Local vs. Global Ancestry: regions deviating from genome wide admixture in a composite cattle breed

N. Khayatzadeh, G. Mészáros, Y.T. Utsunomiya, J.F. Garcia,
U. Schnyder, B. Gredler, I. Curik, J. Sölkner

Introduction

Admixture

❑ Genetic exchange between two or more previously separated populations

❑ Break up in the genome as a result of recombinations at meiosis

❑ Mosaic of ancestral segments after generations in the genome of admixed individuals

Timing of admixture depend on the size of on-recombined chromosomal segments

genome of admixed individual
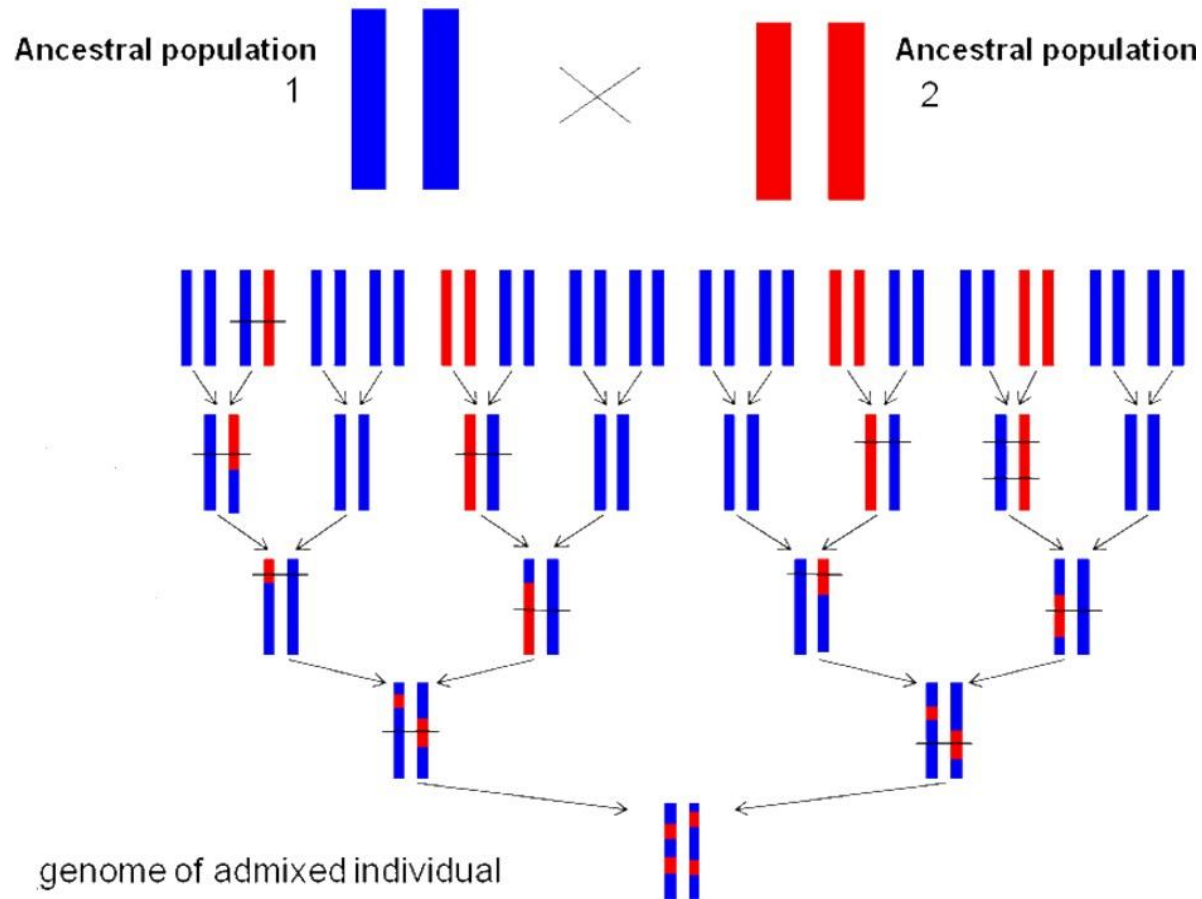
Admixture levels

❑ **Global genetic ancestry:** ancestry proportions averaged across the genome of an individual

❑ **Local genetic ancestry:** ancestral origin of distinct chromosomal segments within an individual genome

▪ Variation in admixture proportions among loci: deviation of local from global estimates

Sources of variations

- ❑ **Genetic drift** with influence on the whole genome

- ❑ **Selection**

  - ▪ Effect on specific gene regions

  - ▪ Increase in the frequency of the desired alleles

  - ▪ Cause of excess or deficiency of ancestral proportions

Aims

- ❑ Extremes as an indicator of selection signatures

- ❑ Statistical methods to define the significant threshold for signals

Materials and Methods

❑ Illumina Bovine SNP50 BeadChip data from Swissherdbook cooperative Zollikofen

91	300	101
Simmental	Admixed	Red Holstein Friesian

- **Swiss Fleckvieh** 0.125-0.875 RHF ancestry (Swissherdbook)
- **Admixed Animals** 0.020-0.990 RHF ancestry (this study)

❑ **Standard Quality Control** 39,525 SNPs and 485 animals after QC (PLINK 1.07)

- **ADMIXTURE** (Alexander et al., 2009) unsupervised global ancestry proportion estimation

- **LAMP** (Sankararaman et al., 2008) estimation of ancestry for every locus along each chromosome

Δ ancestry deviations of the locus-specific ancestry averaged across the admixed animals from the genome wide ancestry (Tang et al., 2007)

$$\delta_k^m = \frac{1}{I} \sum_{i=1}^{I} (q_k^{i,m} - \bar{q}_k^i) = \tilde{q}_k^m - \bar{q}_k$$

I: the number of admixed animals (1,...,i); m: number of SNPs (1,...,k); q: local admixture

Definition of the significance threshold

❑ **Normal distribution hypothesis tests**

▪ Scaling local admixture by standard deviation (0.040)

▪ Admixture LD in the genome of admixed populations

▪ Assuming 1000 and 5000 independent segments, Bonferroni correction test (Bhatia et al., 2014)

▪ Scaled local ancestry > 4.06 SD (p < 1 × 10^{-5}): 1000 independent tests

> 4.42 SD (p < 5 × 10^{-5}): 5000 independent tests

□ **Permutations tests** (Doerge & Churchill 1996)

- Concatenating Δ ancestries, cutting the circularized genome randomly and rearranging the two resulting pieces

- Considering 5% (-0.17 to 0.17)and 1 % (-0.16 to 0.15) of permutation tests, as the extreme deviations of Δ ancestries from normal distribution.
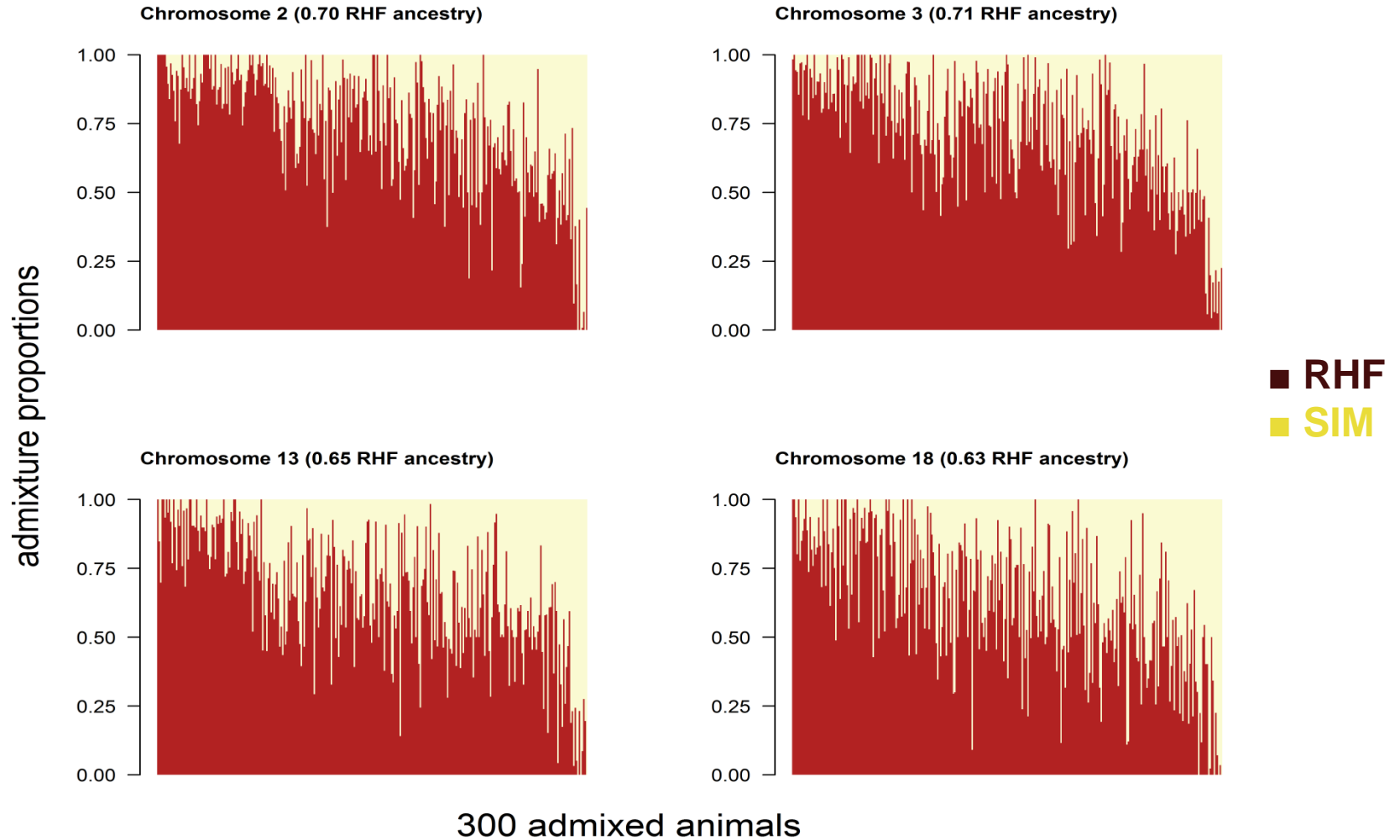
Results

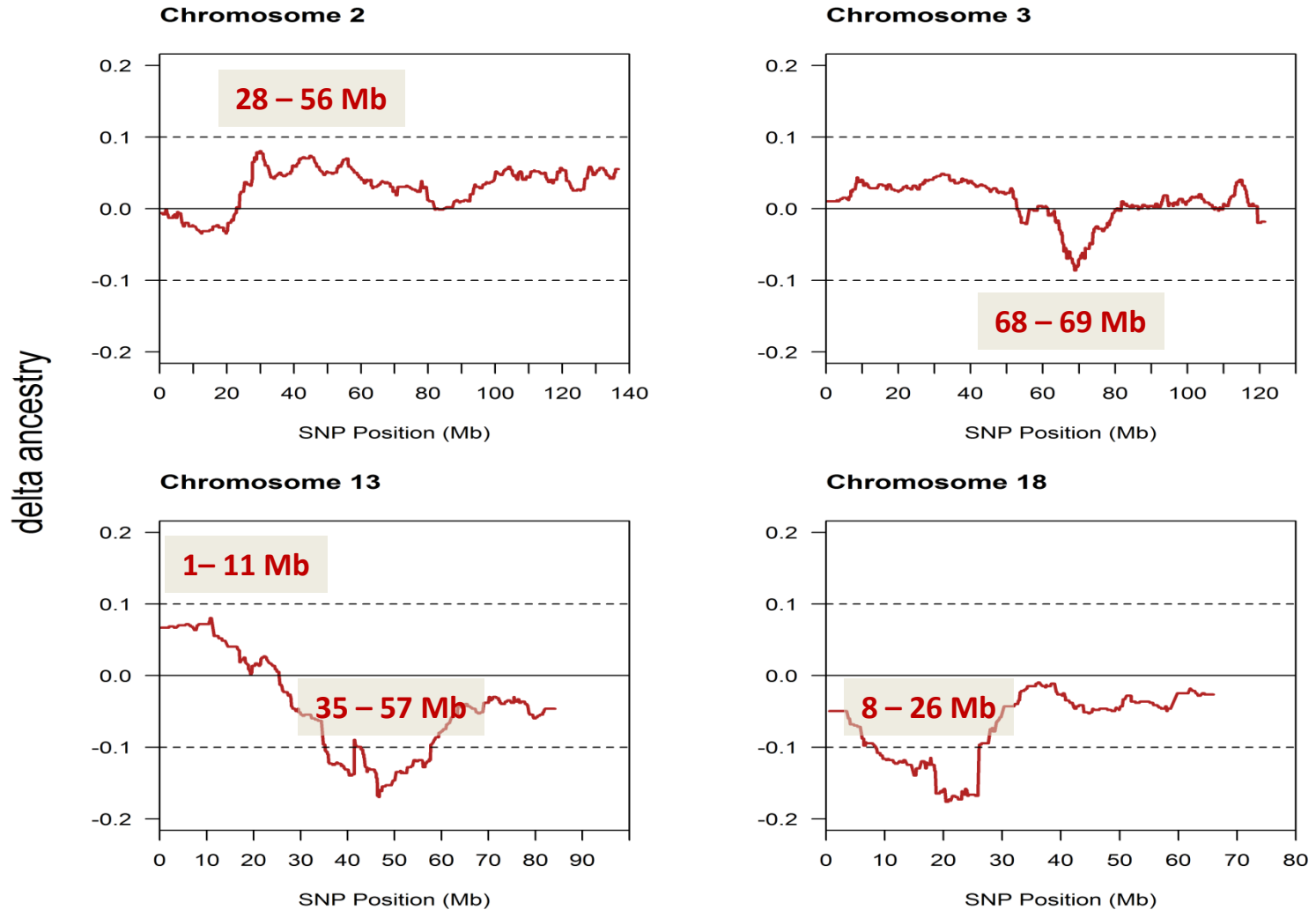Global admixture across the admixed animals calculated by

ADMIXTURE and LAMP

The average RHF ancestry proportion by ADMIXTURE and LAMP

Method	Ancestry Proportion
ADMIXTURE	0.68 (0.19 SD)
LAMP	0.70 (0.04 SD)

Average ancestry estimates (LAMP)

Chromosome 2 (0.70 RHF ancestry)

Chromosome 3 (0.71 RHF ancestry)

Chromosome 13 (0.65 RHF ancestry)

Chromosome 18 (0.63 RHF ancestry)

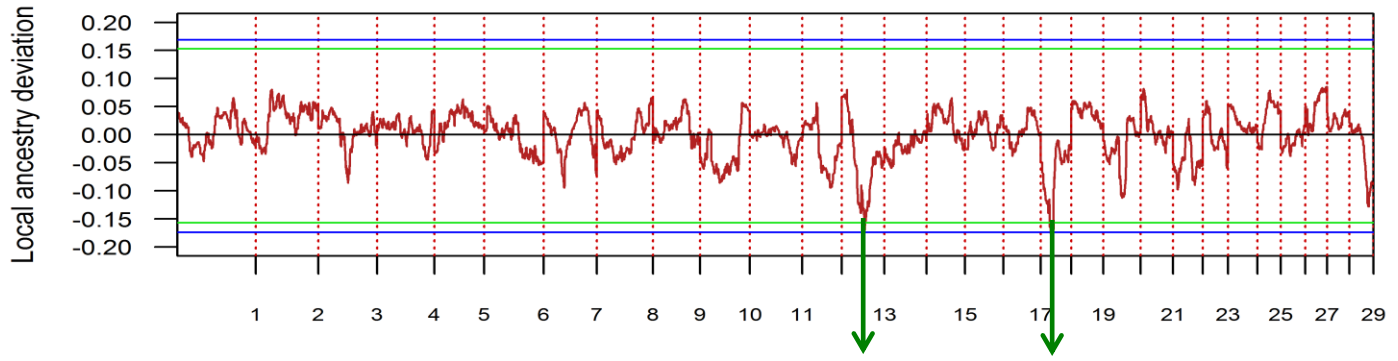admixture proportions

300 admixed animals

■ **RHF**
■ **SIM**

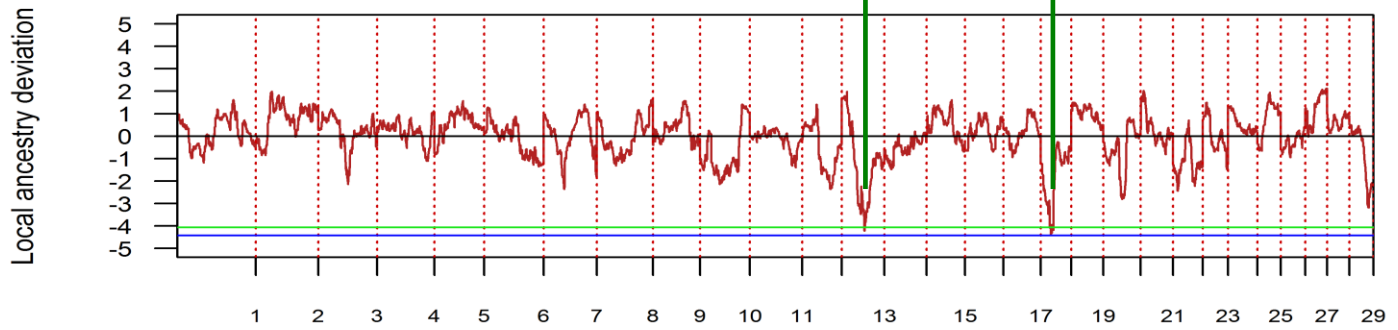# Genome wide variation of RHF ancestry at SNP level

Local ancestry deviations based on the permutation and multiple hypothesis tests

(a)

(b)

46.3-47.3 Mb (Chr13)

18.7-25.9 Mb (Chr18)

Discussion

- **Significant peaks**
  - In direction of increased Simmental ancestry
  - Selection signatures for very recent selection

- **Relatively long segments**
  - Limited recombination events due to small number (~10) of generations after admixture
  - Not enough time for selection to sharpen the signals

- Both hypothesis tests are conservative enough to detect the selection signatures.

Conclusions

❏ Strong excess and deficiency of local ancestry from global estimation is not likely by chance

❏ Long segments of excess/deficiency as the indicator of recent admixture

❏ Hypothesis tests applied here are conservative enough to detect the significant selection signatures

❏ Wideness of signals results in inclusion of many candidate genes in this particular composite

Thank you your attention