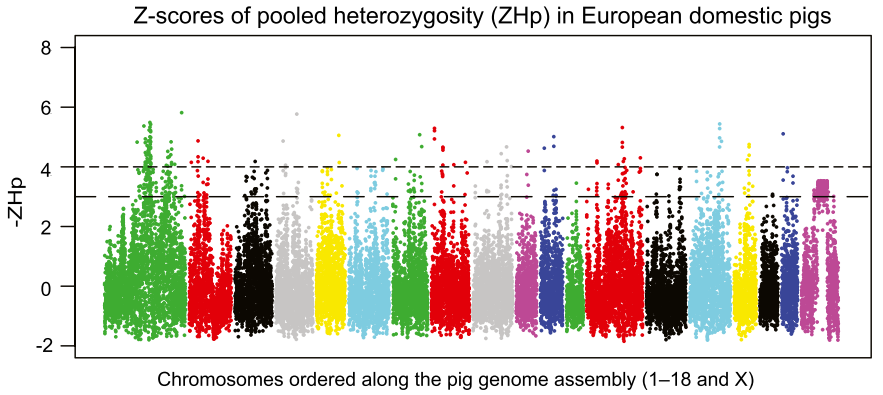
# Supporting Information

## Rubin et al. 10.1073/pnas.1217149109




| Chromosome 1 | Chromosome 1 (cont.) | Chromosome 2 | Chromosome 5 | Chromosome 9 | Chromosome 15 |
|---|---|---|---|---|---|
| NP_001106154.1 | U6 | NAA40 | DEPDC4 | SC5DL | AOX1 |
| AQP9 | PHLPP1 | RCOR2 | UHRF1BP1L | CD36 | FAM126B |
| FAM169B | ZCCHC2 | MARK2 | ACTR6 | PAPPA2 | ORC2 |
| PGPEP1L | TNFRSF11A | C11ORF84 | SCYL2 | RALGPS2 | NDUFB3 |
| IGF1R | KIAA1468 | SWAP70 | | ANGPTL1 | FLIP-L |
| MAGEL2 | PIGN | SBF2 | Chromosome 7 | VWC2 | CASP10 |
| MKRN3 | CDH20 | SH2D3A | RANBP9 | | CASP8 |
| CHRNA7 | 7SK | TRIP10 | SIRT5 | Chromosome 11 | BZW1 |
| KLF13 | MC4R | EMR1 | GFOD1 | FOXO1A | IDH1 |
| OTUD7A | PMAIP1 | GPR108 | VSX2 | UCHL3 | C2ORF80 |
| NFATC1 | CCBE1 | VAV1 | LIN52 | | |
| ATP9B | LMAN1 | | | Chromosome 13 | Chromosome 16 |
| SALL3 | CPLX4 | Chromosome 4 | Chromosome 8 | NAALADL2 | PIKFYVE |
| TSHZ1 | RAX | ODFP1 | NCAPG | OSTN | MAP1B |
| CNDP2 | SMARCA2 | ODF1 | LCORL | CCDC50 | MRPS27 |
| CNDP1 | VPS13A | KLF10 | CPE | PIGX | FAM196B |
| CYB5 | GNA14 | RRM2B | MSMO1 | CEP19 | CCDC99 |
| CYB5A | CYLC2 | PLAG1 | SC4MOL | LRRC33 | |
| C18ORF63 | PSB7 | MOS | KLHL2 | PAK2 | Chromosome 18 |
| C18ORF55 | PSMB7 | RS20 | TMEM192 | PCYT1A | CCT8L2 |
| FBXO15 | GPR144 | RPS20 | ZNF827 | SEC22A | |
| SOCS6 | SF-1 | snoU54 | | ADCY5 | |
| TMX3 | NR6A1 | CHCHD7 | | RUNX1 | |
| 5S_rRNA | STF1 | LYN | | | |
| SERPINB2 | NR5A1 | | | | |
| VPS4B | ssc-mir-181b | | | | |
| SERPINB5 | NEK6 | | | | |
| BCL2 | | | | | |
| KDSR | | | | | |
| SERPINB7 | | | | | |

**Fig. S1.** Genes located within or in the vicinity of the candidate selective sweeps identified. Genes residing within ±20 kb of putative selective sweeps (at least one window with ZHp ≥ 4) are presented for each chromosome. The genes are ordered according to their location (smallest to largest coordinate on each chromosome).

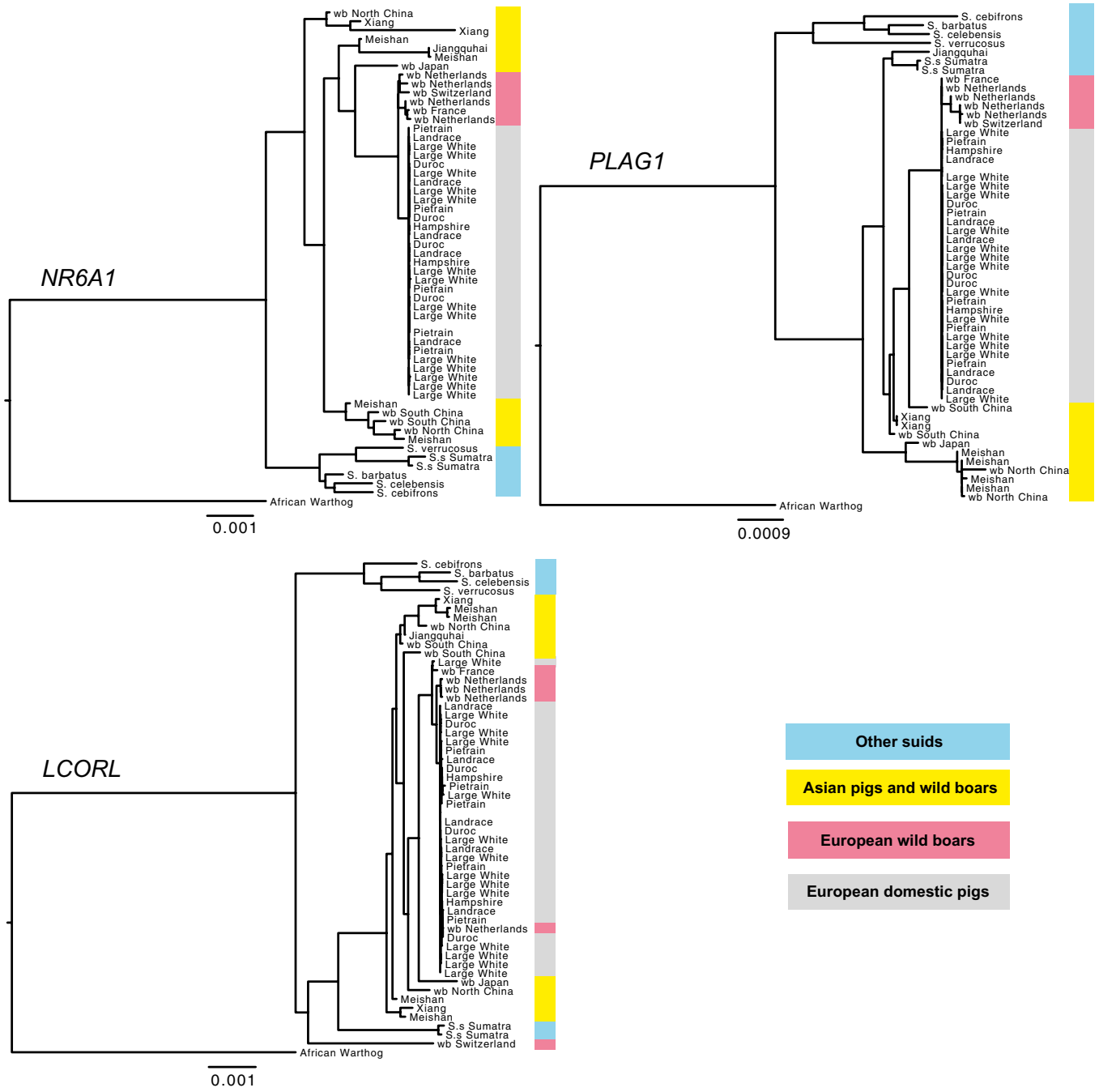



**Fig. S2.** Phylogenetic analysis of the haplotypes associated with *NR6A1*, *PLAG1*, and *LCORL* in individually sequenced pigs. For all three loci, the trees show that the haplotype swept among European domestic pigs is most similar to the haplotypes in European wild boars, thus implying that the selection for an elongated body was initiated in Europe.

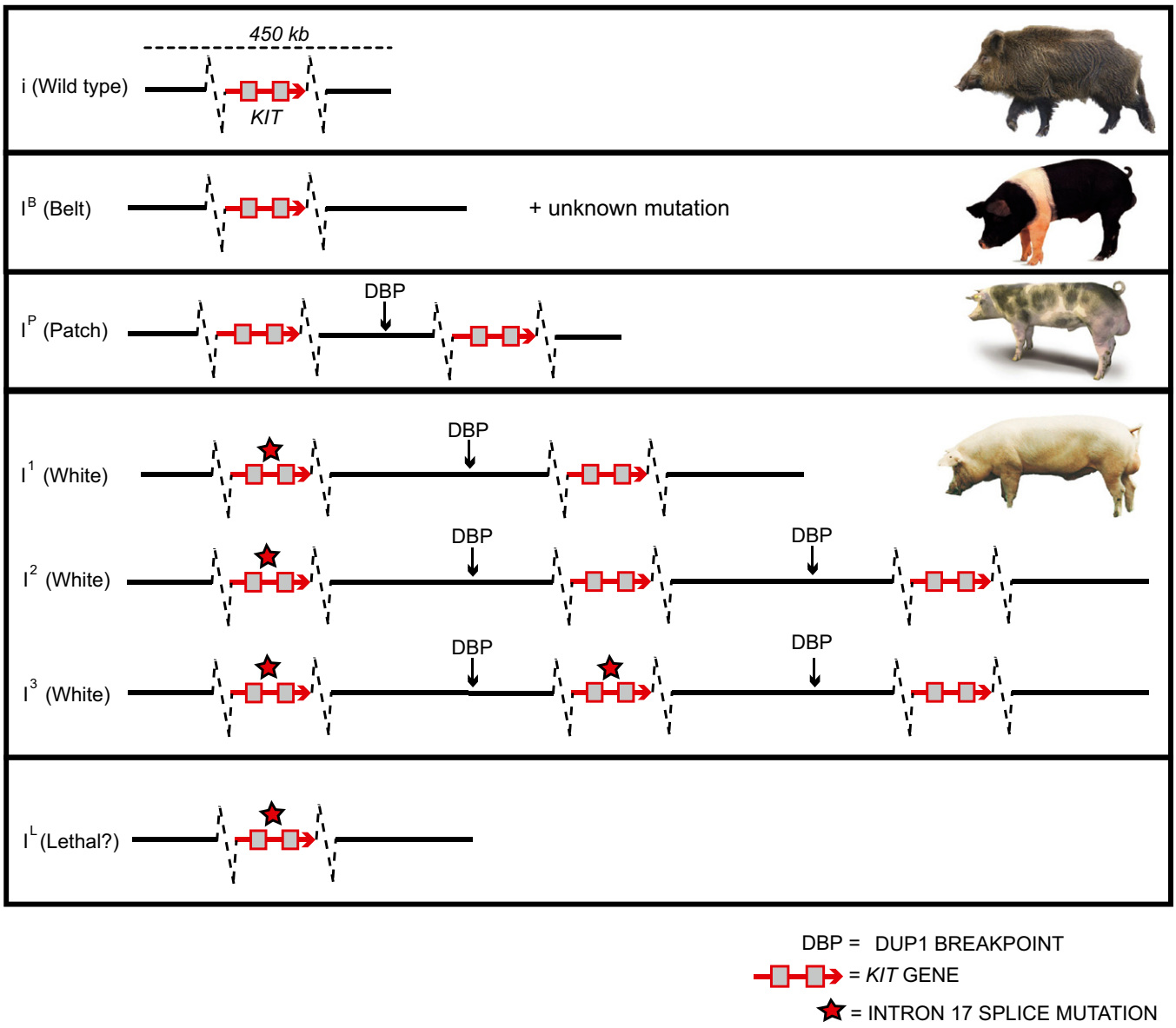



**Fig. S3.** Summary of *KIT* alleles in domestic pigs. Dominant white color in pigs is associated with the presence of a large 450-kb duplication that encompasses the entire gene and which shows copy number variation (two to three copies) between haplotypes (1). One or two of the *KIT* copies present on Dominant white haplotypes carries a splice mutation causing exon skipping and the expression of a KIT protein lacking an essential part of the tyrosine kinase domain. *Patch* is a second variant *KIT* allele causing a partial white spotting pattern, and this allele involves the 450-kb duplication but not the splice mutation (2). A third phenotype, the Belt, is characterized by a white belt across the forelegs, and a previous study demonstrated that *Belt* is a *KIT* allele but did not reveal any causative mutations in coding sequence implying one or more regulatory mutations (3).

1. Giuffra E, et al. (2002) A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking *KIT*. *Mamm Genome* 13(10):569–577.
2. Marklund S, et al. (1998) Molecular basis for the dominant white phenotype in the domestic pig. *Genome Res* 8(8):826–833.
3. Giuffra E, et al. (1999) The Belt mutation in pigs is an allele at the *Dominant white* (*I*/*KIT*) locus. *Mamm Genome* 10(12):1132–1136.

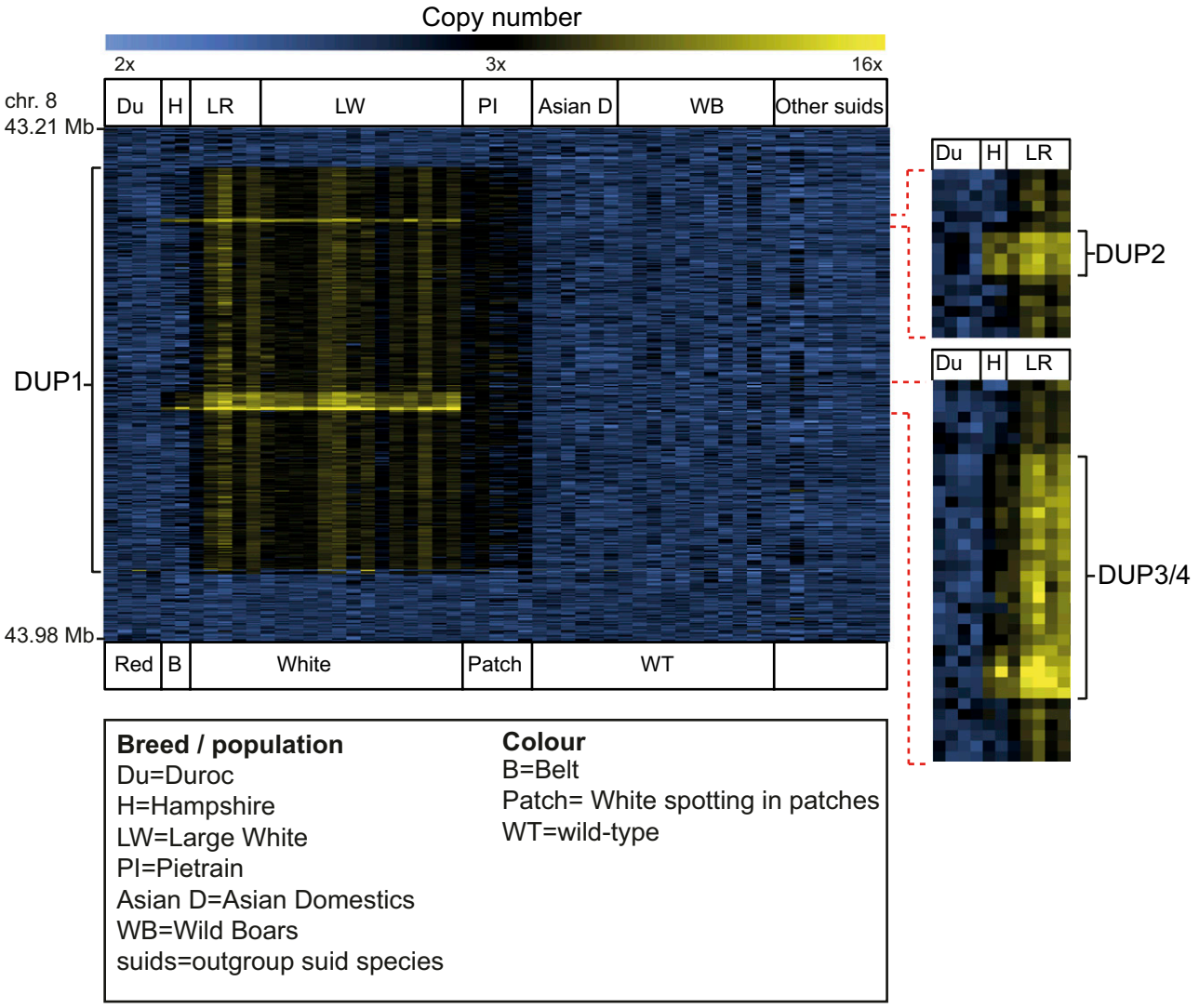



**Fig. S4.** Heat map depicting estimated copy numbers along the *KIT* locus for individually sequenced pigs. For each individual, copy numbers were estimated for each 1-kb window by normalizing observed depths to the average depth of 50 kb of sequence flanking DUP1 on each side. Copy numbers over the DUP2–4 regions for selected individuals are shown in magnifications to the right.

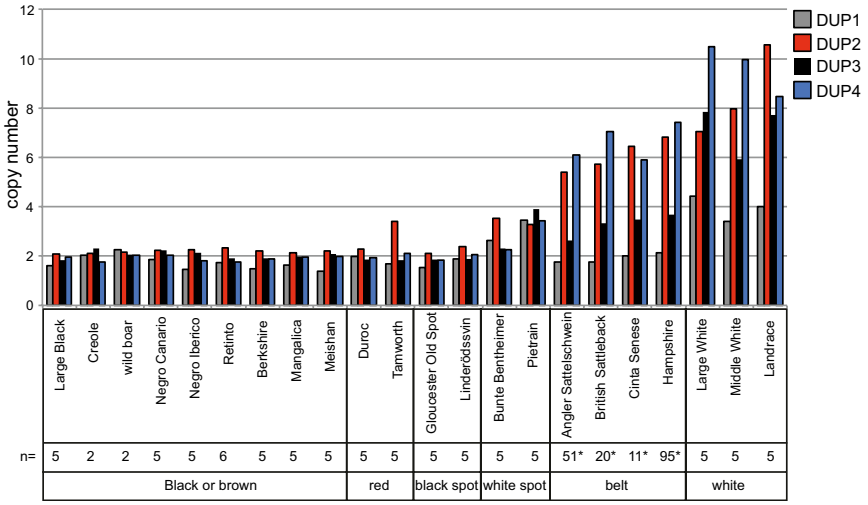



**Fig. S5.** Results from genomic copy number qPCR analysis of DUP1–4 in wild boars and in 21 breeds of domestic pig. Average copy numbers of DUP1–4 are shown for each breed. The numbers of individuals analyzed as well as the breed color phenotype characteristic is shown for each breed. *For Belted breeds, DUP4 was analyzed in five individuals per breed. In some of the Belted breeds, individuals that do not show the breed-characteristic belt are occasionally observed. We did not have access to pictures of all individuals from these breeds, so we could not conclude a minimum number of copies needed to express the belt. The Belt phenotype however is fixed in the Hampshire breed. Observed ranges for Hampshire (number of individuals = 95 for DUP 2 and DUP3 and ($n$ = 5) for DUP4) were (2N copy numbers): DUP2: 5–11× (average = 6.8×), DUP3: 3–5× (average = 3,7×), DUP4: 7–8 × . Observed ranges for white breeds were (2N copy numbers): DUP2: 6–15×, DUP3: 6–10×, DUP4: 3–15 × . All individuals from white breeds had >2 copies of DUP1. Copy numbers of the local duplications relative to the copy numbers of DUP1 in white breeds were as follows: DUP2/DUP1 = 1.5–3; DUP3/DUP1 = 1.5–3. Considering the extent of copy number variation at the *KIT* locus, the numbers of possible genotype combinations of DUP1–4 and the splice mutation are extensive.

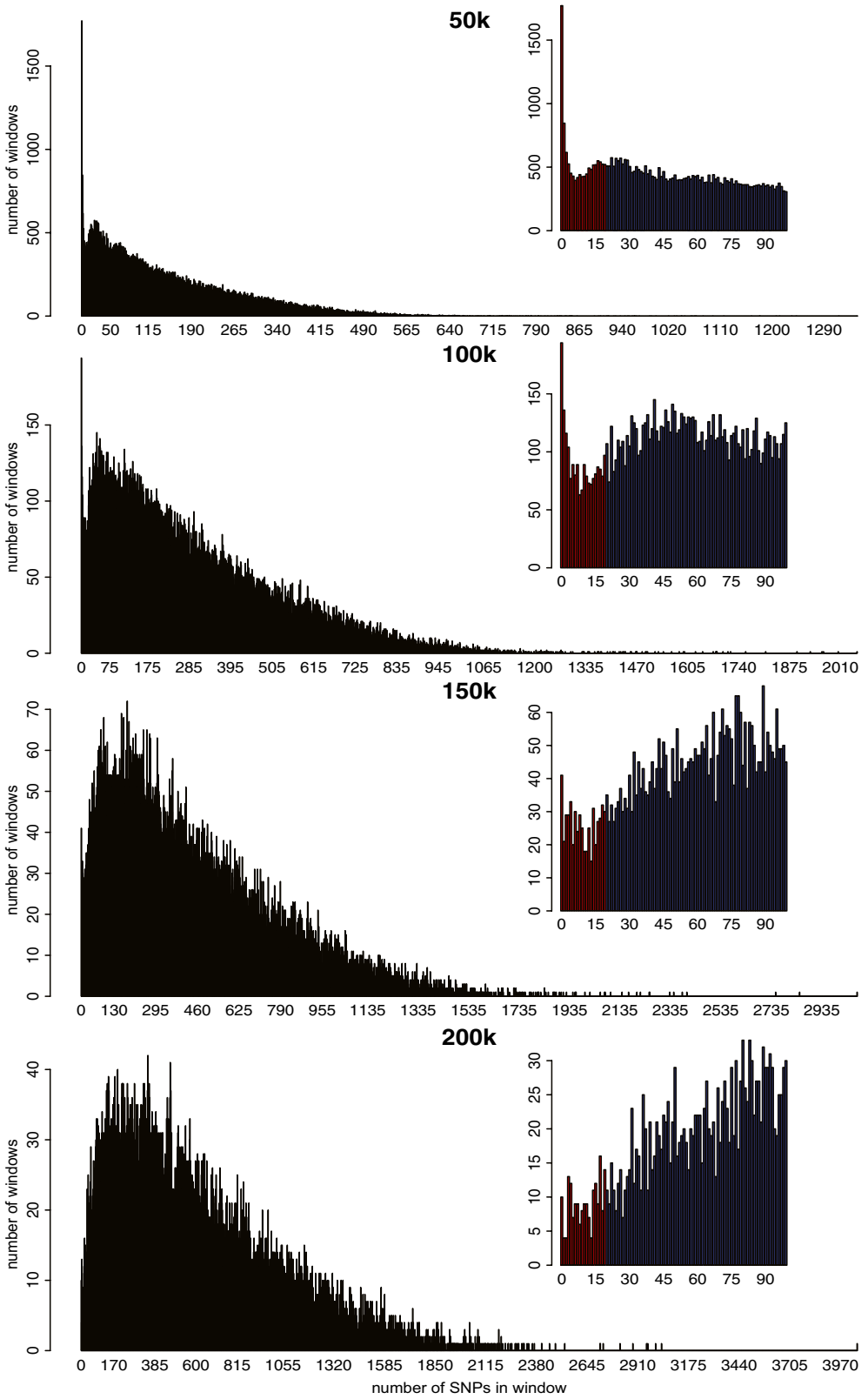



**Fig. S6.** Histograms showing SNP counts in windows sized 50, 100, 150, and 200 kb. The magnified distribution plots have been clipped at 100 SNPs to show the parts relevant for determining an appropriate window size. Red bars indicate windows with <20 SNPs.

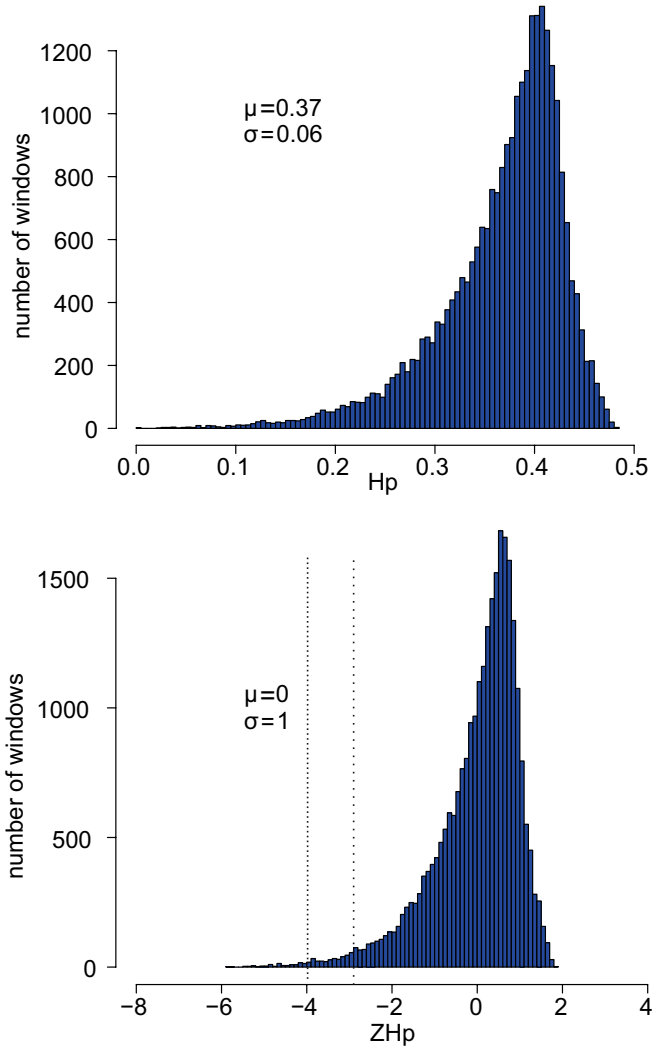



**Fig. S7.** Histograms of genome-wide Heterozygosity values (Hp) and Z-scores thereof (ZHp) based on all 150-kb windows on autosomes in pools of European domestic pigs. Dotted lines indicate the thresholds for inclusion (ZHp ≤ −4) or bridging (ZHp ≤ −3) candidate selective sweeps.

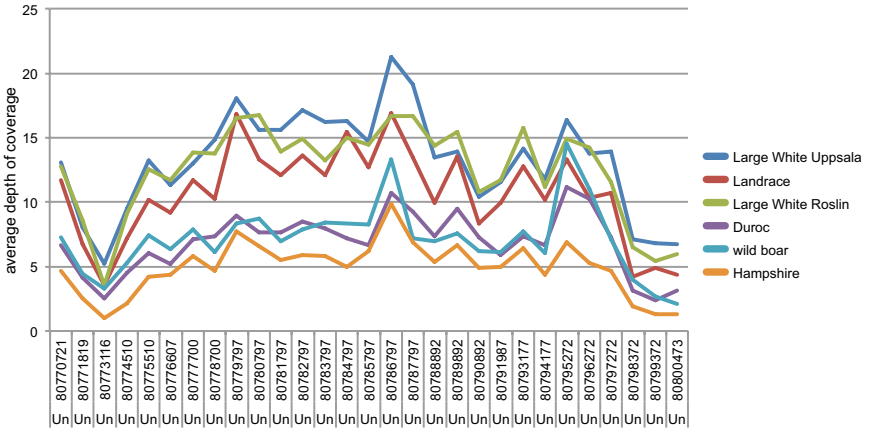



**Fig. S8.** Average depths of coverage for 1-kb windows along the unplaced Illumina scaffold GL895466 in the Sscrofa10.2 genome assembly. Three lines of evidence place this scaffold (size = 31,180 bp) within the assembly gap upstream of *KIT* (Fig. 3*A*). Nucleotides 27,253–31,180 of GL895466 matches SSC8 from 43,537,655–43,541,694 with some gaps in the alignment. BLAST search of GL895466 against the human genome results in hits to a ∼32-kb region immediately upstream of, and extending into, the *KIT* transcription unit. The Dominant white pools (Large Whites and Landrace) have approximately twofold higher depths than non-Dominant White pools, which is consistent with GL895466 being a part of DUP1 and that GL895466 therefore constitutes at least parts of the assembly gap. The observed depth of sequence coverage across this region in Hampshire indicates that no additional large structural variants reside within the assembly gap.

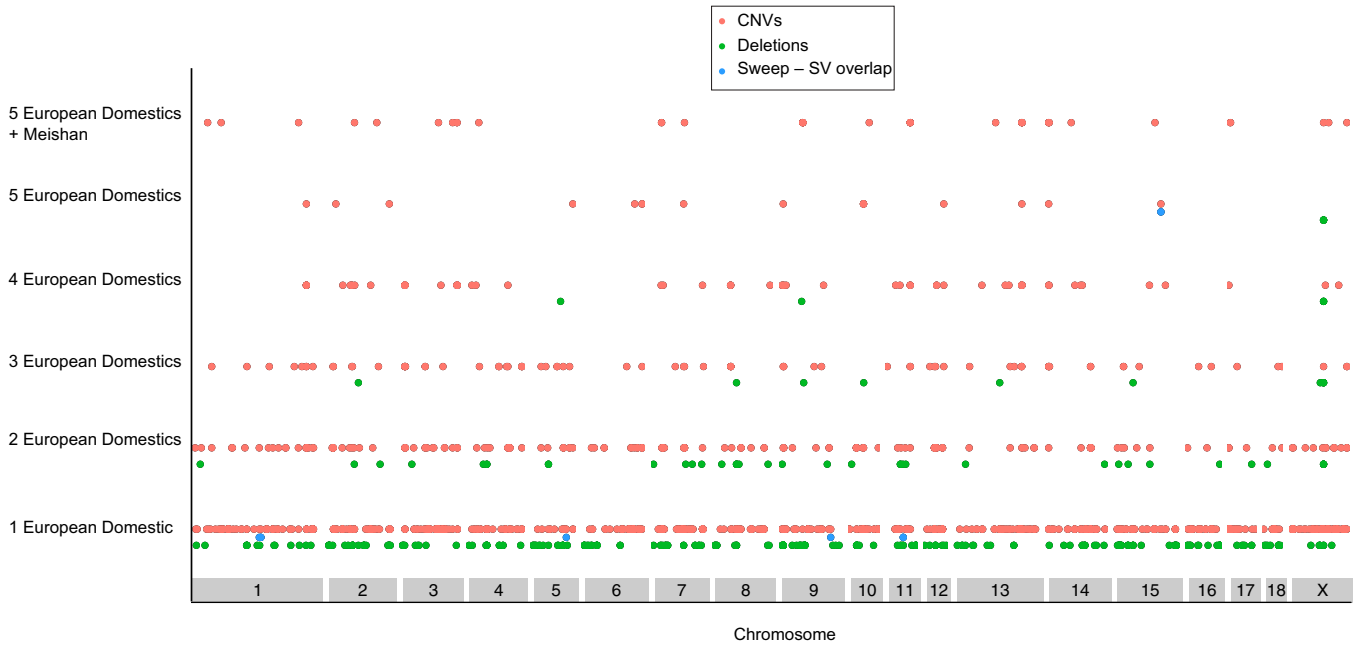



**Fig. S9.** Distribution of identified structural variants along the reference genome assembly. Red and green dots indicate locations of CNVs and deletions, respectively according to the chromosomes given on the *x* axis. The *y* axis indicates whether the structural variants are shared between one and five European domestic populations or between all domestic pig populations analyzed (including the Chinese breed Meishan). The statistical significance for a structural variant was first evaluated independently within each population, and then the overlap between populations was assessed without any restriction as regards the length of overlap. Blue dots indicate six structural variants overlapping putative selective sweeps (SWE). These six structural variants have the following coordinates: Chr1: 172,750,126–172,753,125 bp (duplication); Chr1: 176,848,126–176,851,125 bp (duplication); Chr5: 89,805,376–89,808,375 bp (duplication); Chr9: 129,826,126–129,829,125 bp (duplication); Chr11: 52,557,438–52,557,848 bp (deletion); Chr15: 116,193,376–116,202,375 bp (duplication); the latter duplication is located in an intron of *CASP10*.

## Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOC)
Table S3 (DOCX)
Table S4 (DOCX)
Table S5 (DOCX)