

GENETIC DIFFERENTIATION OF SWAMP BUFFALOES IN CALAYAN AND TAWI-TAWI USING 90K SNP GENOTYPING PANEL FOR INCLUSIVE BREEDING CONSERVATION

Pauleen S. Pineda^{1*}, Maureen B. Gajeton¹, Sherwin D. Matias¹, and Ester B. Flores¹

ABSTRACT

Water buffaloes are a significant livestock resource that helps rural farmers by providing draft power, milk, meat, and hides. However, crossbreeding between the two water buffalo types—swamp and river—has raised concerns about preserving the genetic identity of native swamp populations. In this study, 58 water buffaloes from an institutional swamp herd, Tawi-Tawi, Calayan Islands, and crossbreds were genotyped using a 90K single-nucleotide polymorphism (SNP) buffalo genotyping panel. A total of 54,481 polymorphic SNPs were used after filtering. The mean fixation index (F_{ST}) is 0.37, indicating substantial genetic differentiation between swamp populations. A high-impact SNP predicted to disrupt gene function was identified in *ADCY2*, a gene previously linked to meat quality traits. Additionally, 148 SNPs were shared exclusively among swamp buffalo groups and absent in crossbreds, highlighting potential markers specific to swamp buffalo that may help in breed traceability. Principal Component Analysis (PCA) revealed distinct clustering among swamp populations, particularly between those from Calayan and Tawi-Tawi. However, the panel's applicability to swamp buffaloes is limited, as only 36% SNPs had a minor allele frequency (MAF) > 0.05 . Nevertheless, the SNP data effectively revealed population structure among swamp and crossbred buffaloes, offering valuable insights for conservation and breeding strategies.

Keywords: animal breeding, genetic conservation, genotyping, SNPs, variations

INTRODUCTION

Water buffaloes, or commonly known as carabaos in the Philippines, are a significant livestock in the Philippines, providing farmers with draft power and additional income through meat, milk, and hide. There are two types of water buffaloes: the swamp type, which was found natively in the Philippines, and the river type that was imported to the Philippines to enhance milk production. (Pineda *et al.*, 2021). To improve the productivity of water buffaloes, crossbreeding was implemented by initially crossing swamp and river types, followed by backcrossing with swamp buffaloes to improve growth traits, and with river buffaloes to boost milk production and adaptability to the Philippine climate (PCC, *n.d.*). Backcrossing is done to maintain its reproductive ability, as a hybrid swamp and river is reported to have reduced fertility. (Pineda *et al.*, 2024). While genetic improvement

¹Philippine Carabao Center National Headquarters and Genepool, Science City of Muñoz, Nueva Ecija Philippines 3120 (*email: polenpineda@gmail.com)

programs are ongoing for water buffaloes in the Philippines, it is also important to identify and characterize genetic clusters within swamp buffalo local populations to maintain the conservation of their native genetic diversity.

One of the valuable approaches to assess the genetic differentiation of an animal is the use of SNP genotyping panel, which screens the animals at the genomic level and is cheaper than whole-genome sequencing (Crooijmans *et al.*, 2025; Berry and Spangler, 2023; Bertola *et al.*, 2022). For instance, the analysis of 333 water buffalo samples from 31 populations using a 90K SNP genotyping panel revealed patterns of diversity, which also provided insights into their domestication and migration routes, as well as introgression of the two water buffalo types. These findings highlight the value of the SNP genotyping panel for monitoring introgression and conservation of the animal's genetic background (Colli *et al.*, 2018). Additionally, the buffalo genotyping array enabled analysis of the signatures of selection, genetic diversity assessment, and runs of homozygosity, elucidating the adaptation, dispersal, and intermixing of the different water buffalo types and breeds (Saravanan *et al.*, 2022; Thakor *et al.*, 2021; Macciotta *et al.*, 2021). Other SNP genotyping panel such as the use of the BovineHD genotyping panel with 777K SNPs, also provided insights into the differentiation of the two water buffalo types (Pérez-Pardal *et al.*, 2018). Therefore, employing a SNP genotyping panel enables an effective approach to assess the genetic diversity and understand adaptation, migration, and evolutionary history of water buffalo populations, providing genomic information for effective management and conservation of the water buffalo populations.

The Philippine swamp buffaloes are considered a homogeneous breed locally and are genetically closely related to those in Taiwan and China, but also exhibit greater genetic diversity than buffalo populations in Taiwan, Myanmar, and Malaysia (Paraguas *et al.*, 2018; Villamor *et al.*, 2024). The study by Villamor *et al.* (2024) also identified distinct genetic clusters among swamp buffaloes in the Philippines, challenging the assumption that they form a homogeneous breed. However, samples were not collected from Tawi-Tawi, the southernmost island of the Philippines. Samples were obtained from Calayan, the northernmost island, where swamp buffaloes were found to be morphologically larger and have been recommended for ex-situ conservation (Villamor *et al.*, 2024).

To enhance productivity and upgrade swamp buffaloes in the Philippines without compromising their distinct population genetics, it is essential to identify each genetic cluster and establish appropriate conservation sites. Therefore, this study aims to identify the genetic clusters of swamp buffaloes from Calayan, Tawi-Tawi, and the Institutional Herd, and to assess the effectiveness of the water buffalo genotyping panel in detecting crossbred individuals. Additionally, we intend to determine candidate genetic markers that can support selective breeding strategies to improve productivity and contribute to the sustainable management, conservation, and genetic enhancement of the Philippine swamp buffaloes.

MATERIALS AND METHODS

Sample collection and DNA extraction

Approximately 4 mL of whole blood was collected from the jugular vein of water buffaloes using EDTA Vacutainer tubes. For swamp buffaloes owned by smallholders, animals were guided to lift their heads since they are generally tamed, while institutional animals

were restrained in an animal chute. There were a total of 58 water buffaloes from 4 groups comprising 14 crossbreds (swamp x river buffaloes) and native swamp buffaloes – 10 from Calayan, 7 from Tawi-Tawi, and 27 from the Philippine Carabao Center Institutional Herd in Cagayan (PCC at CSU). The sampling included animals from diverse regions: Calayan in the northernmost and Tawi-Tawi in the southernmost part of the Philippines, crossbreds, and native swamp in the gene pool to capture potential genetic differences. Genomic DNA was extracted from 1 mL of whole blood using Promega Wizard Kit following the optimized protocol of the PCC Molecular Genetics Laboratory.

Genotyping and SNP quality filtering

The samples were submitted to Affymetrix for genotyping using the Axiom 90k Buffalo Genotyping Array (Iamartino *et al.*, 2017). The raw data files were cleaned and converted to a vcf file using Axiom Analysis Suite with default settings. Further filtering of the SNPs was done using PLINK (v1.90) (Purcell *et al.*, 2007) with the parameters: –cow –nonfounders –allow-no-sex –autosome –geno 0.1 –mind 0.1 –maf 0.05. The number of SNPs and heterozygosity were computed using BCFtools stats (v1.17) (Danecek *et al.*, 2021).

Genetic Differentiation Analysis and SNPs Annotation

To identify the genetic clustering of the four groups, principal component analysis (PCA) was performed using PLINK (v1.90) (Purcell *et al.*, 2007). The SNPs were pruned based on linkage disequilibrium with the parameter ‘–indep 50 5 2’ and the PCA analysis were plot in R Studio. FST score was computed using VCFtools (v0.1.17) using ‘–weir-fst-pop’ with four populations. Minor allele frequency of the SNPs was identified using a custom Rscript. The functional effect of the SNPs was predicted and annotated using Variant Effect Predictor (VEP) (v113.0) (McLaren *et al.*, 2016) using NCBI water buffalo reference annotation from UOA_WB_1 (Low *et al.*, 2019), which is the same reference genome used to design the SNP panel. VEP will categorize SNPs based on their predicted impact on gene functions as high, moderate, low, and modifier. High-impact variants can disrupt protein functions, moderate-impact variants cause amino acid changes but are less likely to disrupt the gene functions, low-impact variants are typically synonymous mutations, and modifiers are variants occurring in non-coding regions. Analysis of breed/population-specific SNPs was identified through a custom R script.

RESULTS AND DISCUSSION

A total of 54,481 high-quality single-nucleotide polymorphisms (SNPs) were identified across the four water buffalo populations (Calayan, Tawi-Tawi, PCC at CSU, and crossbreds) following quality control and filtering. These SNPs effectively distinguish crossbreds from pure swamp-type buffaloes, with crossbreds exhibiting a heterozygosity level of 45%, compared to only 11% in the pure swamp population. However, as the genotyping panel was designed using a river buffalo population, there is limited applicability for swamp-type buffaloes (Iamartino *et al.*, 2017). A total of 35,092 of the SNPs in pure swamp types were below 0.05 minor allele frequency (MAF), indicating that most of the SNPs in the genotyping panel were not polymorphic. While crossbreds showed evenly distributed 49,112 SNPs with MAF > 0.2 (Figure 1a). The low MAF in swamp-type buffaloes and the

more uniform MAF in river-types result in an overall balanced MAF distribution across all samples. The low MAF for the swamp type is unsurprising and agrees with the previous studies evaluating the utility of the genotyping panel for swamp type buffaloes (Colli *et al.*, 2018; Herrera *et al.*, 2016; Pineda *et al.*, 2024).

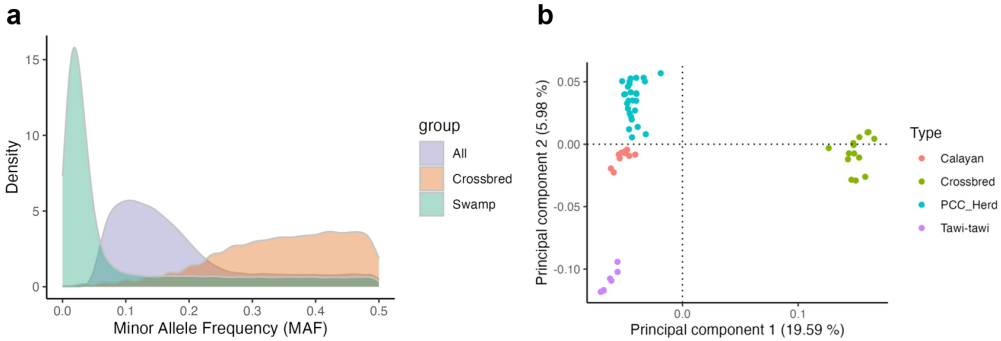


Figure 1. (a) Density plot showing the minor allele frequency (MAF) distribution of SNPs in pure swamp buffaloes, crossbred buffaloes, and their combined dataset. (b) Principal Component Analysis (PCA) of the four water buffalo populations showing distinct clustering patterns, indicating genetic differentiation among the populations.

Nevertheless, clear genetic clustering is observed among the populations, with the Tawi-Tawi and crossbred groups forming distinct clusters separate from the Calayan and PCC at CSU populations (Figure 1b). In contrast, Calayan and PCC at CSU swamp buffaloes cluster closely together, indicating a higher degree of genetic similarity, which is in agreement with previous studies (Paraguas *et al.*, 2018). The geographic location of the Tawi-Tawi and Calayan populations may explain the observed genetic differentiation. Tawi-Tawi is in the southernmost part of the Philippines, while Calayan is in the northernmost part and is closer to PCC at CSU. This geographical separation may explain the distinct genetic clustering between Tawi-Tawi and Calayan, and the possible shared genetic ancestry between Calayan and PCC at CSU populations. The mean fixation index (F_{ST}) of 0.37 and the weighted F_{ST} of 0.38 indicates a high level of genetic differentiation among the swamp populations.

Analysis of the SNPs revealed that about 10,682 (20%) were polymorphic across all populations, while about 24,609 (45%) were polymorphic only in crossbred buffaloes – almost half of the filtered SNPs from the genotyping panel (Figure 2). This result is also reflected in their MAF distribution. Whereas, 212 SNPs were identified across the pure swamp buffalo populations, with 148 of them shared among them, which can be utilized to further identify crossbreds from pure swamp type. However, since the genotyping panel was designed based on river buffaloes, its utility to accurately assess the genetic purity of swamp buffaloes may be limited. Still, these SNPs may be valuable genetic information for characterizing swamp buffalo populations.

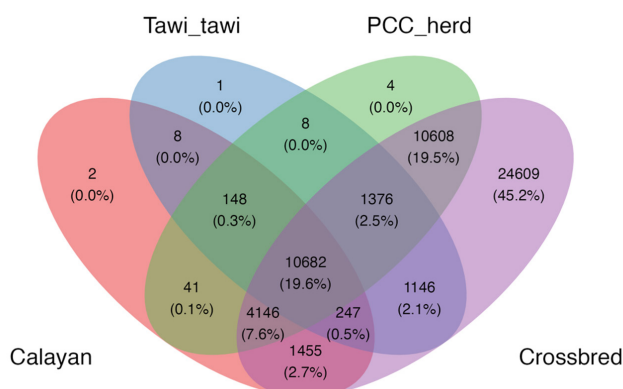


Figure 2. Venn diagram showing SNPs that are unique or shared among the different populations. This showed the distribution of polymorphic variants across the populations.

Annotation of the SNPs on their predicted effects on gene function based on their impact on transcription revealed three variants classified as “high-impact”. Two of the variants had stop-gained consequences, leading to premature termination of the transcripts in *PGPEP1* and *FAM228A* genes, while one affected a splice acceptor site affecting the *ADCY2* gene. From the genes identified, *ADCY2* is involved in many cellular processes, specifically acting as a catalyst in the formation of cAMP (Gray *et al.*, 2024). This gene has been associated with meat tenderness in cattle (Li *et al.*, 2014), reproductive traits (Shao *et al.*, 2021), and was also associated with milk and fat yield, and milking temperament or the level of comfort they experience during the milking process (Laodim *et al.*, 2019; Mincu-Iorga *et al.*, 2025). Given the association of this gene with economically important traits in bovids, and its predicted functional effect in water buffaloes, the identified variant in this gene can serve as a potential marker for genetic improvement programs. The other two genes with high-impact on gene function, the *PGPEP1* was found to may have a regulatory role in gene networks of compensatory growth, or the accelerated growth response from nutritional restriction in cattle (Keogh *et al.*, 2019). Additionally, the *PGPEP1* was identified to be a site of insertion for bovine leukemia virus (BLV), commonly observed in cattle with persistent lymphocytosis (Babii *et al.*, 2022). The *FAM228A* was found to have an association with carcass and body traits in cattle (Liu *et al.*, 2021; Rafter *et al.*, 2021). Although there is no available information for buffaloes in these three genes, this may also occur in buffalo due to their close genetic similarity to cattle. The identified high-impact genes are all associated with meat-related traits, highlighting the potential of these markers for swamp buffaloes in enhancing meat production.

CONCLUSION

The findings in this study revealed the genetic differences of four water buffalo populations in the Philippines, specifically the clustering of Tawi-Tawi water buffaloes distinct from swamp buffaloes kept in the institutional herd. We also identified markers associated with production traits. While functional studies in buffaloes are still limited, there

is strong evidence from cattle suggesting the variants identified are candidate markers for selective breeding strategies and improving the productivity and sustainability of swamp buffaloes while maintaining their native genetics in the Philippines.

REFERENCES

- Babii AV, Arkhipova AL, and Kovalchuk SN. 2022. Identification of novel integration sites for bovine leukemia virus proviral DNA in cancer driver genes in cattle with persistent lymphocytosis. *Virus Res* 317:198813.
- Berry DP and Spangler ML. 2023. Animal board invited review: Practical applications of genomic information in livestock. *Animal* 17:100996.
- Bertola LD, Vermaat M, Lesilau F, Chege M, Tumenta PN, Sogbohossou EA, Schaap OD, Bauer H, Patterson BD, White PA, de longh HH, Laros JFJ and Vrieling K. 2022. Whole genome sequencing and the application of a SNP panel reveal primary evolutionary lineages and genomic variation in the lion (*Panthera leo*). *BMC Genom* 23:231.
- Colli L, Milanesi M, Vajana E, Iamartino D, Bomba L, Puglisi F, Del Corvo M, Nicolazzi EL, Ahmed SSE, Herrera JRV, Cruz L, Zhang S, Liang A, Hua G, Yang L, Hao X, Zuo F, Lai SJ, Wang S, Liu R, Gong Y, Mokhber M, Mao Y, Guan F, Vlaic A, Vlaic B, Ramunno L, Cosenza G, Ahmad A, Soysal I, Ünal EÖ, Ketudat-Cairns M, Garcia JF, Utsunomiya YT, Baruselli PS, Amaral MEJ, Parnpai R, Drummond MG, Galbusera P, Burton J, Hoal E, Yusnizar Y, Sumantri C, Moioli B, Valentini A, Stella A, Williams JL, and Ajmone-Marsan P. 2018. New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. *Front Genet* 9:53.
- Crooijmans RPMA, Gonzalez Prendes R, Colli L, Del Corvo M, Barbato M, Somenzi E, Tosser-Klopp G, Meszaros G, Ajmone-Marsan P, Weigend S, Wallner B, McCue ME, Orlando L, Bradley D, Hiemstra SJ, Schokker D, Peynot N, Stella A, Restoux G, Groenen MAM, and Tixier-Boichard M. 2025. IMAGE001: A new livestock multispecies SNP array to characterize genomic variation in European livestock gene bank collections. *Anim Genet* 56(5):e70039.
- Danecek P, Bonfield, JK., Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, and Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10(2):giab008.
- Gray M, Nash KR, and Yao Y. 2024. Adenylyl cyclase 2 expression and function in neurological diseases. *CNS Neurosci Ther* 30(7): e14880.
- Herrera JR, Flores E, Gondro C, and Van Der Werf J. 2016. Performance of the Axiom 90k buffalo genotyping array in four Philippine water buffalo populations. *Ces Med Vet Zootec* 11(3): 210-210.
- Iamartino D, Nicolazzi EL, Van Tassell CP, Reecy JM, Fritz-Waters ER, Koltjes JE, Biffani S, Sonstegard TS, Schroeder SG, Ajmone-Marsan P, Negrini R, Pasquariello R, Ramelli P, Coletta A, Garcia JF, Ali A, Ramunno L, Cosenza G, de Oliveira DAA, Drummond MG, Bastianetto E, Davassi A, Pirani A, Brew F, and Williams JL. 2017. Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). *PLoS One* 12(10): e0185220.
- Keogh K, Kenny DA, and Waters SM. 2019. Gene co-expression networks contributing to

- the expression of compensatory growth in metabolically active tissues in cattle. *Sci Rep* 9(1):6093.
- Laodim T, Elzo MA, Koonawootrittriron S, Suwanasopee T, and Jattawa D. 2019. Pathway enrichment and protein interaction network analysis for milk yield, fat yield and age at first calving in a Thai multibreed dairy population. *Asian-Australas J Anim Sci* 32(4):508-518.
- Li YX, Jin HG, Yan CG, Ren CY, Jiang CJ, Jin CD, Seo KS, and Jin X. 2014. Molecular cloning, sequence identification, and gene expression analysis of bovine *ADCY2* gene. *Mol Bio Rep* 41(6):3561-3568.
- Liu D, Chen Z, Zhao W, Guo L, Sun H, Zhu K, Liu G, Shen X, Zhao X, Wang Q, Ma P, and Pan Y. 2021. Genome-wide selection signatures detection in Shanghai Holstein cattle population identified genes related to adaption, health and reproduction traits. *BMC Genom* 22(1):747.
- Low WY, Tearle R, Bickhart DM, Rosen BD, Kingan SB, Swale T, Thibaud-Nissen F, Murphy TD, Young R, Lefevre L, Hume DA, Collins A, Ajmone-Marsan P, Smith TPL, and Williams JL. 2019. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat Commun* 10(1):260.
- Macciotta NPP, Colli L, Cesarani A, Ajmone-Marsan P, Low WY, Tearle R, and Williams JL. 2021. The distribution of runs of homozygosity in the genome of river and swamp buffaloes reveals a history of adaptation, migration and crossbred events. *Genet Sel Evol* 53(20).
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, and Cunningham F. 2016. The ensembl variant effect predictor. *Genome Biol* 17(1):122.
- Mincu-Iorga M, Mizeranschi AE, Gavojdian D, Nicolae I, Kusza S, and Ilie DE. 2025. Detection of genetic variants associated with behavioural response during milking in simmental dual-purpose cows. *Animals* 15(12):1766.
- Paraguas AM, Cailipan TPC, Flores EB, and Villamor LP. 2018. Morphology and phylogeny of swamp buffaloes (*Bubalus bubalis*) in Calayan Island, Cagayan. *Philipp J Vet Anim Sci*, 44(1):59-67.
- Pérez-Pardal L, Chen S, Costa V, Liu X, Carvalheira J, and Beja-Pereira A. 2018. Genomic differentiation between swamp and river buffalo using a cattle high-density single nucleotide polymorphisms panel. *Animal* 12(3):464-471.
- Philippine Carabao Center (PCC). n.d. *PCC improving native carabao breed for greater productivity*. Retrieved on July 29, 2025 from <https://www.pcc.gov.ph/pcc-improving-native-carabao-breed-for-greater-productivity/>.
- Pineda PS, Flores EB, Herrera JRV, and Low WY. 2021. Opportunities and challenges for improving the productivity of swamp buffaloes in Southeastern Asia. *Front Genet* 22:12:629861.
- Pineda PS, Flores EB, Villamor LP, Parac CJM, Khatkar MS, Thu HT, Smith TPL, Rosen BD, Ajmone-Marsan P, Colli L, and Williams JL. 2024. Disentangling river and swamp buffalo genetic diversity: initial insights from the 1000 Buffalo Genomes Project. *Gigascience* 13:giae053.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, and Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559-575.

- Rafter P, Gormley IC, Purfield D, Parnell AC, Naderi S, and Berry DP. 2021. Genome-wide association analyses of carcass traits using copy number variants and raw intensity values of single nucleotide polymorphisms in cattle. *BMC Genom* 22(1):757.
- Saravanan KA, Rajawat D, Kumar H, Nayak SS, Bhushan B, Dutt T and Panigrahi M. 2023. Signatures of selection in riverine buffalo populations revealed by genome-wide SNP data. *Anim Biotechnol* 34(8):3343–3354.
- Shao B, Sun H, Ahmad MJ, Ghanem N, Abdel-Shafy H, Du C, Deng T, Mansoor S, Zhou Y, Yang Y, Zhang S, Yang L, and Hua G. 2021. Genetic features of reproductive traits in bovine and buffalo: lessons from bovine to buffalo. *Front Genet* 12:617128.
- Thakor PB, Hinsu AT, Bhatia DR, Shah TM, Nayee N, Sudhakar A, Rank NR and Joshi CG. 2021. High-throughput genotype-based population structure analysis of selected buffalo breed. *Trans Anim Sci* 5(2):txab033.
- Villamor LP, Cuanang AJE, Cailipan TPC, Paraguas AM, Flores EB, Villanueva MA, Balbin MM, Belotindos LP, and Rellin FT. 2024. Population structure, morphology, and health assessment of Philippine swamp buffalo (*Bubalus kerabau*, Fitzinger, 1860) in Calayan Island, Cagayan. *Ruminants* 4(4):556-564.