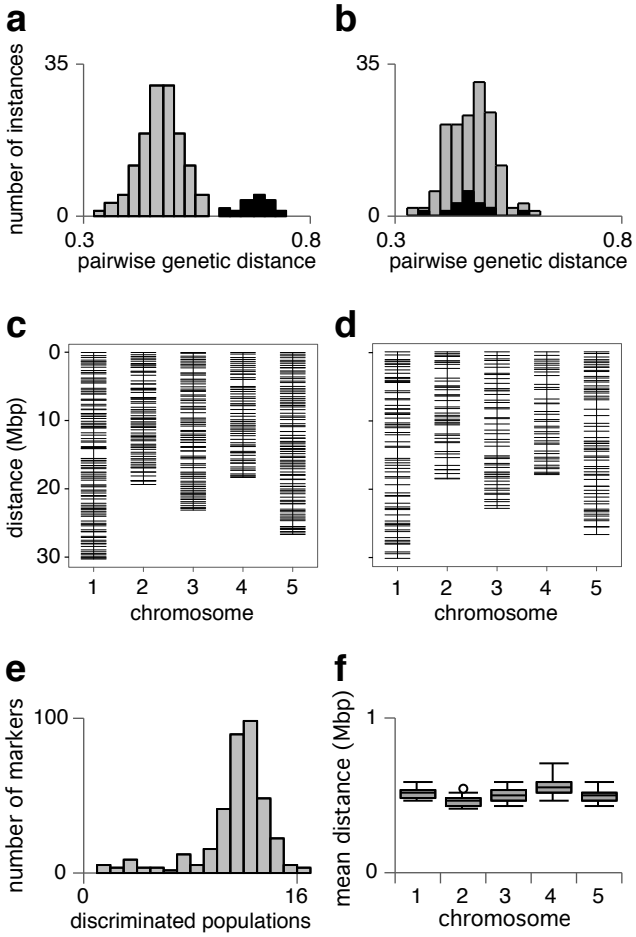
Supplementary Figure 1. Genotype error corrections.
Supplementary Figure 2. Generation of physical maps for multiple populations with a single SNP set.
Supplementary Figure 3. Lack of phasing information and selection of double XOs.
Supplementary Figure 4. The frequency of multiple XO events is correlated with physical chromosome length.
Supplementary Figure 5. Higher recombination rates adjacent to centromeres.
Supplementary Figure 6. Crossover interference for chromosomes 3, 4 and 5.
Supplementary Figure 7. Individual crossovers from a double crossover pair are not randomly distributed.
Supplementary Figure 8. Comparisons of genetic maps of the P2 population estimated by the Haldane, Kosambi and Carter-Falconer map functions.
Supplementary Figure 9. Correlation of genetic map lengths and recombination rates across F_2 and RIL populations: law of large numbers.
Supplementary Figure 10. Recombination rates are locally, but not globally correlated with sequence differences between parental accessions.
Supplementary Figure 11. Comparison of XO frequencies in F_2, RIL and AIL populations.

4 supplementary tables:

Table S1. Summary of informative SNP numbers, per population and per chromosome.
Table S2. Mean XO numbers.
Table S3. R^2 values for linear regressions between SNP frequency and recombination rates adjacent to the centromeres. Data were pooled in 1 Mbp windows, with a 200 kbp slide.
Table S4. R^2 values for linear regressions between SNP frequency and recombination rates away from the centromeres. Data were pooled in 1 Mbp windows, with a 200 kbp slide.

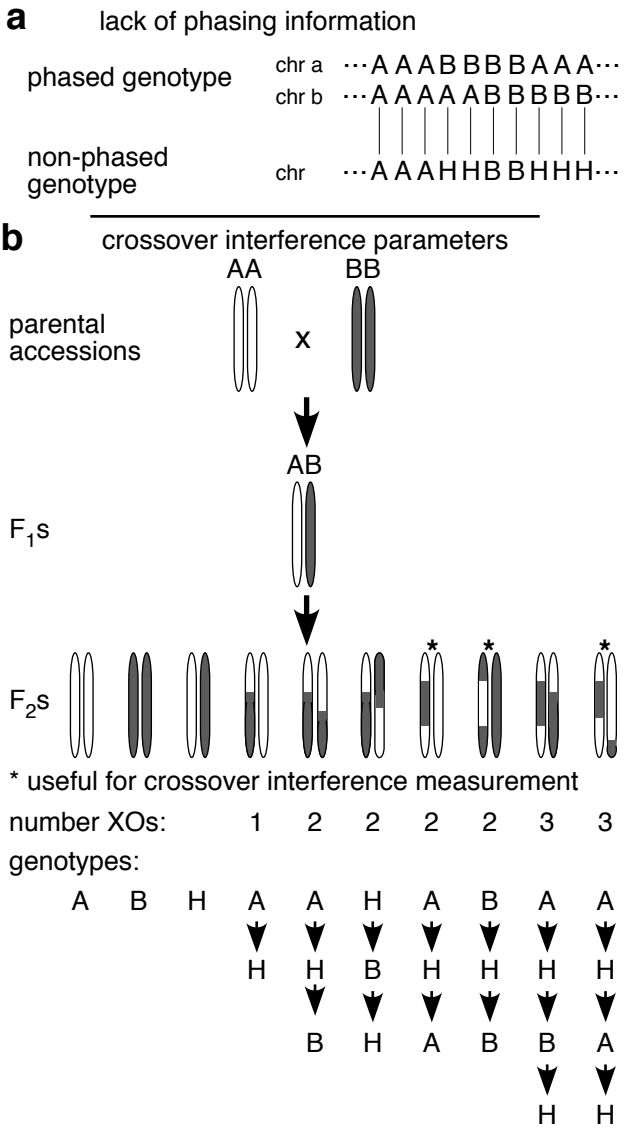


Supplementary Figure 1. Genotype error corrections.
a. Example of systematic marker failure along chromosome 1 in the P3 population, indicated by the red asterisks.
Distribution of inter-XO distances of apparent double crossovers occurring on the same chromosome on raw (**b**) and cleaned data (**c**).



Supplementary Figure 2. Generation of physical maps for multiple populations with a single SNP set.

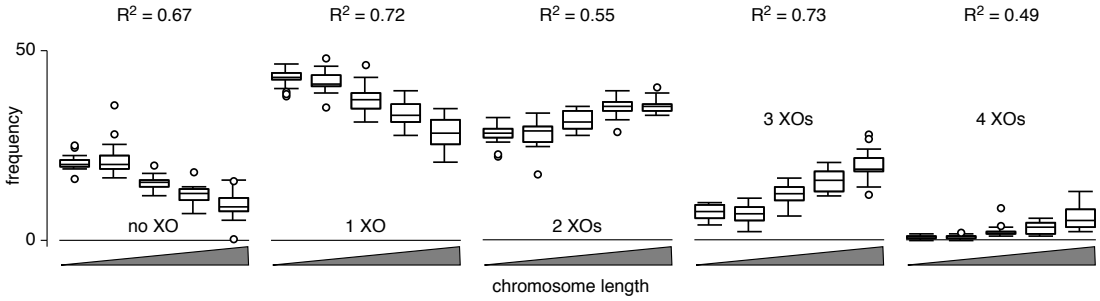
a. Pairwise differences between the 18 parental accessions with a set of 402 optimized SNPs.
b. Pairwise differences with a set of 149 non-optimized, intermediate-frequency SNPs (Warthmann *et al*, 2007). In **a** and **b**, Pairwise differences for parents of the 17 F_2 populations are in black.
c. Physical location of the 402 SNPs designed for this study on the reference genome.
d. Distribution of 257 informative SNPs in the P3 population.
e. Number of populations informed by each of the 370 SNPs retained after genotyping. The bests SNPs can discriminate between 16 populations.
f. Mean inter-SNP distances. Each dot corresponds to the mean for one population, and the horizontal line marks the mean across all populations.



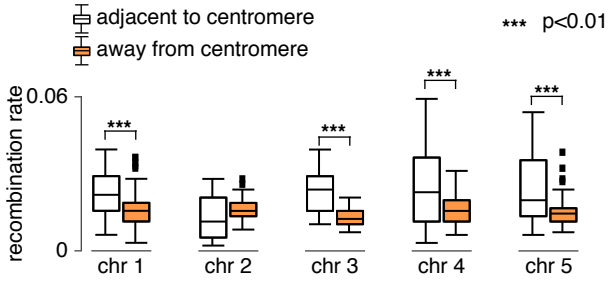
Supplementary Figure 3. Lack of phasing information and selection of double XOs.

a.The genotype information of each chromosome in a chromosome pair is known in phased genotypes. However, the genotypes of our F_2 plants are not, since both chromosomes of a given chromosome pair were genotyped at once.

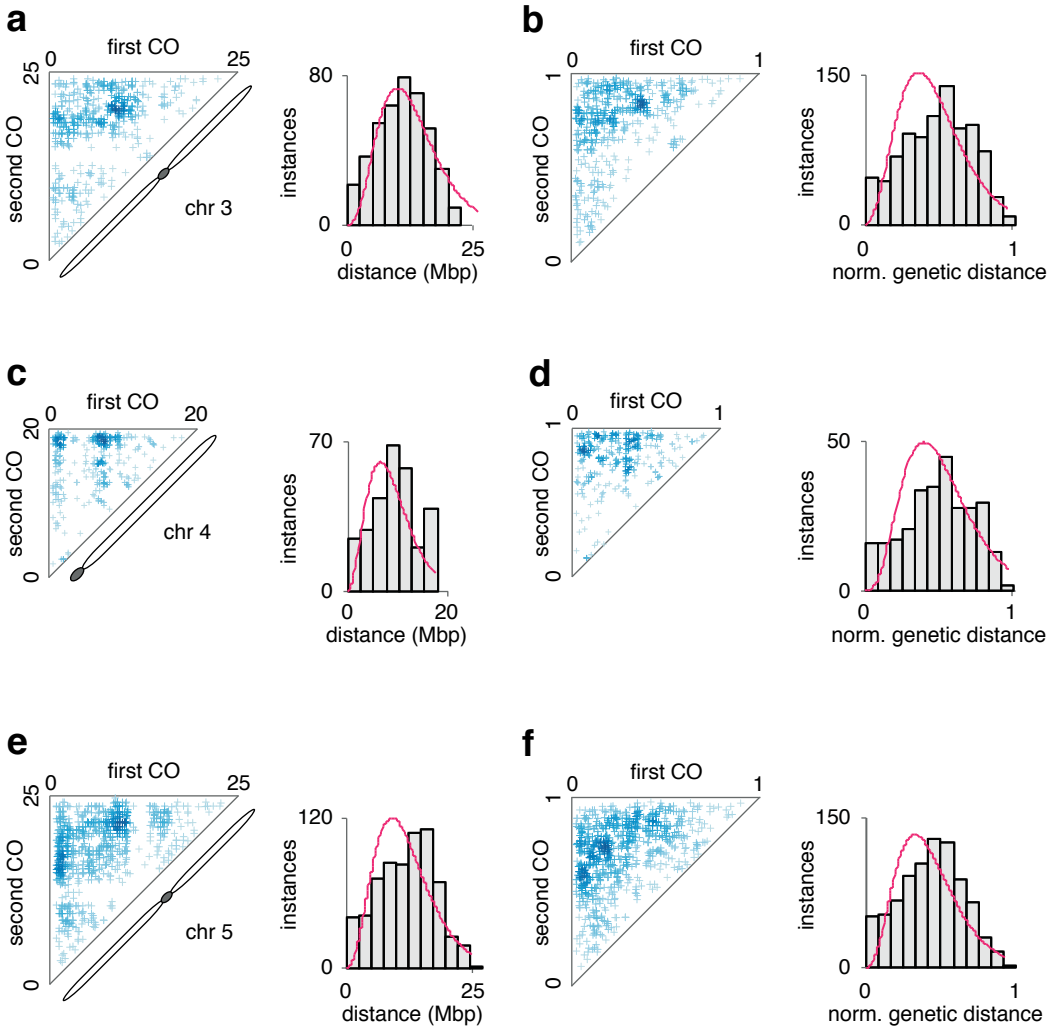
b. Selection of useful double XOs for measure of XO interference. Only a subset of chromosome pairs carrying 2 or 3 XOs were selected; the selected pairs carry 2 XOs formed during the same meiosis on a single chromosome, as opposed to two independent XOs each formed during the meiotic division of the female and male gametes. Not drawn are the remaining informative chromosome pairs whose genotypes read B➔H➔B➔H, H➔A➔H➔A and H➔B➔H➔B.



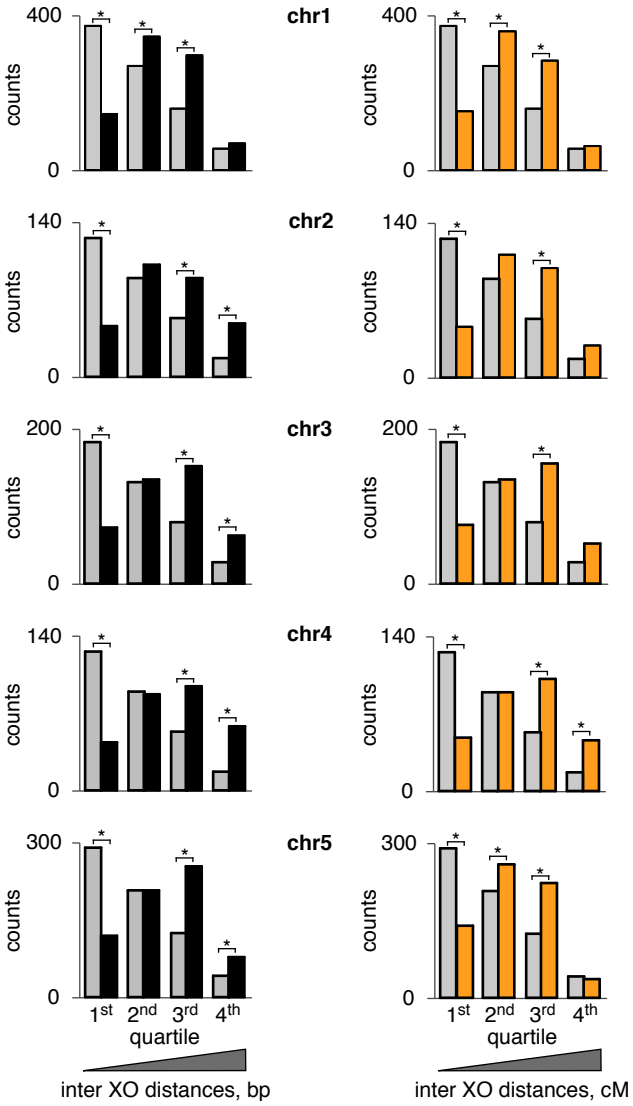
Supplementary Figure 4. The frequency of multiple XO events is correlated with physical chromosome length.
Box and whiskers plot of the frequency of crossover event along each chromosome, for XO number between 0-4. The correlation coefficient between chromosome length and frequency of crossover event is reported above each graph; for all, the associated p-values (after Bonferroni correction for multiple testing) are below 0.0001.



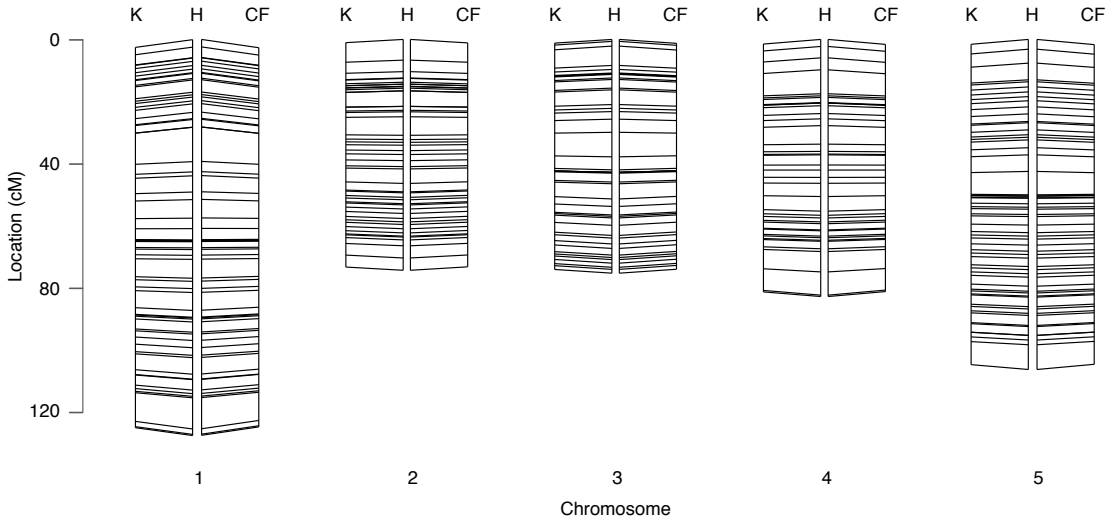
Supplementary Figure 5. Higher recombination rates adjacent to centromeres.
Recombination rates near centromeric regions were compared to the recombination rates away from centromeres, for mean recombination rates across all 17 F_2 populations. Genomic regions that are adjacent to or away from the centromeres are as shown in Figure 4.
p-values are for Student's t-test with Bonferroni correction for multiple testing.



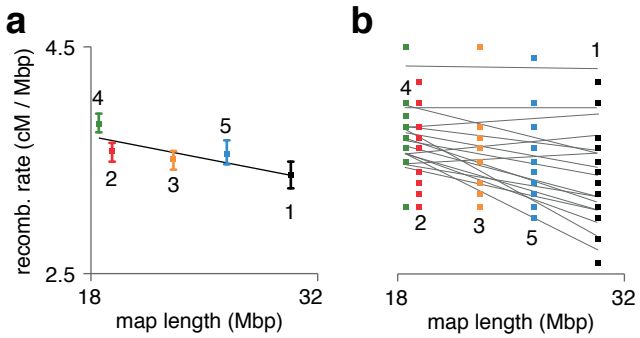
Supplementary Figure 6. Crossover interference for chromosomes 3, 4 and 5.
Positions of first and second crossovers for all double crossover pairs, according to their physical (**a, c, e**) or genetic (**b, d, f**) positions along the chromosome.
Left panels show the density of double XOs, while the right panels report the distribution of inter-crossover distances separating the two crossovers of a double XO pair. Magenta line: gamma distribution of scale = $(sd^2)/mean$, and shape = $(mean/sd)^2$.



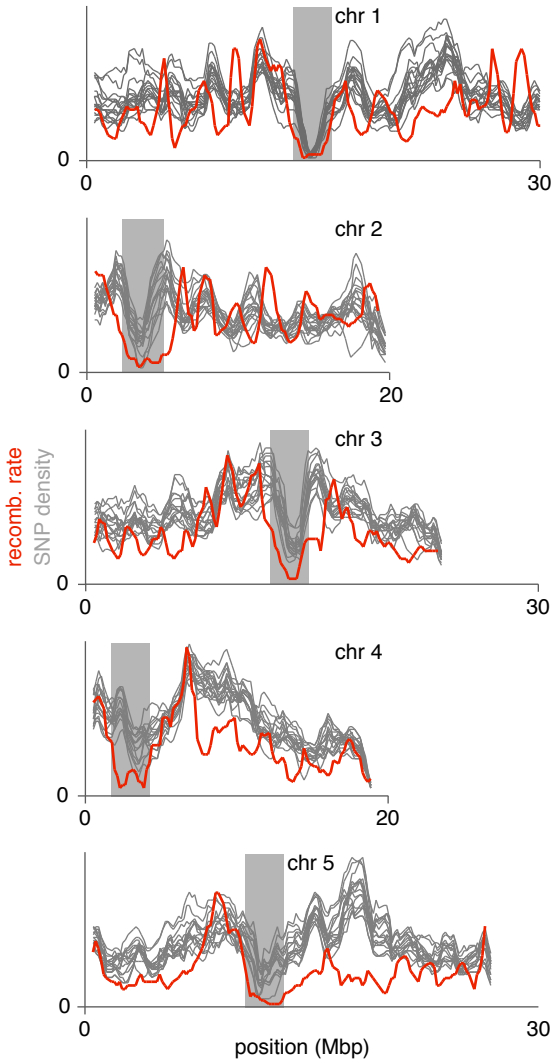
Supplementary Figure 7. Individual crossovers from a double crossover pair are not randomly distributed.
Observed inter-XO distances separating the two XOs from a double crossover pair are plotted as quartiles against the expected distribution of inter-XO distances. Left panels: distributions based on lengths of inter-XOs in bp; right panels" distributions based on lengths of inter-XOs in cM. Asteriks indicate when the observed distribution is significantly different from expectations ($p<0.001$, Pearson's Chi-squared test with Yates' continuity correction, and Bonferroni correction for multiple testing).
Observed inter-XO distances: black bars (left panel) or orange bars (right panel).
Expected inter-XO distances: grey bars. The expected distribution of inter-XO distances was determined according to (Drouaud *et al*, 2007).



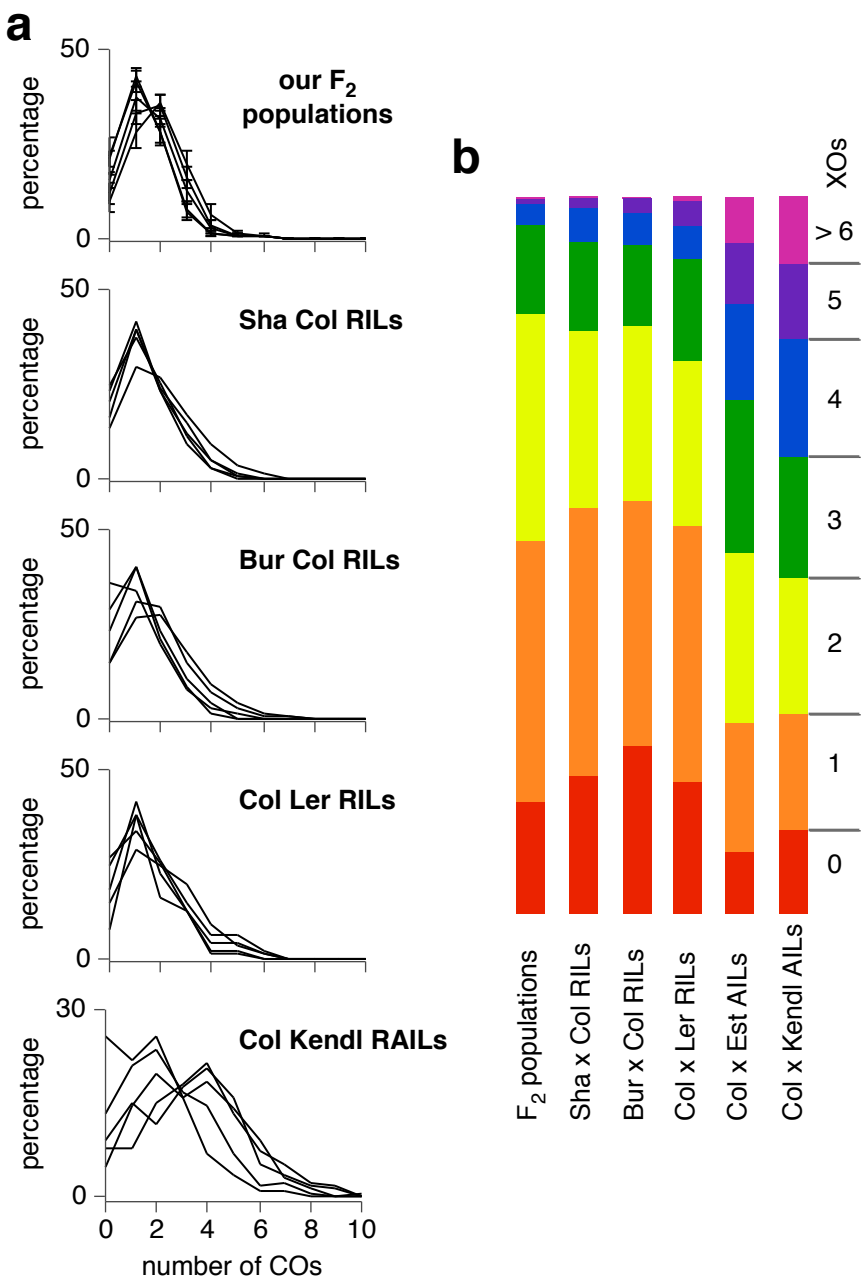
Supplementary Figure 8. Comparisons of genetic maps of the P2 population estimated by the Haldane (H), Kosambi (K) and Carter-Falconer (CF) map functions.
Genetic maps were calculated in R/qtl with the *est.map* function



Supplementary Figure 9. Correlation of genetic map lengths and recombination rates across F_2 and RIL populations: law of large numbers.
a. Recombination rates (in kbp/cM) along individual chromosomes for each F_2 population were plotted as a function of genetic map length (in cM). A multiple linear regression analysis demonstrated that the linear fits or chromosomes 2 and 4 (the two acrocentric chromosomes) share the same slope. Linear fits for chromosomes 1, 3 and 5 themselves share the same slope, but this slope is slightly different from that for chromosomes 2 and 4. R^2 value for multiple linear regression is 0.975, p-value < 2.2 e-16. Right panel: mean recombination rate across all 17 F_2 populations for each chromosome.
b. Recombination rates as a function of genetic map length for 12 RIL populations. R^2 value for multiple linear regression is 0.972, p-value = 2.9 e-16. Right panel: mean recombination rate across all 12 RIL populations for each chromosome. Inset: mean genetic map length, in cM, for F_2 populations plotted against mean genetic map length for RIL populations.



Supplementary Figure 10. Recombination rates are locally, but not globally correlated with sequence differences between parental accessions.
SNP density for all 17 F_2 populations was compared to each population's recombination rate (1 Mbp windows with a slide of 200 kbp). Mean recombination rate across all 17 F_2 populations is replotted from Figure 3 as a red line.
The position of the centromere is indicated by the grey area.



Supplementary Figure 11. Comparison of XO frequencies in F_2, RIL and AIL populations.
The Sha x Col-0 and Bur-0 x Col-0 RIL data are from SIMON *et al.* (2008), Col x Ler RIL data from LISTER *et al.* (1993), and Col x Kend AIL data from BALASUBRAMANIAN *et al.* (2009).
a. XO distribution per chromosome.
b. Frequencies of mean XO number (from no XO to >6).

TABLES

Table S1. Summary of informative SNP numbers, per population and per chromosome.

Pop.	genotyped plants	markers	chr 1	chr 2	chr 3	chr 4	chr 5	AA	AB	BB
P2	443	254	63	47	47	39	58	24.8%	50.9%	24.3%
P3	395	257	66	43	47	39	62	23.6%	51.6%	24.7%
P6	455	219	54	41	44	32	48	24.6%	48.4%	27.0%
P7	418	215	56	42	40	31	46	23.6%	50.5%	25.9%
P8	427	250	64	42	51	35	58	24.7%	49.2%	26.1%
P9	458	245	67	48	40	34	56	25.0%	47.7%	27.3%
P10	410	236	65	39	44	36	52	26.4%	50.3%	23.3%
P12	453	242	62	39	47	36	58	23.7%	51.0%	25.3%
P15	438	233	62	37	42	38	54	26.2%	50.6%	23.1%
P17	377	242	61	43	54	32	52	19.9%	53.1%	27.0%
P19	386	245	59	48	50	30	58	25.6%	50.7%	23.8%
P20	346	224	51	35	44	33	61	25.0%	48.9%	26.1%
P35	443	236	59	43	44	33	57	23.8%	49.8%	26.3%
P66	446	227	61	41	48	27	50	24.7%	49.0%	26.2%
P129	462	232	57	46	48	32	49	23.3%	50.9%	25.8%
P145	239	238	58	41	52	35	52	23.8%	50.2%	26.1%
P169	449	240	56	46	51	35	52	23.9%	51.4%	24.7%

Table S2. Mean XO numbers.

Population	chr 1	chr 2	chr 3	chr 4	chr 5	Whole genome
P2	2.1	1.2	1.2	1.2	1.7	7.4
P3	1.7	0.9	1.4	1.2	1.8	7.1
P6	1.6	1.0	1.3	1.2	1.4	6.5
P7	1.8	1.2	1.3	1.1	1.5	7.0
P8	1.7	1.1	1.3	1.2	1.3	6.6
P9	1.5	1.1	1.3	1.1	1.5	6.5
P10	1.6	1.0	1.3	1.0	1.4	6.4
P12	1.5	1.1	1.3	1.1	1.4	6.4
P15	1.7	1.0	1.3	1.2	1.6	6.7
P17	1.9	1.3	1.5	1.1	1.6	7.3
P19	2.0	1.3	1.5	1.2	1.7	7.7
P20	2.0	1.2	1.6	1.3	1.9	8.0
P35	1.6	1.1	1.3	1.0	1.4	6.4
P66	1.8	1.1	1.3	1.2	1.7	7.2
P129	1.3	1.0	1.2	1.0	1.3	5.8
P145	1.8	1.1	1.5	1.1	1.5	7.0
P169	1.6	1.1	1.2	1.2	1.4	6.5
mean	1.7	1.1	1.3	1.1	1.5	6.9

Table S3. R^2 values for linear regressions between SNP frequency and recombination rates adjacent to the centromeres. Data were pooled in 1 Mbp windows, with a 200 kbp slide.

population	chr1	chr2	chr3	chr4	chr5
P2	-0.03	0.04	0.15	0.10	0.007
P3	0.14	0.17	0.10	0.37	0.02
P6	0.30	-0.016	0.10	0.65	0.20
P7	0.038	-0.036	-0.013	-0.016	-0.02
P8	0.057	0.17	0.038	0.043	-0.02
P9	0.32	-0.035	-0.024	-0.028	-0.004
P10	0.15	0.028	-0.002	-0.043	-0.02
P12	0.12	-0.036	-0.017	-0.001	0.15
P15	0.03	0.033	0.11	0.19	0.002
P17	0.035	-0.032	0.032	-0.01	-0.02
P19	-0.001	-0.023	-0.014	-0.012	0.01
P20	0.07	-0.006	-0.026	-0.032	0.02
P35	0.009	-0.03	0.010	0.08	-0.026
P66	-0.022	0.03	-0.006	0.17	0.010
P129	0.03	-0.035	-0.026	-0.04	-0.021
P145	-0.015	0.20	-0.004	0.11	0.04
P169	0.30	0.036	0.018	0.17	0.003
mean	0.11	-0.036	-0.008	0.037	-0.015

p-values levels of significance:

0.001
0.01
0.05

Table S4. R^2 values for linear regressions between SNP frequency and recombination rates away from the centromeres. Data were pooled in 1 Mbp windows, with a 200 kbp slide.

population	chr1	chr2	chr3	chr4	chr5
P2	0.10	0.040	-0.017	0.003	-0.01
P3	-0.01	0.025	0.21	0.31	-0.001
P6	-0.01	0.015	0.032	0.005	-0.006
P7	0.003	-0.019	-0.015	0.35	0.003
P8	-0.001	0.153	0.040	0.28	0.20
P9	-0.010	0.033	-0.016	0.09	-0.007
P10	0.016	-0.021	-0.016	-0.02	0.006
P12	0.015	-0.011	0.0001	0.19	0.11
P15	-0.01	0.14	0.13	0.007	-0.01
P17	0.006	-0.01	-0.006	0.30	0.36
P19	0.034	-0.01	0.013	0.24	-0.01
P20	-0.011	0.10	-0.01	0.03	-0.01
P35	0.019	0.008	0.021	0.08	0.056
P66	-0.01	0.11	-0.007	0.36	-0.012
P129	-0.01	0.09	0.042	0.16	0.015
P145	0.001	0.08	-0.017	0.18	0.026
P169	-0.01	0.12	0.065	-0.01	-0.012
mean	-0.004	0.08	-0.013	0.47	-0.012

p-values levels of significance:

0.001
0.01
0.05
0.1

Supplementary references:

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN *et al* (2009). QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. *PLoS ONE* **4:** e4318.

Drouaud J, Mercier R, Chelysheva L, Berard A, Falque M, Martin O *et al* (2007). Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4. *PLoS Genet* **3:** e106.

Lister C, Dean C (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in Arabidopsis thaliana. *Plant J* **4:** 745-750.

Simon M, Loudet O, Durand S, Berard A, Brunel D, Sennesal FX *et al* (2008). Quantitative trait loci mapping in five new large recombinant inbred line populations of Arabidopsis thaliana genotyped with consensus single-nucleotide polymorphism markers. *Genetics* **178:** 2253-2264.