

Predicting breed composition in an Angus-Brahman crossbred population using genomic data

M. Gobena, J. Leal, M. A. Elzo, and R. G. Mateescu

Department of Animal Sciences, University of Florida, Gainesville, FL



Introduction

Pedigree information has conventionally been used to estimate breed composition. It has been suggested that breed composition or ancestry information estimated from genomic data can be more accurate than estimates based on pedigree information. It can also be especially useful when pedigree information is missing or incomplete. The goal of this study was to examine the feasibility and accuracy of using genotype data to estimate breed composition and subsequently identify the minimum number of markers needed to determine breed composition with adequate accuracy.

Materials and Methods

Sampling, Genotyping and Quality Control: A total of 782 cattle consisting of purebred Angus, purebred Brahman and crossbreeds across the spectrum were included in the study. DNA was extracted from blood or semen samples and genotyped with the Illumina Bovine 250k SNP chip (GGP F250). After applying quality control measures (minor allele frequency, call rate, LD pruning, HWE deviation and genotyping completion), 54,728 SNP and 676 animals remained and were used in subsequent analysis.

Population Structure Identification: Individual breed membership coefficients were estimated using the projection analysis feature of the software ADMIXTURE1.3 (K=2) which minimizes confounding effects of familial relationships¹. Principal component analysis (PCA) was also performed on a set of mutually unrelated samples (pairwise kinship < 0.1) in order to identify major axes of variation using the R function 'pcair'². Principal component values for the rest of the samples were then predicted based on genetic similarity to the unrelated set.

Selection of Ancestrally Informative Markers (AIM): Twenty subsets of SNP were selected based on Fixation Index (Fst) starting from the top 5 and increasing the number by 5 up to 100. For each subset of SNP, 5-fold cross-validation was performed to assess the accuracy of prediction which was measured as correlation with breed composition estimates by ADMIXTURE using full genotype data.

Results

There was a very strong correlation between the breed composition estimates obtained from pedigree information and estimates from ADMIXTURE (Correlation 0.965; **Figure 1**) and the first principal component (PC1) from PCA (Correlation 0.965; **Figure 2**). There was an almost perfect correlation (0.999) between PC1 and breed composition estimates from ADMIXTURE. As small as 5 SNP selected based on Fst were sufficient to predict breed composition with accuracy of > 0.9 (**Figure 3**).

Conclusion

There was strong agreement between breed composition estimates from pedigree and population structure identified by both ADMIXTURE and PCA. The results show that a very small number of SNP were sufficient to predict breed composition in the study population with high accuracy which will reduce the cost of genotyping. However, cross validation on a separate population has to be performed to make sure markers identified in this study can be used in other Angus-Brahman crossbred populations.

Reference

1. Shringarpure, Suyash S., Carlos D. Bustamante, Kenneth L. Lange, and David H. Alexander. 2016. "Efficient Analysis of Large Datasets and Sex Bias with ADMIXTURE." *bioRxiv* 1(1):1-10.
2. Conomos, Matthew P., Mike Miller, and Timothy Thornton. 2016. "Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness." 39(4):276-93.

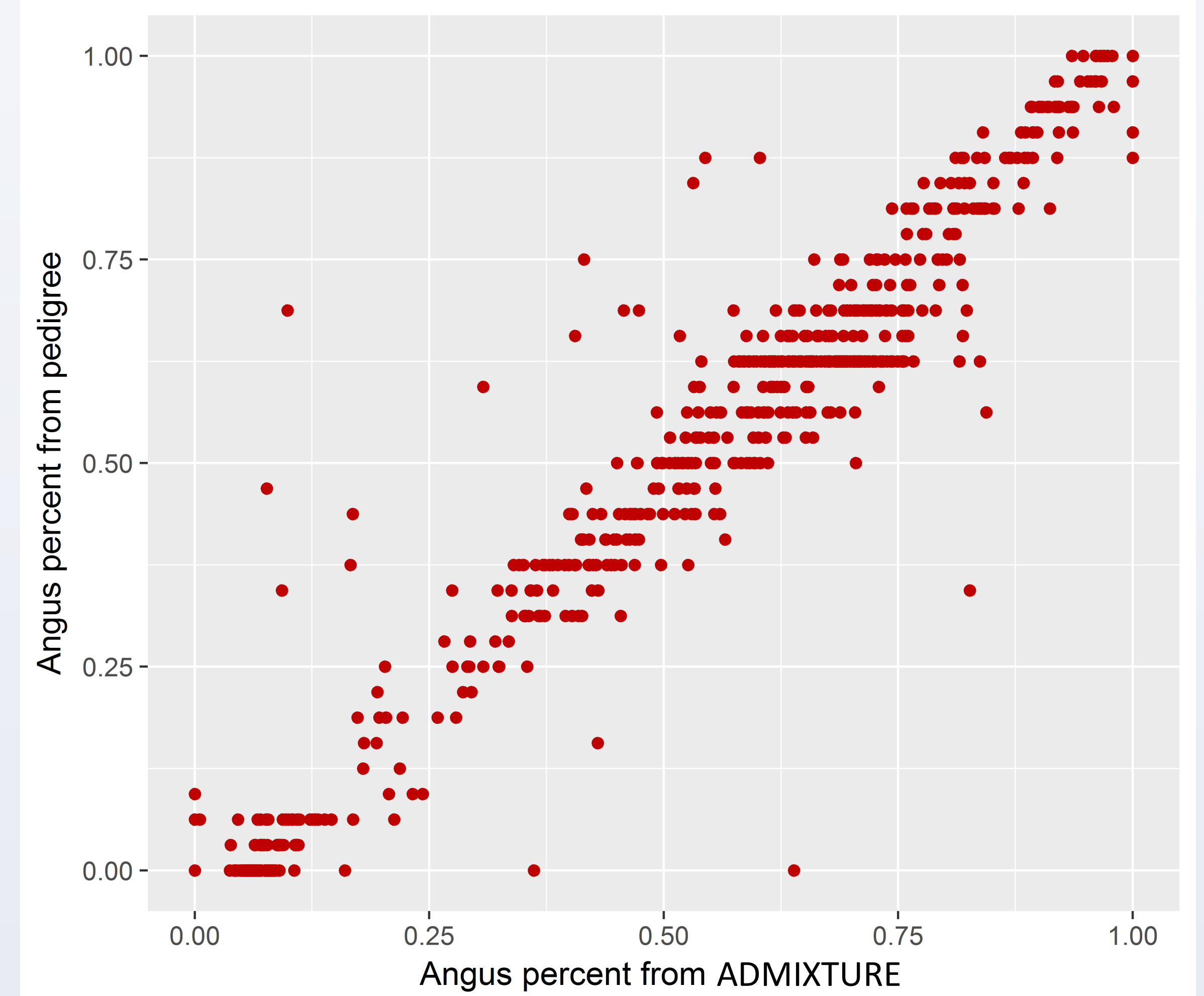


Figure 1: The scatter plot shows a very strong positive relationship (correlation of 0.965) between Angus ancestry estimates from Admixture and pedigree.

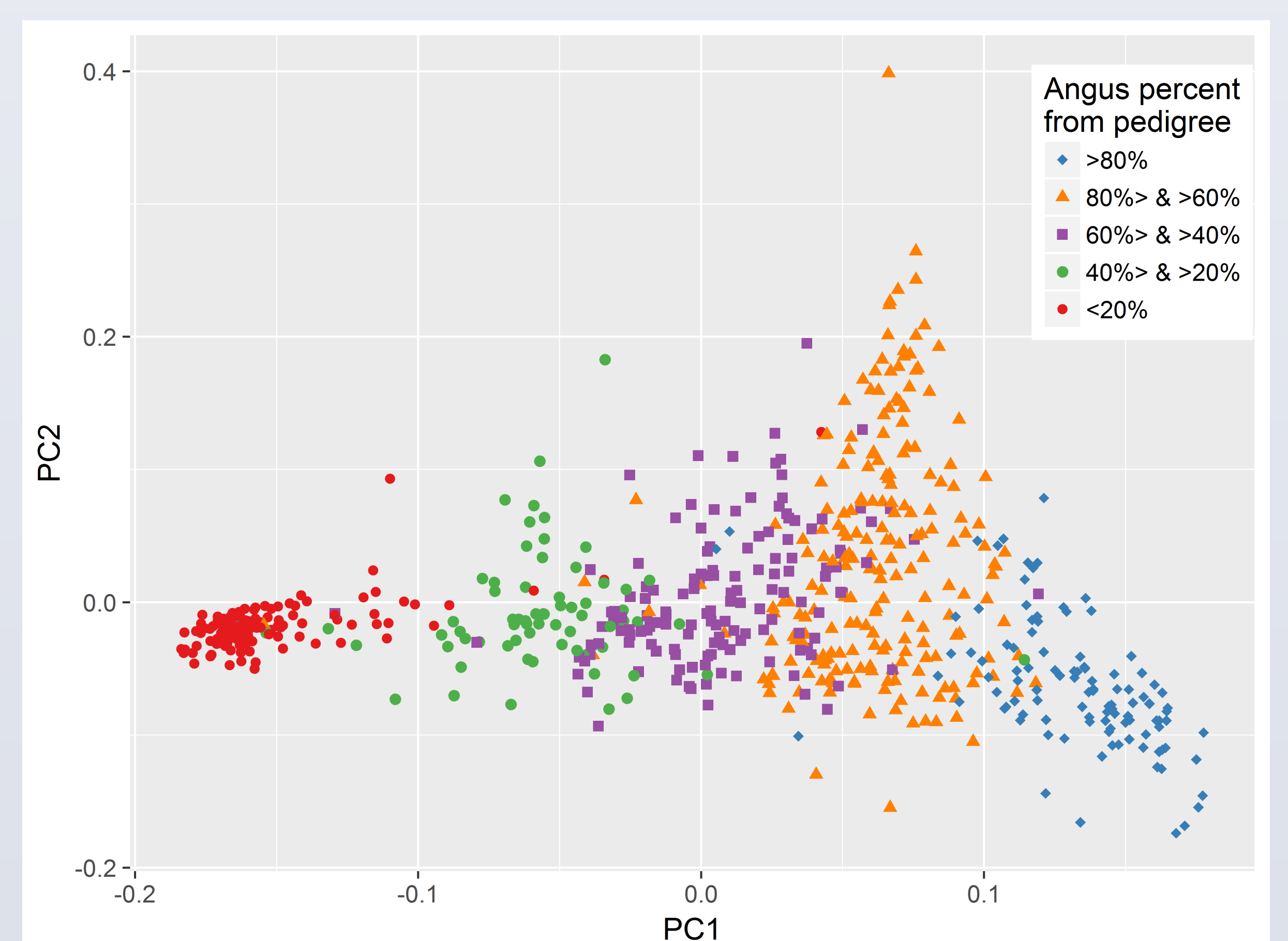


Figure 2: The figure shows how the first PC resolved the samples in a manner consistent with grouping based on Angus percent from pedigree.

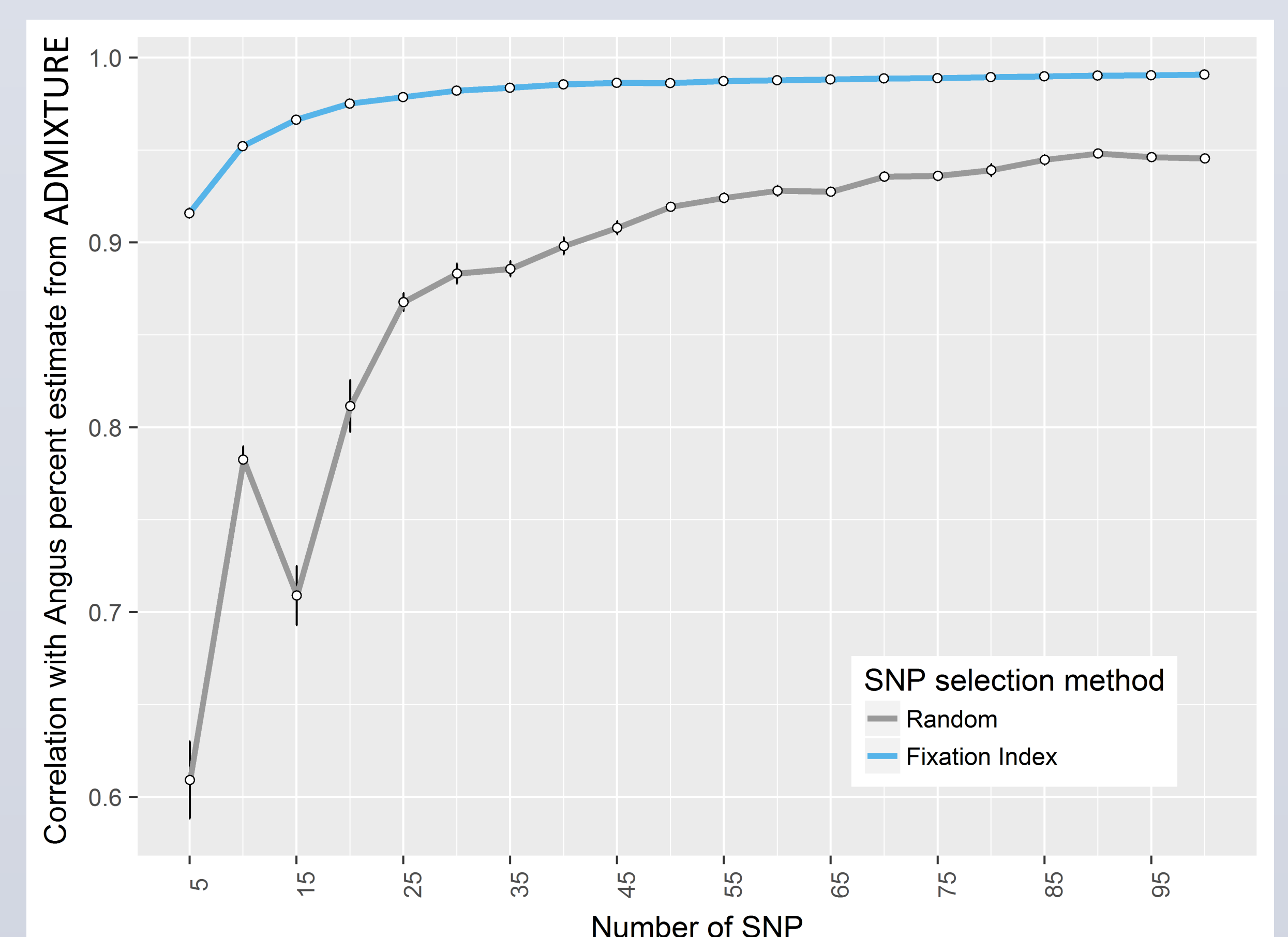


Figure 3: The graph shows results from 5-fold cross-validation applied to SNP sets of different size selected randomly and based on Fst value.

Acknowledgements and Contact

Financial support: Florida Beef Council, Florida Cattlemen's Association, UF IFAS

Contact: mgobena@ufl.edu

