

Supplemental Figure 1. Schematic representation of synthetic mosaic genome simulation scheme used to evaluate GWBC. Red boxes indicate breakpoints generated along the genome to define segments that were stitched together to form a synthetic genome. In brief, in each simulation instance, two animals (Purebred 1 and Purebred2) were randomly sampled from two of the four MARC datasets (*MARChampshire*, *MARClandrace*, *MARCyorkshire*, *MARCduroc*). Subsequently, the number of breakpoints was sampled from a Poisson distribution and the position of each genomic breakpoint was sampled from a uniform distribution. Finally, the genomes of the two selected animals (Purebred 1 and Purebred 2) were segmented at the sampled breakpoints and re-connected to form a synthetic mosaic genome. For instance: in the figure, four breakpoints were sampled, resulting in five segments, segments 1,3, and 5 were obtained from Purebred 1 and segments 2 and 4 were obtained from Purebred 2. Thus, the realized % of genomic composition for purebred 1 was proportional to the sum of the length of the segments that were selected from that individual, and the % composition from Purebred 2 was the complement: % Purebred 2= 100-% Purebred 1. This resulted in synthetic individuals composed of two breeds with known percentage of each breed, ranging from 1% to 99% for each breed.

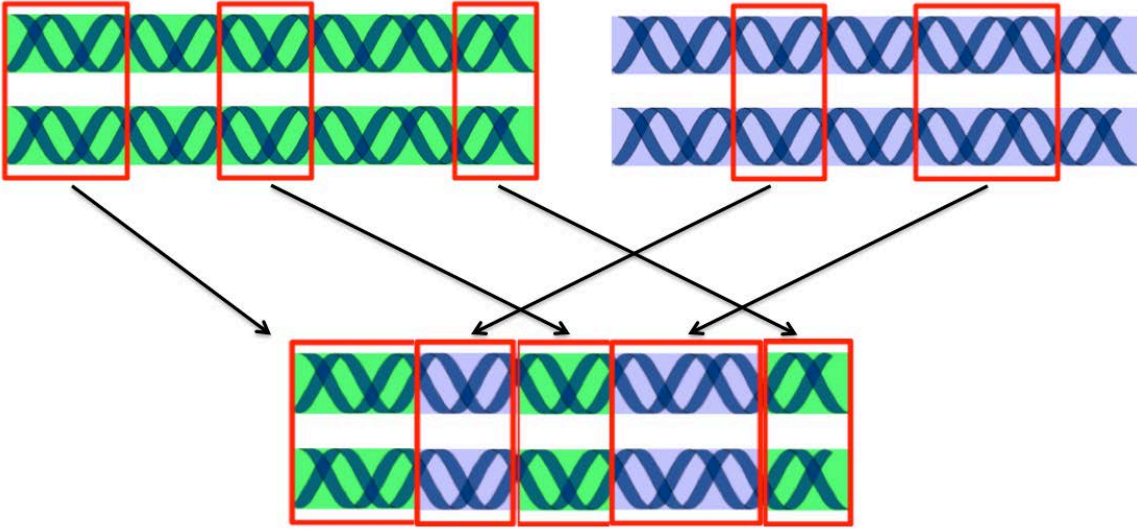
Supplemental Figure 2. Estimated genome-wide breed composition compared to actual genome-wide breed composition of 1000 simulated hybrid genomes, where each hybrid contains a variable proportion of genomic segments of Pietrain origin. Color of each point indicates the breed of one of the parent genotypes used to construct the hybrid, and the composition that is estimated for that breed. Red line has a y-intercept of 0 and slope of 1.

Supplemental Figure 3. R^2 values plotted against actual breed composition of 1000 simulated Pietrain hybrids. Color indicates the non-Pietrain breed present in the hybrid, and the breed that is estimated. Simple linear regression line shown in black with a coefficient estimate of 0.369 and standard error 0.0105.

Supplemental Figure 1

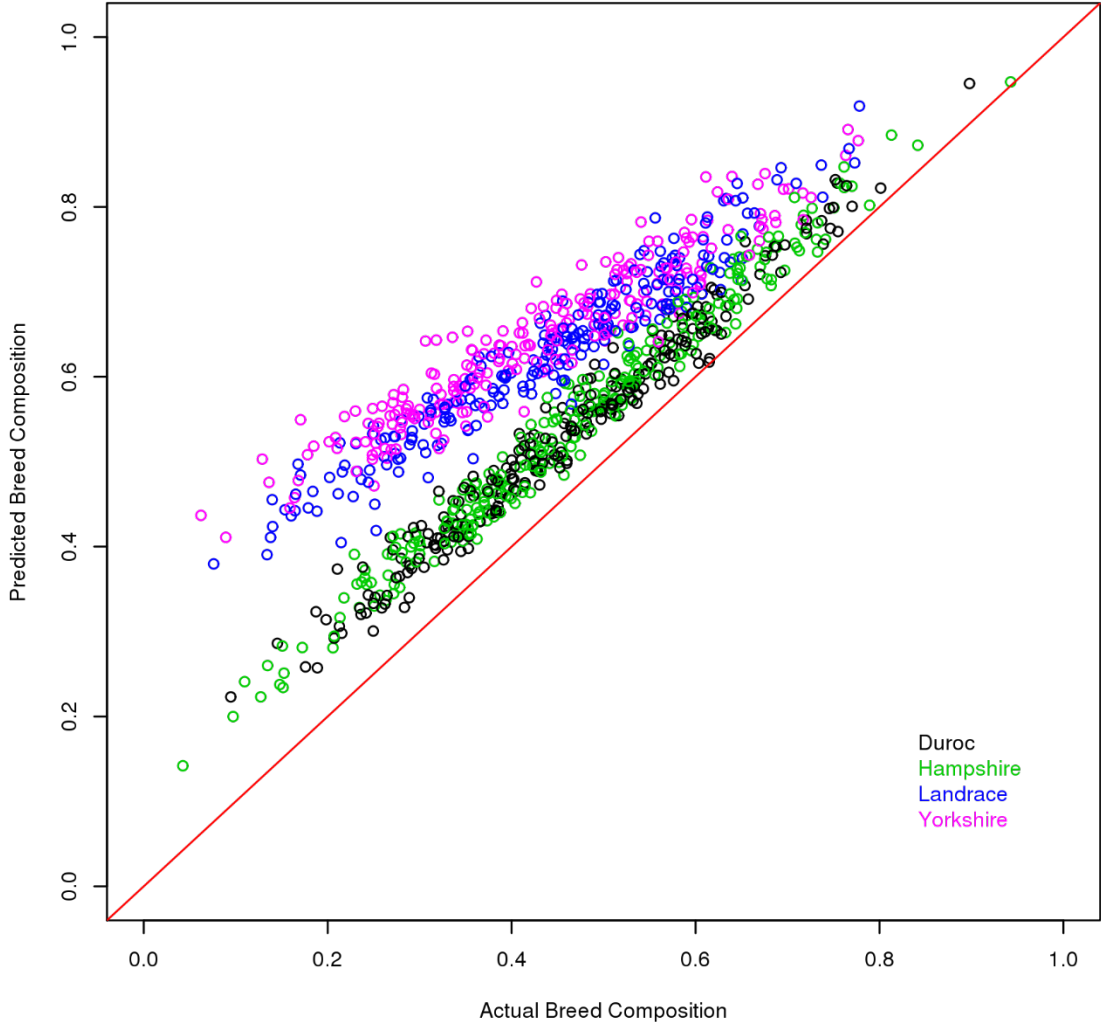
Purebred 1 - e.g. Yorkshire

Purebred 2 - e.g. Hampshire

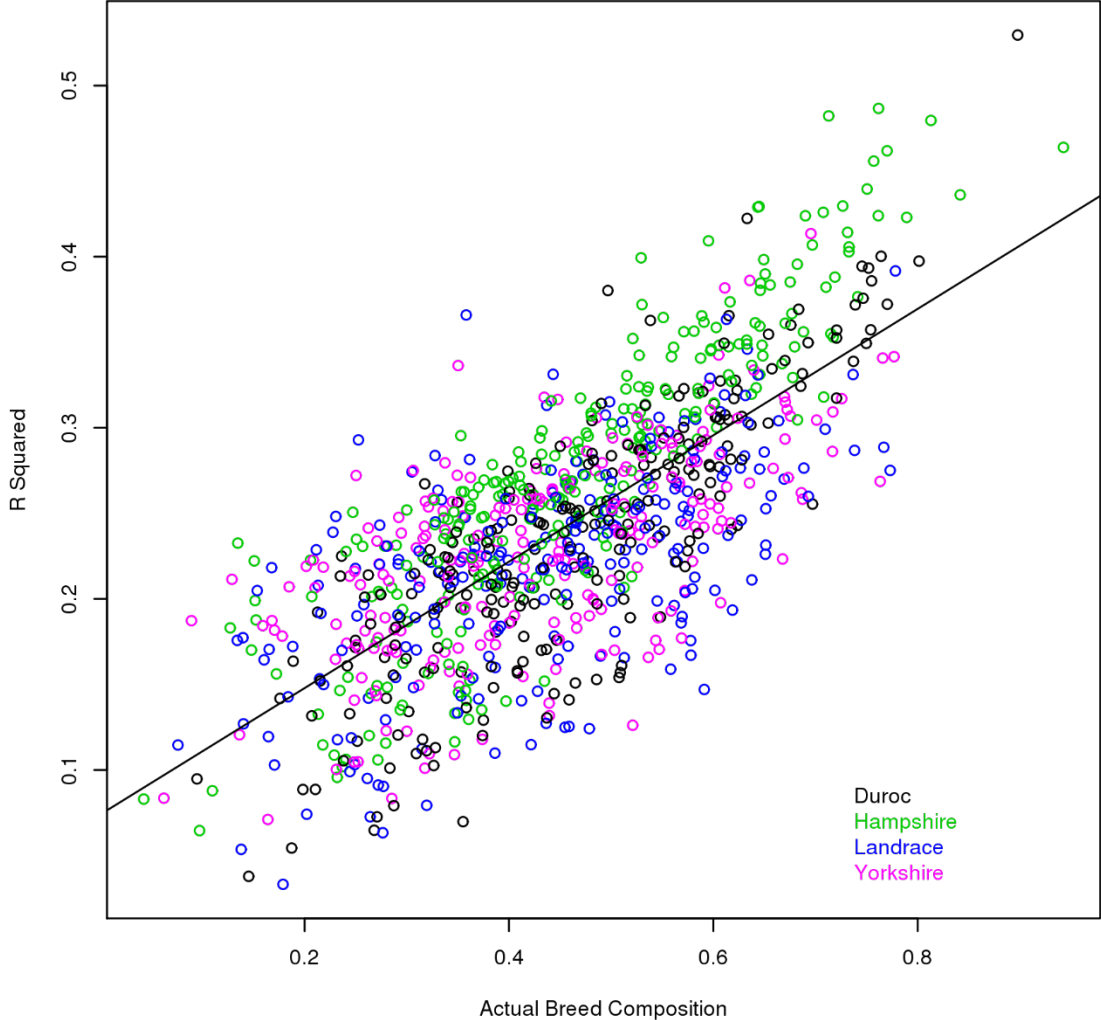


"Synthetic genome"

Supplemental Figure 2



Supplemental Figure 3



Field	Description
Animal	The test animal in question
GWBC_Duroc	Duroc genome-wide breed composition
GWBC_Hampshire	Hampshire genome-wide breed composition
GWBC_Landrace	Landrace genome-wide breed composition
GWBC_Yorkshire	Yorkshire genome-wide breed composition
KBP_White	Kit-based breed probability for purebred white breed (Yorkshire or Landr
KBP_Duroc_White	Kit-based breed probability for Duroc/White hybrid
KBP_Hampshire_White	Kit-based breed probability for Hamsphire/White hybrid
KIT_inconclusive	Whether the KIT test for the animal in question is inconclusive by way of
Comments	Any additional comments for why an animal might fail purity screening, c

race)

missing SNPs in KIT. In this case the KBP results are averaged across all possibilities other than a "yes" in the KIT_inconclusive field.