The Pharma Innovation Journal 2023; SP-12(9): 1800-1804

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; SP-12(9): 1800-1804

www.thepharmajournal.com
Received: 23-06-2023
Accepted: 28-07-2023

Sylvia Lalhmingmawii
Division of Animal Genetics and Breeding, National Dairy Research Institute, Karnal, Haryana, India

Zarzoliani
Department of Veterinary Pharmacology and Toxicology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Dharamshaw CA
Division of Animal Genetics and Breeding, National Dairy Research Institute, Karnal, Haryana, India

Corresponding Author:
Sylvia Lalhmingmawii
Division of Animal Genetics and Breeding, National Dairy Research Institute, Karnal, Haryana, India

Exploring copy number variation and its implication in livestock

Sylvia Lalhmingmawii, Zarzoliani and Dharamshaw CA

Abstract
Copy number variation (CNV) is a term used to describe structural variations involving DNA segments varying from 50bp to several mega base pairs due to loss by deletion or gain by duplication and insertional translocation. It has been demonstrated as a crucial substrate for diversity, evolution, adaptation and breed development. It is an important source of genetic variation in an individual and is now being utilized more than single nucleotide polymorphisms (SNPs), as it covers more genomic region. Several mechanisms such as non-allelic homologous recombination (NAHR), non-homologous end- joining (NEJH), fork stalling and template switching (FosTes) and L-1 mediated retro transposition are involved in its genesis. In livestock animals, numerous researches had been done to evaluate the impact of CNV and its implication. A wide range of variation in copy number in both intra and inter population for different species had been reported occupying nearly 10% of the genome. Moreover, CNV was also determined to involve in the adaptation of animals to different climatic conditions. Overall, copy number variation is a phenomenon capable of influencing the genome structure and deciphering them is a must to not miss the forest for the trees.

Keywords: Copy number, genetic variation, genome

1. Introduction
Genetic variations refer to variances between individuals at the genome level and are important since they define our individuality. In breeding programmes designed to enhance the genetic diversity of a population, these genetic differences are also the main focus. The development of karyotyping in the mids of the 20th century allowed for the identification of structural variants, which were later linked in humans to various disorders such as down syndrome, cri-du-chat etc. Over the time, SNPs were found through technological advancements and are now widely employed in GWAS and genomic prediction. Then gradually, large segment mutations were found and eventually, they became what is known as copy number variation. Thus, CNV was explained as a molecular phenomenon in which sequences of the genome are present at a variable number when compared to a reference genome [11, 27].

They are genomic variations that reveal differences between and within species [8]. The size varies from >1kb to several mega base pairs but in recent studies, variants as small as 50bps were detected in recent studies [4, 26, 29]. Despite the occurrence of large CNV, high resolution studies revealed that smaller CNV with a size of 50 kb or less are more typical. They are found in all chromosomes and dispersed throughout the genome in a non-random manner stating that they follow a heterogenous distribution. Compared to other areas of the genome, they are discovered to occur more frequently in G-C rich regions [11]. The majority of CNVs discovered were deletions, and they are known to constitute a significant class of genetic variations that exhibit extensive base pair losses and gains [38, 32]. These CNV's are becoming an important topic in the field of genomics as several CNV's were found to be linked with mendelian and complex genetic disorders [37, 7]. Gene dosage and structure can be impacted by significant genomic CNVs, which can then affect gene expression [10]. They are also found to play an essential role for animals' adaption to various environmental conditions [39] and helped to explain breed and inter-individual variances [30]. CNVs make up about 7% of the mouse and cattle genomes [25], which suggests that they can significantly affect an individual's genotype and phenotype. Moreover, they are found to increase the accuracy of genomic prediction by combining it with SNP data [15].

2. History

The first association of CNV with a phenotype reported in a non-human species was the bar eyes phenotype in Drosophila Melanogaster due to single duplication of bar gene [31]. In 2004, Iafrate *et al.* [16] and Sebat *et al.* [30] first reported a large scale polymorphism, calling it copy number polymorphism. Then in 2006, Redon *et al.* [28] introduced the term copy number variation and defined it as a DNA segment >1kb. So, research on CNV continues until 2011 when Alkan *et al.* [1] in their review, re-defined CNV as DNA segments <50bp to several mega base pairs and the definition has remained the same ever since.

3. Mechanism of formation

The genesis and processes of CNV formation are one of the key questions in understanding their biology. Sequence recombination and replication error are the two main processes that were reported. Non-Allelic Homologous [13].

Recombination (NAHR) and Non-Homologous End Joining (NHEJ) are examples of sequence recombination, while replication error includes fork stalling and template switching (FoSTeS) and L1-mediated retro transposition [14].

4. Non-Allelic Homologous Recombination (NAHR)

A type of homologous recombination that happens when two DNA segments with high sequence homology that are not alleles and produce more deletion than duplication interact [26]. Inter-chromosomal, Inter-chromatid, and intra-chromatid recombination are the three forms, and they can take place during both meiosis and mitosis [2]. In inter-chromatid and intra-chromosomal recombination, there is an increase in DNA segment at the expense of another resulting in duplication and deletion whereas Intra-chromatid, there is inversion of chromosome segment due to presence of homologous segments in the same chromatid, forming an inversion then deletion [13].

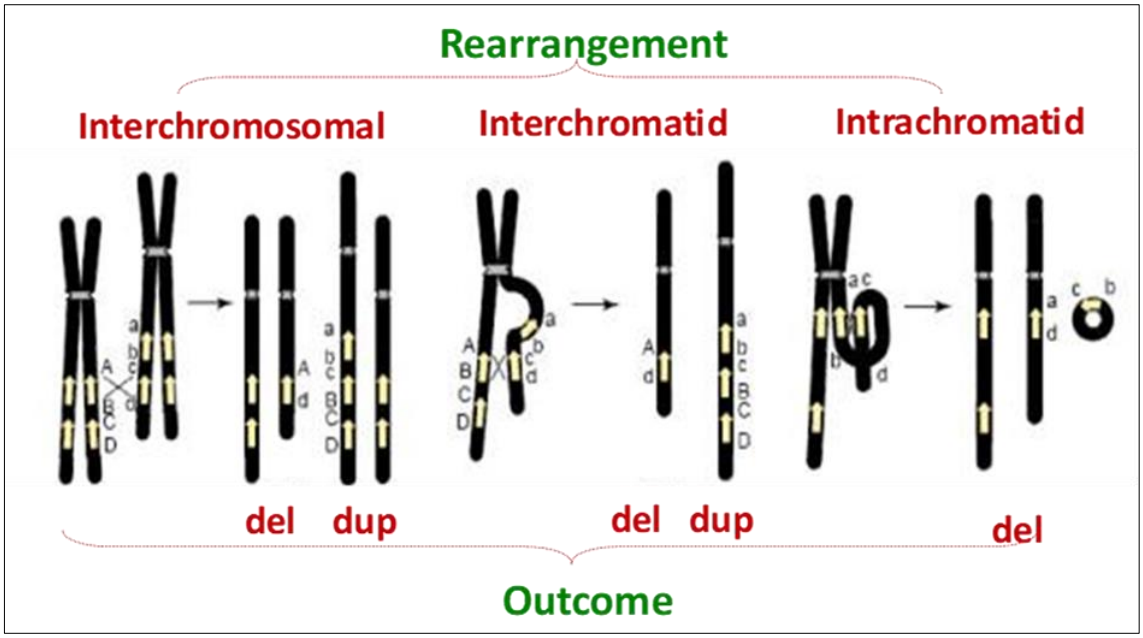

Fig 1: Non-Allelic Homologous Recombination [13]

5. Non-Homologous End Joining (NHEJ)

It is a DNA repair mechanism that operates to repair double stranded break in DNA. Cells use the NHEJ pathway to repair physiological double stranded break (DSBs) such as variable (diversity) joining [V(D)J] recombination as well as DSBs brought on by ionising radiation or reactive oxygen species. It has two distinguishing characteristics i.e., it doesn't require substrates with extensive homology and it can leave an "information scar" at the joining point in the form of nucleotide loss or addition. Occasionally, activation of the NHEJ DNA repair mechanism in a cell might result in errors or mutations at the repair site, resulting in CNV [23]

6. Fork Stalling and Template Switching (Fostes)

This DNA replication phenomenon, which involves the temporary stalling or pausing of a DNA replication and then transferring the replication fork to a different template, can result in complex genomic rearrangements. Replication forks stop or pause at DNA lesions when genomic instability develops, and from the lagging strand, they serially disengage and move to another neighbouring template at another active replication fork that may be moving in either direction. Microhomology is required for this switching. After switching, DNA would be copied at this second sequence and the nascent strand might disengage again after a short time [22].

7. L1-mediated Retro Transposition

A type of retrotransposons in which DNA sequence has the ability to copy itself and insert into target sites. L1 replication cycle starts with synthesis of bi-cistronic mRNA coding for two L1 proteins- ORF1p and ORF2p. After the synthesis, L1 RNA is exported to the cytoplasm then ORF1p and ORF2p proteins are translated and bind to newly transcribed L1 RNA forming L1 ribonucleoprotein particles (RNP) complex. This L1 RNP is imported into the nucleus and integration and reverse transcription occur at the genomic target site. L1 reverse transcriptase (RT) initiates the reverse transcription of L1 RNA of L1 RNP forming cDNA. The L1 RNP complex also contains an endonuclease activity associated with ORF2p. This endonuclease cleaves the genomic DNA at a target site where the L1 is inserted and creates a DNA double-

strand break at the target site. The cDNA binds to the DNA double-strand break at the target site created by the endonuclease activity. Partial reverse transcription can lead to 5′-truncated L1 copies [33].

8. Commonly used softwares

8.1 Penn CNV

SNP array data are used by PennCNV, which employs the Hidden Markov Model (HMM) technique. The advantage of this software is that it integrates data from several sources, including pedigree information, the distance between neighbouring SNPs, the Log R ratio (LRR), B allelic frequency (BAF), and the allele frequency of SNPs [34].

8.2 CNV Partition

This software, which operates from within the Genome Studio Genotyping module, is best for array-CGH data. In order to predict the LRRs and BAFs of various copy number scenarios, the algorithm compares the observed LRR and BAF for each locus to estimate copy number [17].

8.3 Quanti SNP

It utilises the Objective Bayes Hidden-Markov Model (OB-HMM) technique and is used for SNP data. LRR and BAF frequency are regarded as separate variables in this programme, but they are integrated in PennCNV [6].

8.4 CN. MOPS (Mixture of Poisson S)

It is a pipeline of R packages and use Bayesian approach to analyse CNV in NGS data. It locates overlapping sequences and calculates the copy number for each allele [19]

8.5 CNV Finder

A python package for CNV detection on whole exome sequencing data generated using amplicon-based enrichment technologies [12].

9. Implications of CNV in livestock

Copy number variation (CNV), which contributes to the genetic variation seen within populations, has a profound impact on genetic diversity. A few thousand base pairs to millions of base pairs can make up these variances. With the use of whole genome sequencing, 5845 Holstein Friesian cattle were analysed in which 23,256 CNVs were found [3] Array CGH was also done in Tharparkar, a *Bos indicus* breed where 447 CNV's were detected from 447 animals [20]. It was also discovered through analysis of the Swamp buffalo and Murrah population that CNV had a 5–10% genome coverage [9,21]. In sheep and goat, CNV's were identified were OvineSNP50K and CaprineSNP50K, respectively, where CNVs have a 4-8% genome coverage [10,24]. These few instances collectively demonstrate that one of the primary implications of CNV is increasing genotypic variation within a population.

Table 1: List of CNVs identified in different species

Species	Breed (No. of individuals)	No. of CNV	Genome Coverage	Method	Author
Cattle	Holstein (5,845)	23,256		WGS	Butty *et al*., 2021 [3]
	Tharparkar (72)	447	2.17%	aCGH	Kumar *et al*., 2023a [20]
Buffalo	Swamp buffalo (20)	587	5.82%	WGS	Dash *et al*., 2023 [9]
	Murrah (279)	7937	10.33%	WGS	Kumar *et al*., 2023b [21]
Pig	Anhui ind (150)	3863	10.9%	Porcine 80K SNP	Xu *et al*., 2023 [36]
Sheep	Italian sheep breeds (468)	7208	4.05%	OvineSNP50K	Di Gerlando *et al*., 2022 [10]
Goat	Goat (1023)	6286	8.96%	Caprine SNP50K	Liu *et al*., 2019 [24]

CNV was also believed to cause phenotypic variation. For instance, in cattle, the MUC19 gene's CNV and growth traits in 624 individuals were analysed, and it was discovered that there was a significant link between the MUC19-CNV and body length and hip width ($p<0.05$). This finding suggests that CNV can also lead to phenotypic variation as in comparison to other animals, the individuals who had the MUC-19 gene duplication performed better in terms of hip width ($p<0.01$) [5]. Poultry provided yet another illustration of phenotypic diversity. According to reports, massive amplification of duplicated sequence in intron 1 of the gene for SOX5 transcription factor causes the phenotypic pea-comb character in poultry. A closer examination at the duplicated sequence revealed that it comprises a short CpG island and is notably GC-rich in nature. The CpG island has two copies on the wild-type chromosome while a massive copies of about 30 were identified on the pea-comb chromosome [35]. The case of an animal's disease resistance is another consequences of CNV. For example, the DNA sequences of the domesticated pig population and the Eurasian wild boar population were distinguished. A CNV was found in CLEC4E gene of chromosome no.5, which produces the immunological modulator protein Mincle (macrophage inducible C-type lectin) and the copy number of this gene were lost in domestic groups. The greater copy number of CLEC4E in wild boars may be used as proof of adaptation to mycobacterial infection widespread in the wild environment since they are natural reservoirs of mycobacterial infections [18]. Animal reproductive performance has also been found to be impacted by CNVs. CNV was discovered in the promoter region upstream of the LALBA gene, overlapping with two distinct transposons (RTE-BovB and Bov-Ta3) in a comparative examination between swamp buffaloes and river buffaloes. 95.08% of river buffaloes had this CNV, but swamp buffaloes did not. This finding implies that CNV contributes to the increased milk production of riverine buffaloes compared to swamp buffaloes because the LALBA gene is known to be associated with milk production [37]. Furthermore, it has been discovered that CNV may be involved in animals' adaptation to different types of environments. For instance, when individual CNVs were identified, it was discovered that Tibetan cattle have less copies of the TXNRD2 and STUB1 genes. As these genes are are responsible for degrading HIF-1Alpha which is involved in hypoxia induced apoptosis, it was suggested that reduced dosage may promote the adaptation to hypoxia [39].

10. Conclusion

Copy number variation, an interesting phenomenon, having a great impact in the field of genomics is an open eye in understanding the genotypic variation in the population. These segmental rearrangements have been found to be a

source of alteration of genetic structure and further influenced the protein production. Because of its strong impact in the genome structure, incorporating the results of the current research on identification of CNV in genomic selection will change the concept of breeding for genetic improvement.

11. References

1. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet. 2011;12:363e76.
2. Beckmann J, Estivill X, Antonarakis S. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet. 2007;8:639-646.
3. Butty AM, Chud TC, Cardoso DF, Lopes LS, Miglior F, Schenkel FS, *et al.* Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle. Journal of dairy science. 2021;104(7):8050-8061.
4. Chen L, Chamberlain AJ, Reich CM, Daetwyler HD, Hayes BJ. Detection and validation of structural variations in bovine whole-genome sequence data. Genetics Selection Evolution. 2017;49(1):1-13.
5. Chen Y, Peng W, Zhang Z, Liu X, Yang P, Fu C, *et al.* The relationship between MUC19 copy number variation and growth traits of Chinese cattle. Gene; c2022, 47010.
6. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, *et al.* Quanti SNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic acids research. 2007;35(6):2013-2025.
7. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, *et al.* Origins and functional impact of copy number variation in the human genome. Nature. 2010;464(7289):704-712.
8. Dash A, Dhandapani S, Gupta ID, Sivalingam J. Delineating copy number variations in livestock animals. The Pharma Innovation Journal. 2022;11(11S):2289-2297.
9. Dash A, Sivalingam J, Bidyalaxmi K, Sukhija N, Kumar R, Niranjan SK, *et al.* Stepwise Detection of Copy Number Variations in Whole Genome Sequence of Buffalo. Indian Journal of Veterinary Sciences & Biotechnology. 2023;19(3):30-33.
10. Di Gerlando R, Mastrangelo S, Tolone M, Rizzuto I, Sutera AM, Moscarelli A, *et al.* Identification of Copy Number Variations and Genetic Diversity in Italian Insular Sheep Breeds. Animals (Basel). 2022;12(2):217.
11. Fadista J, Thomsen B, Holm LE, Bendixen C. Copy number variation in the bovine genome. BMC Genomics. 2010;11(1):284
12. Fiegler H, Redon R, Andrews D, Scott C, Andrews R, Carder C, *et al.* Accurate and reliable high-throughput detection of copy number variation in the human genome. Genome research. 2006;16(12):1566-1574.
13. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. Pathogenetics. 2008;1:4.
14. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet. 2009;10(8):551-64.
15. Hay EHA, Utsunomiya YT, Xu L, Zhou Y, Neves HH, Carvalheiro R, *et al.* Genomic predictions combining SNP markers and copy number variations in Nellore cattle. BMC genomics. 2018;19:1-8.
16. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, *et al.* Detection of large-scale variation in the human genome. Nat Genet. 2004;36:949e51.
17. Illumina-Sequencing and Array-Based Solutions for Genetic Research. [(accessed on 6 August 2023)]. Available online: http://www.illumina.com
18. Jang J, Kim B, Jhang SY, Ahn B, Kang M, Park C, *et al.* Population differentiated copy number variation between Eurasian wild boar and domesticated pig populations. Sci Rep. 2023;13(1):1115.
19. Klambauer G, Schwarzbauer K, Mayr A, Clevert DA, Mitterecker A, Bodenhofer U, *et al.* cn. MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic acids research. 2012;40(9):e69-e69.
20. Kumar H, Panigrahi M, Saravanan KA, Rajawat D, Parida S, *et al.* Genome-wide detection of copy number variations in Tharparkar cattle. Animal biotechnology. 2023(a);34(2):448-455.
21. Kumar H, Panigrahi M, G Strillacci M, Sonejita Nayak S, Rajawat D, Ghildiyal K, *et al.* Detection of genome-wide copy number variation in Murrah buffaloes. Anim Biotechnol. 2023(b); Jun 29:1-13.
22. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell. 2007 Dec 28;131(7):1235-1247.
23. Lieber MR. The mechanism of human non-homologous DNA end joining. The Journal of Biological Chemistry. 2008;283(1):1-5.
24. Liu M, Zhou Y, Rosen BD, Van Tassell CP, Stella A, Tosser-Klopp G, *et al.* Diversity of copy number variation in the worldwide goat population. Heredity. 2019;122(5):636-646.
25. Locke MEO, Milojevic M, Eitutis ST, Patel N, Wishart AE, Daley M, *et al.* Genomic copy number variation in Mus musculus. BMC Genomics. 2015;16:497.
26. Mielczarek M, Frąszczak M, Giannico R, Minozzi G, Williams JL, Wojdak-Maksymiec K, *et al.* Analysis of copy number variations in Holstein- Friesian cow genomes based on whole-genome sequence data. Journal of Dairy Science. 2017;100(7):5515-5525.
27. Pös O, Radvanszky J, Buglyó G, Pös Z, Rusnakova D, Nagy B, *et al.* DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. Biomed J. 2021;44(5):548-559.
28. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, *et al.* Global variation in copy number in the human genome. Nature. 2006;444(7118):444–454.
29. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen); c2020. p. 245-257.
30. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, *et al.* Large-scale copy number polymorphism in the human genome. Science (New York, N.Y.). 2004;305(5683):525-528.
31. Tice SC. A new sex-linked character in drosophila. Biol Bull. 1914;26:221e30.
32. Upadhyay MR, Chen W, Lenstra JA, Goderie CRJ,

MacHugh DE, Park SDE, *et al*. Genetic origin, admixture and population history of aurochs (*Bos primigenius*) and primitive European cattle. Heredity. 2017; 118:169-176
33. Viollet S, Monot C, Cristofari G. L1 retrotransposition: The snap-velcro model and its consequences. Mob. Genet. Elements. 2014;4(1):e28907.
34. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17(11):1665-74.
35. Wright D, Boije H, Meadows JR, Bed'Hom B, Gourichon D, Vieaud A, *et al.* Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS genetics. 2009;5(6):e1000512.
36. Xu C, Zhang W, Jiang Y, Zhou M, Liu L, Su S, *et al.* Genome-Wide Detection and Analysis of Copy Number Variation in Anhui Indigenous and Western Commercial Pig Breeds Using Porcine 80K SNP Bead Chip. Genes. 2023;14(3):654.
37. Yang L, Han J, Deng T, Li F, Han X, Xia H, *et al.* Comparative analyses of copy number variations between swamp buffaloes and river buffaloes. Animal Genetics. 2023;54(2):199-206.
38. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annual review of genomics and human genetics. 2009;10:451-481.
39. Zhang Y, Hu Y, Wang X, Jiang Q, Zhao H, Wang J, *et al*. Population structure, and selection signatures underlying high-altitude adaptation inferred from genome-wide copy number variations in Chinese indigenous cattle. Frontiers in genetics. 2020;10:1404.