

Supplementary Material

Figure S1. Individual estimates from LDhat under neutrality and selection

The individual estimates from the 250 replicates underlying the $s = 0$ (black circles) and $s = 0.1$ (red circles) boxplots from Figure 2 of the main text.

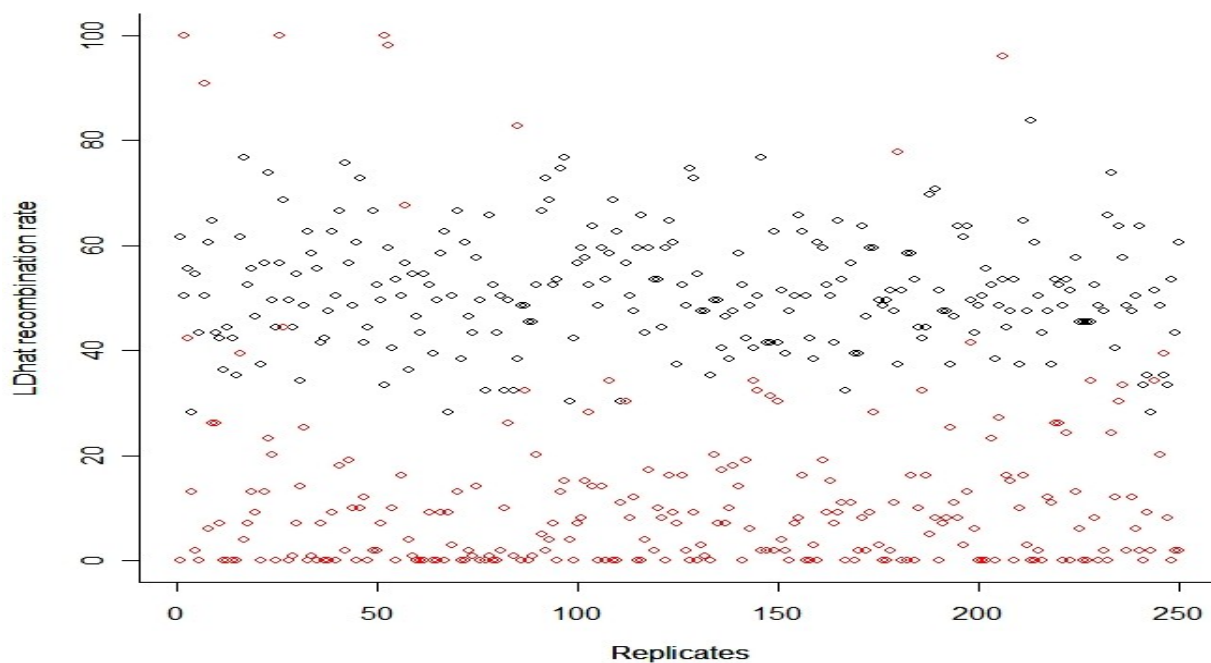


Figure S2: The effect of no recombination on Tajima's D under neutrality

The density of Tajima's D in 300 kb regions approximated from SelSim simulations (see Methods). The density in bold green corresponds to neutrality and no recombination; the dashed green curves show the same density but scaled by a factor of 10 and 100, illustrating that extreme values of Tajima's D can arise relatively often under neutrality if low recombination regions outnumber those under strong positive selection.

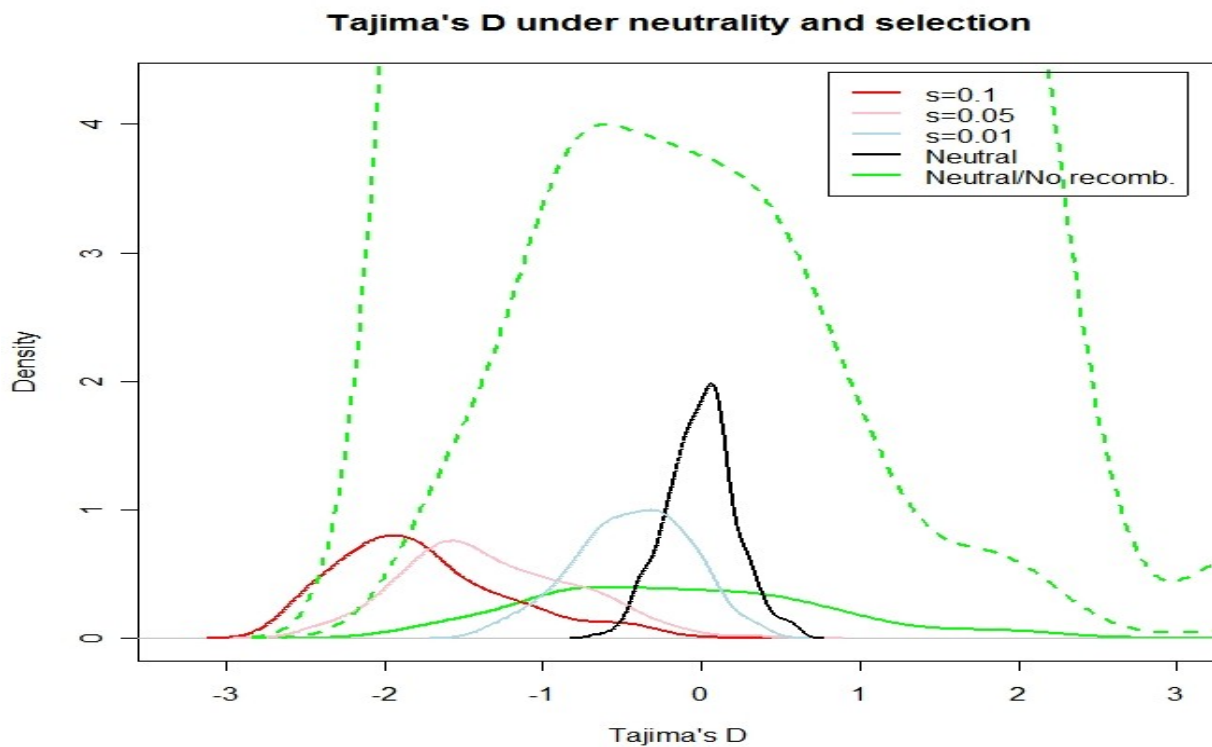
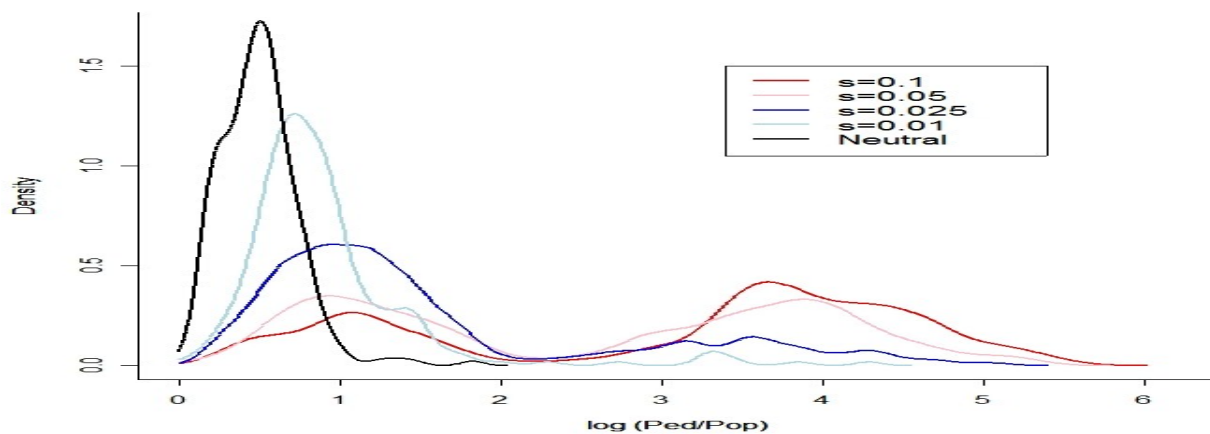


Figure S3. Ped/Pop with low and high recombination rates

The results gained by repeating the simulation study in the main text (see Fig. 5) using low and high rates of recombination in the simulation of population data from SelSim (Spencer and Coop, 2004). The low and high rates of recombination correspond to the 20% and 80% points of the population-based recombination rate estimates used in our scans for selection (ie. from the HapMap and Perlegen datasets). Note: the densities relating to the low recombination rate are only shown to the right-side of zero (spikes of density exist at $-\infty$ due to zero draws from the Poisson distribution that generates pedigree estimates).

Low recombination rate:-



High recombination rate:-

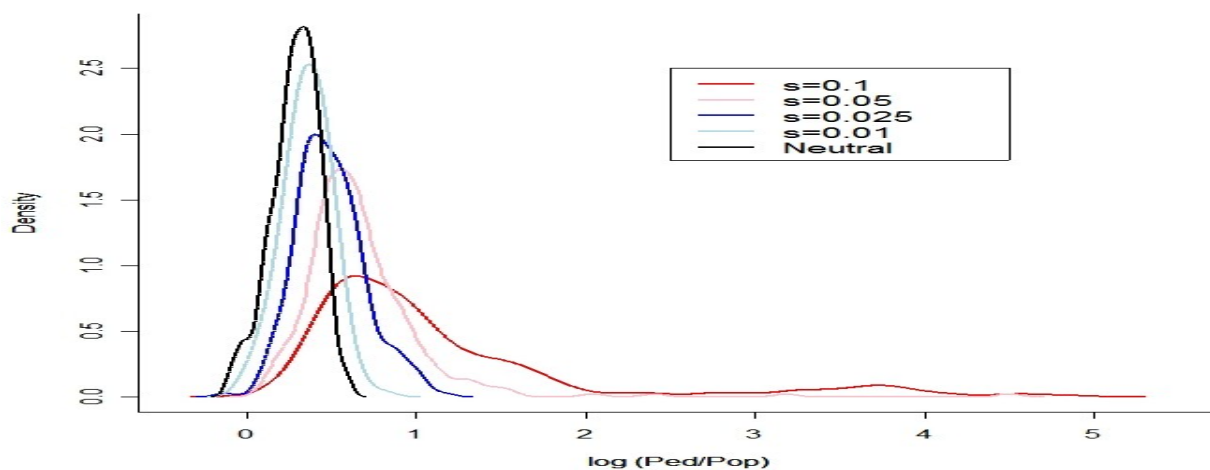
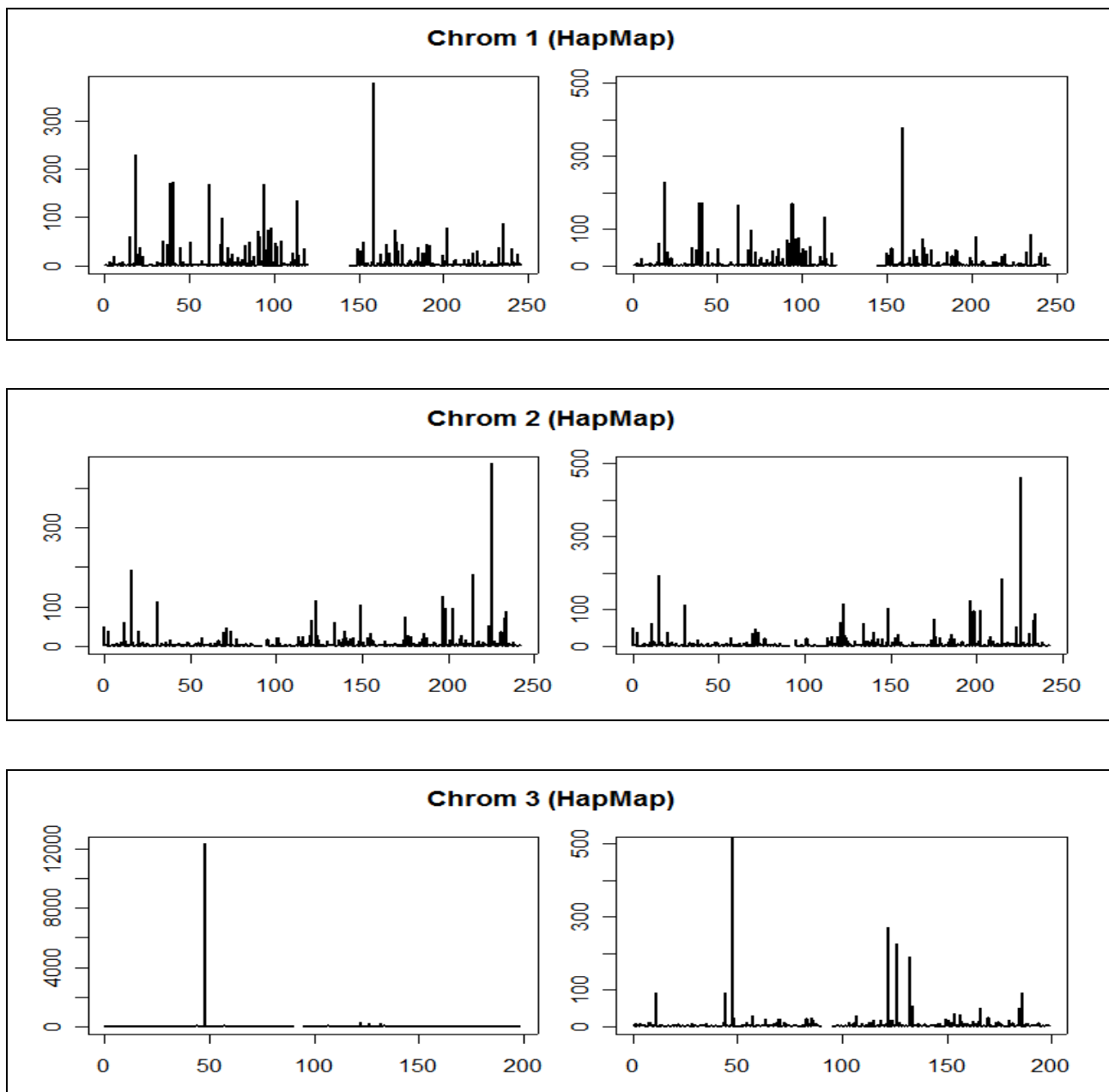
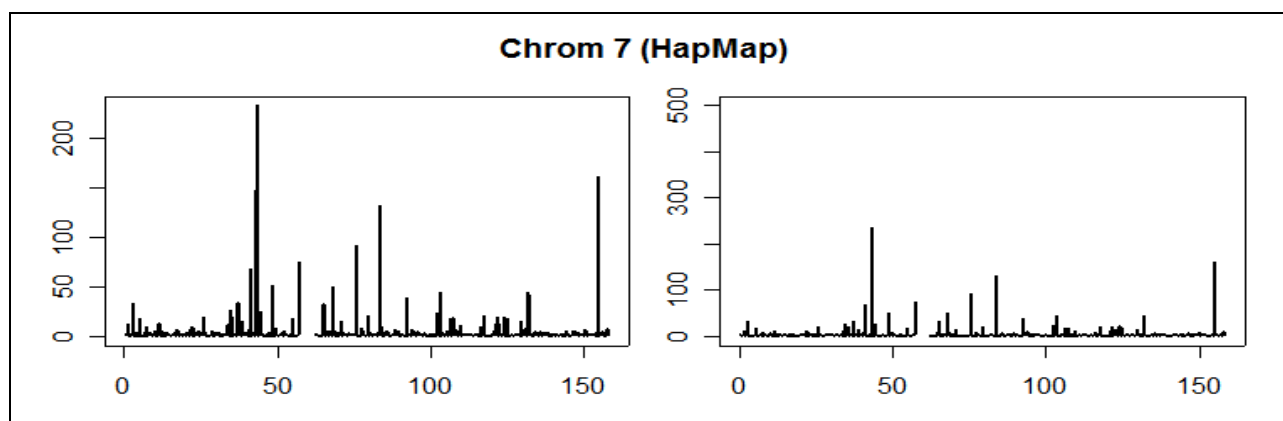
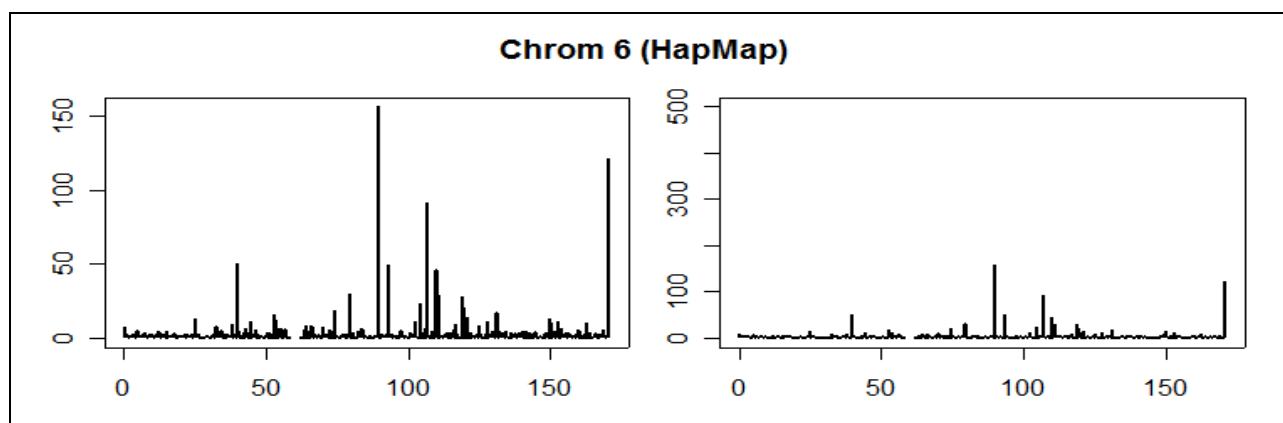
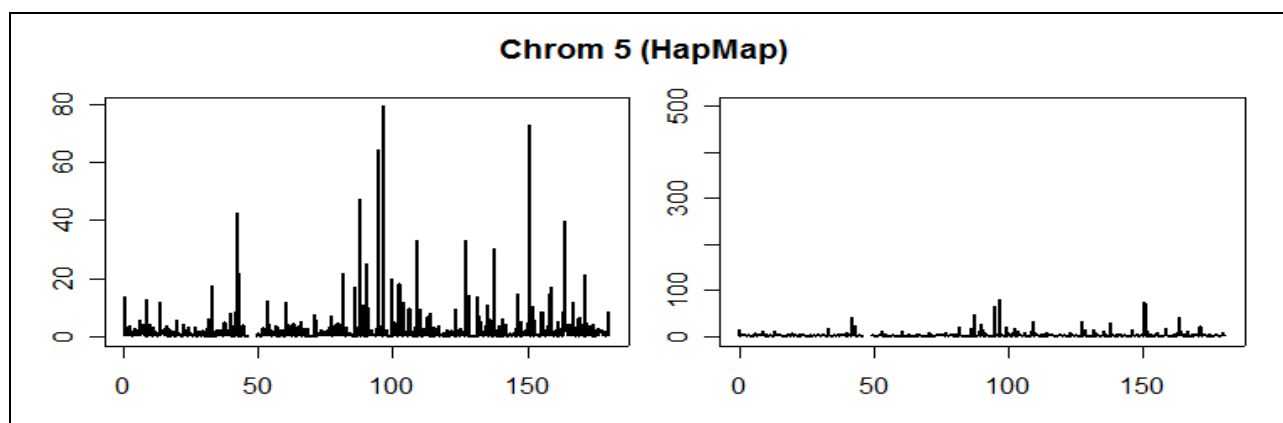
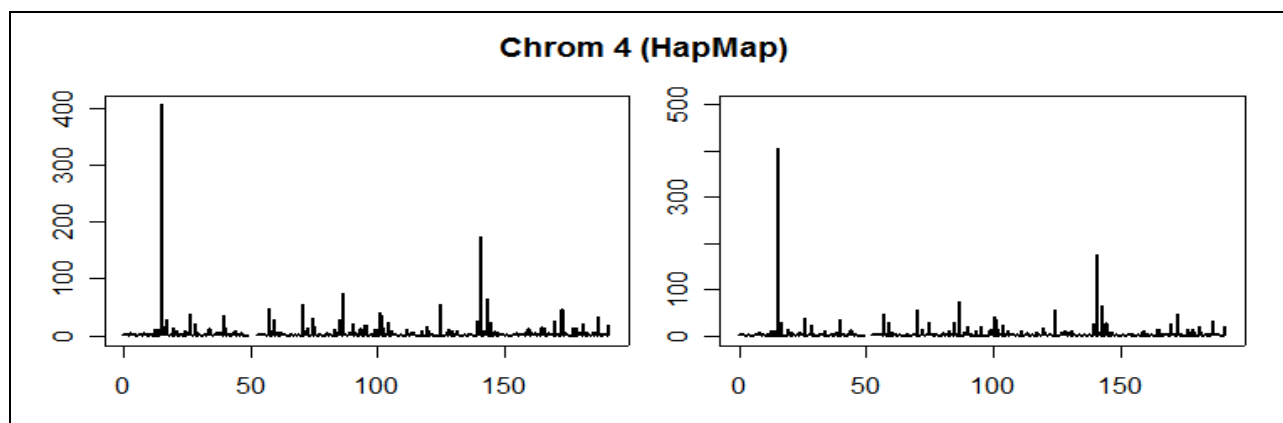
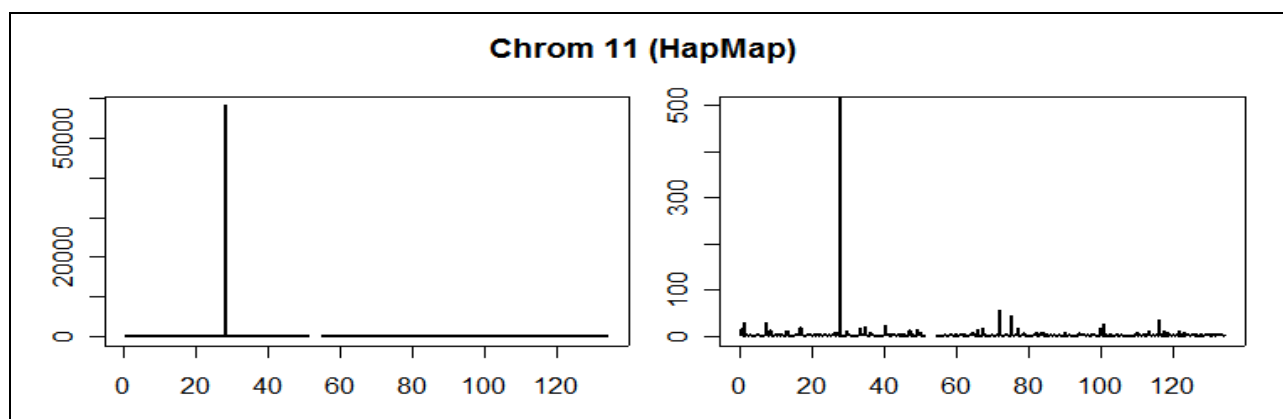
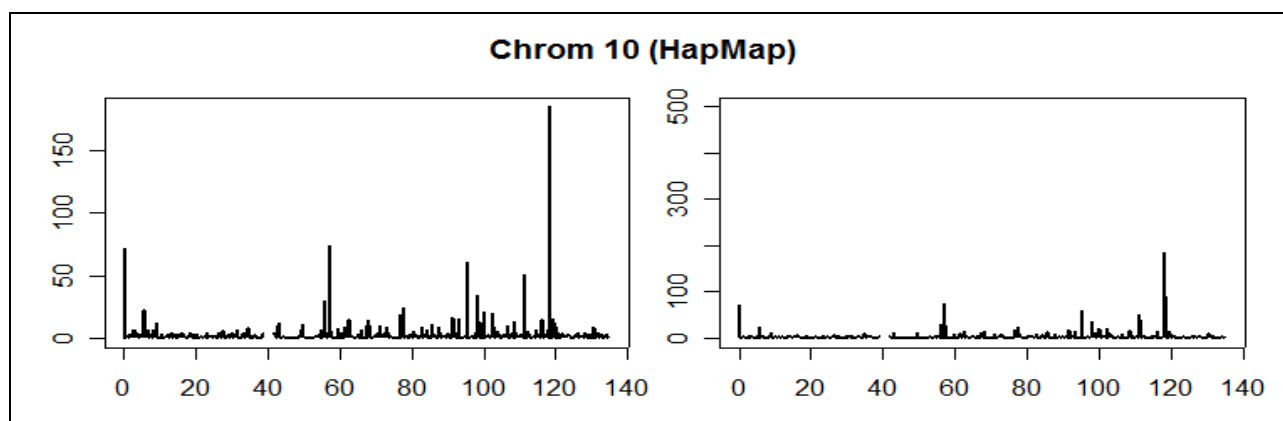
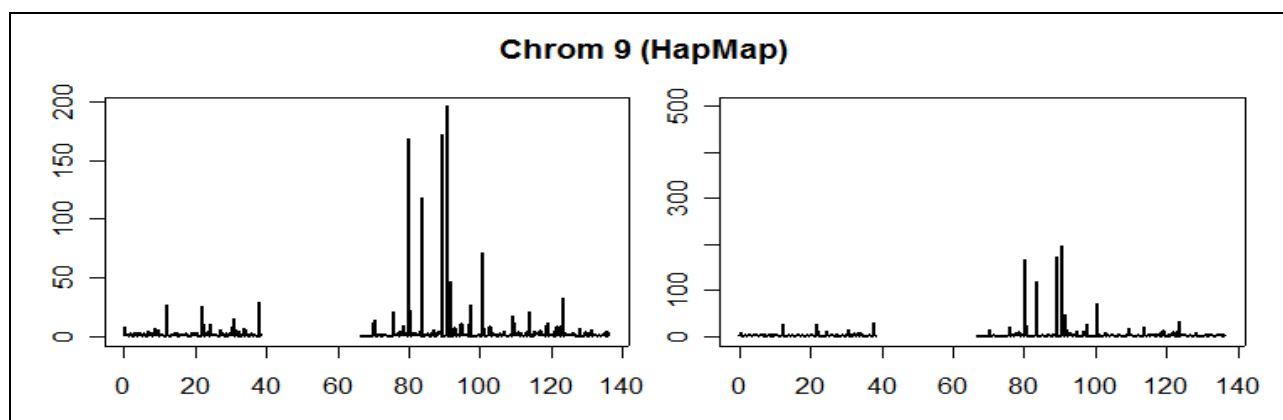
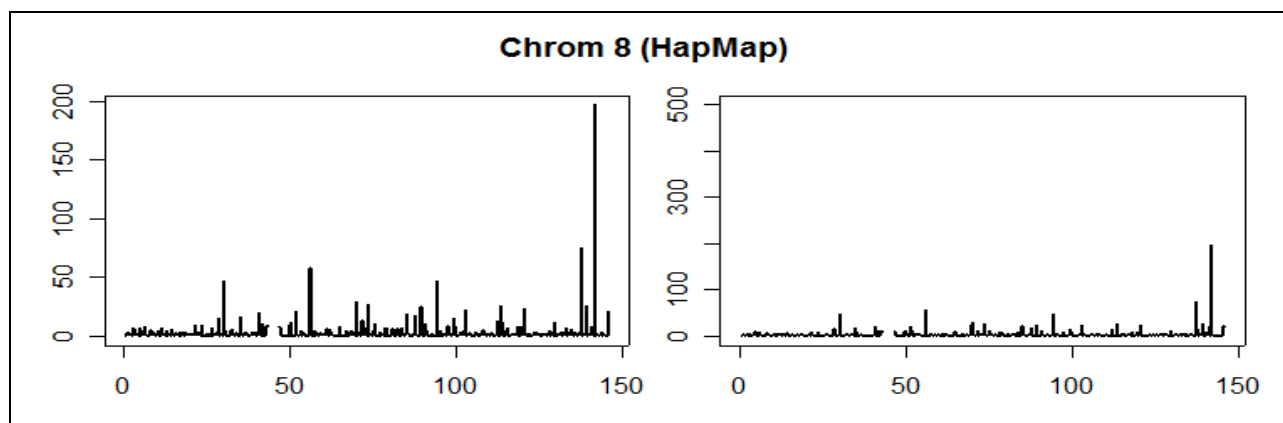


Figure S4. Scans for selection over the human genome – using HapMap data

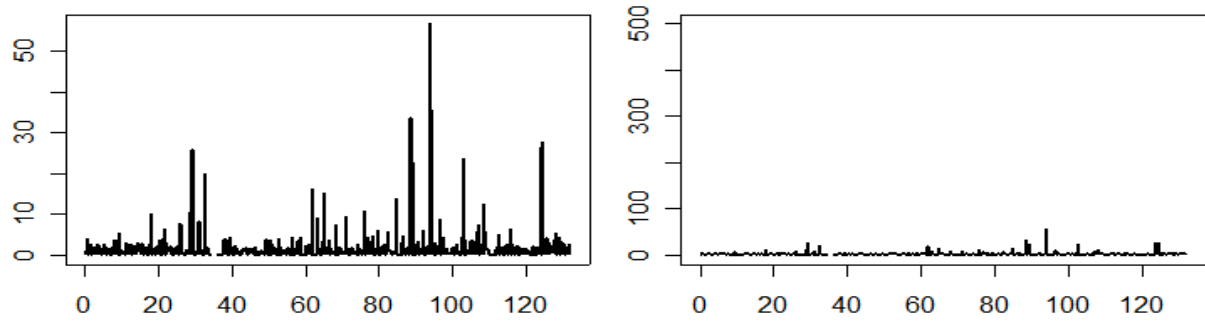
The scans for selection that were gained by applying the Ped/Pop approach using the HapMap data to procure population-based recombination rate estimates (pages 4-9) . Those in Figure S5 (pages 10-15) were gained from using the population data from the Perlegen genotyping project (see Methods). The x-axes give the physical position (in Mb) along the genome, while the y-axes give the Ped/Pop value. Each scan shows one plot with y-axis scaled according to the maximum value of Ped/Pop on the chromosome (left), and the corresponding plot where the y-axis takes a maximum value of 500 (right) to aid comparison between chromosomes.



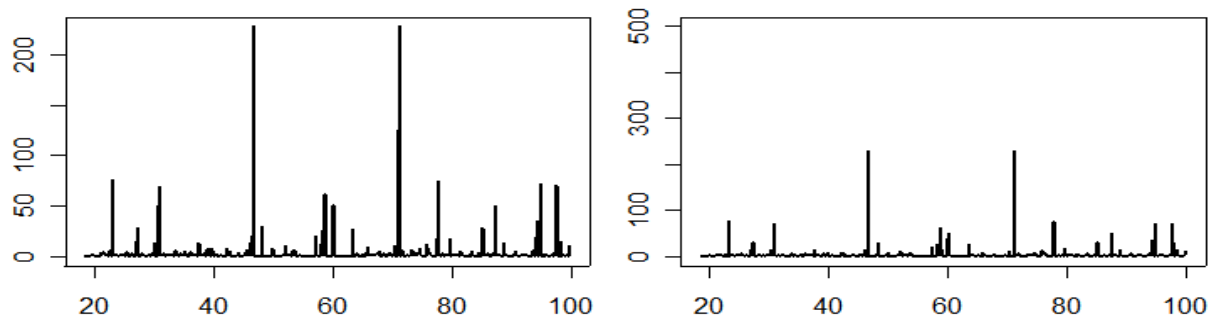




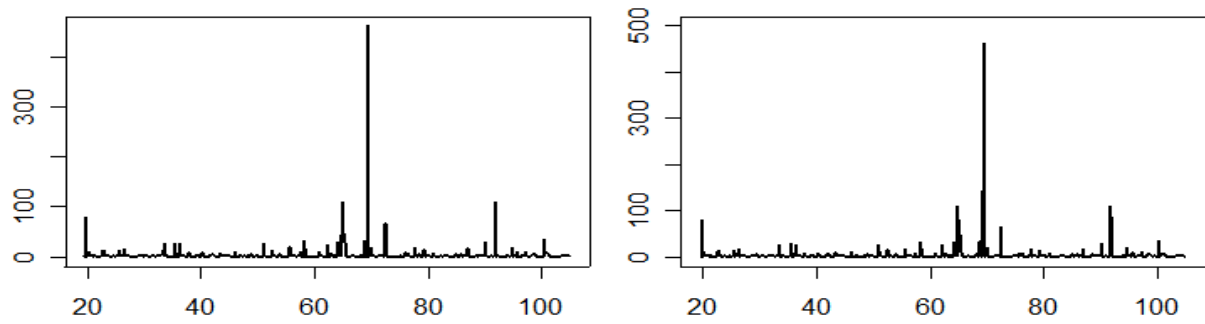
Chrom 12 (HapMap)



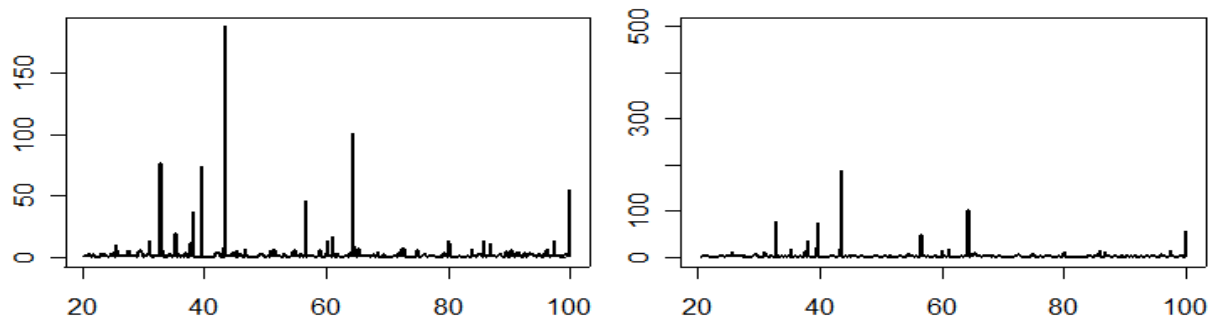
Chrom 13 (HapMap)



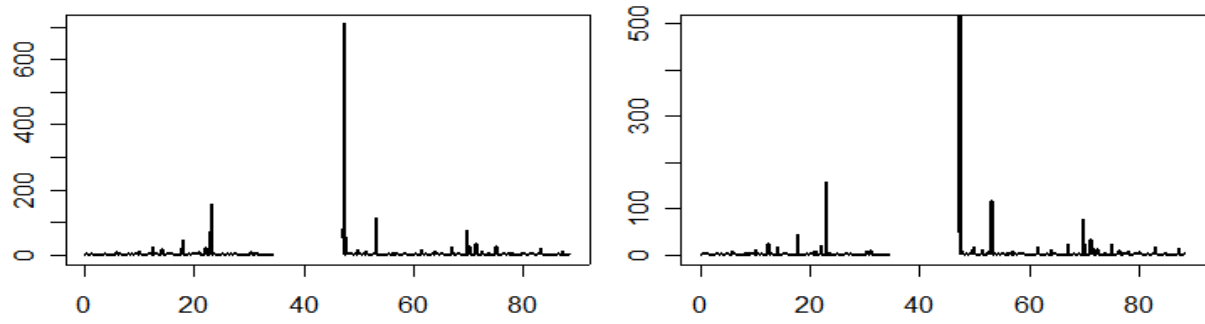
Chrom 14 (HapMap)



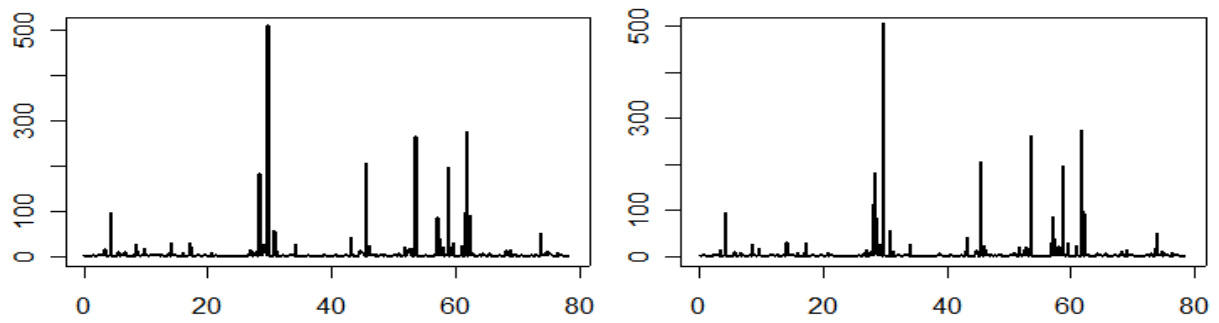
Chrom 15 (HapMap)



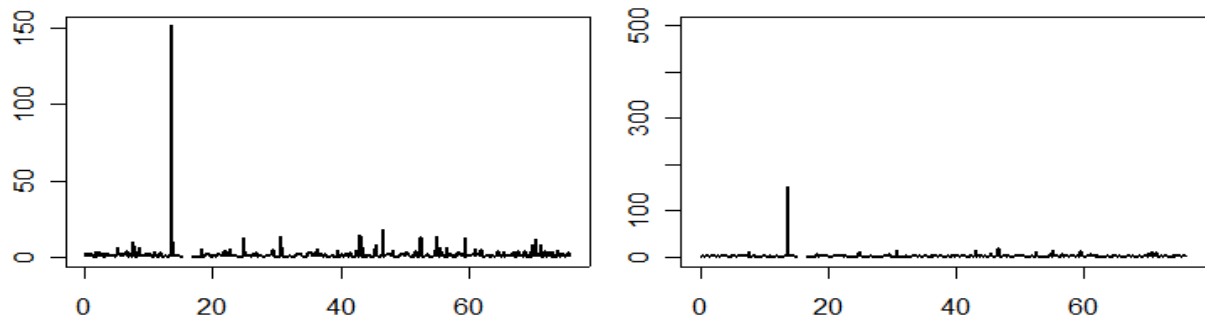
Chrom 16 (HapMap)



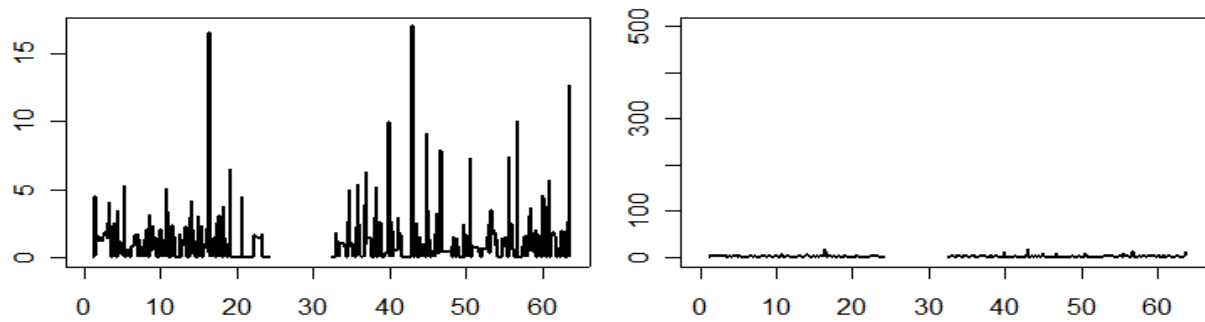
Chrom 17 (HapMap)



Chrom 18 (HapMap)



Chrom 19 (HapMap)



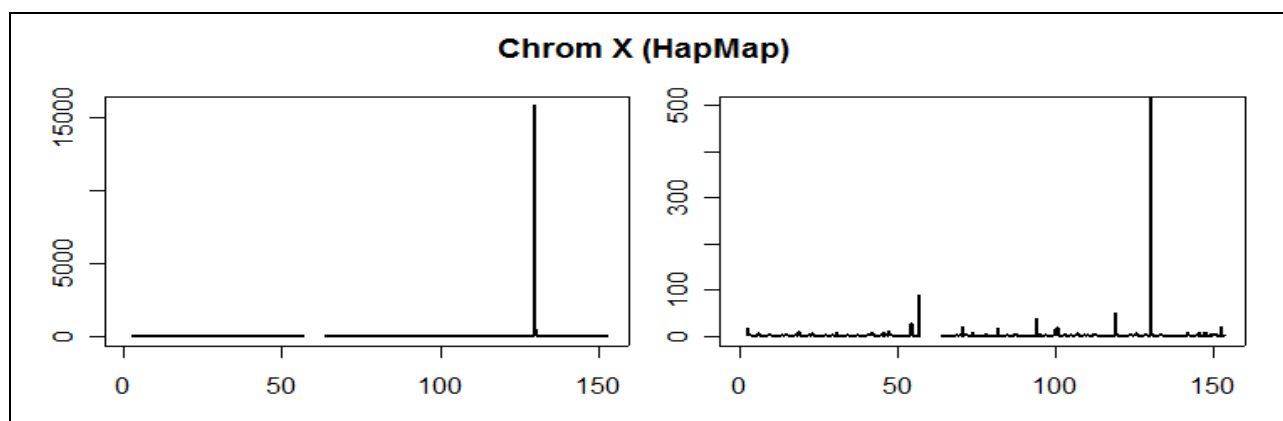
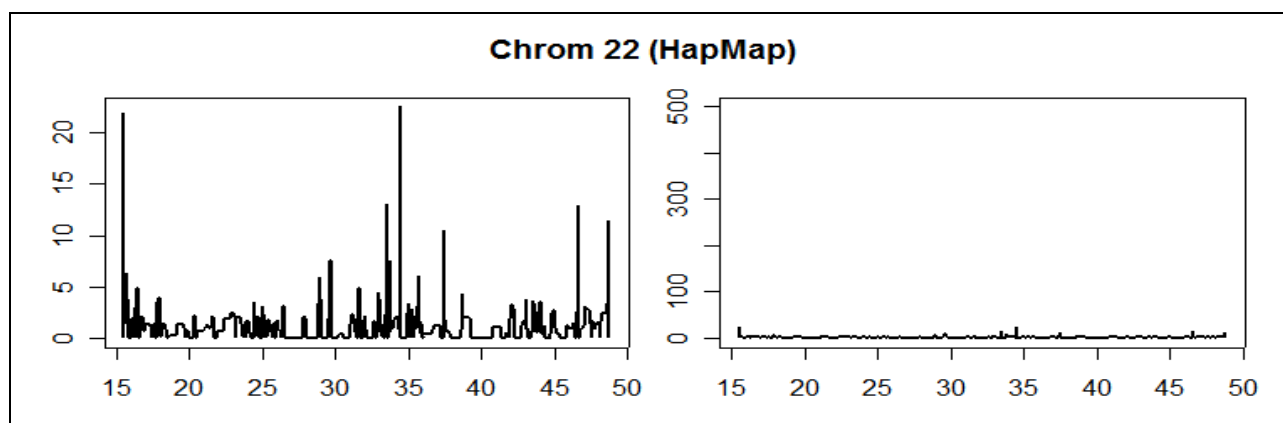
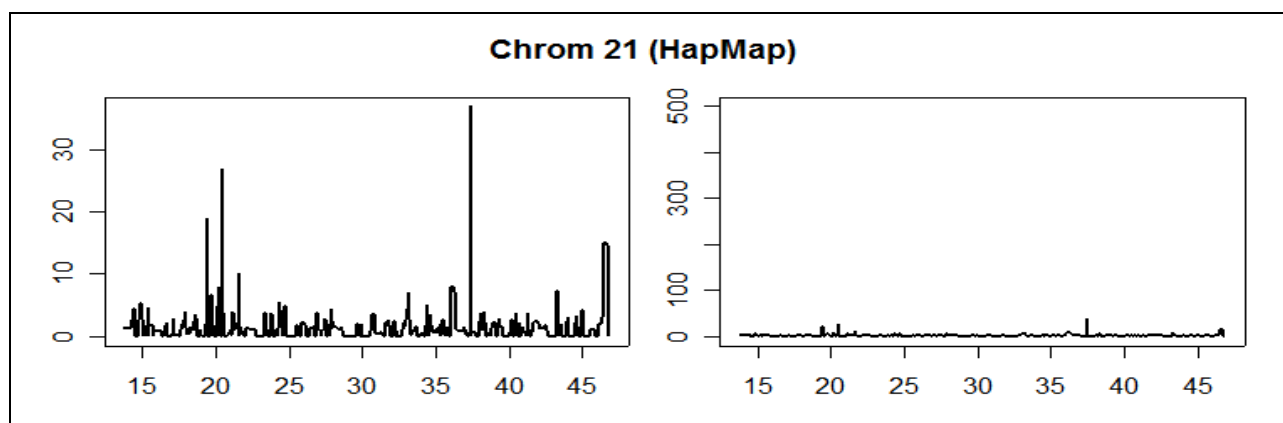
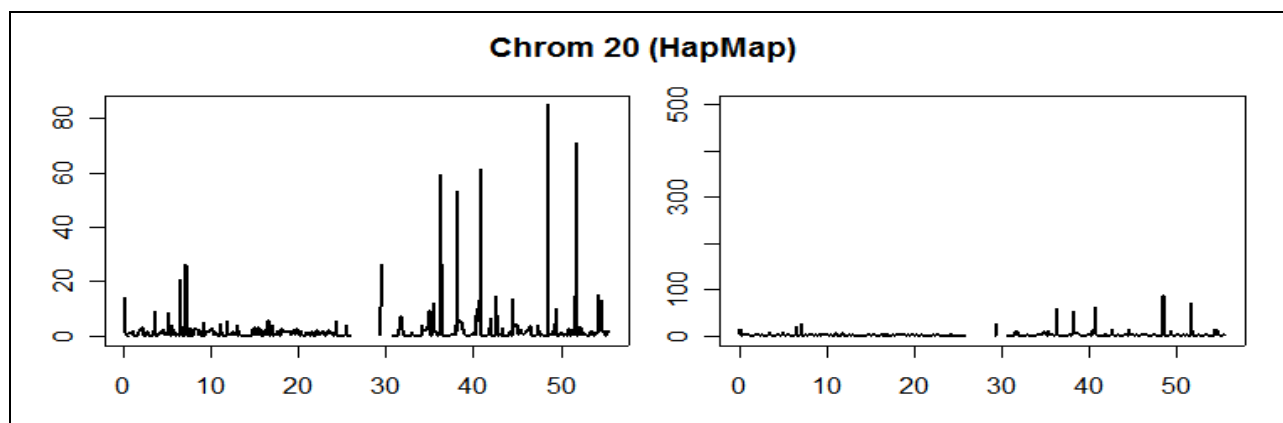
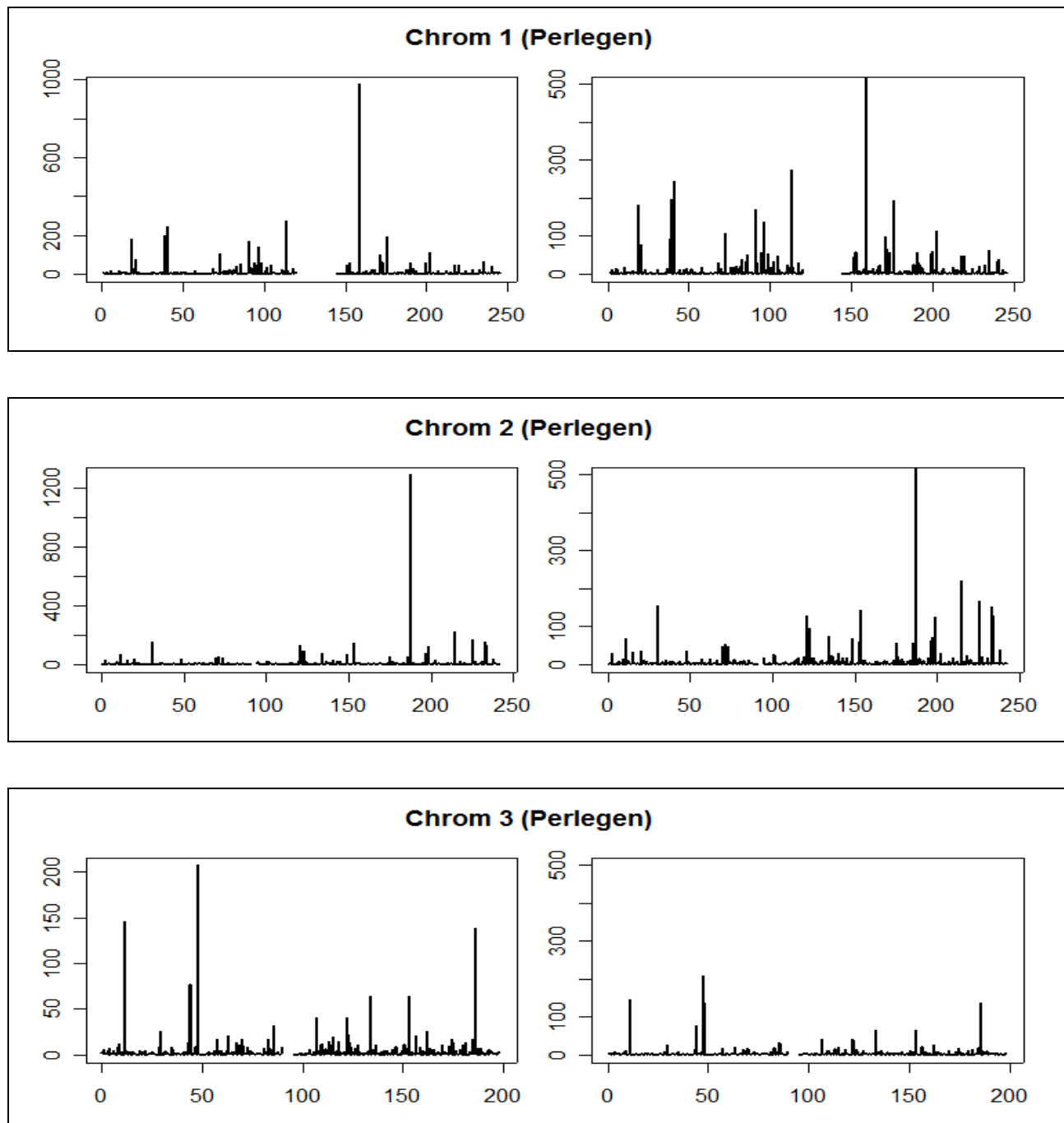
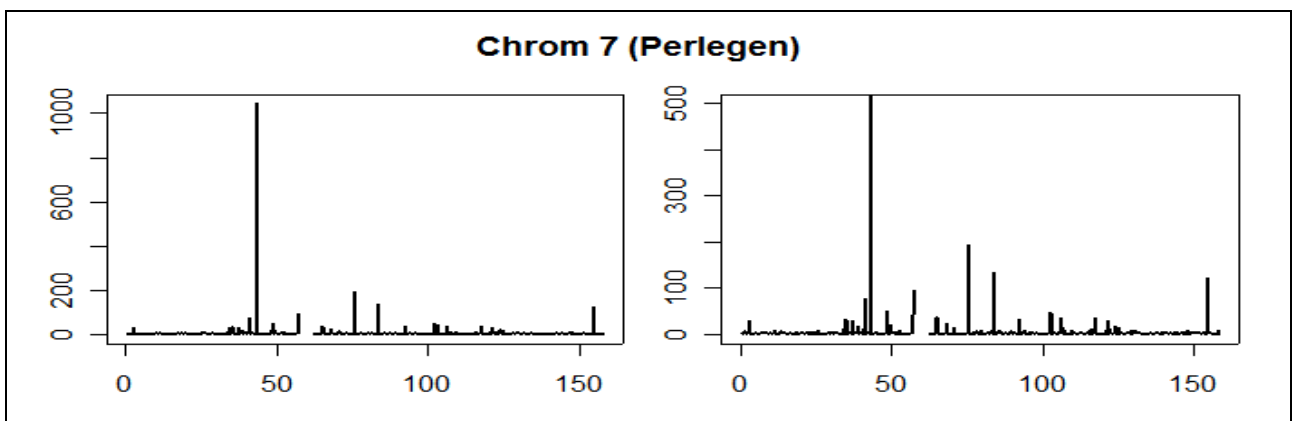
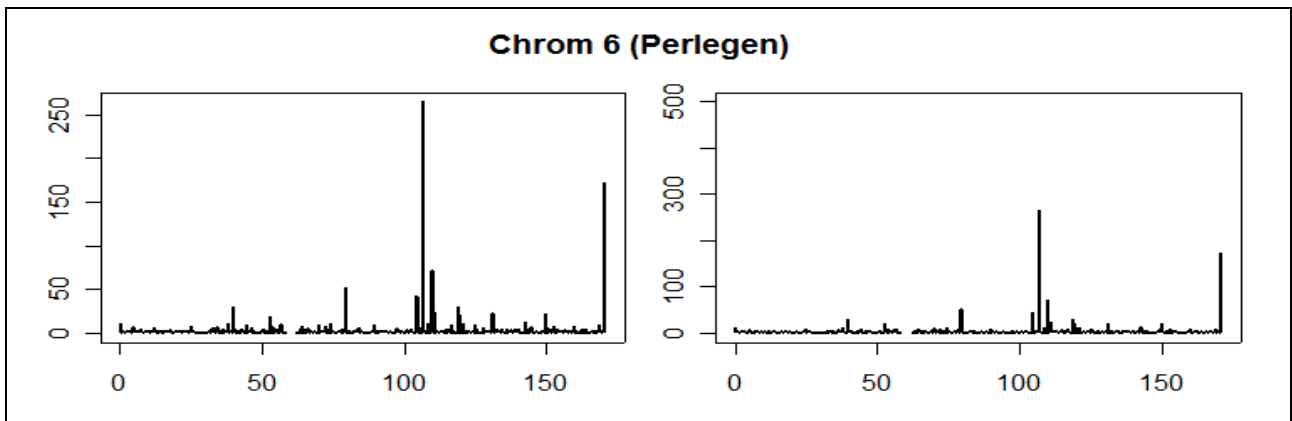
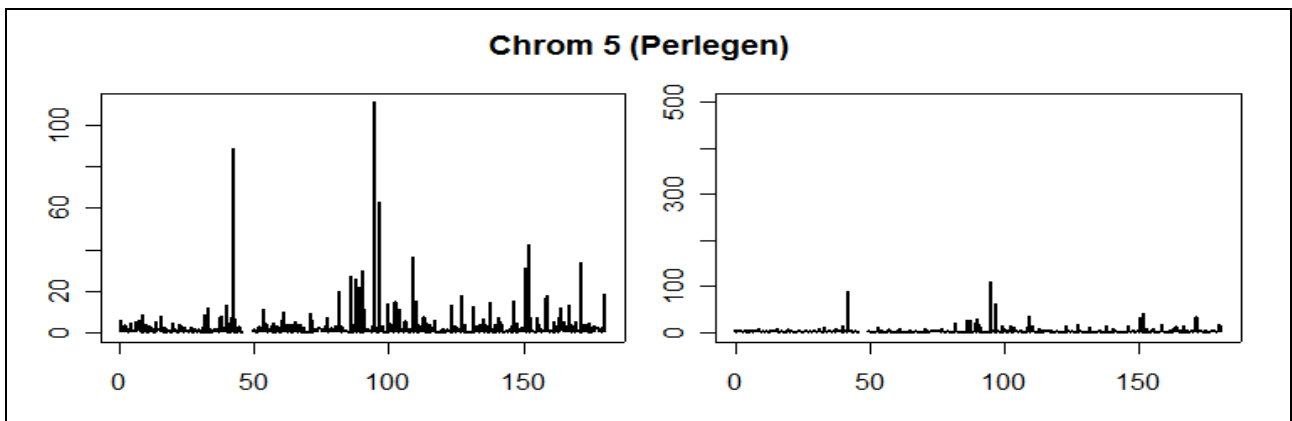
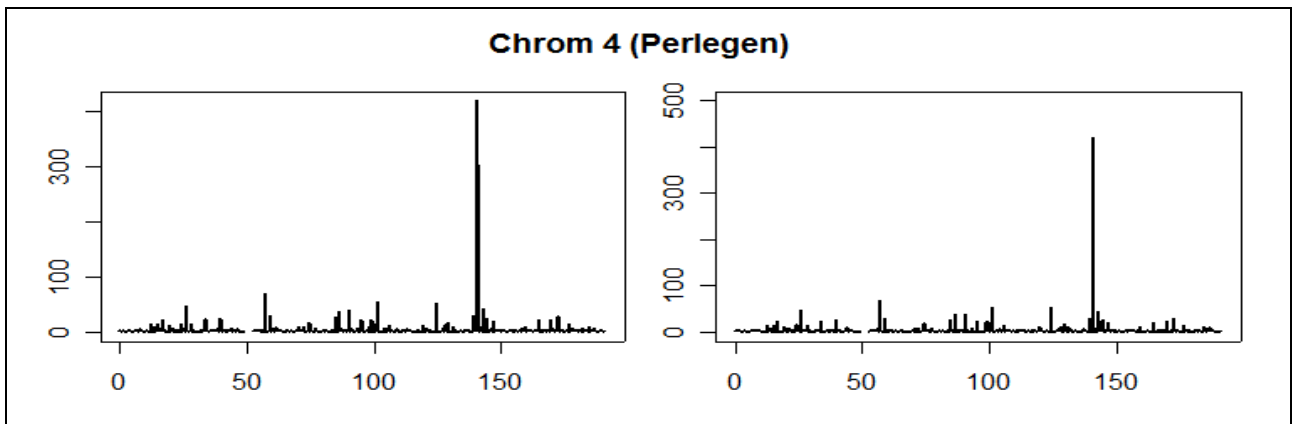
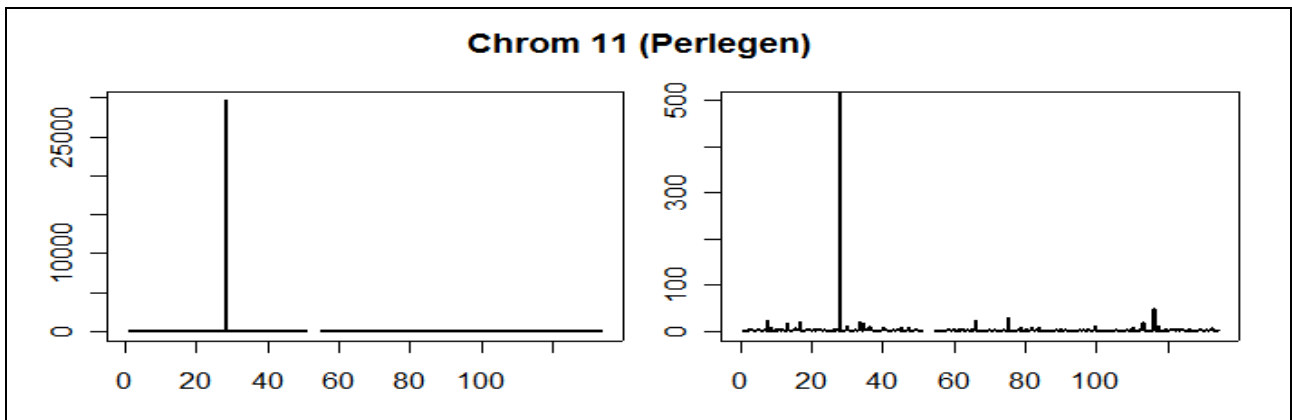
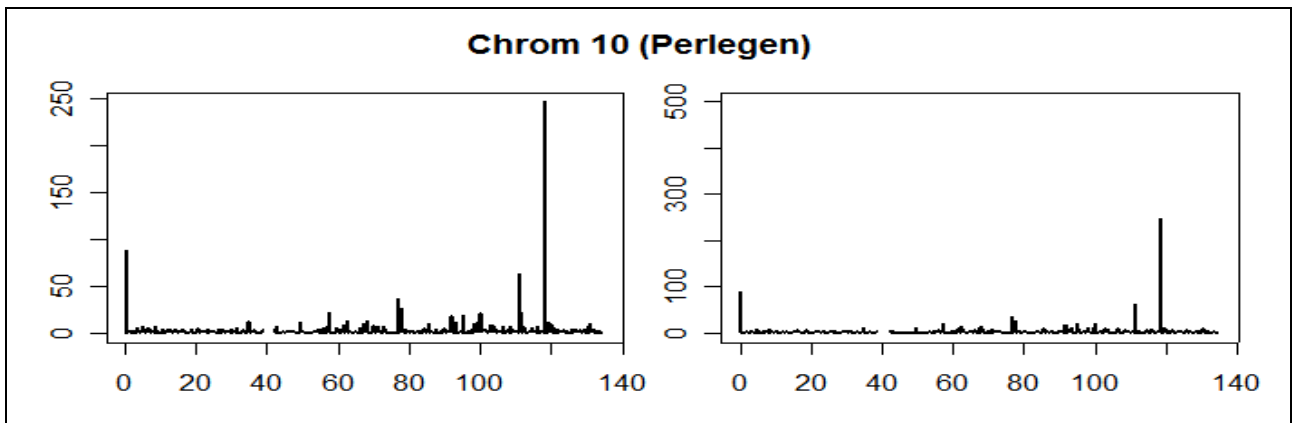
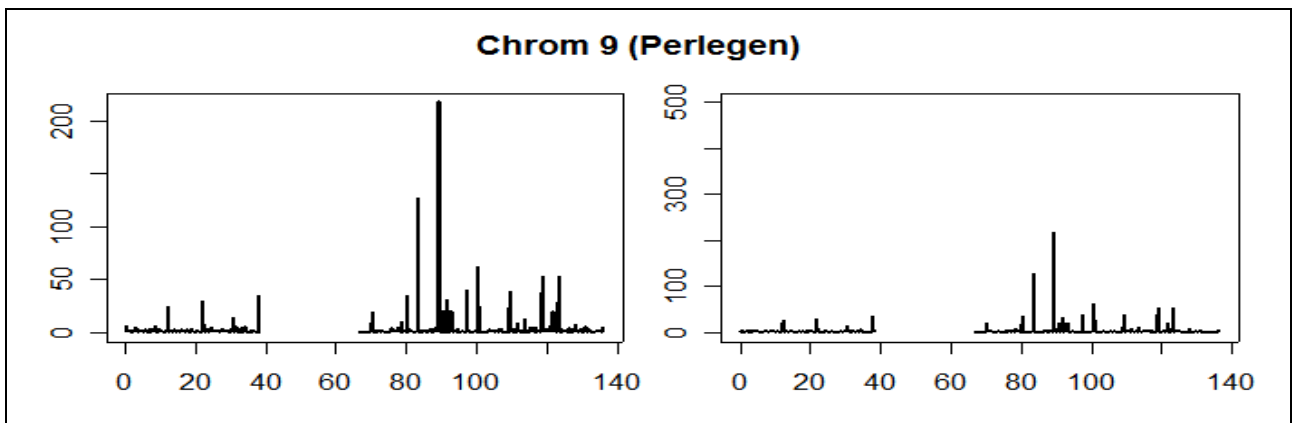
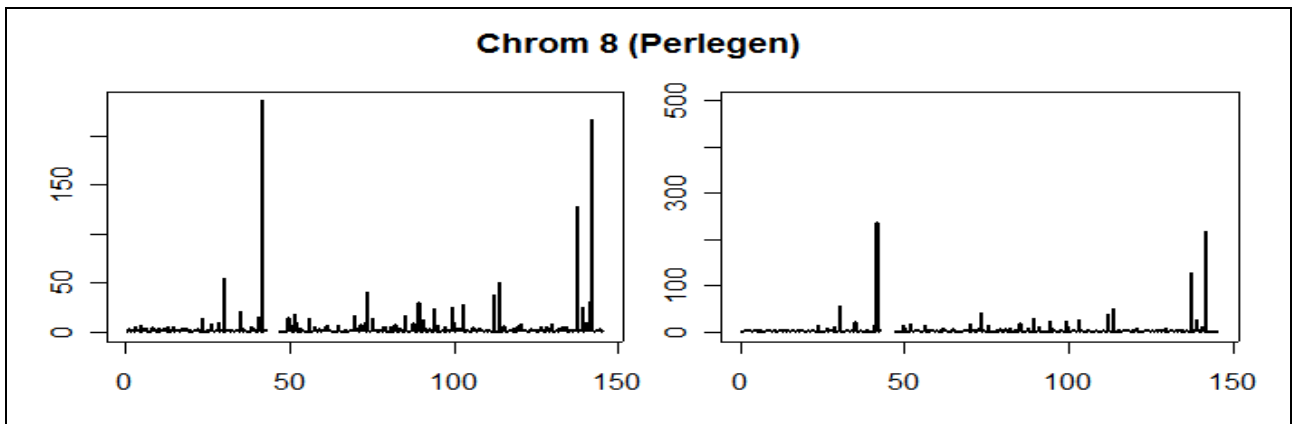


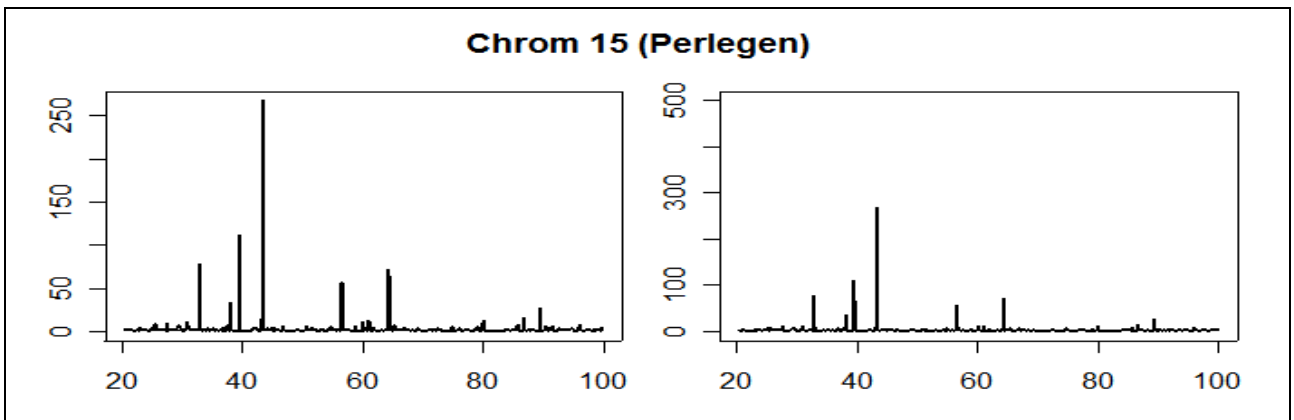
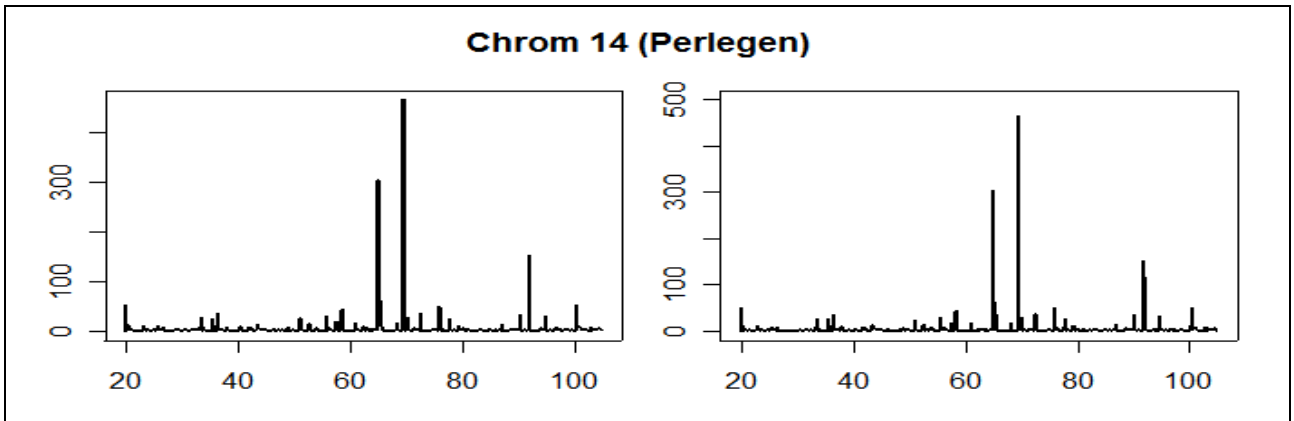
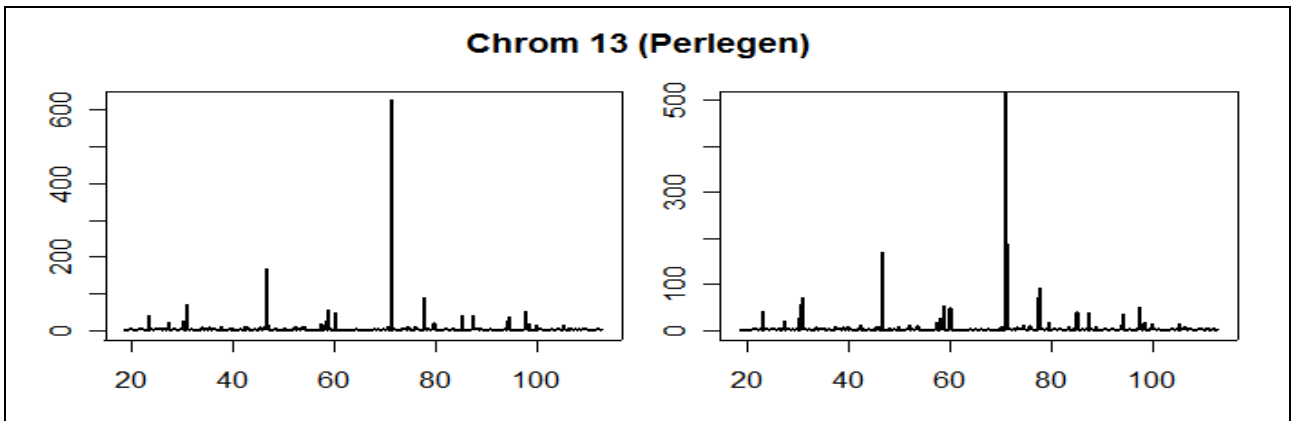
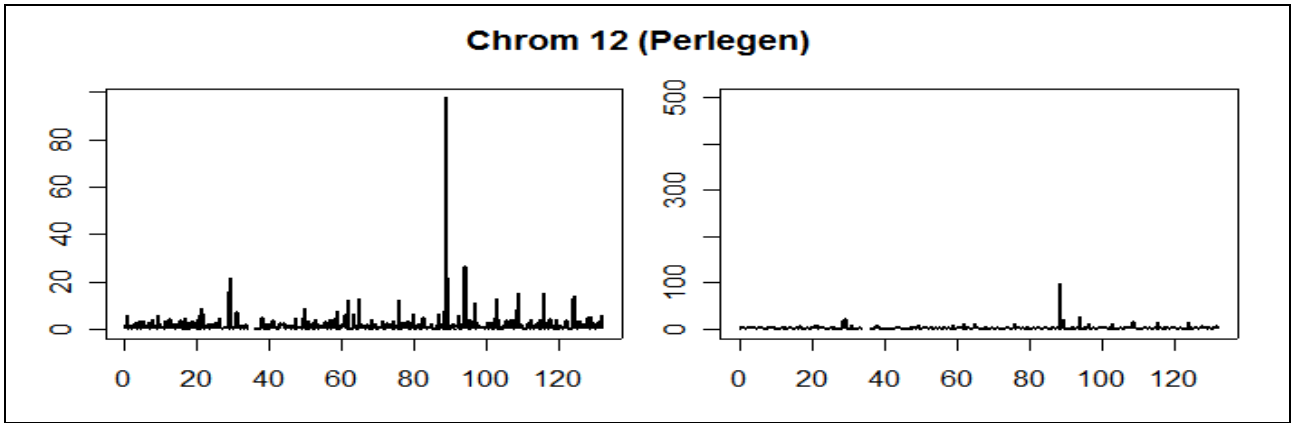
Figure S5. Scans for selection over the human genome using Perlegen data

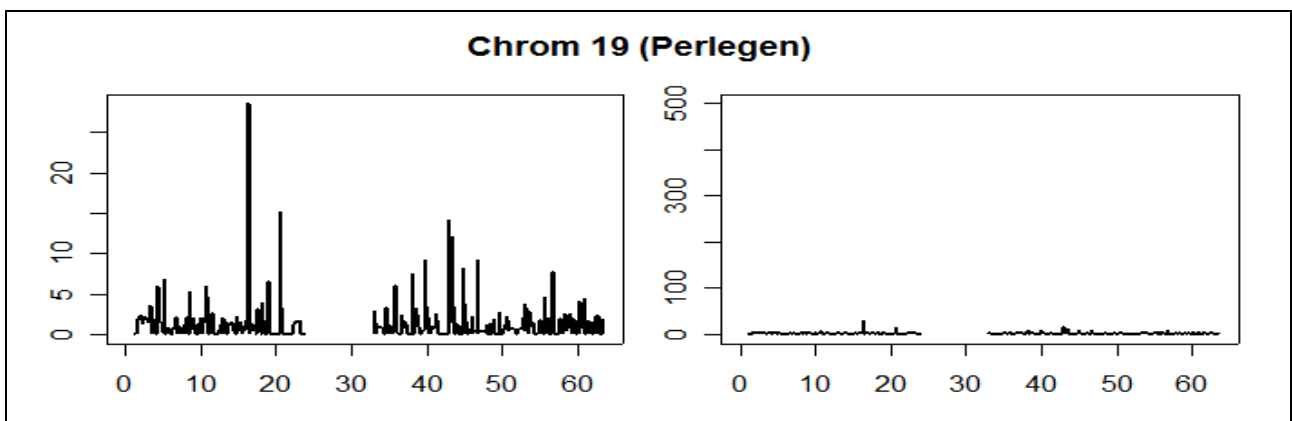
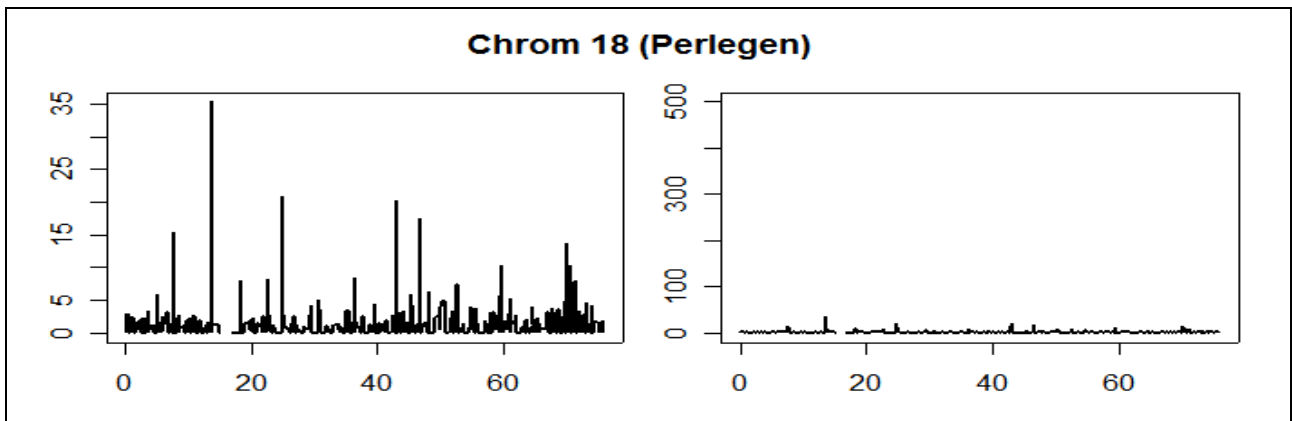
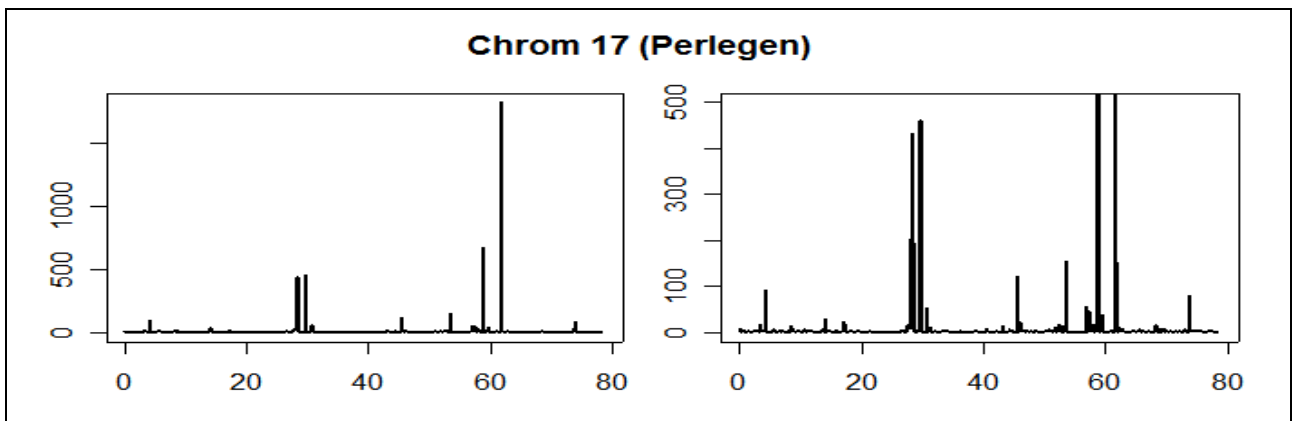
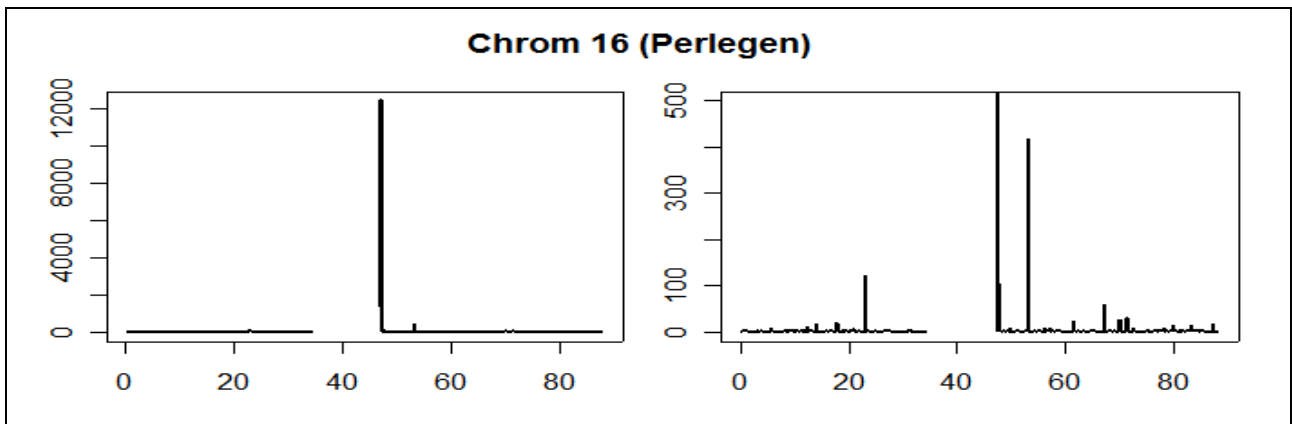
See legend for Figure S4 (above) for details.











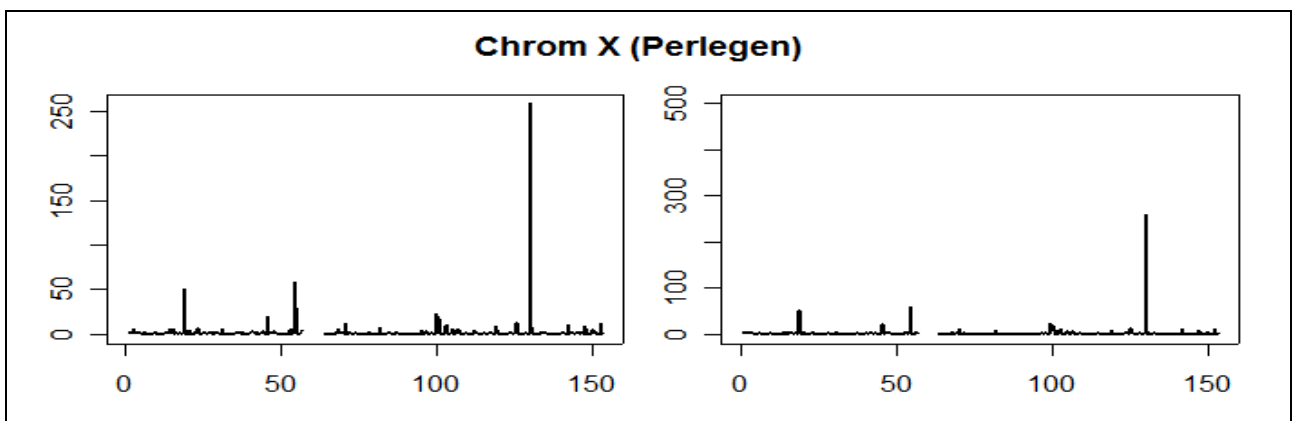
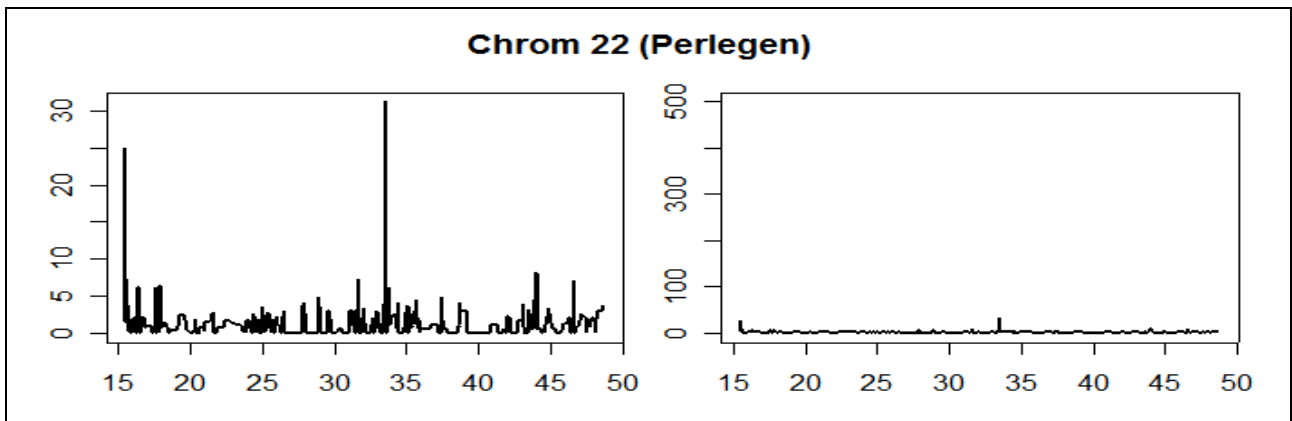
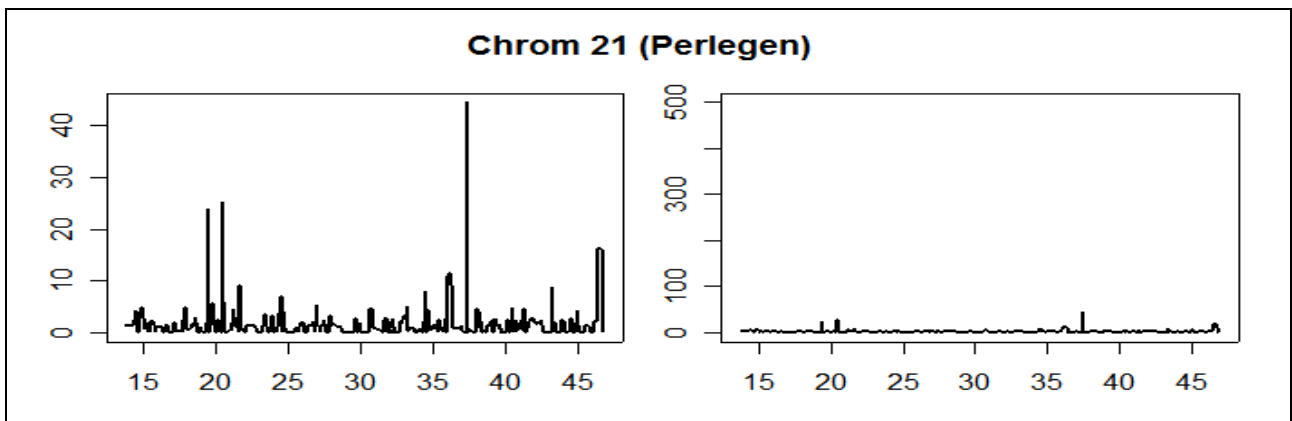
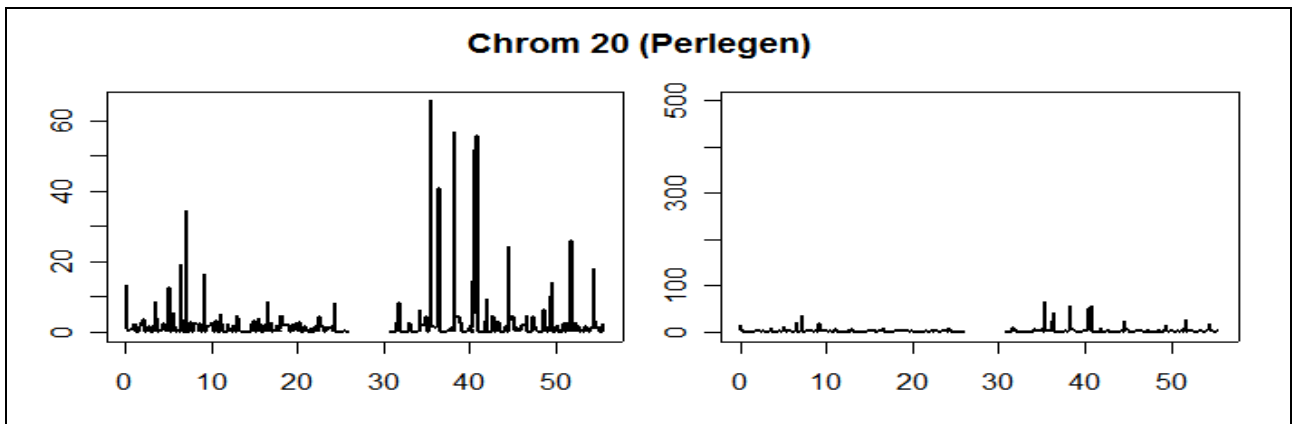


Table S1: Top 100 candidates

The table below provides an ordered list of the top 100 candidates from the scans for selection over the HapMap and Perlegen datasets together (plots of these scans are shown in Fig. S4 and S5). The Ped/Pop value shown in bold corresponds to the higher of the HapMap and Perlegen scan signals at that locus, and thus is the value responsible for placing the locus at that position in the candidate table. Ped/Pop values are only given for both the HapMap and Perlegen scans if both are greater than 20 (which we consider a threshold of interest; see main text). * Generally the physical position of the selection signals in both datasets coincide, but when they differ the position according to the HapMap is given on the left of the forward slash (/) while that according to the Perlegen data is on the right.

Rank	Chrom.	Position (Mb) *	HapMap Ped/Pop	Perlegen Ped/Pop
1	11	28.18 / 28.20	58352	29657
2	X	130.11	15798	259
3	16	47.41 / 47.23	710	12461
4	3	47.93 / 47.95	12353	208
5	17	61.81	275	1825
6	2	187.40	22	1294
7	7	43.46	234	1045
8	1	159.00	377	978
9	17	58.85	197	672
10	13	71.27 / 71.23	228	627
11	17	29.77	508	460
12	14	69.45	462	465
13	2	225.60	463	167
14	17	28.44 / 28.43	181	432
15	4	140.88	175	419
16	16	53.26	114	418
17	4	15.39	406	
18	14	64.95	111	302

19	1	113.60	135	275
20	3	122.12	271	41
21	15	43.52	188	268
22	6	106.75	92	265
23	17	53.61	263	154
24	1	41.10	174	245
25	8	/ 41.85		236
26	1	18.97	228	182
27	13	46.75	228	168
28	3	126.33	225	
29	2	214.59	183	220
30	9	89.37	172	218
31	8	141.90	197	217
32	17	45.60 / 45.59	206	123
33	9	90.73 / 90.81	196	21
34	1	39.11	171	196
35	2	15.61	193	33
36	7	75.82	92	193
37	1	176.05 / 176.04	44	192
38	3	132.17	191	
39	6	170.70	122	172
40	1	91.25	72	170
41	1	94.27 / 94.36	168	56
42	9	80.11 / 80.48	168	35
43	1	62.35	167	
44	7	154.90	161	121
45	16	23.11 / 23.10	158	122
46	6	89.83	157	
47	2	30.79	112	155

48	18	13.69 / 13.71	152	35
49	14	91.81	108	151
50	2	233.24	72	151
51	3	11.45	93	145
52	2	153.28 / 153.30	22	142
53	3	185.95	92	138
54	1	96.70	74	137
55	7	83.84 / 83.85	132	135
56	8	137.59	75	127
57	2	120.77 / 120.78	66	127
58	2	233.86	88	127
59	9	83.64	118	127
60	2	196.84 / 196.87	126	72
61	2	198.60 / 198.56	96	124
62	2	122.90	116	94
63	1	202.35	79	113
64	5	94.94 / 94.95	64	111
65	15	39.66	74	111
66	1	72.81	39	106
67	2	149.01 / 148.96	105	67
68	15	64.43	101	71
69	1	70.04	99	
70	1	171.40 / 171.39	74	98
71	12	88.67	33	98
72	2	202.49 / 202.42	97	28
73	17	4.38 / 4.39	95	92
74	7	57.42 / 57.44	75	94
75	3	44.28	93	76
76	13	/ 77.77		91

77	5	42.29	43	89
78	10	0.23 / 0.24	72	88
79	X	56.95	88	
80	1	235.25	87	62
81	17	57.18 / 56.98	85	56
82	20	48.55	85	
83	17	73.91 / 73.92	51	81
84	14	19.87	80	51
85	5	96.74	79	63
86	1	98.50 / 98.57	78	54
87	15	32.93	76	77
88	1	20.86	37	76
89	7	41.35	68	76
90	13	23.26	76	41
91	16	69.83 / 69.84	76	26
92	2	175.27 / 175.26	75	55
93	13	77.82	75	
94	4	86.59	75	37
95	2	134.34	61	74
96	5	150.81	73	31
97	6	109.99	45	71
98	13	95.01	71	
99	20	51.76	71	26
100	9	100.62	71	62

Table S2: Top 20 candidates in separate populations

The tables below show the top 20 candidates for selection according to the Ped/Pop statistic in the 3 separate worldwide populations (with HCB and JPT grouped). The results show a strong overlap between extreme outliers of the Ped/Pop statistic among the separate populations, but also show a number of candidates that appear to be only extreme outliers in one or two populations.

European Top 20:

Rank	Chrom.	Position (Mb)	Ped/Pop	JPT+HCB top 20?	YRI top 20?	Global Rank (top 100)
1	16	47.41	100 000	Yes	Yes	3
2	8	141.90	88 000	No	No	31
3	14	69.45	70 000	No	Yes	12
4	11	28.18	45 000	Yes	Yes	1
5	17	61.81	45 000	No	Yes	5
6	2	15.61	29 000	Yes	No	35
7	2	225.60	20 000	No	Yes	13
8	X	130.11	18 000	Yes	Yes	2
9	3	122.12	12 000	No	Yes	20
10	4	140.88	7 000	Yes	No	15
11	17	58.85	7 000	No	No	9
12	14	72.6	6 000	No	No	NA
13	3	47.93	4 000	Yes	Yes	4
14	X	70.6	4 000	Yes	No	NA
15	15	43.52	3 000	No	No	21
16	13	71.27	2 900	No	Yes	10
17	11	48	1 000	No	No	NA
18	9	83.64	600	No	No	59
19	4	15.39	550	No	No	17
20	17	29.77	550	Yes	Yes	11

Asian Top 20:

Rank	Chrom.	Position (Mb)	Ped/Pop	CEU top 20?	YRI top 20?	Global Rank (top 100)
1	11	28.18	234 000	Yes	Yes	1
2	16	47.41	45 000	Yes	Yes	3
3	17	30.9	28 000	No	No	NA
4	3	47.93	25 000	Yes	Yes	4
5	X	130.11	18 000	Yes	Yes	2
6	5	94.94	17 000	No	No	64
7	8	85.4	16 000	No	No	NA
8	1	39.11	14 000	No	No	34
9	2	15.61	10 000	Yes	No	35
10	1	92.6	9 000	No	No	NA
11	17	29.77	9 000	Yes	Yes	11
12	17	57.18	8 000	No	Yes	81
13	X	70.6	5 000	Yes	No	NA
14	11	50.1	4 000	No	No	NA
15	4	140.88	2 500	Yes	No	15
16	20	36.4	600	No	No	NA
17	10	95.4	550	No	No	NA
18	X	53.3	500	No	No	NA
19	17	53.61	327	No	Yes	23
20	7	154.90	320	No	No	44

African Top 20:

Rank	Chrom.	Position (Mb)	Ped/Pop	CEU top 20?	JPT+HCB top 20?	Global Rank (top 100)
1	X	130.11	92 000	Yes	Yes	2
2	11	28.18	47 000	Yes	Yes	1
3	2	225.60	40 000	Yes	No	13
4	3	47.93	25 000	Yes	Yes	4
5	3	122.12	12 000	Yes	No	20
6	17	57.18	8 900	No	Yes	81
7	1	41.10	6 800	No	No	24
8	1	159.00	600	No	No	8
9	14	69.45	364	Yes	No	12
10	7	43.46	330	No	No	7
11	2	30.79	295	No	No	47
12	2	214.59	270	No	No	29
13	17	53.61	256	No	Yes	23
14	13	71.27	255	Yes	No	10
15	18	13.69	239	No	No	48
16	16	47.41	228	Yes	Yes	3
17	17	29.77	225	Yes	Yes	11
18	6	106.75	217	No	No	22
19	6	89.83	209	No	No	46
20	17	61.81	200	Yes	No	5

Figure S6. The Ped/Pop density: Separate population scan

The first plot depicts the Ped/Pop densities for the scans over the separate populations of the HapMap data (see legend for Table S2). Black – CEU population; Blue – HCB+JPT population; Red – YRI population. The plot below that shows the upper tail of the densities to highlight the different sizes of the extreme tail that we expect to harbour candidates for recent positive selection.

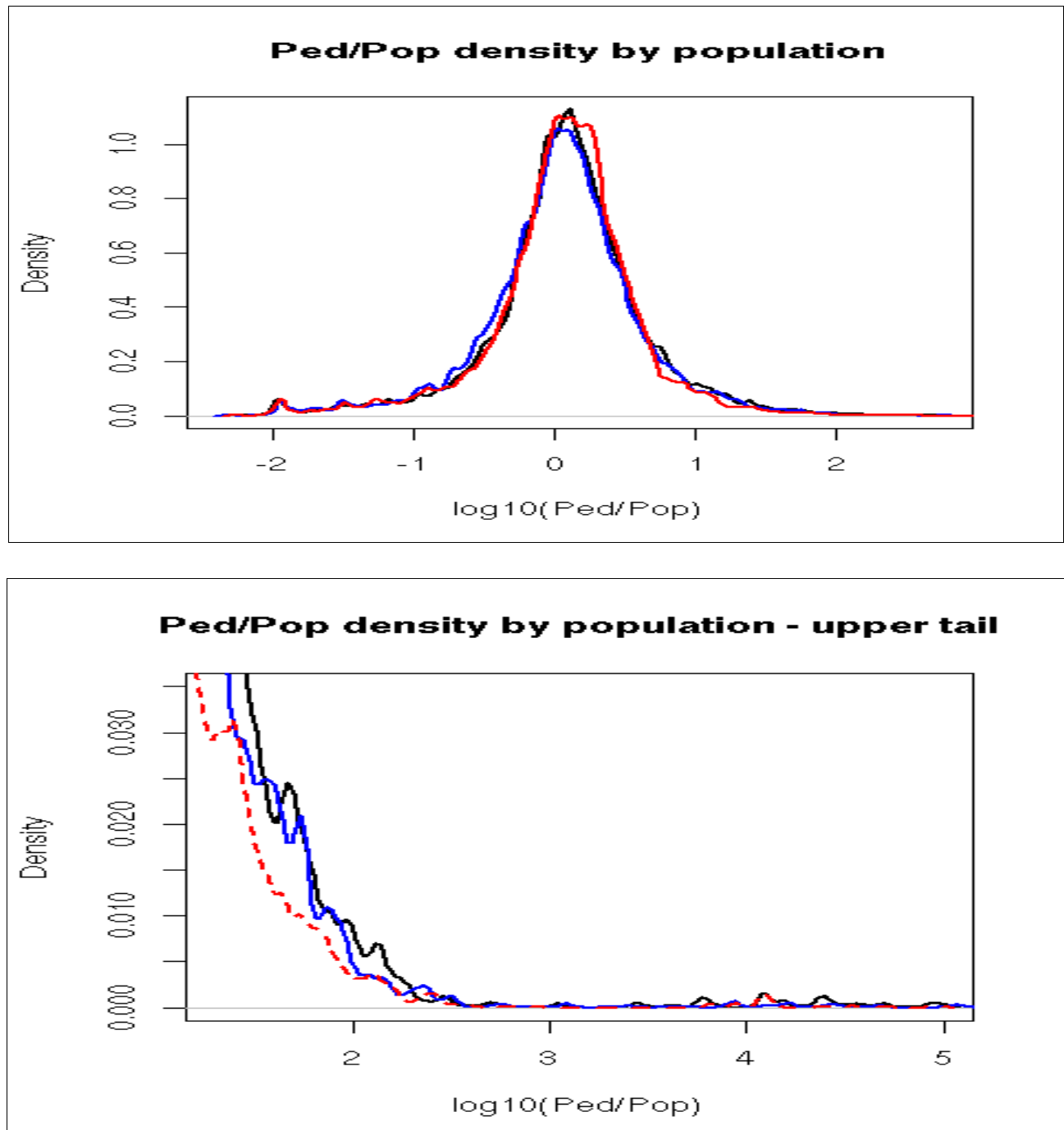
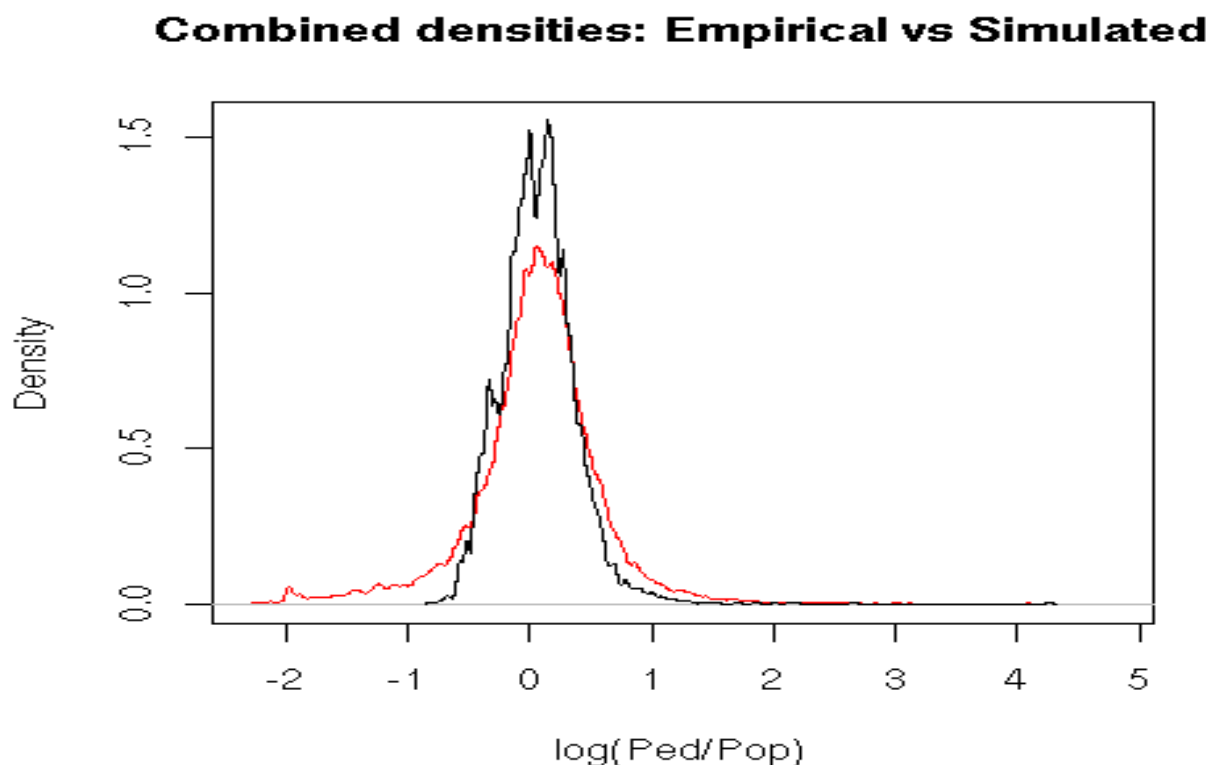


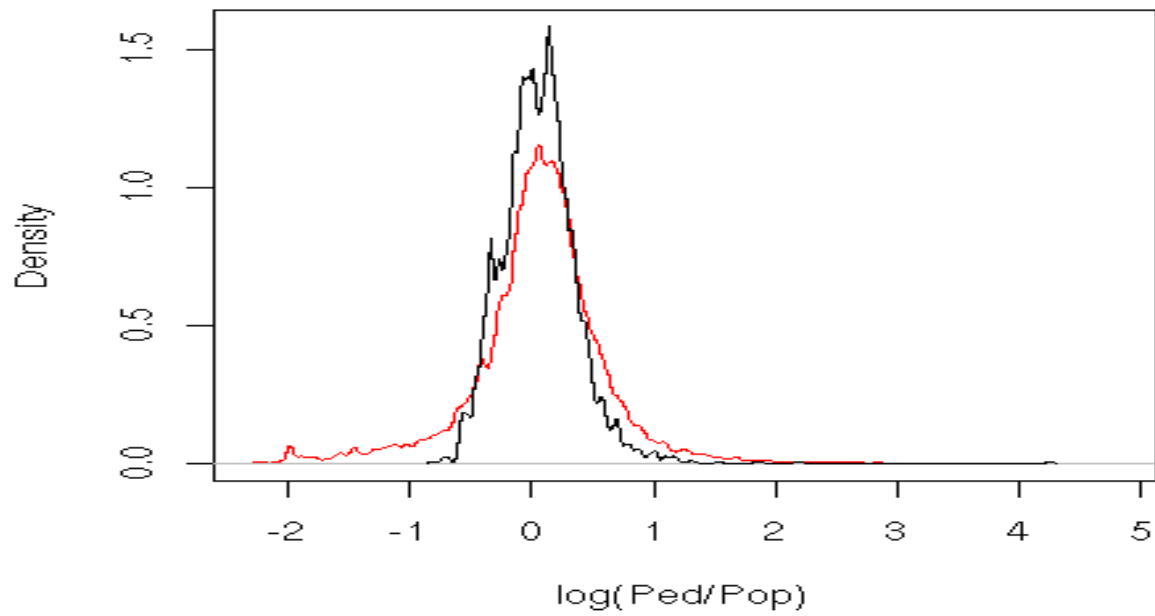
Figure S7. The Ped/Pop Density: Empirical Vs Schaffner Null

Here we show density plots comparing the distribution of the Ped/Pop statistic from the main empirical scans for selection (Fig. S4, S5), in red, with that of the null distribution simulated according to the Schaffner model (Schaffner et al, 2005) described in the main text and Methods section, in black. Using the Schaffner model we are able to simulate HapMap and Perlegen datasets separately. All densities correspond to $\text{Ped/Pop} > 0$. The proportion of 0 counts in the empirical and simulated densities are 42.5% and 44.5% respectively. The marginally larger 0 count in the simulated data suggests that the null distribution is conservative, since more 0 counts correspond to greater variation in pedigree-based recombination estimates overall.

In addition to a comparison of the combined densities (corresponding to HapMap and Perlegen results together), we also depict the densities of Ped/Pop from the separate datasets (see overleaf), which indicate that there is little difference in the power to detect selection between the two sources. Note that the size of the lower tail is exaggerated by the log transform.



HapMap densities: Empirical vs Simulated



Perlegen densities: Empirical vs Simulated

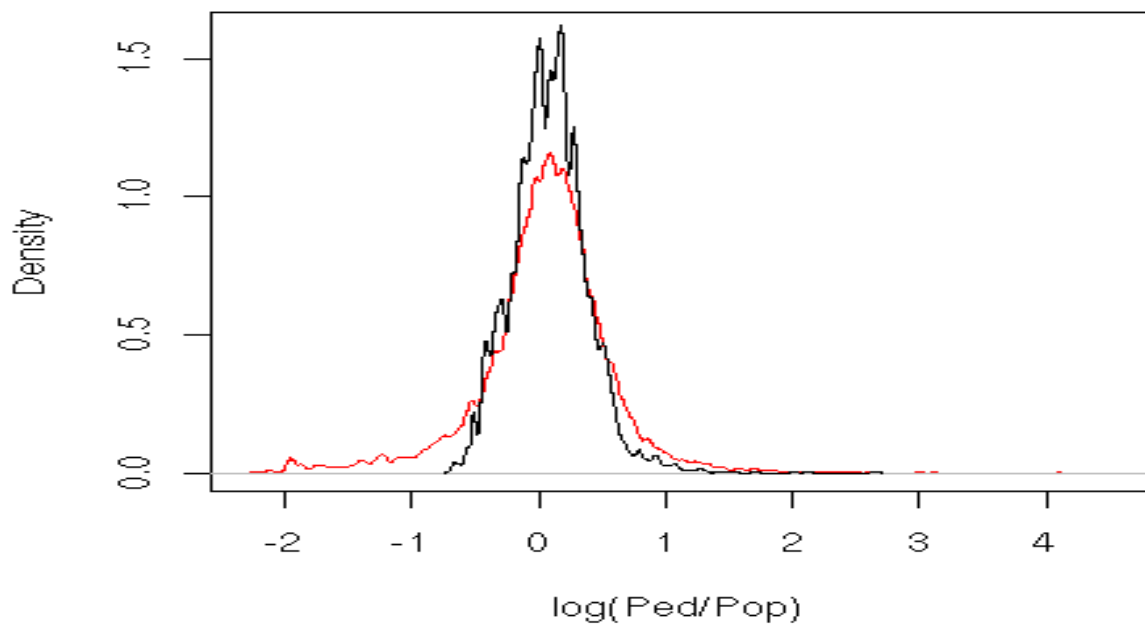


Figure S8. The CLR statistic under the Schaffner Model

Our simulations suggest that extreme values of the CLR statistic are often highly localised, meaning that exploring the robustness of the method under different demographic scenarios by inspecting the full densities, as in Williamson et al. (2007), can obscure differences in the extreme tail. The plot below illustrates the distinct difference between full densities of the statistic (in red), calculated every 2 kb for gridsize=50, for a set of 100 kb sequences and densities corresponding to maximum values calculated for each 100 kb region (in black), which are relevant here since extreme values of the statistic are considered potential candidates irrespective of neighbouring values.

With the behaviour in the tail of the distribution highlighted by using maxima (in black) rather than the full densities (in red) it is visible that the 'Schaffner null' distribution of the CLR statistic has a heavier tail in Asians (main) than Europeans (insert). Using the smallest value of CLR from the selection candidate list in Williamson et al. (2007) as an approximate threshold of interest, there are in fact > 50% more extreme CLR values in Asians than Europeans. It should be noted that these densities are not directly comparable to those relevant to the study of Williamson et al. (2007) since we use all SNPs in our analysis, while that study used ascertained SNP data which effects the null distribution of the statistic.

