

Genomic and genetic variability of six chicken populations using single nucleotide polymorphism and copy number variants as markers

M. G. Strillacci^{1†}, M. C. Cozzi¹, E. Gorla¹, F. Mosca¹, F. Schiavini², S. I. Román-Ponce³, F. J. Ruiz López³, A. Schiavone⁴, M. Marzoni⁵, S. Cerolini¹ and A. Bagnato¹

¹Department of Veterinary Medicine, Università Degli Studi di Milano, via Celoria 10, 20133 Milano, Italy; ²Department of Health, Animal Science and Food Safety (VESPA), Università degli Studi di Milano, via Celoria 10, 20133 Milano, Italy; ³Centro Nacional de Investigación en Fisiología y Mejoramiento Animal, INIFAP, Km.1 Carretera a Colón, Auchtitlán, 76280 Querétaro, México; ⁴Department of Veterinary Science, Università degli Studi di Torino, Largo Paolo Braccini 2, 10095 Grugliasco, Italy; ⁵Department of Veterinary Science, Università degli Studi di Pisa, Viale delle Piagge 2, 56124 Pisa, Italy

(Received 3 May 2016; Accepted 9 September 2016)

Genomic and genetic variation among six Italian chicken native breeds (Livornese, Mericanel della Brianza, Milanino, Bionda Piemontese, Bianca di Saluzzo and Siciliana) were studied using single nucleotide polymorphism (SNP) and copy number variants (CNV) as markers. A total of 94 DNA samples genotyped with Axiom® Genome-Wide Chicken Genotyping Array (Affymetrix) were used in the analyses. The results showed the genetic and genomic variability occurring among the six Italian chicken breeds. The genetic relationship among animals was established with a principal component analysis. The genetic diversity within breeds was calculated using heterozygosity values (expected and observed) and with Wright's F-statistics. The individual-based CNV calling, based on log R ratio and B-allele frequency values, was done by the Hidden–Markov Model (HMM) of PennCNV software on autosomes. A hierarchical agglomerative clustering was applied in each population according to the absence or presence of definite CNV regions (CNV were grouped by overlapping of at least 1 bp). The CNV map was built on a total of 1003 CNV found in individual samples, after grouping by overlaps, resulting in 564 unique CNV regions (344 gains, 213 losses and 7 complex), for a total of 9.43 Mb of sequence and 1.03% of the chicken assembly autosome. All the approaches using SNP data showed that the Siciliana breed clearly differentiate from other populations, the Livornese breed separates into two distinct groups according to the feather colour (i.e. white and black) and the Bionda Piemontese and Bianca di Saluzzo breeds are closely related. The genetic variability found using SNP is comparable with that found by other authors in the same breeds using microsatellite markers. The CNV markers analysis clearly confirmed the SNP results.

Keywords: single nucleotide polymorphism, copy number variation, poultry, biodiversity, genetic variability

Implications

The aim of this study was to assess the genetic diversity of six Italian chicken breeds in order to define the status of *in situ* genetic collections and study their conservation potential. The genetic and genomic structure of the six Italian native chicken populations reported here will contribute to design coherent programmes for *in vivo* and *in vitro* conservation, valorisation and utilisation of the breeds. As these breeds represent a unique animal resource, these findings will impact the economic value and environmental sustainability of traditional food production.

Introduction

Genetic makeup of populations is the result of a long-term process of adaptation to specific environments and

ecosystems and, of artificial selection. Local populations are usually well adapted to environment and capable to express optimal functionality of life cycle events, as reproduction and resistance to diseases despite environmental challenges and, at the same time, to exhibit a good food production (i.e. meat and eggs).

The Food and Agricultural Organization of United Nation (FAO) definition of animal genetic resources eligible for conservation includes animal populations with economic potential, scientific and cultural interest (Food and Agricultural Organization (FAO) of United Nation, 2009). In most of the World about 50% of documented breeds have been classified as extinct, at critical survival or endangered (Hammond, 1996); furthermore 31% of cattle breeds, 35% of pig breeds and 38% of chicken breeds are at risk of extinction. In addition, especially in poultry, local breeds have often been diluted by indiscriminate cross-breeding with imported stocks (FAO of United Nation, 2009).

[†] Email: maria.strillacci@unimi.it

As a consequence, the conservation of domestic animal biodiversity has become a priority to develop sustainable, safe and diversified products and production systems. Considering that the 68% of the 53 Italian chicken breeds were classified as extinct (Zanon and Sabbioni, 2001), efforts for conservation of the remaining local populations are urgently required. Recently, national initiatives (Mosca *et al.*, 2015) have been undertaken in Italy to characterise local populations for resilience and for the nutritional properties of their primary production used as basis of regional food products often related to gastronomic traditions.

In the last decades, microsatellite markers have been used to perform phylogenetic analysis and studies on genetic variability in the chicken breeds (Strillacci *et al.*, 2009; Al-Qamashoui *et al.*, 2014; Ceccobelli *et al.*, 2015). The availability of high-density single nucleotide polymorphism (SNP) arrays has opened the possibility to investigate the genetic structure of a population on a very large number of markers having uniform distribution on all chromosomes. Moreover, these arrays permit to identify and map copy number variants (CNV) on the genome. CNV are distributed over the whole genome in all species, and are defined as large-scale genome mutations ranging from 50 bp to several Mb (Mills *et al.*, 2011) compared with a reference genome (insertions, deletions and more complex changes). Involving large genomic regions, CNV may affect gene structure and determine expression and/or regulation gene changes (Redon *et al.*, 2006). Although CNV were recently mapped in several livestock species (Han *et al.*, 2014; Schiavo *et al.*, 2014; Bagnato *et al.*, 2015), their use as markers to explain intra-breeds genetic diversity has been explored only in few species (Gazave *et al.*, 2011; Xu *et al.*, 2016).

The aim of this study was to analyse the genomic and genetic variation in order to describe the existing variability among individuals of six Italian chicken breeds using both SNP and CNV as markers. We will then test the hypothesis that genetic variation exists among the six breeds considered in this study, highlighting that the new knowledge gained thanks to high throughput genotyping (SNP, CNV) strongly contribute to the characterisation of genetic diversity among them. The knowledge of the genetic structure of these breeds may be used to preserve the genetic variability and the phenotypic features peculiar of each population.

Material and methods

Sampling and genotyping

In this study, six Italian chicken breeds were used: Livornese (LI) from Tuscany, Milanino (MI) and Mericanel della Brianza (MB) from Lombardy, Bionda Piemontese (PI) and Bianca di Saluzzo (SA) from Piedmont and Siciliana (SI) from Sicily (Supplementary Tables S1 and S2). All the populations are ancient Italian breeds except the composite MI. The MB is the only Italian bantam breed, with an official recognised standard.

In total, 96 blood samples (16 per breed) were randomly selected among blood bio-banks (stored in 0.5 M ethylenediaminetetraacetic acid at -20°C) representative of flock

nucleus conserved within the universities of Milano, Torino and Pisa. Genomic DNA was isolated using the NucleoSpin[®] Blood kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) according to the manufacturer's instructions. DNA concentration was determined with the Qubit[®] dsDNA HS Assay kit (Invitrogen[™], Carlsbad, CA, USA) using the proper Qubit[®] fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA); purity was assessed through the evaluation of A260/280 and A260/230 ratios on the Infinite[®] 200 PRO NanoQuant spectrophotometer (Tecan Trading AG, Switzerland) and integrity verified running samples E-Gel[®] 48 Agarose Gels, 1% (Invitrogen).

All DNA samples were genotyped using the Axiom[®] Genome-Wide Chicken Genotyping Array (Affymetrix, Thermo Fisher Scientific Inc.) including 580 961 SNP markers, distributed across the genome with an average spacing of 1.7 Kb (galGal4 assembly). Axiom[™] Analysis Suite software (Affymetrix, Thermo Fisher Scientific Inc.) was used to run raw intensity data Quality Control and Genotyping Algorithms. Default quality control settings were applied to filter for low-quality samples before running the genotyping analysis. Axiom CNV summary tool was used to generate input files for CNV prediction analysis software.

Single nucleotide polymorphism analyses

SNP allele frequencies, expected (H_e) and observed (H_o) heterozygosity were computed separately for each breed using the PEAS software (Xu *et al.*, 2010). Genetic diversity within and among breeds was determined estimating the Wright's *F*-statistics fixation index (F_{ST}) and inbreeding coefficient of an individual relative to a subpopulation (F_{IS}) on SVS Golden Helix software 8.3.1 (Golden Helix Inc., Bozeman, MT, USA) (SVS). The genetic structure of Italian chicken populations was analysed using:

- (i) SVS: principal component analysis (PCA) based on SNP allele frequencies.
- (ii) ADMIXTURE ver. 1.3.0 software (Alexander *et al.*, 2009): population structure analysis with a number of ancestral populations *K* that ranged from two to eight. To evaluate optimal number of ancestors, cross-validation error values were computed for each *K* using a fivefold cross-validation procedure, as reported by Nicoloso *et al.* (2015). Each inferred chicken population structure was visualised using an R script.
- (iii) PEAS software: individual tree using Neighbour-Joining (NJ) algorithm. The NJ tree, constructed based on the allele sharing distance as the genetic distance between individuals, was graphically represented using FigTree version 1.4.2 (Rambaut, 2014).

Copy number variants and copy number variants region analyses

Both the log *R* ratio (LRR) and the B-Allele Frequency (BAF) values of each sample were obtained from the Axiom[®] CNV Summary Tool software. LRR and BAF were used in the individual-based CNV calling performed by PennCNV

software (Wang *et al.*, 2007) on chromosomes 1 to 28, using the default parameters of the HMM: SD of LRR <0.30, BAF drift as 0.01 and waviness factor at 0.05. The CNV regions (CNVR) were defined in each breed using the BedTools software, through merging overlapping CNV by at least 1 bp, as described by Redon *et al.* (2006).

Clustering analysis using copy number variants region. A clustering analysis for all samples was performed considering the identified CNVR as genetic markers (Tian *et al.*, 2013). A scoring matrix of the CNVR data was built by encoding a value of '0' or '1' according to the absence or presence for each individual of any mapped CNV in the pertinent CNVR. A hierarchical agglomerative clustering was applied on the scoring matrix using the pvclust function from the pvclust R package (Suzuki and Shimodaira, 2006). Multiscale bootstrap resampling was applied to calculate the Approximately Unbiased *P*-value (AU-*P*) using 10 000 bootstraps to assess the robustness of branches. Agglomerative method chosen was Unweighted Pair Group Method with Arithmetic mean.

Results and discussion

Single nucleotide polymorphism analyses

SNP analyses and the CNV detection were performed on 94 quality-filtered samples, as two samples belonging to MB and PI breeds were discarded for low raw signal intensity. SNP with minor allele frequency value ≤ 0.01 , SNP with Hardy–Weinberg equilibrium ≤ 0.00001 , SNP not on first 28 autosomal chromosomes and SNP having a call rate <99% were excluded, reducing to 412 336 SNP markers the number of loci used in the analysis. The number of polymorphic sites within breed ranged from 197 099 (47.8%) to 383 086 (92.8%) for SI and SA, respectively (Table 1).

For each breed, the effective number of polymorphic SNP (number of SNP in which at least one heterozygous individual was identified) represents >99% of polymorphic sites (Table 1). The H_o and H_e ranged from 0.210 and 0.170 (SI) to 0.345 and 0.320 (SA), whereas the F_{IS} values ranged from -0.192 (SI) to 0.094 (LI). The SI H_o and H_e values (0.210; 0.170) reflect the highest percentage of monomorphic SNP (52.2%) and the low variability within the breed. On the contrary, the SA breed has a low F_{IS} value (-0.045) and the

highest H_o and H_e values confirming results previously obtained by Sartore *et al.* (2016) using microsatellite markers.

In the LI breed, despite the high percentage of polymorphic SNP (75.9%), the H_o and H_e values are quite low (0.232 and 0.249), although the F_{IS} value (0.094) indicates a low level of inbreeding. Ceccobelli *et al.* (2015) reported for the same breed similar F_{IS} value and higher H_o and H_e values obtained using microsatellites data. The low genetic variability measured in both LI and SI birds is suggested to be related to the small size of the population under conservation for many years, situation generally known to be associated with relevant value of inbreeding. The H_o and H_e values for the bantam breed MB (0.243 and 0.221) are lower than those obtained by Tadano *et al.* (2008) on Japanese bantam breeds using a panel of 40 microsatellites. The F_{IS} value for MB (-0.060) is very low and quite similar to that identified in the Japanese Bantam breed Tosa-Jidori (Tadano *et al.*, 2008).

Except for LI and PI, the negative F_{IS} values detected in all other breeds reflect an excess (increasing) of heterozygosity, probably due to outbreeding (Tadano *et al.*, 2007). The heterozygous SNP were classified into three classes according to the number of individuals resulted heterozygous at the same locus: '01 to 05', '06 to 10' and '11 to 16'. In fact, for MB, MI, SA and SI breeds (MB = 4.7%, MI = 6.6%, SA = 6% and SI = 7.4%) respect to LI and PI (LI = 2.1%, PI = 1.8%) a higher proportion of SNP were heterozygous in >10 samples (class '11 to 16') (Figure 1). The same distribution applies for class of individuals '6 to 10'. On the contrary, if we consider the class of individuals '1 to 5' the two breeds LI and PI are those with the largest proportion of heterozygous SNP. This behaviour in SNP heterozygous loci agree well with the F_{IS} values found here.

The pairwise fixation indexes (F_{ST}) among the six Italian chicken breeds are presented in Figure 2. The F_{ST} values range from 0.082 (PI *v.* SA) to 0.439 (SI *v.* MB). The largest differences were between the SI breed and the other populations, with F_{ST} values ranging from 0.290 (SA) to 0.439 (MB). The F_{ST} values >0 can be related to the effect of genetic isolation respect to the other populations, which can lead to homozygous excess over time. As expected by their origin (i.e. same geographical region), the PI and SA breeds are closely related ($F_{ST} = 0.082$) and their F_{ST} values against the other populations are very low. Sartore *et al.* (2016) considered PI as the ancestral population of the present day

Table 1 Single nucleotide polymorphism (SNP) statistics, observed heterozygosity (H_o), expected heterozygosity (H_e) and inbreeding coefficient (F_{IS}) values for the six Italian chicken populations

Breeds	Size	No. of polymorphic SNP	No. of heterozygote SNP	H_o	H_e	F_{IS}
LI	16	312 823	310 782	0.232	0.249	0.094
MB	15	263 920	262 346	0.243	0.221	-0.060
MI	16	270 881	270 039	0.258	0.237	-0.055
PI	15	366 337	364 921	0.312	0.304	0.008
SA	16	383 086	382 286	0.345	0.320	-0.045
SI	16	197 099	196 845	0.210	0.170	-0.192

LI = Livornese; MB = Mericanel della Brianza; MI = Milanino; PI = Bionda Piemontese; SA = Bianca di Saluzzo; SI = Siciliana.

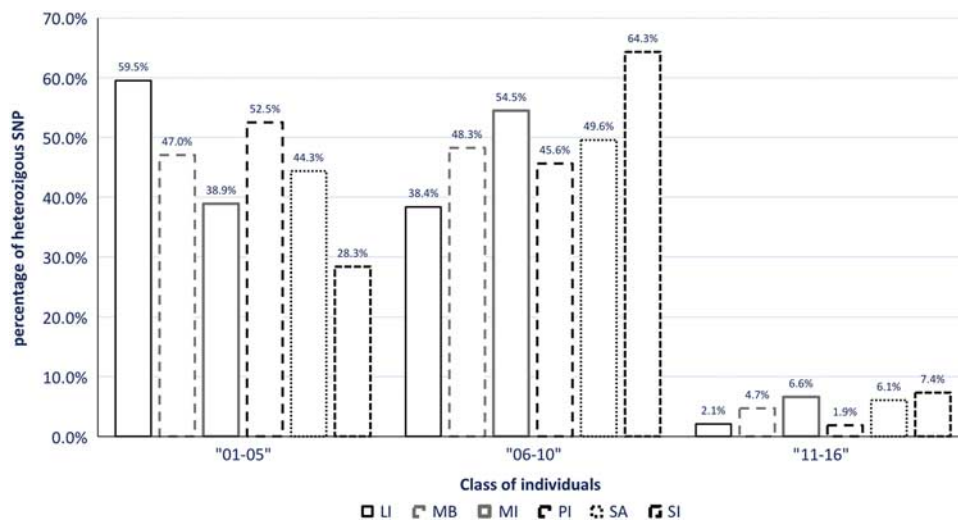


Figure 1 Proportion of heterozygous single nucleotide polymorphism (SNP) classified into three classes according to the number of individuals resulted heterozygous at the same locus: '01-05', '06-10' and '11-16'. LI = Livornese, MB = Mericanel della Brianza, MI = Milanino, PI = Bionda Piemontese, SA = Bianca di Saluzzo and SI = Siciliana.

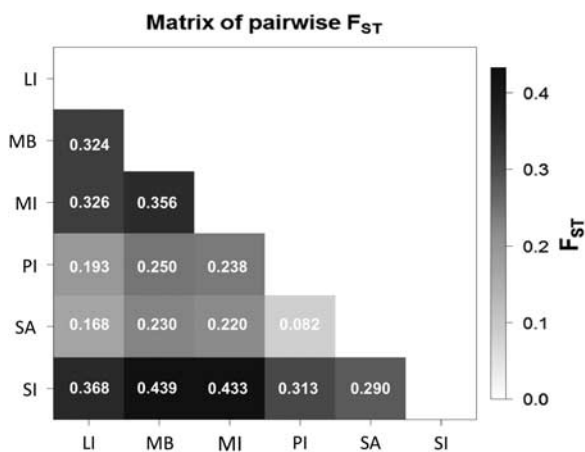


Figure 2 Matrix of pairwise fixation index F_{ST} among the six Italian chicken breeds. LI = Livornese, MB = Mericanel della Brianza, MI = Milanino, PI = Bionda Piemontese, SA = Bianca di Saluzzo and SI = Siciliana.

SA. These authors also report a similar F_{ST} value for the same breeds using a panel of 32 microsatellite markers.

The MI breed is relatively similar to PI and SA and differs from all other breeds in terms of genetic structure (Figure 2). The bantam breed MB differs from the MI and LI breeds ($F_{ST} = 0.356$ and $F_{ST} = 0.324$), but is relatively similar to the Piedmont PI and SA breeds ($F_{ST} = 0.250$ and $F_{ST} = 0.230$). MB is a very common breed in north-east of Milan, and it is still not possible to determine the period in which this breed appeared. The breed anyhow is reported to derive from dwarf rural chickens diffused in small rural farms at the beginning of last century (Ceppolina, 2015).

The overall F_{ST} value found across all breeds is 0.253, indicating that 25.3% of the genetic variation is explained by the breed differences, whereas the remaining 74.7% of the variance describes the differences among individuals. This value is >0.15 considered by Frankham *et al.* (2004) as an

indicator of significant differentiation among populations. The genetic variability of local breeds here highlighted must be considered an important genetic resource as indicated by Muir *et al.* (2008). In fact, they reported in a recent analysis using SNP markers, that commercial pure line showed a substantial decrease of genetic diversity compared with non-commercial chicken populations.

The overall F_{ST} value identified here is similar to the previous reported using microsatellites markers in commercial chicken lines (Tadano *et al.*, 2007), British (Wilkinson *et al.*, 2011) and Mediterranean chicken breeds (Ceccobelli *et al.*, 2015). In contrast, lower F_{ST} values were reported in Japanese, Italian and Swedish local populations (Tadano *et al.*, 2008; Zanetti *et al.*, 2010; Abebe *et al.*, 2015). The higher chicken F_{ST} values, highlighted the larger genetic variability of chicken populations, respect to the one found in other livestock species. For instance, Wang *et al.* (2015) reported a F_{ST} value of 0.149 in Chinese pig breeds and Makina *et al.* (2014) a F_{ST} value of 0.149 in South Africa cattle breeds.

The PCA (Figure 3a) disclosed genetic differences among the six breeds and show that all individuals are well clustered by breed. The canonical variable plotted on the y-axis explained 1.93% of the overall SNP variance. On this axis, the LI breed is clearly separated in two different groups according to bird's feather colour (black upper group and white lower group) as well as the PI and SA breeds create two separated clusters closely related. The origin of LI breed is not so clear, probably from Central Italy, obtained from the selection of light chicken reared in Tuscany region. LI is worldwide spread with different colours of livery: black, white and brown (light and dark) (Ceppolina, 2015) and selected according to colour differences for decades.

The distinction among breeds was clearly displayed on the canonical variable plotted as x-axis representing 7.18% of the SNP variance. The SI breed is a distinct group, confirming results of F_{ST} values. In fact, this breed appears to derive

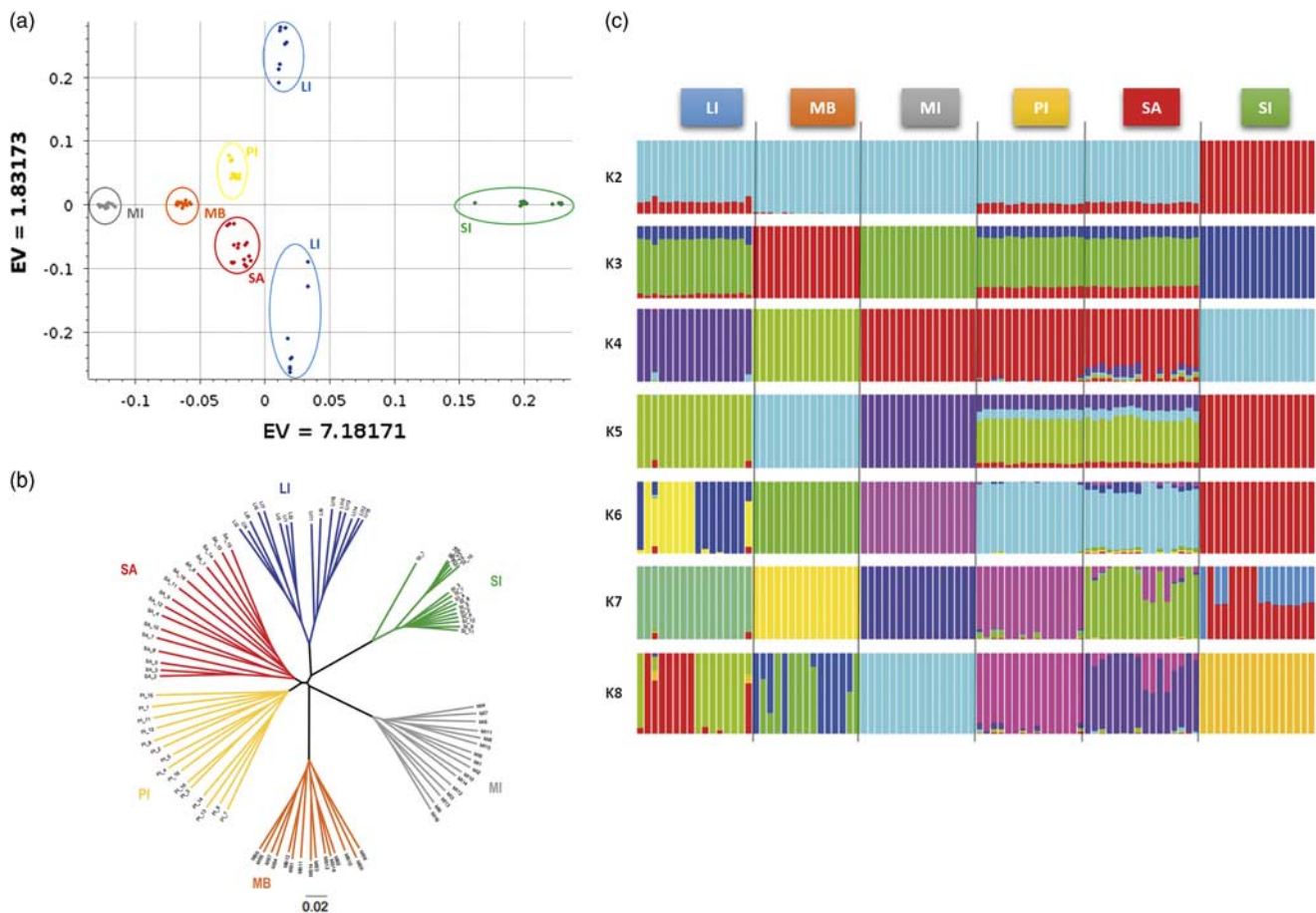


Figure 3 Population genetic analyses of the six Italian chicken breeds (LI = Livornese, MB = Mericanel della Brianza, MI = Milanino, PI = Bionda Piemontese, SA = Bianca di Saluzzo and SI = Siciliana): (a) Scatter plot (EV = Eigenvalues of canonical variables) from a principal component analysis based on SNP frequencies. (b) Neighbour-Joining dendrogram constructed using genetic sharing distances. (c) Admixture plot for all Italian chicken breeds analysed based on different number of assumed ancestors (K).

from ancient inter-breeding of local Sicilian birds with North African stock (Ceppolina, 2015). The PCA plot shows the division of SI samples in three sub-groups. The major distance was identified between MI and SI breeds.

The results of the NJ analysis (Figure 3b), are consistent with those obtained by the PCA. The NJ dendrogram suggests the presence of three distinct clusters: cluster 1 includes the closely related PI and SA breeds (originating in Piedmont), cluster 2 includes the two varieties of LI breed and SI and cluster 3 includes MI and MB breeds (originating in Lombardy).

An increasing number of assumed ancestors, from $K = 2$ to 8 was used for global admixture analysis done by the ADMIXTURE software. The graphical representation of the estimated ancestor fractions in individual genomes is shown in Figure 3c. In fact, at $K = 2$ two distinct ancestors are represented by SI and MB + MI, whereas LI, PI and SA genomes seem to include a major fraction of the MB + MI ancestor and a minor fraction the SI ancestor. $K = 3$ and 4 split MB from MI, and the above three composite breeds now had a major MI and minor MB and SI ancestor components. A similar albeit more complicated figure was kept by $K = 5$. Based on agreement with the PCA and CNV analyses,

the ADMIXTURE software identified $K = 6$ as the most probable number of common ancestors of our samples. At $K = 6$, MI, MB and SI breeds grouped again into independent ancestors, and the LI breed appears to be divided into two genetically distinguishable sub-groups, confirming both PCA results and CNV cluster analysis.

Independently of the K number, individuals belonging to the PI and SA breeds seem to share the same ancestors composition, but when K increased to seven they separated in two distinct groups, while retaining some common genetic features. At $K = 8$ almost all breeds (except for MB) returned to show the same genetic features identified at smaller K s. It is interesting to note that all the grouping strategies identify the MI breed as distinct from the other genetic groups: this is representative of the selection history of the breed initiated at the beginning of 20th century by crossing Valdarnese Bianca males to Horington females (Mosca *et al.*, 2015).

Copy number variants and copy number variants region analyses

In Table 2, the frequency of CNV identified, the mean and median values, as well as the CNV coverage per each breed

Table 2 Descriptive statistics of copy number variant (CNV) identified for each breed

Breeds	No. CNV (min to max)*	No. losses (state 0/1)	No. gains (state 3/4)	CNV min (length (bp))	CNV max (length (bp))	CNV mean (length (bp))	CNV median (length (bp))	Coverage (bp)	Coverage (%)
LI	159 (3 to 17)	97	62	160	265 647	17 919.37	6535	2 849 180	0.31
MB	110 (5 to 10)	75	35	462	240 256	17 587.3	6381	1 934 603	0.21
MI	131 (4 to 29)	69	62	381	171 360	15 032.57	6133	1 969 267	0.21
PI	211 (6 to 28)	131	80	52	356 281	19 241.97	8497	4 060 057	0.44
SA	131 (5 to 11)	41	90	258	384 766	35 254.32	13 306	4 618 316	0.50
SI	261 (7 to 46)	143	118	213	119 253	16 262.30	7910	4 244 461	0.46
Total	1003	556	447	52	384 766	19 617.03	7380	19 675 884	2.14

LI = Livornese; MB = Mericanel della Brianza; MI = Milanino; PI = Bionda Piemontese; SA = Bianca di Saluzzo; SI = Siciliana.

*Min to max = minimum and maximum number of CNV for individual.

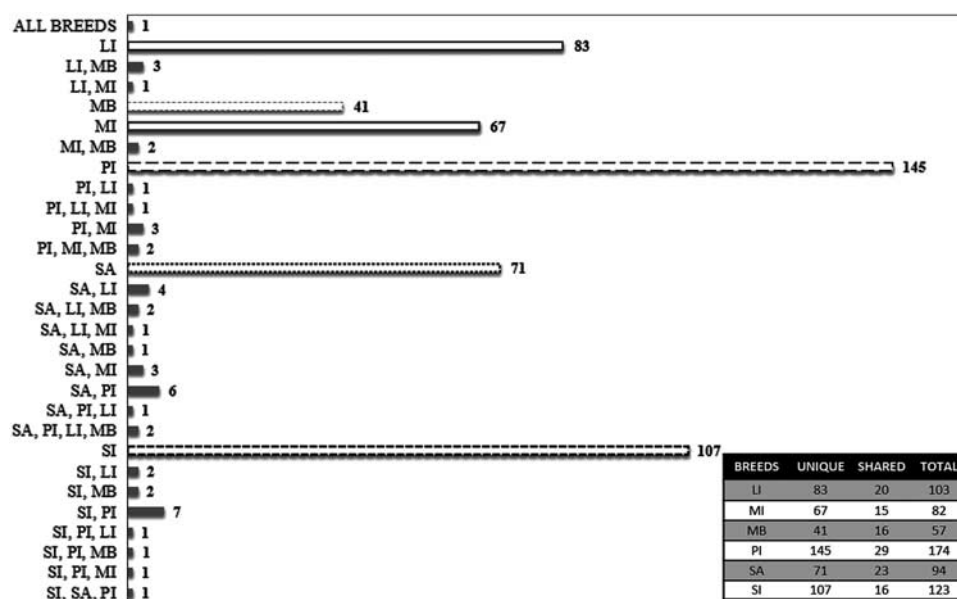


Figure 4 Intra (unique) and inter (shared) breed variation of copy number variant region in the six Italian chicken populations (LI = Livornese, MB = Mericanel della Brianza, MI = Milanino, PI = Bionda Piemontese, SA = Bianca di Saluzzo and SI = Siciliana).

compared with the chicken assembly autosomes are reported. In all breeds, the number of losses (state 0 and 1) is higher than the number of gains (state 3 and 4), except for the SA breed. This is indicated by the deletions/duplications ratios calculated as the total number of losses divided by number of gains: 1.56, 2.14, 1.11, 1.63, 1.12 and 0.45 for LI, MB, MI, PI, SI and SA, respectively. The majority of CNV (i.e. 91% among all breeds) identified in this study, have a length between 1 and 100Kb representing a proportion over the total number of CNV of 87.7% in the MI to 95.4% in the SI.

A total of 564 unique CNVR (344 gains, 213 losses and seven complex) were found among all breeds. These CNVR covered a total of 9.43 Mb of sequence length corresponding to 1.03% of the chicken galGal4 assembly autosome. The total number of CNVR detected for each breed is 103 in LI, 57 in MB, 82 in MI, 174 in PI, 94 in SA and 123 in SI (Figure 4 and Supplementary Table S3). Table 3 shows the number of CNVR for each breed by chromosome. With the exception of

chr21 and chr24, which contain CNVR identified only in two breeds (LI-PI and MI-PI, respectively), all other autosomes include CNVR from at least three breeds. CNVR on chromosomes 1, 2, 3, 4, 5, 8, 12, 14, 16 and 20 have been identified in all breeds. In the PI breed, the identified CNVR map on all chromosomes, with the exception of the chr26, whereas the CNVR identified in the LI breed are distributed on only 12 autosomes.

Among the identified CNVR, 426 (75%) were present in a single individual (singleton), 61 (10%) in two individuals, 23 (4%) in three individuals, 14 (2%) in four individuals and 40 (7%) in more than five individuals. The high proportion of the singleton has been previously reported by Yi *et al.* (2014) (68.8%) and by Han *et al.* (2014) (76.5%), confirming that segregating CNV exist among individuals. The CNVR on chr16 at 215 410 to 330 020 bp was identified in 31 samples across all six chicken breeds (at least two samples per breed) as well as in chicken populations analysed by the latter above-cited authors.

Table 3 Descriptive statistics of copy number variant region (CNVR) identified for each breed by chromosome (CHR)

CHR	Breeds					
	LI	MB	MI	PI	SA	SI
1	24	15	24	40	21	24
2	15	9	9	20	14	20
3	5	9	11	14	5	15
4	6	4	4	15	6	9
5	8	3	6	11	8	12
6	3	0	1	9	3	4
7	3	0	1	3	7	5
8	3	1	2	3	1	2
9	3	2	4	4	0	4
10	1	0	2	5	4	0
11	1	0	2	3	4	3
12	2	2	1	5	2	2
13	5	2	2	5	2	0
14	3	2	3	5	3	2
15	3	0	2	3	0	0
16	1	1	1	1	1	2
17	3	0	0	3	1	0
18	0	2	1	3	1	3
19	1	0	1	2	3	4
20	1	1	1	3	1	1
21	4	0	0	1	0	0
22	0	1	0	2	1	1
23	2	0	1	2	1	2
24	0	0	1	2	0	0
25	1	0	0	3	0	5
26	1	1	1	0	2	3
27	2	2	1	3	2	0
28	2	0	0	4	1	1
Total	103	57	82	174	94	124

LI = Livornese; MB = Mericanel della Brianza; MI = Milanino; PI = Bionda Piemontese; SA = Bianca di Saluzzo; SI = Siciliana.

Comparison of the CNVR in the six breeds (Figure 4) reveals that the number of CNVR shared among the breeds ranged from 15 (MI *v.* others) to 29 (PI *v.* others), whereas the number of intra-breed shared CNVR (mainly contributed by single sample variations) ranged from 41 (MB) to 145 (PI). Considering the CNVR identified by CNV common to individuals of different breeds, the most frequent combinations are: SI-PI ($n = 7$) and SA-PI ($n = 6$). Adding to these combinations those including other breeds, it gives a total of 11 and 10 CNVR common to SI-PI and SA-PI, respectively (Figure 4).

Despite recent studies on CNV in chicken have showed their role in metabolic pathways and their association with innate and adaptive immunity, morphological traits, developmental defects or disease susceptibility (Wang *et al.*, 2014; Yan *et al.*, 2015), the actual knowledge on CNV and their full role in the genomic expression is still limited and do not permit to understand the specific function of CNV here found.

Figure 5 shows the cluster-tree built for the six chicken breeds based on CNVR similarities. In the plot, the branch

length is not directly proportional to the genetic distance estimated among samples. The AU-P and Bootstrap Probability value (BP-P) were shown for each node, as well as the edge numbers. We focussed on the AU-P because the BP-P is considered less accurate than AU-P and according to Suzuki and Shimodaira (2006) the cluster (edges) with AU-P >95% are the most plausible.

Edge numbers represent the order in which the clusters were built. More closely related samples have a smaller edge numbers, whereas higher edge numbers reflect clusters formed later in the breed evolutionary process. As shown in the plot, all samples of SI and almost all samples of MB were assigned to a single breed-cluster. The MI and LI samples are grouped in two distinct clusters each. Instead for PI and SA breeds, three and four clusters were identified respectively, two of which include samples belonging to both breeds.

Conclusion

This research represents a first approach to evaluate the genetic variability and diversity within and between six Italian chicken populations using SNP and CNV markers. The results highlight the existence of genetic variability and a low inbreeding coefficient in all Italian chicken breeds considered. Notably, the pairwise fixation indexes, the PCA and the NJ trees all show the clear separation of the SI breed from the others and in the LI, the presence of two distinct groups corresponding to the white and black varieties. In addition, PI and SA resulted closely related, highlighting the geographic common origin. The genetic variability found using SNP is comparable with the one reported by other authors in the same breeds, using microsatellite markers. In addition, the CNV markers analysis have well separated the breeds in terms of genetic identity, according to their breeding history.

Some of the CNV interestingly maps in chromosomal regions where important functional genes are annotated (e.g. the major histocompatibility complex region on chromosome 16). A follow-up analysis may further investigate functional association between CNV and genes.

Results of this study represent a basis for the Italian chicken population's valorisation as an important reservoir of genetic diversity. In Italy, Avian Research Units within Academic infrastructures are currently involved in *in situ* conservation programmes of Italian poultry populations. Efforts to maintain genetic variability have been implemented and the small poultry flocks available need to be continuously monitored to avoid the loss of biodiversity.

As a conclusion, this manuscript confirm the existence of genetic and genomic variability in the Italian chicken populations suitable for their maintenance and genetic improvement. To enhance this process it is advisable that other researches on a larger population sample disclose the association between SNP and CNV markers with phenotype expression of quantitative traits.

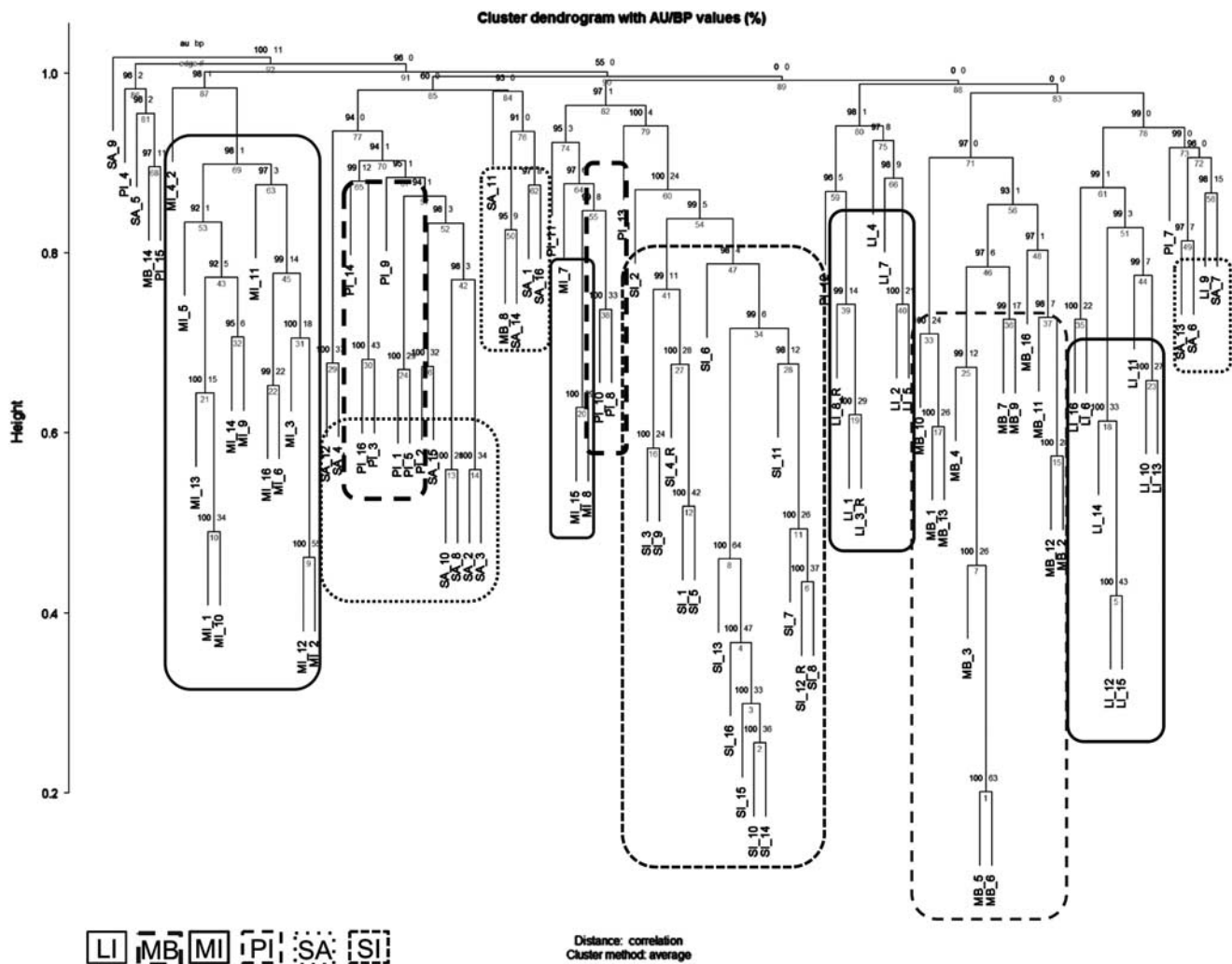


Figure 5 Dendrogram generated by clustering all individuals of the Italian chicken breeds (LI = Livornese, MB = Mericanel della Brianza, MI = Milano, PI = Bionda Piemontese, SA = Bianca di Saluzzo and SI = Siciliana) based on their copy number variant similarities. (i) approximate unbiased (AU) *P*-value in dark grey colour, (ii) bootstrap probability (BP) value in grey colour, (iii) edge in light grey colour.

Acknowledgements

This study was co-funded by project no. M01678 – Minister of Foreign affairs of Italy and Mexico.

Supplementary material

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1751731116002135>

References

Abebe AS, Mikko S and Johansson AM 2015. Genetic diversity of five local Swedish chicken breeds detected by microsatellite markers. *PLoS One* 10, e0120580.

Alexander DH, Novembre J and Lange K 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19, 1655–1664.

Al-Qamashoui B, Simianer H, Kadim I and Weigend S 2014. Assessment of genetic diversity and conservation priority of Omani local chickens using microsatellite markers. *Tropical Animal Health and Production* 46, 747–752.

Bagnato A, Strillacci MG, Pellegrino L, Schiavini F, Frigo E, Rossoni A, Fontanesi L, Maltecca C, Prinsen RTMM and Dolezal MA 2015. Identification

and validation of copy number variants in Italian Brown Swiss dairy cattle using Illumina Bovine SNP50 Beadchip. *Italian Journal of Animal Science* 14, 552–558.

Ceccobelli S, Di Lorenzo P, Lancioni H, Monteagudo Ibáñez LV, Tejedor M, Castellini C, Landi V, Martínez Martínez A, Delgado Bermejo JV, Vega Pla JL, Leon Jurado JM, García M, Attard G, Grimal A, Stojanovic S, Kume K, Panella F, Weigend S and Lasagna E 2015. Genetic diversity and phylogeographic structure of sixteen Mediterranean chicken breeds assessed with microsatellites and mitochondrial DNA. *Livestock Science* 175, 27–36.

Cepolina S 2015. *Standard Italiano delle Razze Avicole*. FIAV Publisher, Sacile (PN), Italy.

Food and Agricultural Organization (FAO) of United Nation 2009. Status and trends report on animal genetic resources – 2008. CGRFA/WG-AnGR-5/09/Inf. 7. FAO Publisher, Rome, Italy.

Frankham R, Ballou JD and Briscoe D A 2004. *A primer of conservation genetics*. Cambridge University Press, Cambridge, UK.

Gazave E, Darré F, Morcillo-Suarez C, Petit-Marty N, Carreño A, Marigorta UM, Ryder OA, Blancher A, Rocchi M, Bosch E, Baker C, Marqués-Bonet T, Eichler EE and Navarro A 2011. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. *Genome Research* 21, 1626–1639.

Hammond K 1996. The status of global farm animal genetic resources. Paper Presented at the Symposium on the Economics of Valuation and Conservation of Genetic Resources for Agriculture, Centre for International Studies on Economic Growth, 13–15 May, Tor Vergata University, Rome, Italy.

- Han R, Yang P, Tian Y, Wang D, Zhang Z, Wang L, Li Z, Jiang R and Kang X 2014. Identification and functional characterization of copy number variations in diverse chicken breeds. *BMC Genomics* 15, 934.
- Makina SO, Muchadeyi FC, van Marle-Köster E, MacNeil MD and Maiwashe A 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Frontiers in Genetics* 5, 333.
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Jakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemes J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stütz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Ye K, Eichler EE, Gerstein MB, Hurler ME, Lee C, McCarroll SA and Korbel JO, and 1000 Genomes Project 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* 470, 59–65.
- Mosca F, Madeddu M, Mangiagalli MG, Colombo E, Cozzi MC, Zaniboni L and Cerolini S 2015. Bird density, stress markers and growth performance in the Italian chicken breed Milanino. *Journal of Applied Poultry Research* 24, 529–535.
- Muir WM, Wong GK, Zhang Y, Wang J, Groenen MA, Crooijmans RP, Megens HJ, Zhang H, Okimoto R, Vereijken A, Jungerius A, Albers GA, Lawley CT, Delany ME, MacEachern S and Cheng HH 2008. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *PNAS* 105, 17312–17317.
- Nicoloso L, Bomba L, Colli L, Negrini R, Milanese M, Mazza R, Sechi T, Frattini S, Talenti A, Coizet B, Chessa S, Marletta D, D'Andrea M, Bordonaro S, Ptak G, Carta A, Pagnacco G, Valentini A, Pilla F, Ajmone-Marsan P and Crepaldi P, and Italian Goat Consortium 2015. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genetic Selection Evolution* 4, 47–62.
- Rambaut A. 2014. FigTree. Retrieved on 16 January 2016 from <http://tree.bio.ed.ac.uk/software/figtree/>.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Flegler H, Shapero MH, Carson AR and Chen W 2006. Global variation in copy number in the human genome. *Nature* 444, 444–454.
- Sartore S, Sacchi P, Soglia D, Maione S, Schiavone A, De Marco M, Ceccobelli S, Lasagna E and Rasero R 2016. Genetic variability of two Italian indigenous chicken breeds inferred from microsatellites marker analysis. *British Poultry Science* 15, 1–9.
- Schiavo G, Dolezal MA, Scotti E, Bertolini F, Calò DG, Galimberti G, Russo V and Fontanesi L 2014. Copy number variants in Italian Large White pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. *Animal Genetics* 45, 745–749.
- Strillacci MG, Marelli SP, Cozzi MC, Colombo E, Polli M, Gualtieri M, Cristalli A, Pignattelli P, Longeri M and Guidobono Cavalchini L 2009. Italian autochthonous chicken breeds conservation: evaluation of biodiversity in Valdarnese Bianca breed (*Gallus gallus domesticus*). *Avian Biology Research* 2, 229–233.
- Suzuki R and Shimodaira H 2006. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 12, 1540–1542.
- Tadano R, Nishibori M, Imamura Y, Matsuzaki M, Kinoshita K, Mizutani M, Namikawa T and Tsudzuki M 2008. High genetic divergence in miniature breeds of Japanese native chickens compared to Red Junglefowl, as revealed by microsatellite analysis. *Animal Genetics* 39, 71–78.
- Tadano R, Nishibori M, Nagasaka N and Tsudzuki M 2007. Assessing genetic diversity and population structure for commercial chicken lines based on forty microsatellite analyses. *Poultry Science* 86, 2301–2308.
- Tian M, Wang Y, Gu X, Feng C, Fang S, Hu X and Li N 2013. Copy number variants in locally raised Chinese chicken genomes determined using array comparative genomic hybridization. *BMC Genomics* 14, 262.
- Wang K, Li M, Hadley D, Liu R, Glessner J, Grant S, Hakonarson H and Bucan M 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research* 17, 1665–1674.
- Wang X and Byers S 2014. Copy number variation in chickens: a review and future prospects. *Microarrays* 3, 24–38.
- Wang Z, Chen Q, Yang Y, Liao R, Zhao J, Zhang Z, Chen Z, Zhang X, Xue M, Yang H, Zheng Y, Wang Q and Pan Y 2015. Genetic diversity and population structure of six Chinese indigenous pig breeds in the Taihu Lake region revealed by sequencing data. *Animal Genetics* 46, 697–701.
- Wilkinson S, Wiener P, Teverson D, Haley CS and Hocking PM 2011. Characterization of the genetic diversity, structure and admixture of British chicken breeds. *Animal Genetics* 43, 552–563.
- Xu L, Hou Y, Bickhart DM, Zhou Y, Hay EH, Song J, Sonstegard TS, Van Tassell CP and Liu GE 2016. Population-genetic properties of differentiated copy number variations in cattle. *Scientific Reports* 6, 23161.
- Xu S, Gupta S and Jin L 2010. PEAS V1.0: a package for elementary analysis of SNP data. *Molecular Ecology Resources* 10, 1085–1088.
- Yan Y, Yang N, Cheng HH, Song J and Qu L 2015. Genome-wide identification of copy number variations between two chicken lines that differ in genetic resistance to Marek's disease. *BMC Genomics* 16, 843.
- Yi G, Qu L, Liu J, Yan Y, Xu G and Yang N 2014. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genomics* 15, 962.
- Zanetti E, De Marchi M, Dalvit C and Cassandro M 2010. Genetic characterization of local Italian breeds of chickens undergoing in situ conservation. *Poultry Science* 89, 420–427.
- Zanon A and Sabbioni A 2001. Identificazione e salvaguardia genetica delle razze avicole italiane. *Annali della Facoltà di Medicina Veterinaria Università di Parma* XXI, 117–134.