


RESEARCH ARTICLE

Open Access



Genomic variability in Mexican chicken population using copy number variants

E. Gorla¹, M. C. Cozzi¹, S. I. Román-Ponce², F. J. Ruiz López², V. E. Vega-Murillo³, S. Cerolini¹, A. Bagnato¹ and M. G. Strillacci^{1*} 

Abstract

Background: Copy number variations are genome polymorphism that influence phenotypic variation and are an important source of genetic variation in populations. The aim of this study was to investigate genetic variability in the Mexican Creole chicken population using CNVs.

Results: The Hidden Markov Model of the PennCNV software detected a total of 1924 CNVs in the genome of the 256 samples processed with Axiom® Genome-Wide Chicken Genotyping Array (Affymetrix). The mapped CNVs comprised 1538 gains and 386 losses, resulting at population level in 1216 CNV regions (CNVRs), of which 959 gains, 226 losses and 31 complex (i.e. containing both losses and gains). The CNVRs covered a total of 47 Mb of the whole genome sequence length, corresponding to 5.12% of the chicken galGal4 autosome assembly.

Conclusions: This study allowed a deep insight into the structural variation in the genome of unselected Mexican chicken population, which up to now has not been genetically characterized. The genomic study disclosed that the population, even if presenting extreme morphological variation, cannot be organized in differentiated genetic subpopulations. Finally this study provides a chicken CNV map based on the 600 K SNP chip array jointly with a genome-wide gene copy number estimates in a native unselected for more than 500 years chicken population.

Keywords: Copy number variant, Chicken, Genetic variability

Background

Copy Number Variants (CNV) are genomic structural variations distributed over the whole genome in all species and refers to genomic segments of at least 50 bp in size [1], for which copy number differences have been observed in comparison to reference genome assemblies (insertions, deletions and more complex changes) [2, 3]. Sequencing of the chicken genome, released in 2004 [4], has facilitated the use of molecular markers for breed/ecotype characterization. Structural variation has been recognized as an important mediator of gene and genome evolution within populations. In the last decades, microsatellite markers have been often used to perform phylogenetic analysis and studies on genetic variability in chicken populations [5–7]. Although numerous studies investigating genetic variation have focused on SNPs, there is a growing evidence for the substantial role of

structural DNA polymorphism in phenotypic diversity [8]. It has been shown that CNVs are ubiquitous in the genome and can contribute substantially to phenotypic variability and disease susceptibility in humans [8, 9] and animals [10, 11]. The underlying assumption is that CNVs are changing the gene structure and dosage and altering the gene regulation [8–13]. Even if CNVs are less frequent than SNPs in terms of absolute numbers, CNVs cover a larger proportion of the genome and have, therefore, a large potential effect on phenotypic variability [14]. Compared with humans and other model organisms, there is limited research on the extent and impact of CNVs in the chicken genome.

In Mexico the poultry population, even if it shows large morphological variability, is not divided into breeds or strains and, possibly, can be considered as a unique widespread Creole chicken population (*Gallus gallus domesticus*), as the result of undefined crosses among different breeds imported into Mexico from Europe for almost 500 years [15, 16]. Creole chickens include, in fact a wide variety of biotypes with different colors of

* Correspondence: maria.strillacci@unimi.it

¹Department of Veterinary Medicine, Università degli Studi di Milano, Via Celoria 10, 20133 Milan, Italy

Full list of author information is available at the end of the article



plumage and morphological features that are widely distributed in the country [17]. In the absence of comprehensive breed characterization data and documentation of the origin of breeding populations, DNA polymorphism provides the most reliable estimates of genetic diversity within and between a given set of populations [18].

Several studies have been developed in the recent past to detect CNV in poultry using low-density 60 K SNP chips [19] or aCGH [20–22]. The major limits of these studies reside in the density of the spots of the used arrays and the limited sample size. It has been already suggested by Jia et al. [23] that the use of the 600 K SNP array can improve the efficiency of the CNV detection in the poultry species. The whole genome sequence data can improve the detection of small CNVs but, even if desirable and employed by some authors [24, 25], is still economically too demanding to be realized over a large number of samples.

The aim of this study was to map the CNV in the Mexican chicken population with an unprecedented resolution using high density SNP chip (i.e. 600 K Affymetrix SNP chip) on a large number of individuals (i.e. 256) and to characterize the genetic variability of the Mexican Creole chicken's population using CNV as genomic markers.

Methods

Sampling and genotyping

In this study a collection of 265 individuals of the Mexican chicken population, from different farms across 26 states of United States of Mexico, was previously sampled by Instituto Nacional de Investigaciones Forestales, Agrícola y Pecuarias (INIFAP) within the institutional activities of the Centro Nacional de Recursos Genéticos at Tepatitlán, Jalisco. As mentioned hereinbefore, a classification of the Mexican population in breeds does not exist. For this reason, the birds have been considered as a unique Creole population and sampled in several states of Mexico.

Samples were processed and genotyped within the framework of a previous project of INIFAP using the 600 K Affymetrix Axiom® Chicken Genotyping Array, containing 580,954 SNPs distributed across the genome, with an average spacing of about 1.8 kb and data made available for the present study. A commercial service provider performed the genotyping and the DNA extraction from feathers. The galGal4 chicken assembly was used in this study as reference genome.

Quality assurance of CNV raw data and CNV detection

The CNV detection was performed on a total of 471,730 SNPs on the first 28 chicken autosomes.

The Axiom® Analysis Suite software (Affymetrix) was used to perform raw intensity data Quality Control and

run the genotyping algorithms. Default quality control settings were applied to filter for low quality samples before running the genotyping analysis, to exclude the ones with call rates <97% and Dish Quality Control <0.82. The Axiom® CNV summary software tool was used to generate input files for CNV prediction analysis.

The CNV detection was performed with PennCNV software [26] using Log R Ratio and the B allele frequency [27] obtained with the Axiom® CNV Summary Tool software. The individual-based CNV calling was performed using the default parameters of the Hidden Markov Model (HMM): standard deviation of LRR <0.30, BAF drift as 0.01 and waviness factor at 0.05 and a minimum of 3 SNP was required to define a CNV. The distribution of CNV per individual spanned from 0 CNV to more than 100. Up to 79 CNV the distribution was continuous, while a step to more than 100 CNV was detected in 9 birds. To avoid the introduction of possible false positive and a bias in the CNV interpretation they were then filtered out as the number of CNVs detected appeared to be outlier respect to the CNV distribution, leaving 256 samples for further analyses. It is worth to mention that Zhang et al. [19] have performed a validation of the CNV called by PennCNV, using the CNVPartition program obtaining an overlapping of results of 99%. Additionally recent studies in cattle [28] have used two software to map CNV based on different algorithms: the HMM of PennCNV, based on the CNV identification on B allele frequency and Log R ratio, and the CNAM of SVS (Golden Helix) basing the identification only on Log R ratio. These studies provide an additional empirical evidence of the results provided by Xu et al. [29] that in their study concluded that using multiple CNV calling algorithms might also increase the false positive rate.

In addition to detect the outliers as hereinbefore described, in order to minimize the false positive callings, the PennCNV was run using different ".hmm" files (agre.hmm, affygw6.hmm, hh550.hmm), which is known that may affect substantially the false positive as well as the false negative rate. The online PennCNV manual (<http://penncnv.openbioinformatics.org/en/latest/>) in fact instruct the user that the agre.hmm file produces an excess of false positive calls respect to the default affygw6.hmm file, which has been criticized to produce a low number of CNV calls (i.e. excess of false negative) respect to other calling software and algorithms. Additionally we used the hh550.hmm file in the calling process, which is based on a chip with the closest number of SNPs respect to the SNP chip used here. To reduce the false calling rate we have then considered valid only the CNV calls obtained both with the agre.hmm and the hh550.hmm files. The number of CNV calls resulted using

the affygw6.hmm files were negligible respect to other two files, but anyhow present in the consensus here obtained. The hmm file supplied to the HMM of PennCNV, (<http://penncnv.openbioinformatics.org/en/latest/>), provides to the algorithm the expected signal intensity values for different states of CNV and the expected probability for the transition in different copy number state. As described in the PennCNV user manual, however, the transition probability is a function of the distance between neighboring markers. This makes the choice of a correct hmm file, in respect to the density of markers, a critical step in the mapping of CNV to control false positive and negative calls.

CNVR definition

The CNV regions (CNVRs) were obtained using the BedTools software (`-mergeBed` command) [30], through merging overlapping CNVs by at least 1 bp, as described by Redon et al. [12]. CNVRs were classified as gain, loss and complex CNVRs (i.e. a CNVR comprising both gain and loss events).

CNVR annotation

After downloading the list of chicken autosome galGal4 genes (GCA_000002315.2) from Ensembl database (Release 88) (<http://www.ensembl.org>), the gene annotation was performed using the software Bedtools, *intersect* command [30], identifying the genes fully included in, or partially overlapping, the defined CNVRs. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed using the Panther database (<http://pantherdb.org>).

Clustering analysis using CNVRs

A clustering analysis was performed considering CNVRs found in this study [31]. A scoring matrix of the CNVRs was constructed, attributing the “0” or “1” values to indicate the absence or the presence of a CNV in a specific CNVR. A hierarchical agglomerative clustering was then applied to the scoring matrix using the pvclust function

of the pvclust R package [32]. Multiscale bootstrap resampling (no. 10,000 bootstraps) was used to obtain the Approximately Unbiased *P*-value (AU), in order to determine the robustness of branches. The Unweighted Pair Group Method with Arithmetic mean (UPGMA) was the Agglomerative method chosen.

Results

CNV and CNVR detection

The Table 1 reports the descriptive statistics of identified CNVs and CNVRs. The HMM of the PennCNV software detected a total of 1924 CNVs; among these, 386 were deletions (i.e. loss state) and 1538 were duplication (i.e. gain state), with a deletions/duplications CNV ratio of 0.25, calculated as the total number of losses divided by the total number of gains.

The CNVs overlapping among samples were summarized across all individuals into 1216 CNVRs (959 gains, 226 losses and 31 complex), covering a total of 47 Mb of sequence length, corresponding to 5.12% of 28 autosomes in the galGal4 assembly (Additional file 1: Sheet 1).

In Fig. 1 the CNVRs map, divided in gain, loss and complex on each chromosome is shown.

In Table 2 the number of CNVRs found is reported, together with the state and the proportion of coverage by chromosome. The coverage proportion is smaller than 10% for all chromosomes, except for 16, 18, 24, 27 ones.

CNVRs were classified as singleton if detected in only one individual. Among the identified CNVRs, 1009 (82.9%) were present in singleton, 127 (10.4%) in two individuals, 30 (2.4%) in three individuals, 11 (0.9%) in four individuals, and 39 (3.2%) in five or more individuals. For every state (i.e. gain, loss, complex) CNVRs were divided according to their length into four classes: <1 kb, 1–10 kb, 10–100 kb, >100 kb; Fig. 2 reports the CNVRs count for each class of CNVRs length.

The majority of the 1065 CNVRs identified in this study had a length comprised between 10 kb and 100 kb, of which 471 comprised between 1 kb and 10 kb

Table 1 Descriptive statistics for Copy Number Variants (CNVs) and Copy Number Variants Regions (CNVRs) identified in the Mexican chicken population

Type	No.	Length	Min length	Max length	Mean length	Median length	Total Coverage
CNVs							
Loss	386	12,575,609	92	574,231	32,579	6038	1.37%
Gain	1538	74,022,420	138	1,345,291	42,129	22,810	8.05%
All	1924	86,598,029	92	1,345,291	45,009	19,273	9.42%
CNVRs							
Loss	226	3,920,955	92	279,420	17,349.36	4950	0.43%
Gain	959	38,550,088	138	1,345,291	40,198.21	15,414	4.19%
Complex	31	4,580,519	3501	607,435	147,758.7	60,250	0.50%
All	1216	47,051,562	92	1,345,291	38,693.72	13,897.5	5.12%

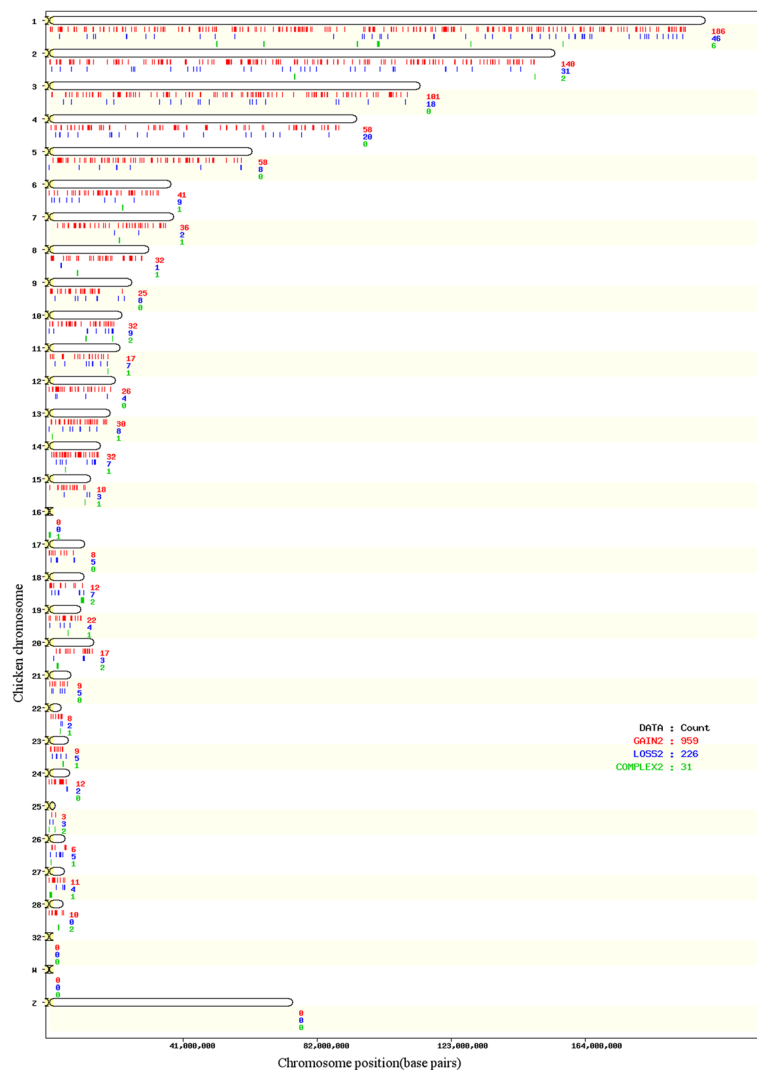


Fig. 1 Physical distribution of the Copy Number Variants Regions (CNVRs) according to states (gain, loss and complex)

and 594 comprised between 10 kb and 100 kb. A total of 39 CNVRs had a length lower than 1 kb while 112 CNVRs showed a size longer than 1 Mb (Fig. 3). The highest number of gain and complex CNVRs are those with a length of 10–100 kb, while the loss CNVRs were present at largest frequency within a length of 1–10 kb (Fig. 3).

The regions mapping in a large number of individuals were: the CNVR on chromosome 1 at 42.96–43.13 Mb, identified in 61 samples; the CNVR on chromosome 12 at 1.12–1.22 Mb, identified in 56 samples; the CNVR on chromosome 16 at 1253–533,589 bp, identified in 53 samples; the CNVR on chromosome 1 at 193.13–193.24 Mb, identified in 52 samples.

The Fig. 3 shows the sample count for every CNVR state according to the previously defined 4 CNVR length classes (as shown in Fig. 2). The sample count classes were defined as: 1 (singleton), 2–5, 6–20 and >20. The

gain CNVRs (Fig. 3a) have a sample count distribution with most of the regions falling into the 10–100 kb class. The loss CNVRs (Fig. 3b) have a sample count distribution with most of the regions falling into the 1–10 kb class. Class 1 mostly represents the gain regions. Furthermore, class 1 is the most frequent in all length classes. The highest length and sample classes mainly belong to the gain regions. In the complex region (Fig. 3c) the class mostly represented is the 10–100 kb one. More precisely, the most represented sample class is the 2–5 class falling mainly within the 10–100 kb length class. Furthermore, class 2–5 is the most frequent. Lastly, all the sample classes are distributed mostly within the 1–10 and 10–100 length classes.

CNVR annotation

The intersection analysis performed between the chicken gene database (Ensembl galGal4) and our CNVRs allowed

Table 2 Number and proportion of genome covered (Coverage %) by Gain, Loss and Complex Copy Number Variants Regions per chromosome (CHR)

CHR	Gain (%)	Loss (%)	Complex (%)	Total	Coverage (%)
1	186 (3.94)	46 (0.38)	6 (0.29)	238	4.61
2	140 (4.78)	31 (0.38)	2 (0.14)	173	5.29
3	101 (3.02)	18 (0.11)	0 (0)	119	3.13
4	58 (3.40)	20 (0.36)	0 (0)	78	3.75
5	58 (6.43)	8 (0.15)	0 (0)	66	6.58
6	41 (3.61)	9 (0.15)	1 (0.15)	51	3.91
7	36 (4.03)	2 (0.02)	1 (0.46)	39	4.51
8	32 (4.55)	1 (0.30)	1 (0.68)	34	5.53
9	25 (3.22)	8 (0.23)	0 (0)	33	3.45
10	32 (5.06)	9 (0.79)	2 (1.11)	43	6.96
11	17 (2.64)	7 (0.78)	1 (0.19)	25	3.61
12	26 (2.73)	4 (0.16)	0 (0)	30	2.89
13	30 (3.88)	8 (1.05)	1 (0.52)	39	5.45
14	32 (7.72)	7 (2.05)	1 (0.20)	40	9.97
15	18 (1.90)	3 (0.12)	1 (0.31)	22	2.33
16	0 (0)	0 (0)	1 (81.60)	1	81.60
17	8 (2.28)	5 (0.97)	0 (0)	13	3.26
18	12 (3.54)	7 (2.06)	2 (5.03)	21	10.63
19	22 (8.32)	4 (0.23)	1 (0.91)	27	9.46
20	17 (3.57)	3 (0.26)	2 (0.39)	22	4.22
21	9 (1.60)	5 (0.30)	0 (0)	14	1.90
22	8 (4.31)	2 (0.74)	1 (0.62)	11	5.67
23	9 (4.78)	5 (0.95)	1 (0.73)	15	6.46
24	12 (9.91)	2 (0.24)	0 (0)	14	10.14
25	3 (2.41)	3 (1.13)	2 (2.39)	8	6.48
26	6 (2.27)	5 (2.11)	1 (1.46)	12	5.84
27	11 (6.04)	4 (3.66)	1 (10.74)	16	20.45
28	10 (3.36)	0 (0)	2 (2.24)	12	5.61
Total	959	226	31	1216	

^aCoverage of CNVR by chromosome and state (gain/loss/complex) relatively to each chromosome length

the identification (within or overlapping the consensus CNVRs) of 1543 Ensemble genes ID, corresponding to 1068 genes with an official gene ID. Out of the 1216 CNVRs identified in this study, 783 (64.4%) encompassed one or more genes, while 433 (35.6%) did not involve any gene. More specifically, among these genes, 1028 (96.25%) were protein-coding genes, 34 (3.1%) were miRNAs and 6 (0.56%) were small nuclear RNAs (Additional file 1: Sheet 4).

The Panther database provided the annotation information, according to GO terms and KEGG pathways, for only 865 chicken genes. The Additional file 1 reports the annotation output including only terms

resulted statistically significant after Bonferroni correction (p -value <0.05): 27 classified as Cellular Component, 11 as Molecular Function, and 28 as Biological Process. The significant GO terms were mainly involved in muscle contraction, sensory perception of sound, response to stimulus, cellular component morphogenesis and movement, and cell communication (Additional file 1: Sheet 5). Instead, the KEGG pathway analysis indicated that these genes are involved in 166 pathways, but none of which was significant after Bonferroni correction.

Clustering analysis using CNVRs

The Fig. 4 shows the cluster-tree built for the chicken Mexican Creole population based on CNVRs similarities.

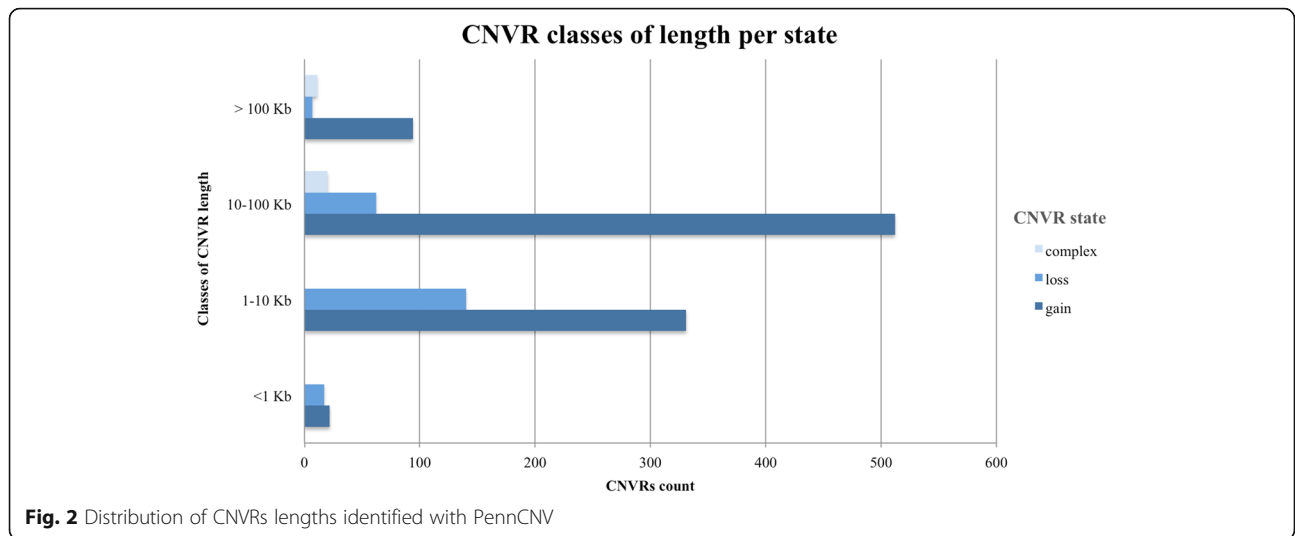
In the plot (Fig. 4), the branch length is not directly proportional to the genetic distance estimated among samples. The Approximately Unbiased P -value (AU- P in red) and Bootstrap Probability value (BP- P in green), indicative of how strongly the cluster is supported by the data, were shown for each node, as well as the Edge numbers (in light grey). As can be read from Fig. 4 mostly all AU- P and BP- P values are zero, showing no difference among branch in which individuals are clustered in: there is no cluster with both AU- P and BP- P values greater than 0.

Discussion

CNV and CNVR detection

The use of a high density SNP chip allows to disclose smaller CNVs compared to studies performed in the recent past that were based on a 60 K SNP chip [19] or on aCGH [20–22]. The average probe distance in the SNP chip used here is in fact more or less 1,8 kb (galGal4) allowing the identification of short CNVs. The smaller CNV (i.e. 92 bp.) that was detected in this study (Table 1), according to the criteria of minimum 3 SNPs to map a CNV, overlaps with the one mapped by Yi et al. [24] using a sequencing approach.

Chromosome 16 is the only one with a very large proportion of length covered by CNV, i.e. 81% (Table 2). This may be due to the small length of the autosome and to the presence of the Major Histocompatibility Complex (MHC), which is known to be affected by variation in genome copy number as reported by Fulton et al. [33]. The region is resulting in this study as a complex CNVR, but having the majority of individual CNVs (46 over 52) to be gain variant (45 heterozygous duplications, 1 homozygous duplication). The existence of such a CNV is possibly due to the importance that the MHC has in immune resistance. As it is known by literature in fact, the high number of polymorphic sites, closely associated with resistance against infection diseases (e.g.

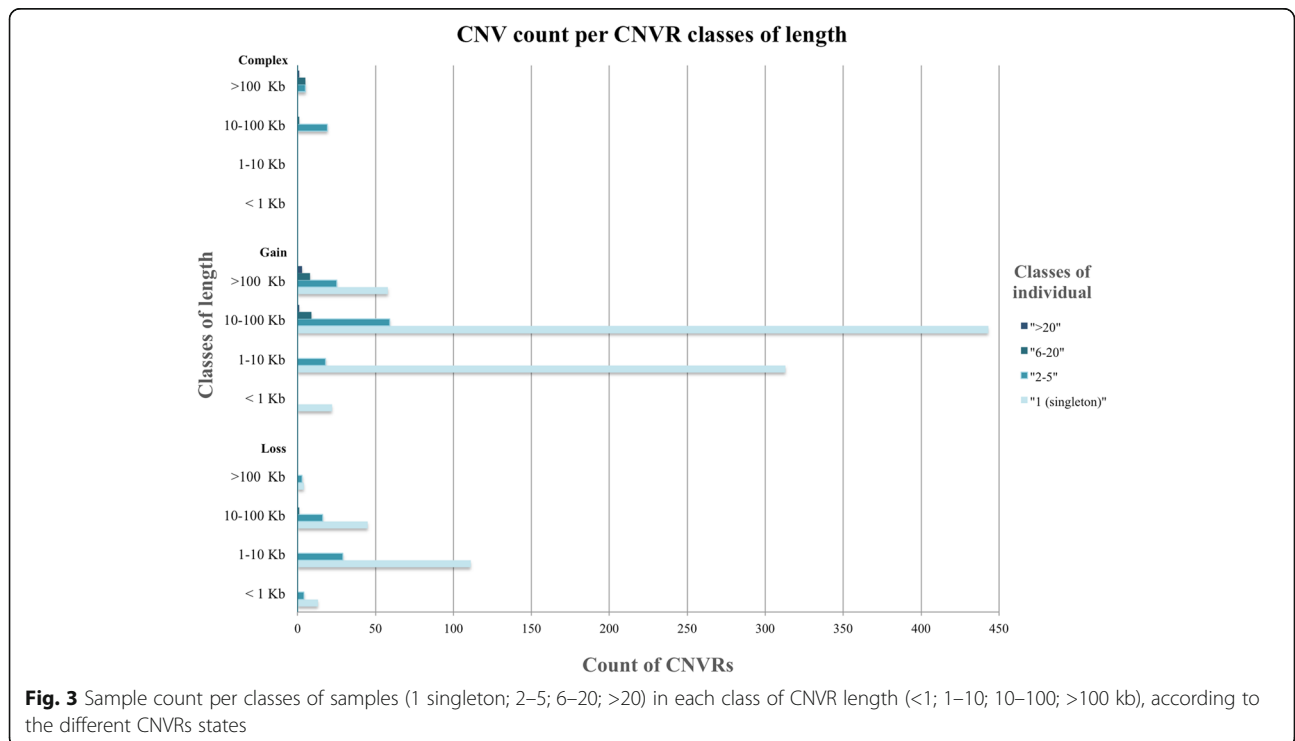


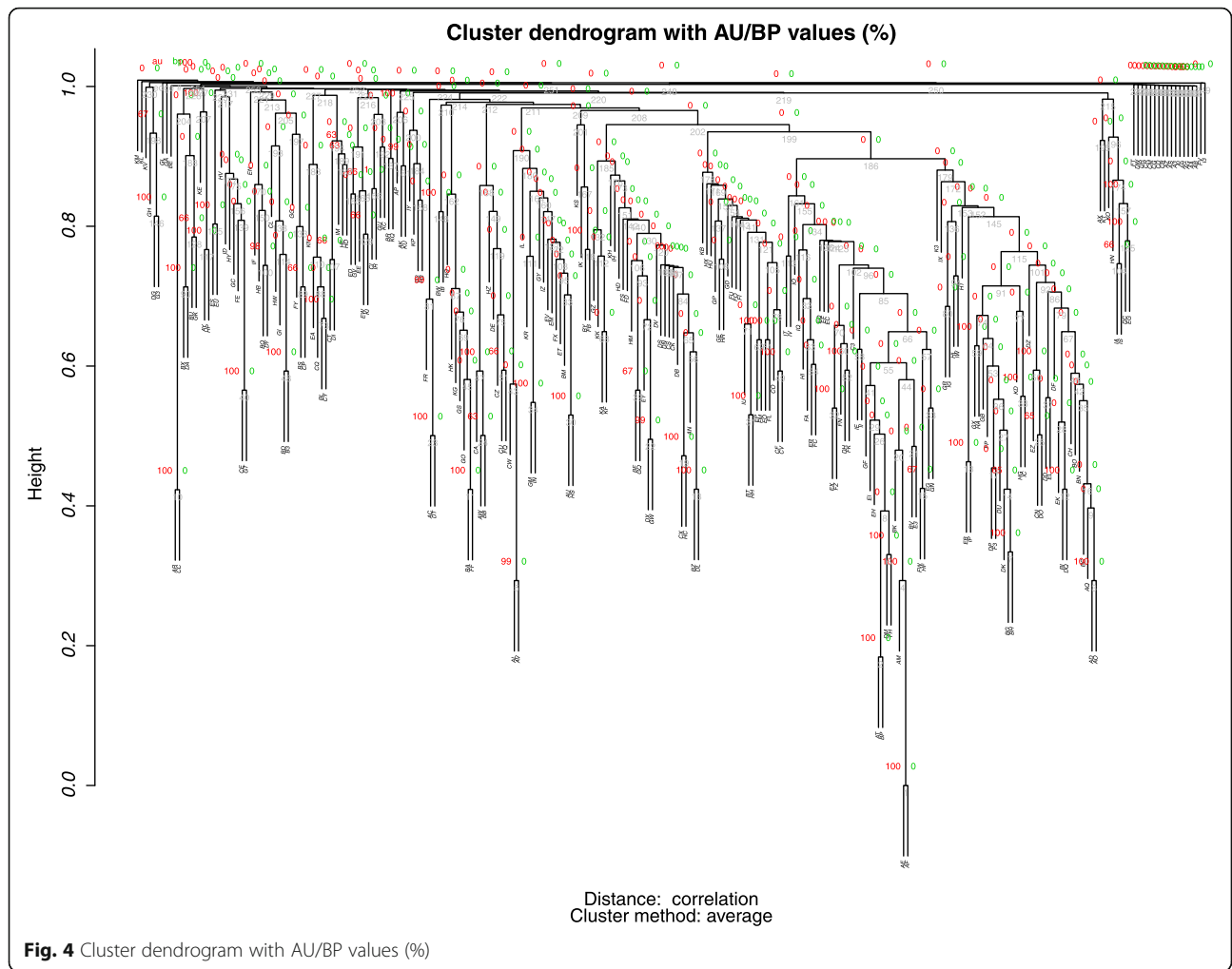
Marek’s disease, avian Influenza, Rous sarcoma disease, avian leukosis, infectious bursal disease, avian infectious bronchitis, *Salmonella enteritidis*, *E. coli* and other bacterial diseases), characterizes this complex [34, 35].

The large proportion of singleton CNVRs has been previously reported in chicken populations also by Yi et al. [24], Han et al. [22] and Strillacci et al. [36], finding a total fraction of 68.8%, 76.5% and 75%, respectively. Our findings confirm their results and showed that also in the Mexican chicken population the segregation of CNVs exists among individuals.

Comparison with previous chicken CNV studies

In order to perform a comparison with previous studies mapping CNVs in chicken, we migrated autosomal CNVRs coordinates from galGal3 to galGal4 for the CNVRs identified by Tian et al. [21], by Crooijmans et al. [20] and by Han et al. [22] using the UCSC liftOver tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). In total 201 out of 308 (65%) autosomal CNVRs detected by Tian et al. [21], 837 out of 1504 (56%) mapped by Crooijmans et al. [20] and 134 out of 264 (50.75%) identified in Han et al. [22] were converted successfully.





The comparison among the CNVRs found in this study and those found in other 7 studies [19–36] is reported in Table 3 and in the Additional file 1: Sheet 2 showing the number of CNVRs overlapping among the studies.

The 1216 CNVRs detected in this study overlap with 617 mapped by others confirming that a proportion of

51% of them were found by independent methods and in other populations (Additional file 1: Sheet 2).

As reported in Table 3, the proportion of overlapping CNVRs between this study and each of the other 7 studies ranged from 2.38% to 35.19%. Independently from the breeds included in all studies, the CNVRs detection is mainly influenced by the sample size and by

Table 3 Comparison between CNVRs detected in this study and in other 4 published ones

Study	Method	Samples	Breeds	CNVR	Length overlap (Mb)	Common CNVR	Overlap (%)
Crooijmans et al. [20]	aCGH	64	7	837 ^a	4.49	92	7.57
Tian et al. [21]	aCGH	22	11	201 ^a	0.969	29	2.38
Zhang et al. [19]	SNP chip (60 K)	475	11	438	19.903	80	6.58
Han et al. [22]	aCGH	10	4	134 ^a	1.311	29	2.38
Yi et al. [24]	Sequencing	12	12	8487	10.424	428	35.19
Yan et al. [25]	Sequencing	6	2	5009	2.933	256	21.05
Strillacci et al. [36]	SNP chip (600 K)	96	6	564	3.855	109	8.96
This Study	SNP chip (600 K)	256	1	1216	47.05		

^aThis value refers to the number of CNVRs after the shifting to galGal4

the algorithm and the technology used to CNVs mapping (i.e. aCGH vs. SNP or whole genome sequence). The largest overlap rates occurred in fact when the comparison is done with studies using in their analyses a large sample of individuals [24–25]. On the contrary, a low overlap occurred when the comparison was performed with studies that employed a low number of samples, when CNVs were detected with different technical methods (i.e. aCGH or whole genome sequencing) and calling algorithms.

No CNVR is simultaneously common to this and to all the 7 other studies here considered. The Additional file 1: Sheet 3 reports the list of CNVRs simultaneously shared by our study and at least 3 other ones among the 7 here considered, and the annotated genes found in the regions. As shown, the CNVR common among 7 studies are 4 and are located on chromosome 1 at 42.96–43.13 Mb, chromosome 5 at 2.6–3.9 Mb, chromosome 8 at 15.45–15.47 Mb and chromosome 9 at 3.42–3.49 Mb.

In particular the CNVR on chromosome 1 is common to 7 studies and includes the *KITLG* (*KIT ligand*), a pigmentation candidate gene that has a role in controlling the migration, survival and proliferation of melanocytes. Rare mutations in the mouse homolog of *KITLG* are known to affect coat color [37]. Additionally Metzger et al. [38] highlighted the importance of this gene in the reproduction efficiency in horses claiming its general effect in all livestock populations.

The CNVR on chromosome 5 (2.60–3.95 Mb) (Additional file 1: Sheet 3) harbors the *BDNF* (*brain derived neurotrophic factor*) gene, which seems to be involved in chicken heat stress response. In fact, Lamont et al. [39] reports that early thermal conditioning allows increased transcription of the *BDNF* gene in response to heat stress later in the bird's life. Furthermore, previous findings indicate that *BDNF* prevents the death of cultured chick retinal ganglion cells and, as reported by Herzog et al. [40], the tightly controlled expression of the *BDNF* gene might be important in the coordinated development of the visual system in chicks. Also, in the same CNVR on chromosome 5 lies the *LGR4* (*leucine rich repeat containing G protein-coupled receptor 4*) gene that in human is associated with low bone mineral density [41]

In the region on chromosome 8 no genes were annotated, while in the region on chromosome 9 the *IMP4* (*U3 small nucleolar ribonucleoprotein*) and the *VPS8* (*Vacuolar Protein Sorting-Associated Protein 8 Homolog*) genes are annotated, but there are no studies that associate these genes to specific traits.

CNVR annotation

Additionally, quantitative trait loci (QTL) from chicken QTLdb (<http://cn.animalgenome.org/cgi-bin/QTLdb/GG/>

index) were downloaded in order to examine their overlapping with the identified CNVRs. Because the confidence intervals of some QTL were too large, we considered QTL less than 5 Mb of length. A total of 656 CNVRs overlapped with 918 QTL, corresponding to 172 different traits that included mainly: body weight, body size, carcass traits, fatness traits, Marek's disease-related traits, and egg shell (Additional file 1: Sheet 6).

Some of the genes identified in our CNVR have already been associated with functional traits in chickens in previous studies. The region identified on chromosome 4 at 80.75–81.02 Mb contains the gene *SORCS2* (*sortilin related VPS10 domain containing receptor 2*) associated with aggressive behavior traits in males [42]. The region on chromosome 1 at 130.82–130.89 Mb includes the gene *OCA2* (*oculocutaneous albinism II*). This gene had highly significant effects on body weight in weeks 11–12 in chicken, as reported by Gu et al. [43] and is also involved in pigmentation [44]. The CNVRs on chromosome 1 at 65.63–65.98 Mb and at 66.02–66.03 Mb harbor *SOX5* (*SRY-box 5*) gene, which is involved in chicken the Pea-comb expression. In fact, Pea-comb is caused by a duplication located near conserved non-coding sequences in intron 1 of the gene [45]. Three regions on chromosome 1 at 146.55–146.59 Mb, at 147.08–147.13 Mb and at 147.78–147.80 Mb harbor the *glypican 6* (*GPC6*) gene, *glypican 5* (*GPC5*) gene, which are located within the QTL for bodyweight identified in previous studies [46, 47].

The CNVR on chromosome 18 (5.00–5.02 Mb) includes the *FASN* (*fatty acid synthase*) gene that has been identified as one of the gene that control fat deposition in chickens (i.e. fat bandwidth, abdominal fat percentage and abdominal fat weight) [48].

Finally, some genes included in 10 different CNVRs found in this study are classified into the pathway for salmonella infection (http://www.genome.jp/dbget-bin/www_bget?gga05132). These genes are: *IFNG* (*interferon gamma*) (chromosome 1 at 34.95–35.16 Mb), *DYNC2H1* (*dynein cytoplasmic 2 heavy chain 1*) (chromosome 1 at 182.31–182.3 Mb), *WASF1* (*WAS protein family member 1*) (chromosome 3 at 66.86–66.87 Mb), *ARPC2* (*actin related protein 2/3 complex subunit 2*) (chromosome 7 at 22.60–22.70 Mb), *TJP1* (*tight junction protein 1*) (chromosome 10 at 6.08–6.11 Mb), *DYNC1LI2* (*dynein cytoplasmic 1 light intermediate chain 2*) (chromosome 11 at 11.42–11.51 Mb), *FLNB* (*filamin B*) (chromosome 12 at 8.87–8.87 Mb), *RAB7A* (*member RAS oncogene family*) (chromosome 12 at 9.15–9.15 Mb), *ARPC1B* (*actin related protein 2/3 complex subunit 1B*) (chromosome 14 at 4.38–4.38 Mb), *PLEKHM2* (*pleckstrin homology and RUN domain containing M2*) (chr21 at 4.21–4.22 Mb).

Clustering analysis using CNVRs

The results of this study suggest that there is not a clear division in classifiable subpopulations based on the CNVR characterization and, thus, that the Mexican Creole chicken population can be considered a unique genetic mix. These results are different to the ones recently found by Strillacci et al. [36] using the same approach in Italian well defined chicken breeds clearly clustered by CNVRs classification and by Tian et al. [21] and Wang et al. [49] in chicken and pigs respectively, showing additional evidence of the usefulness of CNV as markers for differentiating individuals. To provide a validation of the approach here used to cluster individuals of the Mexican population with CNVs we performed a PCA and a hierarchical clustering using the SNP genotypes: no clustering was obtained and the population resulted as for CNVs a unique genetic mix (Additional file 2: Fig. S1).

Conclusion

This study is the first CNV genomic analysis on a large sample of individuals of the Mexican chicken population based on high-density SNP chips. It provides insights into the genetic and genomic architecture of the Mexican Creole chicken population, providing valuable genomic source of structural variation to enrich the chicken CNV map, helping future CNV association studies for important traits in chickens.

The major result resides in the disclosure of the genetic homogeneity of the Mexican chicken population. This result allows to consider all individuals of population as a unique genetic mix deriving from the introduction of chicken in the American continent, following the colonization from Europe. According to our results the CNV variation in the population does not allow to disclose breeding strategy addressed to specific selection criteria. The same method, we used here based on the CNV, was able to dissect properly different Italian breeds in a previous study [36]. The results of this study, thus, suggest that there is not a clear division in classifiable subpopulations based on the CNVR characterization and that the Mexican Creole chicken population can be considered a unique mix of genetics.

Most of the 1216 CNVRs detected were novel variants disclosed thanks to the HD SNP chips here used, which enrich the current poultry CNV database. This mapping is having a particular value because it is based on a unique poultry population, that we assumed to own a larger genetic variability respect to selected commercial population, as reproduction is based on an outbreeding mating system by more than 500 years.

Finally we detected 1543 Ensemble genes ID overlapping with CNVRs, including genes involved in well-known phenotypes such as *KITLG* and *OCA2* on chromosome 1,

SORCS2 on chromosome 4, *FASN* on chromosome 18. Also, some genes included in 10 different CNVRs found in this study, belong to the pathway for salmonella infection. The MHC region on chromosome 16, which has great interest for disease resistance, lies on a region that is in common among the CNVRs of four studies.

Online reference

- 2007-2017 "PennCNV". <http://penncnv.openbioinformatics.org/en/latest/>.
- Thomas lab at the University of Southern California. <http://pantherdb.org>. December 28, 2016.
- University of California, Santa Cruz. <https://genome.ucsc.edu/cgi-bin/hgLiftOver>. December 13, 2016.
- NAGRP - Bioinformatics Coordination Program. <http://cn.animalgenome.org/cgi-bin/QTLdb/GG/index>. December 30, 2016.
- Kanehisa, M.; "Post-genome Informatics". http://www.genome.jp/dbget-bin/www_bget?gga05132. January 1, 2017.

Additional files

Additional file 1: Sheet 1. List of CNVRs detected in this study; Sheet 2: comparison of the CNVR detected in this study with published ones; Sheet 3: annotation of CNVRs common among this and published studies; Sheet 4: annotated gene in CNVRs; sheet 5: Significant Go Term_Panther; sheet 6: QTLs mapped within CNVRs (XLSX 501 kb).

Additional file 2: Figure S1. Dendrogram of hierarchical cluster analysis of SNP genotypes (PDF 20 kb).

Abbreviations

AU-P: Approximately Unbiased *P*-value; BP-P: Bootstrap Probability value; CNV: Copy number variant; CNVR: Copy number variant region; GO: Gene ontology; HMM: Hidden Markov Model; INIFAP: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias; KEGG: Kyoto Encyclopedia of Genes and Genomes; MHC: Major histocompatibility complex; QTL: Quantitative trait loci; QTLdb: Quantitative trait loci database; UCSC: University of California Santa Cruz; UPGMA: Unweighted Pair Group Method with Arithmetic

Acknowledgments

Authors kindly acknowledge the Centro Nacional de Recursos Genéticos at Tepatlapan, Jalisco.

Funding

This study was Co-funded by project n° M01678 of the "Executive program of scientific and technical cooperation between Italian Republic and United States of Mexico" - Minister of Foreign affairs of Italy and United States of Mexico who provided economic support for the exchange of researchers among Mexico and Italy, and by internal funds of authors organisations.

Availability of data and materials

The datasets supporting the results and conclusions of our study are included within the article and in the additional files.

Authors' contributions

EG, MGS: run the data analyses, wrote the manuscript; SIRP, FJRL, VEVN, MCC, SC: collected the samples, provided the genotypes and participated in data analyses; AB, MGS: Conceived the experiment, interpreted the results,

wrote the manuscript and supervised the research group. All authors have read and approved the final version of the manuscript.

Competing interests

The authors declare they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The samples are owned by the CNRG who control their access and reuse. Original owners of individuals have donated the samples to CNRG who gave consent for re-use for research purposes.

The study did not require any ethical approval according to national rules, according to EU regulation as it does not foresee sampling from alive animals. The University of Milan permit for the use of collected samples in existing bio-banks was released with n. OPBA-56-2016.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Veterinary Medicine, Università degli Studi di Milano, Via Celoria 10, 20133 Milan, Italy. ²Centro Nacional de Investigación en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícola y Pecuarias (INIFAP), Km.1 Carretera a Colón, Auchtitlán, 76280 Querétaro, CP, Mexico. ³Centro Nacional de Investigación en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícola y Pecuarias (INIFAP), Melchor Ocampo # 234 Desp. 313, Col. Centro Veracruz, C.P. 91700 Veracruz, Mexico.

Received: 5 April 2017 Accepted: 12 June 2017

Published online: 03 July 2017

References

1. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011;470:59–65.
2. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7:85–97.
3. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, et al. Copy number variation: new insights in genome diversity. *Genome Res*. 2006;16:949–61.
4. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 2004;432:695–716.
5. Al-Qamashouy B, Simianer H, Kadim I, Weigend S. Assessment of genetic diversity and conservation priority of Omani local chickens using microsatellite markers. *Trop Anim Health Prod*. 2014;46:747–52.
6. Strillacci MG, Marelli SP, Cozzi MC, Colombo E, Polli M, Gualtieri M, et al. Italian autochthonous chicken breeds conservation: evaluation of biodiversity in Valdarnese Bianca breed (*Gallus gallus Domesticus*). *Avian Biol Res*. 2009;2:229–33.
7. Ceccobelli S, Di Lorenzo P, Lancioni H, Monteagudo Ibáñez LV, Tejedor M, Castellini C, et al. Genetic diversity and phylogeographic structure of sixteen Mediterranean chicken breeds assessed with microsatellites and mitochondrial DNA. *Livest Sci* 2015; 175: 27–36.
8. Zhang F, Gu W, Hurlles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10: 451–81.
9. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet*. 2007;39:37–42.
10. Wang X, Nahashon S, Feaster TK, Bohannon-Stewart A, Adefope N. An initial map of chromosomal segmental copy number variations in the chicken. *BMC Genomics*. 2010;11:351.
11. Yalcin B, Wong K, Agam A, Goodson M, Keane TM, Gan X, et al. Sequence-based characterization of structural variation in the mouse genome. *Nature*. 2011;477:326–9.
12. 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;444:444–54.
13. 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;464:704–12.
14. Yang Z, Zhuang B, Yan Y, Jiang S, Wang T. Integrated analyses of copy number variations and gene differential expression in lung squamous-cell carcinoma. *Biol Res*. 2015;48(1):47.
15. Segura-Correa JC, Sarmiento-Franco L, Magaña-Monforte JG, Santos-Ricalde R. Productive performance of Creole chickens and their crosses raised under semi-intensive management conditions in Yucatan, Mexico. *Br Poult Sci*. 2004;45(3):342–5.
16. Rodriguez JC, Allaway CE, Wassink GJ, Segura JC, Rivera T. Estudio de la Avicultura de traspatio en el municipio de Dzunuzcán. *Yucatán Vet Mex*. 1996;27(3):215–9.
17. Segura-Correa JC, Juarez-Caratachea A, Sarmiento-Franco L, Santos-Ricalde R. Growth of Creole chickens raised under tropical conditions of Mexico. *Trop Anim Health Prod*. 2005;37(4):327–32.
18. Ceccobelli S, Lorenzo PD, Lancioni H, Castellini C, Ibáñez LV, Sabbioni A, et al. Phylogeny, genetic relationships and population structure of five Italian local chicken breeds. *Italian J AnimSci*. 2013;12(3):e66.
19. Zhang H, Du ZQ, Dong JQ, Wang HX, Shi HY, Wang N, et al. Detection of genome-wide copy number variations in two chicken lines divergently selected for abdominal fat content. *BMC Genomics*. 2014;15:517.
20. Crooijmans RP, Fife MS, Fitzgerald TW, Strickland S, Cheng HH, Kaiser P, et al. Large scale variation in DNA copy number in chicken breeds. *BMC Genomics*. 2013;14:398.
21. Tian M, Wang Y, Gu X, Feng C, Fang S, Hu X, et al. Copy number variants in locally raised Chinese chicken genomes determined using array comparative genomic hybridization. *BMC Genomics*. 2013;14:262.
22. Han R, Yang P, Tian Y, Wang D, Zhang Z, Wang L, et al. Identification and functional characterization of copy number variations in diverse chicken breeds. *BMC Genomics*. 2014;15:934.
23. Jia X, Chen S, Zhou H, Li D, Liu W, Yang N. Copy number variations identified in the chicken using a 60K SNP BeadChip. *Anim Genet*. 2012;44: 276–84.
24. Yi G, Qu L, Liu J, Yan Y, Xu G, Yang N. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genomics*. 2014;15(1):962.
25. Yan Y, Yang N, Cheng HH, Song J, Qu L. Genome-wide identification of copy number variations between two chicken lines that differ in genetic resistance to Marek’s disease. *BMC Genomics*. 2015;16:843.
26. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant S, 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.
27. Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 2006; 16: 1136–1148.
28. Durán Aguilar M, Román Ponce SI, Ruiz López FJ, González Padilla E, Vásquez Peláez CG, Bagnato A, et al. Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *J Anim Breed Genet*. 2017;134(1):49–59.
29. Xu L, Hou Y, Bickhart DM, Song J, Liu GE. Comparative analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP data. *Microarrays*. 2013;2(3):171–85.
30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
31. Gazave E, Darré F, Morcillo-Suarez C, Petit-Marty N, Carreño A, Marigorta UM, et al. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. *Genome res*. 2011;21(10):1626–39.
32. Suzuki R, Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 2006;12:1540–2.
33. Fulton JE, McCarron AM, Lund AR, Pingear KN, Wolc A, Chazara O, et al. A high-density SNP panel reveals extensive diversity, frequent recombination and multiple recombination hotspots within the chicken major histocompatibility complex B region between BG2 and CD1A1. *Genet Sel Evol*. 2016;48:1.
34. Wang X, Byers S. Copy number variation in chickens: a review and future prospects. *Microarrays*. 2014;3:24–38.
35. Garcia-Camacho L, Schat KA, Brooks R Jr, Bounous DI. Early cell-mediated immune responses to Marek’s disease virus in two chicken lines with

- defined major histocompatibility complex antigens. *Vet Immunol Immunopathol.* 2003;95(3):145–53.
36. Strillacci MG, Cozzi MC, Gorla E, Mosca F, Schiavini F, Román-Ponce SI, et al. Genomic and genetic variability of six chicken populations using single nucleotide polymorphism and copy number variants as markers. *Animal.* 2017;11(5):737–45.
 37. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genet.* 2007;39(12):1443–52.
 38. Metzger J, Karwath M, Tonda R, Beltran S, Águeda L, Gut M, et al. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics.* 2015;16(1):764.
 39. Lamont SJ, Coble DJ, Bjorkquist A, Rothschild MF, Persia M, Ashwell C, et al. Genomics of heat stress in chickens. Proceedings, 10th World Congress of Genetics Applied to Livestock Production. Vancouver, BC, Canada. August 17–22, 2014.
 40. Herzog KH, Bailey K, Barde YA. Expression of the BDNF gene in the developing visual system of the chick. *Development.* 1994;120(6):1643–9.
 41. Styrkarsdottir U, Thorleifsson G, Sulem P, Gudbjartsson DF, Sigurdsson A, Jonasdottir A, et al. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature.* 2013;497(7450):517–20.
 42. Li Z, Zheng M, Abdalla BA, Zhang Z, Xu Z, Ye Q, et al. Genome-wide association study of aggressive behaviour in chicken. *Sci Rep.* 2016;6:30981.
 43. Gu X, Feng C, Ma L, Song C, Wang Y, Da Y, et al. Genome-wide association study of body weight in chicken F2 resource population. *PLoS One.* 2011; 6(7):e21872.
 44. Zhang J, Liu F, Cao J, Liu X. Skin Transcriptome profiles associated with skin color in chickens. *PLoS One.* 2015;10(6):e0127301.
 45. Wright D, Boije H, Meadows JR, Bed'Hom B, Gourichon D, et al. Copy number variation in intron 1 of SOX5 causes the pea-comb phenotype in chickens. *PLoS Genet.* 2009;5(6):e1000512.
 46. Sewalem A, Morrice DM, Law A, Windsor D, Haley CS, Ikeobi CO, et al. Mapping of quantitative trait loci for body weight at three, six, and nine weeks of age in a broiler layer cross. *Poultry Sci.* 2002;81(12):1775–81.
 47. Carlborg Ö, Hocking PM, Burt DW, Haley CS. Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. *Genet Res.* 2004;83(03):197–209.
 48. D'Andre HC, Paul W, Shen X, Jia X, Zhang R, Sun L, Zhang X. c. *J Anim Sci Biotechnol.* 2013; 4(1): 43.
 49. Wang Z, Chen Q, Yang Y, Liao R, Zhao J, Zhang Z, et al. Genetic diversity and population structure of six Chinese indigenous pig breeds in the Taihu Lake region revealed by sequencing data. *Anim Genet.* 2015;46(6):697–701.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

