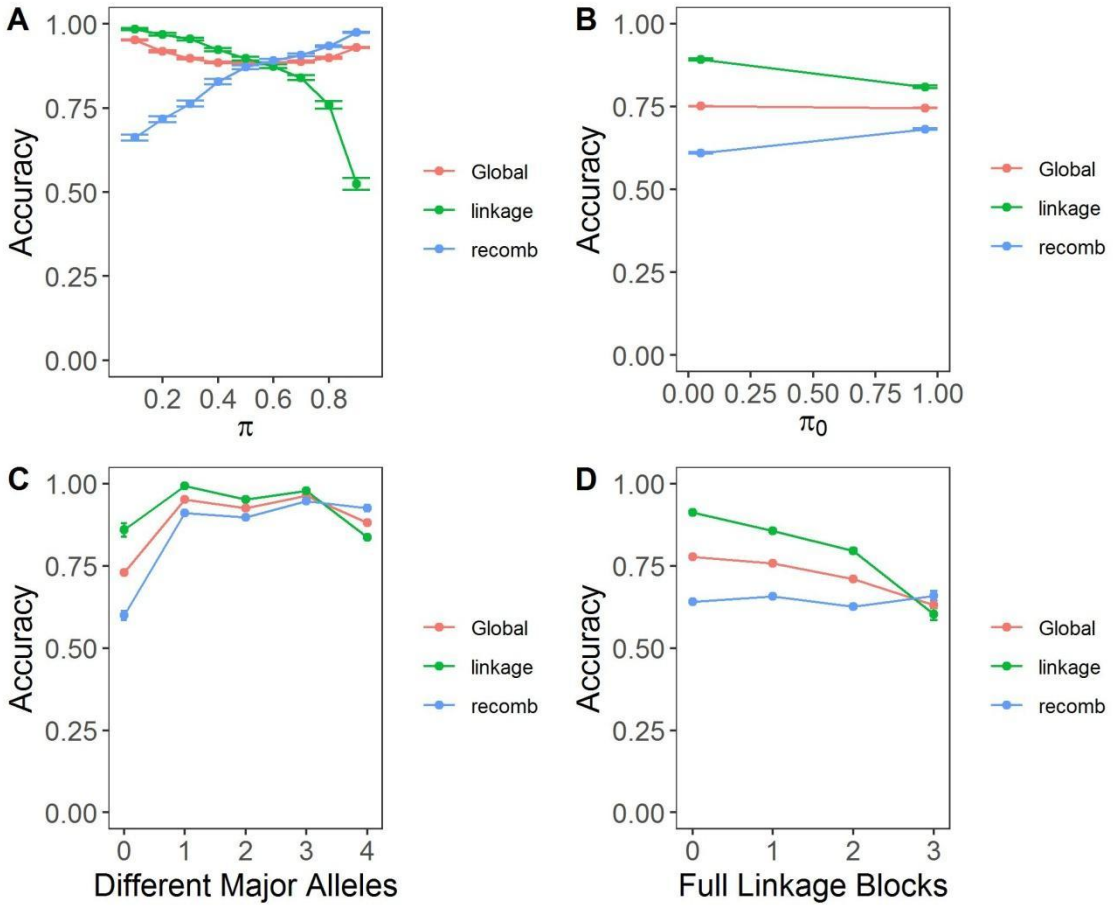
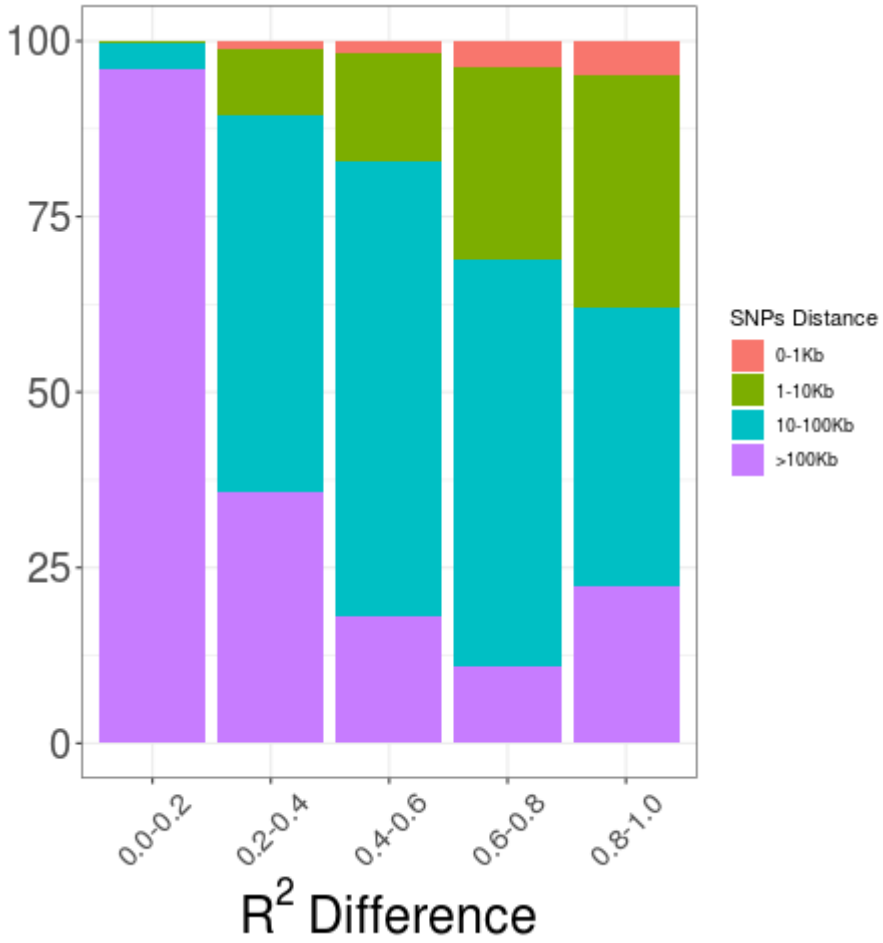
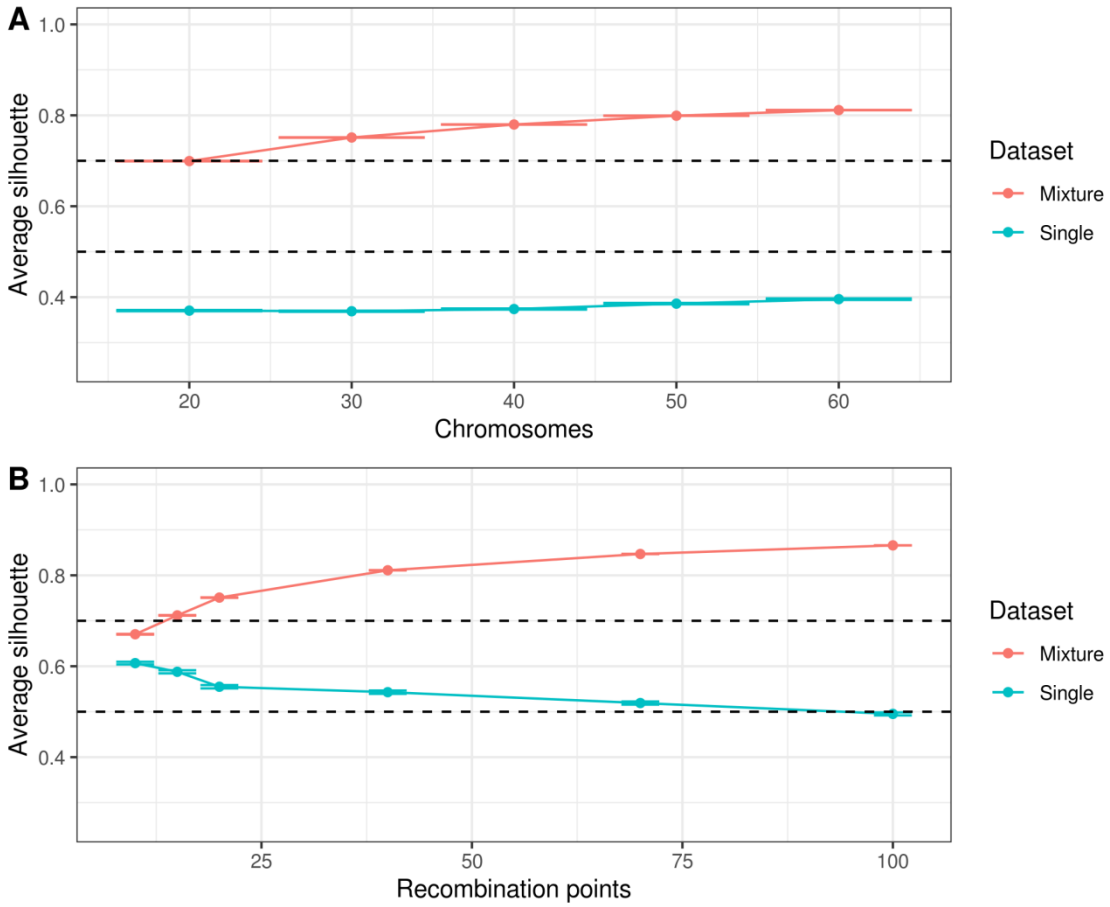


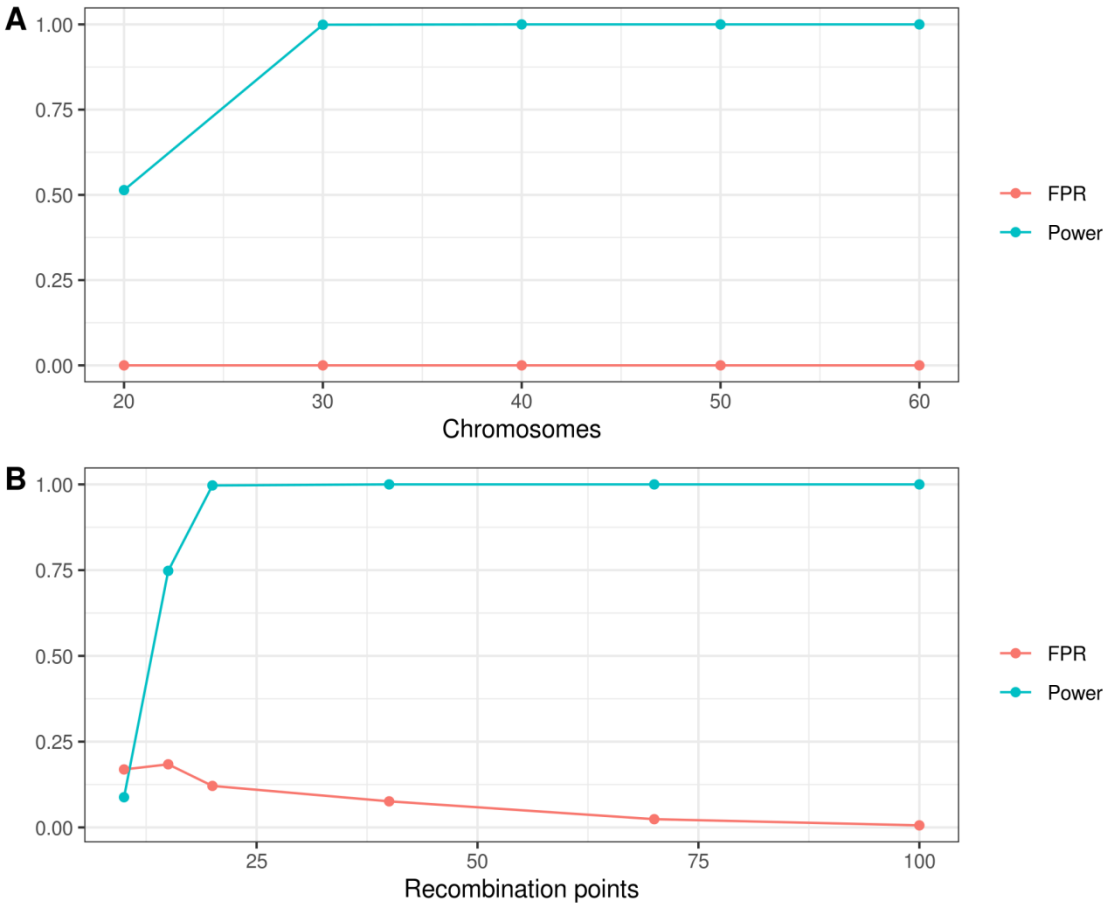
Sup Fig S1: Evaluation of the mixture model under different simulated scenarios. Each plot contains the accuracy for classifying all chromosomes and the accuracies to classify *recomb* or *linkage* chromosomes (i.e. sensibility). π: proportion of chromosomes belonging to *recomb* population. π_0: *recomb* proportion initial value. Different Major Alleles: SNPs having different major alleles in *recomb* and linkage populations. Full Linkage Blocks: Number of blocks without variability, i.e. they can be considered as a single SNP. Accuracies contain the mean and standard error computed from 200 simulations of each scenario.



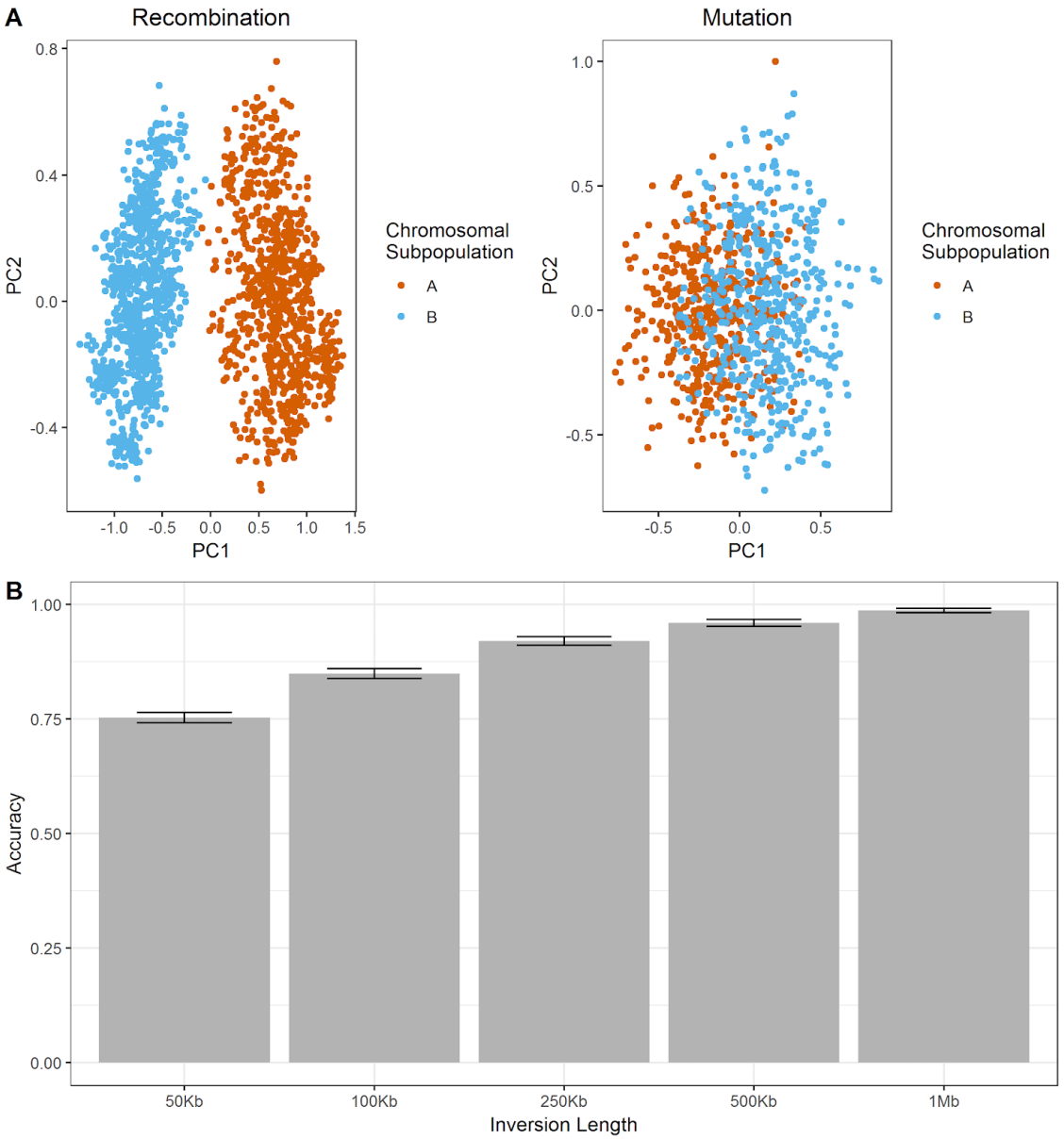
Sup Fig S2: Distribution of difference in linkage disequilibrium (R^2) per distance bins. These results are based on human inv8p23.1. We computed the R^2 for all SNP-pairs in the inversion region, independently for inverted and standard chromosomes. R^2 difference is the absolute difference between the R^2 in standard and inverted. SNP-pairs were grouped in 4 bins, according the distance between the SNPs (0-1Kb, 1-10Kb, 10-100Kb and more than 100Kb).



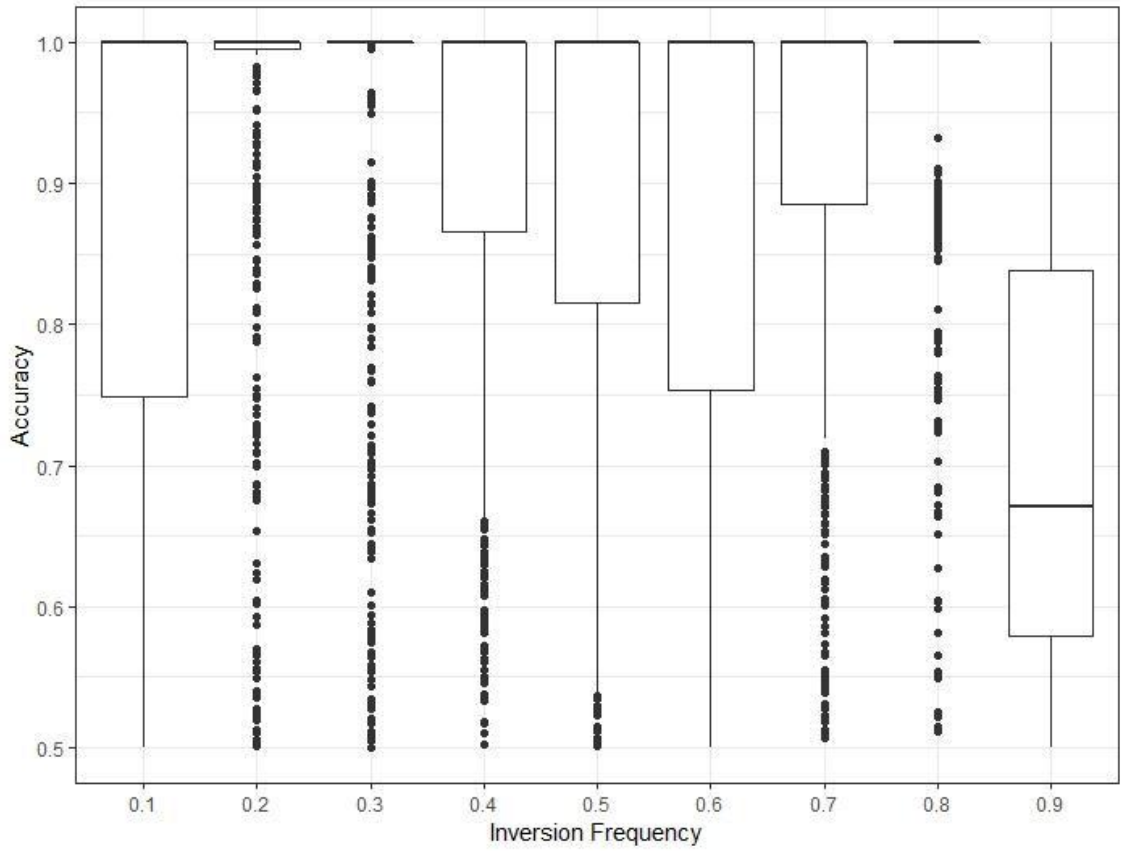
Sup Fig S3: Accuracy of *recombClust* to detect datasets with a population mixture. Average silhouette value indicates how reliable was *recombClust* clustering in a given dataset. Dashed lines mark critical cut-off (>0.7: very reliable structure; 0.5-0.7: reliable structure; <0.5: unreliable clustering). Mixture datasets contains two subpopulations with non-overlapping recombination points, while single datasets contain one population. Average silhouette values contain the mean and standard error computed from 1000 simulations of each scenario.



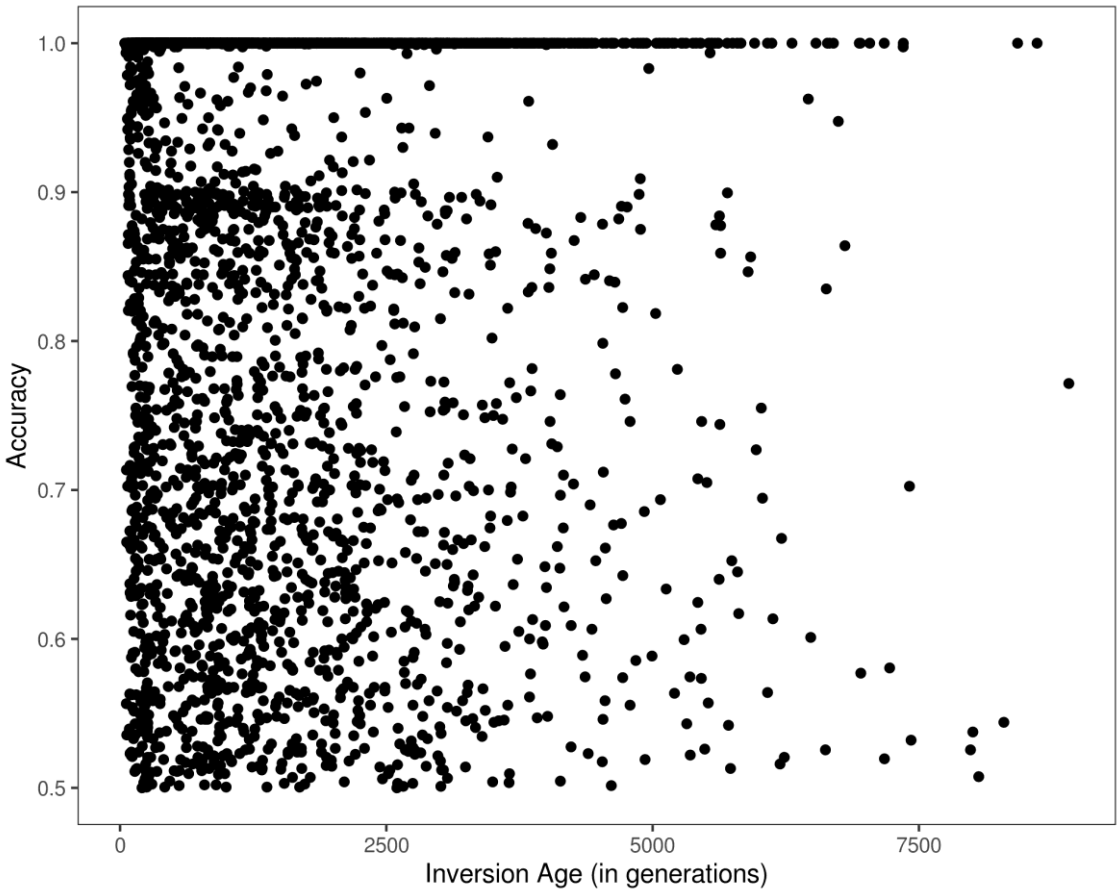
Sup Fig S4: *recombClust* evaluation in simulated datasets with a known population mixture. We computed FPR and power based on simulated datasets. Half of the datasets contained a mixture population and the other half one population. A dataset with an average silhouette value of 0.7 was considered as supporting a mixture population by *recombClust*. Each point is the FPR and power of 2000 simulations. FPR: False Positive Rate. Proportion of regions called as having two subpopulations, among all the regions with just one population.



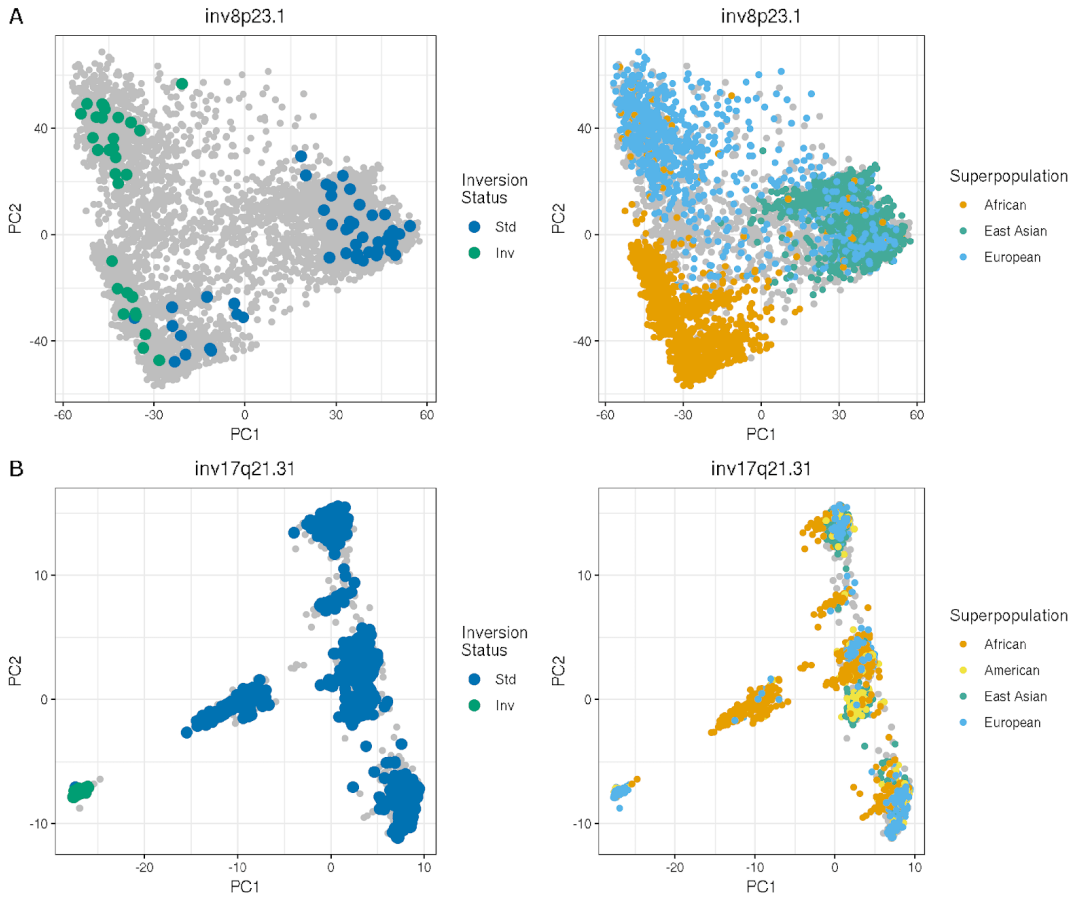
Sup Fig S5: *recombClust* accuracy for detecting subpopulations with different recombination patterns. A-left) Detection of recombination patterns on simulated data by clusters on the PCs of the prediction matrix across mixture models on 10 recombination points. Five different recombination points were simulated for each subpopulation A and B, where the other subpopulation remained in linkage. The first PC shows a clear separation of the subpopulations. A-right) To test mutation differences between the subpopulations, we computed the PC for the genotype matrix of the markers flanking the 10 recombination points. In this case the PCs did not showed a clear separation between the chromosome subpopulations. B) The figure shows the match between the chromosome subpopulations as obtained by *recombClust* and inversion status of chromosomes, for 9,000 simulated inversions at a given size (1000 simulations at 9 different inversion frequencies). The figure shows the mean accuracy and standard error. *recombClust* identifies inversion status by recombination differences with high accuracy, particularly for inversions > 250Kb.



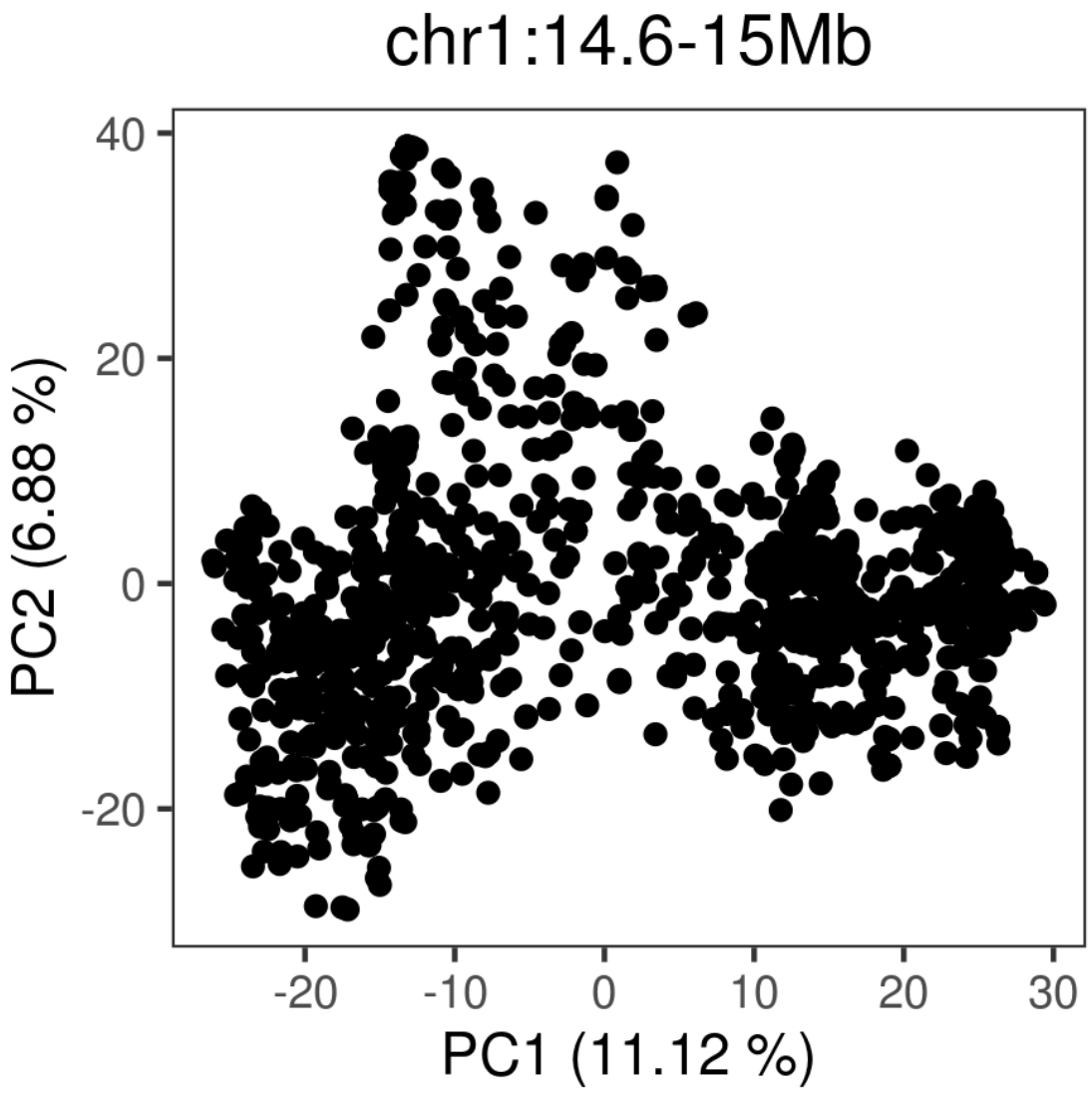
Sup Fig S6: *recombClust* accuracy for different inversion frequencies. Accuracy is the proportion of phased chromosomes correctly classified. Each boxplot includes 500 simulations.



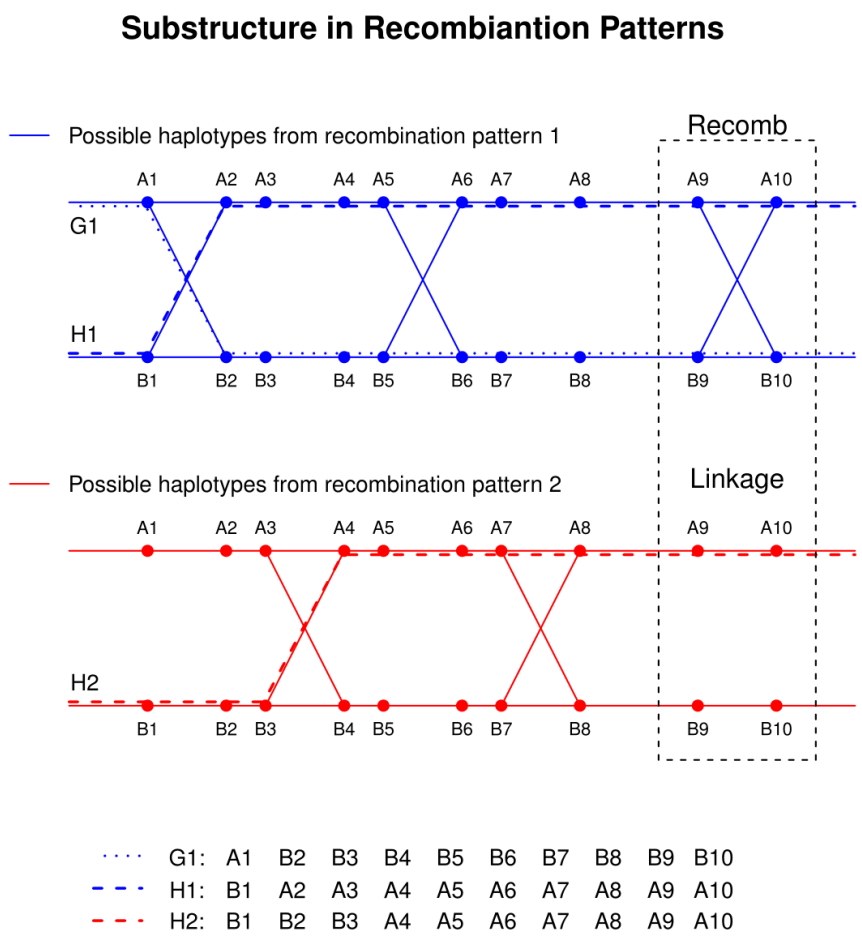
Sup Fig S7: *recombClust* accuracy for different inversion ages. Accuracy is the proportion of phased chromosomes correctly classified. Inversion age is measured in the number of generations since the inversion appeared and until the simulation stops. Each point is the accuracy of an independent simulation.



Sup Fig S8: Identification of chromosomal subpopulations of different ancestries from differences in the recombination patterns within two inversions. The figures show the first two PCA components for the all mixture model predictions at numerous recombination points across inv-8p23.1 and inv-17q21.31, computed for all 1000 Genomes ancestries. Chromosomes are clearly separated by inversion status (Std, Inv) and ancestry. For inv-8p23.1 clear ancestral groups are identified within inversion status whereas ancestry is mixed within each inv-17q21.31 status. Colored points indicate experimentally validated observations of inversion status and ancestry.



Sup Fig S9: *recombClust* clustering in a randomly selected region (chr1:14.6-15Mb). *recombClust* was run on a region with the same of length TAR region but where no structure is expected.



Sup Fig S10: Representation of two chromosomal subpopulations with different recombination patterns in a genomic segment. Lines represent the possible chromosomes present in population 1 (blue) and population 2 (red). Each SNP has two alleles (A and B) and is labelled with a number. Recombination points are placed between SNPs where A and B alleles are joined by a line. G1 and H1 are two possible chromosomes from population 1 and H2 is one of the possible chromosomes from population 2. The dotted box contains a recombination point present in population 1 but not in population 2.

Sup Table S1: Chromosomal inversions analyzed in this manuscript. Coordinates in human refer to GRCh37 build and in *Drosophila melanogaster* to dm6 build. Age in generations was inferred from age in years, assuming 30 years/generation in humans and 15 generations/year in *Drosophila melanogaster*.

	Inversion	Coordinates	Age in years	Age in generations	Selection
Human	inv17q21.31	Chr17:43.66-44.37 Mb	3,000,000[1]	100,000	Balancing Selection[1]
	inv8p23.1	Chr8:8.06-11.98 Mb	370,000[2]	12,000	
Drosophila melanogaster	In(2L)t	2L:2.23-13.15 Mb	69,398[3]	1,040,970	Balancing Selection[4]
	In(2R)NS	2R:11.28-16.16 Mb	178,886[3]	2,683,290	Balancing Selection[4]
	In(3R)Mo	3R:17.23-24.86 Mb	2,861[3]	42,915	Balancing Selection[4]

1. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).

2. Salm, M. P. A. *et al.* The origin, global distribution, and functional impact of the human 8p23 inversion polymorphism. *Genome Res.* **22**, 1144–1153 (2012).

3. Corbett-Detig, R. B. & Hartl, D. L. Population Genomics of Inversion Polymorphisms in Drosophila melanogaster. *PLoS Genet.* **8**, (2012).

4. Kapun, M., Fabian, D. K., Goudet, J. & Flatt, T. Genomic Evidence for Adaptive Inversion Clines in *Drosophila melanogaster*. *Mol. Biol. Evol.* **33**, 1317–1336 (2016).

Sup Table S2: *recombClust* allele frequencies in LCT locus for different European populations. Allele 1 is more frequent in all populations by TSI, the only population that do not show a selection mark based on iHS. CEU: Utah residents (CEPH) with Northern and Western European ancestry. FIN: Finnish in Finland. GBR: British in England and Scotland. IBS: Iberian populations in Spain. TSI: Toscani in Italy.

	Allele 1 Frequency	Allele 2 Frequency
CEU	79.8%	20.2%
FIN	65.7%	34.3%
GBR	74.2%	25.8%
IBS	59.8%	40.2%
TSI	28.5%	71.5%