Characterization of Linkage disequilibrium in a Danish, Swedish and Finnish Red Breed Cattle Population

Elisenda Rius-Vilarrasa[1], J. Vilkki[2], I. Strandén[2], N. Schulman[2], T. Iso-Touru[2], B. Guldbrandtsen[3], E. Strandberg[1], M.S. Lund[3] and W.F. Fikse[1]

[1]Swedish University of Agricultural Sciences, Uppsala, Sweden
[2] MTT Agrifood Research Finland, Jokioinen, Finland
[3] Aarhus University, Tjele, Denmark

11th EAAP, 31st August 2011

Stavanger, Norway

Acknowledgments

The Swedish Farmers' Foundation for Agricultural Research (SLF)

Nordic Genetic Evaluation

Svensk Mjölk

Viking Genetics

Swedish University of Agricultural science

Danish Directorate for Food, Fisheries and Agri. Business

Aarhus University, Denmark

MTT Agrifood Research Finland

Background

- Multi-breed genomic predictions in the Nordic Red dairy cattle populations, are influenced by the genetic ties between breeds.

- Linkage disequilibrium and persistency of marker-QTL phase might differ between populations.

- This can compromise the effectiveness of pooling genetic data together in a multi-breed genomic prediction analysis.

- Assessing the differences in the patterns of LD between populations can contribute to:
  - The information on the marker density requirements (de Roos *et al.* 2008)
  - The optimization of the genomic prediction models for multi-breed reference populations .

Aims

Explore linkage disequilibrium (LD) and persistency of phase

Estimate the effective population size

Materials and Methods

- Genotypes
  - Illumina 50K SNP array → 38.647 SNPs after editing (MAF>5%)
  - 6,107 bulls (RDM = 21%, SRB = 32% and FAY = 47%)

- LD measures *(Lewontin, 1964; Hill & Robertson 1968)*

$$r^2 = \frac{D^2}{p_{A1} p_{A2} p_{B1} p_{B2}} \qquad r = \frac{D}{\sqrt{p_{A1} p_{A2} p_{B1} p_{B2}}}$$

  - $D = p_{A1B1} p_{A2B2} - p_{A1B2} p_{A2B1}$ and p_{A1}, p_{A2}, p_{B1} and p_{B2} are the frequencies of alleles *A1, A2, B1* and *B2*, respectively.

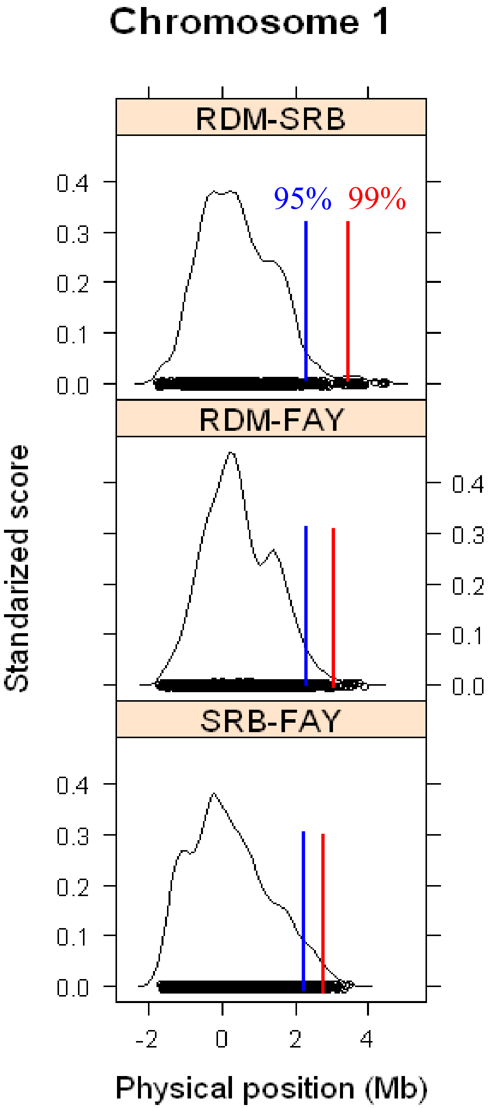
  - varLD program *(Ong and Teo 2010, Gen. Res.)*

VarLD program

- Identify genomic regions with LD difference between populations
    - *Calculate the extent of LD (r^2) between every pair of SNP of pre-defined sliding windows → LD matrix for each window/population.*
    - *varLD scores: comparing the extent of departures between the eigenvalues of the two matrix*
    - *The magnitude of the varLD scores as a measure of dissimilarity between populations*

varLD program

- *Candidate regions displaying significant LD difference between populations are obtained by truncation of the distribution of scores at 95% and 99% percentiles.*



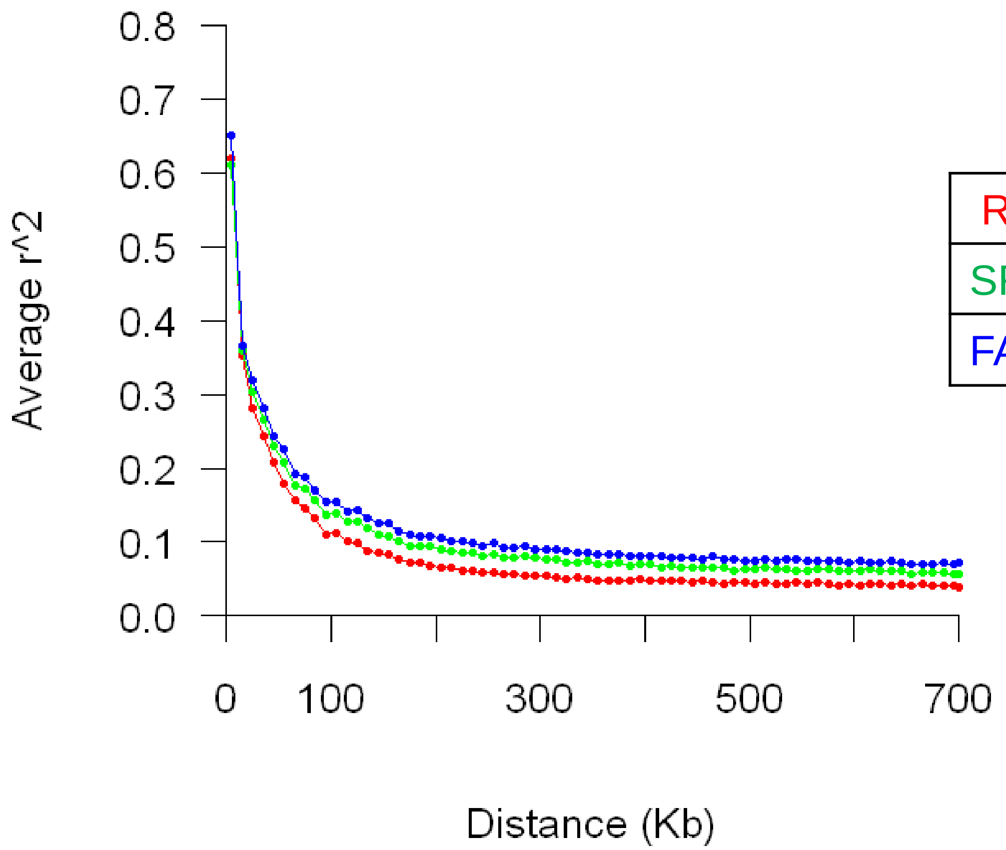
Effective population size: Ne

- Equations *(Hill W. 1975, Hayes et al. 2003; Goddard et al. 2006)*

$$E(r^2) = 1 / (4Ne_t c+2) \qquad Ne_t = (4\bar{c})^{-1}\left[(\bar{r}_c^{\,2})^{-1} - 2\right]$$

  - ***c*** is the marker distance, assuming that 1Mb ≈ 1cM
  - $\boldsymbol{r^2}$ is the average LD at different marker distances
  - $\boldsymbol{Ne_t}$ is the effective population size, *T* generations ago, where *T* = 1/(2***c***)

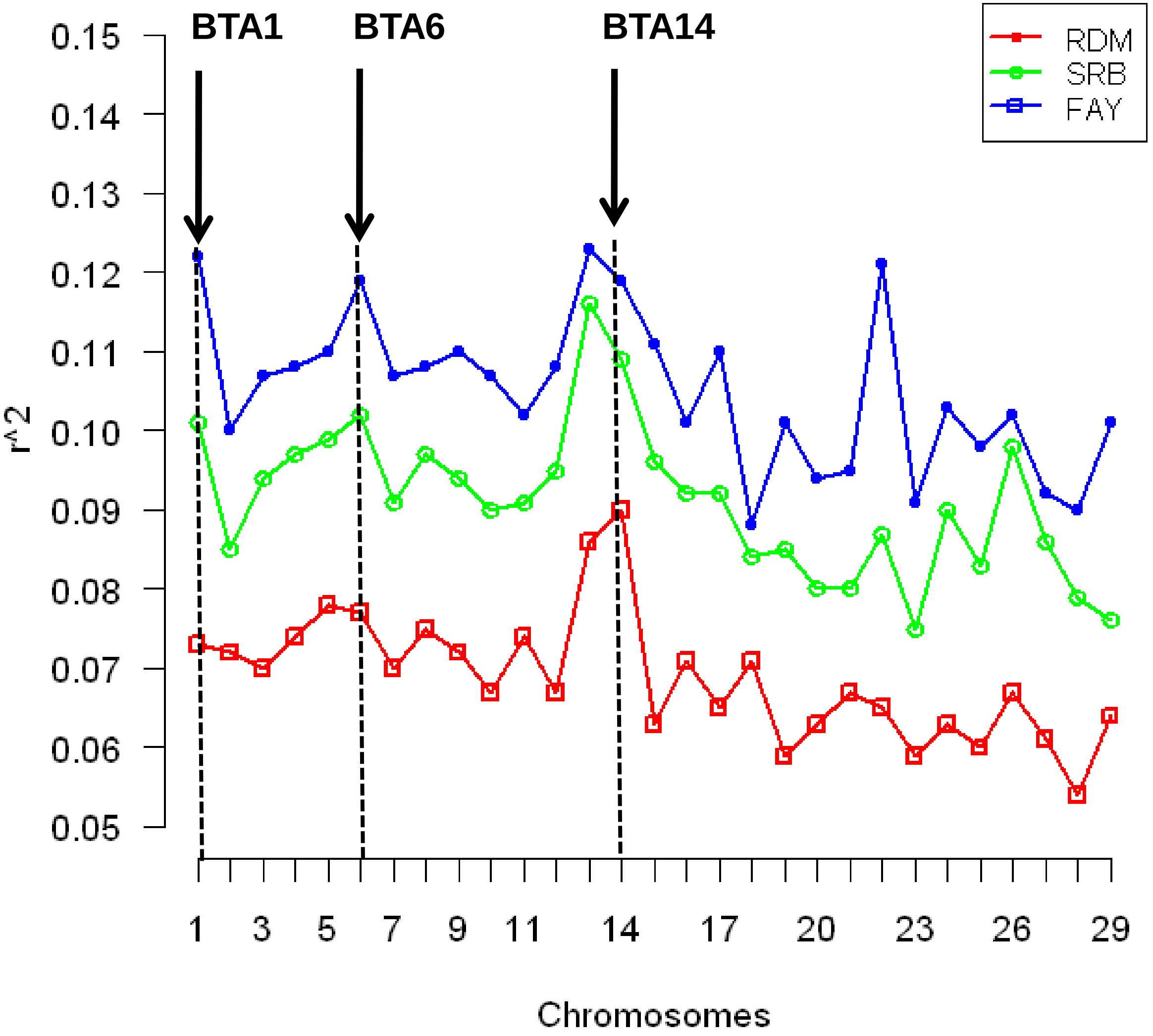
Results$_{(1)}$ - Average r^2 across marker distance



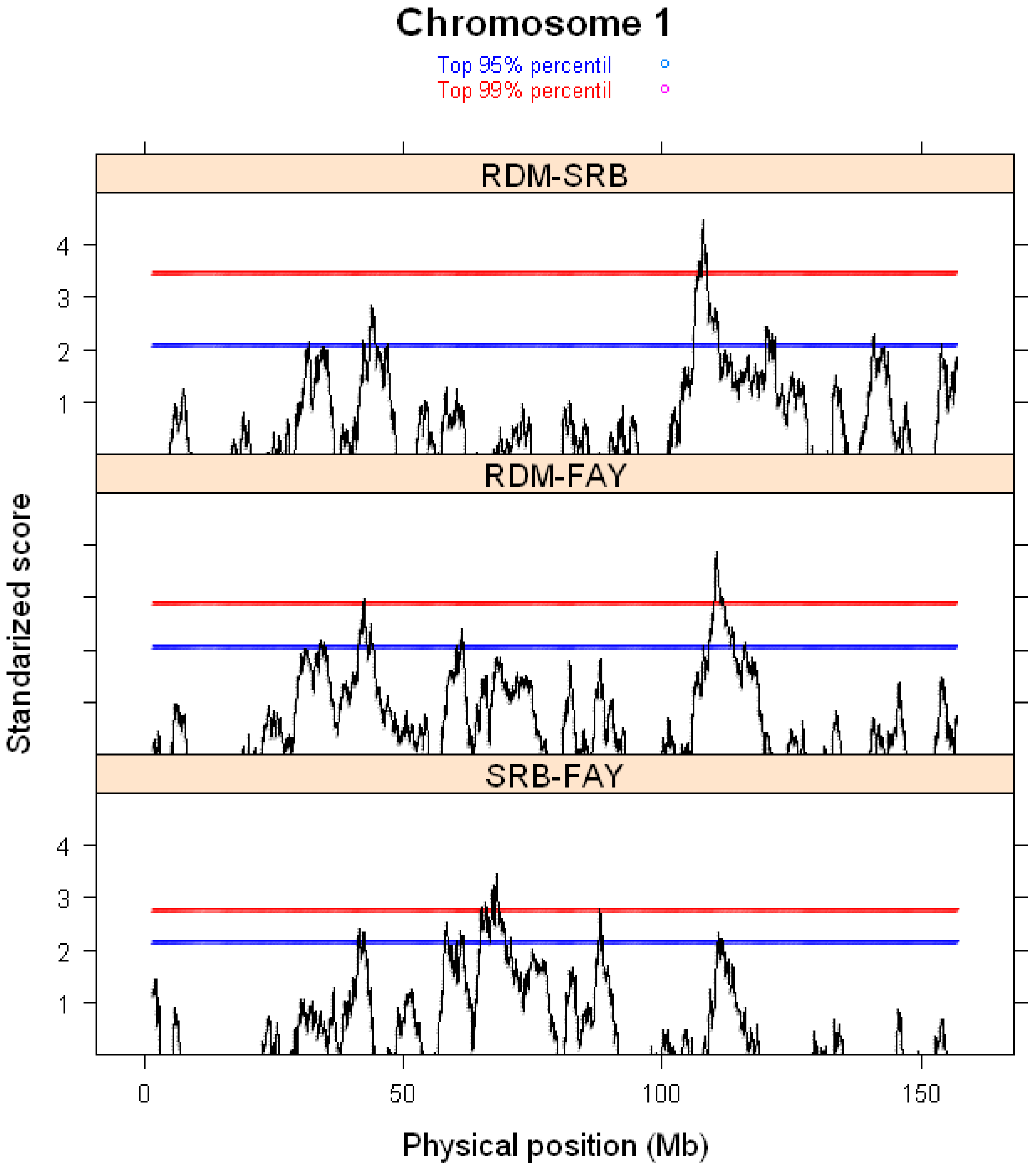
	Average LD (r^2)	
	>0 – 700 Kb	**>0 – 100 Kb**
RDM	0.07	0.22
SRB	0.091	0.24
FAY	0.11	0.25

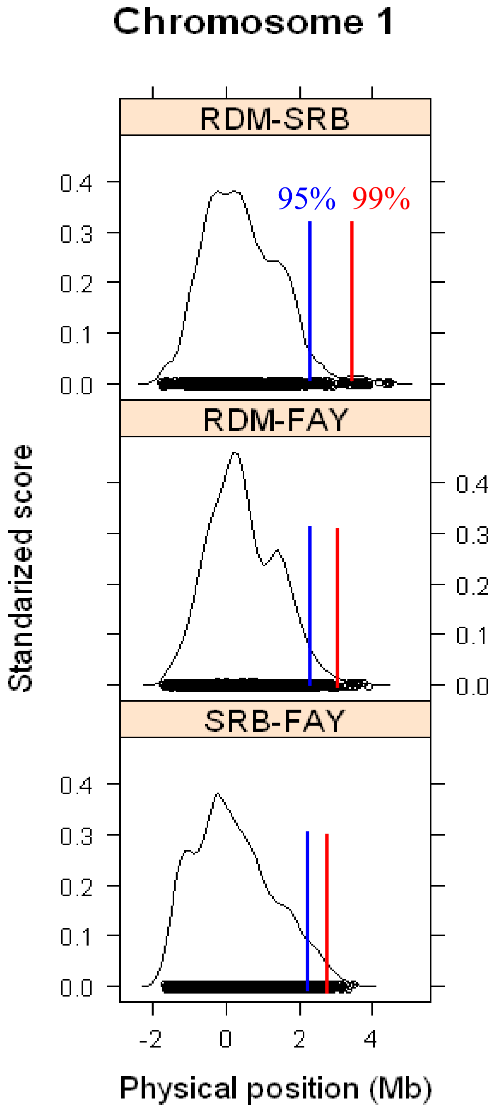
Useful LD of 0.20 → 54,000 SNPs (SRB/FAY) and 75,000 SNPs (RDM) to achieve accuracies from genomic prediction models of ≥ .85 *(Meuwissen et al. 2001)*

Results$_{(2)}$ – Average r^2 per chromosome

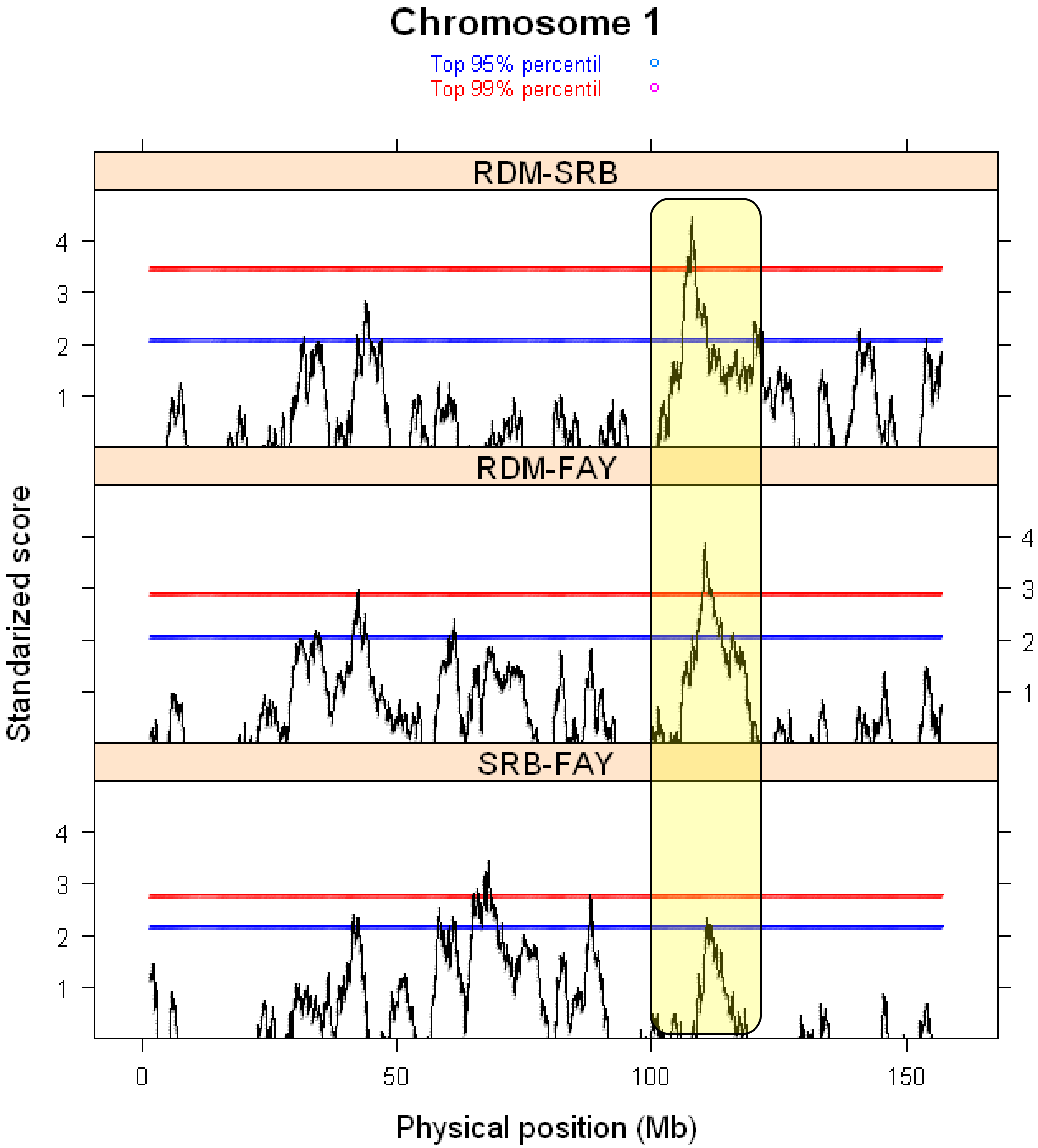


Results$_{(3)}$ – Patterns of LD between populations





Results$_{(3)}$ – Patterns of LD between populations

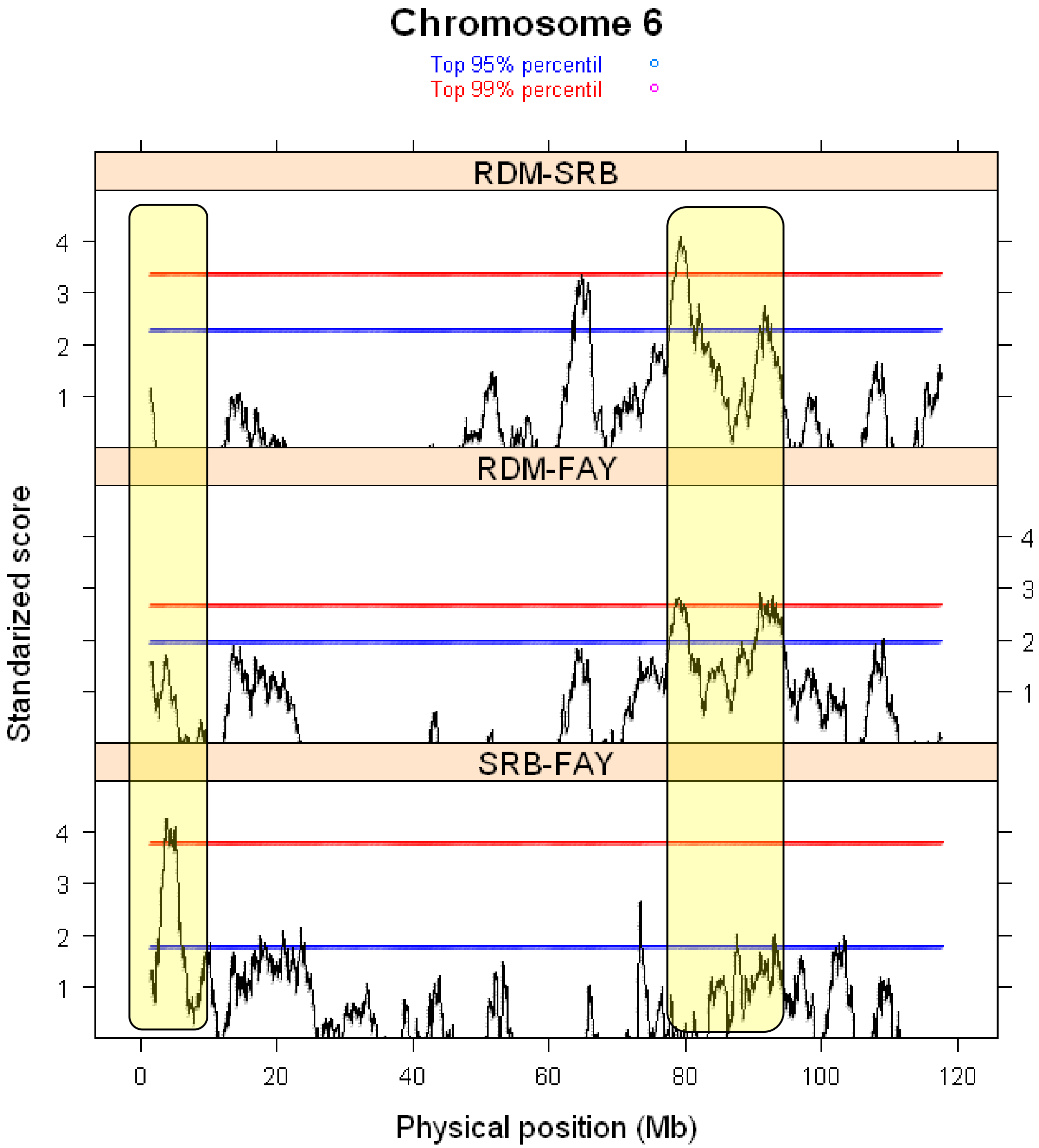


QTL

Milk yield (FAY)

(Viitala S.M et al. 2003, J Dairy Sci.)

Results$_{(4)}$ – Patterns of LD between populations

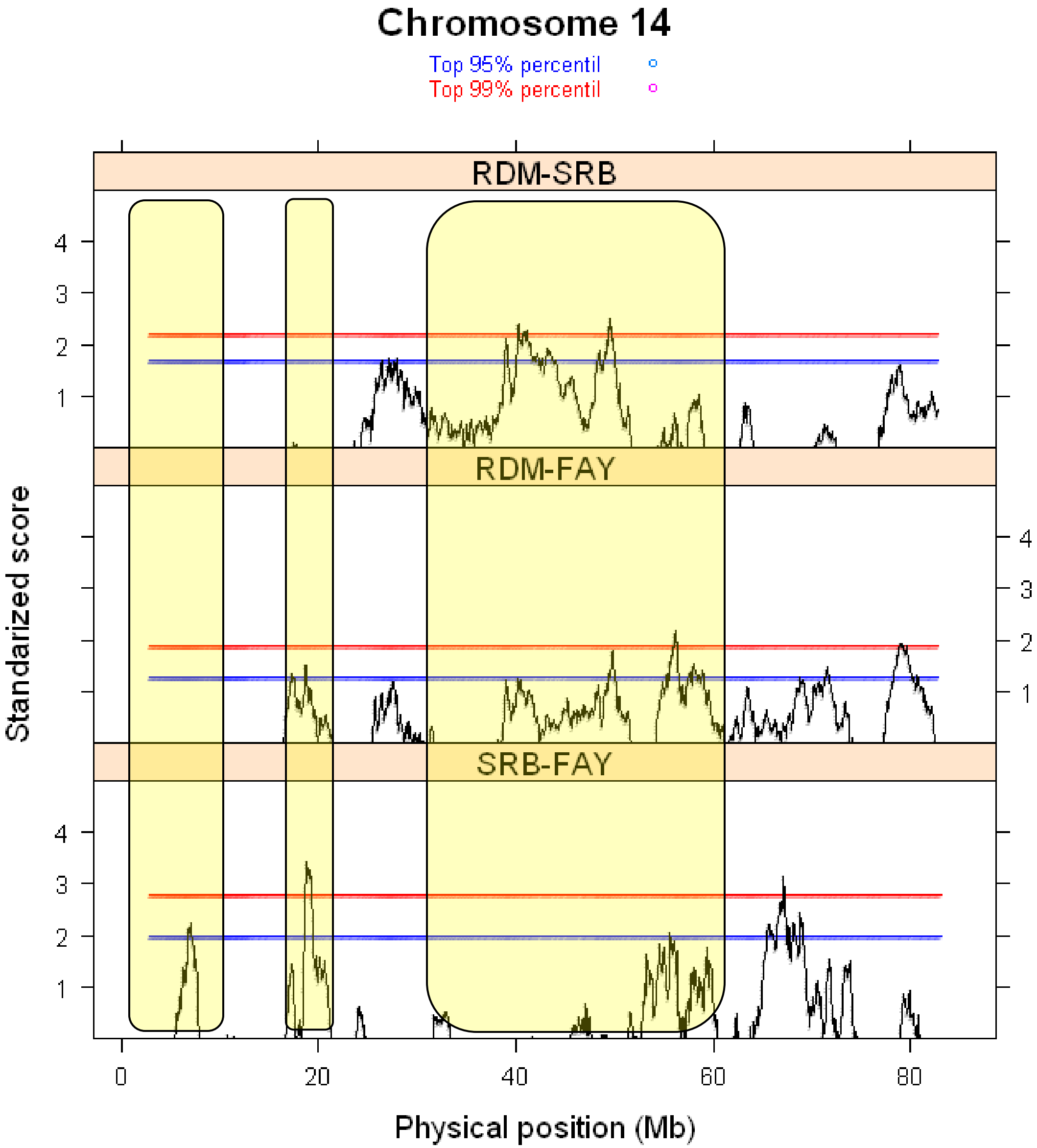


QTLs:

Somatic cell count (NRF) *(Klungland H. et al. 2001 Mamm. Genome)*

Milk protein % (FAY) *(Velmala R.J et al. 1999, Animal Genetics)*

Results$_{(5)}$ – Patterns of LD between populations



QTLs:

Milk fat yield (FAY)

(Viitala S.M et al. 2003, J. Dairy Sci.)

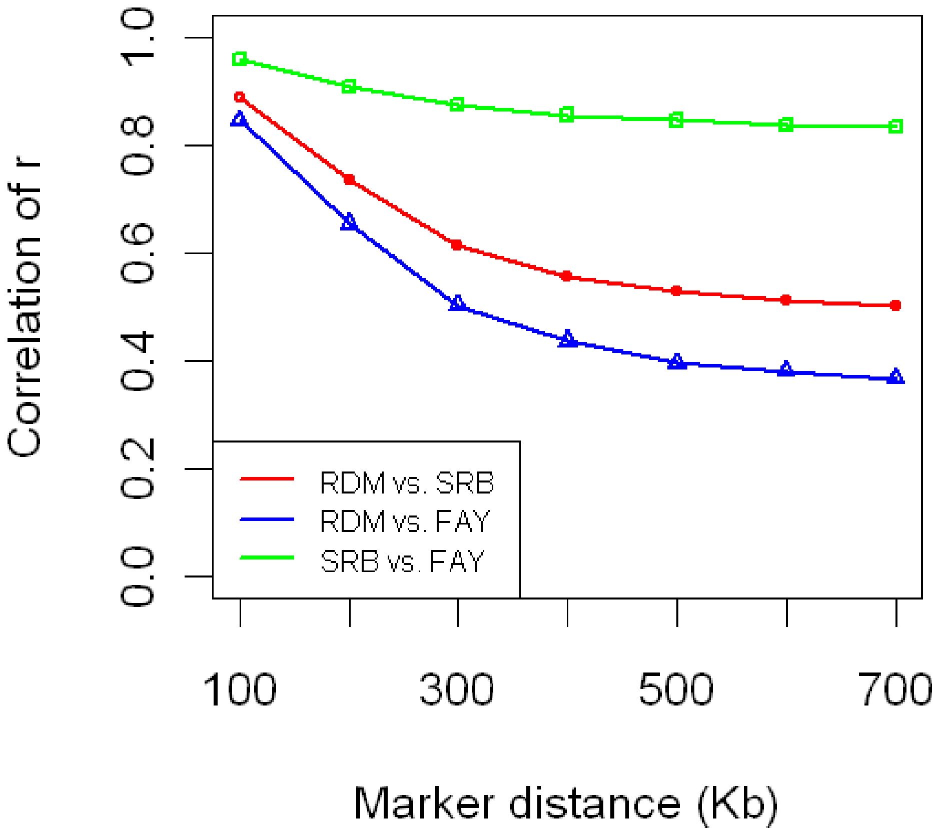
Non-return rate (FAY)

(Shulman N. et al 2008, GSE)

Somatic cell score (RDM, SRB and FAY)

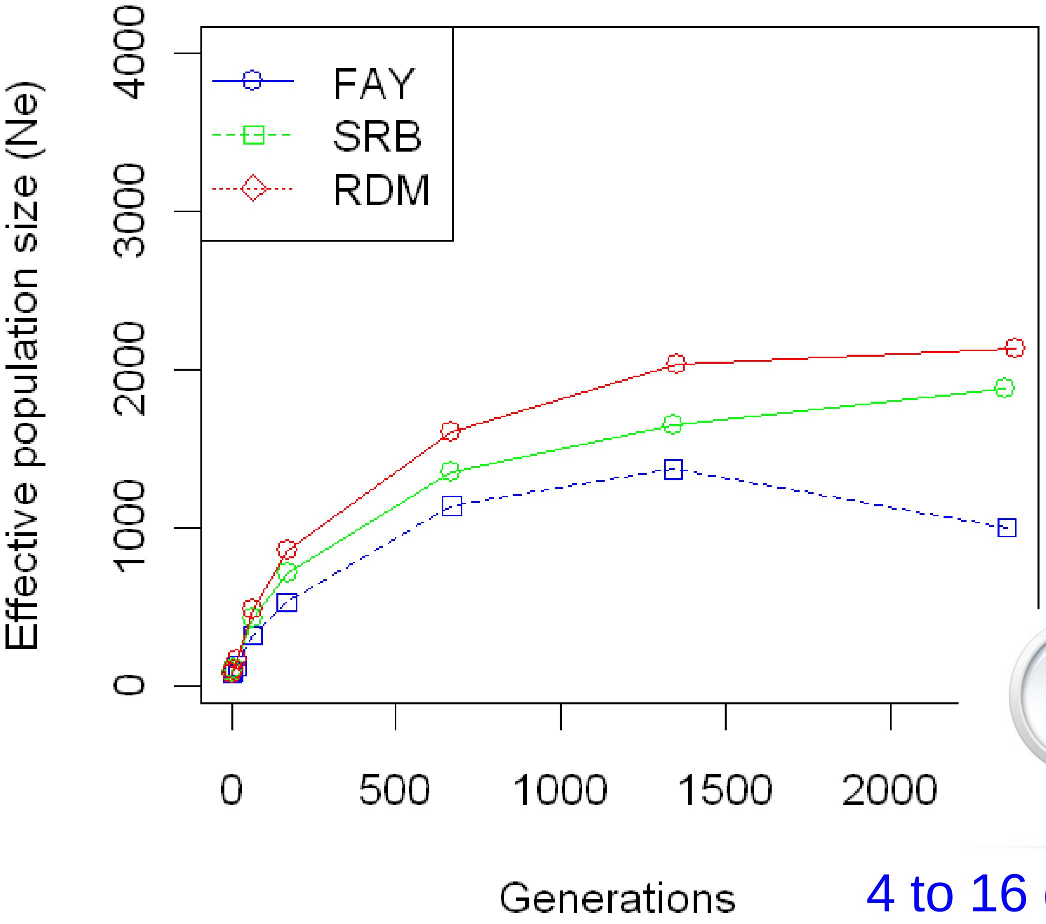
(Lund M. S et al. 2007, J. Dairy Sci.)

Results$_{(6)}$ – Persistency of marker phase



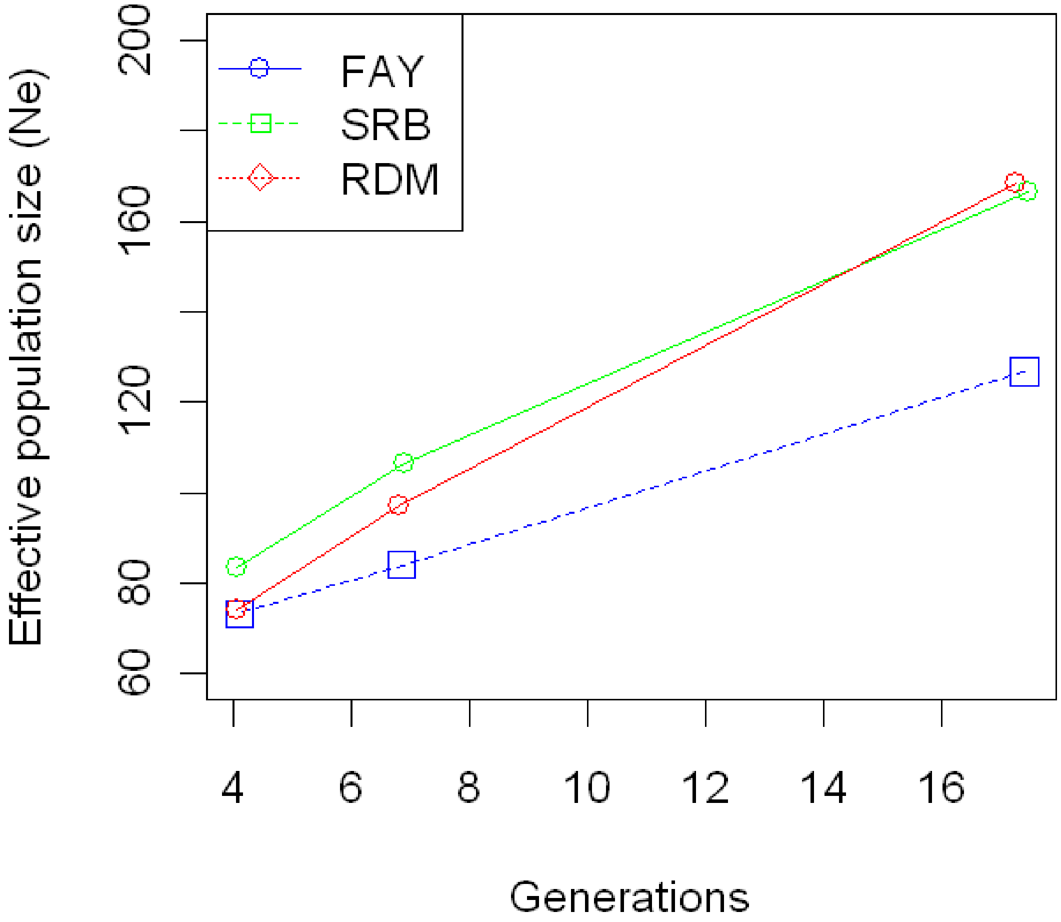
	RDM-SRB	RDM-FAY	SRB-FAY
P. Phase (0-100kb)	29.1% (16.5%)	33.2% (19.0%)	15.6% (9.1%)
Corr. of r (0-100kb)	0.68 (0.89)	0.58 (0.85)	0.89 (0.96)

Results$_{(7)}$ – Estimates of Ne



4 to 16 generations

Results$_{(7)}$ – Estimates of Ne of recent generations



Holstein < 50

Summary

- Average LD was lowest for RDM (0.07) and highest for FAY (0.11)
  - North American Holstein $r^2 > 0.20$, (Bohmanova J. et al. 2010)
- More markers to increase accuracies from genomic prediction models.
- Regional LD differences within reported QTL positions.
- Quantify regional LD variation to help:
  - Imputation strategies.
- The loss of persistency was highest between RDM-FAY and lowest between SRB-FAY
- Ne up to 4 generations ago was 84 for SRB, 74 for RDM and 73 for FAY.

Thank you