Archiv Tierzucht 56 (2013) 19, 200-207

Short communication

Looking for breed differentiating SNP loci and for a SNP set for parentage testing in Mangalica

Attila Zsolnai[1], Gábor Tóth[2], János Molnár[3], Viktor Stéger[2], Ferenc Marincs[2], Anna Jánosi[4], Gabriella Ujhelyi[4], Erika Koppányné Szabó[4], Anita Mohr[3], István Anton[1], Réka Szántó-Egész[3], Rita Sipos[3], István Egerszegi[1], Klára Dallmann[3], Péter Tóth[5], Adrienn Micsinai[3], Klaus-Peter Brüssow[6] and József Rátky[1]

[1]Research Institute for Animal Breeding and Nutrition (ÁTK), Herceghalom, Hungary, [2]Agricultural Biotechnology Center, Gödöllő, Hungary, [3]BIOMI Kft, Gödöllő, Hungary, [4]Central Food Research Institute, Budapest, Hungary, [5]Olmos és Tóth Kft, Debrecen, Hungary, [6]Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany

Abstract

The whole genome of Mangalica animals has been screened on the Illumina porcine chip giving the possibility (1) to replace the previously applied ten microsatellite markers by nine SNP loci to classify the Blond, Swallow-Belly and Red Mangalica individuals into three different breed groups ($P>0.95$); (2) to propose 54 SNP loci for parentage testing in Mangalica pigs where the exclusion probability is 0.999115 if one parent is known and the probability of identity is 1.54×10^{-23}.

Keywords: genome wide screening, breed differentiation, parentage testing, SNP

Abbreviations: BM: Blond Mangalica, F_{ST}: fixation index, RM: Red Mangalica, SBM: Swallow-Belly Mangalica, SNP: single-nucleotide polymorphism

Introduction

Since the commercial availability of 40K-60K or even 500K SNP chips, genome wide screening or genome wide association studies have been applied in domestic animals to identify SNPs to establish SNPs as genetic markers or to identify genes by causative SNPs which are responsible for a given trait (Andersson 2009). To identify loci associated with mono- or multigenic traits, the number of selected animals for genome wide association studies varied from 10 to 50 in dogs (Andersson 2009), horses (Orr *et al.* 2010) and cattle (Huang *et al.* 2010)

Archiv Tierzucht 56 (2013) 19, 200-207
doi: 10.7482/0003-9438-56-019

Received: 20 March 2012
Accepted: 17 September 2012
Online: 1 March 2013

Corresponding author:
Klaus-Peter Brüssow; email: bruessow@fbn-dummerstorf.de
Leibniz Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

or up to 311-820 individuals in cattle and pig (Kim *et al.* 2011, Fan *et al.* 2011). Searching for SNPs sensitive to the population structure in sheep (Kijas *et al.* 2009) or pig (Matsumoto *et al.* 2012) has also been successful, and the average number of animals here in a breed was 14 and 17, respectively.

SNP chips used in genome wide association studies are also useful to select for those SNPs which are applicable in parentage testing (Matukumalli *et al.* 2009). In pigs, based on Duroc, Landrace, Hampshire and Yorkshire breeds, 60 SNPs were proved to be more powerful than ten microsatellite loci (Rohrer *et al.* 2007).

The present study was aimed to build the basis to replace our previously applied microsatellite set used for both differentiation of the three Mangalica breed variants (Zsolnai *et al.* 2006) and for assignment of individuals in paternity or forensic tests (unpublished). This effort is described herein starting with the comparison of genetic distances obtained by different markers both in type and numbers, continuing with genome wide screening for handful SNP loci capable to fulfil breed differentiation and screening for an SNP set to perform parentage or identity tests.

As previously described (Egerszegi *et al.* 2003), Mangalica was bred by crossing of already extinct Hungarian and Mediterranean pig breeds in the 19th century (Figure 1). This breed was kept in very large numbers between the late 1800s and mid 1900s. However, twice in its history, it nearly disappeared. Scientific and breeding efforts supported its new development (Brüssow *et al.* 2005). Nowadays, Mangalica has three existing varieties, Blond (BM), Red (RM) and Swallow-Belly (SBM), and the Mangalica population is considerably growing in Hungary. This is due to the high quality of meat and meat products which is favoured in different cuisines and for niche food products.

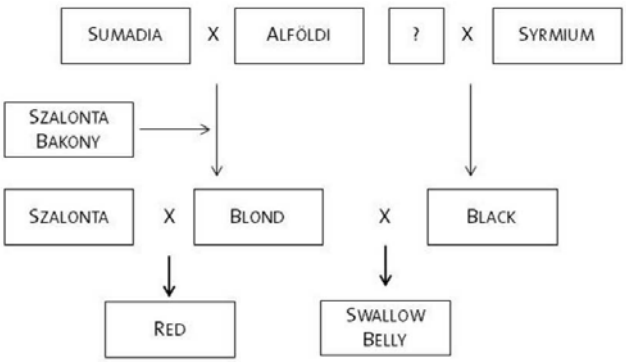

Figure 1
Breeding scheme according to herd books

Materials and methods

Samples

Blood samples were obtained from the Hungarian Pig Tissue Biobank collected by the research consortium MANGFOOD at different farms of 80 Mangalica, including 24 BM, 33 SBM and 23 RM, and of 63 non-Mangalica (White pigs) including 10 Pietrain, 12 Large White, 3 H39 Hybrid, 12 Landrace, 12 Hampshire and 14 Duroc pigs. Samples were stored at –20 °C

until DNA preparation. DNA was isolated from the samples using the Genomic DNA Maxi Kit (Geneaid, New Taipei City, Taiwan) according to the manufacturer's Frozen Blood Kit Protocol.

Genotyping

Microsatellite genotyping was performed using the S0005, S0090, S0101, S0155, S0355, S0386, SW24, SW240, SW857 and SW951 markers (Nechtelberger *et al.* 2001). The Mangalica samples were checked by microsatellites within themselves and against our core population data from the year 2006 before they were genotyped on 60K pig SNP (Illumina, San Diego, CA, USA). SNP genotyping was carried out by Aros Applied Biotechnology AS (Aarhus, Denmark).

Analysis

For calculation of pairwise population F_{ST}, those SNPs were selected from the entire SNP dataset for breed-pairs, of which call rate was 1 and the minor allele frequency was greater than 0.05. The number of the selected SNPs was 15297, 18208, 21057, 32273 and 33872 for the BM–SBM, BM–RM, SBM–RM, BM–Duroc and BM–Large White pairs, respectively. To calculate the F_{ST} values between breeds, a preinstalled Eigensoft package (Patterson *et al.* 2006) was used on a BioSmack linux platform (Hong *et al.* 2012).

To find SNP loci which are able to differentiate Mangalica breeds, we have searched for differences in the allele effect in pairwise comparison, where Blond, Swallow-Belly and Red Mangalica were set up against all other breeds: BM vs. (RM+SBM+White), RM vs. (BM+SBM+White) and SBM vs. (RM+BM+White). In the experiment we have used SVS7 (Golden Helix, Bozeman, MT, USA) for principal component analysis aided quality check and genotype association tests; Genalex (Peakall & Smouse 2006) for building Structure format datasets and for calculating exclusion and identity probability values; and Structure (Falush *et al.* 2003) for population assignment, where 10 000 burn-in and 50 000 MCMC steps were applied. In SVS program the effect of alleles were tested using genotypic model and Chi-square test. Missing genotypes were not used as predictors.

For parentage testing, SNP alleles with call rate equal to 100 % and minor allele frequency higher than 0.40 were selected. The chromosomal assignment and position of the 202 loci meeting these criteria were obtained from the marker list of the PorcineSNP60 chip (http://www.illumina.com/products/porcinesnp60_dna_analysis_kit.ilmn). The calculated physical distance of adjacent SNPs positioned on the same chromosome was based on that list. Data filtering was performed by SVS7 software.

Results

Before any SNP genotyping of Mangalica samples, they were checked by microsatellite loci. Population assignments against our core population data (Zsolnai *et al.* 2006) had proved that the animals have matched into their expected category ($P>0.9$, data not shown). F_{ST} values were calculated pairwise for the three Mangalica breed variants using SNPs selected for the described pairs, and were compared with the previously obtained microsatellite F_{ST} values (Zsolnai *et al.* 2006). The microsatellite and SNP based F_{ST} values displayed a strong correlation ($r^2=0.788$). F_{ST} values are presented in Table 1.

Table 1
Pairwise F_{ST} values between Mangalica breed variants

	SBM	BM
BM	$0.064^{m}/0.062^{SNP}$	–
RM	$0.099^{m}/0.091^{SNP}$	$0.095^{m}/0.075^{SNP}$

^{m}values based on ten microsatellite markers, ^{SNP}values based on SNP markers with minor allele frequency greater than 0.05, in the BM, SBM and RM groups, respectively.

The SNP-based genetic distances between BM and other breeds were 0.24 to Duroc and 0.18 to Large White which is similar to the microsatellite-based F_{ST} values reported by others (0.27 to Duroc and 0.21 to Large White; Garcia *et al.* 2006).

The next experimental setups have yielded 24 SNPs with high $-\log_{10}P$ values (Table 2.)

Table 2
24 breed discriminating SNP loci. A given Mangalica breed variant was compared to the other two Mangalica plus the white groups involved in the study to identify potential breed-discriminating SNP loci.

Mangalica	Marker	Chromosome	Position	$-\log_{10}P$
	DRGA0009603	1	174 282 114	14.70
	ALGA0007555	na	na	14.63
	ALGA0118931	na	na	13.77
BM	ALGA0108666	na	na	13.00
	ASGA0068481	17	18 490 103	13.00
	H3GA0056684	2	12 338 278	12.96
	ALGA0054195	9	52 225 904	12.85
	ASGA0042399	9	50 169 218	12.80
	MARC0055761	na	na	18.95
	ALGA0103326	na	na	17.91
	MARC0058565	17	30 360 527	17.20
SBM	MARC0017379	4	37 772 563	17.18
	ASGA0019473	na	na	16.73
	ASGA0098145	7	61 582 321	16.26
	ASGA0037579	17	31 062 240	16.21
	INRA0026120	17	40 157 380	15.89
	ALGA0061440	11	19 031 846	24.27
	MARC0029343	13	19 560 975	19.53
	ASGA0097149	na	na	18.95
RM	ALGA0086070	15	53 096 678	18.89
	SIRI0001427	15	86 674 393	18.82
	ASGA0091586	15	32 670 468	17.93
	DRGA0012847	13	75 221 598	17.84
	MARC0036482	15	83 943 521	17.71

na: data not available

These SNPs were tested by Structure program to determine their ability to separate Mangalica breed variants. All White animals were also incorporated into the iterations. The 24 SNPs were useful to perform assignment of the individuals to their corresponding groups with a probability higher than 0.8. Then systematically excluding the lower valued ($-\log_{10}P$) SNPs from the iterations, twelve, nine and six SNPs sets were used in the calculations and the assignment probabilities were determined. The best result for all animals (P>0.95) were

achieved by nine SNPs (Figure 2). In case of the top ranked six SNPs the group identification of Mangalica individuals was further improved (P>0.98), but one White animal became misclassified (data not shown).

Figure 2
Assignment of Mangalica and White pigs into four clades using nine selected SNPs using the Structure program (n=23 RM, 24 BM and 33 SBM, and 63 White crossbred pigs). Each animal refers to a vertical bar broken into four segments, representing an individual's estimated membership fraction in each of the four clusters.

For pedigree control 202 markers with call rate equal to one and with minor allele frequency higher than 0.4 in each Mangalica breed have been selected from the entire SNP data set. The 202 loci were reduced to 54, based on the chromosomal location of the markers and the distance between two adjacent SNPs. The average physical distance between two adjacent loci on a given chromosome was 3.7×10^7 bp. The 54 loci covered all but the Y chromosome of the swine genome (Table 3).

Using our set of 54 SNPs, the number of loci needed for both identity (the chance that two animals have the same genotype) and exclusion (excluding one animal as a parent) probabilities were set at different levels. To determine identity, the numbers of SNPs needed at 0.01, 0.001 and 0.0001 probability levels were five, eight and ten, respectively. For exclusion, 23 and 34 SNPs were needed at the 0.99 and 0.999 probability levels, respectively, when both parents were known.

When only one parent was known in the simulation, the corresponding numbers of SNPs were 36 and 54, respectively. When all 54 SNPs were included in the calculation, the exclusion probability levels were 0.999115 and 0.999985 for one and two parents, respectively, while the probability level for identity was 1.54×10^{-23}. In white pigs, using a panel of 60 SNPs (Rohrer *et al.* 2007), the corresponding values were 0.997391 and 0.999982 for one- and two-parents exclusion, respectively. The probability level for identity was 4.55×10^{-23}.

Discussion

We have found in this breed differentiation study that F_{ST} values obtained by thousands of SNP loci and by microsatellites (Zsolnai *et al.* 2006) are highly correlated. The observed F_{ST} values are very likely relying on the breeding history of Mangalica (see Figure 1). It is assumed that BM was developed the earliest by crossbreeding older Hungarian pig varieties with the Serbian Sumadia breed and then BM was used to breed both RM and SBM (Egerszegi *et al.* 2003). The F_{ST} data also indicated that the strongest and weakest genetic relationship is between the BM and SBM and the RM and SBM, respectively, while the strength of the BM–RM relationship is between them. Our study indicated that analysis of thousands of SNP loci

Table 3
Chromosomal positions of the 54 SNPs used for parentage testing in Mangalica pigs

Marker	Chromosome	Position	Distance to the previous loci (bp)
ALGA0001311	1	12 898 775	
ALGA0002972	1	36 428 348	23 529 573
H3GA0002674	1	103 544 479	67 116 131
ASGA0005840	1	196 235 537	92 691 058
MARC0111536	2	5 714 719	
ALGA0013102	2	22 683 636	16 968 917
ASGA0012938	2	78 689 048	56 005 412
ALGA0017286	3	959 104	
MARC0009789	3	18 366 792	17 407 688
ALGA0020170	3	40 908 089	22 541 297
ASGA0015618	3	46 142 849	5 234 760
ASGA0018835	4	16 679 929	
H3GA0012614	4	38 091 293	21 411 364
ALGA0027902	4	105 422 253	67 330 960
INRA0016870	4	111 339 127	5 916 874
ALGA0122915	5	2 967 683	
ALGA0030323	5	5 889 349	2 921 666
ASGA0025351	5	26 149 428	20 260 079
ALGA0109617	5	98 305 667	72 156 239
INRA0021220	6	4 136 665	
ALGA0036985	6	34 162 176	30 025 511
M1GA0024350	6	171 673 133	137 510 957
ASGA0031365	7	13 969 600	
MARC0056863	7	22 278 959	8 309 359
ASGA0032847	7	39 056 179	16 777 220
ASGA0038698	8	18 014 482	
DRGA0008675	8	41 223 881	23 209 399
MARC0076960	8	62 872 348	21 648 467
ASGA0043688	9	37 725 359	
ASGA0044267	9	59 946 882	22 221 523
ALGA0055749	9	75 339 172	15 392 290
ASGA0046475	10	8 522 122	
ALGA0058903	10	29 249 092	20 726 970
ALGA0112577	10	57 773 515	28 524 423
ASGA0049827	11	10 131 894	
ASGA0050560	11	22 195 749	12 063 855
ASGA0052353	11	63 188 981	40 993 232
H3GA0034325	12	20 480 737	
ASGA0055555	13	617 259	
ALGA0068017	13	7 699 226	7 081 967
ASGA0059979	13	116 044 203	108 344 977
H3GA0038403	14	4 224 189	
M1GA0018459	14	28 495 575	24 271 386
MARC0055325	14	85 285 041	56 789 466
ASGA0067052	14	135 760 672	50 475 631
MARC0045727	15	31 740 704	
ASGA0071775	15	95 597 162	63 856 458
H3GA0053443	15	161 449 543	65 852 381
ALGA0090112	16	18 208 098	
M1GA0021543	16	50 414 878	32 206 780
ASGA0075033	17	2 969 921	
M1GA0022894	17	55 975 265	53 005 344
ALGA0097941	18	21 384 842	
ASGA0080447	18	34 881 514	13 496 672

in dozens of animals per Mangalica breed groups were as useful as using ten microsatellite loci with more than 50 animals per breed for characterisation. Similar SNP based approaches have been also performed in Meishan and White crossbred pigs (Matsumoto *et al.* 2012).

Identification of a few SNPs for genotyping is desirable (Wang & Shete 2011). Whole SNP panelling is still expensive, while using only a few SNPs is more feasible both technically and financially. It was reported (Wilkinson *et al.* 2011) that among sixteen closely related cattle breeds about 200 SNPs are needed for separation. Frkonja *et al.* (2011) have systematically reduced the number of SNPs to 96 and 48 to detect the Red Holstein Friesian ratio in Swiss Fleckvieh individuals. We could demonstrate that as few as nine SNPs were sufficient for breed separation at a $P>0.95$ probability level in Mangalica pigs. However, it is mentionable here that such nine-SNP approach is no more able to describe the genetic variability in the studied breeds. On the other hand, these nine SNPs were chosen just for differentiation purposes. They can distinguish not only between Mangalica varieties, but also separate White pigs from Mangalica. It is suggested that after validation, using intensity values (Huang *et al.* 2010) of such SNPs, Mangalica and non-Mangalica ingredients can be distinguished and calculated in a processed food. Thereby, quality control of trade mark (Mangalica) food products can be promoted. Garcia *et al.* (2006) described a similar application to determine the ratio of Iberico in ham products by microsatellites. In addition, an effective SNP panel could also be useful for forensic applications where degraded samples might prohibit microsatellite typing (Dixon *et al.* 2006).

The SNP based comparison of breeds applied in our study could also be useful, if ancestry linked loci are requested amongst genetically more distinct breeds. In our example, a list of loci could be drawn to show the greatly influence of selection procedures in the Mangalica breed.

We have also identified SNPs for parentage testing in Mangalica. The identity and exclusion testing can be achieved at a very high probability level with the application of the selected 54 SNP loci. Rohrer *et al.* (2007) reported that 60 SNPs were useful in White pigs for parental exclusion, and the exclusion and identity probabilities are similar to the values obtained in our study.

In summary, it becomes evident that microsatellite genotyping can be replaced successfully by SNP genotyping. Employing the Illumina PorcineSNP60 chip, breed characterization and parentage testing could be done to describe in more detail the Mangalica breed variants.

Acknowledgements

The research is supported by Hungarian National Innovation Office (TECH_08-A3/2-2008-0405).

References

Andersson L (2009) Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. Genetica 136, 341-349

Brüssow KP, Egerszegi I, Ratky J, Torner H, Toth P, Schneider F (2005) Reproduction in the Hungarian Mangalica pig. Pig News and Information 26, 23N-28N

Dixon LA, Dobbins AE, Pulker HK, Butler JM, Vallone PM, Coble MD, Parson W, Berger B, Grubwieser P, Mogensen HS, Morling N, Nielsen K, Sanchez JJ, Petkovski E, Carracedo A, Sanchez-Diz P, Ramos-Luis E, Brion M, Irwin JA, Just RS, Loreille O, Parsons TJ, Syndercombe-Court D, Schmitter H, Stradmann-Bellinghausen B, Bender K, Gill P (2006) Analysis of artificially degraded DNA using STRs and SNPs–results of a collaborative European (EDNAP) exercise. Forensic Sci Int 164, 33-44

Egerszegi I, Rátky J, Solti L, Brüssow K-P (2003) Mangalica - an indigenous swine breed from Hungary (Review). Arch Tierz 46, 245-256

Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics 164, 1567-1587

Fan B, Onteru SK, Du Z-Q, Garrick DJ, Stalder KJ, Rothschild MF (2011) Genome-Wide Association Study Identifies Loci for Body Composition and Structural Soundness Traits in Pigs. PLoS One 6, e14726

Frkonja A, Gredler B, Schnyder U, Curik I, Sölkner J (2011) How to Use Fewer Markers in Admixture Studies? ACS 76, 187-190

Garcia D, Martinez A, Dunner S, Vega-Pla JL, Fernandez C, Delgado JV, Canon J (2006) Estimation of the genetic admixture composition of Iberian dry-cured ham samples using DNA multilocus genotypes. Meat Sci 72, 560-566

Hong CB, Kim YJ, Moon SH, Shin YA, Go MJ, Kim DJ, Lee JY, Cho YS (2012) BioSMACK: a Linux Live CD for genome-wide association analyses. BMB reports 45, 44-46

Huang W, Kirkpatrick BW, Rosa GJM, Khatib H (2010) A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in Holstein cattle. Anim Genet 41, 570-578

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy VH, Nicholas FW, Raadsma H for the International Sheep Genomics Consortium (2009) A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. PLoS ONE 4, e4668

Kim Y, Ryu J, Woo J, Kim JB, Kim CY, Lee C (2011) Genome-wide association study reveals five nucleotide sequence variants for carcass traits in beef cattle. Anim Genet 42, 361-365

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TPL, Sonstegard TS, Van Tassell CP (2009) Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS One 4, e5350

Matsumoto T, Okumura N, Uenishi H, Hayashi T, Hamasima N, Awata T (2012) Population structure of pigs determined by single nucleotide polymorphisms observed in assembled expressed sequence tags. Anim Sci J 83, 14-22

Nechtelberger D, Kaltwasser C, Stur I, Meyer J-N, Brem G, Mueller M, Mueller S (2001) DNA Microsatellite Analysis for Parentage Control in Austrian Pigs. Anim Biotechnol 12, 141-144

Orr N, Back W, Gu J, Leegwater P, Govindarajan P, Conroy J, Ducro B, Van Arendonk JAM, MacHugh DE, Ennis S, Hill EW, Brama PAJ (2010) Genome-wide SNP association–based localization of a dwarfism gene in Friesian dwarf horses. Anim Genet 41 (Suppl.), 2-7

Patterson N, Price AL, Reich D (2006) Population Structure and Eigenanalysis. PloS Genet 2, e190

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6, 288-295

Rohrer GA, Freking BA, Nonneman D (2007) Single nucleotide polymorphisms for pig identification and parentage exclusion. Anim Genet 38, 253-258

Wilkinson S, Wiener P, Archibald AL, Law A, Schnabel RD, McKay SD, Taylor JF, Ogden R (2011) Evaluation of approaches for identifying population informative markers from high density SNP Chips. BMC Genet 12, 45

Wang J, Shete S (2011) A powerful hybrid approach to select top single-nucleotide polymorphisms for genome-wide association study. BMC Genet 12, 3

Zsolnai A, Radnóczy L, Fésüs L, Anton I (2006) Do Mangalica Pigs of Different Colours Really Belong to Different Breeds? Arch Tierz 49, 477-483