PL-ISSN 0015-5497 (print), ISSN 1734-9168 (online)
Folia Biologica (Kraków), vol. 61 (2013), No 3-4
doi: 10.3409/fb61_3-4.173

The Evaluation of Bovine SNP50 BeadChip Assay Performance in Polish Red Cattle Breed

Artur GURGUL, Kacper ŻUKOWSKI, Klaudia PAWLINA, Tomasz ZĄBEK, Ewelina SEMIK, and Monika BUGNO-PONIEWIERSKA

Accepted May 15, 2013

GURGUL A., ŻUKOWSKI K., PAWLINA K., ZĄBEK T., SEMIK E., BUGNO-PONIEWIERSKA M. 2013. The Evaluation of Bovine SNP50 BeadChip Assay Performance in Polish Red Cattle Breed. Folia Biologica (Kraków) **61**: 173-176.

According to increasing interest in the use of high density SNP (single nucleotide polymorphism) genotyping assays for genome-wide genetic studies in farm animals, there is a need to assess the usefulness of currently available genomic tools for application in different breeds. The performance of the assays may differ between the breeds because of discrepancies in allele frequencies of the polymorphisms or differences in linkage disequilibrium patterns. In this study we attempted to test the performance of the Bovine SNP50 v2 genotyping assay (Illumina) for population genetic and other applications in the Polish Red cattle breed. We found that 37,977 of the 53,438 autosomal markers included in the assay give high quality genomic information and can be used for different applications concerning this breed. The remaining markers were denoted as "of limited use" or redundant because of their weak performance, deviation from Hardy-Weinberg equilibrium, low minor allele frequency, high linkage disequilibrium with other neighboring markers or location on sex chromosomes.

Key words: Cattle, Polish Red, SNP50 BeadChip.

Artur GURGUL, Klaudia PAWLINA, Tomasz ZĄBEK, Ewelina SEMIK, Monika BUGNO-PONIEWIERSKA, National Research Institute of Animal Production, Laboratory of Genomics, Krakowska 1, 32-083 Balice, Poland.
E-mail: agurgul@izoo.krakow.pl
Kacper ŻUKOWSKI, National Research Institute of Animal Production, Departament of Animal Breeding and Genetics, Krakowska 1, 32-083 Balice, Poland.

Recent developments in molecular biology techniques have caused a significant reduction in the costs of genomic analysis and have enabled the use of high density SNP microarrays in studies involving farm animals. Currently available genomic SNP panels for farm animals provide uniform genome coverage and have been proven to be useful tools for population genetics, association studies and quantitative trait *loci* fine mapping in several populations (MORAN *et al.* 2006; NIJMAN *et al.* 2008; MATUKUMALLI *et al.* 2009).

The vital element of each of these applications is adequate marker selection. The selected markers have to be characterized by several parameters such as polymorphism, high minor allele frequency, uniform distribution across the genome, and ease of analysis. Some of these parameters may differ between different breeds or populations and may limit the use of SNP panels as tools for genome-wide association studies or population genetics (WILKINSON *et al.* 2011).

According to increasing interest in the genomic selection of cattle based on SNP arrays (especially BovineSNP50 BeadChip, Illumina) there is a huge amount of genomic information concerning the most popular cattle breeds (HARRIS *et al.* 2008; HAYES *et al.* 2009). However, there is only limited information about the performance of currently available genomic tools in native or rare cattle breeds (MATUKUMALLI *et al.* 2009; MELKA *et al.* 2011). Despite the fact that genomic selection does not concern preserved cattle breeds, with low population sizes, these breeds may be the subject of population genetics and phylogenetic studies. The Polish Red Breed is included in the FAO National Rare Livestock Breeds Preservation Programme, and is characterized by valuable phenotypic traits characteristic for primitive populations. Accordingly, this breed is an interesting subject of studies aimed at the recognition of genetic diversity of cattle breeds and functioning of the bovine genome.

In this study we attempt to evaluate the usefulness of the most popular SNP genotyping assay (BovineSNP50; Illumina) for genomic studies in preserved Polish Red cattle breed. We have also assessed the number of informative and polymorphic SNPs which can be used for population genetics, association studies and other genomic applications in this breed.

Material and Methods

The material was genomic DNA isolated from whole blood of randomly selected 172 cows and 20 sires of Polish Red Breed (RP). Genotypes of 54,609 SNP *loci* were established using Bovine SNP50 BeadChip v2 (Illumina) according to the manufacturer's instructions. Raw data analysis was performed in GenomeStudio software (Illumina). Clustering was based on standard Illumina *.egt* files containing cluster positions. To evaluate marker panel performance in the Polish Red Breed, the following parameters were analyzed: call rate, number of polymorphic/fixed *loci*, minor allele frequency (MAF) and Hardy-Weinberg equilibrium. Because the studied population comprised both cows and sires (unequal sex ratio), after call rate calculation the markers located on sex chromosomes (1,170 SNPs on BTA X and 1 on BTA Y) were analyzed separately. This was also dictated by the fact that SNPs located on sex chromosomes require special treatment in some applications (such as copy number variant identification or the analysis of heterozygosity) and thus may be of limited use in several analytical approaches.

SNP pruning was performed with the PLINK whole genome association analysis toolset (http://pngu.mgh.harvard.edu/purcell/plink/) (PURCELL *et al.* 2007). This operation allows for exclusion of markers in strong linkage disequilibrium LD ($r^2 \geq 0.9$) with neighboring SNPs and provides a panel of unlinked informative SNPs. Average marker distance was calculated based on Illumina-provided genomic coordinates after removing pruned SNPs and SNPs with no genomic coordinates.

Results and Discussion

Call rate

In Illumina's Infinium technology, poor marker performance is expressed by low gencall scores being the result of poorly separated genotype clusters or resulting from the situation in which observed intensities do not match predefined cluster positions. The markers with gencall scores classified beneath the so called no-call threshold are rejected from further analysis and lower the average call rate for each sample (OLIPHANT *et al.* 2002; FAN *et al.* 2003). The call rate measure is highly dependent on the quality of the DNA used for analysis and quality of laboratory work, however, it gives relevant information about general assay performance. In our experiment the call rate integer ranged from 0.9819 to 0.9973 with a mean value of 0.9941 (s. d.= 0.0028). This shows that on average, 54,287 markers were genotyped for each studied animal. The results are comparable or even better than that observed during assay development (MATUKUMALLI *et al.* 2009) in which average call rate was estimated at 97.5%. In our dataset 53,559 markers were genotyped in more than 90% of the animals and only 258 SNPs were not called in any of the animals. Over 90% of the samples were genotyped for more than 99% of markers.

Number of polymorphic and fixed *loci*

Non-polymorphic (fixed) *loci* are non-informative and thus lower the overall value of the genomic SNP panel (RINCON *et al.* 2011). The number of polymorphic *loci* for the SNP50 BeadChip differed significantly between the different cattle breeds and ranged from 29,444 for Brahman cattle to 43,723 for Charolaise cattle (http://www.illumina.com/documents/products/datasheets/datasheet_bovine_snp5O.pdf). The total number of fixed autosomal *loci* for Polish Red breed was 5,178 (9.69 % of all autosomal markers) giving a high number of 48,260 polymorphic *loci*. Such a high number of polymorphic *loci* reported in this work may result from a significantly higher number of studied animals when compared to the testing group used as IIlumina's test panel. Additionally, in the study of MATUKUMALLI *et al*. (2009) the percentage of non-informative *loci* across different breeds ranged from 9-38% with a mean value of ~17%. However, the results are not directly comparable, because they include markers with a frequency up to 1%. The percentage of such markers (fixed and with MAF<1%) in our dataset was calculated at 10.5%.

Minor allele frequency (MAF)

Minor allele frequency limits the use of markers for association studies because the effect of the rare variants is difficult to estimate. Typical genome wide association studies (GWAS) are designed to exclude markers with minor allele frequency below 1% or even 5%. In our experiment the MAF for all autosomal SNPs in RP breed ranged from 0 to 0.5 with an average value of 0.2266 (s.d.= 0.1584) and was slightly lower (MATUKUMALLI *et al.* 2009) or similar (EDEA *et al.* 2012) to that observed in previously studied

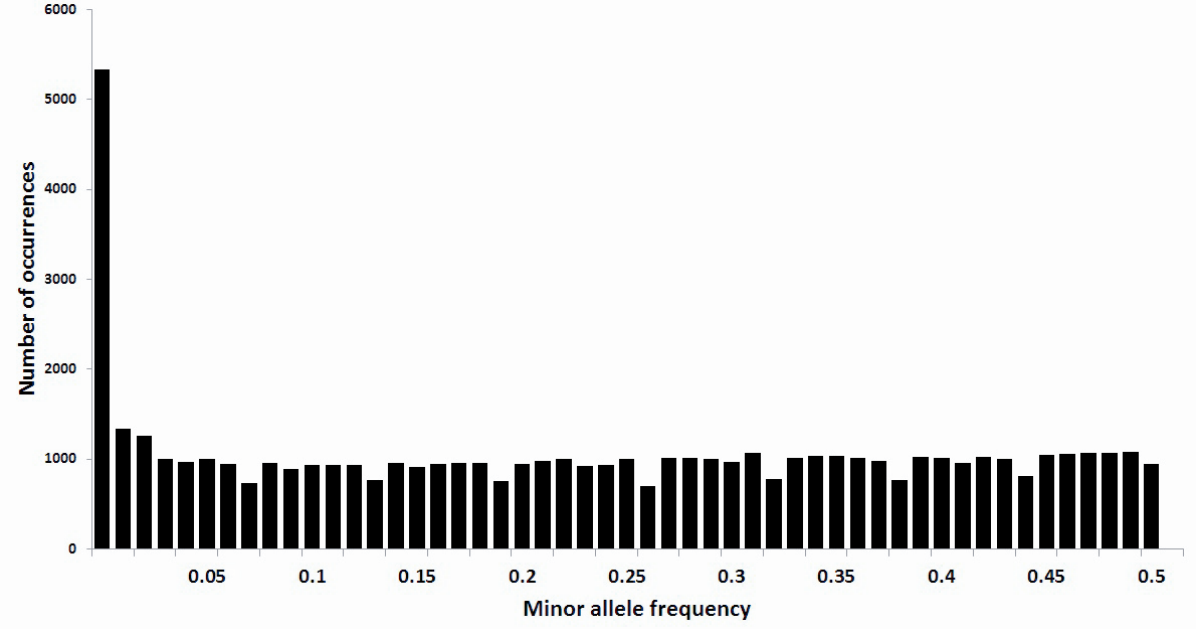

Fig. 1. Minor allele frequency distribution of SNPs included in the Bovine SNP50 genotyping assay for Polish Red breed.

worldwide and indigenous cattle breeds. The average MAF per chromosome ranged from 0.2184 for BTA 22 to 0.2394 for BTA 25. After excluding monomorphic *loci* and *loci* with low polymorphism (MAF<0.01), the average MAF for remaining SNPs increased to 0.2510 (s.d.=0.1473) and ranged from 0.0026 to 0.5. Moreover, 47,815 (89.5%) *loci* had a MAF higher than 1% and 10,219 (19.1%) *loci* had MAF higher than 40%. The MAF distribution in the studied population is shown on Fig. 1.

Hardy-Weinberg equilibrium (HWE)

HWE analysis allows for the identification of *loci* potentially under selection or *loci* prone to genotyping errors (LEAL 2005). In the studied population we found 2,935 *loci* which deviated from HWE (P<0.05) by using Fisher's exact test. The *loci* were distributed uniformly across all 29 autosomes and accounted for 5.5% of all studied markers.

SNPs localized on sex chromosomes

SNPs located on sex chromosomes were analyzed separately because of their specificity and possible impaired performance in males, especially for SNPs with positions outside the PAR (pseudoautosomal) region. Of the 1171 SNPs assigned to the BTAX and BTAY, 884 had MAF higher than 1%. From the initial number, 52 SNPs deviated from HWE (p<0.05). Only 18 SNPs performed poor (were not called in more than 90% of animals). When taking into account all filtration criteria, 828 SNPs showed satisfying performance and could be used for further applications concerning the analysis of sex chromosomes.

SNP pruning

The LD-based approach allows the identification of SNPs that are linked and thus only one of them is necessary to describe genetic variation in a studied genomic region (RINCON *et al.* 2011). Linked SNPs and SNPs being in high LD are presumed to give the same genetic information and thus in some applications (e. g. identification of runs of homozygosity or genomic inbreeding coefficients calculation), LD-based SNP pruning is recommended (PURCELL *et al.* 2007). Here we used the r^2 correlation coefficient to identify SNPs that are linked and to exclude the redundant markers. We used a threshold of r^2=0.9. This approach allowed to prune 7,794 SNPs distributed on all 29 autosomes. For applications such as genetic merit prediction or quantitative trait *loci* fine mapping, the distance between SNPs is crucial. SNP pruning increases average gap size between the markers, nevertheless the pruned markers in most cases do not provide new information to the study. The average distance between the makers for the Bovine SNP50 BeadChip is 49.4 kb according to the marker map provided by Illumina. SNP pruning showed that the distance between informative markers is slightly higher and was calculated at 55.1 kb.

Summary

The performed analysis showed that 11,592 markers included in Bovine SNP50 BeadChip are of limited use for population genetics or association studies in Polish Red cattle breed. This marker set comprises *loci* which were not called, are non-polymorphic, had MAF lower than 1%, were genotyped in less than 90% of studied animals or were pruned based on LD or have no genomic coordinates. This gives 37,977 informative SNPs (71% of all autosomal markers) which can be used for various genomic applications. The results obtained are comparable with previously described observations for different worldwide cattle breeds (MATUKUMALLI *et al.* 2009). Additionally, 828 SNPs with positions on sex chromosomes and 7,794 SNPs that were excluded as linked with other assayed markers can be retained for further analysis, resulting in a set of 46 599 well performing SNPs.

Summarizing, our study shows that the Bovine SNP50 BeadChip assay consists of a large set of informative SNP markers which can be used for genetic studies in the Polish Red cattle breed. Our work will be continued to establish genetic diversity and variability within Polish Red breed and to support the national biodiversity preservation program carried out by our Institute.

References

EDEA Z., DADI H., KIM S.W., DESSIE T., KIM K.S. 2012. Comparison of SNP variation and distribution in indigenous ethiopian and korean cattle (hanwoo) populations. Genomics Inform. **10**: 200-205.

FAN J.B., OLIPHANT A., SHEN R., KERMANI B.G., GARCIA F., GUNDERSON K.L., HANSEN M.S., STEEMERS F., BUTLER S.L., DELOUKAS P., GALVER L., HUNT S., MCBRIDE C., BIBIKOVA M., RUBANO T., CHEN J., WICKHAM E., DOUCET D., CHANG W., CAMPBELL D., ZHANG B., KRUGLYAK S., BENTLEY D., HAAS J., RIGAULT P., ZHOU L., STUELPNAGEL J., CHEE M.S. 2004. Highly Parallel SNP Genotyping. Cold Spring Harbor Symposia on Quantitative Biology, vol. LXVIII, 69-78, January.

HARRIS B.L., JOHNSON D.L., SPELMAN R.J. 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Proceedings of the Interbull Meeting, Niagara Falls, NY.

HAYES B.J., BOWMAN P.J., CHAMBERLAIN A.J., GODDARD M.E. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. Dairy Sci. **92**: 433-443.

LEAL S.M. 2005. Detection of genotyping errors and pseudo-SNPs via deviations from Hardy-Weinberg equilibrium. Genet. Epidemiol. **29**: 204-214.

MATUKUMALLI L.K., LAWLEY C.T., SCHNABEL R.D., TAYLOR J.F., ALLAN M.F., HEATON M.P., O'CONNELL J., MOORE S.S., SMITH TIMOTHY P.L., SONSTEGARD T.S., VAN TASSELL C.P. 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS ONE **4**: e5350.

MELKA H.D., JEON E.K., KIM S.W., HAN J.B., YOON D., KIM K.S. 2011. Identification of genomic differences between Hanwoo and Holstein Breeds using the Illumina Bovine SNP50 BeadChip. Genomics Inform. **9**: 69-73

MORAN J.L., BOLTON A.D., TRAN P.V., BROWN A., DWYER N.D., MANNING D.K., BJORK B.C., LI C., MONTGOMERY K., SIEPKA S.M., VITATERNA M.H., TAKAHASHI J.S., WILTSHIRE T., KWIATKOWSKI D.J., KUCHERLAPATI R., BEIER D.R. 2006. Utilization of a whole genome SNP panel for efficient genetic mapping in the mouse. Genome Res. **16**: 436-440.

NIJMAN I.J., KUIPERS S., VERHEUL M., GURYEV V., CUPPEN E. 2008. A genome-wide SNP panel for mapping and association studies in the rat. BMC Genomics **9**: 95.

OLIPHANT A., BARKER D.L., STUELPNAGEL J.R., CHEE M.S. 2002. BeadArray™ Technology: enabling an accurate, cost-effective approach to high-throughput genotyping. snps; discovery of markers for disease. Biotechniques Suppl: 56-8, 60-1.

PURCELL S., NEALE B., TODD-BROWN K., THOMAS L., FERREIRA M.A.R., BENDER D., MALLER J., SKLAR P., DE BAKKER P.I.W., DALY M.J., SHAM P.C. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. **81**: 559-575.

RINCON G., WEBER K.L., VAN EENENNAAM A.L., GOLDEN B.L., MEDRANO J.F. 2011. Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys. J Dairy Sci. **94**: 6116-6121.

WILKINSON S., WIENER P., ARCHIBALD A.L., LAW A., SCHNABEL R.D., MCKAY S.D., TAYLOR J.F., OGDEN R. 2011. Evaluation of approaches for identifying population informative markers from high density SNP chips. BMC Genet. **12**: 45.