Characterization of additive, dominance, and runs of homozygosity effects inbreeds of dairy cattle

Hafedh Ben Zaabza[1,2], Mahesh Neupane[2], Mohd A. Jaafar[3], Srikanth Krishnamoorthy[3], Stephanie McKay[1], Asha Miles[2], Heather J. Huson[3], Ismo Strandén[4], Harvey Blackburn[5], Curtis P. Van Tassell[2]

[1]Department of Animal and Veterinary Sciences, University of Vermont, Burlington, VT, 05405
[2]Animal Genomics and Improvement, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705
[3]Department of Animal Science, Cornell University, Ithaca, NY 14853
[4]Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland
[5]National Center for Genetic Resources Preservation, Fort Collins, CO 80521

Email: Hafedh.BenZaabza@usda.gov

Background

- There is little consensus surrounding the risk of inbreeding to pure breeds of dairy cattle, and the dynamics of inbreeding at the genomic level remain largely unclear.
- Inbreeding in an individual is usually measured by an inbreeding coefficient based on pedigree information.
- The pedigree-based calculation of the inbreeding coefficient suffers from a lack of reliability because pedigree information is often incomplete and is rarely evaluated for inbreeding coming from ancestors of different generations.
- Several studies have shown that genome-based characterization provides a better measure of inbreeding than pedigree-based estimation.

Background

- Recent reports indicate that the characterization of inbreeding based on runs of homozygosity (ROH) gives a more accurate indication of individual homozygosity than either the pedigree or genomic-based inbreeding coefficient.
- Inbreeding fitness is usually expressed as the change in mean fitness due to inbreeding, $M0 - MF = -2\sum pqd$ (Falconer, 1976). Thus, inbreeding depression is correlated with the degree of dominance for a trait.
- We hypothesize that the characterization of dominance effects will improve our understanding of the genetic architecture of inbreeding depression and that ROH can provide an effective measure of genomic inbreeding.

Objective

- Characterize the genomic landscape of the regions associated with additive, dominance, and ROH effects in U.S. dairy cattle.

Methods: Data

- Pedigree records were obtained from the Council on Dairy Cattle Breeding (CDCB) and consisted of all known ancestors of the available genotyped animals, comprising a total of 4 million Holsteins.
- Over a million Holstein cows had both genotypes and phenotypes were used. All genotypes were imputed to approximately 78k SNPs using findhap.f90 software (VanRaden et al., 2011).
- Phenotypic data included 3 milk production traits (milk, fat, and protein yields), 3 fertility traits (daughter pregnancy rate, cow conception rate, heifer conception rate), and somatic cell score.

Computation

- All genotypes were imputed to approximately 78k SNPs using findhap.f90 software.
- Pedigree was pruned and sorted using relaX2.
- Quality control of the genomic data was performed using the preGSf90 program.
- Regions of homozygosity (ROH) were identified with the plink 2.0 software.
- ROH-based matrix was computed using a program written in Fortran 95.
- Estimation of variance component using GREML software
- GWAS analysis was conducted using WOMBAT
- Probability associated with solutions and t-statistics were calculated in R.

Identification of ROH

- The following criteria were used to define a ROH:
  - sliding window of 10 SNPs
  - A maximum of 1 heterozygous SNP within a window
  - minimum ROH length of 2 Mb
  - maximum gap between 2 sequential SNPs of 500 kb
  - minimum SNP density of 1 SNP/100 kb
  - minimum ROH length of 15 SNPs

Models
Single SNP-GWAS

The following model was used to identify inbreeding depression regions of interest:

$$\mathbf{YD} = \mathbf{1}\mu + \mathbf{u} + \mathbf{a}\alpha + \mathbf{d}\delta + \mathbf{r}\rho + \mathbf{e}$$

Where:

YD is a vector of yield deviations,

1 is a vector of ones,

μ is an overall mean,

u is a vector of additive genetic effects and $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \mathbf{A}\sigma_a^2)$,

a is a vector of allele counts (0, 1, or 2 for number of alternative alleles),

Models
Single SNP-GWAS

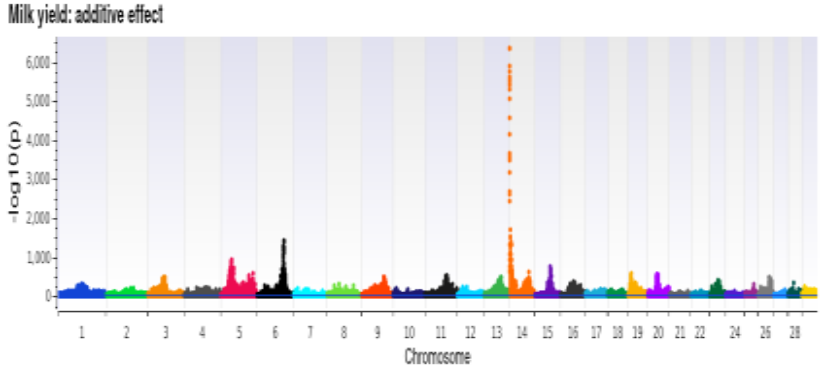
$$\mathbf{YD} = \mathbf{1}\mu + \mathbf{u} + \mathbf{a}\alpha + \mathbf{d}\delta + \mathbf{r}\rho + \mathbf{e}$$

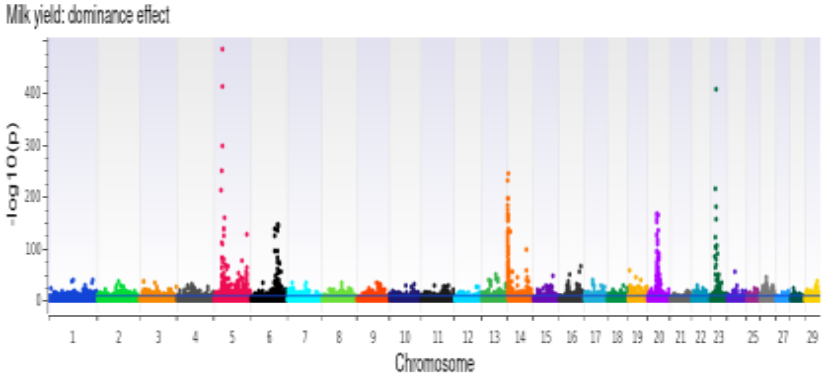
s a vector of dominance coefficients (0 fo
homozygotes, 1 for heterozygotes),

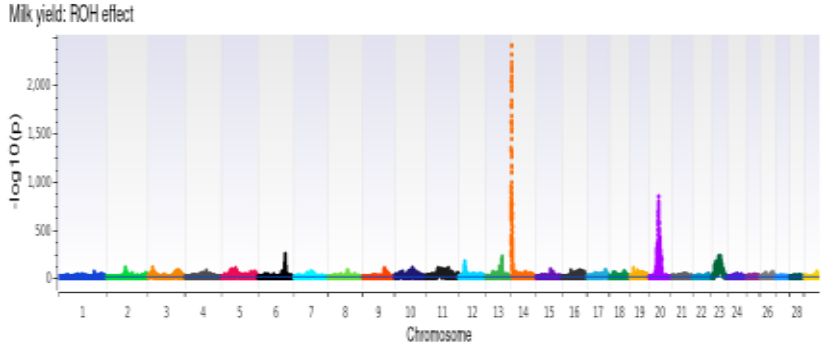
a vector of ROH coefficients (1 if SNP wa
ROH for individual, 0 otherwise),

s the additive effect of the SNP allele,

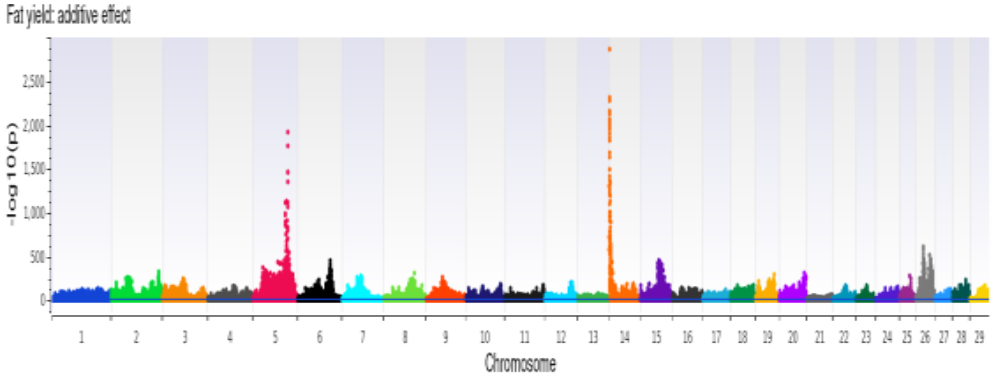
Milk yield

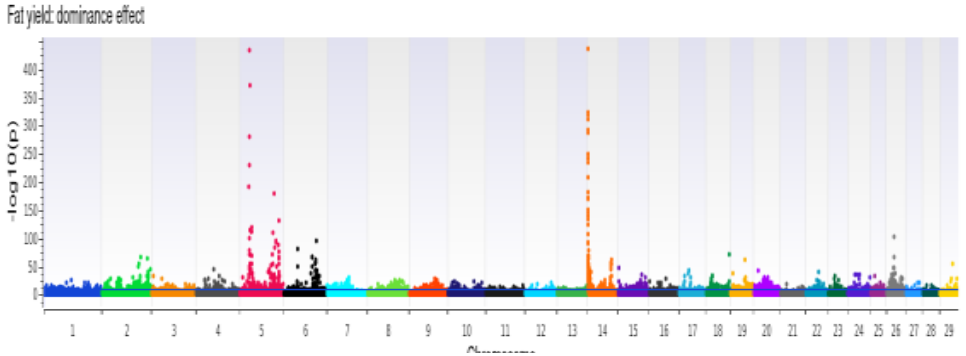


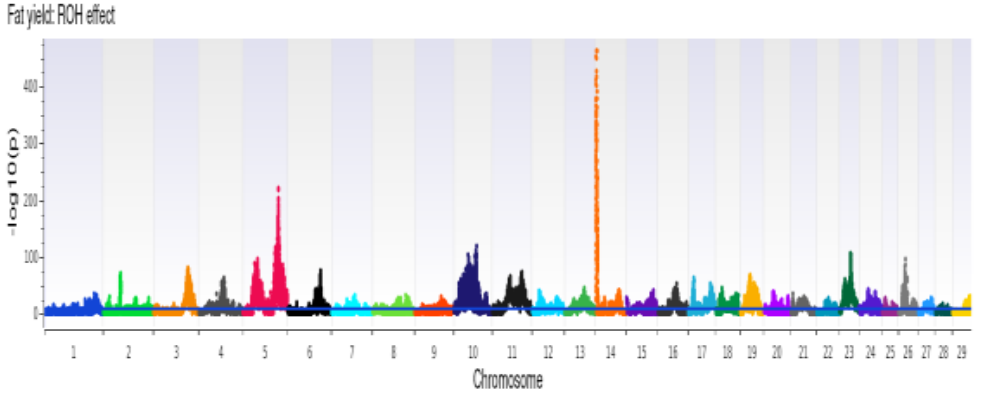




Fat yield

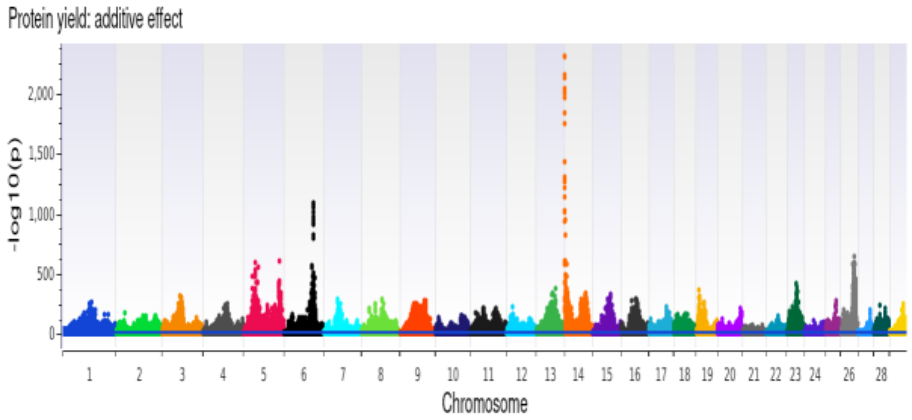


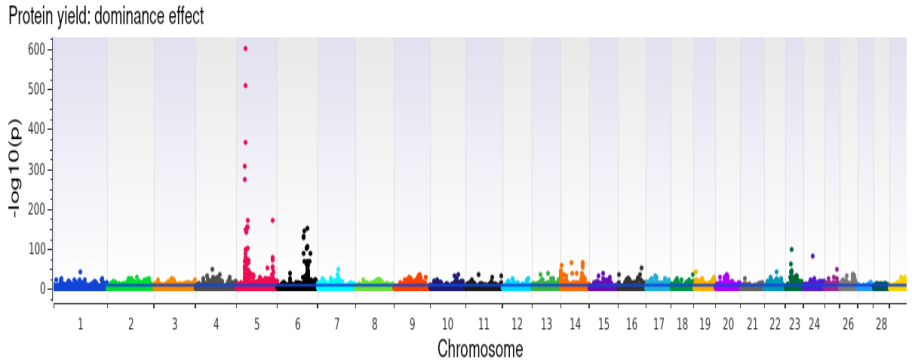


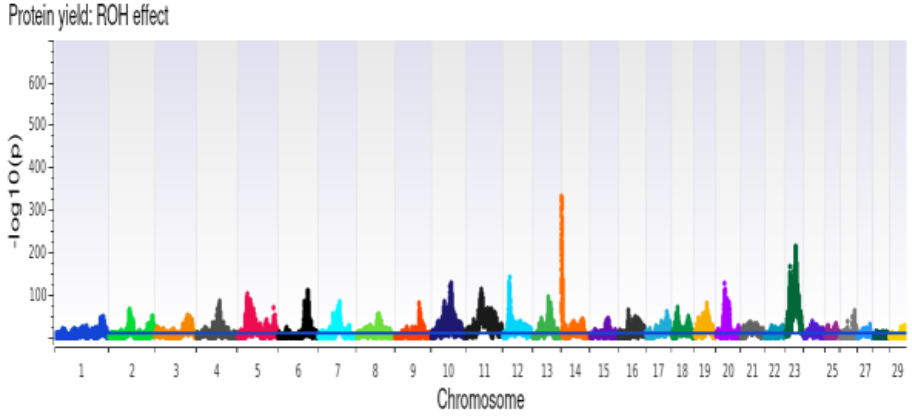


The horizontal line at log10(1/p)= 8 is the threshold for statistical significance of 5% genome-wide false positives.

Protein yield

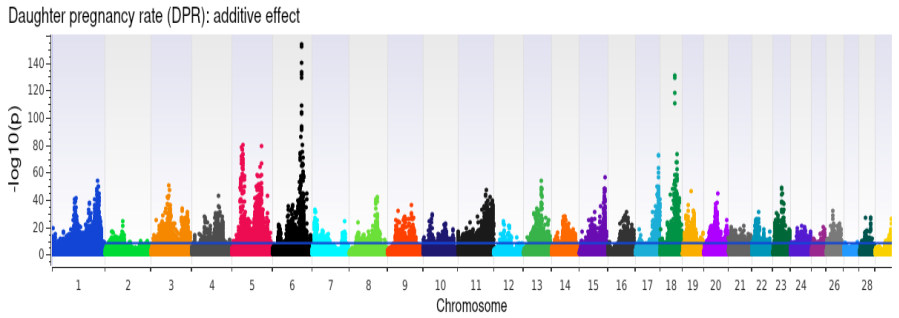


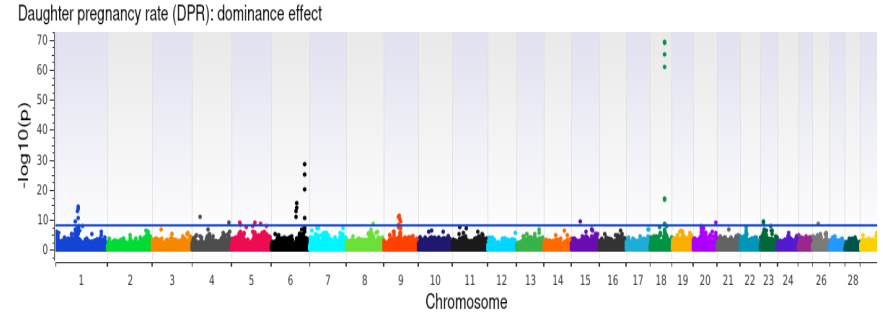


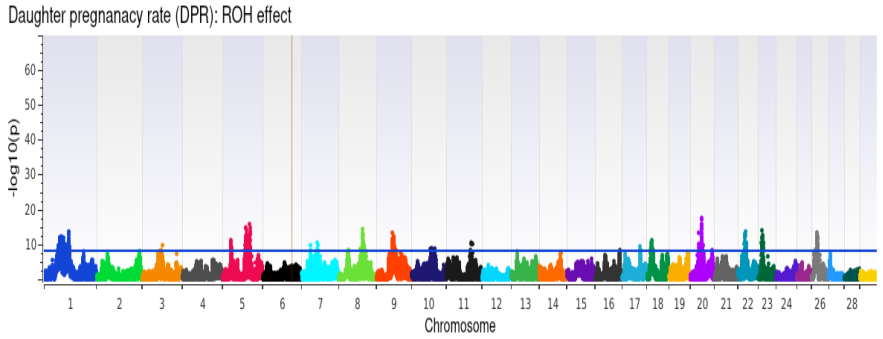


- For the three production traits, the SNPs in 0-1Mb region on BTA14 surrounding the DGAT1 gene had the largest additive effects.
- For fat, BTA5 had the most significant and the largest number of dominance SNP effects.
- For protein BTA14, BTA6, and BTA23 had the largest additive SNP effects.
- ➔ 3566 significant dominance effects were detected for the production traits – most of them are new effects and are in or near genes recognized to affect production traits.

Daughter pregnancy rate (DPR)

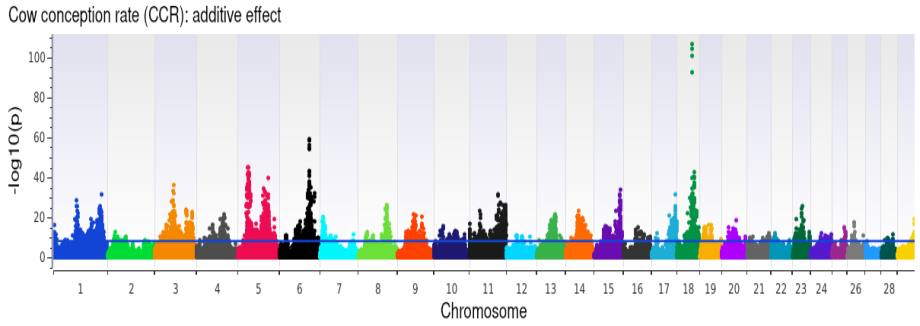


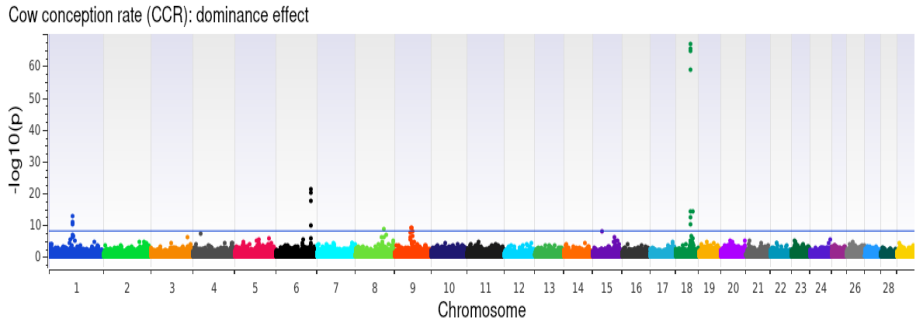


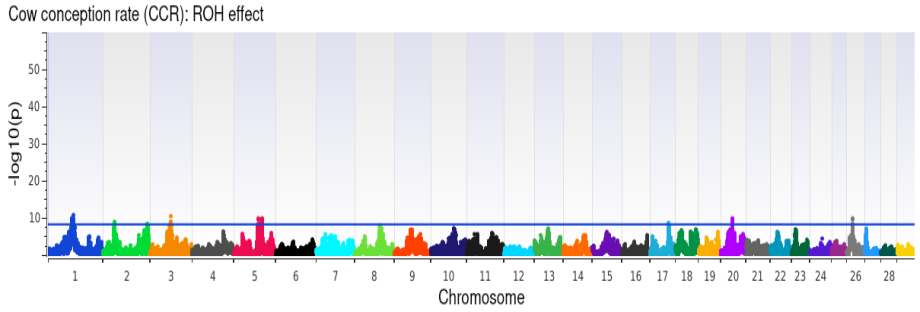


- 8501 significant additive effect, 37 significant dominance effect, and 616 significant ROH effect.
- The SNPs in the 83-90Mb region on BTA6 and in 43-51Mb on BTA18 had the most significant additive effect.
- The SNPs in the 68-70Mb region on BTA1, in the 77-79Mb region on BTA6, and in 43-50Mb region on BTA18 had the most significant dominance effects.
- ROH effects were detected with less statistical significance than the additive and dominance effects.

Cow conception rate (CCR)



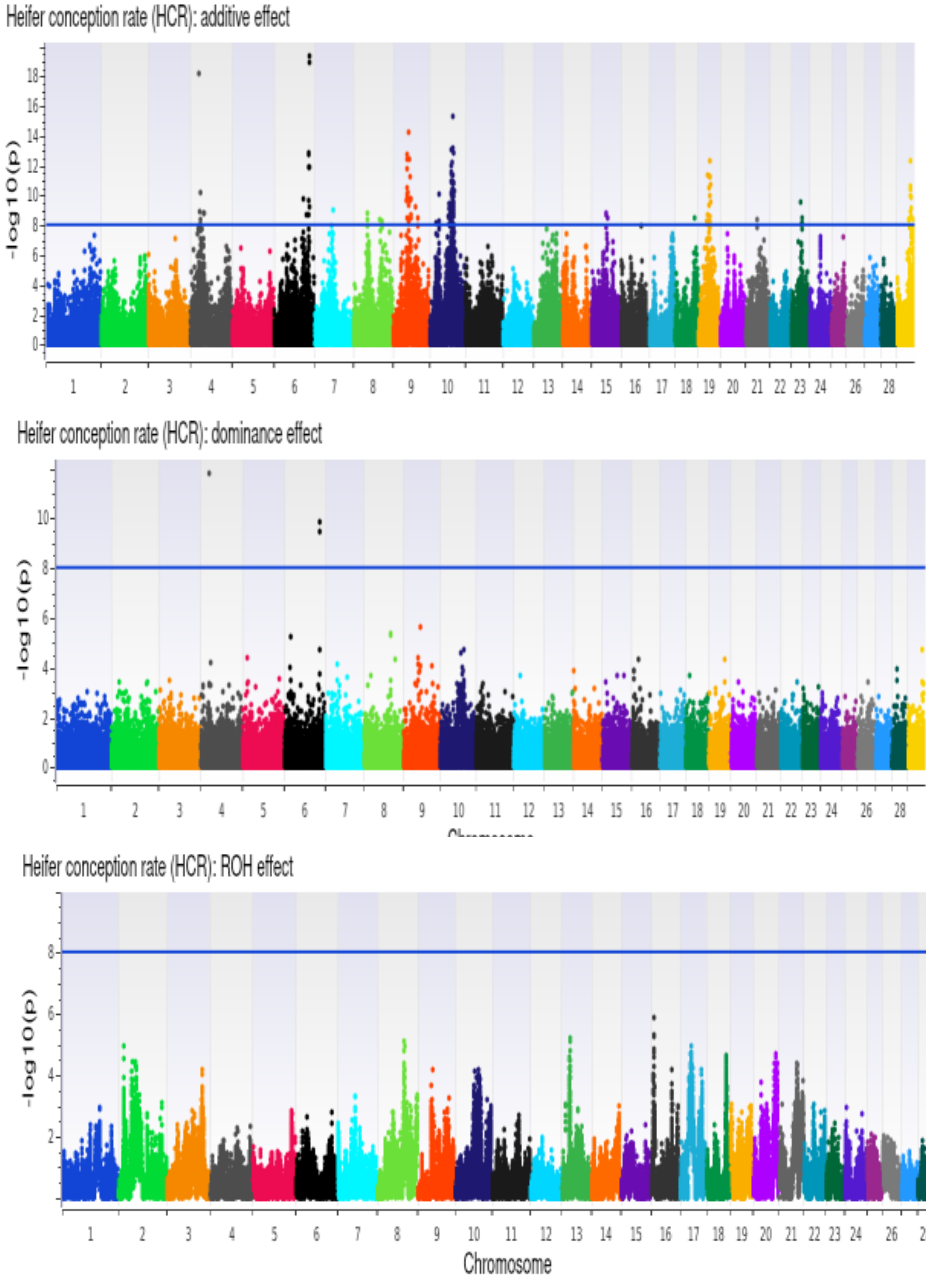




- 3567 significant additive effect, 20 significant dominance effect, 73 and significant ROH effect.
- The SNPs in the 83-93Mb region on BTA6 and in 43-51Mb on BTA18 had the most significant additive effect.
- The SNPs in the 68-69Mb region on BTA1, in the 101-102Mb region on BTA6, and in 43-50Mb region on BTA18 had the most significant dominance effects.
- ROH effects were detected with much less statistical significance than the additive and dominance effects.

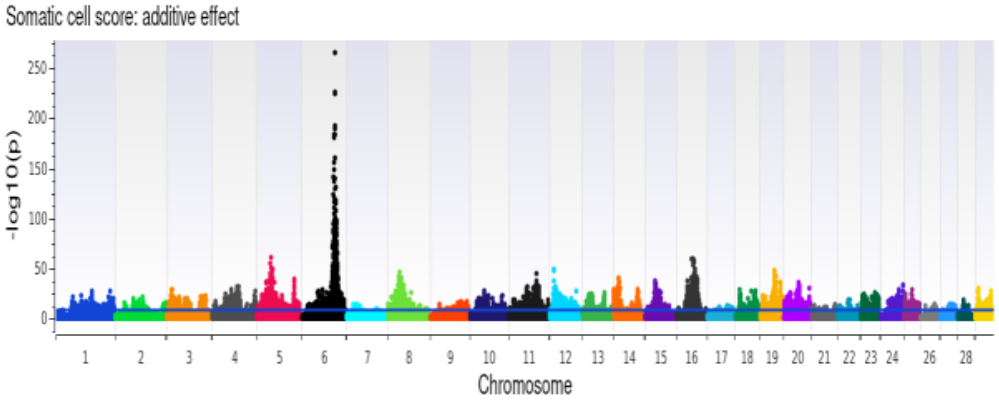
The top significant SNPs in DPR are found to be significant in CCR as well, but with different ranks.

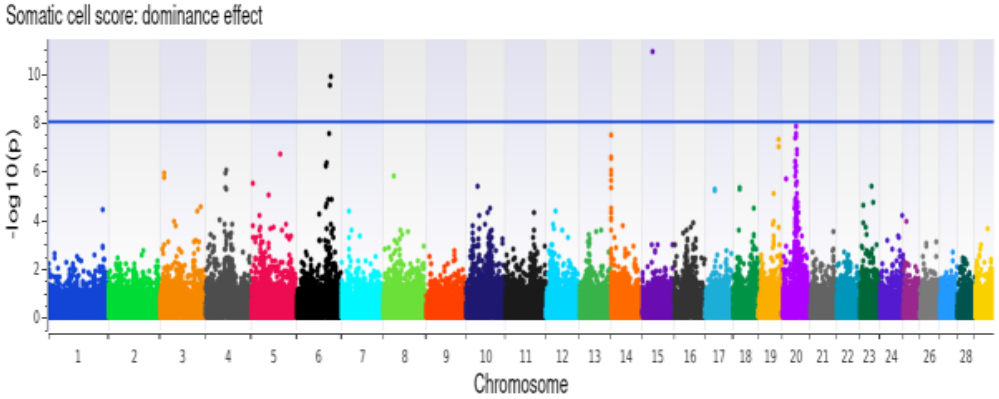
Heifer conception rate (HCR)

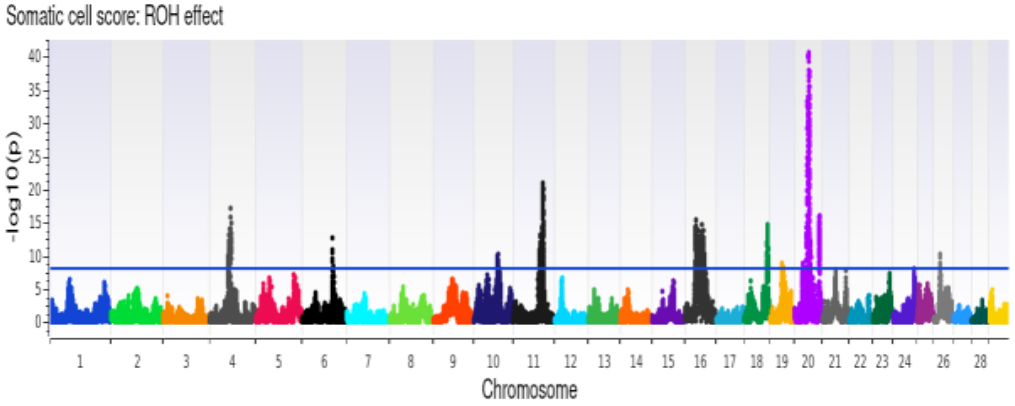


- A small number of dominance effects were detected for the HCR with lower statistical significance than the additive effects.
- No significant ROH were found.
- Interestingly, the 2 SNPs "BTB-00277427" and "chr6_103844298" are the top 2 SNPs on BTA6 for dominance effects for DPR, CCR, and HCR.

Somatic cell score



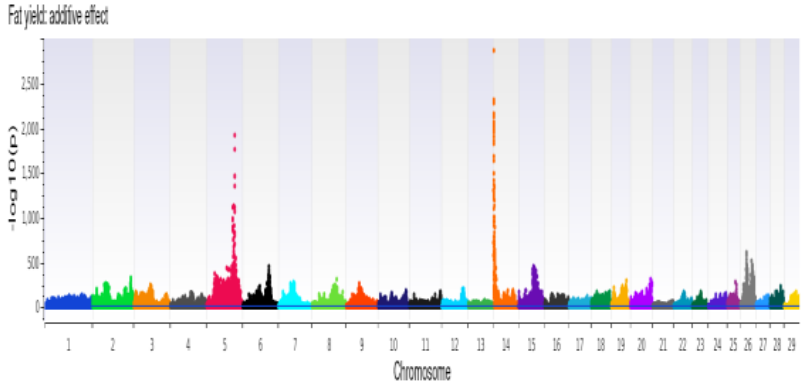


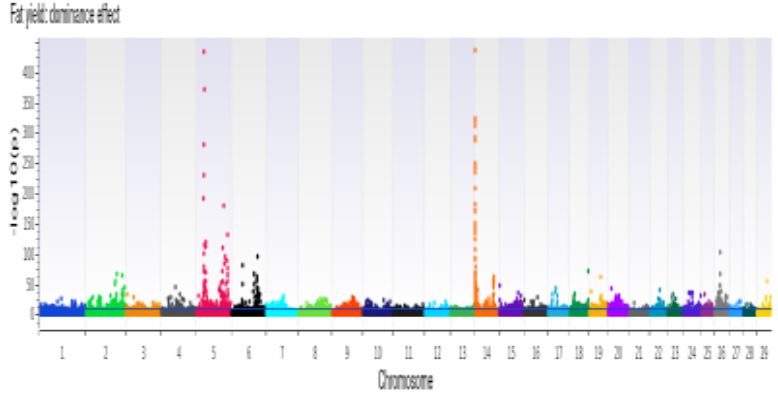


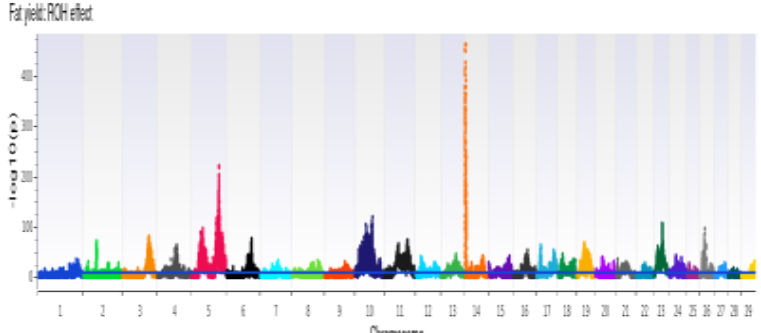
- For somatic cell score, BTA6 and BTA15 had the most significant dominance SNP effects.
- NO dominance effects exceeded the Bonferroni significance ($p<1e-07$) in Jian et al.(2019) using ADD-DOM model on U.S Holstein Cattle.

3-covariable model (ADD-DOM-ROH) vs. 2-covariable model (ADD-DOM): Fat yield

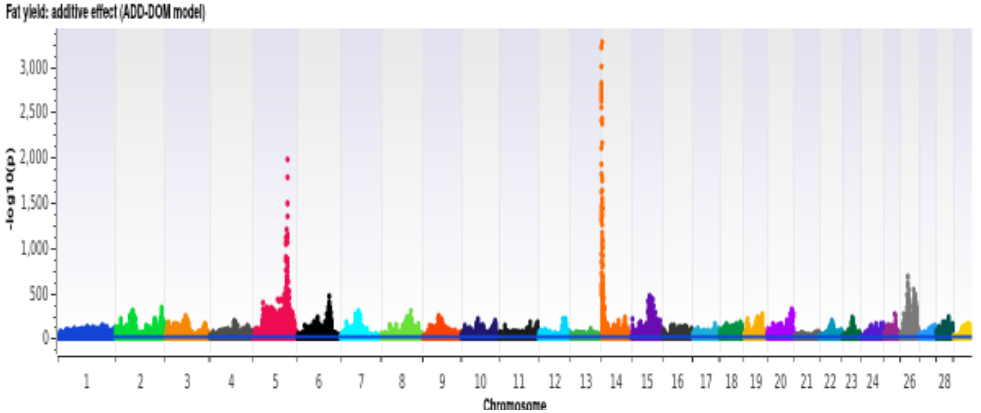
ADD-DOM-ROH model

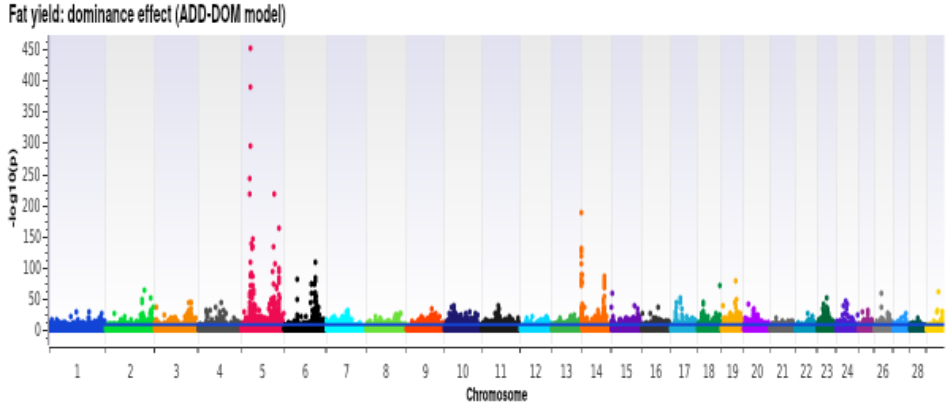






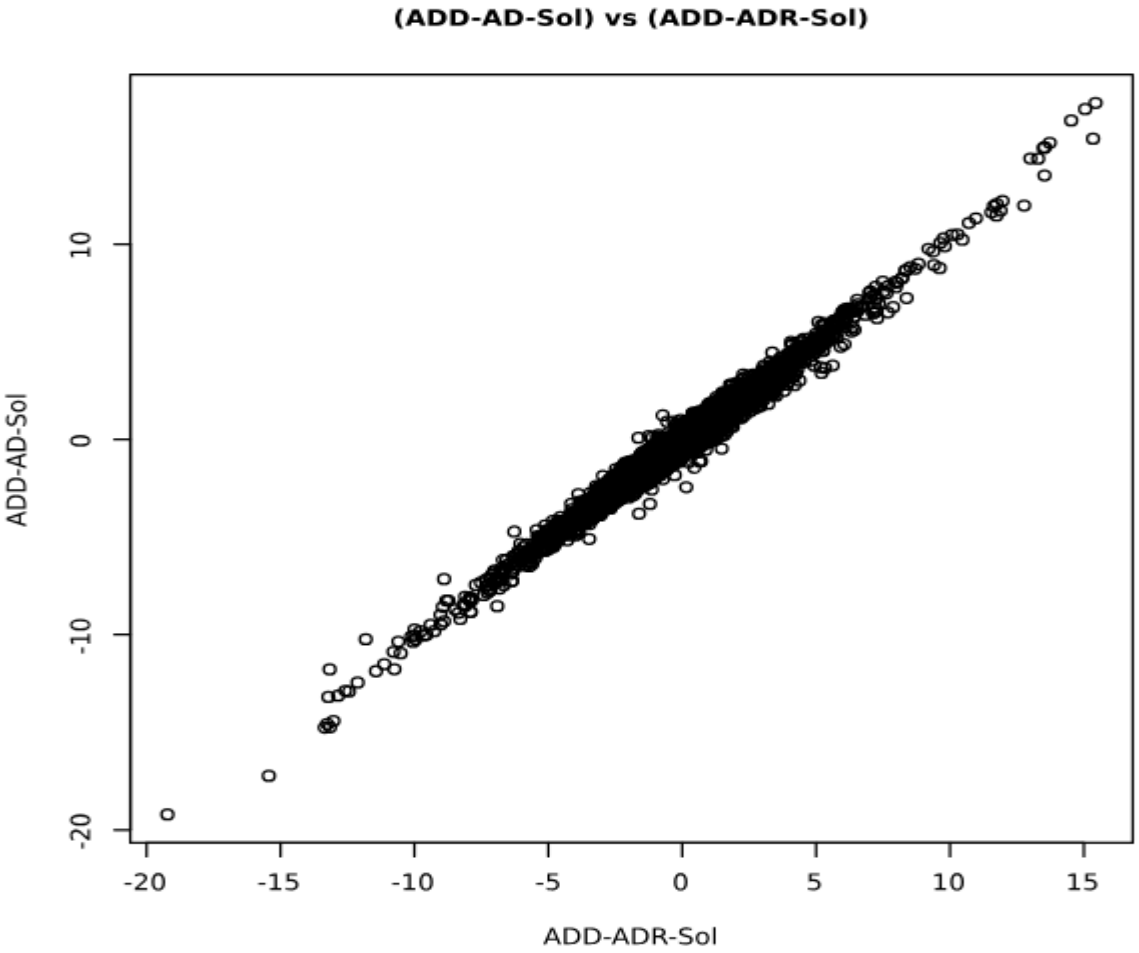
ADD-DOM model





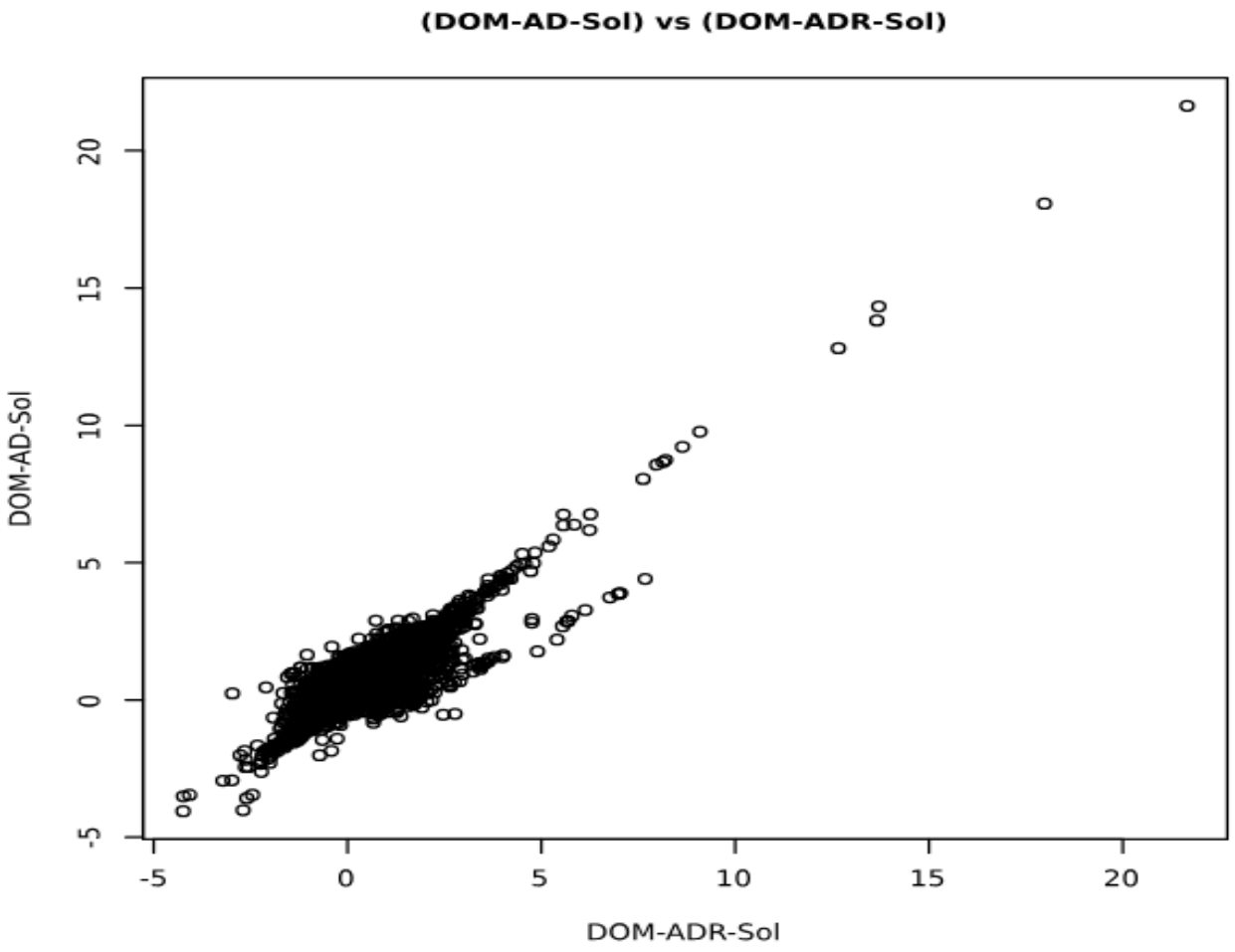
- Does the DOM variance from the ADD-DOM model appear to be split over DOM and ROH components?
- Do DOM and ROH capture the same variation?

Adding ROH had minor effects on:



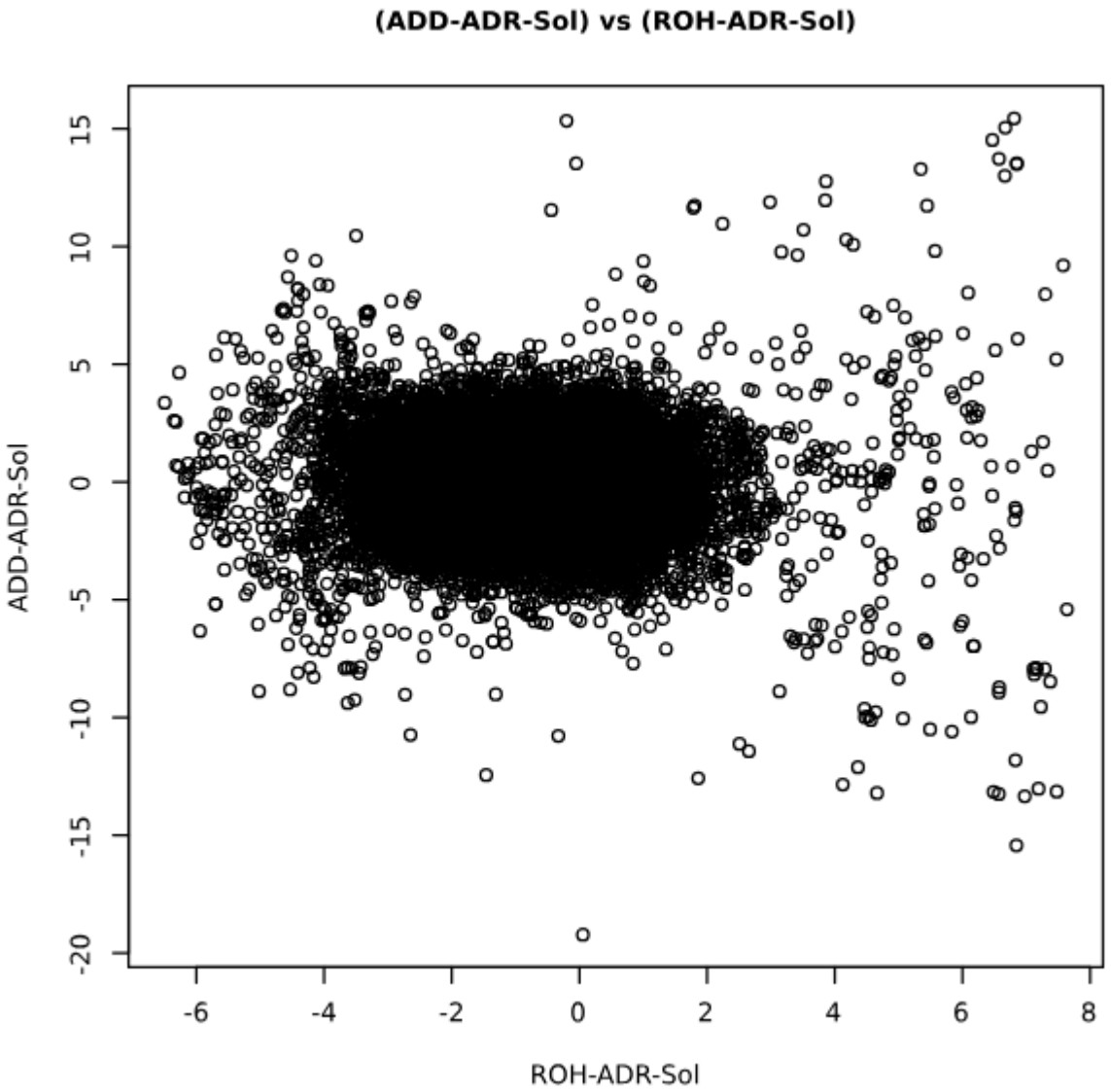
Additive effects: r = 0.994

Adding ROH had moderate effects on:



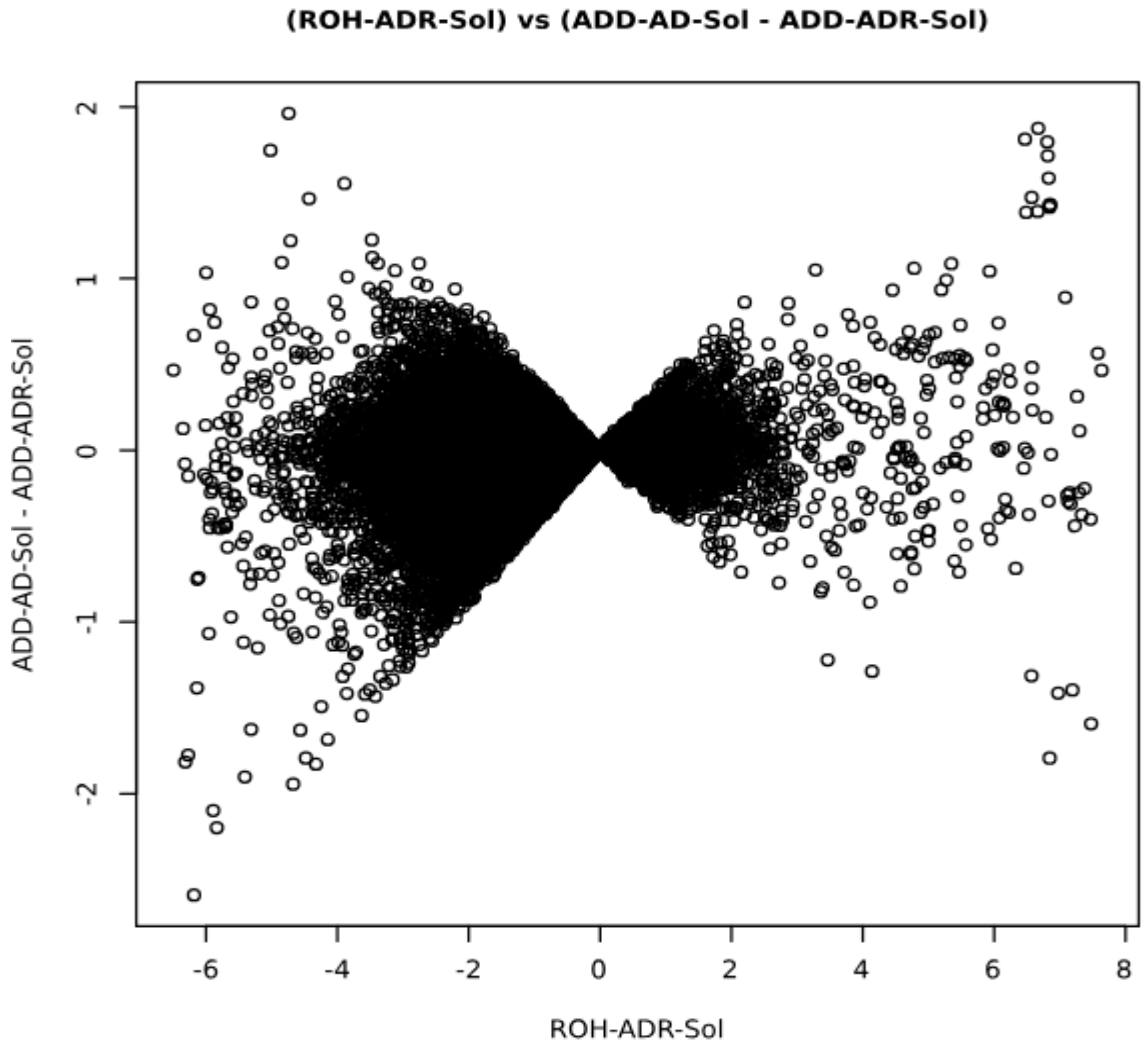
Dominance effects: r = 0.82

ROH effects uncorrelated with

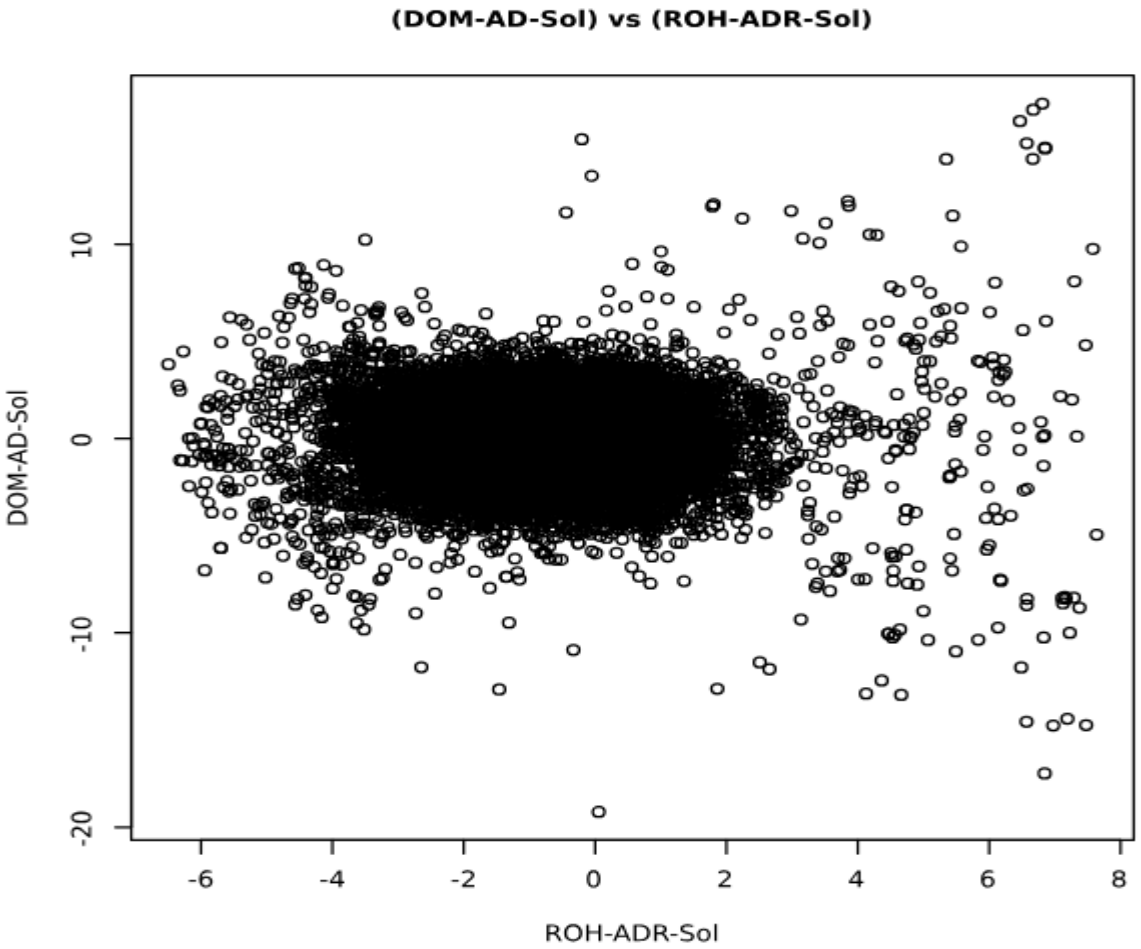


Additive effects with ROH: r = -0.047

Constrained changes in additive effects (Bow tie plot)

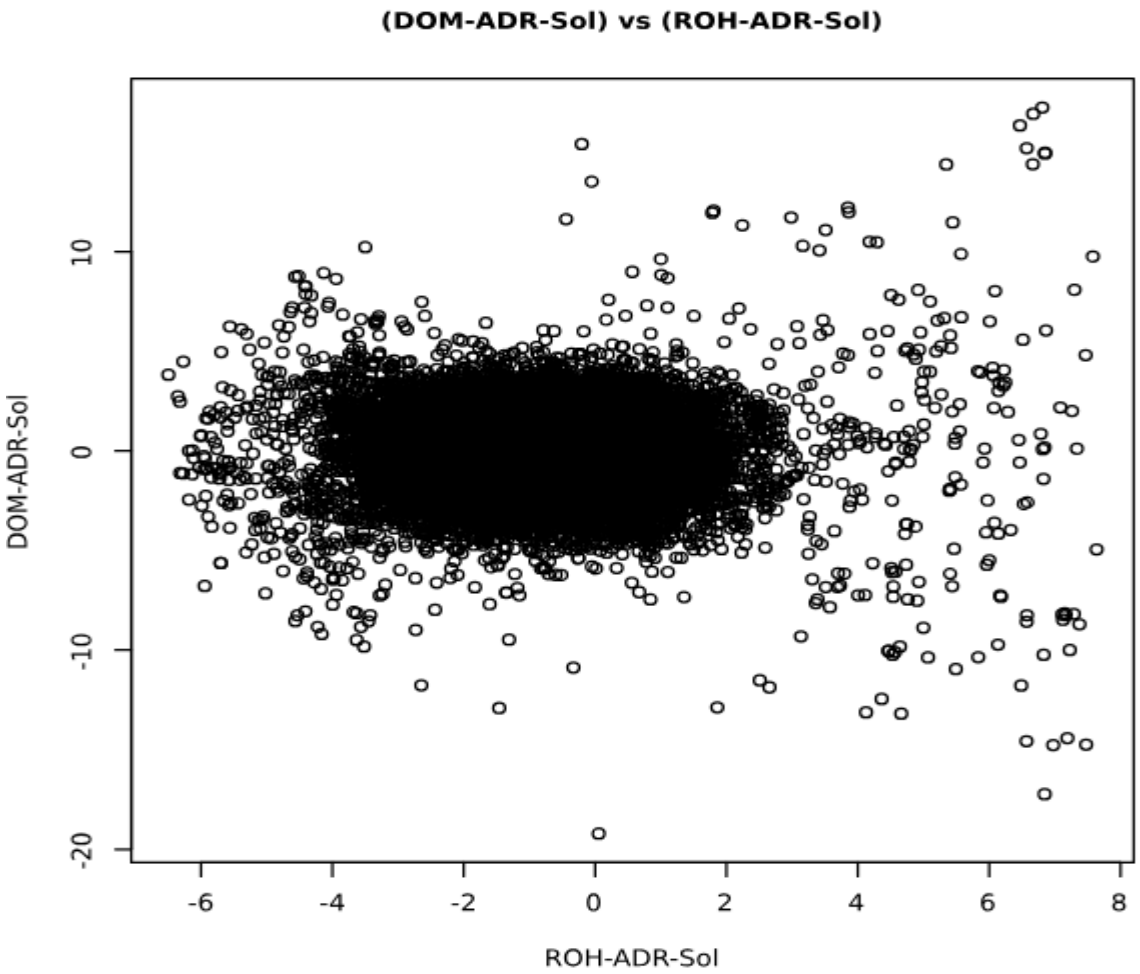


ROH effects uncorrelated with dominance effects



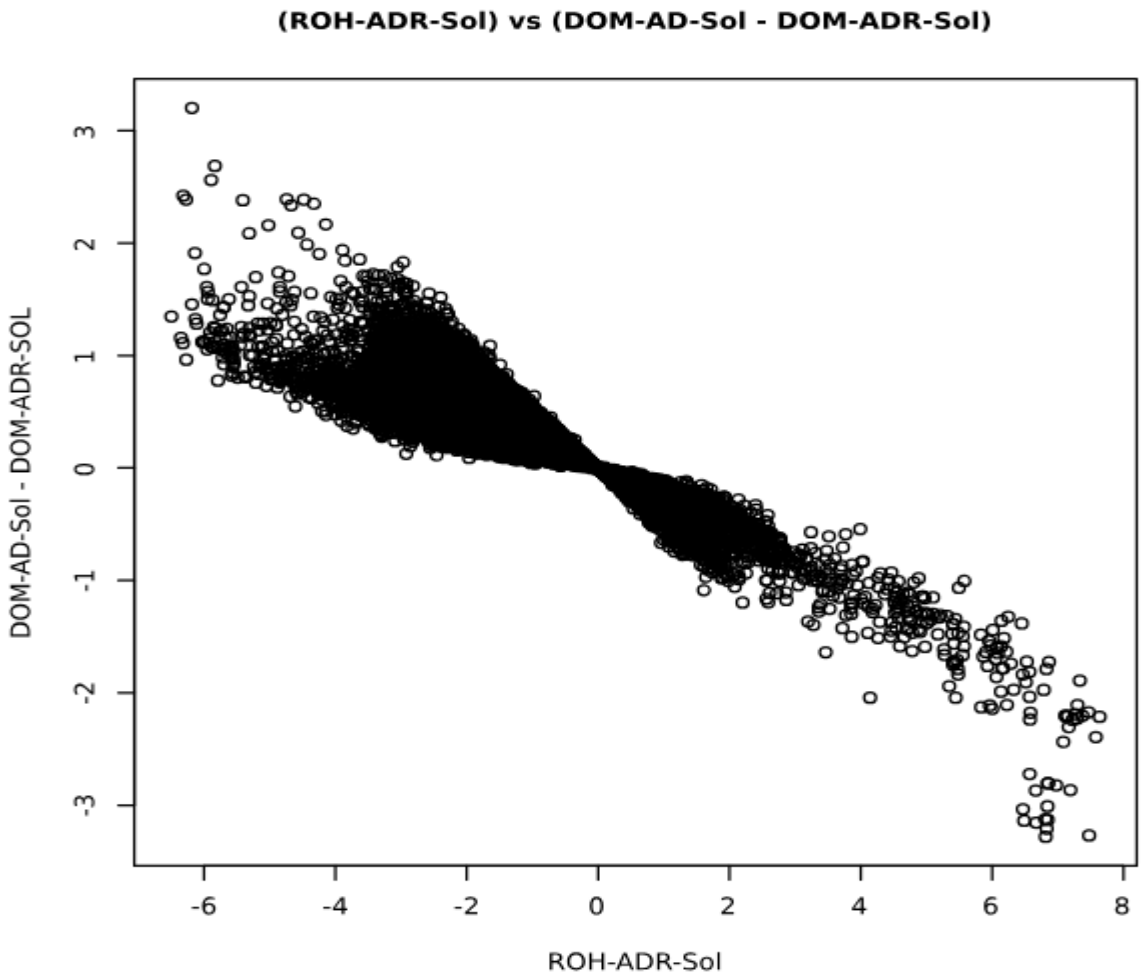
Dominance effects with ROH: r = -0.013

ROH effects moderately correlated with dominance effects



Dominance effects with ROH: r = 0.412

Changes in dominance effects correlated with ROH (Twisted bow tie plot)



Conclusions

- Over a million-cow GWAS identified many additive and dominance effects for production traits.
- Small number of dominance effects were detected for fertility traits with lower statistical significance than the additive effects.
- ROH effects were detected with less statistical significance than the additive and dominance effects, except for somatic cell score.
- We detected novel dominance effects for somatic cell score.
- A substantial correlation of DOM-ADR with ROH, especially when compared to DOM-AD with ROH.
- It seems to be some confounding between dominance effects and ROH, but we had to dig pretty deeply to find evidence of that.

Acknowledgments

- Members of the Council on Dairy Cattle Breeding (CDCB) are acknowledged for providing the data for the GWAS analysis.
- Karin Meyer for making her software, Wombat, available.
- Ignacy Misztal and Daniela Lourenco for making their software, preGSf90, available.
- This project was supported by the USDA-ARS Food Systems Research Center at the University of Vermont.

Thank you for your attention

ROH=0 is likely in chromosome segments having markers with high MAF,
because it will more likely give DOM=1. Low MAF gives more likely DOM=0 & ROH=1.
So, high MAF markers are more likely heterozygous (DOM=1) and ROH=0.
For ROH=1, it is important for many consecutive markers to have DOM=0,
but for ROH=0, it is enough to have some DOM=1.
If many consecutive markers had DOM altering between 0 and 1, then ROH would be 0,
and the confounding would be small. However, if nearby markers are similar
(either DOM=1 or 0), DOM & ROH can become confounded to some degree.

So, DOM=0 means a high probability for ROH=1,
DOM=1 means a high probability for ROH=0. The probability depends on the
conditions that we have for the sliding window (number of markers, length, space between markers etc.).
You can easily make a 2 by 2 table to see the dependency:

	ROH=1	ROH=0
DOM=0	a	b
DOM=1	c	d

The expectation is that the numbers of markers on the diagonal (a & d) are much
higher than off-diagonal (b & c).
We can even compute a Chi-square test from this kind of table.
Computing this to all markers is a lot of computation!