Performance of the Axiom 90k Buffalo Genotyping Array in four Philippine water buffalo populations

Jesus Rommel Herrera[1,2], Ester Flores[3], Cedric Gondro[2] and Julius van der Werf[2]

[1] *Philippine Carabao Center-University of the Philippines, College, Laguna 4031 Philippines*

[2] *School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia 2350*

[3] *Philippine Carabao Center National Headquarters Muñoz, Nueva Ecija 3120 Philippines*

Abstract

The Axiom Buffalo Genotyping Array, a high-density single nucleotide polymorphism (SNP) chip for water buffalo (*Bubalus bubalis*) was used on four water buffalo breeds in the Philippines to test the suitability of the SNP chip on the local buffalo population. A total of nine hundred eighteen (918) DNA samples from both male and female buffaloes were submitted for genotyping. The Axiom analysis suite software was used to analyze the raw data for quality control metrics and generating genotype calls. For downstream analysis, principal component analysis (PCA) and heatmap based on genomic relationship matrix (GRM) using R was done to visualize the relationship structure among the four buffalo breeds.

Polymorphic (PHR) SNPs useful for downstream study were identified for the four Philippine water buffalo populations using the Axiom 90K Buffalo Genotyping Array. The number of polymorphic markers for the 3 riverine breeds is higher ranging from 57,094 to 67,810, compared to those of the Philippine swamp population with only 16,573 PHRs since the SNPs included in the array all came from riverine breeds. The genotyping array will be useful for genomic studies in the four buffalo populations. However, for the swamp population with a longer inter-marker distance in autosomes, the array's usefulness will be limited to population diversity studies. Common PHR SNPs among the three riverine breeds was determined to be 46,445. There were 10,443 PHR SNPs common to the four populations and could serve as a basis for the design of a lower density SNP chip specific for Philippine buffalo populations. The PCA plot and heatmap generated from GRM using only the PHR SNPs clearly separated the riverine and swamp populations and showed the genetic relationship among the riverine breeds.

Key words: Water buffaloes, Axiom Buffalo Genotyping Array, PCA plot

Introduction

The Axiom Buffalo Genotyping Array (Affymetrix, Inc., Sta. Clara, CA) is the first commercially available high-density single nucleotide polymorphism (SNP) chip for water buffalo (*Bubalus bubalis*). It contains 123,040 probes, representing 90,000 SNPs, ~ 33,000 are duplicate probes (Nicolazzi et al., 2014). Of these 90,000 SNPs, 84,820 are found in the 29 autosomes, 5,132 in the X chromosome, 12 in the Y chromosome and 36 have no chromosome position. Given this, the total length covered by the SNPs (genome coverage) is 2.6MB with an average inter-probe spacing of 21.6 kb per chromosome (Table 1).

The SNPs included in the array are based on buffalo sequencing data; but for SNP positioning, the cattle genome (UMD3.1 assembly) was used as the reference genome. SNPs used in the array came from 4 riverine breeds: Mediterranean, Murrah, Jaffarabadi, and Nili-Ravi, in the ratio 30:30:20:20 (Iamartino et al., 2013). Content validation was performed on a riverine population (samples coming from 10 countries), a swamp population (samples coming from 5 countries), one South African Cape buffalo population and one Indonesian anoa population. The proportion of markers that was found to be polymorphic were 74.8%, 48.3%, 0.8%, and 2.6% for these four populations, respectively (Nicolazzi et al., 2014). There are still few published reports on the utilization of the chip for genomic analysis. El-Halawany et al. (2015) used 70,182 polymorphic SNPs for a Genome wide Association Study (GWAS) for daily milk production in Egyptian buffaloes; and de Camargo et al. (2015) utilized 61,145 SNPs for GWAS and gene network analysis in Brazilian buffaloes.

The Philippines has both swamp and riverine buffalo populations which include a small number of Brazilian and Italian Mediterranean breeds and except for the latter two breeds, the SNP chip has not been tested on these local breeds. The objective of this study was to present the performance of the Axiom 90k Buffalo Genotyping Array in four Philippine water buffalo populations in terms of QC call rates, heterozygosity and number of polymorphic markers identified for downstream analysis on a per population basis.

Table 1. Content of the Axiom 90k Buffalo Genotyping Array. Data

Chromosome	Number of probes	Number of SNPs	Length (Mb)	Ave. probe distance (kb)
1	7,396	5,404	158.2	21.4
2	6,416	4,673	137.0	21.4
3	5,637	4,130	121.3	21.5
4	5,547	4,070	120.7	21.8
5	5,606	4,096	121.1	21.6
6	5,485	3,992	119.4	21.8

7	5,161	3,826	112.6	21.8
8	5,206	3,800	113.3	21.8
9	4,776	3,492	105.6	22.1
10	4,793	3,500	104.2	21.7
11	4,940	3,581	107.2	21.7
12	4,168	3,061	91.1	21.9
13	3,923	2,822	84.2	21.5
14	4,022	2,901	84.6	21.0
15	3,900	2,877	85.2	21.8
16	3,789	2,779	81.6	21.5
17	3,525	2,558	75.1	21.3
18	3,122	2,260	65.9	21.1
19	3,037	2,167	63.7	21.0
20	3,218	2,392	72.0	22.4
21	3,269	2,402	71.5	21.9
22	2,809	2,051	61.3	21.8
23	2,423	1,764	52.5	21.7
24	2,955	2,122	62.7	21.2
25	2,086	1,507	42.8	20.5
26	2,403	1,756	51.6	21.5
27	2,060	1,516	45.3	22.0
28	2,149	1,586	46.2	21.5
29	2,348	1,735	51.4	21.9
X	6,823	5,132	148.7	21.8
Y	12	12	-	-
No chr position	36	36	-	-
Total	123,040	90,000	2658.2	NA

Materials and Methods

Blood samples from 612 Bulgarian Murrah, 156 Brazilian Murrah and 55 Italian Mediterranean animals were used in the study. These three riverine breeds are selected on milk traits. Ninety-five (95) blood samples from a population of the Philippine swamp (SP) buffalo selected for growth traits were also included. All animals are part of the institutional herds managed by the Philippine Carabao Center. Genomic DNA was extracted using the Promega ReliaPrep Blood gDNA Miniprep System according to the manufacturer's protocol. DNA quantification was done using the Promega Quantus Fluorometer. Samples were first subjected to RNA purification, a requirement of the USDA, prior to shipment to Affymetrix, Inc., Sta. Clara, California. Submitted samples were genotyped using the Axiom Buffalo Genotyping Array. Generated ".cel" files were analysed using the Axiom Analysis Suite software. The configuration used is the Best Practices Workflow mode. Moreover, some default settings for the sample and SNP QC used were: dish QC ≥ 0.82, QC_call_rate ≥ 97,

plate QC_percent samples passed ≥ 95, plate_QC_average call rate ≥ 98.5, species type is diploid and number of minor allele ≥ 2.

Venn diagrams were generated using R to show common polymorphic SNPs among the populations. Principal Component Analysis (PCA) and heatmap based on the genomic relationship matrix (GRM) were generated using R to visualize the relationship of the river and swamp types, and the relationship structure among the four buffalo populations.

Results

Table 2 shows the sample quality control (QC) results as generated by the Analysis Suite. Of the 918 samples submitted for genotyping, only 904 passed for genotyping QC. Eight samples failed the dish QC cut-off of 0.82 while six failed the QC call rate of 97. Dish QC is the single-sample measure of "interference" between foreground and background signal distributions.

Average QC call rates of the passing samples for all four populations are 99.6-99.7%, comparable to what the Affymetrix, Inc. found in Murrah and Mediterranean samples (The Axiom Buffalo Genotyping Array DataSheet).

In terms of heterozygosity for all SNPs, values (%) for the three riverine populations (Bulgarian Murrah, Brazilian Murrah, Italian Mediterranean) are similar (34.1, 34.2 and 31.2, respectively,) while the value for the Philippine swamp population is only 9.6%. This low value can be explained by the fact that no swamp buffalo samples were included in the design of the array. All SNPs included came from riverine breeds: Mediterranean, Murrah, Jaffarabadi, and Nili-Ravi, in the ratio 30:30:20:20 (Iamartino et al., 2013).

Table 2. Sample QC results of the four Philippine water buffalo populations

Population	Number of samples	Samples passing DQC, QC CR and Plate QC	Average QC CR for the passing samples (%)	Heterozygosity (%)
Bulgarian Murrah	612	602	99.7	34.1
Brazilian Murrah	156	155	99.7	34.2
Italian Mediterranean	55	54	99.6	31.2
Philippine Swamp	95	93	99.6	9.6

Table 3 presents the QC results and genotype statistics of the 90K buffalo genotyping array in four Philippine water buffalo populations. For diploid species, the Poly High Resolution (PHR) SNPs are recommended for statistical tests in downstream study as these are the polymorphic SNPs that are informative (Axiom Analysis Suite 1.1 User Guide 2015). The number of PHR SNPs for the three riverine populations (Bulgarian Murrah, Brazilian Murrah, Italian Mediterranean) are 67,810, 66,902 and 57,094, respectively. In the

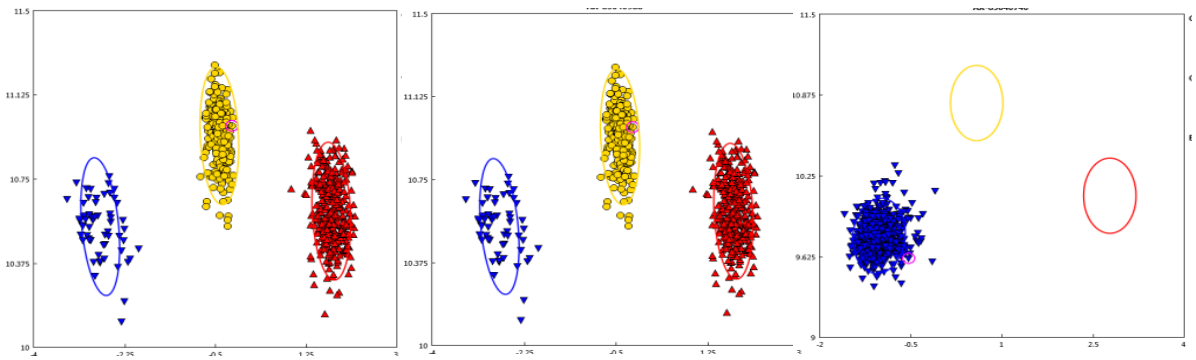
case of the Philippine swamp buffalo, the number of PHR SNPs is only 16,573. Most of its SNPs belong to the Mono High Resolution (37.6%) and the No minor homozygote (29.9%) categories. The high number of PHR SNPs identified in riverine breeds and the low number of PHR SNPs in the swamp population is due to the design of the SNP chip wherein only SNPs from riverine breeds were included. Among the three riverine breeds, the Bulgarian Murrah and Brazilian Murrah populations have higher PHR SNPs compared to the Italian Mediterranean population since the first 2 populations have both Murrah and Mediterranean blood (Borghese, 2005); both of these breeds were included in the design of the SNP chip.

Table 3. Performance of of the Axiom 90k Buffalo Genotyping Array in four Philippine water buffalo populations.

Population	SNP count based on genotype cluster category (% in brackets)						
	PHR	NMH	MHR	Call rate below threshold	OTV	Hemi	Other
Bulgarian Murrah (n=602)	67,810 (75.3)	1,113 (1.2)	7,059 (7.8)	3,687 (4.1)	705 (0.8)	12 (0.001)	9,614 (10.7)
Brazilian Murrah (n=155)	66,902 (74.3)	1,712 (1.9)	6,929 (7.7)	3,983 (4.4)	690 (0.8)	12 (0.001)	9,772 (10.8)
Italian Mediterranean (n=54)	57,094 (63.4)	9,748 (10.8)	8,555 (9.5)	3,166 (3.5)	884 (1.0)	12 (0.001)	10,541 (11.7)
Philippine Swamp (n=93)	16,573 (18.4)	26,935 (29.9)	33,892 (37.6)	1,665 (1.8)	1,194 (1.3)	12 (0.001)	9,729 (10.8)

PHR - polyhigh resolution, NMH - no minor homozygote, MHR - monohigh resolution, OTV - off targt variant, Hemi - hemozygous

Figure 1 shows cluster plots representative of three of the genotype cluster categories generated by the Axiom Buffalo Genotyping Array. The informative markers are those that cluster well as seen in the PolyHighResolution category.



SNP in the PolyHighResolution category- with good cluster resolution and at least two examples of the minor allele.

SNP in the NoMinorHom category-with 2 clusters with no minor allele

SNP in the MonoHighResolution category- with good cluster resolution but less than two examples of the minor allele

Figure 1. The cluster plots on the left, middle and right are representative examples of Poly High Resolution, No Minor Homozygote and Mono High Resolution genotypes.

Common PHR SNPs among the 3 riverine breeds are shown in Figure 2.1, totaling 46,445. If the PHR SNPs of the Philippine swamp population are included, common PHR

SNPs went down to 10,443 (Figure 2.2). These common SNPs could be the basis for the development of a lower density SNP chip specific for Philippine buffalo populations.

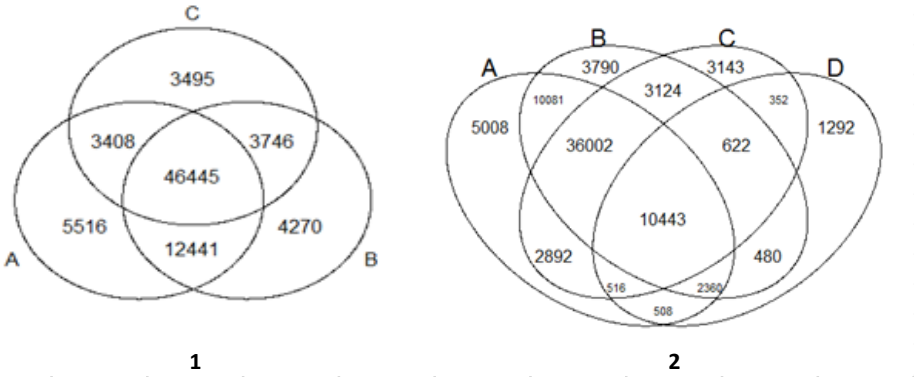


Figure 2. (1) Venn diagram of Polymorphic High Resolution (PHR) SNPs from 3 Philippine riverine populations. (2) Venn diagram of Polymorphic High Resolution (PHR) SNPs from 4 Philippine buffalo populations. A=Bulgarian Murrah, B=Brazilian Murrah, C=Italian Mediterranean, D=Philippine Swamp.

Figure 3 shows a Principal Component Analysis (PCA) plot generated based on the genomic relationship matrix (GRM) using only the PHR SNPs (n=67,569) of the 4 Philippine buffalo populations if all samples are used at the same time in the Analysis Suite. The first principal component (PC1), with a variance of 89.5%, splits the data into the riverine type (BrMB,BMB,ItMB) on the left side and the swamp type (SP1) to the right. The variation due to PC2 is only 2.45%. BMB and BrMB samples overlapped one another, since the 2 breeds have both Murrah and Mediterranean blood.

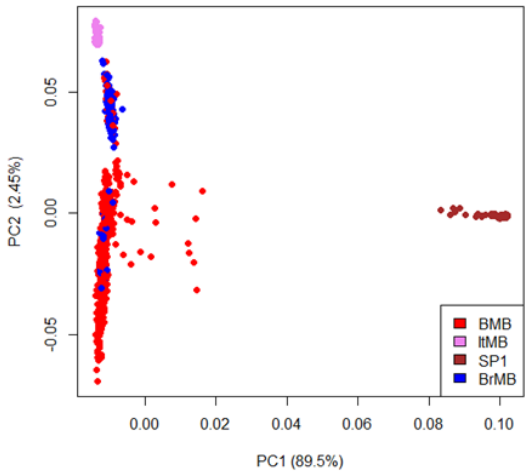


Figure 3. PCA plot among animals belonging to 3 riverine buffalo populations and 1 swamp buffalo population (n=902). BMB =Bulgarian Murrah, BrmB= Brazilian Murrah, ItMB= Italian Meditteranean, SP1= Philippine Swamp. PCA plot generated using R scripts.

A heatmap was also generated based on the GRM using the very same SNPs used in the PCA plot (Figure 4). Again, there is a clear separation between the riverine and

swamp types. The Italian Mediterranean population is clearly a distinct population although is closely related to the Bulgarian Murrah population. Again, overlapping of Bulgarian and Brazilian Murrah animals were observed. From the above information, the reference population for GWAS and prediction of genomic breeding values will be composed of Bulgarian and Brazilian Murrah populations.

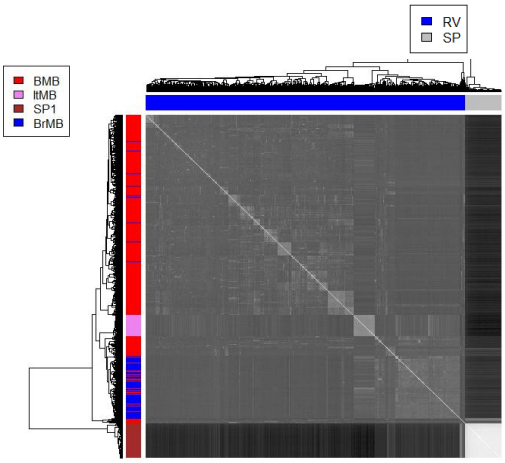


Figure 4. Heatmap of relationship among animals belonging to 3 riverine buffalo populations and 1 swamp buffalo population (n=902). RV=Riverine type, SP=Swamp type, BMB =Bulgarian Murrah, BrmB= Brazilian Murrah, ItMB= Italian Meditteranean, SP1= Philippine Swamp

Considering only SNPS with MAF $\geq$ 0.01 and HW p-value $\geq$ 0.01, the number of autosomal polymorphic markers in the Bulgarian Murrah, Brazilian Murrah, Italian Mediterranean and Philippine swamp populations were 63,252, 62,709, 53,917 and 14,578 respectively. Average inter-marker in autosomes in kb are 39.6, 39.9, 46.3 and 170.0 for the Bulgarian Murrah, Brazilian Murrah, Italian Mediterranean and Philippine swamp populations, respectively.

Table 4. Number of autosomal polymorphic SNPs in four water buffalo populations in the Philippines.

Population	Total no. of Polymorphic SNPs	Polymorphic SNPs with MAF$\geq$ 0.01 and HW p-value $\geq$ 0.01	Autosomal Polymorphic SNPs	Average inter-marker distance in autosomes (kb)
Bulgarian Murrah	67,810	64,561 (71.7)	63,252 (70.3)	39.6
Brazilian Murrah	66,902	65,298 (72.6)	62,709 (69.7)	39.9
Italian Mediterranean	57,094	56,106 (62.3)	53,917 (59.9)	46.3
Philippine Swamp	16,573	14,920 (16.6)	14,578 (16.2)	170

MAF - minor allele frequency, Hardy-Weinberg

Conclusions

Polymorphic (PHR) SNPs useful for downstream study were identified for the four Philippine water buffalo populations using the Axiom 90K Buffalo Genotyping Array. The number of polymorphic markers for the 3 riverine breeds is higher compared to those of the Philippine swamp population since the SNPs included in the array all came from riverine breeds. The genotyping array will be useful for genomic studies in the four buffalo populations. However, for the swamp population with a longer inter-marker distance in autosomes, the array's usefulness will be limited to population diversity studies. Common PHR SNPs among the four populations could serve as a basis for the design of a lower density SNP chip specific for Philippine buffalo populations. Based on PCA and heatmap of relationship, the GRM derived from the PHR SNPs clearly separated the riverine and swamp populations and showed the genetic relationship among the riverine breeds.

Funding acknowledgements
This research is funded by the Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development- Department of Science and Technology (PCAARRD-DOST).

References

Affymetrix Axiom Buffalo Genotyping Array DataSheet.

Affymetrix Axiom Analysis Suite 1.1 User Guide 2015.

Borghese, A. (2005). Buffalo Production and Research. FAO Regional Office for Europe Inter-Regional Cooperative Research Network on Buffalo (ESCORENA). http://www.fao.org/3/a-ah847e.pdf

de Camargo GMF, Aspilcueta-Borquis RR, Fortes MRS, Porto-Neto R, Cardoso DF, Santos DJA, SA Lehnert, Reverter A, Moore and Tonhati. H. Prospecting major genes in dairy buffaloes BMC Genomics DOI 10.1186/s12864-015-1986-22015.

El-Halawany NK, Abdel-Shafy H, El-Monsif A. , Shawky AE and Al-Tohamy AFM (2015). A genome-wide association study for daily milk production in Egyptian Buffalo (2015). ISAFG 2015 Conference. Piacenza, Italy. 27th-29th July 2015.

Iamartino, D; de Oliveira, DAA; Coletta, A; Garcia, JF; Ali, A; Ramunno, L; Pasquariello, R; Drummond, MG; Bastianetto, E; Fritz, E; Knoltes, J; Williams, JL; Sonstegard, T; Reecy, J; Van Tassell, C; Nicolazzi, EL; Biffani, S; Biscarini, F; Schroeder, S (2013).The Buffalo Genome and the Application of Genomics in Animal Management and Improvement Buffalo Bulletin 2013 Vol.32 (Special Issue 1): 151-158.

Nicolazzi EL, VanTassell CP, Iamartino D, Reecy JM, FritzWaters E, Sonstegard TS, Koltes JE, Schroeder SG, Ahmad A, Garcia JF, Ramunno L, Cosenza G, Williams J, and the International Buffalo Consortium (2015) Using the 90k Buffalo SNP array. Plant and Animal Genome, San Diego CA, USA, 10-14th January 2015.

R: A Language and Environment for Statistical Computing. http://www.Rproject.org.