

Prediction of breed composition in an admixed cattle population

Frkonja Anamarija, Gredler Birgit,
Schnyder Urs, Curik Ino, Sölkner Johann

EAAP, Stavanger, August 31th

Session 30/7

Introduction

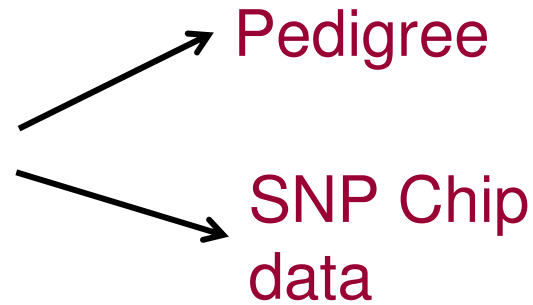
- What are the proportions of particular breeds in this animal?



EAAP, Stavanger, August 31th

Introduction

- Levels of crossbreeding



- Methods implemented for predicting levels of admixture:

HMM methods

Genomic selection methods

Aim

- Compare the performance of
 - HMM (**STRUCTURE** software)
 - Methods used in genomic selection:
 - Bayes B**
 - partial least squares regression (**PLSR**)
 - least absolute shrinkage and selection operator (**LASSO**)
- Results were correlated with pedigree based admixture calculations

Data

- Swiss Fleckvieh – genotyped Illumina 54k-Chip
- Cross/composite: Simmental and Red Holstein Friesian
- Genotypes of founder animals- not available



EAAP, Stavanger, August 31th



Selection of SNPs

- Full set of 40492 SNPs
- Subsets:
 - 50% (20246),
 - 20% (8098),
 - 10% (4049),
 - 5% (2024)
 - 1% (404) evenly distributed SNPs

Selection of SNPs

- F_{ST} chosen SNPs
 - > 25% (5635 SNPs),
 - > 30% (3904 SNPs),
 - > 35% (2620 SNPs)
- 48 and 96 SNPs with the highest F_{ST}

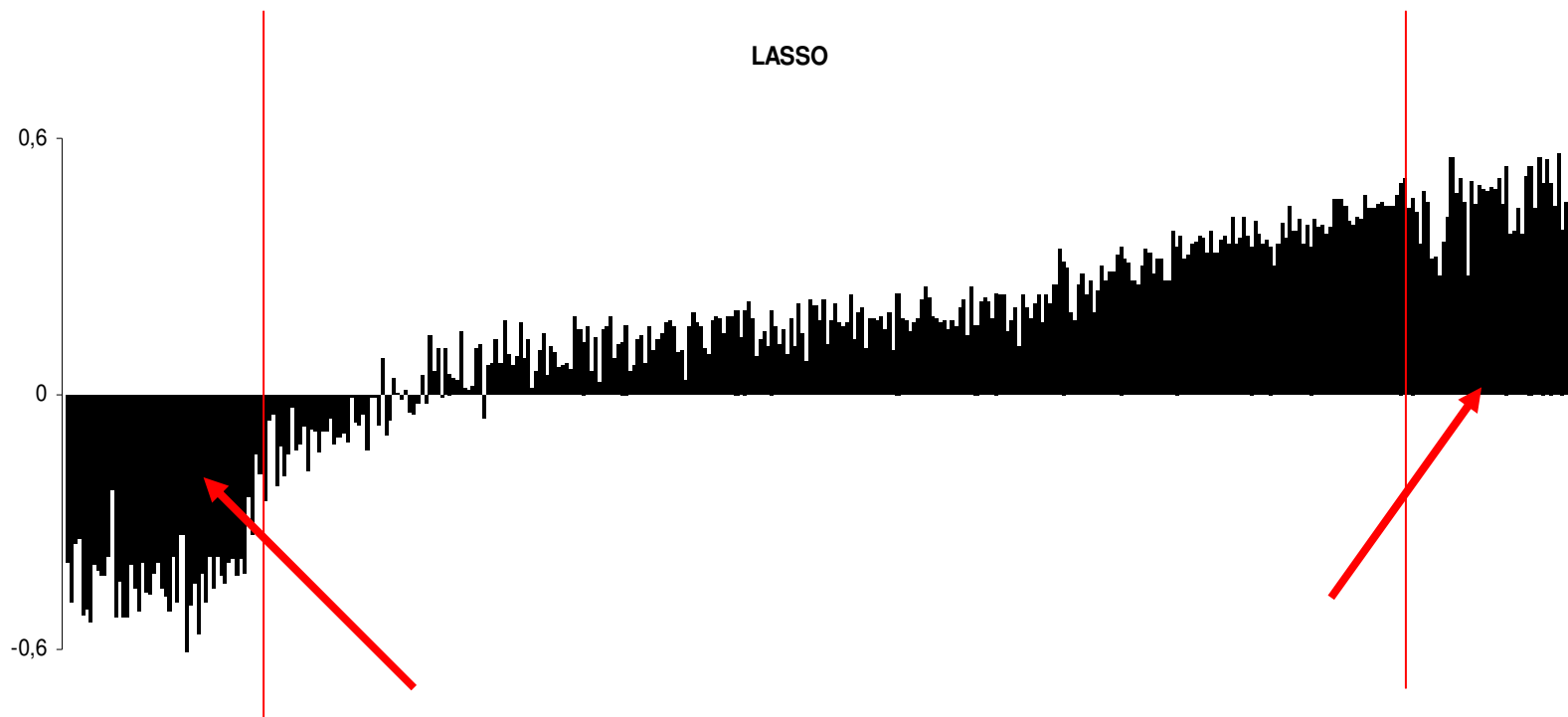


Selection of SNPs

- SNPs chromosomes:
1, 8, 15, 22 and 29
- Set of SNPs from chromosomes
1+15+29

Results

- Correlations of results obtained with the listed methods and pedigree based calculations



EAAP, Stavanger, August 31th

Results of evenly spaced SNPs

- Full set of SNPs- similar performance of methods: ~ 0.97 (except LASSO 0.93)
- Evenly spaced SNPs down to 2024 SNPs- results not significantly different for STRUCTURE, PLSR, LASSO
- BayesB breaking point - 4049 SNPs

Results of F_{ST} and chromosome picked SNPs

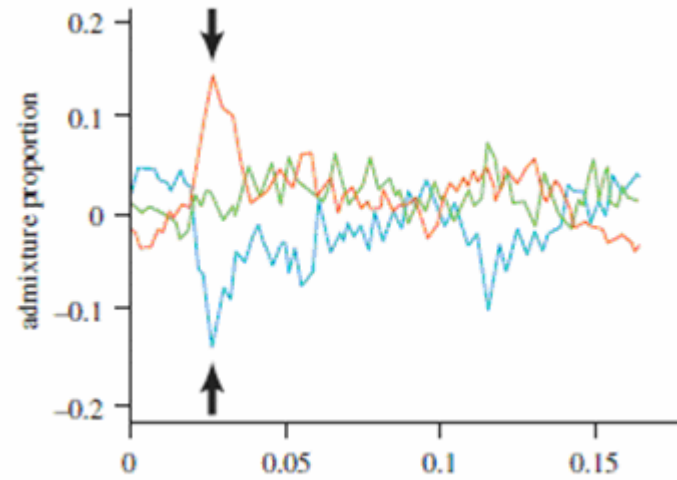
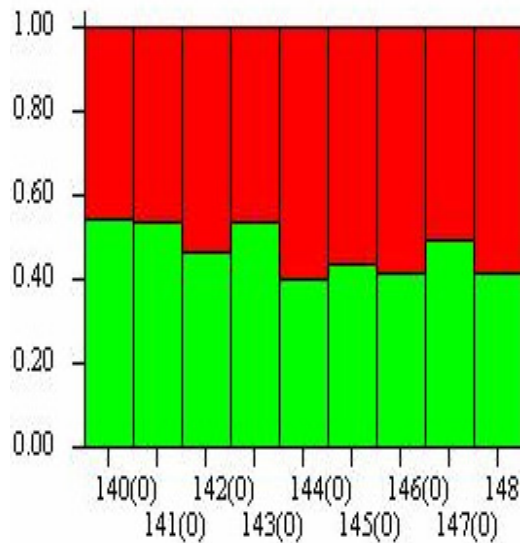
- F_{ST} chosen SNPs- results very similar to using the full set (except highest F_{ST} 48, 96 SNPs)
- LASSO- no reliable results obtained
- Full sets of SNPs extracted from individual chromosomes- ~ 0.73
- STRUCTURE - best performing method
- Combined set of 3 chromosomes ~ 0.85

Conclusion

- All methods except LASSO with full set of SNPs **gave almost identical results**
- FST chosen SNPs- results very similar to using the full set
- Using 96 and 48 SNPs with the highest FST, correlations with pedigree based admixture calculations dropped
- Particular chromosomes separately did not give satisfactory results

Future research

- Admixture of particular regions on the chromosomes



Oleksyk et al. (2010)

EAAP, Stavanger, August 31th



Acknowledgements

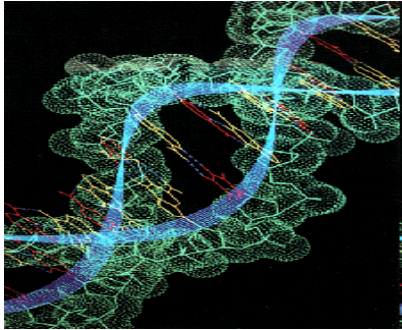
- Swiss Fleckvieh genotypes were provided by swissherdbook cooperative

Thank you for
your attention



Conclusion

- All methods except LASSO with full set of SNPs **gave almost identical results**
- FST chosen SNPs- results very similar to using the full set
- Using 96 and 48 SNPs with the highest FST, correlations with pedigree based admixture calculations dropped
- Particular chromosomes separately did not give satisfactory results



Results

Prediction of breed composition in an admixed cattle population

Frkonja Anamarija, Gredler Birgit,
Schnyder Urs, Curik Ino, Sölkner Johann

Number of SNPs

STRUCTURE

PLSR

BayesB (1% important SNPs assumed)

LASSO (96)

40492

_a0.972^a

_a0.976^a

_a0.974^a

_a0.934^b

20246

_a0.971^a

_{ab}0.968^a

_{ab}0.972^a

_b0.907^b

8098

_a0.969^a

_{ab}0.968^a

_{ab}0.966^a

_b0.907^b

4049

_a0.969^a

_a0.974^a

_b0.961^a

_{ab}0.920^b

3k (2694)

0.957

0.965

0.905

0.901

2024

_b0.950^a

_{ab}0.966^a

_c0.890^b

_{cb}0.892^b

404

_b0.929^a

_b0.955^b

_c0.860^c

_c0.847^c

F_{ST} subsets

5635 ($F_{ST} > 0.25$)	${}_a0.97^a$	${}_a0.976^a$	${}_{ab}0.966^a$	${}_a0.934^b$
3904 ($F_{ST} > 0.30$)	${}_a0.971^a$	${}_a0.974^a$	${}_{ab}0.965^a$	${}_a0.934^b$
2620 ($F_{ST} > 0.35$)	${}_a0.969^a$	${}_a0.974^a$	${}_{ac}0.952^b$	${}_a0.934^b$
1677 ($F_{ST} > 0.40$)	${}_a0.968^a$	${}_a0.973^a$	${}_c0.949^b$	${}_a0.934^b$
1028 ($F_{ST} > 0.45$)	${}_a0.966^a$	${}_{ab}0.968^a$	${}_b0.967^a$	${}_a0.934^b$
594 ($F_{ST} > 0.50$)	${}_a0.961^a$	${}_b0.957^a$	${}_{bc}0.955^{ab}$	${}_a0.934^b$
96 ($F_{ST} > 0.623$)	${}_b0.924$	${}_c0.916$	${}_d0.926$	0.918
48 ($F_{ST} > 0.651$)	${}_b0.907$	${}_c0.903$	${}_d0.906$	0.903

Chromosome subsets

1 (2115 SNP)	0.794	0.739	0.699	0.724
8 (1568 SNP)	0.716	0.711	0.690	0.701
15 (1077 SNP)	0.747	0.744	0.678	0.712
22 (586 SNP)	0.683	0.667	0.636	0.669
29 (657 SNP)	0.638	0.654	0.665	0.643
1/15/29 (3867 SNP)	0.878 a	0.892 a	0.863 a	0.766 b

Residuals (SNP based admixture – pedigree based admixture calculations) for pure animals represented as admixed using different methods

		Red Friesian		Simmental	
		mean	sd	mean	sd
STRUCTURE	$F_{ST} = 0.30$	-0.070	0.043	0.030	0.027
STRUCTURE	2024 SNPs	-0.089	0.060	0.031	0.028
BayesB	$F_{ST} = 0.30$	-0.030	0.049	0.010	0.034
BayesB	2024 SNPs	-0.055	0.056	0.019	0.045
LASSO	$F_{ST} = 0.30$	-0.046	0.070	0.048	0.082
LASSO	2024 SNPs	-0.042	0.105	0.041	0.074
PLSR	$F_{ST} = 0.30$	-0.025	0.041	0.004	0.029
PLSR	2024 SNPs	-0.050	0.052	0.019	0.038

- Pearson correlations of admixture based on pedigree and STRUCTURE and PLSR with different sets of pure animals in the training sets
- For sets of 10 animals each, 5 samples were run, numbers in brackets indicate the range of correlations

Number of pure animals per breed	STRUCTURE	PLSR
100	0.971	0.974
50	0.971	0.973
20	0.971	0.971
10	0.970 (0.9702-0.9704)	0.970 (0.969-0.971)
