Genome-wide Analysis of Genetic Diversity in Autochthonous Spanish Populations of Beef Cattle

A. González-Rodríguez[1], **M.A. Toro**[2], **L. Varona**[1], **M. J Carabaño**[3], **J. J. Cañas-Álvarez**[4], **J. Altarriba**[1], **T.B R. da Silva**[3,5], **J. A. Baró**[6], **A. Molina**[7], **and C. Díaz**[3].

[1]Universidad de Zaragoza, Spain, [2]Universidad Politécnica de Madrid, Spain, [3]INIA, Spain, [4]Universitat Autònoma de Barcelona, Spain, [5]UNESP-Jaboticabal, Brazil, [6]Universidad de Valladolid, Spain, [7]Universidad de Córdoba, Spain

ABSTRACT: The BovineHD 770K BeadChip was used on 116 sire/dam/offspring triplets from five Spanish local beef cattle breeds (Asturiana de los Valles, Avileña-Negra Ibérica, Bruna dels Pirineus, Pirenaica and Retinta). The parents were chosen to be as unrelated as possible. We calculated the molecular coancestry for every pair of individuals (within and between populations). From them, Reynolds, Nei's minimum and standard distances and average F_{ST} were calculated for each subset of markers at every Mb along the autosomal chromosomes. The results of the total genetic diversity showed an average F_{ST} value of 0.049 (0.015). However, the F_{ST} vary enormously across genomic regions, showing remarkable differences with respect to the average estimate. In particular, regions of the chromosomes 2, 5, 6, 11, 13 and 18 showed a global F_{ST} greater than 0.15. In these regions a total of 77 genes were identified.

Keywords: beef cattle; molecular coancestry; genetic diversity

Introduction

After the process of domestication of *Bos Taurus Taurus,* selection and differentiation has resulted in the formation of a wide range of breeds (Scherf (2000)). Among them, most of the Spanish autochthonous cattle breeds are now fully oriented to beef production under a variety of production systems. They are considered an important reservoir of genetic variability besides having cultural and environmental values. The origin of all these populations is common, and the process of differentiation probably started not further away than the XVIII century, when some groups of individuals became genetically isolated. Later on, subsequent processes of selection and adaptation defined the current standards that characterize these breeds (Beja-Pereira et al. (2003)).

The objective of this study is to locate genome regions, along the autosomal chromosomes, associated with genetic differentiation between five of these populations (Asturiana de los Valles –AV–, Avileña-Negra Ibérica –ANI–, Bruna dels Pirineus –BP–, Pirenaica –Pi– and Retinta –Re–) with the aim of identifying potential genes involved in that process.

Materials and Methods

Animals and sample size. A total of 116 triplets (sire/dam/offspring) were collected from five Spanish beef cattle populations (AV, n=25; ANI, n=24; BP, n=25; Pi, n=24; Re, n=18). The parents were chosen to capture the existing variability in the breeds, by minimizing the genealogical coancestry among them.

SNP genotyping and phasing. Genomic DNA was extracted by standard protocols. High density SNP genotyping was performed by using the BovineHD 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 those at repeated positions were excluded. Additional requirements were Mendelian error rate lower than 0.05, and genotyping rates over 95% for both, individuals and SNPs. The quality control was made using PLINK software (Purcell et al. (2007)). Finally, 706,978 SNPs were retained, covering 2,510,395 kb, that represent an average density of one marker every 3.551 kb. The phases of the parental chromosomes were established by means of the Beagle software (Browning and Browning (2009)).

Molecular coancestry. Firstly, we calculated the molecular coancestry for every pair of individuals (within and between populations) for each SNP. Following Caballero and Toro (2002), the molecular coancestry was obtained applying Malécot's (1948) definition, by using the difference in identity by state instead of the identity by descent. Afterwards, the molecular coancestry for every pair of individuals (*i* and *j*) at the m^{th} region of 1 Mb of the genome (see Table 1) was calculated averaging the molecular coancestries as:

$$f_{ijm} = \frac{1}{N_m}\sum_{k=1}^{N_m} f_{ijk}$$

where f_{ijk} is the molecular coancestry for the k^{th} marker in the m^{th} region and N_m is the number of markers within the region.

Genetic distances. Given the molecular coancestry between all pairs of breeds, we calculated the following genetic distances between populations:

a) Nei's minimum distance (Nei (1987)):

$$D_{NABm} = \left(\frac{\bar{f}_{Am} + \bar{f}_{Bm}}{2}\right) - \bar{f}_{ABm}$$

where D_{NABm} is the distance between breeds A and B for the m^{th} region of the genome, $\bar{f}_{Am}$ and $\bar{f}_{Bm}$ are the average molecular coancestries between individuals of the A and B populations, respectively, and $\bar{f}_{ABm}$ is the average molecular coancestry between all individuals of the A and B populations.

b) Reynolds, Weir and Cockerham's distance (Reynolds et al. (1983)):

$$D_{RABm} = \frac{D_{NABm}}{(1 - \bar{f}_{ABm})}$$

Table 1. Number of regions (1Mb) for each chromosome with mean and standard deviation for number of SNPs by region.

Chr	Nr	Mean (SD)	Chr	Nr	Mean (SD)	Chr	Nr	Mean (SD)
1	159	281.21 (50.78)	11	108	286.57 (57.89)	21	72	282.94 (68.85)
2	137	282.55 (49.46)	12	92	270.81 (83.25)	22	62	279.58 (55.98)
3	122	281.59 (54.93)	13	85	269.36 (62.11)	23	53	272.51 (64.88)
4	121	277.31 (52.24)	14	85	282.4 (60.42)	24	63	285.97 (54.00)
5	122	274.78 (54.04)	15	86	273.93 (62.41)	25	43	290.49 (47.71)
6	120	285.11 (53.31)	16	82	283.29 (65.16)	26	52	290.28 (52.91)
7	113	281.35 (64.64)	17	76	279.43 (66.70)	27	46	273.96 (62.07)
8	114	285 (64.28)	18	66	282.11 (71.86)	28	47	264.55 (79.28)
9	106	281.39 (52.20)	19	65	281.32 (68.12)	29	52	269.35 (70.71)
10	105	278.81 (71.59)	20	72	287.90 (59.47)			

Chr = Chromosome, Nr = Number of regions by chromosome.

where D_{RABm} is the distance between populations A and B for the m^{th} region of the genome.

c) Nei's standard distance (Nei (1987)):

$$D_{SABm} = -log\left(\frac{\bar{f}_{ABm}}{\sqrt{\bar{f}_{Am}\,\bar{f}_{Bm}}}\right)$$

where D_{SABm} is the distance between populations A and B for the m^{th} region of the genome.

Total genetic diversity. From the information of the genetic distances, a summary of total genetic diversity for the m^{th} region of the genome is calculated by the F_{ST} statistics:

$$F_{STm} = \frac{V_B}{V_T}$$

We calculated the total variance (V_T) as:

$$V_T = V_B + V_W$$

where, V_B is the variance between breeds and is equal to

$$V_B = \sum_{i=1}^{Np} \frac{1}{N_p{}^2} \sum_{j=1}^{N_p} D_{Nijm}$$

with Np the number of populations and V_W, the within breed variance calculated as:

$$V_W = 1 - \left(\frac{1}{N_p}\sum_{i=1}^{N_p} \bar{f}_i\right)$$

being $\bar{f}_i$ is the average molecular coancestry within the i^{th} breed.

Identification of genome regions. Genome regions with the highest genetic differentiation (F_{ST} values) were cross-referenced with the ENSEMBL database (Flicek et al. (2013)) to locate potential differentiation target genes.

Results and Discussion

Genome-wide analysis of genetic diversity. A plot of the F_{ST} values for 1 Mb genome regions along autosomal chromosomes is presented in Figure 1. The average F_{ST} value for all regions is 0.049 (0.015), very close to the estimates obtained by Cañas-Álvarez et al. (2014) between pairs of breeds. However, the F_{ST} varies enormously across genomic regions, and some of them showed remarkable differences with respect to the average estimate. Among them, we selected 6 regions of the chromosomes 2, 5, 6, 11, 13 and 18 (see Figure 1).

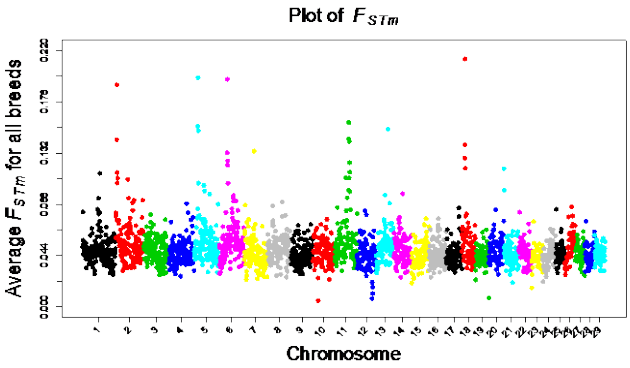

Figure 1. Plot of F_{STm} by regions (1Mb) along the autosomal chromosomes.

Between breed differences. The estimates of pairwise Nei's minimum, Reynolds and Standard Nei distances were coherent between them. Table 2 shows the more remarkable distances between pairs of breeds that are within the six selected regions.

Table 2. Higher estimates of genetic distance within selected genome regions.

Chromosome	Breeds	D_N	D_R	D_S
BTA2 (6-7 Mb)	AV-Pi	0.13	0.44	0.17
	AV-Re	0.13	0.40	0.18
	AV-BP	0.10	0.34	0.13
	AV-ANI	0.09	0.32	0.12
BTA5 (17-20 Mb)	BRU-Re	0.14	0.40	0.21
	ANI-BP	0.14	0.44	0.18
	AV-BP	0.10	0.33	0.13
	Pi-Re	0.10	0.27	0.16
	ANI-Pi	0.09	0.23	0.14
BTA6 (38-39 Mb)	BP-Re	0.18	0.46	0.24
	ANI-BP	0.17	0.45	0.23
	Pi-Re	0.10	0.27	0.15
	ANI-Pi	0.10	0.26	0.14
BTA11 (67-68 Mb)	BP-Re	0.13	0.42	0.17
	Pi-Re	0.08	0.31	0.10
	AV-BP	0.14	0.36	0.19
	BP-Pi	0.10	0.31	0.13
BTA13 (57-58 Mb)	BP-Re	0.13	0.42	0.17
BTA18 (14-15 Mb)	BP-Pi	0.09	0.35	0.12

D_N = Nei's minimum distance, D_R = Distance of Reynolds, D_S = Nei's standard distance.

The results of the 7th genome region of the chromosome 2 (in positions greater than 6Mb and smaller than 7Mb) showed important differences between AV and the rest of the populations (Pi, Re, BP and ANI). This is probably due to the presence of the Myostatin gene in that region, because it is segregating in the AV population (Dunner et al. (2003)).

The genome region between positions 17-20Mb in the chromosome 5 showed important differences between BP-Re, BP-AV, BP-ANI, Pi-Re and Pi-ANI. In addition,

the region of the chromosome 6 between positions 38-39 Mb also showed remarkable differences between BP and Pi with respect to Re and ANI. It should be noted that BP and Pi are located in the Pyrenean regions with a high altitude environment.

Further, the region of the chromosome 11 (67-68Mb) exhibited strong divergence between population Pi and Re, AV and BP and BP and PI. Finally, genomic regions of the chromosomes 13 (57-58Mb) and 18 (14-15Mb) show differentiation between BP-Re and BP-Pi, respectively.

Gene identification. Table 3 presents the list of genes located at the selected regions in the ENSEMBL database (Flicek et al. (2013)). A total of 77 genes were identified. However, further research is necessary to link the functional annotation of genes with the selection and differentiation process of these Spanish autochthonous beef cattle breeds.

Table 3. Genes located in selected genome regions.

Chromosome	Genes		
BTA2 (6-7 Mb)	*C2H2orf88*	*ORMDL1*	*ASNSD1*
	MSTN	*OSGEPL1*	*SLC40A1*
	PMS1	*ANKAR*	*WDR75*
BTA5 (17-20 Mb)	*C12orf50*	*DUSP6*	*KITLG*
	C5H12orf29	*WDR51B*	*ATP2B1*
	TMTC3	*CEP290*	*5S_rRNA*
BTA6 (38-39 Mb)	*PKD2*	*LAP3*	*HCAP-G*
	SPP1	*MED28*	*LCORL*
	MEPE	*FAM184B*	
	IBSP	*DCAF16*	
BTA11 (67-68 Mb)	*ARHGAP25*	*FOV*	*NFU1*
	BMP10	*ANTXR1*	*AAK1*
	GKN2	*GFPT1*	
BTA13 (57-58 Mb)	*CDH26*	*PHACTR3*	*ATP5E*
	FAM217B	*EDN3*	*TUBB1*
	SYCP2	*ZNF831*	*CTSZ*
	PPP1R3D	*SLMO2*	*NELFCD*
BTA18 (14-15 Mb)	*PIEZO1*	*ANKRD11*	*FANCA*
	APRT	*SPG7*	*SPIRE2*
	CDT1	*RPL13*	*TCF25*
	GALNS	*CPNE7*	*MC1-R*
	TRAPPC2L	*DPEP1*	*TUBB3*
	PABPN1L	*CDK10*	*DEF8*
	CBFA2T3	*SPATA2L*	*DBNDD1*
	ACSF3	*CHMP1A*	*GAS8*
	CDH15	*VPS9D1*	*SHCBP1*
	MGC157263	*ZNF276*	

Conclusion

We have found heterogeneity of specific genetic distance and F_{ST} estimates among autochthonous Spanish beef cattle breeds within 1 Mb regions of the autosomal chromosomes. In fact, six genomic regions were highlighted as clearly different between populations. These regions harbor up to 77 genes in the ENSEMBL database.

Aknowledgments

The financial support of Spanish AGL2010-15903 and UE FP7-289592-GENE2FARM grants made possible this research. T.B.R da Silva thanks Capes-Brazil. A. Gónzalez-Rodríguez and J. J. Cañas-Álvarez acknowledge the financial support given by the BES-2011-045434 and COLCIENCIAS Francisco José de Caldas 497/2009 fellowships respectively.

Literature Cited

Beja-Pereira, A., Alexandrino, P., Bessa, I. et al. (2003). *J. Hered*., 94:243-250.

Browning, B. L., and Browning, S. R. (2009). *Am. J. Hum. Genet.*, 84:210-223.

Caballero, A. and Toro M. A. (2002). *Conserv. Genet*., 3:289-299.

Cañas-Álvarez, J. J., González-Rodríguez, A., Varona, L. et al. (2014). 10th WCGALP. (Submitted).

Dunner, S., Miranda, M. E., Amigues, Y. et al. (2003). *Genet. Sel. Evol*., 35:103-118.

Flicek, P., Ahmed, I., Amode, M. R. et al. (2013). *Nucleic Acid. Res*. 41(D1): D48-D55.

Malécot, G. (1948). Paris, France*:* Masson et Cie.

Nei, M. (1987). New York, USA: Columbia Univ. Press.

Purcell, S., Neale, B., Todd-Brown, K. et al. (2007). *Am. J. Hum. Genet.*, 81:559-575.

Reynolds, J., Weir B. S. and Cokerham, C. C. (1983). *Genetics*, 105:767-779

Scherf, B. D. (2000). Rome, Italy: FAO, United Nations.