

Xiao-Lin Wu<sup>1,2\*</sup>, Zhi Li<sup>1</sup>, Yanfang Wang<sup>2,3</sup>, Jun He<sup>1,4</sup>, Guilherme J. M. Rosa<sup>2</sup>, Ryan Ferretti<sup>1</sup>, John Genho<sup>1</sup>, Richard G. Tait Jr.<sup>1</sup>, Jamie Parham<sup>1</sup>, Tom Schultz<sup>1</sup>, Stewart Bauck<sup>1</sup>  
 1 Biostatistics and Bioinformatics, Neogen GeneSeek Operations, Lincoln, NE 68504, USA; 2 Department of Animal Sciences, University of Wisconsin, Madison, WI 53706, USA;  
 3 Ministry of Education Key Laboratory of Marine Genetics and Breeding, Ocean University of China, Qingdao, Shandong 266100, China;  
 4 College of Animal Science and Technology, Hunan Agricultural University, Changsha, Hunan 410128, China.

## Introduction

- Genomic breed composition (GBC) of an individual animal refers to the partition of its genome according to the inheritance from its ancestors or ancestral breeds.
- Various statistical methods have been proposed to estimate GBC (Pritchard et al., 2000; Tang et al., 2005; Bansal and Libiger, 2015; Frkonia et al., 2012), but the interpretations of estimates have varied across methods. For example, linear regression estimated the GBC of an individual by adjusted regression coefficients of coded genotypes of each animal as the progeny on the ancestral allele frequencies (Chiang et al., 2010; Kuehn et al., 2011; VanRaden and Cooper, 2015).
- The admixture model, which postulates that an observed genotype is an instance of a multinomial distribution with the genotype probability being a mixture of those of their ancestors. In this case, the GBC of an individual animal is estimated by the weights of the admixture (Bansal and Libiger, 2015).
- In the present study, we proposed the use of path analysis to decompose the causality relationships between composite (or crossbred) animals and their putative ancestors (or reference breeds) and to estimate GBC of individual animals in terms of the relative determination of respective ancestral (or reference) breeds.

## Materials and Methods

### Number of SNPs and animals

**Table 1.** Number of genotyped animals and number of SNPs on GeneSeek Genomic Profiler (GGP) 30K (GGP 30K), 40K (GGP 40K), and 50K (GGP 50K) SNP chips used in the present study<sup>1,2</sup>

Type	Breed	GGP30K		GGP40K		GGP50K		Animals	
		nAnim	nSNP	nAnim	nSNP	nAnim	nSNP	Before DC	After DC
Composite	Beefmaster	23	32,179	300	40,663	7,282	49,463	7,605	7,605
	Brangus	1,319	32,179	3,053	40,660	3,605	49,463	7,969	7,969
	Angus	6,839	32,179	18,198	40,660	20,359	49,463	45,396	45,367
Ancestral	Beefmaster	---	---	1,811	30,720	509	43,984	2,320	2,271
	Hereford	4,000	32,179	4,000	40,660	2,423	49,463	10,423	10,414
	Shorthorn	---	---	335	40,660	1,232	49,463	3,587	1,577
Non-ancestral	Galloway	2,763	32,179	5,498	40,660	9,508	49,463	17,769	17,735
	Limousin	373	32,179	2,264	40,660	5,043	49,415	7,680	7,677
Non-ancestral	Simmental	3,130	32,179	5,838	40,660	14,734	49,463	23,722	23,697
	Wagyu	1,463	32,179	1,506	40,660	23,720	49,463	26,009	26,344
Sum		19,919	---	42,823	---	88,435	---	152,160	150,676

### Statistical method

- Linear regression

$$y_i = \mu + \sum_{j=1}^k b_j x_{ij} + e_i$$

where y is a vector of coded genotypes for the i-th animal, and x is a vector of allele A frequencies for the reference populations.

- Admixture model

$$Pr(y_i | f_i) = \begin{cases} f_i^2 & \text{if } y_i = 2 \\ 2f_i(1-f_i) & \text{if } y_i = 1 \\ (1-f_i)^2 & \text{if } y_i = 0 \end{cases}$$

$$L(W) = \sum_{i=1}^M [y_i \ln(f_i) + (2 - y_i) \ln(1 - f_i)] + C, \text{ where } C = \sum_{i=1}^M \ln \binom{2}{y_i}$$

- Path analysis

$$p_{y_i x_j} = b_j \times \frac{p_{y_i x_j}}{p_{y_i x_j}}$$

$$D\text{-GBC} = \frac{p_{y_i x_j}^2}{\sum_{j=1}^k p_{y_i x_j}^2}$$

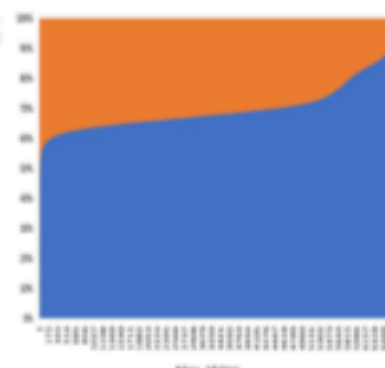
$$C\text{-GBC} = \frac{(p_{y_i x_j}^2 + \sum_{j \neq k} p_{y_i x_j} p_{y_i x_k} p_{y_i x_j})}{\sum_{j=1}^k (p_{y_i x_j}^2 + \sum_{j \neq k} p_{y_i x_j} p_{y_i x_k} p_{y_i x_j})}$$



## Results

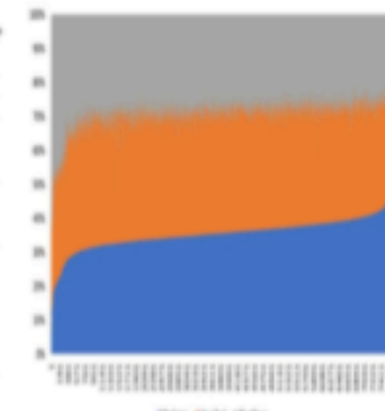
**Table 2.** Comparison of estimated GBC for 7,969 Brangus with genotype data, obtained by the admixture model, linear regression, and path analysis techniques, respectively, using only Angus and Beefmaster in the reference breed set.

Model	Panel	GGP 30K, GGP 40K				GGP 50K			
		Angus	Beefmaster	Angus	Beefmaster				
Admixture	1K	69.9%	7.3%	30.1%	7.3%	70.3%	7.1%	29.7%	7.1%
	5K	69.8%	6.8%	30.2%	6.8%	70.1%	6.8%	29.9%	6.8%
	10K	68.8%	7.1%	31.2%	7.1%	69.1%	7.0%	30.9%	7.0%
Linear Regression	1K	70.0%	7.6%	30.0%	7.6%	70.4%	7.6%	29.6%	7.6%
	5K	69.5%	7.4%	30.5%	7.4%	69.9%	7.5%	30.2%	7.5%
	10K	68.6%	7.3%	31.4%	7.3%	69.0%	7.6%	31.0%	7.6%
Path Analysis (D-GBC)	1K	71.8%	11.9%	28.2%	11.9%	71.3%	12.3%	28.5%	12.3%
	5K	69.6%	11.8%	30.4%	11.8%	70.2%	12.4%	29.8%	12.4%
	10K	69.5%	11.7%	30.5%	11.7%	70.2%	12.3%	29.8%	12.3%
Path Analysis (C-GBC)	1K	70.9%	11.7%	29.1%	11.7%	70.6%	12.1%	29.4%	12.1%
	5K	68.7%	11.3%	31.3%	11.3%	69.3%	12.0%	30.7%	12.0%
	10K	68.2%	11.3%	31.8%	11.3%	68.9%	11.8%	31.2%	11.8%



**Table 3.** Path analysis using the correlation data for 7,605 Beefmaster animals with three ancestral breeds as the reference and three SNP panels. The Beefmaster animals were genotyped on either GGP 30K, GGP 40K bovine SNP chip or GGP 50K bovine SNP chip, respectively.

Statistics		GGP30K, GGP 40K			GGP 50K		
		1K	5K	10K	1K	5K	10K
Correlation with Beefmaster	Beefmaster	0.532	0.549	0.536	0.544	0.561	0.570
	Hereford	0.511	0.504	0.504	0.549	0.548	0.543
	Shorthorn	0.485	0.443	0.483	0.518	0.477	0.520
Path coefficient	Beefmaster	0.513	0.514	0.517	0.536	0.522	0.526
	Hereford	0.375	0.381	0.371	0.420	0.417	0.398
	Shorthorn	0.275	0.263	0.276	0.310	0.282	0.298
D-GBC	Beefmaster	54.9%	55.2%	55.6%	51.3%	51.9%	52.8%
	Hereford	29.3%	30.3%	28.6%	31.3%	33.0%	30.2%
	Shorthorn	15.7%	14.5%	15.8%	17.2%	15.2%	16.9%
C-GBC	Beefmaster	50.1%	51.3%	51.0%	46.0%	47.6%	48.0%
	Hereford	31.1%	31.9%	29.8%	33.4%	34.1%	31.4%
	Shorthorn	18.8%	17.3%	19.2%	20.6%	18.3%	20.6%



## Conclusions

- Path analysis provides an alternative interpretation and an estimation method of GBC. Two measures of GBC were proposed: D-GBC considered only direct path effects of each reference breed, whereas C-GBC also included indirect path effects due to the correlation between reference breeds.
- Estimated GBC for the Brangus were comparable between path analysis, and linear regression and the admixture model, but large differences were observed with the estimated GBC for the Beefmaster due to large correlated effects.