

## Genetic Properties of Copy Number Variations in Some Pakistani Cattle Breeds

Hamid MUSTAFA<sup>1,3</sup>  Waqas Ahmad KHAN<sup>2</sup> Kim EUISOO<sup>3</sup> Zulfiqar Hussain KUTHU<sup>4</sup>  
Adeela AJMAL<sup>1</sup> Afzal ALI<sup>1</sup> Khalid JAVED<sup>1</sup> Tad S. SONSTEGARD<sup>3</sup>

<sup>1</sup> Department of Livestock Production, Faculty of Animal Production and Technology, University of Veterinary and Animal Science, Lahore - PAKISTAN

<sup>2</sup> Department of Biotechnology, Faculty of Science, University of Sargodha, Sargodha, Pakistan.

<sup>3</sup> Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland 20705, USA

<sup>4</sup> Faculty of Veterinary and Animal Sciences, University of Poonch Rawalakot, Azad Kashmir - PAKISTAN

Article Code: KVFD-2018-19282 Received: 02.01.2018 Accepted: 19.05.2018 Published Online: 22.05.2018

### How to Cite This Article

Mustafa H, Khan WA, Euisoo K, Kuthu ZH, Ajmal A, Ali A, Javed K, Sonstegard TS: Genetic properties of copy number variations in some Pakistani cattle breeds. *Kafkas Univ Vet Fak Derg*, 24 (4): 599-602, 2018. DOI: 10.9775/kvfd.2018.19282

### Abstract

A copy number variation (CNV) is information of DNA segment containing deletion, duplication or insertion exist within diverse populations. This study detected 53 CNVs regions that overlapped with immune response, body size and parasitic resistance traits in Pakistani cattle breeds. This study characterized genetic diversity and provides lineage differentiated CNVs in these indigenous cattle breeds.

**Keywords:** Copy number variation, Diversity, Traits, Breed, Pakistan

## Pakistan'da Bazı Sığır Irklarında Kopya Sayısı Değişikliklerinin Genetik Özellikleri

### Öz

Kopya sayısı değişikliği (CNV) farklı popülasyonlar içerisinde delesyon, duplikasyon ve insersiyonları içeren DNA segmentlerinin bilgisidir. Bu çalışma ile Pakistan sığır ırkları arasında bağışıklık yanıtı, vücut boyutu ve parazit direnci özelliklerini içeren 53 CNV tespit edildi. Bu çalışma; belirtilen yerel sığır ırkları arasında genetik çeşitliliği göstermiş ve hat farklı CNV'leri belirlemiştir.

**Anahtar sözcükler:** Kopya sayısı değişikliği, Çeşitlilik, Özellikler, Irk, Pakistan

## INTRODUCTION

Copy number variations (CNV) are DNA segments containing deletions, insertions and duplications existing as complex allelic variants within the diverse population<sup>[1]</sup>. CNVs markers are more effective as compared to SNPs markers, due to more genomic sequences, including gene structure changing and dosage, changing gene regulation, and recessive alleles exposing<sup>[2]</sup>. Recently, a number of studies confirmed CNVs regions in bovine genome that cover a number of genes related to environment adaptation<sup>[3]</sup>. CNVs regions have tendency to demonstrate breed's history and their formation patterns<sup>[4]</sup>. There is very little information about population genetic properties of CNVs in cattle, unlike microsatellites and SNPs markers<sup>[1]</sup>. The

characterization of diversity and origin of CNVs, population genetic properties and functional impacts of CNVs is still a very dynamic research area in animal sciences. This study reported first comprehensive CNVs regions based on population genetic properties, focusing diversity pattern and population structure, within indigenous Pakistani cattle breeds.

## MATERIAL and METHODS

Selected animal's blood sampling and DNA extraction is already described in a previously study<sup>[5]</sup>. Additional, Angus (n=10) and Holstein (n=10) samples were added for this analysis using the USDA available dataset. Log R Ratio file was imported into SNP and variation suite (SVS



### İletişim (Correspondence)



+92 334 9931006



hamidmustafapasha@gmail.com

8.0) and 440 370 SNPs were successfully mapped onto UMD 3.1 genome assembly. Default GC correlation file was used to normalize Log R Ratio and copy number module (CNM) to segment pairwise chromosomes permutations (n=1000). Three state covariates (0.3) was used across the all selected samples for CNVs genotyping as three levels (loss, neutral and gain). All 154 deletion CNVs recoded manually, converting loss event into "11" and neutral into "12". Multidimensional scaling (MDS) analysis was used to identify individual's clustering. PLINK (9.0 version) was used to determine pairwise relationship of 154 CNVs deletion. The cluster analysis was also performed to verify separation based on the mean Log R Ratio values using R version 13.1. We recovered ENSEMBL and REFRSEQ genes overlapping CNVs regions by at least 1 bp including the 5' and 3' untranslated regions from the available UCSC genome browser and annotated CNVs regions using custom software. Enrichment analysis was performed using PANTHER classification system and only clusters with enrichment scores more than 1 (P value <0.05) were considered. VST was calculated for each CNV by using the following equation:  $(VT-VS)/VT$ , to detect lineage differentiated CNVs. where, VT is the total variance in mean Log R Ratio across all individuals and VS is the average variance in cattle within each breed.

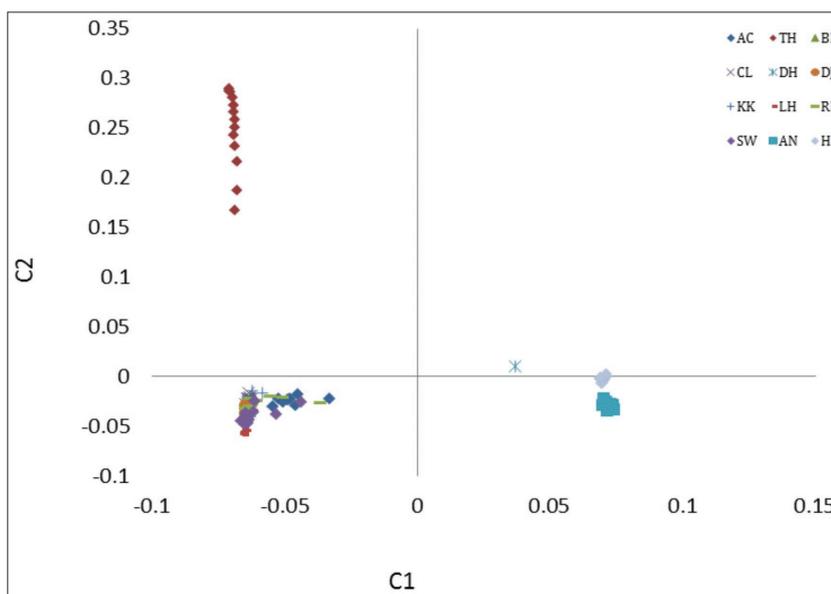
## RESULTS

A total 156 individuals from twelve different breeds were used for CNVs investigation. Total 100,500 CNVs segments were extracted using the default multivariate option of Golden Helix SVS (version 8.0). Above 1% segment frequency were retained for further analysis, where multi-variance option in SVS was actually developed to detect

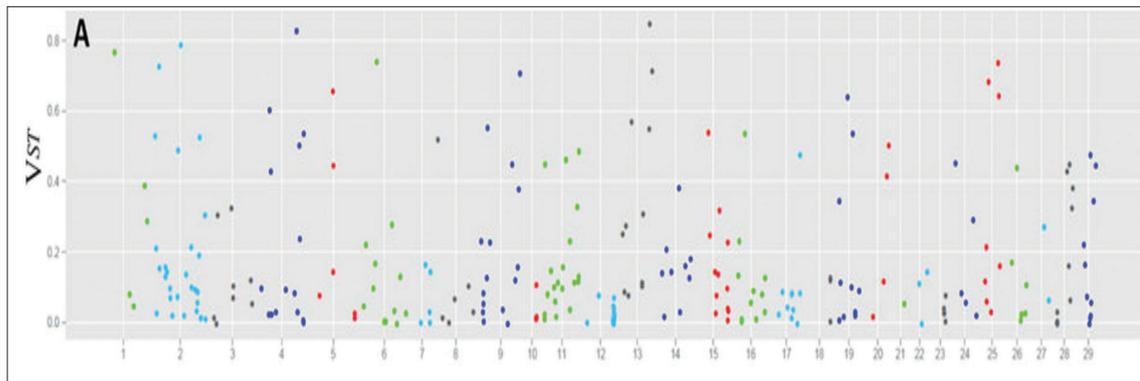
moderate to high CNVs frequency. After filtration total 210 CNVs (total length of 10221 kb) with an average length of 45.2 kb were retained and samples were categorized three type events (loss, neutral and gain). The mean LR ratio  $\pm 0.3$  threshold level was used as three state models, there was 154 CNVs deletion and stored as CNV1 to CNV210 list with descending frequency. However, SVS results of this study indicate low concordant. This may be due the absence of these breeds during the bovine high density SNP bead chip designing or due to low sample size and breed undistinguished in this data set. Two hundred and ten CNVs of this data set were used for VST and frequency calculations.

Multidimensional Scaling (MDS) analysis was performed based on CNVs to investigate the population structure based on 154 CNVs deletion. Three clusters were found, which clearly separate these breeds into three groups. The two distinct groups B (All indigenous Pakistan breeds except Tharparkar) and C (Angus and Holstein) both need more clearance of diversion from other breeds (Fig. 1). The CNVs results based on MDS analysis are reliable with similar results based on SNPs analysis (The Bovine HapMap Consortium, 2009) and suggest that CNVs can be used as genetic markers to separate cattle individuals into distinct groups.

The results of this study revealed that CNVs can also be used effectively as genetic markers in population genetic studies. However, CNVs markers as compared with SNPs have some little resolution issues, may be due to small size sampling. Population diversity and differentiated CNVs were estimated using 210 CNVs. CNVs frequencies were used across all 3 groups (A, B & C). Top four high frequency deletions were on chromosomes 9, 21, 1 and 12 that are



**Fig 1.** MSD analysis based on 154 deletion CNVs and plotted according to their first two components



**Fig 2.** Population differentiations estimated by Vst plotted along each chromosome for all breed comparison (All Pakistan cattle vs Holstein and Angus)

0.90, 0.89, 0.83 and 0.81, respectively and were used to estimate CNVs with differentiated frequencies.

Comparing CNVs regions with the UMD 3.1, UCSC annotation and found 82 CNVs partially overlapping with genes. Gene ontology (GO) was also performed to detect genes enrichments. Response to interferon gamma and response to stimulus were the most enriched biological process. One hundred seventy two CNVs were not encompassed any UCSC genes, but only 38 CNVs overlapped. CNVs without genes were shorter (mean  $22170 \pm 5251$ ,  $n=172$  vs.  $77200 \pm 0.0182$ ,  $n=38$ , t-test,  $P < 0.0001$ ) and at higher frequency ( $0.3234 \pm 0.0173$ ,  $n=172$  vs.  $0.3121 \pm 0.0182$ ,  $n=38$ ,  $P=0.0112$ ), when compared CNVs overlapped with genes and not overlapped with genes. The lack of common deletion CNVs overlapping with genes is positive indicator that population is under selection and deleterious variants removes from the population and already.

Pairwise VST (0-1) was investigated, to test CNVs association with population specific selection (Fig. 2). High VST values indicate regions under increased selection pressure or evolutionary forces (bottlenecks or founder effects). A total of 12, 1 and 14 CNVs were identified at  $VST > 0.6$  threshold and at 0.4 CNVs were found 40, 09 and 41, respectively. The higher VST differential detected in these CNVs regions may suggest the variability of their genomic sequence that could be further involved in phenotypic diversity across these cattle breeds. Seven genes (HIATL1, MRPL48, PRAME, PLCB1, TSPY, ROBO4, and ZNF280B), which overlapped with CNVs regions and potentially impact their functions and evolution.

## DISCUSSION

This study is first comprehensive attempt to explore CNVs as potential genetic markers based on population genetic properties in some indigenous cattle breeds in Pakistan. Previous studies in some species, including Human, Dog, Zebra fish, Stickleback fish and some worldwide cattle breeds have identified CNVs, based on population genetic properties, within and between population [6-8].

Confidence deletion CNVs was used because; it was difficult to detect accurately genotypes complex CNVs events, due to non-biallelic duplication [8]. The CNVs deletion events have some advantage over duplication, due to its bi-allelic nature and compatibility with genetic analysis procedures designed for SNPs markers [9]. In this study, CNVs was used to explore the population genetic properties in some indigenous cattle breeds in Pakistan with moderate intra population frequencies as reported by Bickhart et al.[6]. SVS method, although it reported limited amount of CNVs calls (71.6% deletion and 28.4 duplications). The results of this study revealed that different cattle breeds in different agro-ecological zones clustered roughly and two indicus Lohani and Tharparkar were genetically distinct from other selected cattle populations (Fig. 1) which is similar as reported by previously SNPs based studies in different cattle population worldwide, including taurine, indicus or sanga etc.[7]. Seventy two percent of the CNVs were identified, which were marked by flanking SNPs as reported by Xu et al.[8]. However, the results of this investigation were not efficiently capable to distinguish recent divergent populations of common geographic origin, such as Cholistani and Bhagnari cattle breeds and also reported by Silva et al.[7] and Xu et al.[8] in Holstein and Angus cattle breeds. Human published data also revealed that the CNVs based on population genetic properties can only be investigated with large data set of continent population level [10]. The satisfactory level of population structure identified by SNPs based methods may depends on genotyping accuracy and sample size [11]. Therefore, it is highly recommended that studies based on high throughput SNPs technology will defiantly enhance to make accurately genotype CNVs [12].

Furthermore, estimated CNVs based population differentiation and detected some potential candidate CNVs in these breeds. VST values were calculated and scanned gene enrichments among CNVs regions and found 54 distinctive CNVs with threshold VST value 0.4 as reported in a previous study by Liu et al.[13]. It is

important here that selection pressure, the amount divergence between breeds (divergence time, effective population size and gene flow) can affect VST value and overall differentiation among selected populations<sup>[13,14]</sup>. At this stage a hypotheses can be developed that high VST value between populations does not necessarily involve in divergent (selection for different alleles in both populations) or can occur in the absenteeism of selection by bottlenecks or founder effects by drift<sup>[9]</sup>. All these hypotheses needs further research using large population samples size of these breeds as similar observed by Xu et al.<sup>[8]</sup>.

In this study, only a small amount of lineage differentiated CNVs were detected in all breeds comparison especially Cholistani and Bhagnari and similar was reported in Holstein and Angus Xu et al.<sup>[8]</sup>. These results suggest that these breeds might not have had sufficient time to diversify their CNVs deletion contents and assume that CNVs occurred before split of these breeds and deletion content should eventually fix in these breeds except Tharparkar. However, there are also some evidence of fewer lineage differentiated CNVs in laboratory mouse, some zebra fish strains and cattle breeds which compared with their wild populations<sup>[6]</sup>. This may suggest CNVs differently fixed as large effective population size or indicate inbreeding effects of wild populations<sup>[9]</sup>. There is still very little information about genetic landscape and exact domestication events of these breeds.

In conclusion, CNVs as genetic markers in indigenous cattle breeds in Pakistan raised some important questions about the domestication and divergence of these cattle breeds especially Tharparkar cattle. However, this study provided substantial evidence to support CNVs as potential genetic markers that can be used across population diversity and capture subspecies relationships. Additionally, this study suggest that a comprehensive study of population genetic differentiation in these cattle breeds along with other worldwide cattle breeds will provide a clear picture of diversity pattern and population structure. Finally, some more powerful analytical software tools can maximize the prediction level of population genetics in livestock species and expand the scope with next generation sequencing (NGS) data.

#### ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of Higher Education Commission of Pakistan and Bovine Functional Genomics Laboratory, Agricultural Research Services/United States Department of Agriculture, USA for laboratory facilities. We also appreciate the efforts of Professor Dr. Talat Naseer Pasha for his critical review and suggestions.

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest

#### REFERENCES

- Keel BN, Lindholm-Perry AK, Snelling WM:** Evolutionary and functional features of copy number variation in the cattle genome. *Front Genet*, 7, 207, 2016. DOI: 10.3389/fgene.2016.00207
- Sasaki S, Watanabe T, Nishimura S, Sugimoto Y:** Genome-wide identification of copy number variation using high-density single-nucleotide polymorphism array in Japanese Black cattle. *BMC Genet*, 17, 26, 2016. DOI: 10.1186/s12863-016-0335-z
- Zhang L, Jia S, Plath M, Huang Y, Li C, Lei C, Zhao X, Chen H:** Impact of parental *Bos taurus* and *Bos indicus* origins on copy number variation in traditional Chinese cattle breeds. *Genome Biol Evol*, 7, 2352-2361, 2015. DOI: 10.1093/gbe/evv151
- Wu ZJ, Jin W:** Copy-number variation: A new pattern of structural diversity in genome. *Yi Chuan*, 31, 339-347, 2009.
- Mustafa H, Euisoo K, Heather HJ, Ajmal A, Riley D, Pasha TN, Ali A, Javed K, Sonstegard TS:** Genome-wide SNPs analysis of indigenous zebu breeds in Pakistan. *Biotechnol Anim Husb*, 33, 13-25, 2017. DOI: 10.2298/BAH1701013M
- Bickhart DM, Xu L, Hutchison JL, Cole JB, Null DJ, Schroeder SG, Song J, Garcia JF, Sonstegard TS, Van Tassell CP, Schnabel RD, Taylor JF, Lewin HA, Liu GE:** Diversity and population-genetic properties of copy number variations and multi-copy genes in cattle. *DNA Res*, 23, 253-262, 2016. DOI: 10.1093/dnares/dsw013
- da Silva VH, de Almeida Regitano LC, Geistlinger L, Pertille F, Giachetto PF, Brassaloti RA, Morosini NS, Zimmer R, Coutinho LL:** Genome-Wide detection of CNVs and their association with meat tenderness in Nelore cattle. *PLoS One*, 11, e0157711, 2016. DOI: 10.1371/journal.pone.0157711
- Xu L, Hou Y, Bickhart DM, Zhou Y, Hay el HA, Song J, Sonstegard TS, Van Tassell CP, Liu GE:** Population-genetic properties of differentiated copy number variations in cattle. *Sci Rep*, 6, 23161, 2016. DOI: 10.1038/srep23161
- Zhang X, Wang K, Wang L, Yang Y, Ni Z, Xie X, Shao X, Han J, Wan D, Qiu Q:** Genome-wide patterns of copy number variation in the Chinese yak genome. *BMC Genomics*, 17, 379, 2016. DOI: 10.1186/s12864-016-2702-6
- Zhang Q, Ma Y, Wang X, Zhang Y, Zhao X:** Identification of copy number variations in Qinchuan cattle using BovineHD Genotyping Beadchip array. *Mol Genet Genomics*, 290, 319-327, 2015. DOI: 10.1007/s00438-014-0923-4
- Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, Ding X, Liu J, Zhang Q:** Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. *BMC Genomics*, 14, 131, 2013. DOI: 10.1186/1471-2164-14-131
- Salomon-Torres R, Gonzalez-Vizcarra VM, Medina-Basulto GE, Montano-Gomez MF, Mahadevan P, Yaurima-Basaldua VH, Villa-Angulo C, Villa-Angulo R:** Genome-wide identification of copy number variations in Holstein cattle from Baja California, Mexico, using high-density SNP genotyping arrays. *Genet Mol Res*, 14, 11848-11859, 2015. DOI: 10.4238/2015.October.2.18
- Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Licalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LC, Smith TP, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE, McDanel TG, Keele JW:** Analysis of copy number variations among diverse cattle breeds. *Genome Res*, 20, 693-703, 2010. DOI: 10.1101/gr.105403.110
- Öner Y, Yılmaz O, Okut H, Ata N, Yılmazbaş-Mecitoğlu G, Keskin A:** Associations between GH, PRL, STAT5A, OPN, PIT-1, LEP and FGF2 polymorphisms and fertility in Holstein-Friesian heifers. *Kafkas Univ Vet Fak Derg*, 23 (4): 527-534, 2017. DOI: 10.9775/kvfd.2016.17192