Global Journal of Animal Breeding and Genetics

ISSN: 2408-5502 Vol. 3 (4), pp. 159-163, July, 2015.

http://www.globalscienceresearchjournals.org/

Full Length Research Paper

The extent of linkage disequilibrium in the ancient Ukrainian grey cattle breed

Mekki Boussaha[1,2], Rabia Letaief[1,2], Nikolay Yudin[3,4,5], Mikhail Voevoda[3,4,5] and Dominique Rocha[1,2*]

[1]INRA, UMR1313, Unité Génétique Animale et Biologie Intégrative, Domaine de Vilvert, F-78352 Jouy-en-Josas, France
[2]AgroParisTech, UMR1313, Unité Génétique Animale et Biologie Intégrative, Domaine de Vilvert, F-78352 Jouy-en-Josas, France
[3]Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Lavrentyeva Avenue 10, Novosibirsk 630090, Russian Federation
[4]Institute of Internal and Preventive Medicine, Siberian Branch of the Russian Academy of Medical Sciences, Borisa Bogatkova Street 175/1, Novosibirsk 630089, Russian Federation
[5]Novosibirsk State University, Pirogova Street 2, Novosibirsk 630090, Russian Federation

Accepted 21 July, 2015

Abstract

We present here the first genome-wide characterization of linkage disequilibrium (LD) in the Ukrainian Grey (UKR) cattle breed, a very ancient beef breed. The aim of our study was to assess the pattern and extent of LD in Ukraine. We used a sample of 48 animals and 37,576 single nucleotide polymorphisms (SNPs) covering all cattle autosomes. The squared correlation of the alleles at two loci (r^2) was used as a measure of LD. The analysis of adjacent marker pairs revealed that the level of LD decreases rapidly with physical distance between SNPs. Overall mean r^2 was 0.218 (±0.261). Strong LD (r^2> 0.8) and useful LD (measured as r^2 > 0.2) were observed within genomic regions of up to 534 kb and 2 Mb, respectively. Unexpectedly, our results indicate that LD extent found in UKR is comparable to LD measures in modern breeds, particularly in beef breeds, suggesting a bottleneck during conservation herd and/or breed formation. In addition, LD mapping of phenotypic traits in the UKR population should be feasible with an optimized medium-density SNP panel or using high-density genome-wide genotyping tools.

Key words: Beef breed, cattle, linkage disequilibrium, single nucleotide polymorphism, conservation

INTRODUCTION

In recent years, thanks to new genomics tools, several genome-wide surveys have been done in livestock populations to analyse the extent of LD, the non-random association of different genetic loci, in order to determine the optimal number of markers required for genome-wide scans of complex traits (Terwilliger and Weiss, 1998) or for genomic selection (Meuwissen *et al*., 2001). Several studies of the extent of LD have been reported in cattle, mostly in highly-selected breeds (McKay *et al*., 2007; The Bovine HapMap Consortium, 2009; Beghain *et al*., 2013) and there is little knowledge about the degree of genome-wide LD in more primitive breeds (Flury *et al*., 2010; Edea *et al*., 2014; Mastrangelo *et al*., 2014; Meszaros *et al*., 2015).

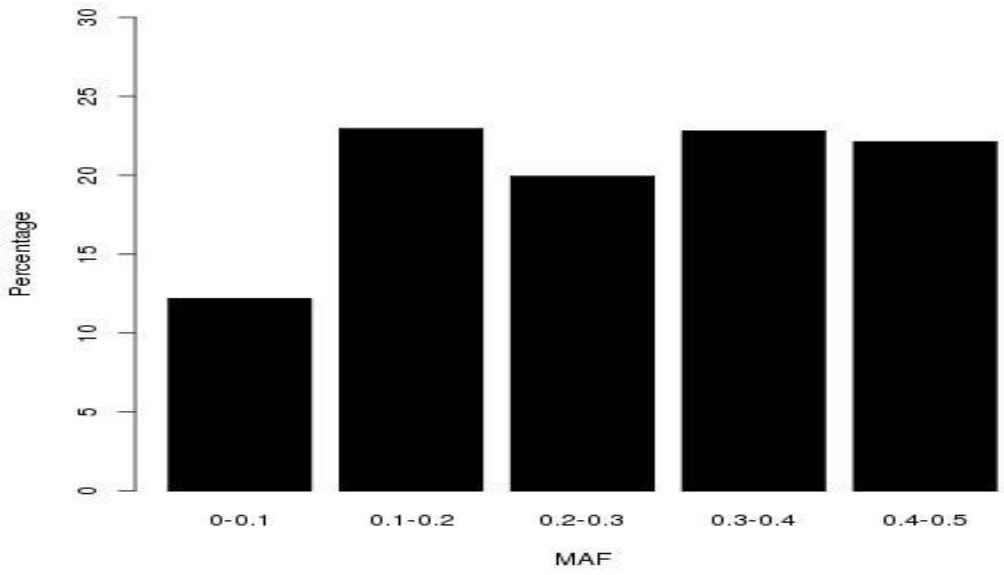
The objective of our study was to determine the extent of LD in the Ukrainian Grey cattle breed, a very ancient beef breed descendent from Grey Steppe cattle from southern and central Europe. This hardy primitive breed,

*Corresponding author. E-**mail**: dominique.rocha@jouy.inra.fr, **Tel**: 33 1 34 65 24 22, **Fax**: 33 1 34 65 24 78

Table 1. Number of single nucleotide polymorphisms (SNPs), length of the chromosome covered, average, median and maximum spacing between adjacent single nucleotide polymorphisms SNPs on different autosomes. Distances are shown in kb..

Chromosome	# SNPs	Length	Mean +/- SD distance	Median	Maximum
1	2,437	158,094.12	64.90 +/- 55.25	45.11	562.99
2	2,016	136,662.11	67.82 +/- 65.92	45.10	643.01
3	1,854	121,144.18	65.38+/- 62.32	44.38	806.72
4	1,817	120,451.33	66.33 +/- 55.00	46.21	507.93
5	1,559	121,043.69	77.69 +/- 79.63	49.75	993.31
6	1,883	119,013.79	63.24 +/- 68.16	43.12	1,601.81
7	1,635	112,357.17	68.76 +/- 99.58	45.08	3,063.80
8	1,744	112,908.18	64.78 +/- 53.66	46.33	557.39
9	1,451	105,464.42	72.73 +/- 70.54	48.41	666.91
10	1,519	103,180.36	67.97 +/- 103.15	45.47	3,393.06
11	1,615	107,096.49	66.35 +/- 58.55	45.39	687.57
12	1,236	90,944.48	73.64 +/- 150.57	47.58	4,331.58
13	1,314	83,835.60	63.85 +/- 56.00	45.60	715.70
14	1,310	83,152.51	63.52 +/- 58.70	44.56	1,077.95
15	1,228	84,444.61	47.95 +/- 38.96	37.26	662.97
16	1,208	81,249.45	67.32 +/- 70.82	45.50	1,316.97
17	1,180	74,887.03	63.52 +/- 66.05	44.78	835.30
18	982	65,509.83	66.78 +/- 69.22	45.10	1,031.96
19	1,032	63,541.21	61.63 +/- 59.79	44.43	1,083.91
20	1,168	71,413.86	61.19 +/- 51.32	43.32	610.93
21	1,028	71,097.98	69.23 +/- 73.94	46.18	1,322.35
22	920	61,121.40	66.51 +/- 56.94	47.05	562.80
23	795	52,096.19	65.61 +/- 68.47	44.88	1,183.63
24	957	62,101.92	64.96 +/- 52.10	48.95	454.80
25	735	42,803.94	58.32 +/- 47.16	41.94	392.57
26	785	50,953.36	64.99 +/- 51.43	47.88	437.18
27	679	45,332.32	66.86 +/- 64.79	47.14	914.35
28	706	46,182.97	65.51 +/- 52.03	46.75	345.19
29	783	50,872.78	65.05 +/- 64.16	48.36	1,060.19
All	37,576	2,498,957.25	66.55 +/- 70.91	45.73	4,331.58

Figure 1: Distribution of minor allele frequency (MAF) for the 37,576 selected SNPs.



closely related to Italy's Podolica cattle or the Hungarian Grey cattle, has been used for meat and draft power. Despite its adaptation to a harsh local environment, the UKR population has been replaced by more productive modern breeds and consequently the total number of animals has fallen to approximately 1,200 heads kept in Ukraine and the Russian Federation (FAO DAD-IS database, 2011). The knowledge of LD is paramount to localizing genes affecting quantitative traits and for implementing genetic selection programmes in UKR.

MATERIAL AND METHODS

Animal Sampling and Genotyping

Forty-eight animals (8 bulls and 40 cows) were selected from a conservation herd established near Novosibirsk. These animals were randomly selected as no pedigree records were available. Whole blood samples were obtained and genomic DNA isolated by the standard method using proteolytic treatment with proteinase K followed by phenol extraction. DNA samples were then genotyped on the BovineSNP50 Genotyping Beadchips (Illumina, San Diego, CA, USA) at LABOGENA (Jouy-en-Josas, France) using the standard operating procedures recommended by the manufacturer.

Estimation of linkage disequilibrium

Linkage disequilibrium between two loci was measured by the squared correlation (r^2) between their genotypes (non-random association of alleles). The Haploview v 4.1 program (Barrett *et al*., 2005) was used for estimation of allele frequencies and tests for Hardy–Weinberg equilibrium (HWE) of genotype frequencies and to compute the r^2 measure between each pair of syntenic markers.
Graphical representations of the patterns of LD were generated using the statistical *R* package (www.r-project.org).

RESULTS AND DISCUSSION

To determine the extent of LD in UKR, 48 animals were genotyped using the Illumina BovineSNP50 Genotyping Beadchips. Among the 54,001 SNPs included in the chip, only polymorphic SNPs mapping to bovine autosomes on the bovine genome assembly UMD3.1 version (Zimin *et al*., 2009), genotyped on >=85% of the individuals, with a minor allele frequency (MAF) larger than or equal to 0.05 and in Hardy–Weinberg Equilibrium (*P*-value < 0.001) were retained for further analysis. A total of 37,576 informative SNPs (~70%) were left after all the filtering step for the study. The total autosomal genome length covered was 2,499 Mb, with the shortest chromosome being 43 Mb and the longest chromosome being 158 Mb (Table 1). The density of SNP loci per chromosome was homogeneous with an average density of 15 SNPs per megabase. The average distance between adjacent SNP pairs for the entire autosomal genome was 66.55 ± 70.91 kb with a median spacing of 45.73 kb. More than eighty-three per cent of the adjacent SNP pairs (31,192 pairs) were <0.1 Mb apart and 13 pairs were more than 1 Mb apart, indicating a relatively dense SNP map. The largest gaps were on BTA12 (4.33 Mb long) and on BTA10 (3.40 Mb long).

The overall mean minor allele frequency (MAF) was 0.273 ± 0.132. The average MAF of the SNPs used in our study is in agreement with those found previously for several breeds (Matukumalli *et al*., 2009; The Bovine HapMap Consortium, 2009; Beghain *et al*., 2013). More than 65% of SNPs had a MAF larger than 0.2, suggesting that the effect of low MAF on the overall LD estimated should be small (Figure 1).

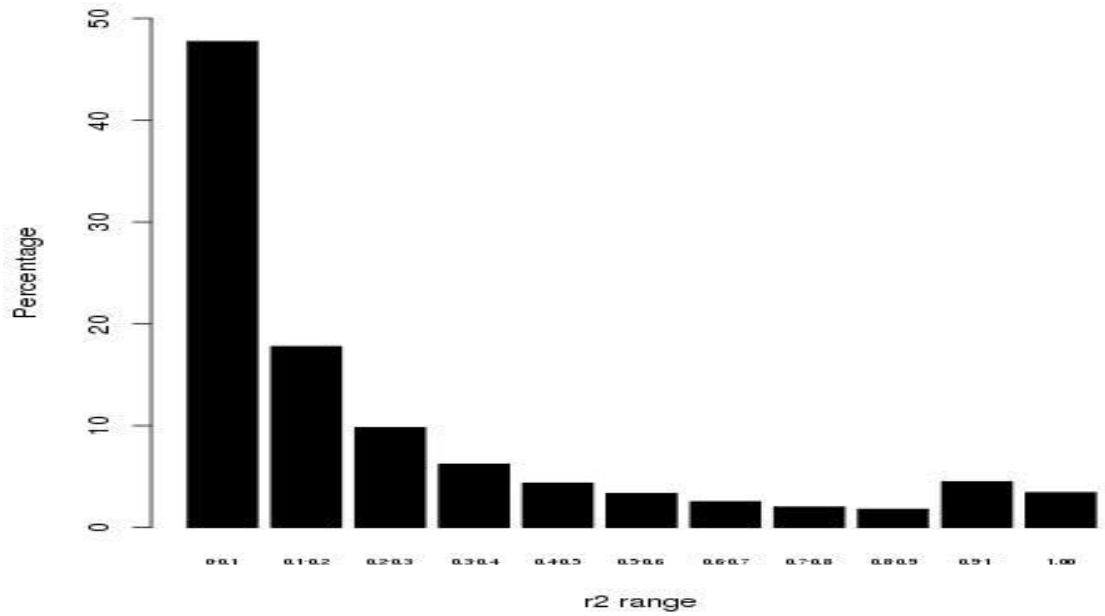
The extent of linkage disequilibrium was evaluated for adjacent marker pairs. r^2 values were estimated for each adjacent SNP pair on each autosome: a total of 36,763 pairs were analysed for all 29 autosomes. The average distances between adjacent informative SNPs and the distribution of MAF of these SNPs were uniform over the genome. The mean value of r^2 pooled over all the autosomes and for individual autosomes is presented in Table 2. The mean r^2 (± SD) between adjacent SNPs ranged from 0.144 (± 0.241) for BTA19 to 0.244 (± 0.282) for BTA10, whereas overall r^2 was 0.218 (± 0.261). The distribution of r^2 between adjacent SNP pairs is shown in Figure 2. To examine LD decay with physical distance, SNP pairs were sorted into intervals based on their inter-marker distance and mean r^2 (± SD) were calculated for each interval; we also calculated the percentage of marker pairs with strong (r^2 > 0.8) and useful LD (r^2 > 0.2) (Table 1). We chose r^2 > 0.2 as a threshold for useful LD according to Meuwissen *et al*. (2001), who predicted genomic breeding values from dense markers across the whole genome with accuracies up to 0.85 when using adjacent markers with an average r^2 of 0.20.

In the population studied, the level of pair-wise LD as measured by r^2 decreased with physical distance between SNPs. High LD values were observed for SNP located in close proximity. A total of 1,283 marker pairs (~3.4%) were in complete LD (r^2 = 1) and complete LD persisted over inter-marker distances of up to 364 kb. (Figure 2); mean and maximum genomic distances between these marker pairs were 38.53 and 364.22 kb, respectively. The mean distance between marker pairs was 56.03 kb at a threshold of r^2 > 0.2 and decreased to 41.70 kb at r^2 > 0.8 (result not shown). The inverse relationship between LD extent and marker distance shown in UKR is in agreement with previous LD studies

Table 2. Mean linkage disequilibrium (r^2) among adjacent single nucleotide polymorphisms(SNPs) for each autosome and pooled over all autosomes and percentage of marker pairs with r^2> 0.2 or 0.8.

Chromosome	#SNP pairs	Mean r^2 +/- SD	% r^2 > 0.2	% r^2 > 0.8
1	2,436	0.239 +/- 0.274	38.42	7.27
2	2,015	0.223 +/- 0.265	35.58	6.50
3	1,851	0.238 +/- 0.274	38.63	7.94
4	1,816	0.203 +/- 0.247	32.98	5.23
5	1,558	0.201 +/- 0.255	31.26	5.91
6	1,882	0.237 +/- 0.278	37.04	7.81
7	1,634	0.235 +/- 0.276	36.96	7.65
8	1,743	0.216 +/- 0.265	33.28	6.25
9	1,450	0.229 +/- 0.274	35.45	7.17
10	1,518	0.244 +/- 0.282	38.41	8.17
11	1,614	0.225 +/- 0.265	36.62	6.51
12	1,235	0.224 +/- 0.263	35.71	6.48
13	1,313	0.200 +/- 0.245	31.07	4.80
14	1,309	0.224 +/- 0.270	34.45	7.10
15	1,227	0.200 +/- 0.242	32.19	4.73
16	1,207	0.215 +/- 0.259	33.97	5.97
17	1,179	0.204 +/- 0.249	32.82	5.51
18	981	0.211 +/- 0.255	33.74	5.81
19	1,031	0.144 +/- 0.241	31.52	4.85
20	1,167	0.204 +/- 0.250	32.65	5.91
21	1,027	0.235 +/- 0.271	37.68	6.62
22	919	0.213 +/- 0.262	31.88	6.96
23	794	0.177 +/- 0.225	28.09	3.27
24	956	0.211 +/- 0.257	32.53	5.96
25	734	0.202 +/- 0.238	32.97	3.81
26	784	0.197 +/- 0.253	30.36	5.87
27	678	0.203 +/- 0.241	34.51	4.72
28	705	0.196 +/- 0.235	30.78	4.54
29	782	0.208 +/- 0.254	32.99	5.75
All	36,763	0.218 +/- 0.261	34.51	6.29

Figure 2. Distribution of r^2between adjacent single nucleotide polymorphism pairs.



in cattle (McKay *et al*. 2007). The extent of LD found in UKR is comparable to LD measured in modern breeds, particularly in beef breeds (McKay *et al*., 2007). The Bovine HapMap Consortium, (Villa-Angulo 2009 *et al*. 2009; Lu *et al*., 2012; Beghain *et al*., 2013). For example, McKay *et al.* (2007) assessed LD using r^2 among all pairs of syntenic markers within eight breeds of cattle from *Bos taurus* and *Bos indicus* subspecies. This study included four beef breeds (Angus, Charolais, Japanese Black and Limousin). Pair-wise r^2 values were estimated using about 2,670 SNPs spanning the entire bovine autosomal genome and they found that the extent of LD was no more than 0.5 Mb in these eight cattle breeds. More recently, we analysed the LD pattern in the French Blonde d'Aquitaine population and found a mean r^2 value (0.205 ± 0.262) between adjacent SNP pairs comparable to the one we estimated in UKR (Beghain *et al*., 2013). The results of these two studies are comparable to the trend we saw in the UKR sampled population.

Unexpectedly, the extent of LD found in UKR is comparable to LD measured in modern breeds. This suggests a small effective population size and/or a strong bottleneck during the development of the current conservation herd and/or at the beginning of the formation of the breed. Interestingly, we saw that useful LD persists up to ~56 kb, on average. The Illumina BovineSNP50 Genotyping Beadchips with a density of, on average, one informative SNP per ~60 kb in UKR would allow capturing approximately 72% of useful LD between SNPs located on the BovineSNP50 Genotyping Beadchip and quantitative trait nucleotides. The current tool should therefore enable association studies as well as selective genomic selection in this beef breed; however, the use of an optimized medium-density SNP panel or of higher-density SNP genotyping tools is recommended to achieve optimum genome-wide association studies and genomic selection in UKR.

CONCLUSIONS

Our results expand on previous studies of genome-wide LD in cattle breeds, particularly for indigenous breeds. We have analysed the extent of linkage disequilibrium in UKR cattle. Our results show that LD decays rapidly with increasing physical distance between SNPs and that the observed pattern of LD is similar to the patterns found previously in highly-selected breeds. The use of an optimized medium-density SNP panel or of higher-density SNP genotyping tools is recommended to achieve association studies and genomic selection. The information generated from this study has important implications for the design and application of conservation and selective breeding programmes for UKR.

ACKNOWLEDGEMENTS

The authors would like to thank Lubov Vasileva for providing DNA samples for the animals analysed in this study, Yves Amigues and colleagues at LABOGENA for their help with the SNP genotyping and Didier Boichard for his support. This work was funded by INRA and RFBR (grant number 13-04-00968).

REFERENCES

Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.

Beghain J, Boitard S, Weiss B, Boussaha M, Gut I, Rocha D (2013). Genome-wide linkage disequilibrium in the Blonde d'Aquitaine cattle breed. J. Anim. Breed. Genet. 130: 294-302.

Edea Z, Dadi H, Kim SW, Park JH, Shin GH, Dessie T, Kim KS (2014). Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in Ethiopia. J. Anim. Breed. Genet. 131: 358-366.

Flury C, Tapio M, Sonstegard T, Drögemüller C, Leeb T, Siniamer H, Hanotte O, Rieder S (2010). Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. J. Anim. Breed. Genet. 127: 339-347.

Lu D, Sargolzaei M, Kelly M, Li C, Van der Voort G, Wang Z, Plastow G, Moore S, Miller SP (2012). Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. Front. Genet. 3: 152.

Mastrangelo S, Saura M, Tolone M, Salces-Ortiz J, Di Gerlando R, Bertolini F, Fontanesi L, Sardina MT, Serrano M, Portolano B (2014). The genome-wide structure of two economically important indigenous Sicilian cattle breeds. J. Anim. Sci. 92: 4833-4842.

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP (2009). Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4: e5350.

McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C, Mannen H, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS (2007). Whole genome linkage disequilibrium maps in cattle. BMC Genet. 8: 74.

Meszaros G, Boison SA, Perez O'Brien AM, Ferencakovic M, Curik I, Da Silva MV, Utsunomiya YT, Garcia JF, Solkner J (2015). Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. Front. Genet. 6: 173.

Meuwissen TH, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Terwilliger JD, Weiss KM (1998). Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr. Opin. Biotechnol. 9: 578–594.

The Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324: 528–532.

Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ (2009). High-resolution haplotype block structure in the cattle genome. BMC Genet. 10: 19.

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marçais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10:R42.