Supplementary information

Supplementary text on deleterious variants - Supplement with additional information regarding the identification, distribution and effects of the deleterious mutations.

Supplementary figure 1A

Supplementary figure 1B

Supplementary figure 2

Supplementary figure 3

Supplementary figure 4

Supplementary figure 5

Supplementary figure 6

Supplementary figure 7

Supplementary figure 8

Supplementary figure 9

Supplementary table 1

Supplementary table 2

Supplementary information on deleterious mutations

Fitness of the individuals during the simulations was assessed based on the frequency of deleterious variants. For this purpose, we used the Ensembl Variant Effect Predictor tool v.74 on the filtered VCF files to assess the nature of the variants (McLaren et al 2010). Variants with a SIFT score close to 0 that were identified as putatively deleterious were extracted (Kumar et al 2009). Two classes of deleterious variants within the populations were found: Homozygous, high-frequency (nearly) fixed deleterious variants, and low-frequency heterozygous deleterious variants. Even though the Pietrain pigs have higher heterozygosity levels throughout their genomes (see main text), the number of deleterious heterozygous variants per individual in *Sus cebifrons* is higher. The total number of homozygous deleterious variants per individual was much higher as well in *Sus cebifrons* than in Pietrain (figure S3).

However, it is acknowledged that the SIFT score of the VEP is sensitive to the alleles that are present in the reference genome in such a way that these alleles are less likely to be predicted as deleterious. The functional predictions for sites where the genome reference carries the derived allele are often classified as benign, even though the overall population frequency would suggest otherwise. This phenomenon has been observed in human and maize as well (Simons 2014, Mezmouk and Ross-Ibarra 2014). In addition, the number of identified deleterious sites is highly dependent on the sequencing depth. The higher the depth, the more putative deleterious sites can be identified. In our analysis, all individuals are sequenced to a depth of ~10x, but we do acknowledge that we have not identified all deleterious alleles because of the unequal distribution of the sequencing depth. Because of these two biases we should be cautious with drawing conclusions on the number of identified deleterious variants directly from the data.

Shared deleterious alleles

Given the total number of unique deleterious homozygous sites in *Sus cebifrons*, most of the deleterious variants are fixed (table S1). This could be the result of past bottlenecks in the *Sus cebifrons* population as observed in the demographic analysis. We are aware of the bias of the predictors and use only the low-frequency deleterious alleles that were observed once or twice in heterozygous state, because this distribution of low-frequency small-effect deleterious alleles fits the Mukai-scenario best that we have used for our data (Mukai et al, 1972). We expect that the technology to identify deleterious mutations in the genome develops quickly. With better predictors not only the mutations causing an undesired effect, but also the magnitude of their effect can be implemented in our management simulations.

For the Pietrain population, the list of deleterious sites was based on the deleterious sites that were actually called within 11 individuals; on average each individual contained 656 deleterious sites. Then the average number of shared deleterious sites between 2,3,4, … individuals was obtained, to be able to infer how many unique deleterious sites each individual would contribute to the total. The actual number of observed deleterious sites among the 11 individuals (=3468) was extrapolated by fitting a power curve to the number of unique contributions per extra individual, so that we expected with 48 individuals a total of ~10.000 deleterious sites (see figure S7). Then, 6532 more sites were randomly extracted from the genome and we assigned a deleterious effect to these sites. Then, for each individual 656 sites were randomly picked from this list, so that to some degree it fitted the observed distribution in the 11 re-sequenced individuals. These 10.000 markers and genotypes for 48 Pietrain pigs were added to the matrix containing the 60K markers for the management. For comparison: If we fit a power curve to the deleterious heterozygous sites in the S*us cebifrons* population we end up with an average of 13504 unique deleterious sites with 48 individuals (see figure S8).

On the distribution of mutational effects

The deleterious mutations were assumed to have mutational effects which followed a Gamma distribution with shape parameter 1 and mean effect 0.005. The Gamma distribution is a good choice as we can represent with it two extreme cases like mutations of large effect and mutations of small effect, by changing the shape of the distribution. A Gamma distribution with shape 1 is an exponential distribution. We chose it as it is linked with scenarios of many mutations of small effect (Caballero & Keightley, 1994; Fernandez & Caballero, 2001; de Cara et al, 2013) which lead to the accumulation of such mutations, as they are not easily removed by natural selection (Mukai et al, 1972).

The deleterious mutations in our study had mean effect 0.005 and mean dominance 0.35 or 0.5. In our Fortran code, we used the subroutine "random gamma" to draw the effect of each marker from a Gamma distribution with such mean and shape 1. We then followed the study by Caballero & Keightley (1994) whereby the dominance of each marker is a random number between 0 and $\exp(-k\ s_i)$, where k is a normalizing constant used to obtain the desired mean dominance and s_i is the effect of marker i.

Fitness of each individual is then $\prod w_i$where $w_i$is each marker fitness and is 1 for homozygous wild, for a heterozygous locus and for a homozygous deleterious locus. In figure S9, we show the theoretical distribution of effects and the actual distribution for one replicate in the *Sus cebifrons* population.

References

Caballero A, Keightley P (1994) A pleiotropic nonadditive model of variation in quantitative traits. Genetics, 138, 883–900.

de Cara MAR, Fernandez J, Toro MA, Villanueva B (2013) Purging deleterious mutations in conservation programmes: combining optimal contributions with inbred matings. Heredity, 110, 530–537.

Fernandez J, Caballero A (2001) Accumulation of deleterious mutations and equalization of parental contributions in the conservation of genetic resources. Heredity, 86, 480–488.

Mukai T, Chigusa SI, Mettler LE, Crow JF (1972) Mutation rate and dominance of genes affecting viability in drosophila melanogaster. Genetics, 72, 333–355.

McLaren W., Pritchard B., Rios D., Chen Y., Flicek P., Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 2010;26:2069-2070.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81

Simons, Y.B., Turchin, M.C., Pritchard, J.K. and Sella G. 2014. The deleterious mutation load is insensitive to recent population history. Nature Genetics. 46: 220-224

Mezmouk S, Ross-Ibarra J (2014) The pattern and distribution of deleterious mutations in maize. G3 4:163-171

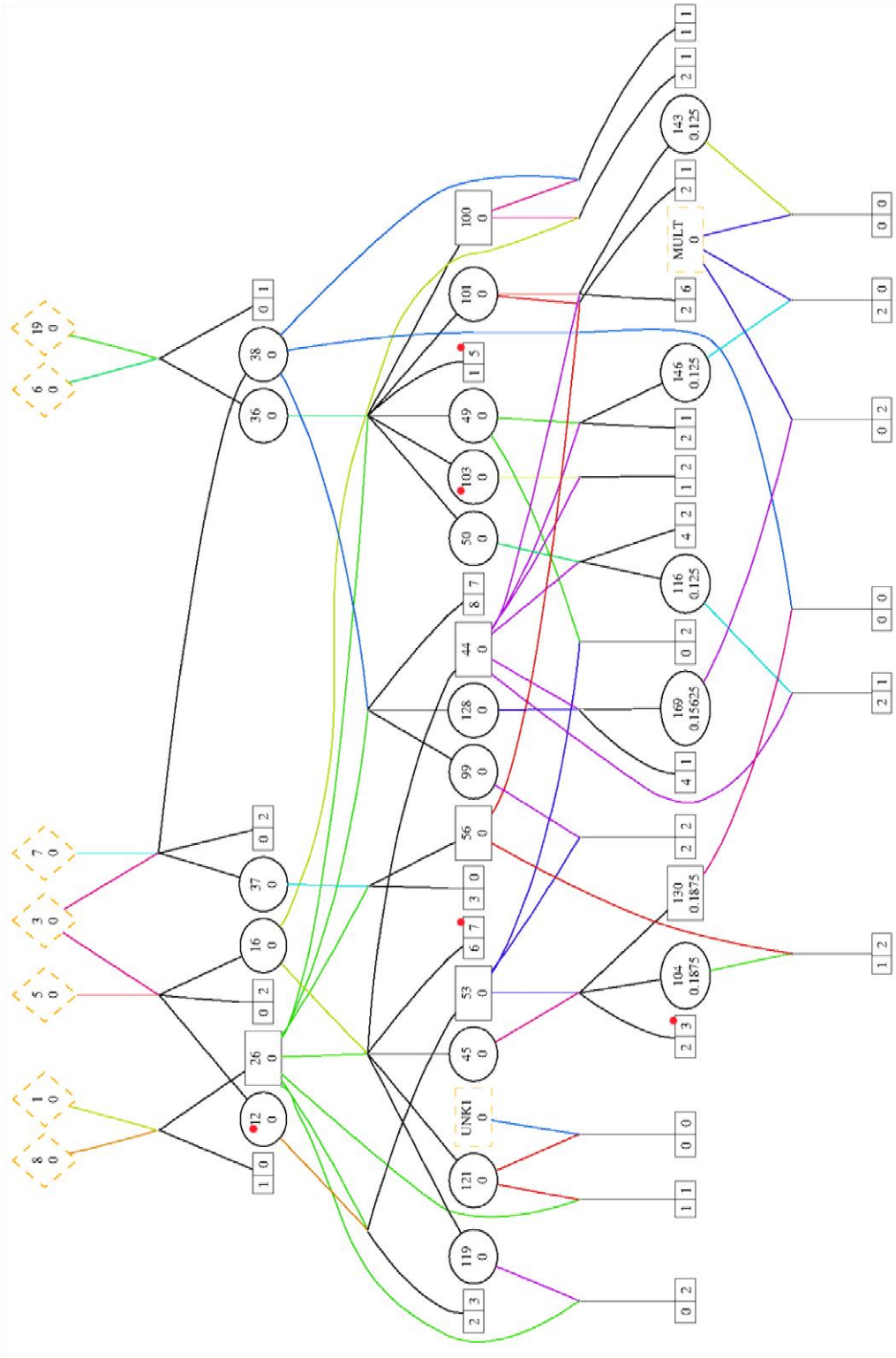


Figure S1A Relationship between the S. cebifrons individuals. S1A. Pedigree of the captive zoo population that the 5 re-sequenced individuals are sampled from. Inbreeding coefficients for breeding individuals are displayed below the number of the individual, and the number of non-breeding offspring from a particular breeding couple is shown within squared boxes. Sampled individuals are indicated with a red dot.

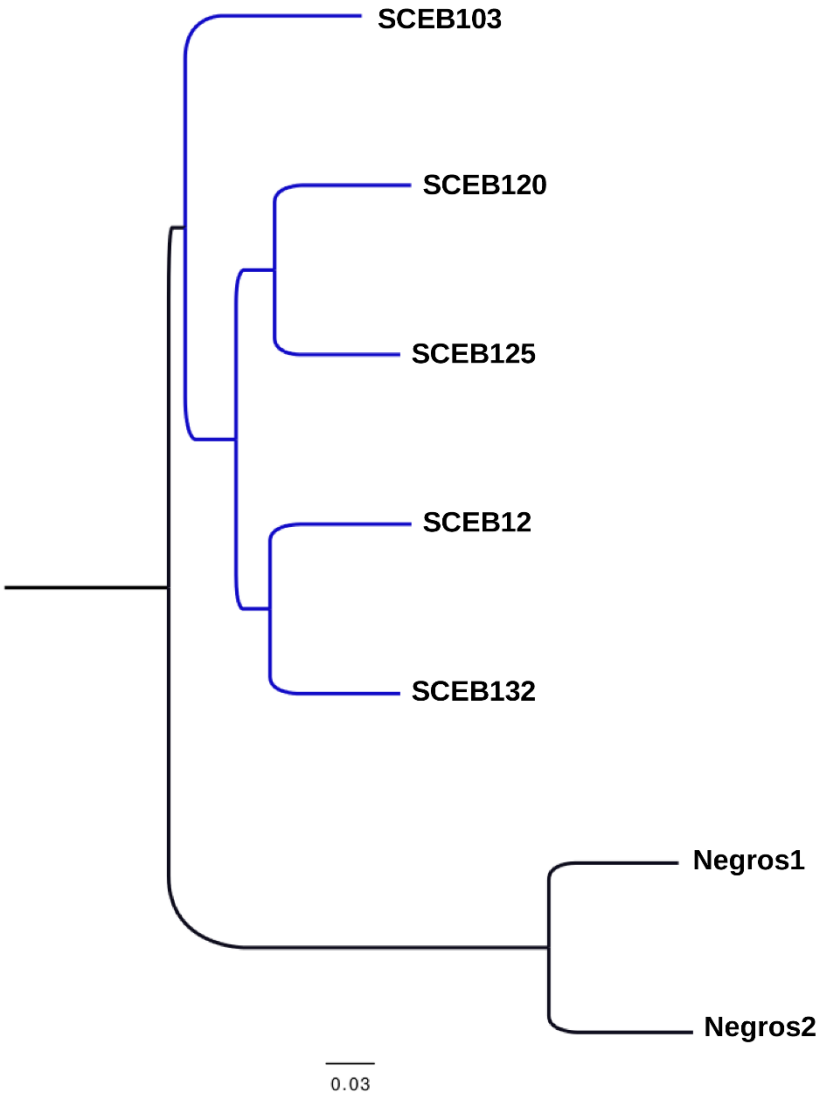


Figure S1B Relationship between the S. cebifrons individuals. S1B. Neighbor-joining phylogenetic tree of the 7 re-sequenced S. cebifrons individuals. Individuals highlighted in blue are from the San Diego Zoo and are used for the in silico management.

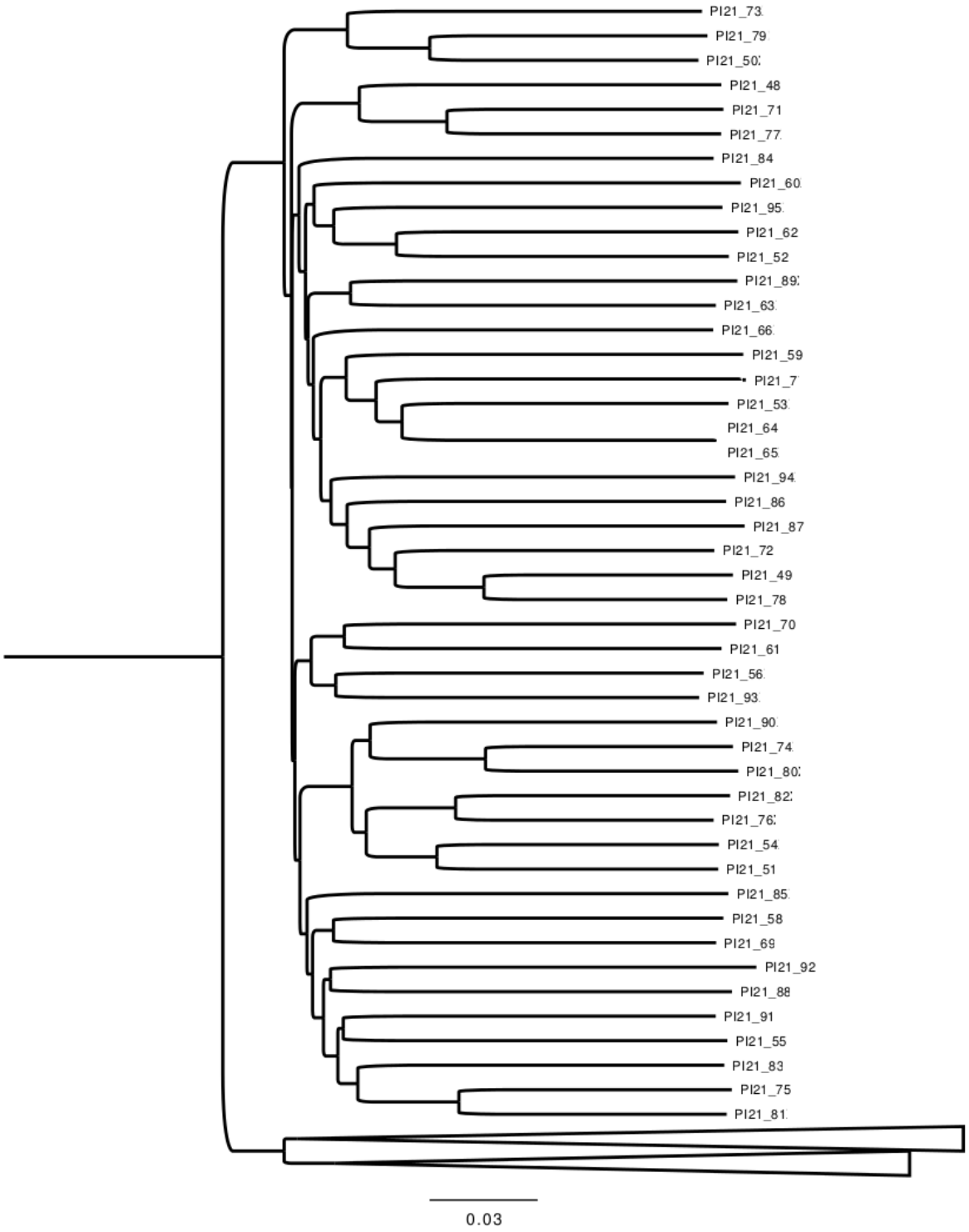


Figure S2 Relationship between all Pietrain individuals. Neighbor-joining phylogenetic tree of all Pietrain individuals that were used for the in silico management. We used Large White and Landrace pigs as outgroup.

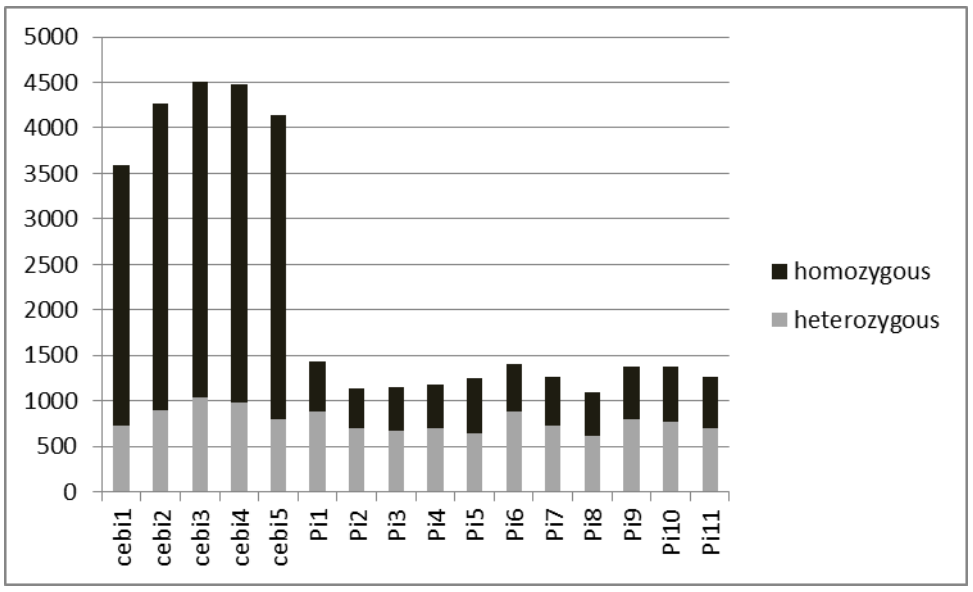


Figure S3: Number of heterozygous and homozygous variants that are predicted to be deleterious for each re-sequenced individual.

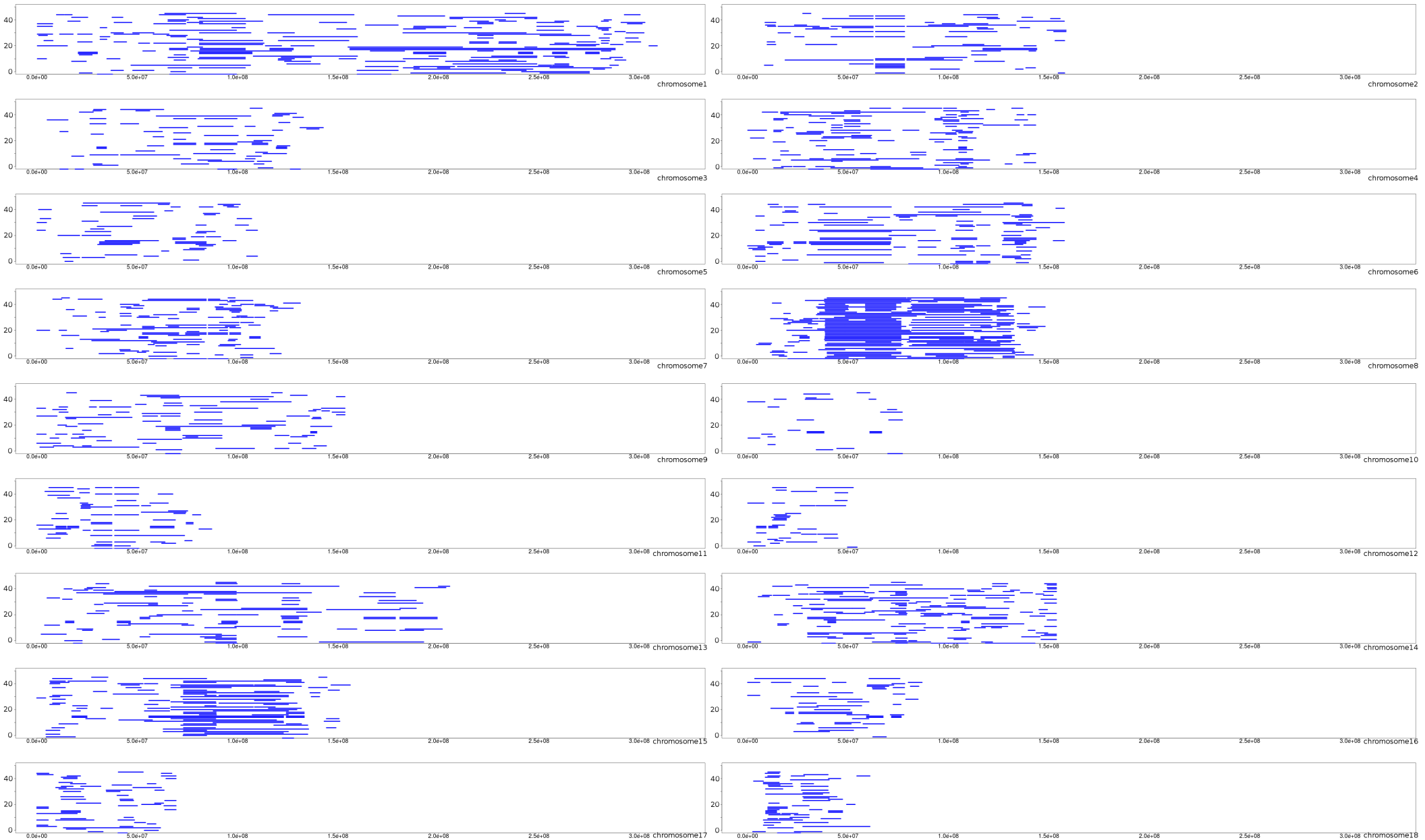

0
20
40
0.0e+00
5.0e+07
1.0e+08
1.5e+08
2.0e+08
2.5e+08
3.0e+08
chromosome1
chromosome2
chromosome3
chromosome4
chromosome5
chromosome6
chromosome7
chromosome8
chromosome9
chromosome10
chromosome11
chromosome12
chromosome13
chromosome14
chromosome15
chromosome16
chromosome17
chromosome18

Figure S4. Runs of homozygosity in the Pietrain population. ROHs within individual genomes in the Pietrain population before the management are displayed per chromosome. The x-axis displays the full length of each chromosome in bp and individuals are listed on the y-axis so that each line represents the genome of one individual. ROHs are indicated in blue.

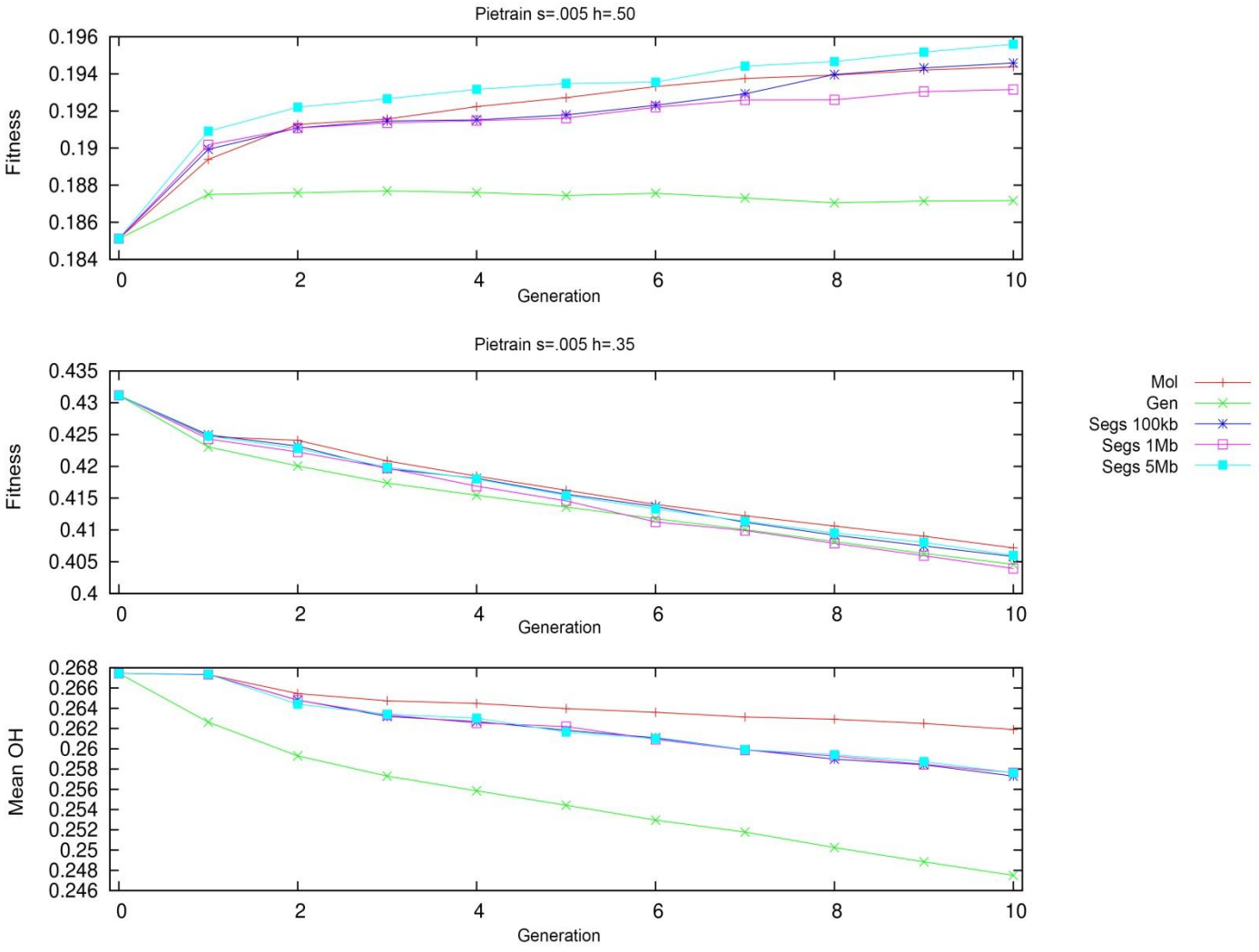
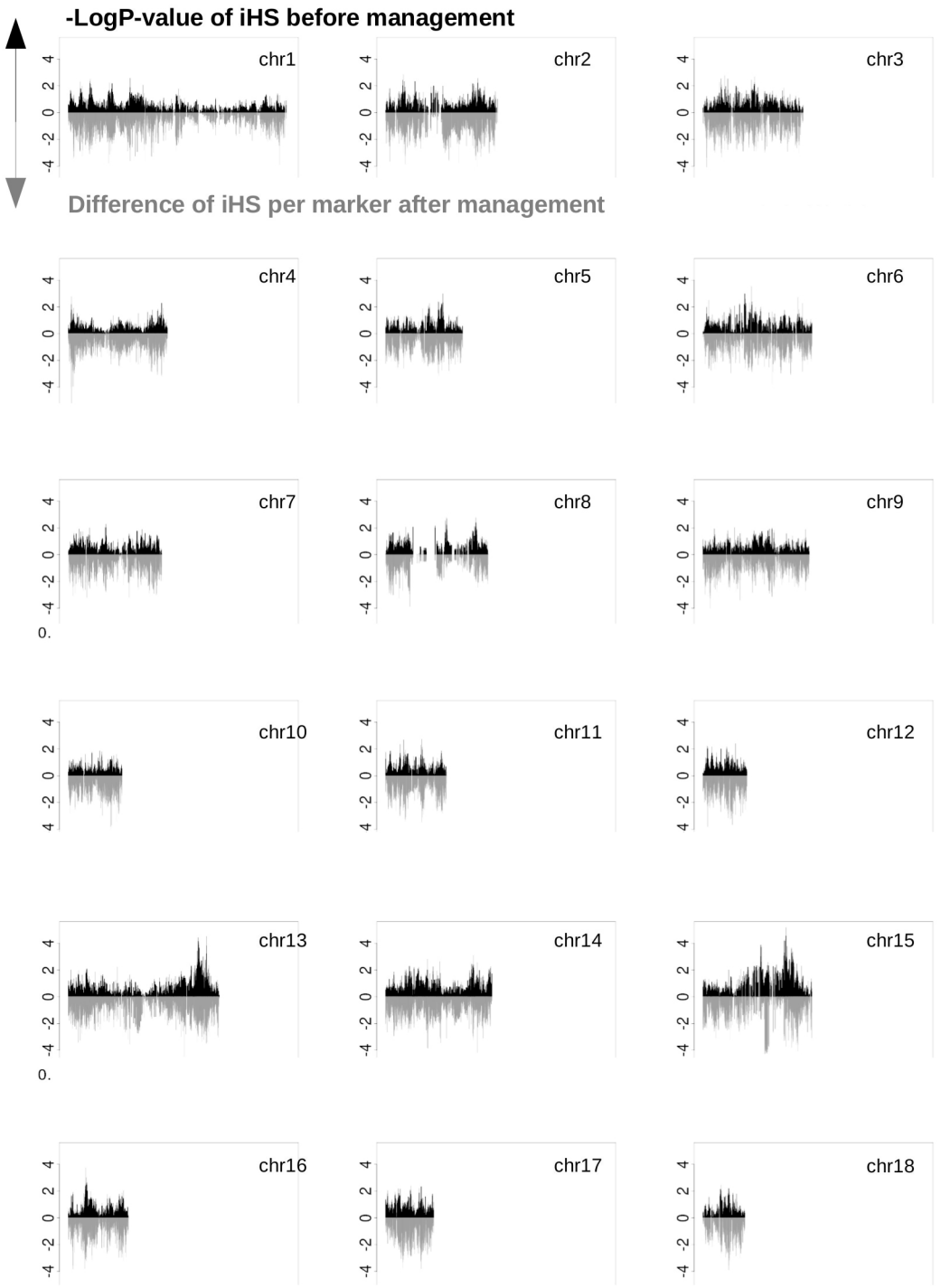
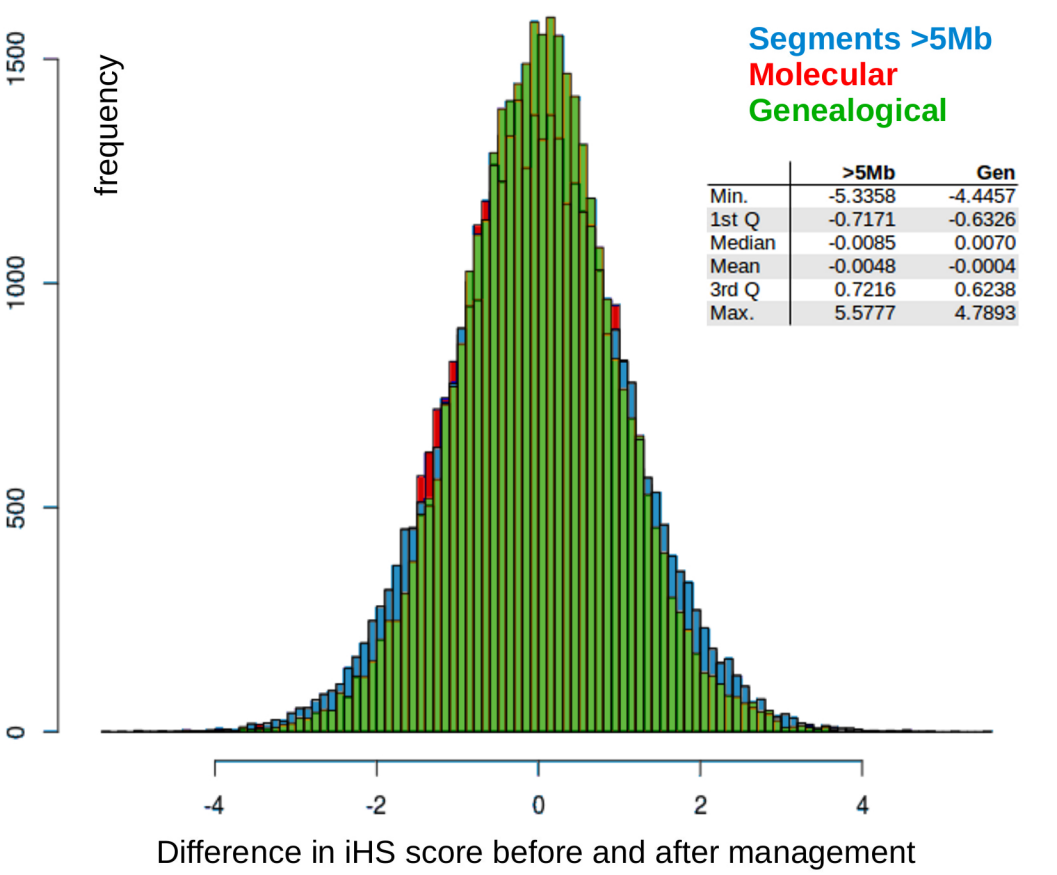


Figure S5 Fitness and diversity during management of the Pietrain population. The change in fitness and observed heterozygosity (OH) during 10 generations of management is displayed for 5 different management strategies. Fitness change over 10 generations of management when a dominance coefficient of 0.5 and selection coefficient of 0.005 is applied for the top panel, and fitness change over 10 generations of management when a dominance coefficient of 0.35 and selection coefficient of 0.005 is applied in the second panel. The lowest panel displays the observed heterozygosity during 10 generations of management.

A



B



	>5Mb	Gen
Min.	-5.3358	-4.4457
1st Q	-0.7171	-0.6326
Median	-0.0085	0.0070
Mean	-0.0048	-0.0004
3rd Q	0.7216	0.6238
Max.	5.5777	4.7893

Figure S6. Effect of management on selective sweeps in the Pietrain population. Signatures of selection are measured as extended haplotype homozygosity (iHS signal) in the Pietrain population before and after management. S6A Genome-wide view of the correlation between the p-value of the iHS signal before management, and the magnitude of difference between iHS signal before and after management. The physical location on the chromosome in bp for each marker is indicated on the x-axis. The significance levels of the iHS signal before management are indicated in black, and displayed on the y-axis (-logp) and range between 0 and 4, so that markers with a signal >2 are considered to be significant. Differences in iHS signal per marker are indicated in grey and range from 0 to -5, with a strong negative number indicating a large difference. S6B displays a histogram of the difference in iHS signal for each marker before and after 5Mb segment-based, molecular-based and genealogical-based management. The table displays numerical details about the distributions.

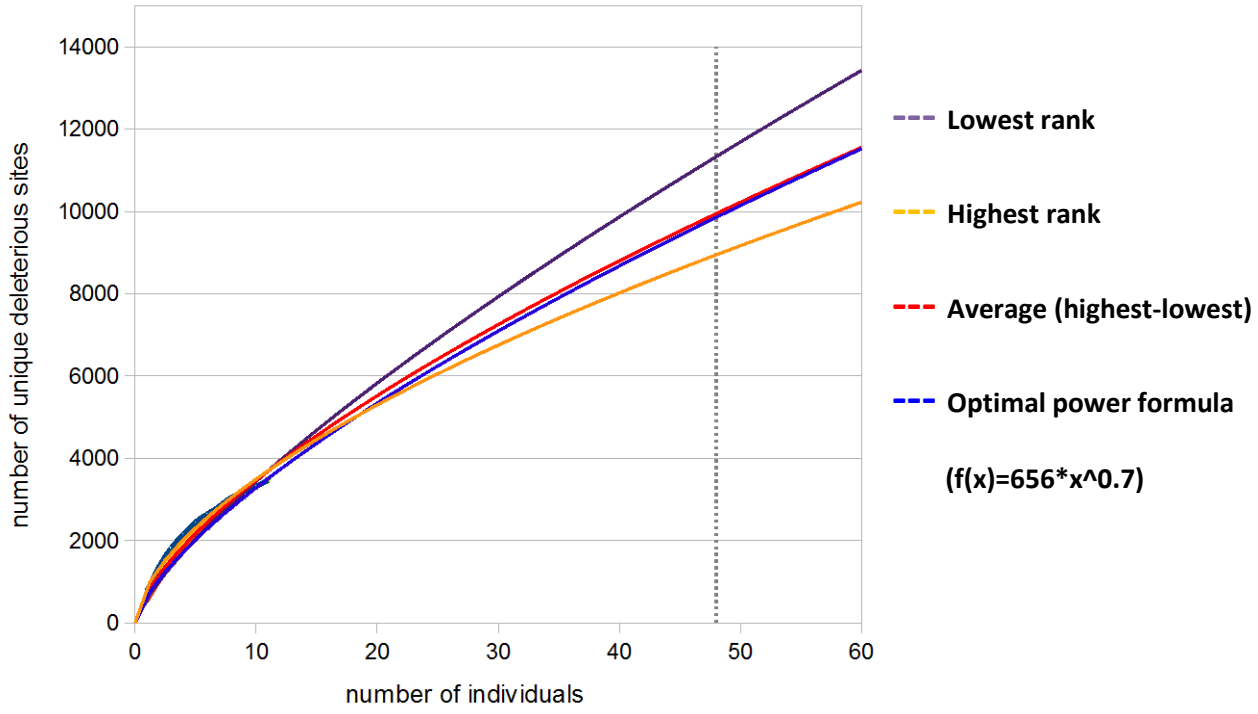


Figure S7. Power curve fitted to the number of unique deleterious sites in the Pietrain population. The curves are based on the deleterious sites that were identified in the 11 re-sequenced individuals. "Lowest rank" corresponds to the best fit when the individuals are listed from lowest to highest number of identified deleterious sites (1..11). "Highest rank" corresponds to the curve fitted to the list of individuals when starting with the individual that has the most identified deleterious sites. "Average" is the curve fitted to the average between the highest and the lowest curve, with the optimal power formula as displayed in blue. Based on this formula, we end up with ~10.000 unique deleterious sites in a population of 48 Pietrain pigs.

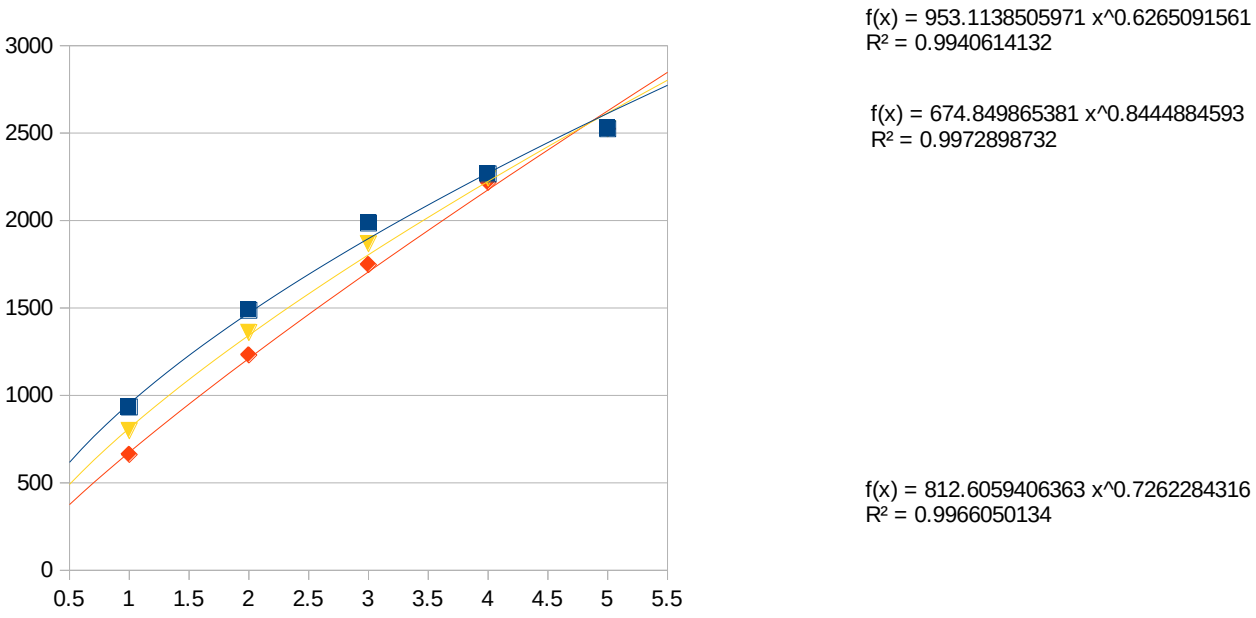


Figure S8. Power curve fitted to the number of unique deleterious sites in the the *Sus cebifrons* population. The curves are based on the deleterious sites that were identified in the 5 re-sequenced *Sus cebifrons* individuals from Panay.

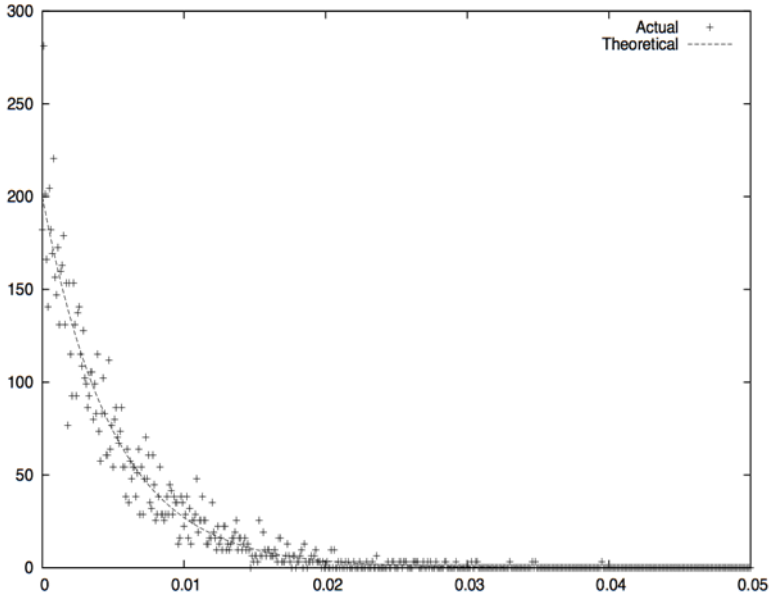


Figure S9: Theoretical Gamma distribution with shape 1 and mean 0.005 and actual distribution from one replicate used in the simulations for the *Sus cebifrons* population with those parameters.

Table S1: Total and unique number of deleterious variants. The *Sus cebifrons* group consists of 5 individuals, and the Pietrain group consists of 11 individuals. T means the total number of identified deleterious sites, and U corresponds to the total number of unique identified deleterious sites in the group.

	Cebifrons	***Pietrain***
Heterzygous T	4430	8082
Heterozygous U	2525	3468
Homozygous T	16554	5285
Homozygous U	4500	1722

Cebifrons

Biological Process	REFLIST	observed	expected	over/under	P-value
RNA metabolic process	2008	190	252.93	-	0.00187
cell adhesion	735	131	92.58	+	0.017
biological adhesion	735	131	92.58	+	0.017
nucleobase-containing compound metabolic process	2982	311	375.62	-	0.0257
primary metabolic process	6302	711	793.82	-	0.0301
transcription, DNA-dependent	1579	151	198.9	-	0.0325
metabolic process	7531	864	948.63	-	0.0365
regulation of biological process	1760	172	221.7	-	0.0399
transcription from RNA polymerase II promoter	1569	151	197.64	-	0.046
response to external stimulus	201	44	25.32	+	0.11
regulation of transcription from RNA polymerase II promoter	1222	117	153.93	-	0.201
fertilization	89	23	11.21	+	0.321
anatomical structure morphogenesis	601	103	75.7	+	0.343
regulation of nucleobase-containing compound metabolic process	1319	130	166.15	-	0.376
sensory perception of sound	64	18	8.06	+	0.423
regulation of catalytic activity	983	157	123.82	+	0.434
cellular component morphogenesis	560	96	70.54	+	0.486
regulation of molecular function	993	158	125.08	+	0.486
blood coagulation	162	35	20.41	+	0.488
cell-matrix adhesion	104	25	13.1	+	0.533
Unclassified	6756	917	851.01	+	0.553
lipid transport	264	51	33.25	+	0.588
regulation of translation	119	5	14.99	-	0.687
nucleobase-containing compound transport	104	4	13.1	-	0.853
defense response to bacterium	24	9	3.02	+	0.991

Pietrain

Biological Process	REFLIST	observed	expected	over/under	P-value
transcription from RNA polymerase II promoter	1579	125	174.1	-	0.00693
transcription, DNA-dependent	1589	126	175.2	-	0.00705
regulation of biological process	1786	145	196.92	-	0.00724
primary metabolic process	6371	624	702.46	-	0.0272
RNA metabolic process	2019	172	222.61	-	0.0284
nucleobase-containing compound metabolic process	3003	271	331.11	-	0.0293
metabolic process	7612	759	839.29	-	0.0326
skeletal system development	301	14	33.19	-	0.0331
Unclassified	6802	823	749.98	+	0.102
regulation of transcription from RNA polymerase II promoter	1229	99	135.51	-	0.104
regulation of nucleobase-containing compound metabolic process	1330	109	146.64	-	0.115
pattern specification process	210	9	23.15	-	0.173
antigen processing and presentation	58	16	6.4	+	0.241
segment specification	146	5	16.1	-	0.312
cellular component morphogenesis	576	88	63.51	+	0.441
anatomical structure morphogenesis	617	93	68.03	+	0.489
cell adhesion	770	112	84.9	+	0.57
biological adhesion	770	112	84.9	+	0.57

Cebifrons

Molecular Function	REFLIST	observed	expected	over/under	P-value
nucleic acid binding	2812	285	354.21	-	0.00346
binding	5102	568	642.66	-	0.036
Unclassified	7852	1071	989.06	+	0.0419
DNA binding	1666	165	209.85	-	0.0722
RNA binding	533	42	67.14	-	0.0927
GTPase activity	230	13	28.97	-	0.112
motor activity	85	22	10.71	+	0.254
translation factor activity, nucleic acid binding	122	5	15.37	-	0.337
nucleic acid binding transcription factor activity	1410	142	177.61	-	0.378
enzyme regulator activity	956	152	120.42	+	0.386
translation initiation factor activity	94	3	11.84	-	0.407
structural constituent of cytoskeleton	711	117	89.56	+	0.412
RNA helicase activity	79	2	9.95	-	0.455
sequence-specific DNA binding transcription factor activity	1401	142	176.47	-	0.492
translation regulator activity	117	5	14.74	-	0.524
lipid transporter activity	63	17	7.94	+	0.54
pyrophosphatase activity	213	42	26.83	+	0.611
serine-type peptidase activity	240	46	30.23	+	0.683

Pietrain

Molecular Function	REFLIST	observed	expected	over/under	P-value
DNA binding	1666	122	185.88	-	0.000019
nucleic acid binding	2812	240	313.74	-	0.000251
nucleic acid binding transcription factor activity	1410	107	157.31	-	0.00104
sequence-specific DNA binding transcription factor activity	1401	107	156.31	-	0.00149
Unclassified	7852	953	876.05	+	0.0433
receptor activity	1340	191	149.5	+	0.0606
binding	5102	507	569.23	-	0.155
serine-type endopeptidase inhibitor activity	82	19	9.15	+	0.451
structural molecule activity	1088	152	121.39	+	0.498
peptidase inhibitor activity	171	32	19.08	+	0.65
voltage-gated ion channel activity	137	6	15.29	-	0.995

Pietrain

PANTHER Protein Class	REFLIST	observed	expected	over/under	P-value
defense/immunity protein	457	93	50.99	+	0.00000943
nucleic acid binding	2248	185	250.81	-	0.000396
transcription factor	1427	109	159.21	-	0.00138
homeobox transcription factor	189	6	21.09	-	0.0197
helix-turn-helix transcription factor	189	6	21.09	-	0.0197
cell adhesion molecule	445	77	49.65	+	0.0285
DNA binding protein	768	55	85.69	-	0.0363
Unclassified	7069	864	788.69	+	0.0552
immunoglobulin receptor superfamily	127	29	14.17	+	0.0614
receptor	1354	192	151.07	+	0.0843
small GTPase	120	3	13.39	-	0.135
cytokine receptor	209	40	23.32	+	0.175
histone	53	0	5.91	-	0.482
extracellular matrix linker protein	21	8	2.34	+	0.517
antibacterial response protein	102	22	11.38	+	0.585
protease inhibitor	171	32	19.08	+	0.731
voltage-gated ion channel	142	6	15.84	-	0.778

Cebifrons

PANTHER Protein Class	REFLIST	observed	expected	over/under	P-value
cell adhesion molecule	445	96	56.05	+	0.0000969
nucleic acid binding	2248	213	283.17	-	0.000384
small GTPase	120	2	15.12	-	0.00615
extracellular matrix glycoprotein	115	30	14.49	+	0.0407
G-protein	197	9	24.81	-	0.0425
RNA binding protein	879	79	110.72	-	0.131
intermediate filament	69	19	8.69	+	0.292
transcription factor	1427	143	179.75	-	0.332
Unclassified	7069	958	890.43	+	0.343
cytoskeletal protein	727	120	91.58	+	0.373
homeobox transcription factor	189	11	23.81	-	0.488
helix-turn-helix transcription factor	189	11	23.81	-	0.488
RNA helicase	79	2	9.95	-	0.511
actin binding motor protein	42	13	5.29	+	0.576
extracellular matrix protein	394	70	49.63	+	0.599
ATP-binding cassette (ABC) transporter	58	16	7.31	+	0.64
DNA binding protein	768	72	96.74	-	0.801

Table S2. GO-enrichment analysis of genes containing deleterious variants. List of gene-ontology enrichment terms for Biological Process, Molecular Function and Protein Class. Those GO-terms are listed that were over- or under-represented in the list of genes that contained deleterious variants in the 5 re-sequenced S. cebifrons pigs and in the 11 re-sequenced Pietrain pigs. Shaded in grey are those GO-terms significant with $p<0.05$.