Genomic diversity using copy number variations in worldwide chicken populations.

E. Gorla[1,2][*], F. Bertolini[2], M.G. Strillacci[1], M.C. Cozzi[1], S.I. Roman-Ponce[3], F.J. Ruiz[3], V.V. Vega[3], C.M.B. Dematawewa[4], D. Kugonza[5], A. Elbeltagy[6], C.J. Schmidt[7], S.J. Lamont[2], A. Bagnato[1], M.F. Rothschild[2]

[1] *Department of Veterinary Medicine, Università degli Studi di Milano, Via Celoria 10, 20133 Milano, Italy.*

[2] *Department of Animal Science, Iowa State University, Ames, Iowa, USA.*

[3] *Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias.*

[4] *University of Peradeniya, Faculty of Agriculture, Department of Animal Science, 20400, Sri Lanka.*

[5] *Department of Animal Science, Makerere University, Kampala, Uganda.*

[6] *Animal Production Research Institute, Department of Animal Biotechnology, Dokki, Cairo, Egypt.*

[7] *University of Delaware, Animal and Food Sciences, Newark, DE, USA.*

Recently, many studies in livestock have focused on the identification of Copy Number Variants (CNVs) using high-density Single Nucleotide Polymorphism (SNP) arrays, but few have focused on studying chicken ecotypes coming from many locations. CNVs are polymorphisms, which may influence phenotype and are an important source of genetic variation in populations (Henrichsen et al., 2009). The aim of this study was to explore the genetic difference and structure, using a high density SNP chip in 936 individuals from seven different countries (Brazil, Italy, Egypt, Mexico, Rwanda, Sri Lanka and Uganda). The DNA was genotyped with the Affymetrix Axiom®600k Chicken Genotyping Array and processed with stringent quality controls to obtain 559,201 SNPs in 915 individuals (Stranger et al., 2007). The Log R Ratio (LRR) and the B Allele Frequency of SNPs were used to perform the CNV calling with PennCNV software based on a Hidden Markov Model analysis and the LRR was used to perform CNV detection with SVS Golden Helix software.

After filtering, a total of 19,027 CNVs were detected with the SVS software, while 9,065 CNVs were identified with the Penn CNV software. The CNVs were summarized in 7,001 Copy Number Variant Regions (CNVRs) and 4,414 CNVRs, using the software BedTool.

The consensus analysis across the CNVRs allowed the identification of 2,820 consensus CNVR, of which 1,721 were gain, 637 loss and 462 complex, for a total length of 53 Mb corresponding to the 5 % of the GalGal5 chicken autosomes (Table 1). Only the consensus CNV regions obtained from both detections were considered for further analysis.

The intersection analysis performed between the chicken gene database (Gallus_gallus-5.0) and the 1,927 consensus CNVRs allowed the identification (within or partial overlap) of a total of

2,354 unique genes with an official gene ID (Quinlan et al., 2010). The CNVRs identified here represent the first comprehensive mapping in several worldwide populations, using a high-density SNP chip.

Table 1: Summary statistic of CNVR identified with of PennCNV, SVS software and consensus, divided according state into Loss, Gain or Complex.

	CNVR Count	Min Length (bp)	Max Length (bp)	Mean Length (bp)	Total Length (bp)	Coverage (%)
			SVS			
ALL	7,001	493	571,113	15,541.61	108,806,878	11.68
Gain	4,640	493	571,113	14,743.86	68,411,533	7.34
Loss	1,573	699	146,115	11,288.70	17,757,126	1.9
Complex	788	1,955	562,765	28,728.70	22,638,219	2.43
			PennCNV			
ALL	4,414	122	2,281,744	51,853.14	228,879,785	17.8
Gain	3,281	122	750,953	47,193.85	154,843,028	12.04
Loss	731	149	181,129	11,661.74	8,524,736	0.66
Complex	402	472	2,281,744	162,965.22	65,512,021	5.09
			Consensus			
ALL	2,820	652	562,765	19,072.43	53,784,265	5.77
Gain	1,721	652	544,758	17,181.11	29,568,693	3.17
Loss	637	1,102	146,115	12,948.43	8,248,151	0.88
Complex	462	2,696	562,765	34,561.51	15,967,421	1.71

References

Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., et al., 2007. Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. Science 315: 848–853.

Henrichsen, C.N., Vinckenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schutz, F., et al., 2009. Segmental copy number variation shapes tissue transcriptomes. Nat. Genet. 41:424–429. doi: 10.1038/ng.345.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. doi: 10.1093/bioinformatics/btq033.

http://penncnv.openbioinformatics.org/en/latest.

http://goldenhelix.com.