

# ESTIMATING AGE OF ADMIXTURE IN A CATTLE POPULATION BASED ON SNP CHIP DATA

Anamarija FRKONJA <sup>1,2</sup>, Tom DRUET <sup>3</sup>, Birgit GREDLER <sup>4</sup>, Ino CURIK <sup>5</sup>, Johann SÖLKNER <sup>2,6</sup>

## ABSTRACT

The aim of this study was to predict individual age of admixture in the crossbred Swiss Fleckvieh population. We checked how well the method is dealing with recent admixture with high throughput single nucleotide polymorphism data from the bovine 50K SNP Chip. A total of 101 Red Holstein, 91 Simmental, and 308 crossed animals were available for analysis. Age of admixture was derived from the complete pedigree and molecular markers. The method applied (using SABER software) based on Markov-hidden Markov model was able to derive age of admixture similar to estimates of pedigree data, however the values were often overestimated. Of 21 investigated cases, results from SNP data reflected paternal and maternal age of admixture well for 9 cases but provided results out of range for the other 12 cases. Alternative methods based on breed-specific haplotype blocks need to be evaluated in the future.

**Key words:** cattle / breeds / Swiss Fleckvieh / age of admixture / breed composition / genetics

## 1 INTRODUCTION

Age of admixture is of big interest in population history reconstruction, in genome wide association studies and admixture mapping. There are many ways of using information about population structure in the field of genome wide association studies.

Freeman *et al.* (2006) used X Chromosome Haplotype Mosaicism to estimate age of admixture. They took closely linked X chromosome microsatellites in cattle populations with differing histories of admixture from Africa, Europe, the Near East, and India. They genotyped male animals from these populations, obtaining unambiguous haplotypes, and measured levels of linkage disequilibrium (LD) and ancestral mosaicism. Extensive LD, likely to be the result of ongoing admixture, was discovered in hybrid cattle populations from the perimeter of the tsetse zone in West Africa. A Bayesian method to

assign microsatellite allele ancestry was used to designate the likely origin of each chromosomal segment and assess the relative ages of admixture in the populations. A gradient of the age of admixture in the African continent emerged, where older admixture has produced more fragmented haplotypes in the south, and longer intact haplotypes, indicating more recent hybridization, feature in the northwest. Age of admixture derived from a very large high throughput number of markers of the autosomal part of genome potentially can provide more precise values when admixture happened.

Tang *et al.* (2007) used a Markov-hidden Markov model (MHMM) which accounts for background linkage disequilibrium (LD) in ancestral populations. Their model estimates allele frequencies in ancestral populations based on pure individuals and then uses that information to estimate population origin at each marker

1 Univ. of Natural Resources and Life Sciences Vienna, Dept. of Sustainable Agricultural Systems, Division of Livestock Sciences, Gregor Mendel Str. 33, A-1180 Vienna, Austria, e-mail: anamarija.frkonja@boku.ac.at

2 Corresponding author

3 Univ. of Liege, Fac. of Veterinary Science, Dept. of Animal Production, 1 Avenue de l'Hopital, Liege Belgium

4 Qualitas AG, Chamerstrasse 56, Ch-6300 Zug, Switzerland

5 Univ. of Zagreb, Fac. of Agriculture, Dept. of Livestock Sciences, Svetosimunska 25, 10000 Zagreb, Croatia

6 Same address as 1, e-mail: johann.soelkner@boku.ac.at

for each admixed individual. In addition, it relies on recombination rate between ancestral haplotypes or distribution of haplotype lengths within one individual. With this model they assessed age of admixture for different simulated data sets; from 10 to 25 generations since admixture.

Swiss Fleckvieh is a good example for such studies because of pedigree for all studied animals is available and it is therefore possible to derive age of admixture from pedigrees. The main aim of this study was to analyse efficiency of assessing age of admixture using SABER software with high throughput single nucleotide polymorphism (SNP) data of Swiss Fleckvieh cattle population. Pedigree derived age of admixture was compared with SNP derived age of admixture.

## 2 MATERIAL AND METHODS

Swiss Fleckvieh was established about 40 years ago from crossing of a large part of the local Simmental breed with Red Holstein Friesian cattle for improved milk production. Illumina Bovine SNP50 Beadchip (Illumina, 2009) data for 500 animals, 101 pure Red Holstein Friesian, 91 Simmental and 308 crosses with a wide range of pedigree based levels of admixture were available. Age of admixture through the pedigree was assessed for 21 crossed animals with wide range of crossing. Quality control for SNP data was performed with PLINK (Purcell *et al.*, 2007), excluding markers with a call rate below 0.90. STRUCTURE software (Pritchard *et al.*, 2000; Falush *et al.*, 2003) was run on SNPs from all autosomes assuming that breeds are separated with two ancestral populations (Frkonja *et al.*, 2012). Individual admixture proportions

**Table 1:** Approximate time of admixture derived out of pedigree and SNP data shown with admixture proportion of 21 Swiss Fleckvieh cattle calculated from pedigree and molecular information

ID	PedAP	MolAP	$\tau$ pedp	$\tau$ pedm	$\tau$ molp	$\tau$ molm
ID13	0.02	0.13	0.00	5.00	0.00	0.00
ID14	0.05	0.05	0.00	5.00	2.83	0.13
ID15	0.10	0.08	0.00	4.00	2.02	0.16
ID16	0.19	0.40	0.00	4.00	1.57	1.05
ID17	0.25	0.38	0.00	1.00	3.98	2.35
ID8	0.45	0.52	0.00	4.60	3.85	3.97
ID1	0.47	0.47	2.50	1.50	5.82	5.94
ID7	0.47	0.54	4.30	4.00	6.68	6.51
ID18	0.50	0.50	0.00	0.00	10.24	10.24
ID2	0.50	0.50	4.38	4.60	4.76	4.76
ID3	0.50	0.50	1.50	1.50	5.17	5.17
ID4	0.55	0.51	0.00	3.00	4.76	4.77
ID5	0.55	0.52	0.00	0.00	3.97	4.06
ID6	0.50	0.58	2.50	2.50	5.95	5.72
ID19	0.66	0.73	2.50	3.33	4.18	4.69
ID20	0.65	0.63	6.00	5.00	4.85	4.91
ID21	0.66	0.75	6.00	5.14	3.31	4.58
ID9	0.54	0.62	0.00	3.83	5.94	5.62
ID10	0.97	1.00	0.00	6.25	0.18	4.96
ID11	0.98	1.00	0.00	7.33	0.06	4.92
ID12	0.97	1.00	0.00	4.00	0.15	4.92

ID – Identification number of animal; MolAP – Molecular admixture proportion calculated with (Q membership) STRUCTURE software; PedAP – admixture proportion calculated with pedigree;  $\tau$  pedp – Estimated age of admixture from pedigree for paternal side;  $\tau$  pedm – Estimated age of admixture from pedigree for maternal side;  $\tau$  molp – Estimated age of admixture from SNP data for paternal side, calculated with assumption that admixture happened 5 generations ago;  $\tau$  molm – Estimated age of admixture from SNP data for maternal side, calculated with assumption that admixture happened 5 generations ago

derived from STRUCTURE were then used as input data for SABER (Tang *et al.*, 2007). The method takes a single admixed individual, and estimates, over all chromosomes, the parameters describing the admixture times of an individual, given ancestral population allele frequencies and other information. The program was repeatedly run for every individual assuming that admixture happened from 1 to 12 generations ago (i.e., input parameters were set from 1 to 12 for each side of pedigree separately).

In Swiss Fleckvieh admixture events frequently happened more than once and admixture on both parental chromosomes does not have to be equal. In order to compare SNP based results with pedigree, pedigree derived age of admixture needed to be calculated. Age of admixture was derived from the paternal and maternal sides of the pedigree separately. For this we counted the number of generations back to each admixture event, defined as a mating involving at least one purebred animal, excluding the mating of two animals of the same breed, and averaged these numbers for all events (Fig. 1).

### 3 RESULTS AND DISCUSSION

Age of admixture derived from molecular data with software SABER obtained with initial values (admixture

**Table 2:** Sensitivity of the method to different assumed initial values (present in numbers in table from 1 to 12, same assumption for both parental sides) of age of admixture shown with proportion of Red Holstein and Simmental calculated with SNP data and pedigree admixture proportion for three animals

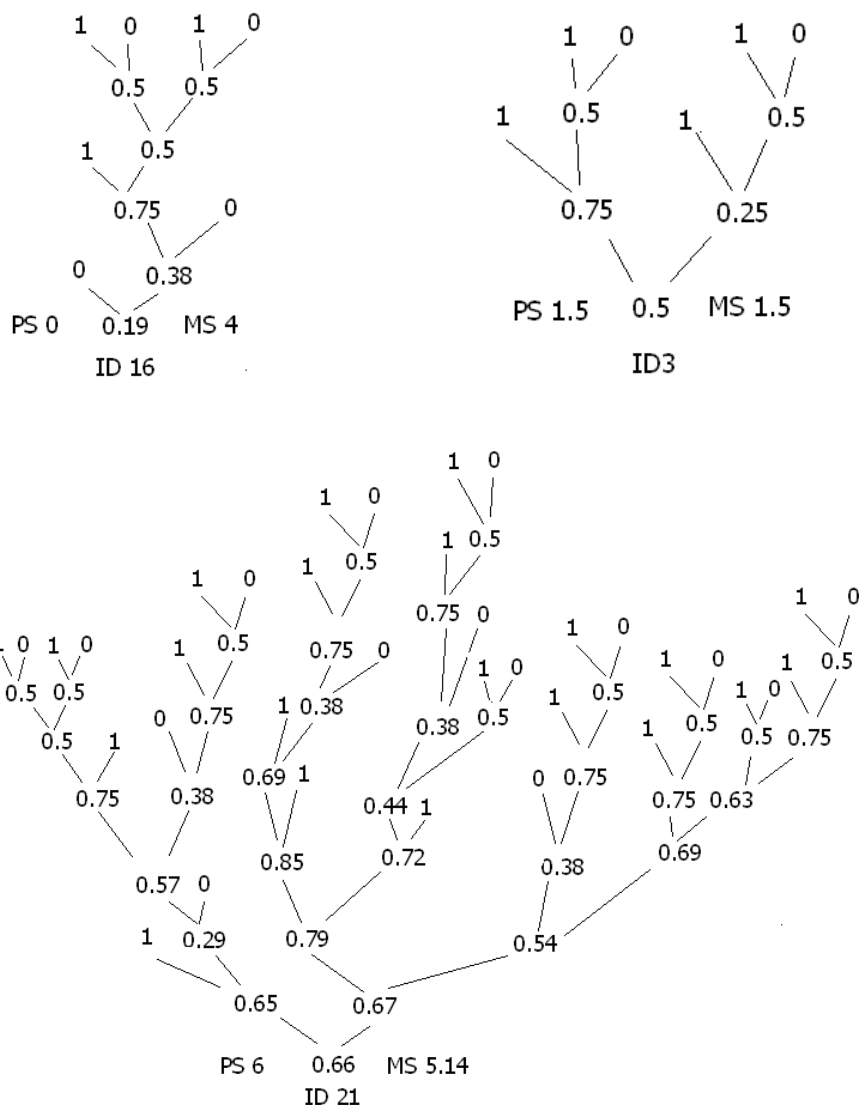
ID	ID16		ID3		ID21	
PedAP	0.19		0.50		0.66	
MolAP	0.40		0.50		0.75	
	P	M	P	M	P	M
$\tau$ ped	0.00	4.00	2.00	2.00	6.00	5.14
$\tau$ mol						
1	1.70	1.30	7.90	7.90	3.10	4.20
2	1.47	1.70	7.90	7.90	3.20	4.20
3	0.00	0.30	5.20	5.20	3.20	4.20
4	1.40	2.00	7.90	7.90	3.20	3.70
5	1.10	1.60	5.20	5.20	4.60	3.30
6	1.00	2.10	5.20	5.20	5.60	3.20
7	1.20	4.10	7.90	7.90	6.60	3.40
12	0.00	6.80	5.20	5.20	9.50	3.00

MolAP – Molecular admixture proportion calculated with STRUCTURE software; PedAP – admixture proportion calculated with pedigree;  $\tau$  ped – Estimated age of admixture from pedigree;  $\tau$  mol – Estimated age of admixture from SNP data; P – paternal; M – maternal

assumed to have happened five generations ago) is shown in Table 1. Of 21 investigated cases, results from SNP data reflected paternal and maternal age of admixture well for 9 cases but provided results out of range for the other 12 cases. F1 animals were sometimes not recognized by SABER and were assigned as older crosses. From further analysis with reduced number of animals in the set we conclude that the method is sensitive to number of purebred animals included in calculation of age of admixture.

Conditional on model parameters, SABER models the ancestral states along the paternal and the maternal chromosomes as two independent and identical Markov processes. The paternal side of the genealogy and the maternal side of the genealogy may have different levels of admixture, and, therefore, the two processes are not necessarily identical. The assumption that matings are random with respect to ancestry is often violated. If animals have big admixture proportion of one breed (e.g. 0.88) we assume that from one parental side the animal will inherit long intact haplotypes of one breed and from the other parental side haplotypes will be shorter. Overestimation was observed with decreasing number of pure animals in the “training” set, using 30 or 10 pure animals each instead of the ~100 available for each breed. With increased admixture proportion impact of distance and big  $\tau$  ( $\tau$  is inverse of chromosome length and can be interpreted as time since admixture happened in generations if admixture happened once) will be smaller. Since in our case admixture often happened more than once  $\tau$  is not the best value to represent all the events.

In Table 2 three animals with different proportions of admixture and different age of admixture are presented. In Fig. 1 corresponding pedigrees are presented. From the different initial values we can obtain sensitivity of the method to the different values. For the 50:50 cross we can see that the algorithm is heavily overestimating  $\tau$ . The algorithm was not able to detect (in most cases) when admixture happened on one side of pedigree. It estimated quite well with a bit of overestimation the time since admixture for other side of the pedigree. Pugach *et al.* (2011) described a principal component analysis-based genome scan approach to analyze genome-wide and local admixture structure, and introduced wavelet transform analysis as a method for estimating the time of admixture. As this method assumes a single event of admixture, we do not feel that this approach would add accuracy in our case of multiple admixture events. A solution to overcome these obstacles would be development of a more sensitive model which could deal with more than one event of admixture.



**Figure 1:** Genealogical information of admixture for animals whose ages of admixture are present in Table 2 (numbers presented are contribution of Red Holstein Friesian). PS – age of admixture from paternal side derived from pedigree; MS – age of admixture from maternal side derived from pedigree.

#### 4 ACKNOWLEDGEMENTS

We want to thank to the European Science Foundation (ESF) for providing the possibility to AF to spend two months at the Unit of Animal Genomics (GIGA-R, University of Liège) to work with TD.

#### 5 REFERENCES

Falush D., Stephens M., Pritchard J. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164: 1567–1587

Freeman A. R., Hoggart C. J., Hanotte O., Bradley D. G. 2006.

Assessing the Relative Ages of Admixture in the Bovine Hybrid Zones of Africa and the Near East Using X Chromosome Haplotype Mosaicism. *Genetics*, 173: 1503–1510

Frkonda A., Gredler B., Schnyder U., Curik I., Solkner J. 2012. Prediction of breed composition in an admixed cattle population. *Animal genetics* (in press). DOI: 10.1111/j.1365-2052.2012.02345.x

Illumina. 2009. Bovine SNP50 Genotyping BeadChip. [http://www.illumina.com/documents/products/datasheets/datasheet\\_bovine\\_snp50.pdf](http://www.illumina.com/documents/products/datasheets/datasheet_bovine_snp50.pdf) (19 Dec. 2010)

Pritchard J. K., Wen X., Falush D. 2010. Documentation for Structure software, Version 2.3. [http://pritch.bsd.uchicago.edu/structure\\_software/release\\_versions/v2.3.3/structure\\_doc.pdf](http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.3/structure_doc.pdf) (14 Jan. 2011)

Pritchard J., Stephens M., Donnelly P. 2000. Inference of popu-

- lation structure using multilocus genotype data. *Genetics*, 155: 945–959
- Pugach I., Matveyev R., Wollstein A., Kayser M., Stoneking M. 2011. Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biology*. 12: R19.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81, 3: 559–575
- Tang H., Choudhry S., Mei R., Morgan M., Rodriguez-Cintron W., Gonza'lez Burchard E., Risch, N.J. 2007. Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. *The American Journal of Human Genetics*, 81: 626–633