Supplementary Material

Supplementary Text

1 - Making a linkage map

The choice of the correct LOD score for each dataset will depend on the family structure. The developer of LepMap3 suggests the use of a LOD score that gives a number and size of linkage groups (LGs) that match the underlying chromosomes. However, in studies with many markers and due to the dubious nature of some parts of the genome assembly the distribution might deviate from the truth. Here we will try to explain and justify our choice of the "correct" LOD score. In classical linkage analysis the usual LOD cutoff is 3 meaning a 1000 to 1 chance of the markers being linked compared to un-linked. If the LOD score chosen is less than 3 then markers that appear linked might in fact belong to different chromosomes and merging of chromosomes into the same LG will be observed. On the contrary if the cutoff is too high then only strongly linked markers will appear as unique LGs meaning that the user will observe over-splitting of LGs. We used an approach to identify the "correct" LOD score based on a few assumptions about the quality of our reference and disregarding a few misplaced markers. A similar approach was used in (Peñalba et al., 2020).

Specifically, we ran the *SeparateChromosomes* module for different LOD scores from 11 to 21. Then we plotted the distribution of markers per linkage group and we compared it to the actual distribution of markers in super-scaffolds (Supplementary Figure 8). While the two

should not always match we did not expect big errors in our assembly since it incorporates information from bionano optical mapping (Machado et al., 2022). We then identified a LOD of 15 as the balanced LOD score between overmerging and a distribution that matched our expectations. We compared that LOD score with the next two (16 and 17). Based on the previously explained idea 15 should exhibit the biggest amount of merging between the three. Thus we manually checked the LGs that contained more than one super-scaffold. We found most merges included one marker from multiple scaffolds and a large number of markers from another one. We assumed that the single markers added were erroneously placed and removed them from that LG. Only one split and two scaffold merges were supported by enough markers.

The split of Super-Scaffold_2 into linkage group 1 and linkage group 24 was further validated by looking into datasets with smaller LOD cutoffs (where overmerging is expected). Because the split was present in LOD 13 we considered it well supported. The scaffolding of the 2 parts of Super-Scaffold_2 was performed with bionano optical mapping with the addition of a long sequence of N nucleotides. Further evidence was provided with synteny with the chicken genome which mapped the two parts of the scaffold on two different parts of Chromosome 1 of the chicken.

One merge was that of Super-Scaffold_100000100064 and Super-Scaffold_200000178 in linkage group 41 but there were not enough markers to verify this.

The other merge was that of Super-Scaffold_3 and Super-Scaffold_49 in linkage group 20. This was verified by looking for this merge in datasets with a bigger LOD score cutoff (where over splitting is expected). The merge appeared up to LOD 18. Synteny analysis showed the two scaffolds mapping side to side on chromosome 12 of the chicken. However, Super-Scaffold_49 was too small to confidently order and was thus removed from downstream analyses.

Supplementary Tables

Table S1. Table of linkage groups in the barn owl assembly.
Length in millions of base pairs and excluding N nucleotides introduced through optical mapping.

Linkage group	Super Scaffold	Starting position	Ending position	Length	Male cM	Female cM	Sex averaged cM	SNPs
1	2	28692481	89706370	60.12	62.22	52.23	57.16	9059
2	14	53821	60417706	60.33	69.13	76.3	72.61	8725
3	6	852023	56510526	55.09	80.91	78.99	79.39	7035
4	40	13170	51299756	51.24	68.24	74.19	71.22	6947
5	9	1015025	43248324	41.47	43.28	44.76	44.05	6292
6	38	47028	40714749	40.47	37.47	43.52	38.61	6279
7	1	47975	41737220	41.69	64.83	67.3	65.74	6127
8	7	213364	44031946	43.82	59.47	73.54	66.39	5887
9	18	70968	35153132	35.08	56.56	77.14	66.81	5693
10	16	208090	34916011	34.71	57.37	64.36	60.81	5511
11	22	161015	41219165	40.69	52.47	50.96	51.75	5404
12	45	296200	42636880	42.34	59.78	71.1	65.59	5412
13	27	18086	33704426	33.69	29.79	28.25	28.99	4853
14	1000006	29863	35914917	35.83	76.44	90.39	83.42	4698
15	10	193481	29879402	29.69	47.17	53.76	50.43	4470
16	21	25830	23008376	22.98	58.87	55.48	56.97	4121
17	48	772976	28607598	27.26	57.18	62.66	59.93	3893
18	23	64769	29188293	28.81	27.33	25.75	25.33	3650
19	26	407561	23966318	23.48	61.92	56.25	59.04	3611
20	3	482286	20936318	20.45	53.78	48.32	51.08	3008
21	17	1222151	23153977	21.22	25.86	35.37	30.56	3308
22	8	39596	24961950	24.89	42.97	53.1	50.76	3044
23	5	35770	24493257	24.31	54.69	65.26	59.81	3025
24	2	726367	28161559	26.37	24.85	26.44	25.42	2962
25	28	288799	19568140	17.83	64.99	56.3	60.52	2955
26	46	12809	18684943	18.6	62.26	44.59	53.43	2937
27	41	236640	24595500	24.32	54.44	60.06	56.46	2745
28	33	220754	18108959	17.89	56	53.45	54.66	2731
29	29	156998	14817099	14.44	54.12	50.25	52.21	2729
30	11	45380	13271628	13.22	53.45	45.16	49.23	2481
31	44	71425	17099552	16.83	18.74	42.24	30.72	2356
32	39	4519	14478862	14.19	54.42	48.08	51.19	2266
33	19	126970	13308235	13.18	58.93	56.66	57.81	2198
34	20000042	263997	15911457	15.61	47.63	61.65	54.68	1987
35	20	2806805	11357357	6.65	47.7	44.7	46.19	1715

36	30	268792	7850256	7.49	45.95	51.91	48.91	1399
37	32	1513738	9418208	6.92	46.52	44.26	44.18	1161
38	12	2239	7405831	7.22	39.98	55.64	47.78	1037
39	34	883598	6637883	5.47	36.06	37.14	36.96	995

Table S2. Sample Metadata.
All samples were used for SNP calling but only some in each analyses. This is indicated in the dataset column with LM3 being the Linkage Map, and the rest used in pyrho as follows: CH - Switzerland full dataset, GB - Great Britain, PT - Portugal, CH13 - first subset of 13 individuals from Switzerland, CH13_2 - second subset of 13 individuals from Switzerland.

See separate file:
https://github.com/EluLava/PhD/blob/main/ANNEXES/ANNEX_I/Annex_I_SupTable.csv

Table S3. **Hyperparameters used in pyrho.**
Abbreviations: CH - Switzerland full dataset, GB - Great Britain, PT - Portugal, CH13 - first subset of 13 individuals from Switzerland, CH13_2 - second subset of 13 individuals from Switzerland.

Population	Window Size	Block penalty
CH	70	15
GB	40	15
PT	60	15
CH13	40	15
CH13_2	40	15

Supplementary Table 4. Number of SNPs that passed filters for each population and for each analysis.
Abbreviations: CH - Switzerland full dataset, GB - Great Britain, PT - Portugal, CH13 - first subset of 13 individuals from Switzerland, CH13_2 - second subset of 13 individuals from Switzerland. Mac - minor allele count, π- nucleotide diversity

Population	**Sample size**	**SMC++** 10% missingness	**Pyrho** 10% missingness, 10bp distance, 2 mac	**π** 50% missingness, 5 mac
CH	76	19’454’720	9’388’296	8’512’324
GB	13	17’348’589	6’441’109	3’211’420
PT	13	19’770’073	4’897’345	3’325’626

CH13	13	19’373’680	5’784’140	3’377’113
CH13_2	13	19’239’643	5’788’185	3’385’148

Supplementary Figures

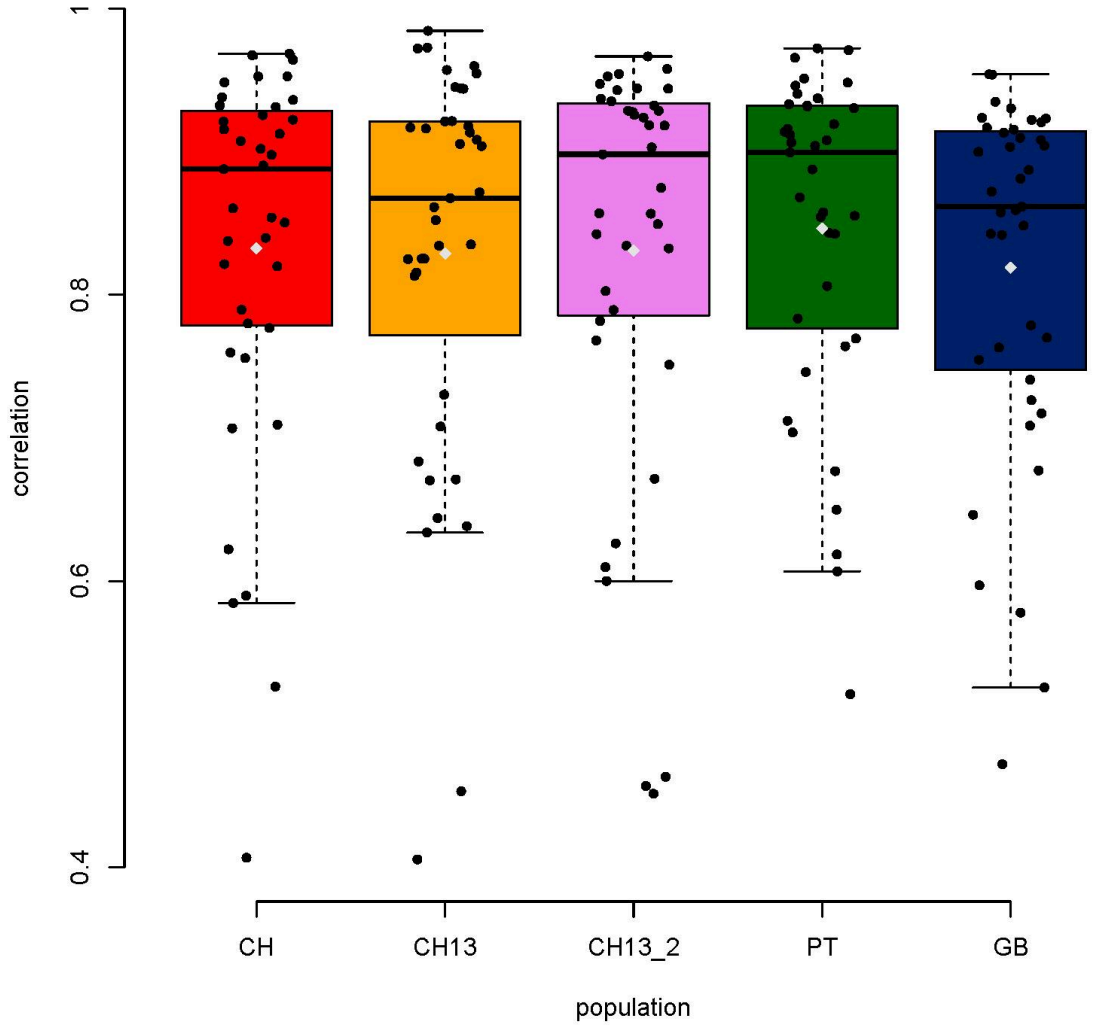

Figure S1 - Correlation between linkage map length and all population recombination rates as estimated from pyrho in 1Mb windows.
Each black point is one linkage group. Grey diamonds signify the average. Abbreviations: CH - Switzerland full dataset, GB - Great Britain, PT - Portugal, CH13 - first subset of 13 individuals from Switzerland, CH13_2 - second subset of 13 individuals from Switzerland.

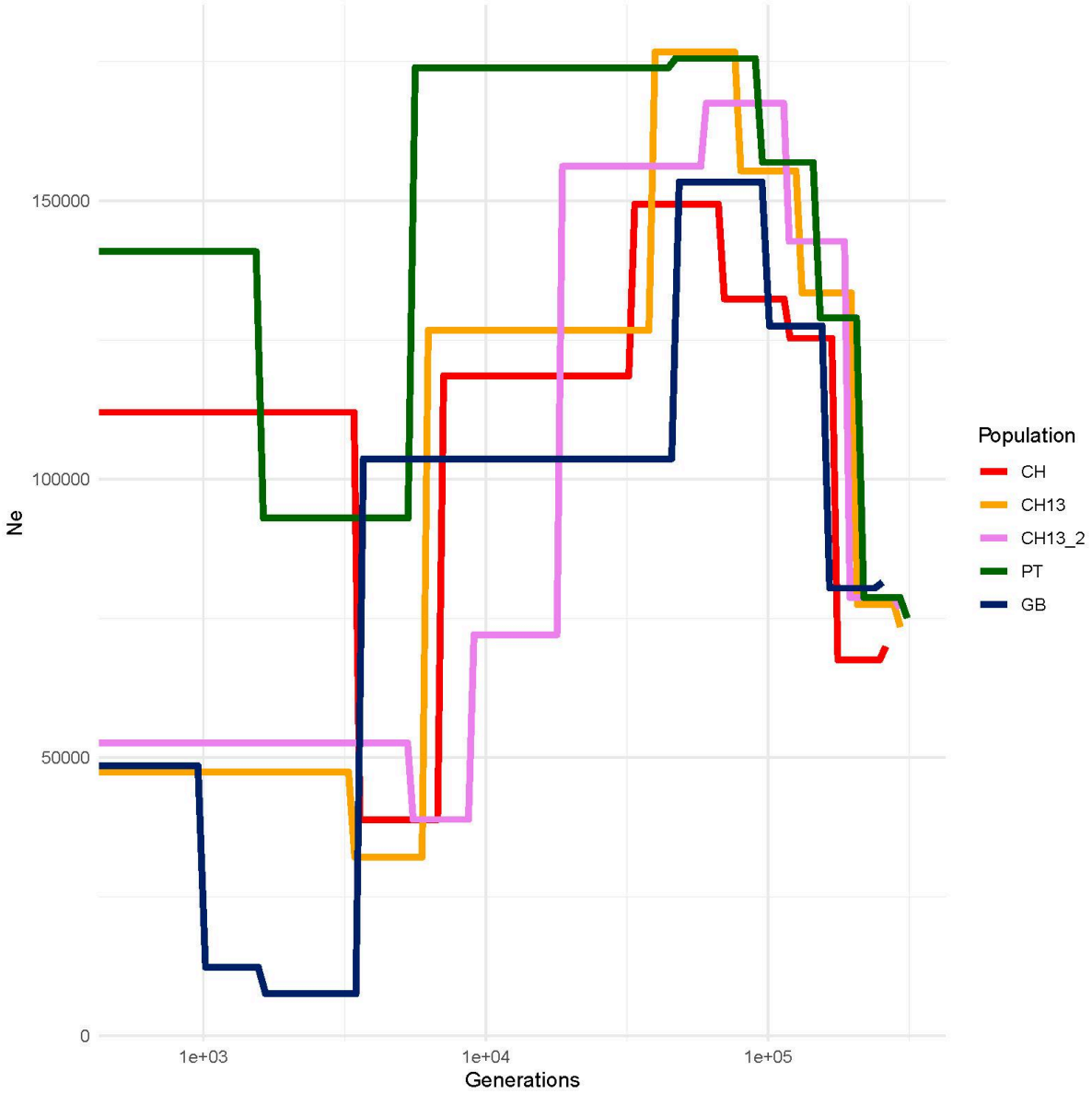

Figure S2. Inferred effective population sizes from SMC++ for each dataset. Abbreviations: CH - Switzerland full dataset, GB - Great Britain, PT - Portugal, CH13 - first subset of 13 individuals from Switzerland, CH13_2 - second subset of 13 individuals from Switzerland.

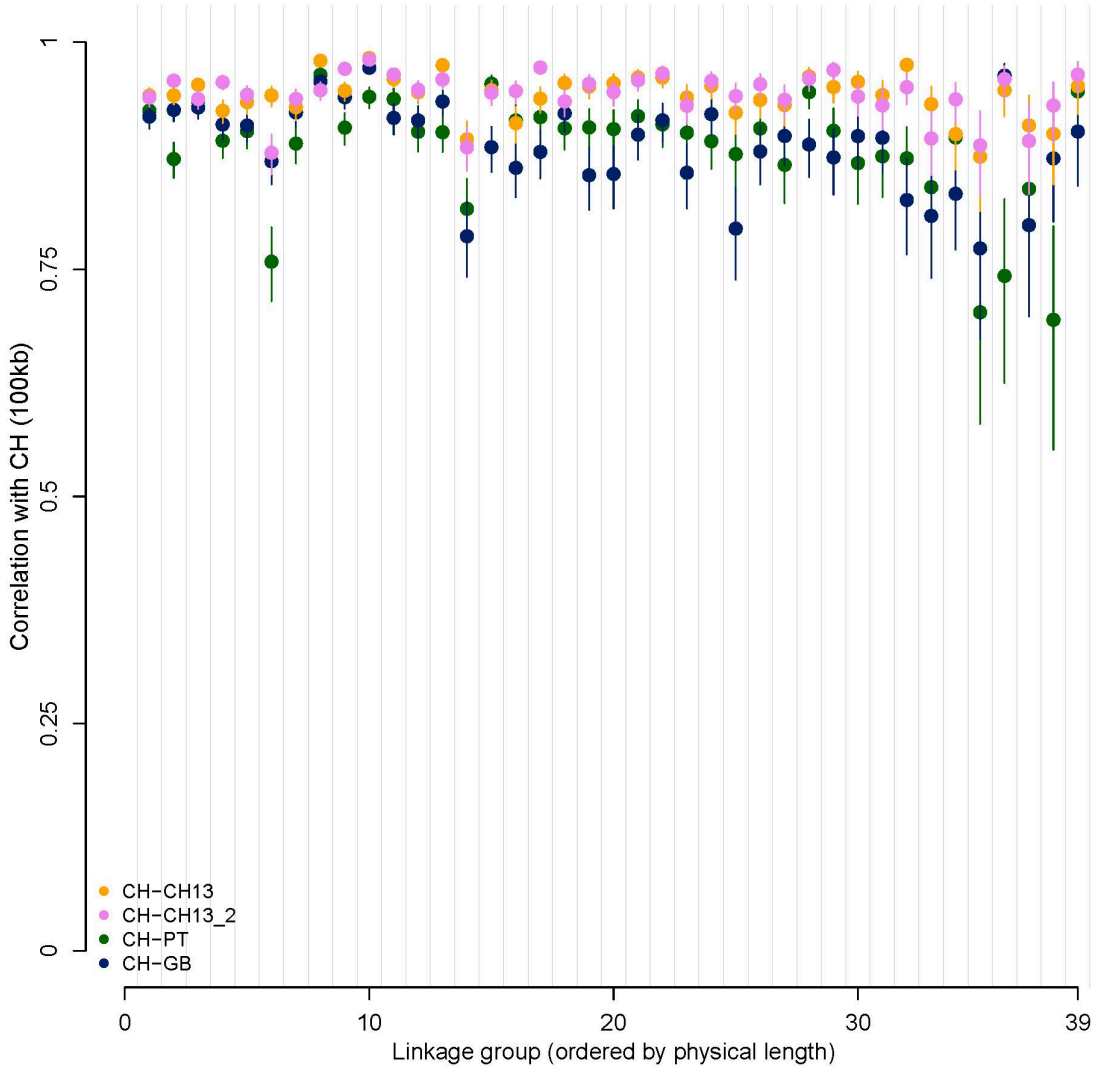

Figure S3 - Pearson correlation between the full Swiss dataset and all other populations per scaffold in 100kb windows.
Correlation with CH13 and CH13_2 is higher than between Switzerland and any other population as expected. Estimates of correlation get less accurate as signified by the 95% confidence intervals as the linkage groups get smaller due to reduced sample size. However, the value of the average correlation is fluctuating around the genome wide average value (main text Figure 4C) depending on the linkage group. Abbreviations: CH - Switzerland full dataset, GB - Great Britain, PT - Portugal, CH13 - first subset of 13 individuals from Switzerland, CH13_2 - second subset of 13 individuals from Switzerland.

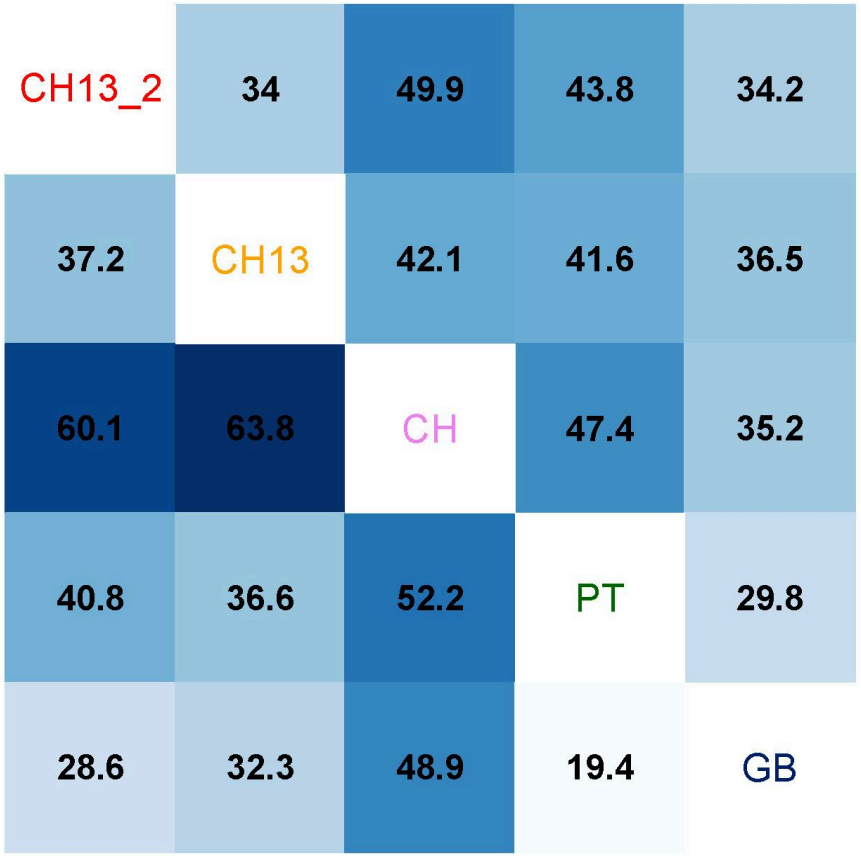

Figure S4. Hotspot sharing between pairs of populations.
Upper diagonal represents the percentage of local hotspots shared. Lower diagonal represents the percentage of global hotspots shared. Abbreviations: CH - Switzerland full dataset, GB - Great Britain, PT - Portugal, CH13 - first subset of 13 individuals from Switzerland, CH13_2 - second subset of 13 individuals from Switzerland.

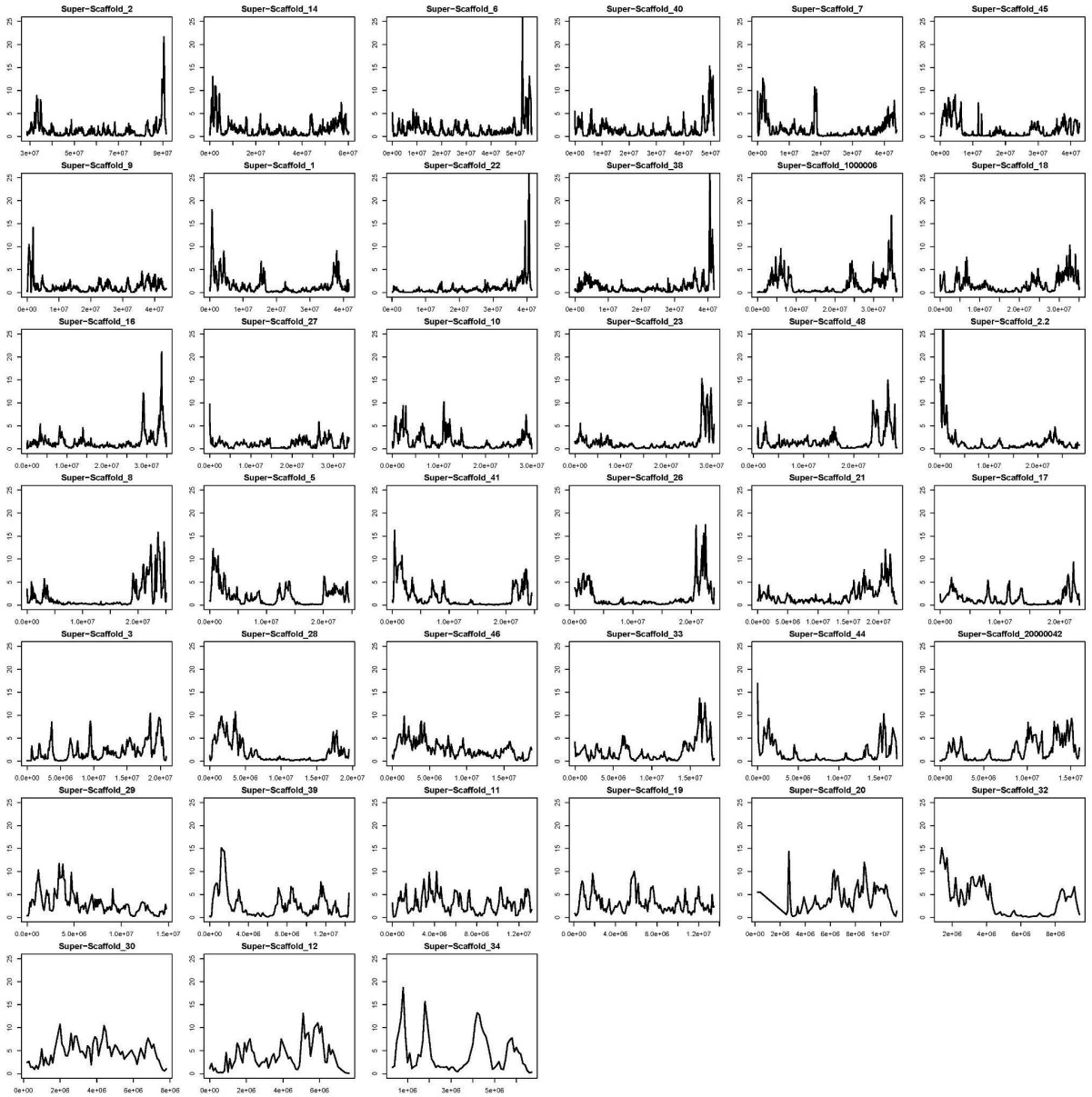

Figure S5. Recombination rates along the physical sequence in the full Swiss dataset (CH - n=76) for all linkage groups in 100 kb windows.

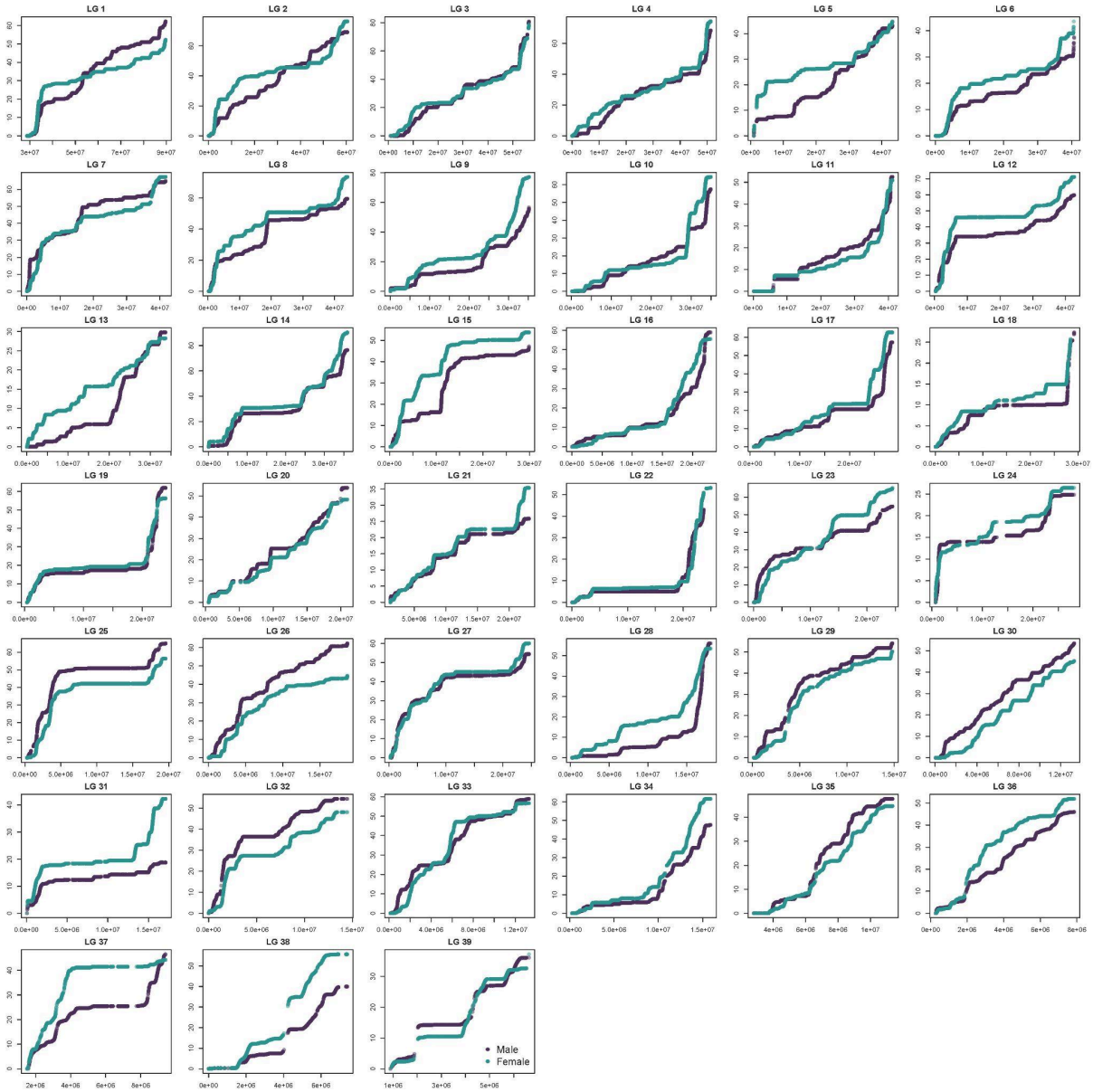

Figure S6. Sex-specific Marey maps.
On the y-axis is the cumulative cM position and on the x-axis the physical positions. Points on the plots are the final monotonically-increasing GAM model fitted values.

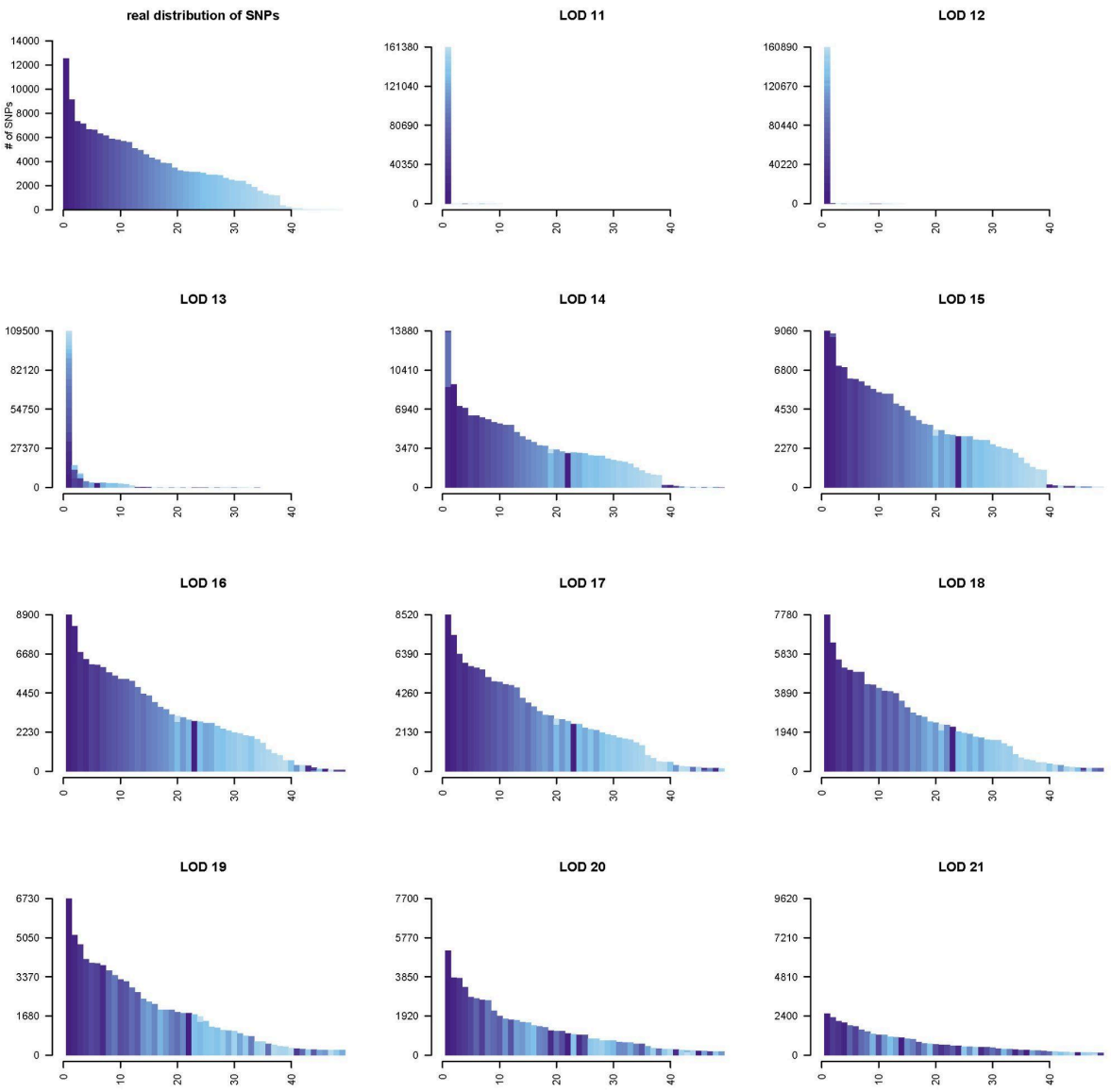

Figure S8. Linkage group length (in # of SNPs) for each LOD score cutoff in LepMap3's SeparateChromosome output.
Y-axis shows the number of SNPs (care for non-standardized axes among plots) and x axis the number (order) of linkage groups. Top-left panel is the observed distribution of SNPs in the assembly scaffolds. Scaffolds are coloured based on their physical length (darker = longer). Note how LOD15 (2nd row, 3rd column) shows a similar profile with the observed scaffolds except for the splitting of the longest scaffold into 2 linkage groups (a split observed in LOD14 too - as a dark line between marks 20 and 30 on the x-axis). The one merged linkage group can be seen around mark 20 on the LOD15 plot (a bar with 2 colours, one very light being the non-ordered very small linkage group). We are confident in this merge since it appears for larger LOD scores where over splitting would be expected.

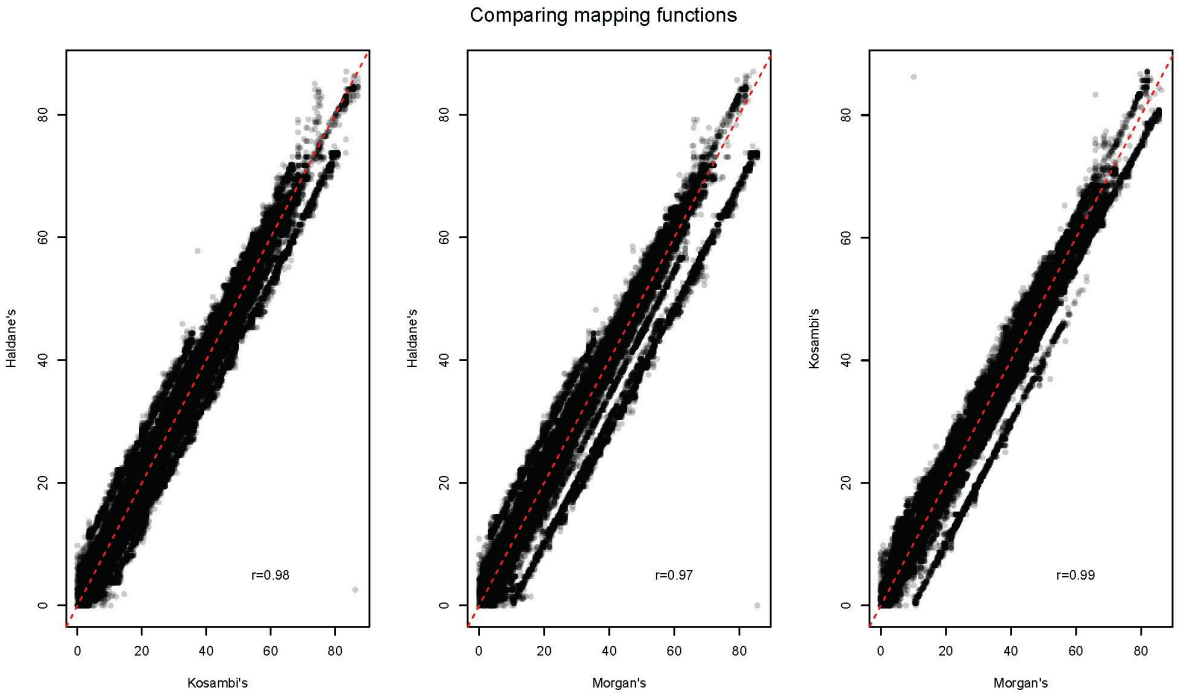

Figure S9 - Comparing mapping functions.
Each dot is one marker used in the linkage map. Different mapping functions give slightly different estimates of cM per scaffold but the same relative positions. Correlation values are shown on each plot. Dashed red lines are the y=x lines.

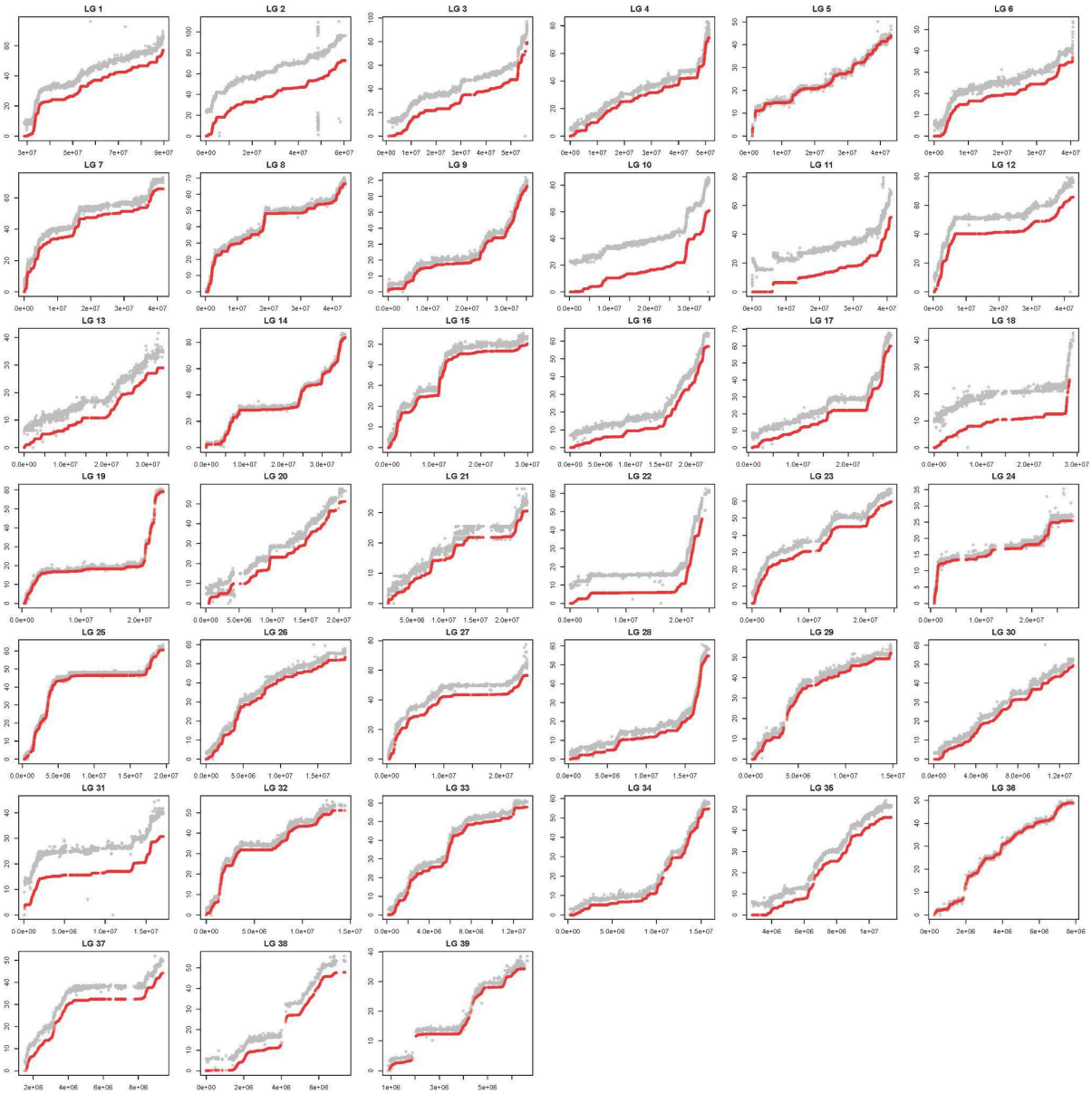

Figure S10. Monotonically increasing GAM models for linkage map construction. Raw sex-averaged cM positions for each marker before any post-ordering filtering are presented in grey points. Red points are the monotonically increasing GAM fitted values after pruning the ends of each linkage group for jumps larger than 2cM and removing residuals of the regression of genetic order on physical order larger than 100 (see methods for more details).