

## Linkage Disequilibrium and Persistence of Phase in Five Spanish Local Beef Cattle Breeds

J.J. Cañas-Álvarez<sup>1</sup>, E.F. Mouresan<sup>2</sup>, L. Varona<sup>2</sup>, C. Díaz<sup>3</sup>, C. Avilés<sup>4</sup>,  
J.A. Baró<sup>5</sup>, J. Altarriba<sup>2</sup>, J. Casellas<sup>1</sup> and J. Piedrafitá<sup>1</sup>

<sup>1</sup>Universitat Autònoma de Barcelona, Spain, <sup>2</sup>Universidad de Zaragoza, Spain,  
<sup>3</sup>INIA, Spain, <sup>4</sup>Universidad de Córdoba, Spain, <sup>5</sup>Universidad de Valladolid, Spain

**ABSTRACT:** We have used the BovineHD Genotyping BeadChip to obtain high density genotypes (>700,000 SNP after quality control) from 116 trios in five Spanish local beef cattle breeds. Linkage disequilibrium (LD) was measured through the  $r^2$  statistic. Average  $r^2$  for adjacent markers in the five breeds were very close, around 0.52, and decreased with increasing distance between markers, although in long distances some LD remained (0.07 and 0.05 for markers 200 kb and 1000 kb apart, respectively). At all distances the standard deviations were large and the shape of the distribution varied depending upon the marker distance. Average  $r^2$  varied also between chromosomes. Pairwise correlations between the  $r$ 's estimated in two breeds at short distances (5 kb) was in the rank of 0.6 – 0.7. Similarly to  $r^2$  estimates, this correlation decreased with increasing marker distance.

**Keywords:** beef cattle; linkage disequilibrium; persistence of phase

### Introduction

Genomic selection was proposed by Mewissen et al. (2001) as a strategy to predict breeding values from a massive set of SNP markers across the genome. SNP chips have been developed in all domestic species, particularly in cattle (Matukumalli et al. (2009)), and recently a more than ten-fold dense chip has been developed by Illumina. The outcome of genomic selection relies on the existence of enough LD between SNP markers and QTLs and QTL such that the marker allele–QTL phase persists across generations. LD markers are always discovered in some reference population in which the initial experiment was conducted. The value of the markers in populations other than the reference population will depend on the persistence of LD phase between the reference population and the second population (Dekkers and Hospital (2002)). The extent of LD has been assessed in several dairy and beef cattle, zebu and African cattle and also in composite breeds (Gautier et al. (2007); De Roos et al. (2008); Villa-Angulo et al. (2009)). Furthermore, persistence of LD phase has been assessed in dairy and beef cattle including several subpopulations of Holstein, Angus and Jersey breeds. Little is known, however, for other local beef breeds, and thus the aim of our research is to assess the magnitude of LD and the persistence of phase, measured from data of a High-density (800 K) chip in the main Spanish beef breeds. This study will provide some insight on the feasibility of applying the genomic selection using data from the individual

breeds or from a metapopulation including all or some of them.

### Materials and Methods

**Animals and sample size.** A total of 116 sire/dam/offspring trios were collected from five Spanish beef cattle populations, including Asturiana de los Valles (AV, n = 25), Avileña – Negra Ibérica (ANI, n = 24), Bruna dels Pirineus (BP, n = 25), Pirenaica (Pi, n = 24) and Retinta (Re, n = 18) breeds. Selected parents were chosen as unrelated as possible.

**SNP genotyping and phasing.** Genomic DNA was extracted by standard protocols. High density SNP genotyping was performed by using the BovineHD Genotyping BeadChip (Illumina Inc, USA) designed to genotype 777,962 SNPs, according to the protocol of the manufacturer at a commercial laboratory (Xenética Fontao, Lugo, Spain). SNPs kept for the study belonged to autosomal chromosomes and were not in repeated positions. Additional requirements were Mendel error rate < 0.05, individual call rate  $\geq 0.95$ , SNP call rate  $\geq 0.95$ , and MAF > 0.01. The quality control was made using PLINK software (Purcell et al. (2007)) and retained 706,704 SNPs, covering 2,510,606 kb, with one marker each 3.553 kb on average. The phases of the parental chromosomes were established by means of Beagle software (Browning and Browning (2009)).

**Linkage disequilibrium decay.** Using PLINK, LD was computed from the founder animals as the  $r^2$  statistic (Hill and Robertson (1968)). Marker pairs were grouped by their pairwise physical distance into bins of 5 kb, starting from 0 to 2000 kb. Average  $r^2$  for SNP pairs in each bin was estimated as the arithmetic mean of all  $r^2$ . The distribution of  $r^2$  estimates and the averages were plotted using an R environment (<http://www.r-project.org>).

**Persistence of LD phase.** Pairwise Pearson correlation coefficients of intermarker  $r_{ij}$  between two populations, k and k', were used to estimate the persistence of phase over bins of 10 kb from 0 to 100 kb, and of 100 kb from 100 to 1000 kb of marker distances. The calculus and figures were developed in an R environment.

### Results and Discussion

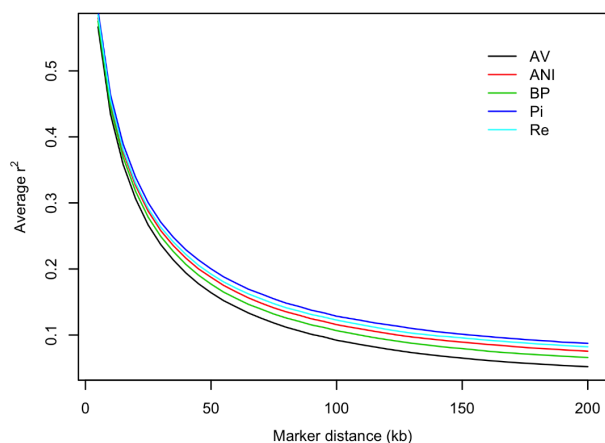
**Linkage disequilibrium.** The number of marker pairs considered to estimate average  $r^2$  ranged from

~ 600,000 for adjacent markers to ~ 900,000 for markers 20 kb apart (not shown in tables). This last figure declined uniformly to ~ 750,000 for markers 1000 kb away. Average  $r^2$  tended to decrease with increasing genomic distance in all studied populations (Table 1, Figure 1). For adjacent markers, average  $r^2$  was 0.531 (Pi), 0.524 (Re), 0.519 (ANI), 0.514 (BP) and 0.504 (AV), in decreasing order. On average, LD decreased up to 0.32, 0.18, 0.12, 0.07, and 0.045 for 20 kb, 50 kb, 100 kb, 200 kb and 1000 kb of marker distance, respectively. At all distances, there were differences among breeds, but the rank of the average  $r^2$  was kept the same that for adjacent markers. As expected from the Sved (1971) formula, the rank of the LD estimates correlated negatively with population effective sizes (Cañas-Álvarez et al. (2014a)) but for Pi. Our LD estimates at various distances were of the magnitude of those reported by de Roos et al. (2008) and Villa-Angulo et al. (2009) in several dairy and beef breeds using less dense SNP panels.

**Table 1. Average  $r^2$  values and (standard deviations) at various distances in five beef breeds.**

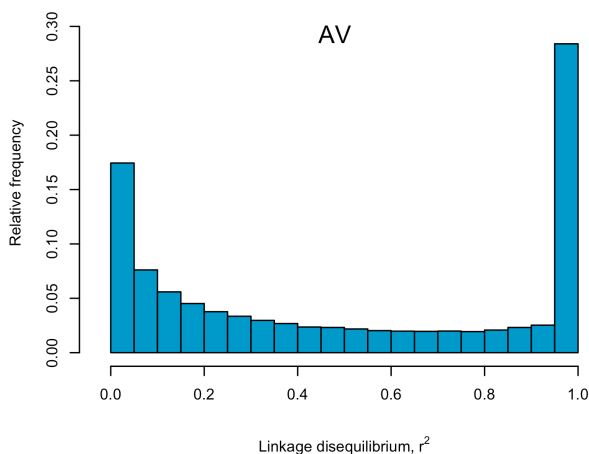
Breed	Adj. (sd)	20 kb (sd)	50 kb (sd)	100 kb (sd)	200 kb (sd)	1000 kb (sd)
AV	0.504 (0.396)	0.306 (0.328)	0.164 (0.231)	0.092 (0.152)	0.052 (0.090)	0.031 (0.045)
ANI	0.519 (0.402)	0.326 (0.341)	0.187 (0.249)	0.116 (0.174)	0.075 (0.115)	0.047 (0.067)
BP	0.514 (0.400)	0.317 (0.336)	0.177 (0.242)	0.106 (0.166)	0.066 (0.107)	0.038 (0.054)
Pi	0.531 (0.404)	0.339 (0.348)	0.200 (0.260)	0.128 (0.187)	0.087 (0.129)	0.058 (0.080)
Re	0.524 (0.403)	0.330 (0.343)	0.192 (0.252)	0.122 (0.179)	0.082 (0.123)	0.056 (0.079)

**Figure 1: Evolution of average  $r^2$  from 0 to 200 kb for all autosomes in five beef breeds**



The standard deviations of  $r^2$  estimates increased with distance (Table 1). This can be related to the distribution of  $r^2$  values as a function of distance (Figure 2). For adjacent markers, a high proportion of complete LD ( $r^2 = 1$ ) estimates is found, followed by markers in complete equilibrium ( $r^2 = 0$ ), the other estimates being almost evenly distributed between these extreme values. As the distance

between markers increases, the proportion of pairwise markers in complete LD decreases and beyond 100 kb no markers in complete LD are found (not shown in figures).



**Figure 2: Distribution of  $r^2$  for adjacent markers**

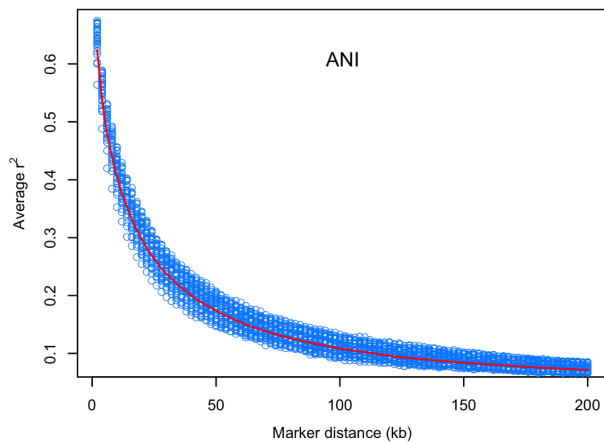
Linkage disequilibrium between adjacent markers declined with the proportion of the SNPs retained for the analysis (Table 2). The average  $r^2$  of 0.52 for adjacent markers in the complete dataset, with an intermarker distance of ~ 4 kb, decreased to 0.30, 0.22 and 0.10 when 10%, 5% and 1% of the markers were included in the analysis. An accuracy of predicted breeding values from dense markers up to 85% was obtained with an average  $r^2$  between adjacent markers of 0.2 (Meuwissen et al. (2001)). In our breeds, a similar average  $r^2$  between adjacent markers would be obtained by using only 5% of the markers of the HD panel (~ 38,000 markers), which corresponds to an average genomic distance of ~ 80 kb.

**Table 2. Average  $r^2$  (average distance in kb) between adjacent markers depending upon the proportion of SNPs retained from the HD Beadchip panel.**

Breed	100%	10%	5%	1%
AV	0.504 (3.97)	0.281 (39.64)	0.206 (78.62)	0.084 (390.56)
ANI	0.519 (4.06)	0.299 (40.41)	0.226 (80.31)	0.106 (399.57)
BP	0.514 (4.02)	0.293 (40.11)	0.217 (79.49)	0.094 (383.50)
Pi	0.531 (4.09)	0.311 (40.92)	0.238 (80.98)	0.113 (401.92)
Re	0.524 (4.08)	0.304 (40.77)	0.232 (80.89)	0.112 (400.20)

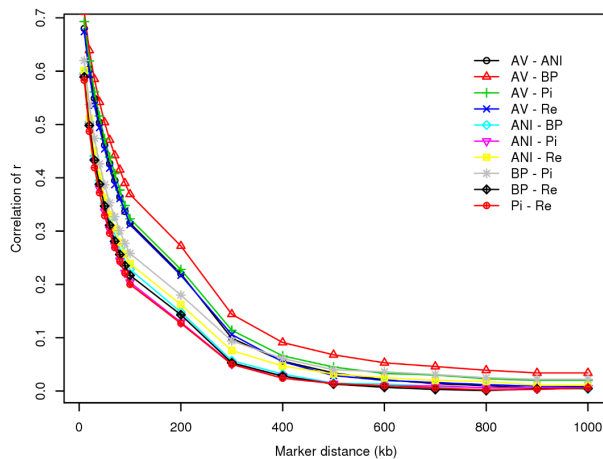
There were differences also in the average  $r^2$  at a particular marker distance depending upon chromosomes. This is shown in Figure 3 where average  $r^2$ 's in the ANI breed for each chromosome at different marker distances are represented as circles. Similar figures were found for the other breeds. This variability among chromosomes leads

to some changes in the ranking of LD among breeds when only the markers at a particular chromosome are considered.



**Figure 3: Average  $r^2$  for each of the 29 autosomes at different distances**

**Persistence of LD phase.** The pairwise correlation of  $r$ 's between AV and the other four breeds at  $< 5$  kb apart was  $\sim 0.7$ , whereas the correlations among the four breeds were  $\sim 0.6$ . This is consistent with the central position of the AV breed found in a PC analysis (Cañas-Álvarez et al. (2014b)). Across all populations, the correlation of  $r$  between populations decreased with increasing marker distance (Figure 4). The estimates of persistence of phase were lower than the average among European cattle breeds (Gautier et al. (2007)), who found an average estimate of 0.77 for markers  $< 10$  kb apart using a 1536 SNP panel, and the estimates of De Roos et al (2008) using panels ranging from 1252 to 5237 SNPs. Simulations will be needed to evaluate whether the estimated correlations would allow the use of a combined training population for genomic selection.



**Figure 4: Pairwise persistence of phase in five beef breeds**

## Conclusion

LD has been estimated in five Spanish beef breeds. Average LD was similar between breeds and decreased with increasing distance in a consistent pattern. The persistence of phase was important at short distances but decreased rapidly. Only 5% of the SNPs would be necessary to achieve an average LD between adjacent markers of 0.22 (intermarker distance of  $\sim 80$  kb). Further studies are necessary to assess the size and structure of the training populations to make genomic selection successful in these populations.

## Acknowledgements

The financial support of Spanish AGL2010-15903 and UE FP7-289592-GENE2FARM grants made possible this research. The COLCIENCIAS Francisco José de Caldas 497/2009 and Spanish government BES-2011-045434 fellowships awarded to J. J. Cañas-Álvarez and to A. González-Rodríguez, respectively, are also recognized.

## Literature Cited

- Browning, B.L., and Browning, S. R. (2009). *Am. J. Hum. Genet.*, 84:210-223.
- Cañas-Álvarez, J. J., González-Rodríguez, A., Martín-Collado, D., et al. (2014a). *J. Anim. Sci.* (Submitted)
- Cañas-Álvarez, J. J., González-Rodríguez, A., Munilla, S., et al. (2014b). 10<sup>th</sup> *WCGALP*. (Submitted)
- Dekkers, J. C. M., and Hospital, F. (2002). *Nat. Rev. Genet.*, 3:22-32.
- De Roos, A. P. W., Hayes, B. J., Spelman, R. J., et al. (2008). *Genetics*, 179:1503-1512.
- Gautier, M., Faraut, T., Moazami-Goudarzi, K., et al. (2007). *Genetics*, 177:1059-1070.
- Hill, W. G., and Robertson, A. (1968). *Theor. Appl. Genet.*, 38:226-231.
- Matukumalli, L. K., Lawley, C. T., Schnabel, L. D., et al. (2009). *PLoS One*, 4:e5350.
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). *Genetics*, 157:1819-1829.
- Purcell, S., Neale, B., Todd-Brown, K., et al. (2007). *Am. J. Hum. Genet.*, 81:559-575.
- Sved, J. A. (1971). *Theor. Popul. Biol.*, 141:125-141.
- Villa-Angulo, R., Matukumally L. K., Gill, C. A., et al. (2009). *BMC Genetics*, 10:19.