SNPs discovery in RRLs from DNA pools of Nero Siciliano pigs with extreme and divergent phenotypes for the Back Fat Thickness (BFT) tract

ANNA MARIA SUTERA[1], ALESSANDRO ZUMBO[1], IRENE SAPIENZA[1], GIUSEPPE TARDIOLO[1], ENRICO D'ALESSANDRO[1]

[1]Department of Veterinary Sciences, Division of Animal Production, University of Messina, Messina, Italy. Via Palatucci, snc, 98168 - Messina (Italy)

SUMMARY

Fat deposition is a key biological process that has implications in pig economic management. This process affects carcass quality and aptitude for the typical product productions such as salami, sausages and of cured ham. This study aimed to detect single nucleotide polymorphisms (SNPs) that could be associated with the backfat thickness (BFT) tract in Nero Siciliano pigs. The Food and Agriculture Organization (FAO) has expressed concern about the lack of interest in local breeds compared to high-output animals and conservation programs have been implemented by various countries worldwide. Genomic DNA from two groups of Nero Siciliano pigs with divergent phenotypes for BFT was pooled and digested with BsuRI (HaeIII) restriction enzyme for preparation of reduced representation libraries (RRLs). The two RRLs produced 4124595 (BFT+) and 4052107 (BFT-) sequenced reads, were mapped on *Sus scrofa* reference genome (Sscrofa 11.1 assembly). SNP calling was performed using SNAPE, a software that implements a Bayesian approach for SNP calling in pooled samples. 47,791 putative SNPs were called by SNAPE, of these 32,235 (67.4%) were polymorphic while 15,556 (32.5%) were monomorphic. Of all SNPs detected in this study, 22 showed enriched alleles in one or the other RRLs. These SNPs, some of these localised in genes involved in fat metabolism, might be potential markers associated with BFT in Nero Siciliano pig.

In this study, we identified SNPs potentially associated with BFT that might be utilized for applications in breeding programs. Attitude to high-fat deposition (in particular in neck, withers and back) for the Nero Siciliano pig is known and our results could contribute to explain the biology of fat metabolism in this breed.

KEY WORDS

Nero Siciliano pig; Back Fat Thickness; SNPs; RRLs; DNA pooling; DNA sequencing.

INTRODUCTION

In the swine production, the deposition of fat is a biological process of fundamental importance for the know implications on the efficiency of animal's growth and on the technological and nutritional characteristics of meat products with significant implications also on business management[1]. Although in the last year's admirable efforts have been made to recover the extremely threatened biodiversity of pigs, today only a few Italian local breeds are able to withstand the competition with foreign commercial breeds (i.e. the Large White, Landrace and Duroc) and/or with other commercial crossbreeds today widespread in the market which combine high reproductive efficiency, good growth performances, good quality of carcasses and excellent quality of meat, are the genetic types on which the production of heavy pigs is based in Italy[2]. Nero Siciliano pig, also known as 'Nero dei Nebrodi', is a local pig breed reared on the island of Sicily mainly under extensive management[3]. Nero Siciliano dates back to ancient times. However, the genetic pool of the breed seems to have been formed mainly during the last few centuries[4]. The breed is known for its aptitude in the accumulation of fat in the region of the neck, back and withers[5]. The remarkable adipogenic capacity of the unimproved breeds is reported in the literature[6,7]. The production aspects and the characteristics of the carcass are complex are partly genetically determined. Although the heritability of these traits is low, increased knowledge of the genome represents an important tool for introducing significant innovations in the management of farms, to better understand their genetic structure[8]. SNPs represent the primary functional basis of genetic variability, which is reflected in the phenotypic differences between and within the breeds. In case of the identified polymorphisms represent exclusive selective imprints of the breed taken in consideration, these markers can be applied efficiently to identify the most favourable genetic variants for the improvement of quantitative traits, or in the identification and traceability of products.[9]

Taking advantage from the sequenced genome of the pig and its reference assembly (Sscrofa 11.1) it is now possible to use next generation sequencing (NGS) platforms to further investigate the level and extent of genetic variability in different breeds and populations[10].The Ion Torrent technology is a promising NGS platform that is based on a semiconductor detection of pH variation during the sequencing process that can be applied in different experimental approaches in which a medium-high throughput is needed[11].

Corresponding Author:
Enrico D'Alessandro (edalessandro@unime.it)

The aim of this study was to detect SNPs that could be associated with the BFT in the Nero Siciliano pig, considering the great importance of fat deposition for the production of hams, and in the known technological effects in the seasoning phase[12].
To perform our investigation the reduced-representation libraries (RRLs) method was used. For the construction of the libraries, the DNA of various pigs was grouped into two Reduced Representation Libraries were obtained by enzymatically digest DNA pools constructed from pigs with an extreme and divergent average of the BFT.

MATERIAL AND METHODS

All the procedures used in this research were in compliances with the European guidelines for the care and use of animals in research (Directive 2010/63/EU 2010).
The study was carried out on 192 Nero Siciliano pigs coming from different swine farms, divided into two homogeneous groups for number (96 pigs), sex (half male and half female), age (10 months) and live weight (95 kg± 5 kg). All pigs were registered in National Herd book of the National Pig Breeder Association (ANAS). The animals were slaughtered on reaching the average live weight of 92 ± 5 kg. From the measurements were extrapolated the averages of the BFT and used for the generation of the extreme and divergent groups, each of 50 pigs (50 with the most negative BFT and 50 with the most positive BFT). Average and standard deviation of BFT of the pigs in the negative and positive tails were 25,70 ± 2,97 mm and 34,40 ± 4,65 mm respectively.
Genomic DNA (gDNA) was extracted from blood sample of two groups of Nero Siciliano pigs with divergent phenotype for average of BFT using the Wizard® Genomic DNA Purification Kit (Promega Corporation, Italy) and quantified by Qubit 2.0 fluorometer with Qubit dsDNA HS Assay Kit (Thermo Fisher, Italy). Genomic DNA digested with BsuRI (HaeIII) restriction enzyme for preparation of reduced representation libraries (RRLs). For each library, 200 ng of DNA gel-purified was enzymatically fragmented by Ion Shear™ Plus 10X Reaction Buffer and Ion Shear™ Plus Enzyme Mix II (Thermo Fisher Scientific). Then, the following steps occurred: the adapters ligation, the nick-repair to make the covalent ligation on both strands (Ion Xpress™ Plus Fragment Library Kit, Thermo Fisher Scientific), and a further purification step of the ligated DNA. Obtained DNA material was size-selected using E Gel® SizeSelect™ Agarose Gel (Thermo Fisher Scientific) and bands corresponding to 200bp of inserts were collected and quantified by qPCR using a StepOnePlus Real-Time PCR System (Life Technologies). Libraries were clonally amplified in emPCR and enriched on Ion OneTouch™ 2 System (Thermo Fisher Scientific) using Ion PGM™ Hi- Q™ OT2 kit (Thermo Fisher Scientific). Each library was loaded on an Ion 318™ Chip V2 BC, previous evaluation of the templating efficiency with the Ion Sphere™ Quality Control kit and the Qubit® 2.0 Fluorometer, and then sequenced on the Ion PGM™ System with Ion PGM™ Hi-Q™ Sequencing Kit. The reads from the two RRLs were checked with FAstaQC program, processed using the Ion Torrent Suite v3.6 software (Life Technologies) and cleaned with Trimmomatic v. 0.36, then mapped on the reference genome (Sscrofa 11.1; GCA_000003025.6) with TMAP 3.6 ALIGNER (https://github.com/iontorrent/TS/tree/master/Analysis/TMAP) using BWA-short[13] , BWA-long[14] and SSAHA[15]. SNP calling was obtained using SNAPE[16]. SNAPE input files (PILEUP format) were obtained using Samtools v.0.1.4[17,18]. To map gene positions and to predict the effect of each substitution was used Variant effect predictor (VEP) tool (http://www.ensembl.org/Sus scrofa/Tools/VEP)[19] and to evaluate if missense mutations could have deleterious effects on the translated proteins was used SIFT[20]. Fisher's exact test was computed for each alternative genomic position covered by a minimum depth of 3x to evaluate differences in allele frequency derived by the number of alternative reads between the two RRLs. All the positions with $P_{Fisher} < 0.05$ were also visually inspected with IGV (IntegrativeGenomicsViewer) software.

RESULTS AND DISCUSSION

The two RRLs produced 4124595 (BFT+) and 4052107 (BFT-) sequenced reads, which after cleaning (filtering and trimming) were mapped on *Sus scrofa* reference genome (Sscrofa 11.1 assembly). Only reads with mapping quality score> 20 were retained for subsequent analysis. SNP calling was performed using SNAPE, a software that implements a Bayesian approach for SNP calling in pooled samples. 47,791 putative SNPs were called by SNAPE, of these 32,235 (67.4%) were polymorphic while 15,556 (32.5%) were monomorphic. The analysis performed with VEP showed that most of the identified variants (about 74%) were already present in dbSNPs (release 151), whereas about 26% were novels. The SNPs was identified in the transcribed regions. Among these located in the coding regions of the genome, 96 were synonymous mutations, 277 missense, 10 stop-gained (e.g. in ADH6, PGM2, TMA16, MRPL1 genes), 9 stop lost (e.g. in WDR25, HECTD1) and 1 stop retained (in ENSSSCG00000036380). Among the aforementioned missense mutations, 38 were classified as deleterious by SIFT. These mutations have been noted in genes that play a role in processes correlated to lipid metabolism (such as MIGA1 and CUBN) and in various cellular mechanisms such as in the activation of the innate immune response (TLR10), in the neuronal migration (SLIT2), and in the apoptotic signal (DTHD1) (data not showed). SNAPE was also used to estimate, from the count of alternative reads, the allelic frequency of each polymorphic genomic position, among the 452288 in common between the two RRLs. Fischer's exact test was performed to evaluate differences in allele frequency estimates as determined by alternative read count between the two libraries. Of all SNPs detected in this study, 22 showed enriched alleles in one or the other RRLs for $P_{Fisher} < 0.05$ (Table. 1). The SNPs identified were located in autosomal chromosomes in which candidate genes (SSC1, SSC6, SSC12, SSC14, SSC18) and Quantitative Trait Loci (QTLs) (SCC12, SCC14) for BFT have been mapped (Table. 2). Among the 22 identified variants about 31% were intergenic variants, about 31% were located in the intronic regions, 4% were downstream variants and 3% upstream. The downstream and upstream variants were located in annotated six genes with known functions such as Melanocortin-4 receptor (MC4R), Fatty acid synthase (FASN), Phosphodiesterase 6C (FFAR4), Stearoyl-CoA desaturase (SCD), whereas the intronic variants in genes Leptin (LEP) and Leptin Receptor (LEPR). In particular, MC4R gene is a member of the superfamily of G-protein-coupled receptors (GPCRs), which affects body weight, energy homeostasis and food intake in humans and mice and the

Table 1 - SNPs showed $P_{Fisher} < 0.05$ between two RRLs (Positive Library: BFT+, Negative Library: BFT-)

				BFT+		BFT-		
Chr	Genomic postion	Reference	Alt	N_Ref	N_Alt	N_Ref	N_Alt	*PFischer*
1	4,553	A	GA	0	6	8	2	2,86E-02
1	1,082	G	TG	0	7	7	1	4,82E-02
1	284,340	T	TA	29	4	17	7	4,05E-02
1	124,584,196	C	CG	17	2	6	5	1,28E-02
1	160,773,173	A	AG	8	1	0	3	3,37E-02
2	123,750,235	G	GA	13	0	1	4	1,70E-02
2	123,751,016	T	TC	1	8	13	3	4,65E-02
6	956,438	G	GA	9	0	4	6	3,35E-02
6	145,607,675	C	CA	11	1	3	7	3,55E-02
6	145,631,685	G	AG	13	1	4	9	1,27E-02
6	145,654,695	C	CT	15	2	2	4	3,63E-02
6	146,893,537	A	GA	8	0	4	7	1,87E-02
6	67,619,980	G	TG	27	13	17	32	1,03E-02
7	75,888,372	T	TA	11	0	3	7	1,25E-02
12	993,860	T	TG	32	3	8	14	2,63E-02
13	31,684,656	G	GA	29	0	11	2	3,69E-02
14	105,011,727	C	CT	8	13	10	1	4,37E-02
14	111,463,097	C	TC	1	15	7	2	1,29E-02
15	155,198,719	C	CG	0	6	7	1	4,96E-02
15	155,253,410	C	CA	0	7	8	0	2,95E-02
16	36,782,695	T	GT	8	43	12	21	3,27E-02
18	20,123,273	C	CT	5	49	57	13	1,29E-02

Chr = Chromosome number; Genomic position on the Scrofa 11.1; Reference = Reference nucleotide on Sccrofa 11.1 reference genome; N_Ref = number of reads supported reference nucleotide; Alt = alternative variant; N_Alt = Number of alternative reads of reference nucleotide

Asp298Asn polymorphism of the MC4R gene is associated with backfat thickness in different commercial cross-bred pigs[21]. FASN gene codes for a protein that has an essential role in the synthesis of long-chain fatty acids, starting from acetyl CoA and using malonyl-CoA as a carbon 2 donor[22]. This gene has been mapped to chromosome 12[23], in a region that includes numerous QTLs for traits related to the adiposity of the carcass (http://aaa.animalgenome.org/cgi- bin/gbrowse/pig/?name = Chr.12). LEP (Leptin) gene codes for a 146 amino acid protein, mainly expressed in adipose tissue[24]. This hormone acts on the hypothalamic receptor by modulating the sense of satiety and therefore regulating body weight and energy balance. Several mutations at this locus have been associated with traits related to average daily weight gain, intramuscular fat and back fat thickness[25,26] in the Casertana, Duroc, Large White and Meishan breeds, SCD gene is a candidate gene for fatty acid composition. It is located on pig SSC14 in a region where quantitative trait loci (QTL) for fatty acid composition were previously detected in a Duroc purebred population[27]. LEPR gene plays an important role in the regulation of fat deposition and other commercially important traits in pigs and this regulation is known to be breed-specific[28]. Balatsky et al., (2018) suggest that LEPR SNP c.2856C > T can be considered as a genetic marker for subcutaneous fat deposition and average daily weight gain in Ukrainian Large White pigs.

CONCLUSIONS

Fatness-related traits, in particular the backfat thickness, are very important in pig production since they influence meat quality and technological processes adopted for long matured prod-

Table 2 - Significant SNP for back fat thickness allocated in candidate gene.

CHR	PosSNP	GENE	Position Gene	Acc. Num.
1	160,773,173	MC4R	160,772,013-160,774,124	397359
6	146,893,537	LEPR	146,802,297-146,896,152	396836
12	993,860	FASN	992,405-997,560	397561
14	105,011,727	FFAR4	105,010,974-105,122,628	100135678
14	111,463,097	SCD	111,461,570-111,478,033	396670
18	20,123,273	LEP	20,106,867-20,124,071	396832

Chr = chromosome number; PosSNP = nucleotide position on the Sscrofa 11.1 reference genome of the SNP having $P_{Fisher} < 0.05$; *Pos SNP located within QTL regions

ucts such as cured ham. For valorisation of Nero Siciliano's productions, a request to label the fresh meat with the Protected Denomination of Origin (PDO) was issued in 2005 while for Nero Siciliano's cured ham the request for the PDO has been started in 2011. In this study, we identified SNPs potentially associated with BFT that might be utilized for applications in breeding programs. Attitude to high-fat deposition (in particular in neck, withers and back) for the Nero Siciliano pig is known and our results could contribute to explain the biology of fat metabolism in this breed.

References

1. D'Alessandro E., Sapienza I., Giosa D., Giuffrè L., Zumbo A. (2019). In silico analysis of meat quality candidate genes among Nero Siciliano, and Italian heavy pigs genomes. Large Animal Review, 25(4):137-140.
2. (ANAS) Associazione Nazionale Allevatori Suini (2016.) Notizie: La selezione per il prosciutto DOP. Una sfida tecnica per un orizzonte dilungo periodo. Retrieved on 15 January 2018 from http://www.anas.it/circolari/20170001A.PDF Google Scholar.
3. Guastella A.M., Criscione A., Marletta D., Zuccaro A., Chies L., Bordonaro S. (2010). Molecular characterization and genetic structure of the Nero Siciliano pig breed. Genet Mol Biol, 33:650-656
4. D'Alessandro E., Sottile G., Sardina M.T., Criscione A., Bordonaro S., Sutera A.M., Zumbo A., Portolano B., Mastrangelo S. (2020). Genome wide analyses reveal the regions involved in the phenotypic diversity in Sicilian pigs. Anim Genet, 51(1):101-105. doi: 10.1111/age.12887
5. Chiofalo L. (2007). Nero Siciliano pig. Proceedings of 6th International Symposium on the Mediterranean Pig. Capo d'Orlando (ME), Italy.
6. Maiorano G. (2009). Swine production in Italy and research perspectives for the local breeds. Slovak J Anim Sci, 42 (4): 159-166.
7. Chiofalo B., Lo Presti V., Savoini G., D'Alessandro E., Chiofalo V., Liotta L. (2011). Nucleotides in Brolier chicken diet: Effect on breast muscles quality. Czech J. Food Sci, 29(4):308-317.
8. D'Alessandro E., Giosa D., Sapienza I., Giuffrè L., Aiese C. R., Romeo O., Zumbo A. (2019). Whole genome SNPs discovery in Nero Siciliano pig. Genet Mol Biol, 42(3):594-602. doi.org/10.1590/1678-4685-GMB-2018-0169
9. Fontanesi L., D'Alessandro E., Scotti E., Liotta L., Crovetti A., Chiofalo V., Russo V. (2010). Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. Anim Genet, 41:478-492.
10. Bosse M., Megens H.J., Frantz L.A., Madsen O., Larson G., Paudel Y., Duijvesteijn N., Harlizius B., Hagemeijer Y., Crooijmans R.P., Groenen M.A. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nature Commun, 5:4392-2014.
11. Rothberg J.M., Hinz W., Rearick T.M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature, 475:348-352.
12. Bosi P., Russo V. (2004). The production of the heavy pig for high quality processed products Ital J Anim Sci, 3:309-321.
13. Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.
14. Li H., Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics, 26: 589-595.
15. Ning Z., Cox A.J., Mullikin J.C. (2001). SSAHA: a fast search method for large DNA databases. Genome Research 11, 1725-9.
16. Raineri E., Ferretti L., Esteve-Codina A., Nevado B., Heath S., Pérez-Enciso M. (2012). SNP calling by sequencing pooled samples. BMC Bioinformatics, 13:239
17. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G.,Durbin.(2009).The sequence alignment/Map format and SAM tools. Bioinformatics, 27:863-864.
18. Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21):2987-93.
19. McLaren W., Pritchard B., Rios D., Chen Y., Flicek P., Cunningham F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics, 26:2069-2070.
20. Ng P.C., Henikoff. S. (2011). Predicting deleterious amino acid substitutions. Genome Res, 1:863-874.
21. Lyadskiy, I. K., Getya, A. A., Pochernyaev, K. F. (2011). Association of the Asp298Asn polymorphism in the mc4r gene with back fat thickness in pigs of the large white breed. Cytology and Genetics, 45(2):106-109.
22. Menendez J.A., Vazquez-Martin A., Ortega F.J., Fernandez-Real J.M. (2009). Fatty acid synthase: Association with insulin resistance, type 2 diabetes, and cancer. Clin Chem, 55:425-438.
23. Muñoz G., Alcazar E., Fernandez A., Barragan C., Carrasco A., de Pedro E., Silio L., Sanchez J.L., Rodriguez M.C. (2011). Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc X Iberian cross are pigs, Meat Science, 88:169-173.
24. D'Andrea M., Pilla F., Giuffra E., Waddington D., & Archibald A. L. (2008). Structural analysis and haplotype diversity in swine LEP and MC4R genes. J Anim Breed Genet, 125:130-136.
25. Chao Z., Wang F., Deng C.Y., Wei L.M., Sun R.P., Liu H.L., Liu Q.W., Zheng X.L. (2012). Distribution and linkage disequilibrium analysis of polymorphisms of MC4R, LEP, H-FABP genes in the different populations of pigs, associated with economic traits in DIV2 line, Mol Biol Rep, 19:111-120.
26. Hirose K., Ito T., Fukawa K., Arakawa A., Mikawa S., Hayashi Y., Tanaka K., (2014): Evaluation of effects of multiple candidate genes (LEP, LEPR, MC4R, PIK3C3, and VRTN) on production traits in Duroc pigs. Anim Sci J, 85:198-206.
27. Uemoto, Y., Nakano, H., Kikuchi, T., Sato, S., Ishida, M., Shibata, T., Suzuki, K. (2012). Fine mapping of porcine SSC14 QTL and SCD gene effects on fatty acid composition and melting point of fat in a Duroc purebred population. Anim Genet, 43(2):225-228.
28. Balatsky, V., Oliinychenko, Y., Sarantseva, N., Getya, A., Saienko, A., Vovk, V., & Doran, O. (2018). Association of single nucleotide polymorphisms in leptin (LEP) and leptin receptor (LEPR) genes with backfat thickness and daily weight gain in Ukrainian Large White pigs. Livestock science, 217:157-161.