

# The estimation of linkage disequilibrium decay in Croatian Holstein cattle: potential for genomic selection

Marija Špehar<sup>1</sup>, Zdenko Ivkić<sup>1</sup>, Drago Solić<sup>1</sup>, Ante Kasap<sup>2</sup>

<sup>1</sup>Croatian Agency for Agriculture and Food, Svetošimunska 25, 10000 Zagreb, Croatia

<sup>2</sup>University of Zagreb Faculty of Agriculture, Svetošimunska 25, 10000 Zagreb, Croatia

# Linkage disequilibrium (LD)

- Non-random association of any two loci within the population
- Degree of connectedness between two loci
- Usual indicators of LD –  $r^2$  and  $D'$
- $r^2$  interpreted as a squared correlation (mutual relationship between alleles in two loci)
  - range between 0 and 1
  - $r^2 = 1$  implies the markers provide exactly the same information

# Applications of LD

- Evolutionary biology – reflections on past events and insight into evolutionary history
- Genetic diversity – computation of effective population size
- Genome wide association studies (GWAS) – search for causal genes
- Genomic selection (GS) – determine how many SNPs we need for a proper analysis

# Objective

- To estimate **the genome-wide and chromosome level** of LD in Croatian Holstein cattle



- In order to examine **the potential of GWAS and GS**

# Material

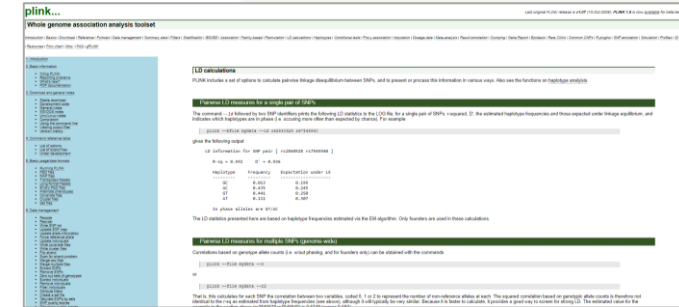
- 101 Holstein females
  - from 32 herds
  - progenies of 45 sires
- Illumina BovineSNP50K BeadChip – 52,445 SNPs
- Quality control
  - autosomal SNPs only
  - MAF < 5%
  - >10% missing genotypes
  - deviation from Hardy-Weinberg equilibrium ( $P < 10^{-6}$ )
  - **40,763 SNPs**



# Method

Calculation of “raw LD” using the  $r^2$  statistics

- **PLINK**



Post hoc statistical analysis of genome wide LD

- **Dplyr and ggplot packages**
- bins spanning 100 Kbp ([0,100].....(900-1000])
- bins spanning 20 Kbp ([0,20].....(980-1000])



# Results



# Genome wide LD

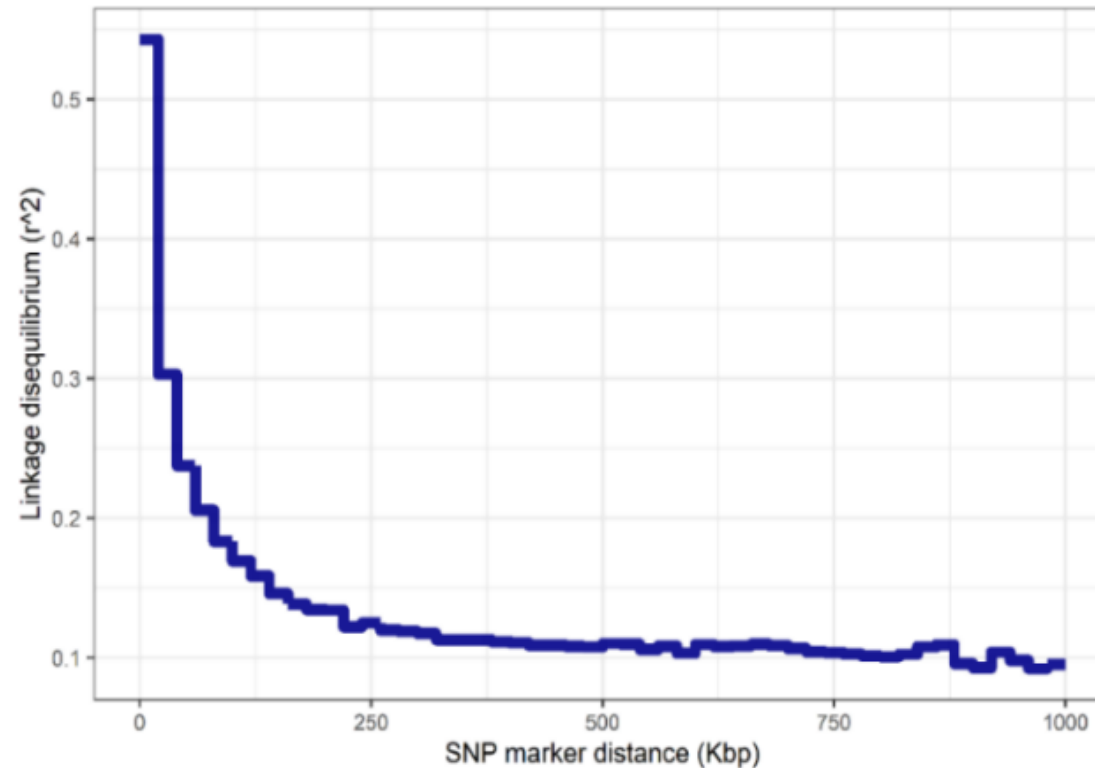
- Overall average genome-wide LD - 0.16

Distance (Kbp)	N	avg ( $r^2$ )	sd ( $r^2$ )	Distance (Kbp)	N	avg ( $r^2$ )	sd ( $r^2$ )
[0,100]	77153	<b>0.30</b>	0.32	(500,600]	25619	0.11	0.14
(100,200]	65967	0.15	0.19	(600,700]	15121	0.11	0.14
(200,300]	62968	0.12	0.16	(700,800]	8818	0.10	0.13
(300,400]	55611	0.11	0.14	(800,900]	5229	0.10	0.13
(400,500]	40370	0.11	0.14	(900,1000]	2952	0.09	0.13



# Genome-wide LD

- Overall average genome-wide LD - 0.16
- Inter-marker distance of 100 Kbp



# Chromosome LD

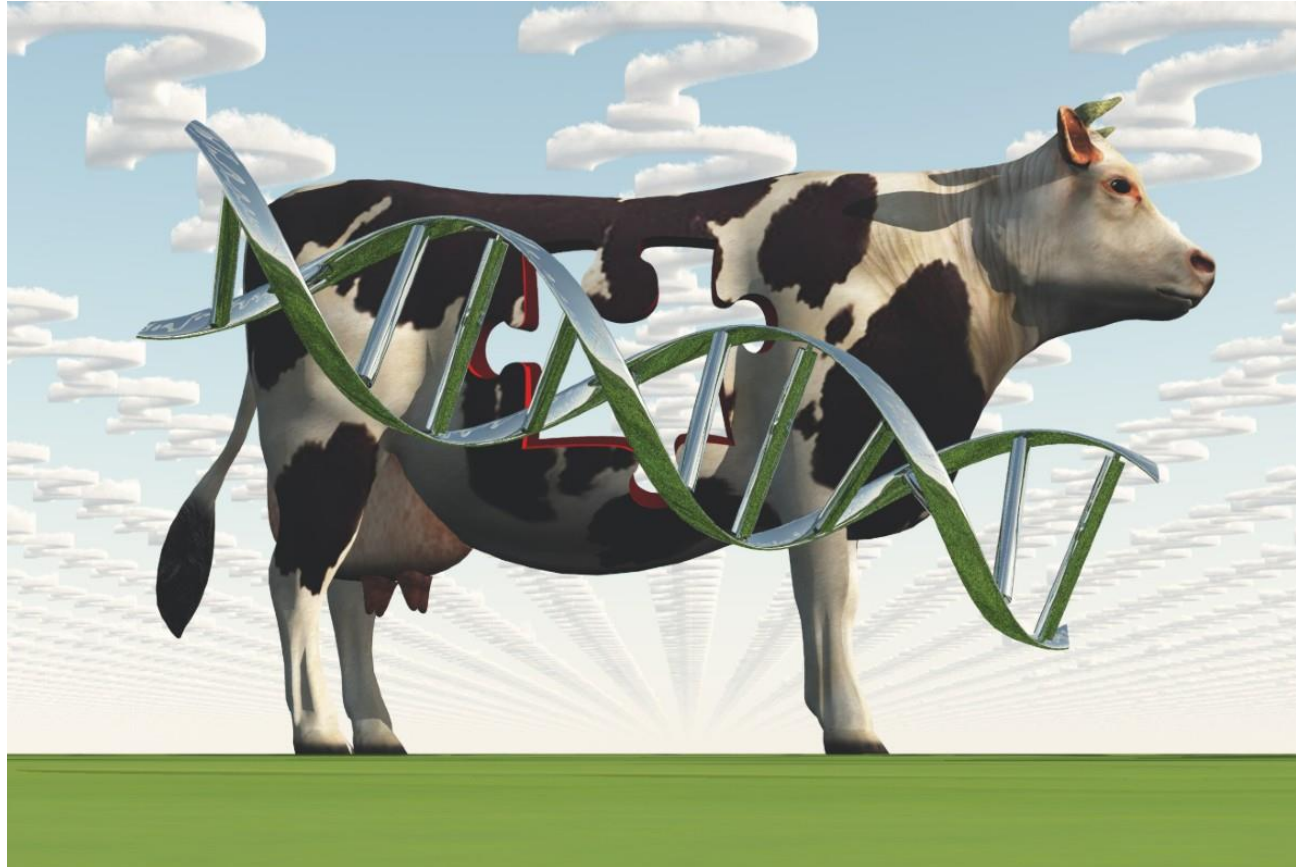
- Inter-marker distance of 100 Kbp on 29 autosomes (BTA)

BTA	avg ( $r^2$ )	sd ( $r^2$ )	BTA	avg ( $r^2$ )	sd ( $r^2$ )	BTA	avg ( $r^2$ )	sd ( $r^2$ )
1	0.16	0.22	11	0.14	0.19	21	0.18	0.26
2	0.17	0.22	12	0.13	0.18	22	0.15	0.21
3	0.15	0.20	13	0.20	0.26	23	0.15	0.21
4	0.14	0.20	14	0.18	0.22	24	0.14	0.18
5	0.15	0.20	15	0.13	0.17	25	0.13	0.17
6	0.17	0.23	16	0.19	0.24	26	0.15	0.20
7	0.16	0.21	17	0.15	0.20	27	0.14	0.22
8	0.16	0.22	18	0.13	0.18	28	0.12	0.17
9	0.15	0.21	19	0.21	0.28	29	0.15	0.21
10	0.17	0.21	20	0.23	0.28			

# Conclusions

- Moderate extend of LD level
  - implies that the **50K SNP chip should be sufficient** for the **GS** and **GWAS** in Croatian Holstein cattle
- More samples should be used to get more information on this issue
- Other breeds for comparison
  - Simmental

# Thank You For Listening



<https://www.icbf.com/?p=5831>