

Linkage disequilibrium, persistence of phase, and effective population size in Spanish local beef cattle breeds assessed through a high-density single nucleotide polymorphism chip¹

J. J. Cañas-Álvarez,*² E. F. Mouresan,†³ L. Varona,†‡ C. Díaz,§
A. Molina,# J. A. Baro,|| J. Altarriba,†‡ M. J. Carabaño,§ J. Casellas,¶ and J. Piedrafita*⁴

*Grup de Recerca en Remugants, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193-Bellaterra, Barcelona, Spain; †Departamento de Anatomía, Embriología y Genética, Universidad de Zaragoza, 50013-Zaragoza, Spain; ‡Instituto Agroalimentario de Aragón (IA2), 50013-Zaragoza, Spain; §Departamento de Mejora Genética Animal, INIA, 28040-Madrid, Spain; #MERAGEM, Universidad de Córdoba, 14071-Córdoba, Spain; ||Departamento de Ciencias Agroforestales, Universidad de Valladolid, 34004-Palencia, Spain; and ¶Grup de Millora Genètica Molecular, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193-Bellaterra, Barcelona, Spain

ABSTRACT: Linkage disequilibrium (LD) and persistence of phase are fundamental approaches for exploring the genetic basis of economically important traits in cattle, including the identification of QTL for genomic selection and the estimation of effective population size (N_e) to determine the size of the training populations. In this study, we have used the Illumina BovineHD chip in 168 trios of 7 Spanish beef cattle breeds to obtain an overview of the magnitude of LD and the persistence of LD phase through the physical distance between markers. Also, we estimated the time of divergence based on the persistence of the LD phase and calculated past N_e from LD estimates using different alternatives to define the recombination rate. Estimates of average r^2 (as a measure of LD) for adjacent markers were close to 0.52 in the 7 breeds and decreased with the distance between markers, although in long distances, some LD still remained (0.07 and 0.05 for markers 200 kb and 1 Mb apart, respectively). A panel with a lower boundary of 38,000 SNP would

be necessary to launch a successful within-breed genomic selection program. Persistence of phase, measured as the pairwise correlations between estimates of r in 2 breeds at short distances (10 kb), was in the 0.89 to 0.94 range and decreased from 0.33 to 0.52 to a range of 0.01 to 0.08 when marker distance increased from 200 kb to 1 Mb, respectively. The magnitude of the persistence of phase between the Spanish beef breeds was similar to those found in dairy breeds. For across-breed genomic selection, the size of the SNP panels must be in the range of 50,000 to 83,000 SNP. Estimates of past N_e showed values ranging from 26 to 31 for 1 generation ago in all breeds. The divergence among breeds occurred between 129 and 207 generations ago. The results of this study are relevant for the future implementation of within- and across-breed genomic selection programs in the Spanish beef cattle populations. Our results suggest that a reduced subset of the SNP panel would be enough to achieve an adequate precision of the genomic predictions.

Key words: beef breeds, effective population size, high-density single nucleotide polymorphism chip, linkage disequilibrium, persistence of phase

© 2016 American Society of Animal Science. All rights reserved. J. Anim. Sci. 2016.94:2779–2788
doi:10.2527/jas2016-0425

¹This work was supported by a Ministerio de Ciencia e Innovación AGL 2010-15903 grant from the Spanish government, and by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number 289592 - Gene2Farm. The collaboration of breed societies and Juan José Arranz in collecting samples and the support of FEAGAS are also acknowledged.

²J. J. Cañas-Álvarez was awarded a COLCIENCIAS Francisco José de Caldas 497/2009 fellowship.

³E. F. Mouresan acknowledges a contract under grant agreement number 289592 - Gene2Farm.

⁴Corresponding author: jesus.piedrafita@uab.es

Received March 1, 2016.

Accepted May 1, 2016.

INTRODUCTION

A successful genomic selection scheme relies on the availability of dense-enough marker information to ensure persistent linkage disequilibrium (**LD**) and of a sufficiently large reference population to ensure accurate estimation of marker effects. Both premises have been largely studied and successfully applied in large single-breed dairy cattle populations such as the Holstein breed, where the commercial chip containing around 50,000 SNP has proven to be sufficient for within-breed genomic selection (de Roos et al., 2008). However, application of genomic selection in small census populations, such as local breeds, is more controversial, and differences in LD between markers and QTL and in phase persistence across breeds need to be assessed in order to determine the required marker density that will be optimal for genomic selection. Few studies have dealt with the determination of LD and LD persistence across breeds and no studies have been found comparing LD phase persistence across small local populations using high-density SNP panels. However, there are many cases in which small local breeds could benefit from a joint genomic selection scheme. This is the case of the local beef breeds in the Spain, where a number of independent selection schemes have been developed during the last decades for breeds maintained genetically separated but of a likely common origin. The main objective of this research is to assess the magnitude of LD and the persistence of phase among local, small-sized beef breeds in the Iberian Peninsula, to estimate past effective population size from LD, and to understand novel aspects of its evolutionary history, measured from data of a high-density chip (more than 777,000 SNPs). This can provide information on the behavior of these parameters of beef cattle of limited censuses and under a nonintensive directional selection. From this overview, we also attempt to arrive at recommendations on the SNP panel size needed for both within- and across-breed genomic selection.

MATERIAL AND METHODS

Animal Samples and SNP Genotyping

A total of 504 animals belonging to 168 trios (sire/dam/son) from 7 Spanish beef cattle breeds, including Asturiana de los Valles (**AV**; $n = 75$), Avileña–Negra Ibérica (**ANI**; $n = 72$), Bruna dels Pirineus (**BP**; $n = 75$), Morucha (**Mo**; $n = 75$), Pirenaica (**Pi**; $n = 72$), Retinta (**Re**; $n = 69$), and Rubia Gallega (**RG**; $n = 66$), were chosen from different and separated geographical areas, taking care to avoid known relationships. The blood samples were collected by veterinary practitioners in

commercial farms from the caudal vein of animals in 5-mL tubes with EDTA as the anticoagulant, following the recommendations of the Joint Working Group on Refinement (1993). This procedure did not require approval from an ethical committee. Genomic DNA was extracted following the protocol described in PrepFiler Forensic DNA Kit of Applied Biosystems (Foster City, CA), using MagMAX Express-96 Magnetic Particle Processor of Applied Biosystems. High-density genotypes (777,962 SNP) were obtained by using the BovineHD Genotyping BeadChip (Illumina Inc., San Diego, CA). The SNP mapped to the UMD3.1 assembly (Zimin et al., 2009). All laboratory work was done at a commercial laboratory (Xenética Fontao, Lugo, Spain). Markers kept for the study belonged to autosomal chromosomes. The SNP that have the same genomic location (3,014 SNP, 0.387% of the total, the same number for all breeds) and those with across-breed Mendelian error rates greater than 5% were removed. Additional requirements were individual call rate ≥ 0.95 , SNP call rate ≥ 0.95 , and minor allele frequency (**MAF**) > 0.05 . Quality control was performed using PLINK version 1.07 software (Purcell et al., 2007). After quality control, we had 629,869 markers, the same in all breeds, distributed across the 29 autosomal chromosomes (80.96% of the initial information), covering a total of 2,508,474.8 kb with 1 marker every 3.98 kb, on average.

Linkage Disequilibrium Decay

Sampling animals in trios yields improved accuracy of the estimated haplotypes (Marchini et al., 2006). Parental haplotypes within each breed were inferred by using the “trio” option of Beagle software version 3.3.2 (Browning and Browning, 2009). To increase the phasing accuracy, we ran 100 iterations of the phasing algorithm and sampled 100 haplotype pairs for each individual during each iteration of the phasing algorithm. Linkage disequilibrium was computed from the phases of the parents’ trios across all autosomal chromosomes within each breed by using PLINK version 1.07 software (Purcell et al., 2007). The LD measure adopted was the r^2 statistic (Hill and Robertson, 1968), which is generally accepted as the most robust and best interpretable LD parameter (Kruglyak, 1999). It was computed as

$$r^2 = (p_{11} - p_1 p_2)^2 / (p_1 q_1 p_2 q_2),$$

in which p_1 and q_1 are the minor and major allele frequencies in SNP 1, respectively; p_2 and q_2 are the minor and major allele frequencies in SNP 2, respectively; and p_{11} corresponds to the frequency of haplotypes with minor alleles in SNP 1 and SNP 2 throughout the whole population. Average r^2 at a variety of distances was

computed for each breed by stacking all SNP in bins of 5 kb across all autosomes. The number of haplotype pairs considered to estimate average r^2 ranged from approximately 920,000 for markers 5 kb apart to approximately 760,000 for markers 2 Mb apart. The average r^2 for SNP pairs in each bin was plotted against physical distance using an R environment (R Core Team, 2014). The average r^2 between adjacent markers for different densities after thinning for a certain percentage of markers chosen at random (10, 5, and 1%) was also estimated.

Persistence of Linkage Disequilibrium Phase

Persistence of phase quantifies the extent to which there is an excess of the same haplotypes in all breeds. In this case the signed square root of r^2 was used as a measure of LD, $r^2 = (p_{11} - p_1p_2)^2/(p_1q_1p_2q_2)^{1/2}$, computed with the PLINK software (Purcell et al., 2007). If r^2 between 2 markers is equal in 2 populations but their corresponding r has the opposite sign, the gametic phase is reversed. Persistence of phase was estimated as the Pearson correlation coefficient between the LD between 2 specific markers in 2 different breeds as in Badke et al. (2012):

$$r_{k,k'} = [\sum (r_{(k)} - \bar{r}_{(k)}) (r_{(k')} - \bar{r}_{(k')})] / S_{(k)} S_{(k')},$$

in which $r_{k,k'}$ is the correlation of phases between $r_{(k)}$ in population k and $r_{(k')}$ in population k' ; $S_{(k)}$ and $S_{(k')}$ are the SD of $r_{(k)}$ and $r_{(k')}$, respectively; and $\bar{r}_{(k)}$ and $\bar{r}_{(k')}$ are the average of r across all SNP within the intervals mentioned above for population k and k' , respectively. To represent the evolution of the persistence of phase with marker separation, the results were stacked in bins of 10 kb from 0 to 100 kb of marker distance and in bins of 100 kb from 100 kb to 1 Mb of marker distance. In addition, to highlight some results, a heat map graph was used to represent the ranges of correlation of r among populations. Calculations and subsequent plots were constructed using an R environment (R Core Team, 2014).

Past Effective Population

Size and Time since Divergence

To estimate the effective population size, N_e , a correction for sample size was applied for all r^2 values following Nilsen et al. (2008):

$$r^2_{\text{corrected}} = [r^2_{\text{computed}} - (1/n)] / [1 - (1/n)],$$

in which n represents the number of haplotypes in the sample. When the corrected r^2 was negative, $r^2_{\text{corrected}}$ was set to 0. The N_e was then estimated by applying the approximate expectation of r^2 in absence of mutation (Sved, 1971):

$$r^2 = 1/(1 + 4N_e c) \approx N_e = (1/4c)[1/(r^2 - 1)],$$

in which N_e is the effective population size, r^2 is the corrected LD estimate as indicated before, and c is the recombination frequency, which is replaced by map distance in Morgans. Furthermore, N_e was estimated at different generations in the past (T), calculated by the approximation $T = 1/2c$ (Hayes et al., 2003). To estimate N_e , 3 models were used. The first model (**M1**) assumes the generalization 1 Mb = 1 cM; in the second model (**M2**), specific recombination rates for each chromosome in cattle were used (Arias et al., 2009), as suggested by Qanbari et al. (2010) and Corbin et al. (2010); the third model (**M3**) is a nonlinear least squares approach also using specific recombination rates for each chromosome:

$$r^2_i = 1/(a + 4bd_i) + e_i \approx (1/4d_i)(1/(r^2_i - a)).$$

In this equation, b estimates the N_e , r^2_i is the estimate of r^2 for the i th marker pair at a distance d_i in Morgans, and the a parameter takes values of 1 or 2 when the mutation is ignored (in the case of the Sved formula) or is taken into consideration (Tenesa et al., 2007), respectively. Calculations and subsequent plots were constructed using an R environment (R Core Team, 2014; Warnes et al., 2015).

Furthermore, the time since breed divergence (T) was estimated using the whole genome. Assuming that the variance of r remains constant in a pair of breeds, the expected correlation of $r_{k,k'}$ equals e^{-2cT} , in which $r_{k,k'}$ is the persistence of phase in populations k and k' , c is the recombination distance between markers (in Morgans), and T is the number of generations since the breed's divergence (de Roos et al., 2008). A linear regression of the natural logarithm of correlation of phases on the pairwise distance between markers, ranging between 5 and 400 kb, was fitted and the slope of this regression was an estimate of $-2T$.

RESULTS AND DISCUSSION

Linkage Disequilibrium

Linkage disequilibrium maps are fundamental tools for exploring the genetic basis of economically important traits in livestock species (Andreescu et al., 2007; McKay et al., 2007). Average r^2 tended to decrease with increasing distance between pairs of markers in all populations studied, the most rapid decline being seen over the first 100 kb. On average, LD decreased from 0.58 to 0.32, 0.18, 0.11, 0.07, and 0.045 for 5 kb, 20 kb, 50 kb, 100 kb, 200 kb, and 1 Mb of marker distance, respectively (Table 1). From 1 to 2 Mb of marker distance, the estimates showed almost no change. The SD of LD estimates were similar

Table 1. Average r^2 values and SD across all autosomes at various marker distances in 7 Spanish beef breeds

| Breed ¹ | 5 kb ² (SD) | 20 kb (SD) | 50 kb (SD) | 100 kb (SD) | 200 kb (SD) | 1 Mb (SD) | 2 Mb (SD) |
|--------------------|------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| AV | 0.571 (0.388) | 0.309 (0.328) | 0.165 (0.230) | 0.092 (0.152) | 0.052 (0.088) | 0.031 (0.044) | 0.029 (0.040) |
| ANI | 0.582 (0.395) | 0.328 (0.341) | 0.188 (0.250) | 0.116 (0.174) | 0.076 (0.115) | 0.047 (0.067) | 0.042 (0.058) |
| BP | 0.579 (0.392) | 0.320 (0.336) | 0.179 (0.242) | 0.107 (0.167) | 0.066 (0.105) | 0.038 (0.054) | 0.033 (0.046) |
| Mo | 0.577 (0.395) | 0.319 (0.337) | 0.180 (0.243) | 0.108 (0.167) | 0.067 (0.105) | 0.041 (0.058) | 0.037 (0.052) |
| Pi | 0.593 (0.396) | 0.340 (0.348) | 0.201 (0.260) | 0.129 (0.187) | 0.088 (0.129) | 0.058 (0.080) | 0.051 (0.071) |
| Re | 0.583 (0.396) | 0.328 (0.342) | 0.190 (0.251) | 0.120 (0.177) | 0.079 (0.120) | 0.053 (0.075) | 0.047 (0.066) |
| RG | 0.581 (0.394) | 0.325 (0.338) | 0.185 (0.245) | 0.115 (0.172) | 0.075 (0.114) | 0.053 (0.073) | 0.049 (0.067) |

¹AV = Asturiana de los Valles; ANI = Avileña–Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

²The averages correspond to intervals including a range of 5 kb below the specified number.

among breeds, ranging from around 0.39 at distances of 5 kb to around 0.05 at distances of 2 Mb (Table 1). The SD diminished in parallel to the LD estimates, although the CV were greater at longer distances to the point that values of LD at larger distances could overlap with LD at short distances. This scenario might highlight the variable effect on the LD structure of recent events in these breeds. The larger SD at shorter distances can be related to the shape of the distribution of r^2 values as a function of distance (Supplemental File S1; see the online version of the article at <http://journalofanimalscience.org>). For adjacent markers (at an average distance of approximately 4 kb), 27 to 32% complete LD ($r^2 > 0.95$) was found across breeds followed by markers close to complete equilibrium ($r^2 < 0.05$), with percentages in the 16 to 17% range. Intermediate estimates were almost evenly distributed between these extreme values. As the distance between markers increased, the proportion of pairs of markers in complete LD decreased, and beyond 100 kb, no markers in complete LD were found.

Some authors have investigated the potential effect of MAF on LD estimates (Lu et al., 2012; Espigolan et al., 2013). It has been shown that SNP of divergent MAF, on average, have different LD properties (Pritchard and Przeworski, 2001). Therefore, to determine how estimates of LD could be affected by MAF in these populations, a more restrictive MAF (<0.15) was imposed for the quality control. As expected, larger LD estimates were obtained (around 15 to 20% of increase for distances of 5 to 100 kb apart and less for longer distances; results not shown). A similar pattern was observed by Espigolan et al. (2013) in Nellore cattle, where they found an increase in LD estimates depending upon the MAF criterion. Furthermore, we found that the dependence of r^2 on distance is stronger than its dependence on the difference in frequencies (results not shown), as found by Qanbari et al. (2010).

There were differences in average r^2 at a particular marker distance depending upon chromosomes. This is shown in Supplemental File S2 (see the online version of the article at <http://journalofanimalscience.org>), where average r^2 in all breeds for each chromosome at

different marker distances are represented as circles of different colors. This variability among chromosomes leads to some changes in the ranking of breeds LD when only the markers at a particular chromosome are considered. Differences in LD at the chromosome level have been observed in different species such as cattle (Lu et al., 2012) and chickens (Andreescu et al., 2007) and, to some extent, could be attributed to changes in the density of markers among different chromosomes (Andreescu et al., 2007). Moreover, a wide variation in autosomal recombination rates can lead to a marked diversity in the pattern of LD in different genomic regions (Arias et al., 2009) and chromosomes, but the differences can also be due to heterozygosity levels, genetic drift, or selection (Qanbari et al., 2010). For differences in LD across genomic regions, the causes of LD may have acted differently at specific genomic regions at singular locations among these populations (González-Rodríguez et al., 2014). Differences across breeds in demographic history and genetic drift, admixture or migrations, population subdivision, selection (hitchhiking effect and epistatic selection), variable mutation rates, and gene conversion could be modifying the relationship between LD estimates and the physical distance between loci (Ardlie et al., 2002) and explain the observed variability among breeds. The differences found suggest that additional studies are needed to identify particular regions that may exhibit different degrees of LD and to evaluate its potential use for detection of QTL and signatures of selection (Pérez O'Brien et al., 2014) and their consequences for the implementation of a genomic selection program.

Overall, the differences in LD estimates across Spanish beef breeds were very small at short distances and mildly increased when markers distance was greater than 50 kb. The ranking of the average was kept the same across distance, with Pi as the breed with the greatest LD and AV that with the lowest (Table 1; Fig. 1).

Our results are in the range of those obtained in Australian taurine cattle (Porto-Neto et al., 2014) and other cattle populations. A study of 19 breeds including continental European and British dairy and beef cattle, African and indicine cattle, and composite breeds

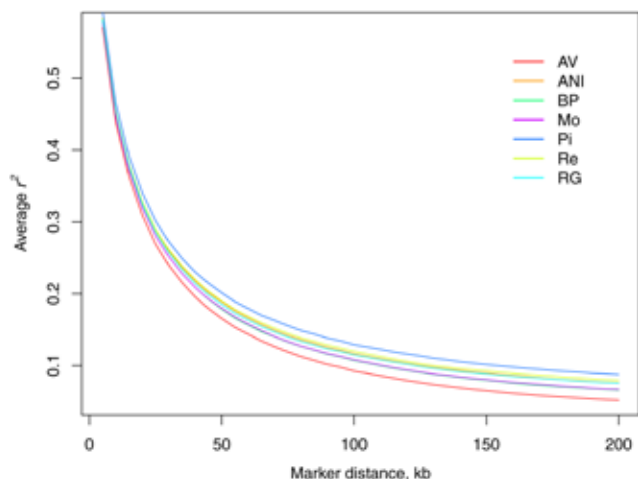


Figure 1. Decay of average r^2 from markers 0 to 200 kb apart across all autosomes in 7 Spanish beef breeds. AV = Asturiana de los Valles; ANI = Avileña–Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

showed LD values ranging from 0.37 to 0.65 at 5 kb of distance between markers and from 0.08 to 0.25 for markers 10 kb apart (Salomon-Torres et al., 2014). In Nellore, the LD estimates were lower than ours at short intervals (5 kb) and similar at a distance between markers 100 kb apart (Espigolan et al., 2013), which was in agreement with the lower LD at short marker distance found in indicine vs. taurine breeds by Porto-Neto et al. (2014). In both studies, the BovineHD Genotyping BeadChip from Illumina was used. With a 50,000 SNP chip, in German Holstein-Friesian cattle, Qanbari et al. (2010) found a lower LD than in our study at short distances but a greater LD at distances of 75 to 120 kb. The causes of the differences can be due to some ascertainment bias, mainly in the Nellore study, but these differences also probably reflect different evolutionary histories (larger N_e in Nellore; Porto-Neto et al., 2014) and selection practices. On the other hand, the extent of LD in our study was close to the one estimated in humans (Tenesa et al., 2007) but lower than the one estimated in horses (Corbin et al., 2010) and pigs (Badke et al., 2012) at distances of up to 100, 500, and 1,000 kb, respectively. The comparison of LD levels between studies, however, should be taken with caution, as LD depends on sample sizes, marker densities, and recent and historical population demographics (Pritchard and Przeworski, 2001).

The decay of LD in a genome determines the power of QTL detection in genomewide association studies and indicates the required marker density for accurate genomic selection. It has been shown that to have the same power to detect the association between the QTL and the marker, the sample size must be increased by roughly $1/r^2$ when compared with the sample size for directly detecting the association with the causal mutation (Kruglyak, 1999; Pritchard and Przeworski,

Table 2. Average r^2 (average distance in kb) between adjacent markers across all autosomes depending upon the proportion of SNP retained in 7 Spanish beef breeds

| Breed ¹ | 100% | 10% | 5% | 1% |
|--------------------|--------------|---------------|---------------|----------------|
| AV | 0.509 (4.03) | 0.284 (40.20) | 0.208 (80.19) | 0.087 (401.63) |
| ANI | 0.521 (4.09) | 0.302 (40.85) | 0.228 (81.34) | 0.110 (407.54) |
| BP | 0.518 (4.08) | 0.295 (40.65) | 0.218 (81.07) | 0.098 (405.90) |
| Mo | 0.516 (4.09) | 0.295 (40.75) | 0.218 (81.27) | 0.099 (405.30) |
| Pi | 0.533 (4.13) | 0.314 (41.26) | 0.239 (82.20) | 0.120 (411.16) |
| Re | 0.523 (4.11) | 0.304 (40.97) | 0.229 (81.57) | 0.113 (406.95) |
| RG | 0.520 (4.08) | 0.299 (40.74) | 0.226 (81.02) | 0.110 (406.34) |

¹AV = Asturiana de los Valles; ANI = Avileña–Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

2001). In the context of genomic selection, the accuracy depends both on the amount of LD between QTL and markers and the number of records available to estimate marker effects (Toosi et al., 2010).

Linkage disequilibrium was computed between adjacent markers, depending upon the proportion of the number of SNP retained in the analysis (Table 2). The average r^2 across all chromosomes was 0.52 for adjacent markers in the complete data set, with an inter-marker distance of 4 kb, and decreased to 0.30, 0.22, and 0.10 when 10, 5, and 1% of the markers, randomly chosen, were included in the analysis. Large differences were not observed among the 7 breeds evaluated.

A mean r^2 value of 0.20 between adjacent markers is considered enough to achieve an accuracy of 0.85 for genomic breeding value estimation (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 (38,000 markers) of the BovineHD chip, which corresponds to an average genomic distance of 80 kb (Table 2). However, given that markers with $r^2 = 1$ will be excluded in genomic selection and because of the high variability of r^2 at small distances (Corbin et al., 2010), this is probably an underestimation of the actual number of SNP needed for single breed genomic selection. Furthermore, this number can vary according to the relationship between training and evaluation animals (Meuwissen, 2009).

Persistence of Linkage Disequilibrium Phase

For genomic selection, the usefulness of markers in a population other than the reference population will depend on the persistence of the LD phase between the reference population and the second population (Dekkers and Hospital, 2002). The persistence of phase between 2 populations for a given marker distance has been calculated as the correlation of the r estimates, that is, the signed squared roots of r^2 . The correlations of r between populations are a result of their genetic

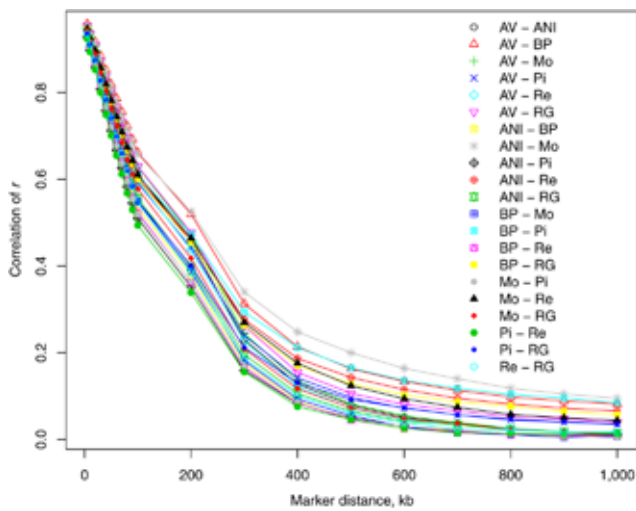


Figure 2. Decay of pairwise persistence of phase across all autosomes in 7 Spanish beef breeds. AV = Asturiana de los Valles; ANI = Avileña–Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

relationship and history. The correlation of r was computed for each marker pair across all chromosomes among all pairwise population combinations over a series of genomic distance intervals using bins of 10 kb from 0 to 100 kb of intermarker distance and bins of 100 kb from 100 to 1,000 kb of intermarker distance. The observed correlation of r between 2 populations decreased with increasing marker distance, with a behavior similar to LD (Fig. 2). At 10 kb apart, the persistence was high and positive, ranging from 0.89 to 0.94, which means that at short distances, there is an excess of similar haplotypes in the same phases among all populations. With increasing distance, the persistence diminished to the ranges 0.70 to 0.82, 0.49 to 0.66, and 0.34 to 0.52 for markers 50, 100, and 200 kb apart, respectively. It is worth mentioning that the estimates of persistence of phase are averages over a certain bin. A positive average estimate does not mean that all estimates of persistence are positive. In fact, at all marker distances, we have found negative estimates. The percentage of negative estimates (reversed phases) increased with marker distance, perhaps as a consequence of recent events, reaching values of 29 to 30%, 36 to 38%, 40 to 43%, and 43 to 48% for average marker intervals of 20 kb, 50 kb, 100 kb, and 1 Mb, respectively, across breeds.

The comparison of heat map graphics corresponding to 2 of the distances analyzed confirms that there is a lower persistence of phase (lighter color) for greater distances among markers (Fig. 3). There are also differences between the pairwise persistence between breeds across distances. At the 2 distances shown, but also at distances not shown in the figures, the group formed by the ANI, Mo, and Re breeds presents the greatest values of persistence among them, indicating their ge-

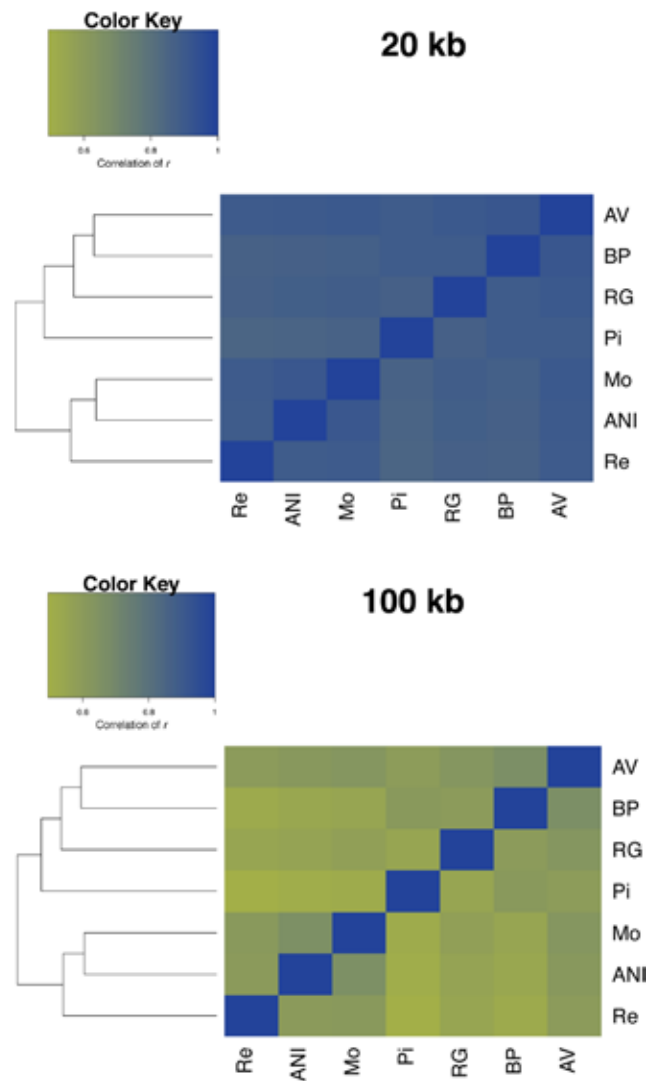


Figure 3. Persistence of phase matrices across all autosomes among 7 Spanish beef breeds evaluated at 20 and 100 kb.

netic proximity, whereas the persistence is the lowest between Pi and BP and the rest of the breeds. Asturiana de los Valles, on the contrary, has a pattern of persistence similar to the rest of the breeds, confirming its central position among the Spanish breeds seen by Cañas-Álvarez et al. (2015). Heat maps also allowed us to present the grouping of breeds by a dendrogram, synthesizing the divergence among breeds. At the different distances, the same 2 clusters are clearly stated. The first one includes ANI, Mo, and Re, whereas the second one includes the rest of the breeds. This second cluster, however, does not provide a consistent grouping with the relationships described in a previous study in these breeds (Cañas-Álvarez et al., 2015).

The estimates of persistence of phase in the Spanish beef breeds were in the upper range of those observed by Gautier et al. (2007) at short marker distances using a 1,536 SNP panel. At greater distances, up to 500 kb, the distances were, in general, lower.

de Roos et al. (2008) found correlations of 0.97 for marker pairs 5 kb apart in closely related populations (black and white Holstein-Friesian bulls and red and white Holstein-Friesian bulls) but greater than 0.80 for less related cattle (Australian Holstein and Australian Angus), using panels ranging from 1,252 to 5,237 SNP. The correlation decreased with the distance in proportion to the degree of divergence of the breeds considered. In pigs, nevertheless, the persistence continued over larger distances, probably due to higher selection pressure and smaller N_e (Harmegnies et al., 2006).

We have described that at short distances, correlations of r between breeds were very high, although it rapidly diminished at longer distances. If we want to find markers that work consistently in 2 populations, we need to set a minimum correlation of 0.8 (as in de Roos et al., 2008). Similar to some results of these authors, in the Spanish beef breeds, correlations greater than 0.8 are found for markers up to 30 to 50 kb apart, depending upon the pair of breeds considered (not shown in tables). Because of that, the number of markers equally distributed across the genome must be in the range of 50,000 to 83,000, depending upon the breed's combination. This implies that, for using SNP information across breeds in genomic selection, for some pairs of breeds, we would need to use at least twice the number of markers needed for within-population genomic selection.

Time since Divergence between Breeds

The slope of the persistence of phase, that is, the correlation of r values, has a strong implication in determining the time in which, possibly, the breeds diverged. Although there are differences among chromosomes in recombination rates, Arias et al. (2009) established that the average genetic distance (cM) per megabase was equal to 1.25. Using this ratio, the breeds diverged between 129.4 (AV–Mo pair) and 206.8 generations ago (Pi–Re pair; Supplemental File S3 [see the online version of the article at <http://journalofanimalscience.org>]). The Pi breed diverged first from the rest of breeds, whereas AV showed the lowest times of divergence from the other breeds. The results parallel those observed in the graphics of persistence of phase between populations and to the neighbor joining representation of the pairwise Nei's distances (Cañas-Álvarez et al., 2015). The range of divergence times found in our study was lower than that observed for the Australian Holstein and Australian Angus ($T = 364$) pair, breeds specialized in dairy and meat production, respectively. However, our estimates were in the order of the divergence time between 2 dairy breeds: New Zealand Friesian and New Zealand Jersey ($T = 191$; de Roos et al., 2008). These results would suggest that the divergence between dairy

and beef breeds preceded the subsequent divergence among breeds of the same aptitude.

If we consider an approximate average generation interval of 6.5 yr for the Spanish beef breeds (Cañas-Álvarez et al., 2014), their divergence occurred between approximately 841 and 1,344 years ago, which corresponds to the years of the first half of the Middle Ages. This fact could be consistent with Spanish history at those times, when there was a political subdivision between the Christian and Muslim kingdoms in the north and south of the country, respectively. The populations located in the ancient Muslim territory (Re, Mo, and ANI) were surely more influenced by the African populations (Decker et al., 2014), whereas the populations located in north Spain were more related to the rest of the European populations. However, after the political reunification of the Spanish kingdoms in 1492, the contact between populations was restored by the instauration of the transhumance, which implies a large movement of animals, especially of the AV, explaining the central position of that population (Cañas-Álvarez et al., 2015).

Effective Population Size

We have estimated the N_e at different generations in the past from LD estimates using different approaches. The first one assumes the generalization widely used in the literature that 1 Mb = 1 cM (M1). The second model takes into account the specific variation in the rate of recombination among chromosomes as in Arias et al. (2009). Both estimates are based on Sved's (1971) formula on the expectation of r^2 . The third approach (M3) is built on the Tenesa et al. (2007) formula and iteratively estimates the values of N_e and the existence of mutation (a parameter). The use of a different recombination rate among chromosomes had a limited impact on the estimation of N_e . In fact, the correlations among the 3 methods across breeds were close to 1. The second model tended to give values similar to those of M1, with inconsistent variations across breeds. As expected, the differences increased with the number of generations (Supplemental File S4; see the online version of the article at <http://journalofanimalscience.org>). The N_e estimates in M2 ranged between 2,079.9 and 2,487.8 for Pi and AV, respectively, at 1,500 generations ago, whereas they ranged between 26.3 and 31.4 1 generation ago for RG and Re, respectively. The third model yielded N_e estimates lower than those of M1 and M2, in general. Specifically, for the same generation and breeds, the estimates of N_e in M3 were 1,753.1 and 2,171.1 at 1,500 generations ago and 24.5 and 31 to the same breeds as in M2. Parameter a determines the value of expected r^2 when the distance between markers is effectively 0 (Corbin et al., 2010). We have found estimates of a ranging from 1.0 to 2.1, with

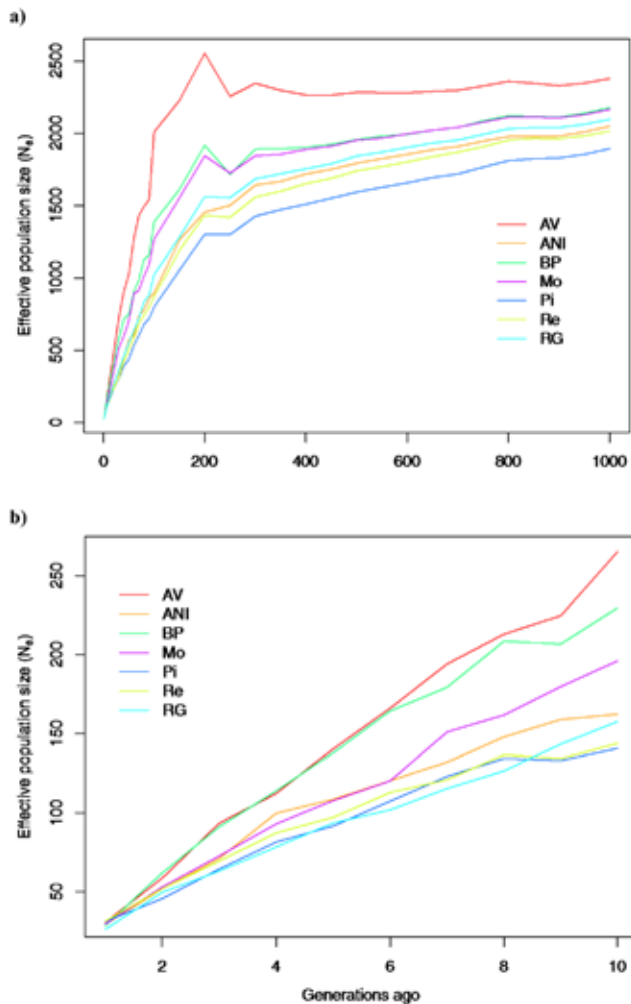


Figure 4. Effective population size estimated using specific recombination rates per chromosome (the second model) in the periods 0 to 1,000 generations ago (a) and 0 to 10 generations ago. AV = Asturiana de los Valles; ANI = Avileña–Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega. (b).

an average of 1.4. This would support, as in Corbin et al. (2010) in Thoroughbred horses, the alternative version of Sved's (1971) equation derived by Tenesa et al. (2007).

The results of historical N_e estimated with M2 from 1,000 generations ago to the present time are graphically shown in Fig. 4a. At the beginning of the interval of time studied, matching with the domestication period that started in the Neolithic period, N_e estimates oscillated from approximately 1,900 to approximately 2,300 for Pi and AV, respectively. Except for AV, a slow decline was observed until 200 generations ago (about 1,300 years ago). From this period, corresponding to the time the breeds diverged, a sharp decline in population size has continued until today. Asturiana de los Valles and BP breeds showed the greatest values of N_e along all generations, particularly in the last 10 (Fig. 4b). The N_e prior to the current generation stood at around 26 to 31 in all breeds studied in this work. These values are lower than those obtained from the analysis of pedigree records in

the same breeds (Cañas-Álvarez et al., 2014), which was expected because the available pedigree files allowed us to trace back a limited number of generations.

One way to test the performance of the 3 methods in the short term is to compare the N_e estimates with those found from average r^2 between nonsyntenic chromosomes to obtain an empirical null distribution. This represents an approximation of the LD that can be obtained by chance, assuming that the markers used have not undergone simultaneous selection. The mean of nonsyntenic r^2 values reflects both sampling of animals and genetic drift and may be expected to become lower with increases in both sample size and N_e (Corbin et al., 2010). The average estimates of nonsyntenic r^2 found in this study were very similar across breeds, in the 0.010 to 0.011 range (Supplemental File S5; see the online version of the article at <http://journalofanimalscience.org>). All LD values between nonsyntenic markers are within the interval of LD at long distances in these breeds (Table 1). The values observed here are much greater than are those observed in Australian Holstein-Friesian cattle (Khatkar et al., 2008) and in Thoroughbred horses (Corbin et al., 2010). The comparatively high values observed in our breeds suggest that some of the LD created by admixture during breed evolution (Cañas-Álvarez et al., 2015) has been kept until the present time. With respect to the effective size, assuming a value of the recombination rate of $c = 0.5$ for the nonsyntenic loci, N_e estimates ranged from 43 (Re) to 47.5 (AV and BP). This would suggest that all 3 methods used above provide downwardly biased estimates of N_e in the last generations or that estimates of N_e very much depend on how well the average LD estimates represent the whole LD patterns in the genome.

Estimates of N_e lower than those presented in our work were obtained in European and West African breeds, where N_e estimates near 500 for 50 generations ago and 80 for 5 generations ago were found (Gautier et al., 2007; de Roos et al., 2008; Qanbari et al., 2010). The observed drop in N_e potentially represents a number of scenarios, including some founder event, the formation of different breeds, and recently artificial reproduction techniques or a combination of any of them (Wang, 2005). To that decline, bottlenecks caused by the Great Famine of 1315 to 1322 in Europe could also have contributed, as reported in other breeds of cattle (Gautier et al., 2007; Villa-Angulo et al., 2009). Regarding the breeds studied, the different N_e found can also be the result of the initial confinement of the populations in different geographical areas; a drop of population censuses due to the partial replacement and uncontrolled crossbreeding of autochthonous breeds with foreign breeds, which occurred in the middle of the 20th century (García-Dory, 1986); and some genetic selection for specific characteristics.

Effective population size also determines the accuracy of genomic selection (Daetwyler et al., 2010), as that accuracy can be expressed as a function of the effective number of loci (Me) in a population (Goddard, 2009). The effective number of loci relates to the number of independent chromosome segments and equals the summation across chromosomes of $2N_eL/\log(4N_eL)$, L being the genome size in Morgans for each specific chromosome. In our case, with chromosome lengths as in Arias et al. (2009) and the current effective sizes calculated from nonsynthetic LD estimates, Me was in the range of 509 (Re) and 552 (AV and BP; Supplemental File S5 [see the online version of the article at <http://journalofanimalscience.org>]). From this value, using the deterministic formula of Daetwyler et al. (2010), we can estimate the size of the training population under different heritability scenarios.

Conclusions

Herein are presented the results of an overview of LD of Spanish beef cattle breeds using a high-density SNP panel. We observed that LD tended to decrease with increasing genetic distance in all populations studied. Similarly, the persistence of phase was high at short distances and decreased at a moderate rate. A minimum of 38,000 and 83,000 SNP markers would be needed for within-breed and across-breed genomic evaluation, respectively.

Effective population sizes of the breeds substantially decreased from 200 generations ago, and N_e in the generation previous to the current one was estimated to be in the interval of 26 to 47. These estimates allow us to predict the size of the training population in different heritability scenarios and also suggest that the use of several of the Spanish breeds can be a solution to generate a reference population because of the small size of each separate breed. The results of this study can help to indicate the breed composition of the joint reference population and also suggest that with the current commercial platforms, a high-density chip would be necessary for across-breed genetic evaluation.

LITERATURE CITED

- Andrescu, C., S. Avendano, S. R. Brown, A. Hassen, S. J. Lamont, and J. C. M. Dekkers. 2007. Linkage disequilibrium in related breeding lines of chickens. *Genetics* 177:2161–2169. doi:10.1534/genetics.107.082206
- Ardlie, K., L. Kruglyak, and M. Seielstad. 2002. Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* 3:299–309. doi:10.1038/nrg777
- Arias, J. A., M. Keehan, P. Fisher, W. Coppieters, and R. Spelman. 2009. A high density linkage map of the bovine genome. *BMC Genet.* 10:18. doi:10.1186/1471-2156-10-18
- Badke, Y. M., R. O. Bates, C. W. Ernst, C. Schwab, and J. P. Steibel. 2012. Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* 13:24. doi:10.1186/1471-2164-13-24
- Browning, B. L., and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84:210–223. doi:10.1016/j.ajhg.2009.01.005
- Cañas-Álvarez, J. J., A. González-Rodríguez, D. Martín-Collado, C. Avilés, J. Altarriba, J. A. Baro, L. F. De la Fuente, C. Díaz, A. Molina, L. Varona, and J. Piedrafito. 2014. Monitoring changes in the demographic and genealogical structure of the main Spanish local beef breeds. *J. Anim. Sci.* 92:4364–4374. doi:10.2527/jas.2013-7420
- Cañas-Álvarez, J. J., A. González-Rodríguez, S. Munilla, L. Varona, C. Díaz, J. A. Baro, J. Altarriba, A. Molina, and J. Piedrafito. 2015. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density chip. *J. Anim. Sci.* 93:5164–5174. doi:10.2527/jas.2015-9271
- Corbin, L. J., S. C. Blott, J. E. Swinburne, M. Vaudin, S. C. Bishop, and J. A. Woolliams. 2010. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* 41:8–15. doi:10.1111/j.1365-2052.2010.02092.x
- Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021–1031. doi:10.1534/genetics.110.116855
- Decker, J. E., S. D. McKay, M. M. Rolf, J. Kim, A. Molina-Alcala, T. S. Sonstegard, O. Hanotte, A. Götherström, C. M. Seabury, L. Praharani, M. E. Babar, L. C. A. Regitano, M. A. Yildiz, M. P. Heaton, W. Lui, C. Lei, J. M. Reecy, M. Saif-Ur-Rehman, R. D. Schnabel, and J. F. Taylor. 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10:e1004254. doi:10.1371/journal.pgen.1004254
- Dekkers, J. C. M., and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3:22–32. doi:10.1038/nrg701
- de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179:1503–1512. doi:10.1534/genetics.107.084301
- Espigolan, R., F. Baldi, A. A. Boligon, F. R. Souza, D. G. Gordo, R. L. Tonussi, D. F. Cardoso, H. N. Oliveira, H. Tonhati, M. Sargolzaei, F. S. Schenkel, R. Carvalheiro, J. A. Ferro, and L. G. Albuquerque. 2013. Study of whole genome linkage disequilibrium in Nelore cattle. *BMC Genomics* 14:305. doi:10.1186/1471-2164-14-305
- García-Dory, M. A. 1986. Las razas bovinas autóctonas de España están en vías de extinción. (In Spanish.) *Quercus* 16:4–9.
- Gautier, M., T. Faraut, K. Moazami-Goudarzi, V. Navratil, M. Foglio, C. Grohs, A. Boland, J.-G. Garnier, D. Boichard, G. M. Lathrop, I. G. Gut, and A. Eggen. 2007. Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* 177:1059–1070. doi:10.1534/genetics.107.075804
- Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximization of long term response. *Genetica (The Hague)* 136:245–257. doi:10.1007/s10709-008-9308-0
- González-Rodríguez, A., M. A. Toro, L. Varona, M. J. Carabaño, J. J. Cañas-Álvarez, J. Altarriba, T. B. R. da Silva, J. A. Baró, A. Molina, and C. Díaz. 2014. Genome-wide analysis of genetic diversity in autochthonous Spanish populations of beef cattle. In: *Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada*. p. 255-257.

- Harmegnies, N., F. Farnir, F. Davin, N. Buys, M. Georges, and W. Coppieeters. 2006. Measuring the extent of linkage disequilibrium in commercial pig populations. *Anim. Genet.* 37:225–231. doi:10.1111/j.1365-2052.2006.01438.x
- Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* 13:635–643. doi:10.1101/gr.387103
- Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38:226–231. doi:10.1007/BF01245622
- Joint Working Group on Refinement. 1993. Removal of blood from laboratory mammals and birds. First report of the BVA/FRAME/RSPCA/UFPAW Joint Working Group on Refinement. *Lab. Anim.* 27:1–22. doi:10.1258/002367793781082412
- Khatkar, M., F. Nicholas, A. Collins, K. Zenger, J. Cavanagh, W. Barris, R. Schnabel, J. Taylor, and H. Raadsma. 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high density SNP panel. *BMC Genomics* 9:187. doi:10.1186/1471-2164-9-187
- Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22:139–144. doi:10.1038/9642
- Lu, D., M. Sargolzaei, M. Kelly, C. Li, G. Vander Voort, Z. Wang, G. Plastow, S. Moore, and S. P. Miller. 2012. Linkage disequilibrium in Angus, Charolais and crossbreed beef cattle. *Front. Genet.* 3(152):1–9. doi:10.3389/fgene.2012.00152
- Marchini, J., D. Cutler, N. Patterson, M. Stephens, E. Eskin, E. Halperin, S. Lin, Z. S. Quin, H. M. Munro, G. R. Abecasis, P. Donnelly, and International HapMap Consortium. 2006. A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.* 78:437–450. doi:10.1086/500808
- McKay, S. D., R. D. Schnabel, B. M. Murdoch, L. K. Matukumally, J. Aerts, W. Coppieeters, D. Crews, D. E. Neto, C. A. Gil, C. Gao, H. Mannen, P. Stothard, Z. Wang, C. P. Van Tassel, J. L. Williams, J. F. Taylor, and S. S. Moore. 2007. Whole genome linkage disequilibrium maps in cattle. *BMC Genet.* 8:74. doi:10.1186/1471-2156-8-74
- Meuwissen, T. H. E. 2009. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41:35. doi:10.1186/1297-9686-41-35
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Nilsen, H., B. Hayes, P. R. Berg, A. Roseth, K. K. Sundsaasen, K. Nilsen, and S. Lien. 2008. Construction of a dense SNP map for bovine chromosome 6 to assist the assembly of the bovine genome sequence. *Anim. Genet.* 39:97–104. doi:10.1111/j.1365-2052.2007.01686.x
- Pérez O’Brien, A. M., Y. T. Utsunomiya, G. Mészáros, D. M. Bickhart, G. E. Liu, C. P. Van Tassel, T. S. Sonstegard, M. V. D. Da Silva, J. F. García, and J. Sölkner. 2014. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet. Sel. Evol.* 46:19. doi:10.1186/1297-9686-46-19
- Porto-Neto, L., J. W. Kijas, and A. Reverter. 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet. Sel. Evol.* 46:22. doi:10.1186/1297-9686-46-22
- Pritchard, J. K., and M. Przeworski. 2001. Linkage disequilibrium in humans: Models and data. *Am. J. Hum. Genet.* 69:1–14. doi:10.1086/321275
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575. doi:10.1086/519795
- Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, and H. Simianer. 2010. The pattern of linkage disequilibrium in German Holstein cattle. *Anim. Genet.* 41:346–356. doi:10.1111/j.1365-2052.2009.02011.x
- R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>. (Accessed July 8, 2014.)
- Salomon-Torres, R., L. K. Matukumalli, C. P. Van Tassel, C. Villa-Angulo, V. M. Gonzalez-Vizcarra, and R. Villa-Angulo. 2014. High density LD-based structural variations analysis in cattle genome. *PLoS One* 9:e103046. doi:10.1371/journal.pone.0103046
- Sved, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Popul. Biol.* 2:125–141. doi:10.1016/0040-5809(71)90011-6
- Tenesa, A., P. Navarro, B. J. Hayes, D. L. Duffy, G. M. Clarke, M. E. Goddard, and P. M. Visscher. 2007. Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17:520–526. doi:10.1101/gr.6023607
- Toosi, A., R. L. Fernando, and J. C. M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. *J. Anim. Sci.* 88:32–46. doi:10.2527/jas.2009-1975
- Villa-Angulo, R., L. K. Matukumalli, C. A. Gill, J. Choi, C. P. Van Tassel, and J. J. Grefenstette. 2009. High-resolution haplotype block structure in the cattle genome. *BMC Genet.* 10:19. doi:10.1186/1471-2156-10-19
- Wang, J. 2005. Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360:1395–1409. doi:10.1098/rstb.2005.1682
- Warnes, G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, and B. Venables. 2015. gplots: Various R programming tools for plotting data. R package version 2.16.0. <https://cran.r-project.org/web/packages/gplots> (Accessed May 14, 2014).
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, R. Hanrahan, G. Pertea, C. P. Van Tassel, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42. doi:10.1186/gb-2009-10-4-r42