

ESTIMATION OF LINKAGE DISEQUILIBRIUM IN THE NERO SICILIANO ITALIAN AUTOCHTHONOUS BREED USING THE ILLUMINA 60K SNP ARRAY

Stefania CHESSA ¹, Andrea CRISCIONE ², Riccardo MORETTI ¹, Salvatore BORDONARO ³, Donata MARLETTA ⁴, Bianca CASTIGLIONI ¹

ABSTRACT

Local breeds represent an important component of the overall farm animal diversity to be maintained and exploited. The new high-throughput molecular technologies allow a wide range of massive, simultaneous genomic analysis. Commercial SNP genotyping platforms are a suitable tool for the genetic characterization and the study of inter-breeds diversity. Linkage disequilibrium, the nonrandom association of alleles at different loci, has received increasing attention in recent years as a result of the availability of genome sequences and large numbers of identified SNP. This study aims to assess the genomic structure of the Nero Siciliano pig, an Italian population reared in eastern Sicily, through the analysis of the extent and range of linkage disequilibrium using the SNP analyzed through the PorcineSNP60 Genotyping BeadChip. Moreover molecular data from other four Italian breeds/populations were also included in the linkage analysis. Linkage disequilibrium may reveal much about breed history, genetic relationships and represent an extremely valuable tool in planning the marker density required to be efficient in marker assisted selection. The Nero Siciliano breed showed the lowest value of average linkage disequilibrium probably due to the lack of systematic selection strategies and variable linkage disequilibrium rates were found in different genomic regions among the analyzed populations.

Key words: Nero Siciliano / linkage disequilibrium /genetic resources

1 INTRODUCTION

Local breeds are a fundamental resource both from the genetic diversity and for the crucial topic related to the maintenance of marginal areas. They represent a repository of allelic combinations, rare or absent in the selected breeds, also they can be successfully associated to typical products helping farmers manage and protect the environment. Nowadays the availability of new massively parallel sequencing technologies allows the multiple analysis of wide genome regions at affordable cost (Glenn 2011), but still requiring a quite big economical effort if searching for polymorphisms due to the high coverage required for variant calling. Anyway, genetic characterization is a fundamental prerequisite for managing genetic resources and can be exploited for setting up molecular authentication protocols. Commercial SNP

genotyping platforms, highly used all over the world as the Illumina BeadChip recently available for pigs (Ramos *et al.*, 2009), seems an highly suitable alternative for the genetic characterization and can be easily used to compare different breeds.

Aim of the present paper was to analyze the genomic structure of the Nero Siciliano pig describing the population-wise level of linkage disequilibrium (LD) using high density genotypes. The knowledge of the extent and range of LD, the non-random association of alleles at two or more loci, is extremely valuable in localizing genes affecting quantitative traits (Pritchard and Donnelly, 2001), identifying chromosomal regions under selection, studying population history, and characterizing genetic resources and diversity (Nordborg and Trave, 2002; Tenesa *et al.*, 2007). Nero Siciliano is an autochthonous black pig breed reared in the eastern part of Sic-

¹ Institute of Agricultural Biology and Biotechnology – National Research Council (IBBA-CNR), via Einstein, 26900 Lodi, Italy, e-mail: chessa@ibba.cnr.it

² Department DISPA, University of Catania, via Val di Savoia 5, 95123 Catania, Italy, e-mail: a.criscione@unict.it

³ Same address as 2, e-mail: s.bordonaro@unict.it

⁴ Same address as 2, e-mail: d.marletta@unict.it

ily (Italy). The breed runs the risk of losing its original traits in the absence of a suitable plan to safeguard and exploit its production.

2 MATERIAL AND METHODS

2.1 BREED DESCRIPTION AND SAMPLE COLLECTION

Nero Siciliano is an ancient black pig breed reared under semi-extensive or extensive system in the mountainous area of Nebrodi, in the North-East of Sicily. It is characterized by rusticity, adaptation to harsh conditions and limited food supply, in addition to being resistant to diseases and it is addressed to the production of high quality meat, including salami and cured ham. The majority of the registered population is kept by small traditional farms. It has registered an important increase in the number of farms and in sow in the last ten years thanks also to the creation of a Protected Designation of Origin label for meat and other related products.

A representative sample of 93 Nero Siciliano (SN) pigs was collected from 22 farms. The sampling has concerned the Nebrodi area which is part of a natural Park

located 100–1700 m above sea level (37°50'–38°9' N; 14°26'–14°54' E).

2.2 DNA EXTRACTION AND GENOTYPING

DNA was extracted from blood using a commercial kit (GE Healthcare, Little Chalfont, UK), checked for quality and submitted to Neogen® Corporation's GeneSeek for the molecular characterization through the PorcineSNP60 Genotyping BeadChip v2 (Illumina, San Diego, CA, USA), containing 61,565 SNP. Genotyping data from other 168 samples were already available from a previous work (Chessa *et al.*, 2011): 96 from a Northern Italy local population called Nero di Garlasco (SG), 24 Italian Large White (LW), 24 Italian Landrace (LA) and 24 Italian Duroc (DU) genotyped by the first version of the Illumina BeadChip, containing 61,263 SNP.

2.3 DATA ANALYSIS

SNP data from the two versions of the Illumina BeadChip were compared and common SNP retained for the following analysis. All SNP were tested for in-

Table 1: Distribution of LD by breed and distance

SNP pair distances	SNP pairs	breed	$r^2 \pm SD$	% of SNP with $r^2 > 0.3$
0.0–50 kb	35,414	SN	0.20 ± 0.29	0.2370
	33,250	LA	0.31 ± 0.33	0.3491
	35,653	LW	0.29 ± 0.32	0.3315
	25,669	DU	0.38 ± 0.37	0.4206
	22,179	SG	0.43 ± 0.38	0.4825
50–150 kb	77,355	SN	0.15 ± 0.23	0.1632
	72,429	LA	0.26 ± 0.30	0.2997
	77,757	LW	0.25 ± 0.29	0.2818
	55,483	DU	0.32 ± 0.34	0.3679
	47,981	SG	0.39 ± 0.36	0.4506
150–300 kb	112,192	SN	0.11 ± 0.19	0.1152
	105,428	LA	0.22 ± 0.26	0.2487
	113,164	LW	0.21 ± 0.25	0.2343
	79,754	DU	0.28 ± 0.31	0.3146
	68,833	SG	0.35 ± 0.34	0.4116
300–1000 kb	520,606	SN	0.08 ± 0.15	0.0710
	486,679	LA	0.18 ± 0.22	0.1890
	524,384	LW	0.16 ± 0.21	0.1662
	351,836	DU	0.21 ± 0.26	0.2361
	313,145	SG	0.30 ± 0.32	0.3642

consistencies in Mendelian segregation and Hardy-Weinberg Equilibrium (HWE). SNP with minor allele frequency (MAF) < 0.05 were excluded due to their possible strong influence on LD (Du *et al.*, 2007). SNP not in HWE ($P < 0.01$), with a call rate (CR) < 0.25 and on sex chromosome were also excluded. Different R packages (<http://www.r-project.org>) were used for data filtering. To evaluate LD we focalized on r^2 measure since it is very useful for biallelic markers, independent from sample size (Devlin and Risch, 1995) and less dependent on allelic frequency than D' , a measure of LD designed for loci with two or more alleles (Du *et al.*, 2007). For each SNP, pairwise LD was calculated for adjacent SNP less than 1 Mb apart. The extent of LD between each marker pair within each breed separately was computed using Haploview 4.1.

3 RESULTS AND DISCUSSION

3.1 GENOTYPING RESULTS AND CHIP VERSION COMPARISON

In the Nero Siciliano breed, 90 subjects had a CR higher than 99%, while just one individual had CR lower than 70% and was excluded from the following analysis. The SNP CR was higher than 95%. In the 94.59% of the

SNP (58.232 on 61.565 total). SNPs that passed the CR threshold were investigated about their frequency and a total of 6.285 SNPs were found to be monomorphic and 13.034 had a MAF lower than 0.05.

Comparing the first (61,263 SNP) and the v2 (61,577 SNP) of the PorcineSNP60 Genotyping BeadChip, 61,177 SNP were in common and were used for the subsequent analysis. In v2 about the 0.03 of the 61,177 SNP was re-located on a different chromosome (0.80 previously assigned to chromosome 0). Within each chromosome (excluding Y) a proportion of SNP ranging from 0.17 in chromosome 17 to 0.22 in chromosome 8 was given a new map position referred to the last swine genome map build. Thus the analysis of LD using v2 SNP map should be more precise than with the previous version.

3.2 ESTIMATION OF LINKAGE DISEQUILIBRIUM

Average r^2 was computed for SNP whose distance spanned from 0 to 1 Mb. Here we summarize the results showing the data of four classes of distances (Table 1). The average r^2 was largest in the Nero di Garlasco animals (from 0.31 to 0.43), as expected because of the high inbreeding. Excluding this population, the highest values were found for Duroc (0.38–0.21), as already described in literature, whereas Nero Siciliano exhibited the small-

Table 2: Distribution of SNP pairs and rate of LD (average r^2) by chromosome and breed

Chr	SN	LA	LW	DU	SG
1	0.12 ± 0.20	0.22 ± 0.26	0.22 ± 0.27	0.27 ± 0.31	0.33 ± 0.33
2	0.10 ± 0.19	0.19 ± 0.24	0.18 ± 0.23	0.24 ± 0.29	0.36 ± 0.35
3	0.07 ± 0.14	0.15 ± 0.20	0.15 ± 0.20	0.20 ± 0.25	0.27 ± 0.32
4	0.09 ± 0.17	0.25 ± 0.28	0.20 ± 0.26	0.20 ± 0.25	0.34 ± 0.33
5	0.08 ± 0.15	0.16 ± 0.21	0.16 ± 0.22	0.23 ± 0.27	0.28 ± 0.30
6	0.07 ± 0.15	0.16 ± 0.21	0.17 ± 0.21	0.23 ± 0.29	0.27 ± 0.32
7	0.13 ± 0.20	0.25 ± 0.27	0.19 ± 0.23	0.29 ± 0.31	0.37 ± 0.33
8	0.09 ± 0.17	0.18 ± 0.22	0.16 ± 0.20	0.25 ± 0.30	0.33 ± 0.35
9	0.09 ± 0.17	0.17 ± 0.22	0.17 ± 0.22	0.21 ± 0.27	0.31 ± 0.33
10	0.07 ± 0.13	0.15 ± 0.19	0.13 ± 0.18	0.21 ± 0.27	0.23 ± 0.28
11	0.08 ± 0.15	0.16 ± 0.21	0.15 ± 0.20	0.19 ± 0.25	0.34 ± 0.33
12	0.08 ± 0.14	0.15 ± 0.19	0.15 ± 0.19	0.20 ± 0.26	0.23 ± 0.27
13	0.10 ± 0.18	0.21 ± 0.25	0.20 ± 0.25	0.24 ± 0.30	0.34 ± 0.33
14	0.15 ± 0.22	0.28 ± 0.29	0.25 ± 0.28	0.32 ± 0.33	0.39 ± 0.36
15	0.08 ± 0.16	0.18 ± 0.23	0.17 ± 0.23	0.23 ± 0.28	0.32 ± 0.34
16	0.09 ± 0.16	0.20 ± 0.23	0.17 ± 0.21	0.21 ± 0.26	0.29 ± 0.32
17	0.09 ± 0.15	0.18 ± 0.22	0.17 ± 0.22	0.21 ± 0.25	0.33 ± 0.32
18	0.06 ± 0.12	0.17 ± 0.21	0.15 ± 0.20	0.19 ± 0.24	0.27 ± 0.29

est average r^2 (0.20–0.08), a result that can be related with the low level of inbreeding in the selected sample together with the absence of SNP typical of this breed in the BeadChip. The SNP with a $r^2 > 0.3$, considered as appropriate threshold of “usable” LD in experimental designs for continuous traits in pigs (Du *et al.*, 2007), were quite frequent in the Duroc (0.27–0.42), whereas in the Nero Siciliano they were much less frequent, especially for long-range distance SNP pairs (0.07).

Table 2 shows the average LD for each chromosome for distances ranging from 0 to 50 kb between SNP. Some variations were observed in the extent of LD in different chromosomes among populations. In the Nero Siciliano and in Duroc chromosomes 1, 7 and 14 had the greatest average LD, whereas for the other breeds the highest values were observed at the chromosomes 4, 7 and 14 for Landrace, 1, 4, 13 and 14 for Large White and 2, 7 and 14 for Nero di Garlasco, a result that should be further investigated.

4 CONCLUSIONS

With the increasing availability of SNPs, the whole genome association studies become increasingly realistic and attractive. The knowledge of the extent and range of LD in animal populations is essential to understand the marker density required to be efficient in marker assisted selection and test the population suitability. The possibility to use the Illumina BeadChip to select marker in linkage with traits of interest for selection purpose in Nero Siciliano is a little lower than in cosmopolitan breeds (0.23 of small-range distance marker was found with an r^2 higher than 0.3). However, the BeadChip can also be really useful to understand the relationship among individuals and so limit the inbreeding within selection programs. Besides, a subset of random SNP can easily be

used to distinguish the different breeds for genetic traceability purposes (Chessa *et al.*, 2013).

5 REFERENCES

- Chessa S., Bordonaro S., Moretti R., Criscione A., Marletta D., Castiglioni B. 2013. Genomic analysis for the valorization of “Nero Siciliano”. *Italian Journal of Animal Science*, 12 (s1): 68
- Chessa S., Stella A., Raschetti M., Passero A., Crepaldi P., Nicolo L., Castiglioni B., Pagnacco G. 2011. Genomic analysis for the valorization of an Italian local swine breed. *Italian Journal of Animal Science*, 10 (s1): 131
- Devlin B., Risch N. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, 29: 311–322
- Du F., Clutter A., Lohuis M. 2007. Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences*, 3: 166–178
- Glenn T.C., 2011. Field guide to next generation DNA sequencers. *Molecular Ecology Resources*, 11: 759–769
- Nordborg M., Tavaré S. 2002. Linkage disequilibrium: what history has to tell us. *Trends in Genetics* 2002, 18: 83–90
- Pritchard J.K., Donnelly P. 2001. Case-control studies of association in structured or admixed populations. *Theoretical Population Biology*, 60: 227–237
- Ramos A.M., Crooijmans R.P.M.A., Amaral A.J., Archibald A.L., Beever J.E., Bendixen C., Dehais P., Hansen M., Hedegaard J., Hu Z., Kerstens H.D., Law A., Megens H-J, Milan D., Nonneman D., Rohrer G., Rothschild M., Smith T., Schnabel R.D., Van Tassell C., Clark, R. Churcher C., Taylor J., Wiedmann R., Schook L.B., Groenen M.A.M. 2009. Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. *PLoS ONE*, 4(8): e6524
- Tenesa A., Navarro P., Hayes B.J., Duffy D.L., Clarke G.M., Goddard M.E., Visscher P.M. 2007. Recent human effective population size estimated from linkage disequilibrium. *Genome Research*, 17: 520–526