Animal Science Papers and Reports vol. 39 (2021) no. 3, 277-295
Institute of Genetics and Animal Biotechnology
of the Polish Academy of Sciences, Jastrzębiec, Poland

Genetic diversity and population structure in the reciprocal cross between a broiler line and indigenous chickens

Hamed Asadollahi[1], Saeid Ansari Mahyari[1]*, Rasoul Vaez Torshizi[2], Hossein Emrani[3], Alireza Ehsani[2]

[1] Department of Animal Science, College of Agriculture,
Isfahan University of Technology, Isfahan, Iran

[2] Department of Animal Science, College of Agriculture,
Tarbiat Modares University, Tehran, Iran

[3] Animal Science Research Institute of Iran, Agricultural Research,
Education and Extension Organization (AREEO), Karaj, Iran

(Accepted July15, 2021)

This study was performed to evaluate the population structure by genome-wide analysis using the Illumina 60K chicken Beadchip. One F2 population derived from the reciprocal cross between Arian fast-growing chickens and Urmia slow-growing indigenous fowls and consisting of 312 F2 birds was investigated. Quality control procedures, population clustering using the multidimensional scaling (MDS) method and allele frequency assessment in half-sib groups were applied. Then the population structure was considered, as provided by the outputs of MDS, Heatmap and neighbor-joining tree, which were consistent with structure analyses by clustering the birds into eight sub-populations of K1 to K8. The average expected heterozygosity in all half-sib families was K1= 0.4677, K2= 0.4635, K3= 0.4949, K4= 0.5235, K5= 0.4613, K6= 0.4273, K7= 0.4618, and K8= 0.4459, respectively. The heterozygosity values were found to be higher than 42% for all the eight clusters in this study, indicating a relatively high genetic variability within the groups. The fixation index (F_{st}) among F2 chicken clusters ranged from 0.01 to 0.29. One possible explanation for the high estimated F_{st} (>0.1) can be a signal of selection in the present population. Our findings can be applied in future GWASs, conservation plans and genetic improvement programs of chickens.

KEY WORDS: population structure / Fixation index / heterozygosity / F2 chicken

*Corresponding authors: s.ansari@iut.ac.ir

Intense artificial selection through the domestication process resulted in a high level of diversity among animal populations. Diversity includes variations in morphology, physiology, production, and fertility traits [Chen *et al*. 2016]. Evaluation of genetic variation patterns is of particular interest when studying domestication, breed formation, population structure and consequences of selection [Kijas *et al.* 2012]. Recent developments in molecular technology have provided new opportunities for assessing genetic variability at the DNA level [Beugin *et al*. 2017, Kévin *et al.* 2019]. Worldwide, numerous studies have been based on microsatellite and SNP markers to explain diversity among chicken breeds [Berthouly *et al.* 2008, Roh *et al*. 2018, Karsli *et al.* 2019, Elferink *et al*. 2012, Granevitze *et al*. 2007, Muir *et al.* 2008, Bodzsar *et al.* 2009]. Single nucleotide polymorphism (SNP) analysis is the most commonly used valuable tool for assessing genetic diversity in chickens. The advantage of SNPs is mainly due to their high numbers compared to other genetic markers. In addition, they are easy to identify and show low mutation rates [Morin *et al.* 2004]. Various measures of genetic differentiation have been provided by researchers, the most common being is the fixation index (F_{st}) [Wright 1951]. Several studies have used F_{st} as a tool to identify patterns of genetic alteration in chickens at a particular locus among populations relative to those within populations [Seo *et al.* 2013, Yamamoto *et al.* 2011, Mwacharo *et al.* 2011, Halima *et al*. 2009, Chen *et al.* 2008, Kanginakudra *et al.* 2008, Berthouly *et al*. 2008]. Thus, F_{st} has also been used as a main statistic to identify signatures of selection in populations [Pintus *et al.* 2014, Wright 1951] using high-throughput SNP genotyping. In this regard the time scale, over which selection has occurred, has a major impact on the ability of this method to detect evidence of selection. In addition, the F_{st} method is best suited to detect events occurring in a distant past [Cadzow *et al.* 2014]. In general traditional methods describing genetic differentiation among populations totally depend on a prior grouping of individuals. Bayesian clustering methods eschew this limitation using linkage disequilibrium (LD) and Hardy-Weinberg disequilibrium to decompose a sample of individuals into genetically distinct groups. Various Bayesian clustering analysis software packages are available, all of which describe a reduction in the ability to distinguish proper clusters as the level of genetic differentiation among populations decreases [Latch *et al*. 2006]. Structure [Pritchard *et al.* 2000] is the most widely used clustering software to identify the genetic structure of a population, and thus it has the highest citation index value. In this software the admixture model is used. Assessing the genetic structure of population has increasingly become the focus, as they can produce valuable insight into patterns of gene flow. Structure analysis creates the clusters based on the Hardy-Weinberg disequilibrium (HWD) and linkage disequilibrium caused by admixture between populations. Structure can identify clusters of the whole population by detecting where both LD and HWD are minimized, therefore it can assign individuals to distinct subpopulations based on the better clustering method [Falush *et al*. 2003]. On the other hand, high LD or a significant departure from HWD could lead to an overestimation of the number of identified clusters [Falush *et al*. 2003]. Structure

analysis deals with two types of LD: the first is the mixture LD, which occurs across loci even if there is less linkage due to the correlation of allele frequencies, but the correlation is due to a mixture of ancestral populations in the studied populations [Falush *et al.* 2003]. The other type of LD is the admixture LD: a correlation that occurs between linked markers in recently admixed populations [Pritchard and Wen 2004]. This LD arose from markers placed on the same chunk of chromosome that branched from ancestral population. The 'admixture model' combines an admixture LD with map distances between markers to improve clustering results [Falush *et al.* 2003]. Furthermore, the amount of LD increases due to the influence of other evolutionary events such as population bottlenecks [Lynch and Walsh 1998] and demographic variations [Jorde 2000]. In several studies, the structure provided robust estimates with no need for prior population grouping of genetic diversity in chicken populations [Palinkas-Bodzsar *et al*. 2020, Wragg *et al.* 2012, Twito *et al.* 2007]. Besides, heterozygosity requires a locus having two different alleles. Consequently, the fraction of heterozygous nucleotides within an individual compared to the parents' heterozygosity, is another important metric for understanding genetic variation in populations [Bryc *et al.* 2013], which can also be estimated by structure. Estimated heterozygosity has been differently reported in various populations of chickens [Zhang *et al.* 2018, Lyimo *et al.* 2013, Muchadeyi *et al.* 2007, Ponsuksili *et al.* 1996, Groen *et al*. 1994]. Other methods such as Multi-Dimensional Scaling (MDS) [Chang *et al.* 2015], Heatmap algorithm [Meyer *et al.* 2020] or neighbor-joining analysis [Luo *et al.* 2020], can be used to study the population structure. This study was designed to investigate population structure. For this purpose, estimates of the expected genetic distance and heterozygosity between and within families were implemented based on a comprehensive analysis of population structure, while the signatures of selection in the F2 population of chickens were considered using high-throughput genomic data. The goal of this study was to estimate the population structure and genetic admixture of each subpopulation.

Material and methods

Experimental population

One F2 population was generated by applying reciprocal crosses between the Arian line (A) and Urmia Iranian native chickens (N), the former as a commercial fast-growing broiler strain and the latter as a slow-growing indigenous population. The F1 chickens were generated from the matings of A ♂ × N ♀ and N ♂ × A ♀ birds. Each F1 male resulted from a reciprocal cross, then mated with four to eight females from the other families. Finally a total of 312 F2 chickens from eight half-sib families were generated in five different hatches. Day-old F_2 chickens were weighed and reared on the floor for 7 days under 24 h light and a brooding temperature of 33°C. This temperature was decreased to 30°C on day 7. On day 8, birds were weighed and moved to individual cages with a temperature of 30°C, which was gradually decreased

to reach a final temperature of 22°C, and a 22 h light and 2 h dark cycle throughout the experimental period. Chickens did not receive vaccines during the rearing period. Feed and water were provided ad libitum. This experiment was conducted to measure the chickens' carcass compositions at 84 days of age.

Genotyping and quality control

DNA was extracted from 312 blood samples by the salting out method and stored at -20°C until further analysis. All the samples were genotyped at Aarhus University, Denmark, using the Illumina Chicken 60K BeadChip containing 54,340 SNP markers provided by Cobb Vantress. These markers covered thirty autosomes and sex chromosomes of the chicken genome. Quality control was performed with the PLINK (v1.9) software package [Chang *et al*. 2015, Purcell *et al.* 2007]. Marker quality control was achieved by excluding SNPs with a minor allele frequency below 5%, Hardy-Weinberg equilibrium below 1×10^{-6}, and the call rate below 95%. After these quality control steps, a total of 48,439 SNPs and 308 individuals remained for further assessments. Distributions of SNPs and the average distance between the adjacent SNPs on each chromosome were determined (Supplementary Table 1) using the synbreed and R3.2.2 software packages [Wimmer *et al.* 2016].

Genetic and population structure

Multi-Dimensional Scaling (MDS) [Chang *et al.* 2015, Sun *et al.* 2013], the Heatmap algorithm [Meyer *et al*. 2020, Emrani *et al.* 2017, Tang *et al*. 2015], neighbor-joining analysis [Luo *et al.* 2020], and a model-based Bayesian method [Li *et al*. 2020, Roh *et al.* 2020, Pritchard *et al.* 2000] were applied to assess the F2 population structure and provide insight into the relationships within and between half-sibs at the genomic level. The population structure was evaluated by multi-dimensional scaling (MDS) using the PLINK software (v1.09) [Chang *et al*. 2015]. Independent SNP markers were obtained for all autosomes using the independence-pairwise option, with a window size of 30 SNPs, a step of five SNPs and an r^2 threshold of 0.2, as suggested by Wang *et al.* [2009]. Then, independent SNP markers were used to estimate the pairwise identity-by-state (IBS) relationship between all individuals, while MDS components were obtained using the MDS-plot option based on the IBS matrix [Sun *et al*. 2013]. Ultimately, the relative kinship matrix was constructed from captured independent markers as suggested by Liu *et al.* [2015]. Cluster analysis was conducted for all genotypes based on genetic distance according to the neighbor-joining method using agglomerative clustering [Luo *et al*. 2020, Bradbury *et al.* 2007]. Additionally, hierarchical clustering of the genomic relationship matrix was performed and the results were plotted as a heatmap to visualize the magnitude of the genomic relationship using the GAPIT package version 2 [Meyer *et al.* 2020, Emrani *et al.* 2017, Tang *et al.* 2015]. To clarify the presence of population structure by assigning individuals to subpopulations a Bayesian approach using marker genotypes data (48,439 SNPs) was used to estimate the historical relationships among

Table 1. Distributions of SNPs before and after quality control and the average distance between adjacent SNPs on each chromosome

Chromosome	SNP Markers in chip	SNP Markers after quality control	Physical Map (Mb)	Average distance (Kb)
1	8298	7546	200.8	26.5
2	6347	5762	152.2	26.7
3	4747	4340	113.5	26.3
4	3869	3549	94.2	26.5
5	2550	2303	62.2	27.1
6	2000	1819	35.8	19.6
7	2092	1907	40.3	20.1
8	1631	1502	30.6	20.1
9	1366	1272	24.0	18.8
10	1545	1378	22.4	16.1
11	1531	1339	21.8	16.4
12	1559	1413	20.2	14.4
13	1371	1241	18.3	14.6
14	1179	1079	15.9	14.3
15	1222	1097	12.9	11.8
16	24	20	0.4	21.7
17	999	898	10.3	11.8
18	1048	950	10.8	11.9
19	973	890	9.8	11.3
20	1807	1567	14.1	8.8
21	901	805	6.9	8.5
22	442	301	3.8	12.6
23	724	635	7.4	9.3
24	853	759	6.4	8.5
25	219	166	2.3	11.5
26	776	696	5.0	7.4
27	566	521	4.8	9.4
28	715	579	4.4	7.6
29	142	108	0.6	7.7
30	7	4	0.02	6.9
Z	2835	1994	73.0	37.5
W	1	0	-	-

Z and W are sex chromosomes.

Table 2. The results of Structure analysis of 312 chickens for the fixation index (F_{st}) (significant divergences) and average distances (expected heterozygosity) in each subpopulation

Subpopulation	F_{st}[1]	Exp. het[2]
K1	0.2186	0.4677
K2	0.2011	0.4635
K3	0.1422	0.4949
K4	**0.0106**	**0.5235**
K5	0.1825	0.4613
K6	**0.2959**	**0.4273**
K7	0.2077	0.4618
K8	0.2653	0.4459

[1]F_{st} is a measure of genetic differentiation; [2]Expected heterozygosity.

Table 3. Net nucleotide distance, computed using point estimates of P (distances among the subpopulations)

Subpopulation	K1	K2	K3	K4	K5	K6	K7	K8
K1	-	0.0710	0.0406	0.0358	0.0672	0.0866	0.0616	**0.0888**
K2	0.0710	-	0.0477	0.0286	0.0403	0.0804	0.0600	0.0823
K3	0.0406	0.0477	-	**0.0148**	0.0416	0.0653	0.0475	0.0630
K4	0.0358	0.0286	**0.0148**	-	0.0223	0.0432	0.0294	0.0402
K5	0.0672	0.0403	0.0416	0.0223	-	0.0697	0.0549	0.0703
K6	0.0866	0.0804	0.0653	0.0432	0.0697	-	0.0743	0.0704
K7	0.0616	0.0600	0.0475	0.0294	0.0549	0.0743	-	0.0782
K8	**0.0888**	0.0823	0.0630	0.0402	0.0703	0.0704	0.0782	-

populations with the structure software [Li *et al.* 2020, Roh *et al.* 2020, Pritchard *et al.* 2000]. In this step individuals was placed in specific clusters (number of clusters = *K*), where the sum of probabilities of belonging to a cluster equals one. Structure analysis would ascribe a probability Pr(*X*|*K*) given the data (X), and the log Pr(*X*|*K*) was used to determine the more likely number of clusters [Pritchard *et al.* 2000]. However, Pr(*X*|*K*) is difficult to estimate, thus Pritchard *et al*. [2000] proposed an approach to approximate the probability of *K* given the genotyping data. Population structure was determined using *k*-values by the assumption of a fixed number of clusters from 1 to 8. Three independent analyses were used for each *k*-value and the program was set at 3 as the burn-in period, followed by 100000 Markov Chain Monte Carlo (MCMC) replications after the burn-in part, in order to have a random starting value for the algorithm [Roh *et al.* 2020, Chen *et al*. 2008]. The best number of k for the current population was determined by the structure harvester [Li *et al.* 2020, Earl and vonHoldt 2012]. The summary statistics of all SNP markers such as the relative membership of each predefined population in each of the clusters, mean value of fixation indexes (F_{st}), expected heterozygosity between individuals in the same cluster (Tab. 2) and divergence among clusters (Net nucleotide distance) (Tab. 3) were also calculated using the structure software V 2.3 [Roh *et al.* 2020, Pritchard *et al.* 2003]. The fixation index (F_{st}) measures the effects of population subdivision related to potential allele fixation relative to the total population [Hartl 1998]. The F_{st} values, as defined in Nei's formula [Nei 1987], was estimated on the basis of subpopulation genetic diversity and total genetic diversity. Finally, the measure of LD was investigated by the standardized disequilibrium coefficient, *D'*, as well as r^2 and *P*-values using the tassel software (Fig. 6) – Bradbury *et al.* 2007.

Results and discussion

Distributions of SNPs before and after quality control and average distances between the adjacent SNPs on each chromosome were determined and summarized in Table 1. Multi-dimensional scaling analysis (MDS) of the SNPs with an r^2 threshold of 0.2 considering the first MDS component clustered chickens in eight families (Fig.

1). The genetic structure of the studied population, i.e. the variance of the data, was captured mainly by the first component of MDS. Furthermore, the Heatmap algorithm captured from hierarchical clustering of the genomic relationship matrix also proved that the 8 half-sib groups were clearly distinguished (Fig. 2). Also the neighbor-joining

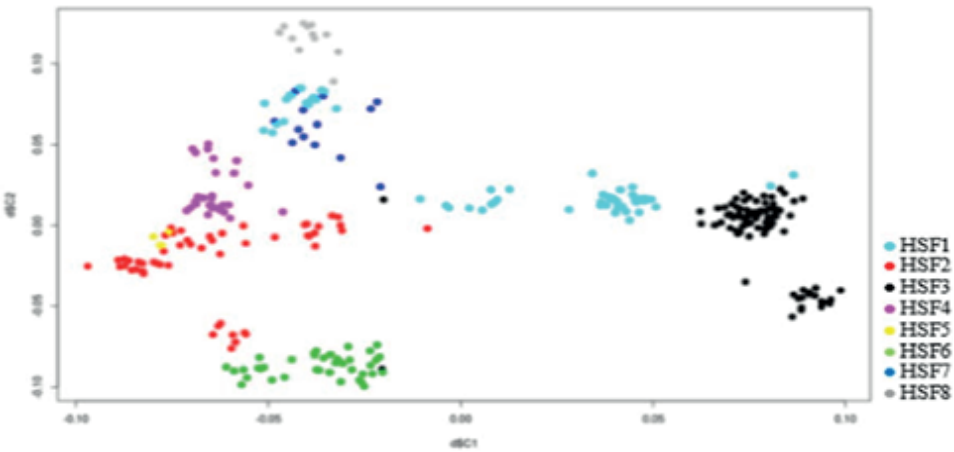

Fig. 1. Population structure identification with multidimensional scaling analysis. Full-sib families are shown in the same color (HSF = half-sibling family).

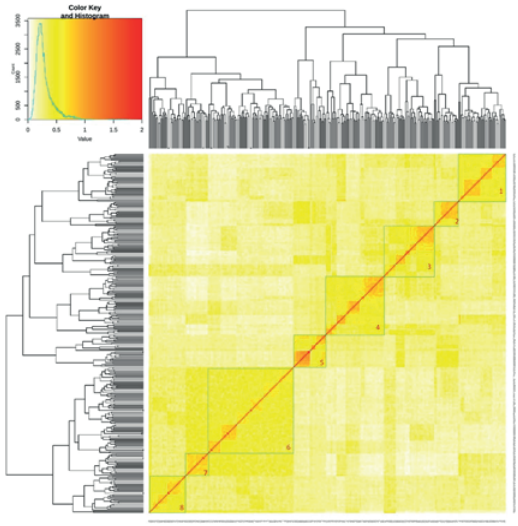

Fig. 2. The kinship matrix displayed virtually via a heatmap and a tree (red = the highest correlation among the pairs of individuals; yellow = the lowest correlation). A hierarchy tree among the individuals presented based on their kinship with the red diagonal = perfect relationship of each individual with itself (The highlighted blocks on the diagonal show clusters of individuals from eight half-sib families for 308 F2 chickens).

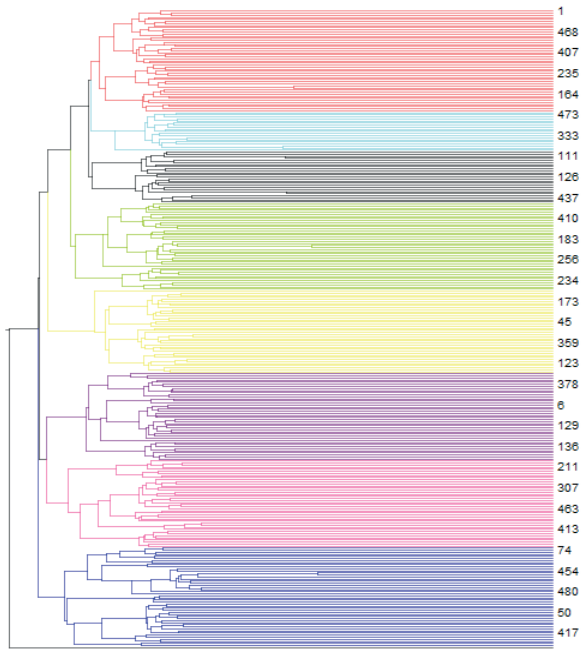

Fig. 3. Genetic relationships among 8 chicken groups constructed using a neighbor-joining phylogenetic tree from shared allele distance, based on 48,439 single nucleotide polymorphisms (SNPs).

tree, based on genetic distance among all genotypes, was created using the same 48,439 SNPs. The findings revealed the existence of eight subgroups in the studied population, which is in agreement with the population structure analysis by MDS (Fig. 3). The eight subpopulations resulted from structure analysis then were used in the structure harvester to calculate genetic diversity. In order to find the efficient value of *K (*number of clusters), it was plotted against ΔK, which depicted a sharp peak at $K = 2$ (Fig. 4). Calculation of Δ*K* from the structure output showed a specific amount of statistics at K = 2 (Fig. 4). Although the height of the specific value of ΔK represents the strength of the population subdivision signal [Evanno *et al.* 2005], here we preferred deep subdivision at $K = 8$, because $K = 2$ may have underestimated the number of subpopulations in the recent population [Janes *et al.* 2017]. Assignment probabilities were used to infer the membership of each individual to the most probable subgroups. Therefore, correlated allele frequencies were applied in the linkage model. The 48,439 SNPs were identified across the eight subgroups of chickens and they are illustrated by the bar plot (Fig. 5). The results of MDS (Fig. 1), Heatmap (Fig. 2), and neighbor-joining tree (Fig. 3) were close to results of structure analyses by clustering the 312 F2 chickens into eight obvious groups through their genotypic information. Average heterozygosity indicated differences between identified clusters. It means that a significant genetic divergence was observed among the subgroups

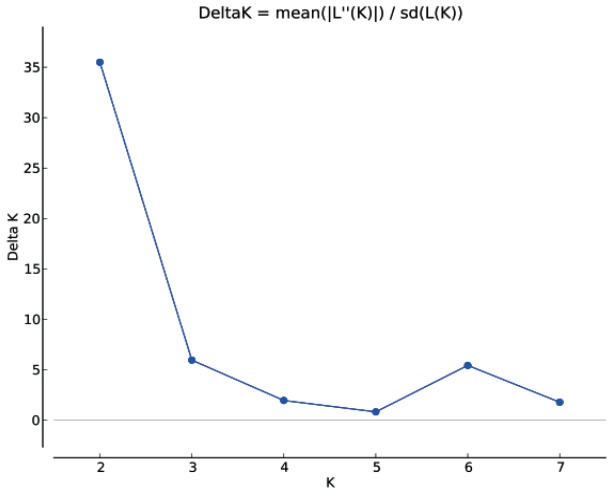

Fig. 4. Delta (Δ)K for differing numbers of subpopulations (k).

and also average distances or expected heterozygosity among genotypes in each subpopulation were statistically significant. The highest and lowest values of expected heterozygosity were observed in *K4* (0.5235) and *K6* (0.4273), respectively (Tab. 2). In this study, genetic differentiation varied throughout the genome. The fixation index (*Fst*) was calculated to evaluate the population substructures and it proved to be the most efficient index for assessing the overall genetic variation among subpopulations. The F_{st} values were 0.2186, 0.2011, 0.1422, 0.0106, 0.1825, 0.2959, 0.2077, and 0.2653 for *K1* to *K8*, respectively (Tab. 2). Generally, the overall extent of LD in the genome and the LD decay over the physical distance throughout the chromosomes are essential to determine the marker density required for population structure studies. Accordingly, 48,439 SNPs in total were used for a schematic demonstration of the LD extent among all paired loci by the structure (Fig. 6). The net nucleotide distance, which is an indicator of genetic divergence among clusters, was calculated applying points of estimated allele frequencies. The shortest distance (0.0148) was observed between clusters of 3 and 4, and the greatest distance between clusters of 1 and 8 (0.0888) (Tab. 3).

As reported by Nejati-Javaremi *et al.* [1997] and Schork [2001], elements of the pedigree numerator relationship matrix represent the expected genome sharing for two individuals, whereas marker-based relationship matrix of whole-genome sharing can be calculated by the summation of allele sharing at numerous loci in the genome. Therefore, the marker approach incorporates kinship variation among related animals with the same degree of relationship (e.g. full- or half-sibs) and thus estimates genome sharing more adequately. Departure from expectation can be attributed to the pivotal factors such as demographic structure, small population size, selection, and Mendelian segregation [AbdollahiArpanahi *et al.* 2014].

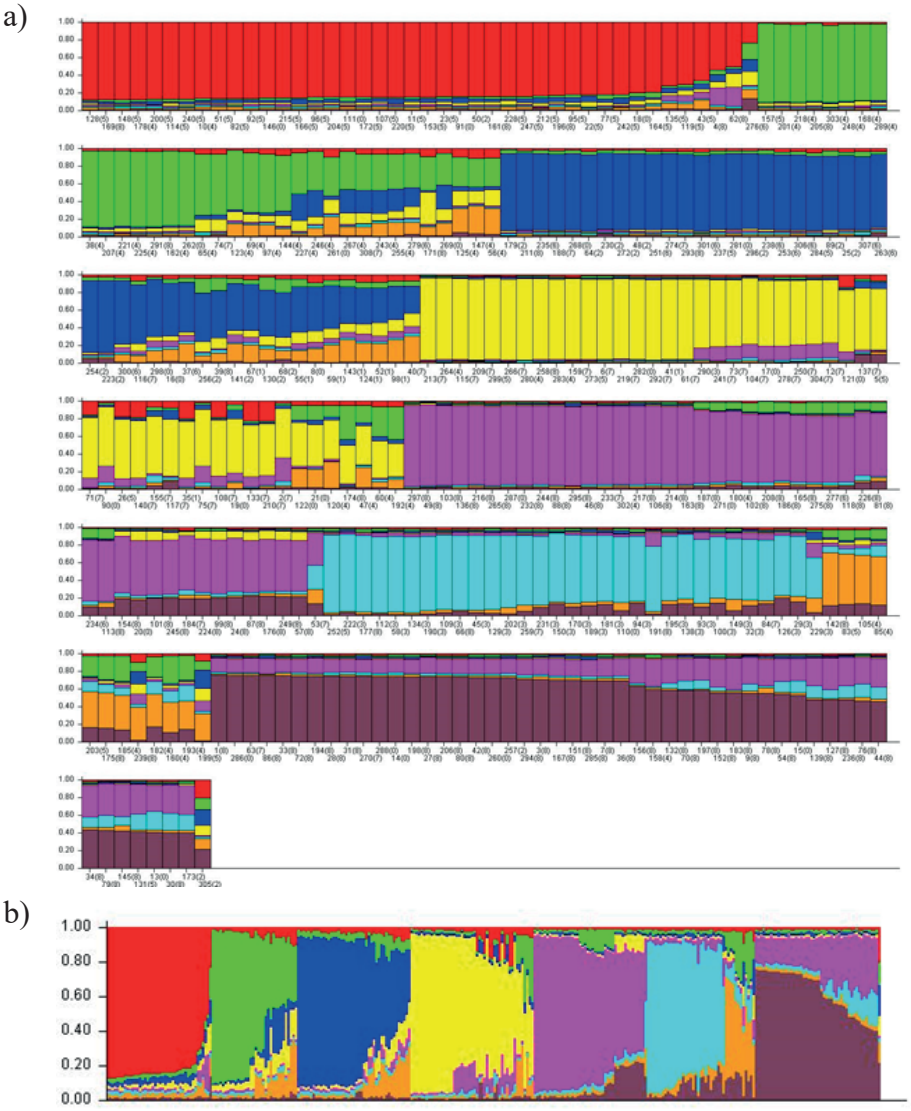

Figure 5. Estimated population structure of 312 F2 chicken genotypes on (k = 8), using Structure 2.3. a) Bar plot 1; b) Bar plot 2. Each color segments represents a subpopulation.

Linkage disequilibrium

The underlying factor responsible for the overestimation of clusters is linkage disequilibrium, which can generate spurious numbers of clusters in the population partitioning [Falush *et al*. 2003]. Structure software constructs population clusters

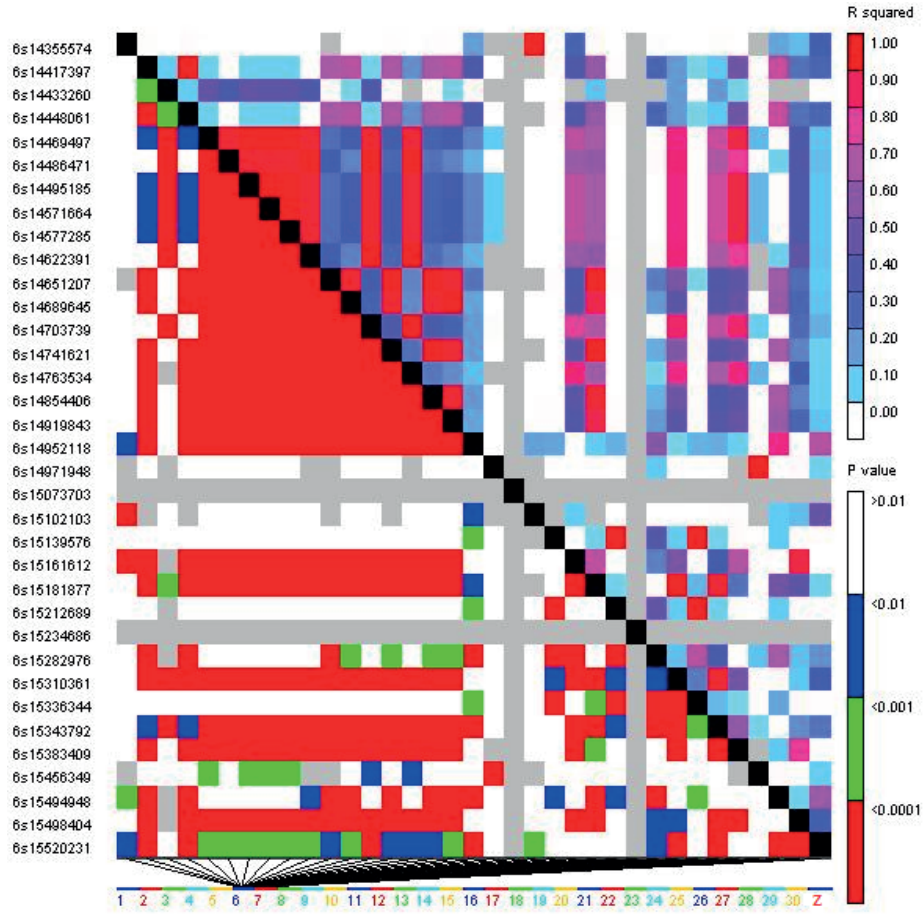

Fig. 6. LD plots with positional information of thirty autosome and one sex (z) chromosomes.

based on transient Hardy-Weinberg (HWD) and linkage disequilibrium (LD) that occurred due to population admixture. Hence, individuals are assigned to groups, where the amount of HWD and LD are minimized in each group. Therefore, this approach enhances the efficiency of clustering outcomes [Falush *et al.* 2003]. Basically, the loci selection based on the distance map is not an efficient approach. The LD strength is never correlated to the physical distance between alleles across the genome (Jorde 2000). Actually, LD not only depends on the distance between two loci, but it can be increased by founder events or be decreased by population dynamics (Slatkin 1994) and the number of alleles at a certain locus [Ott and Rabinowitz 1997]. A triangle plot for pairwise LD between markers in a genome specific fragment is generated by TASSEL, where pairwise LD values of polymorphic sites are plotted

on both the *X* and *Y* axes; above the diagonal it displays r^2 and below the diagonal it displays the corresponding p-values from the rapid permutation test [Abdurakhmonov and Abdukarimov 2008]. In the LD plot each cell represents a comparison of pairwise marker sites with the color codes for the presence of significant LD (Fig. 6). Colored barcodes for the significance threshold levels are shown on both diagonals. The genetic distance scale for a genome fragment was manually drawn. Results indicated that LD was unbiased in the clustering groups.

Genetic structure

According to the results, eight half-sib families in F2 chickens were distinguished by MDS [Chang *et al.* 2015, Sun *et al.* 2013] (Fig. 1), the heatmap from hierarchical clustering of the genomic relationship matrix [Meyer *et al.* 2020, Emrani *et al.* 2017, Tang *et al.* 2015] – Figure 2, and the neighbor-joining tree (Luo *et al.* 2020] – Figure 3. Interestingly, there were one or two individuals of full-sib families within each half-sib family, indicated in clustering. Results were confirmed by the admixture model (Fig. 5). The admixture model supposes that individuals are descended from an ancestral population. In this case, it is modelled by assuming that a certain individual has inherited some fractions of its genome from their ancestors in population *K* [Van Marle-Köster *et al.* 2008]. The structure 2.3 [Li *et al.* 2020, Roh *et al.* 2020] was used to infer the number of subgroups, with a fully Bayesian process described by Pritchard *et al.* (2000). Therefore, this approach was run with various numbers of clusters (*K*). The *K* value that presented the maximum likelihood over the runs retained as the most probable number of clusters [Pritchard and Wen 2004]. A series of models with *K* ranging from 1 to 8 was run using all loci in this research. To verify the consistency of the results, 3 independent runs were applied for each K. Finally, the structure was run using all the different possible pairs of loci and estimated the number of clusters for each of these pairs [Roh *et al.* 2020]. Examination of Ln $P(X|K)$ values from the Structure harvester program also suggested the best level of subdivision at $K = 2$. As has been reported by other researchers [Rosenberg *et al.* 2001, Evanno *et al.* 2005], variance in Ln P(X|K) increased at higher values of K. This variance is thought to prevent the identification of the highest likelihood of *K*. Pritchard *et al*. [2000] suggested that the estimated probabilities should be considered as a guide, which models are consistent rather than accurate estimates of the posterior probabilities of *K*. In order to find the suitable value of K by visualization, the number of clusters (*K*) was plotted against ΔK, which showed a sharp peak at K = 2 (Fig. 4). The ΔK method may underestimate the number of subpopulations. Moreover, it was reported previously in populations with a high admixture level, population structure can be underestimated by the ΔK method [Janes *et al.* 2017]. Although the deep division at $K = 2$ captured, according to Janes *et al*. [2017], Rosenberg *et al.* [2001] and Evanno *et al*. [2005] and also is consistent with the results from the MDS, heatmap and neighbor-joining tree analyses, identified the best number of *K* that clearly defines the true number of subpopulations ($K = 8$) in our research population. In conclusion, results confirmed

that $K = 8$ as the modeled population substructure showed $K = 2$ could underestimate the number of existing subpopulations [Breria *et al*. 2020, Janes *et al.* 2017].

Genetic diversity

In order to measure distances among the subpopulations and to further evaluate population structure, net nucleotide distances between pairs of subpopulations were computed (Tab. 3). The genetic distances between pairs of subpopulations ranged from 0.0148 (between K3 and K4) to 0.0888 (between K1 and K8) (Tab. 3), which indicates low to moderate genetic differentiation among the groups. As reported by Cadzow *et al*. [2014], the level of genetic differentiation between subpopulations (K1 to K8) was low in neutral genomic regions and in regions of balanced selection, and conversely divergence was observed in regions subject to directional selection [Cadzow *et al.* 2014]. The F_{st} quantifies differences in allele frequencies between subpopulations, and it can range from zero to one, the lowest F_{st} of zero implying that there is no differentiation and the highest F_{st} of 1 indicating complete differentiation through the fixation process in the population. Negative F_{st} values have no biological interpretation [Akey *et al.* 2002]. Even though the mean F_{st} values were almost identical between the seven ubgroups (*K1*= 0.2186, *K2*= 0.2011, *K3*= 0.1422*, K5*= 0.1825, *K6*= 0.2959, *K7*= 0.2077 and *K8*= 0.2653), the subgroup *K4*= 0.0106 indicated the lowest mean value of F_{st} among subpopulations. This indicates that F_{st} clearly carries information concerning the population subdivision (*K1, K2, K3, K5, K6, K7* and *K8*), and displays an obvious distinction between these groups with subpopulation 4 [Li *et al*. 2020]. The highest F_{st} values were calculated for the *K6* subpopulation ($F_{st} = 0.29$), followed by subpopulation K8 *(*$F_{st} = 0.26$). This can be explained by high genetic differentiation in these groups (Tab. 2). The high F_{st} values implies a clear population structure of eight groups, which is in agreement with the MDS, heatmap, and neighbor-joining tree results and it is notably similar to the findings of French native chicken populations [Berthouly *et al.* 2008]. One possible explanation of the high estimated F_{st} (>0.1) can be a signal of selection in the present population [Zhao *et al*. 2015, Kijas *et al.* 2012]. The F_{st} index is effective in identifying selection signatures among different groups, [Qanbari *et al.* 2011] and it has been widely employed to determine how the divergent selection can affect the genomic pattern in populations [Zhao *et al.* 2015]. This brings us to the conclusion that alteration in allele frequencies and the observed genetic diversity can be a result of recent selection for different criteria and other non-reported genetic events in the past. The findings of the current study on F_{st} values are in concordance with Indian [Kanginakudra *et al*. 2008], Chinese [Chen *et al*. 2008] and African [Mwacharo *et al*. 2011] indigenous chicken populations, which were reported in several literature sources (0.15 to 0.26), and indicated that comparatively similar genetic distances existed between these populations and the present subpopulations. Halima *et al.* [2009] reported the genetic distances in seven Ethiopian native chicken populations ranged from 0.073 to 0.13. Seo *et al.* [2013] found genetic distance among five Korean native chicken lines to be 0.083 to 0.171. Yamamoto *et al.* [2011] reported

low F_{st} values for Bangladeshi chicken varieties. The results of Yamamoto *et al.* [2011] were supported by the study of Bhuiyan *et al*. [2013], who reported the low genetic differentiation (F_{st}= 0.1) among the different Bangladeshi chicken populations. The genetic distances observed in the present study were higher than those reported values of Halima *et al.* [2009], Seo *et al.* [2013], Yamamoto *et al.* [2011], and Bhuiyan *et al.* [2013]. Altogether, the high genetic differentiation in each subpopulation might be due to the recent use of crossbreeding in the F2 population.

Heterozygosity

The allele frequency-dependent diversity estimator, observed heterozygosity, is a measure of genetic variation, which can be advantageous in comparing populations [Kijas *et al.* 2012]. The heterozygosity values were found to be higher than 42% for all the eight clusters in this study, indicating a relatively high genetic variability within groups. In a study conducted by Ponsuksili *et al.* [1996] that included a number of different local breeds, heterozygosity values ranged from 33.5% for the Dandarawi and Fayomi (35.1%) from Egypt, 50% for the Nunakan from Indonesia, to as high as 62.9% for the Kadaknath from India. Also, it has been reported by Muchadeyi *et al.* [2007] that heterozygosity values for local chickens were all above 50%; about 64-66% for Zimbabwe, 60.7% for Malawi and 56.1% for Sudan chickens. Furthermore, another study has provided evidence that genetic variability for commercial broiler and layer lines was apparently lower, ranging from 28 to 44% [Groen *et al.* 1994]. This has also been explored by Zhang *et al.* [2018], on three different chicken breeds, the amount of heterozygosity in all the three breeds was estimated about 0.22 [Zhang *et al.* 2018]. The findings of the current study on heterozygosity were lower than Tanzanian, Ethiopian and Chinese chicken populations reported in several studies [Lyimo *et al.* 2013, Halima *et al.* 2009, Zhang *et al.* 2002]. However, the heterozygosity observed in the present study was higher than those reported values of Groen *et al.* [1994] and Zhang *et al*. [2018]. In this research, it was found that suitable heterozygosity between individuals in each cluster (42-52%) – Table 2 – from the F2 population resulted from reciprocal crossing between broiler males of the Arian line and the Urmia Iranian native fowl.

These results would support further investigations of population structure and differentiation in the F2 chicken population. Also, the information from our research can be useful for future GWAS studies, conservation plans, and genetic improvement strategies in the chicken breeding industry.

Conclusion

The proposed SNP panel can be used to characterize fast-growing lines, slow-growing indigenous chickens, and their F2 populations. In general, the results showed that the use of genomic data can easily estimate the population structure and genetic distance between different populations, which should contribute to a better

understanding of the population structure in broiler chickens. These findings can accelerate the genetic progress in breeding programs.

Acknowledgements. *The authors gratefully acknowledge the Tarbiat Modares University, Tehran, Iran for financial support, the Staff of the Arian Line Breeding Center and the West Azarbayjan Native Fowls Breeding Center for providing parents of F1 chickens. Genotyping of the birds was supported by Aarhus University, Denmark. The authors would like to thank Dr. Just Jensen for financial support of bird genotyping. Access to the developed 60K SNP Illumina chicken array was kindly provided by the USDA Chicken GWMAS Consortium, Cobb Vantress, and Hendrix Genetics.*

Conflict of interest: Authors declare no conflict of interests for this article.

REFERENCES

1. ABDOLLAHI-ARPANAHI R., PAKDEL A., NEJATI-JAVAREMI A., MORADI SHAHRBABAK M., MOROTA G., VALENTE B.D., KRANIS A., ROSA G.J.M., GIANOLA D., 2014 – Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. ***Journal of Animal Breeding and Genetics*** 131, 1-11.
2. ABDURAKHMONOV I.Y., ABDUKARIMOV A., 2008 – Application of association mapping to understanding the genetic diversity of plant germplasm resources. ***International Journal of Plant Genomics***.
3. AKEY J.M., ZHANG G., ZHANG K., JIN L., SHRIVER M.D., 2002 – Interrogating a high-density SNP map for signatures of natural selection. ***Genome Research*** 12, 1805-1814.
4. BERTHOULY C., BED'HOM B., TIXIER-BOICHARD M., CHEN C.F., LEE Y.P., LALOË D., LEGROS H., VERRIER E. ROGNON X., 2008 – Using molecular markers and multivariate methods to study the genetic diversity of local European and Asian chicken breeds. ***Animal Genetics*** 39, 121-129.
5. BERTHOULY C., BED'HOM B., TIXIER-BOICHARD M., CHEN C.F., LEE Y.P., LALOË D., LEGROS H., VERRIER E. ROGNON X., 2008 – Using molecular markers and multivariate methods to study the genetic diversity of local European and Asian chicken breeds. ***Animal Genetics*** 39, 121-129.
6. BEUGIN M.P., LETTY J., KAERLE C., GUITTON J.S., MUSELET L., QUENEY G. PONTIER D., 2017 – A single multiplex of twelve microsatellite markers for the simultaneous study of the brown hare (Lepus europaeus) and the mountain hare (Lepus timidus). ***Ecology and Evolution*** *7*, 3931-3939.
7. BHUIYAN M.S.A., CHEN S., FARUQUE S., BHUIYAN A.K.F.H., BEJAPEREIRA A., 2013 – Genetic diversity and maternal origin of Bangladeshi chicken. ***Molecular Biology Reports*** 40, 4123-4128.
8. BODZSAR, N., EDING, H., REVAY, T., HIDAS, A. WEIGEND, S., 2009 – Genetic diversity of Hungarian indigenous chicken breeds based on microsatellite markers. ***Animal Genetics*** 40, 516-523.
9. BRADBURY P.J., ZHANG Z., KROON D.E.T., CASSTEVENS M., RAMDOSS Y., BUCKLER E.S., 2007 – TASSEL: software for association mapping of complex traits in diverse samples. ***Bioinformatics*** 23, 2633-2635.

10. BRERIA C.M., HSIEH C.H., YEN J.Y., NAIR R., LIN C.Y., HUANG S.M., NOBLE T.J. SCHAFLEITNER R., 2020 – Population structure of the world vegetable center mungbean mini core collection and genome-wide association mapping of loci associated with variation of seed coat luster. ***Tropical Plant Biology*** 13, 1-12.
11. BRYC K., PATTERSON N., REICH D., 2013 – A novel approach to estimating heterozygosity from low-coverage genome sequence. ***Genetics*** 195, 553-561.
12. CADZOW M., BOOCOCK J., NGUYEN H.T., WILCOX P., MERRIMAN T.R., BLACK M.A., 2014 – A bioinformatics workflow for detecting signatures of selection in genomic data. ***Frontiers in Genetics*** 5, 293.
13. CHANG C.C., CHOW C.C., TELLIER L.C., VATTIKUTI S., PURCELL S.M. LEE J.J., 2015 – Second-generation PLINK: rising to the challenge of larger and richer datasets. ***Gigascience*** 4, 13742-015.
14. CHEN M., PAN D., REN H., FU J., LI J., SU G., WANG A., JIANG L., ZHANG Q. LIU J.F., 2016 – Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. ***Genetics Selection Evolution*** 48, 76.
15. CHEN G., BAO W., SHU J., JI C., WANG M., EDING H., MUCHADEYI F., WEIGEND S., 2008 – Assessment of population structure and genetic diversity of 15 Chinese indigenous chicken breeds using microsatellite markers. ***Asian-Australasian Journal of Animal Sciences*** 21, 331-339.
16. EARL D.A., VONHOLDT B.M., 2012 – Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. Conserv. ***Genetic Resources*** 4, 359-361.
17. ELFERINK M.G., MEGENS H.J., VEREIJKEN A., HU X., CROOIJMANS R.P. GROENEN M.A., 2012 – Signatures of selection in the genomes of commercial and non-commercial chicken breeds. ***PloS one*** 7, 32720.
18. EMRANI H., TORSHIZI R.V., MASOUDI A.A. EHSANI A., 2017 – Identification of new loci for body weight traits in F2 chicken population using genome-wide association study. ***Livestock Science*** 206, 125-131.
19. EVANNO G., REGNAUT S., GOUDET J., 2005 – Detecting the number of clusters of individuals using the software structure: a simulation study. ***Molecular Ecology*** 14, 2611-2620.
20. FALUSH D., STEPHENS M., PRITCHARD J.K., 2003 – Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. ***Genetics*** 164, 1567-1587.
21. GRANEVITZE Z., HILLEL J., CHEN G.H., CUC N.T.K., FELDMAN M., EDING H. WEIGEND S., 2007 – Genetic diversity within chicken populations from different continents and management histories. ***Animal Genetics*** 38, 576-583.
22. GROEN A.F., CROOIJMANS R.P.M.A., VAN KAMPEN A.J.A., VAN DER BEEK S., VAN DER POEL J.J. GROENEN M.A.M., 1994 – August. Microsatellite polymorphism in commercial broiler and layer lines. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Guelph 21, 94-97.
23. HALIMA H.M., NESER F.W.C., DEKOCK A., VAN M.K.E., 2009 – Study on the genetic diversity of native chickens in northwest Ethiopia using microsatellite markers. ***African Journal of Biotechnology*** 8, 13471353.
24. HARTL D.L., 1998 – A primer for population genetics, 2nd ed. Sinauer Associates, Inc., Sunderland, MA. ISBN 0878933018.
25. JANES J.K., MILLER J.M., DUPUIS J.R., MALENFANT R.M., GORRELL J.C., CULLINGHAM C.I. ANDREW R.L., 2017 – The K= 2 conundrum. ***Molecular Ecology*** 26, 3594-3602.
26. JORDE L.B., 2000 – Linkage disequilibrium and the search for complex disease genes. ***Genome Research*** 10, 1435–1444.

27. KANGINAKUDRA S., METTA M., JAKATI R.D., NAGARAJU J., 2008 – Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. ***BMC Evolutionary Biology*** 8, 174.
28. KARSLI T. BALCIOĞLU M.S., 2019 – Genetic characterization and population structure of six brown layer pure lines using microsatellite markers. ***Asian-Australasian Journal of Animal Sciences*** 32(1), 49.
29. KÉVIN K.S., CHARLES D.G.K., VALENTINE Y.G., SOULEYMANE S., MAURICE K. ISSAKA Y.A.K., 2019 – Genetic diversity of Benin cattle populations using microsatellite markers. ***Technology*** 3(1), 7-19.
30. KIJAS J.W., LENSTRA J.A., HAYES B., BOITARD S., NETO L.R.P., SAN CRISTOBAL M., SERVIN B., MCCULLOCH R., WHAN V., GIETZEN K. PAIVA S., 2012 – Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. ***PLOS Biology*** 10, 1001258.
31. LATCH E.K., DHARMARAJAN G., GLAUBITZ J.C. RHODES O.E., 2006 – Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. ***Conservation Genetics*** 7, 295-302.
32. LI D., SUN, G., ZHANG M., CAO Y., ZHANG C., FU Y., LI F., LI G., JIANG R., HAN R. LI Z., 2020 – Breeding history and candidate genes responsible for black skin of Xichuan black-bone chicken. ***BMC Genomics*** 21(1), 1-15.
33. LI J.J., ZHANG L., REN P., WANG Y., YIN L.Q., RAN J.S., ZHANG X.X. LIU Y.P., 2020 – Genotype frequency distributions of 28 SNP markers in two commercial lines and five Chinese native chicken populations. ***BMC Genetics*** 21(1), 12.
34. LIU R., SUN Y., ZHAO G., WANG H., ZHENG M., LI P., LIU L., WEN J., 2015 – Identification of loci and genes for growth related traits from a genome-wide association study in a slow-×fast-growing broiler chicken cross**.** ***Genes Genom*****,** 37, 829-836.
35. LUO W., LUO C., WANG M., GUO L., CHEN X., LI Z., ZHENG M., FOLANIYI B.S., LUO W., SHU D., SONG L., 2020 – Genome diversity of Chinese indigenous chicken and the selective signatures in Chinese gamecock chicken. **S*cientific Reports*** 10(1), 1-14.
36. LYIMO C.M., WEIGEND A., JANßEN-TAPKEN U., MSOFFE P.L., SIMIANER H. WEIGEND S., 2013 – Assessing the genetic diversity of five Tanzanian chicken ecotypes using molecular tools. ***South African Journal of Animal Science*** 43, 499-510.
37. LYNCH M. WALSH B., 1998 – Genetics and Analysis of Quantitative Traits. Sunderland, MA.
38. MEYER R.C., WEIGELT-FISCHER K., KNOCH D., HEUERMANN M.C., ZHAO Y. ALTMANN T., 2020 – Temporal dynamics of QTL effects on vegetative growth in Arabidopsis thaliana. bioRxiv.
39. MORIN P.A., LUIKART G. WAYNE R.K., 2004 – SNPs in ecology, evolution and conservation. ***Trends*** *in* ***Ecology & Evolution*** 19, 208-216.
40. MUCHADEYI F.C., EDING H., WOLLNY C.B.A., GROENEVELD E., MAKUZA S.M., SHAMSELDIN R., SIMIANER H. WEIGEND S., 2007 – Absence of population substructuring in Zimbabwe chicken ecotypes inferred using microsatellite analysis. ***Animal Genetics*** 38, 332-339.
41. MUIR W.M., WONG G.K.S., ZHANG Y., WANG J., GROENEN M.A., CROOIJMANS R.P., MEGENS H.J., ZHANG H., OKIMOTO R., VEREIJKEN A. JUNGERIUS A., 2008 – Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. Proceedings of the National Academy of Sciences of the United States of America 105, 17312-17317.
42. MWACHARO J.M., BJØRNSTAD G., MOBEGI V., NOMURA K., HANADA H., AMANO T., JIANLIN H. HANOTTE O., 2011 – Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. ***Molecular Phylogenetics and Evolution*** 58, 374-382.
43. NEI M., 1987 – Molecular Evolutionary Genetics. Columbia University Press.

44. NEJATI-JAVAREMI A., SMITH C., GIBSON J.P., 1997 – Effect of total allelic relationship on accuracy of evaluation and response to selection. ***Journal of Animal Science*** 75, 1738-1745.
45. OTT J., RABINOWITZ D., 1997 – The effect of marker heterozygosity on the power to detect linkage disequilibrium. ***Genetics*** 147, 927-930.
46. PALINKAS-BODZSAR N., SZTAN N., MOLNAR T. HIDAS A., 2020 – Gene conservation of six Hungarian local chicken breeds maintained in small populations over time. ***PloS ONE*** 15, 0238849.
47. PINTUS E., SORBOLINI S., ALBERA A., GASPA G., DIMAURO C., STERI R., MARRAS G., MACCIOTTA N.P., 2014 – Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds. ***Animal Genetics*** 45(1), 1-11.
48. PONSUKSILI S., WIMMERS K. HORST P., 1996 – Genetic variability in chickens using polymorphic microsatellite makers. ***Thai Journal of Agricultural Science*** 29, 571-580.
49. PRITCHARD J.K., WEN W., FALUSH D., 2004 – Documentation for the STRUCTURE software Version 2. Chicago.
50. PRITCHARD J.K., STEPHENS M., DONNELLY P., 2000 – Inference of population structure using multilocus genotype data. ***Genetics*** 155, 945-959.
51. PRITCHARD J.K., WEN W., FALUSH D., 2003 – Documentation for STRUCTURE software: Version 2.
52. PURCELL S., NEALE B., TODD-BROWN K., THOMAS L., FERREIRA M.A., BENDER D., MALLER J., SKLAR P., DE BAKKER P.I., DALY M.J., SHAM P.C., 2007 – PLINK: a tool set for whole-genome association and population-based linkage analyses. ***American Journal of Human Genetics*** 81, 559-575.
53. QANBARI S., GIANOLA D., HAYES B., SCHENKEL F., MILLER S., MOORE S., THALLER G., SIMIANER H., 2011 – Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. ***BMC Genomics*** 12(1), 318.
54. ROH H.J., KIM K.W., LEE J.W., JEON D.Y., KIM S.C., JEON I.S., KO Y.G., LEE J.H., KIM S.H., BAEK J.J., OH D.Y., 2018 – Genetic Diversity and Relationship of Ogye Population in Korea Using 25 Microsatellite Markers. ***Korean Journal*** *of* ***Poultry Science*** 45, 229-236.
55. ROH H.J., KIM S.C., CHO C.Y., LEE J., JEON D., KIM D.K., KIM K.W., AFRIN F., KO Y.G., LEE J.H., BATSAIKHAN S., 2020 – Estimating genetic diversity and population structure of 22 chicken breeds in Asia using microsatellite markers. ***Asian-Australasian Journal of Animal Sciences*** 33(12), 1896-1904.
56. ROSENBERG N.A., BURKE T., ELO K., FELDMAN M.W., FREIDLIN P.J., GROENEN M.A., HILLEL J., MÄKI-TANILA A., TIXIER-BOICHARD M., VIGNAL A., WIMMERS K., 2001 – Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. ***Genetics*** 159(2), 699-713.
57. SCHORK N.J., 2001 – Genome partitioning and whole-genome analysis. ***Advanced Genetics*** 42, 299-322.
58. SEO D.W., HOQUE M.R., CHOI N.R., SULTANA H., PARK H.B., HEO K.N., KANG B.S., LIM H.T., LEE S.H., JO C., LEE, J.H., 2013 – Discrimination of Korean native chicken lines using fifteen selected microsatellite markers. ***Asian-Australasian Journal of Animal Sciences*** 316322.
59. SLATKIN M., 1994 – Linkage disequilibrium in growing and stable populations. ***Genetics*** 137, 331-336.
60. SUN Y., ZHAO G., LIU R., ZHENG M., HU Y., WU D., ZHANG L., LI P., WEN J., 2013 – The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study. ***BMC Genomics*** 14, 458.
61. TANG Y., LIU X., WANG J., LI M., WANG Q., TIAN F., ZHANG Z., 2015 – GAPIT version 2: an enhanced integrated tool for genomic association and prediction. ***Plant Genome*** 9, 2.

62. TWITO T., WEIGEND S., BLUM S., GRANEVITZE Z., FELDMAN M.W., PERL-TREVES R., LAVI U., HILLEL J., 2007 – Biodiversity of 20 chicken breeds assessed by SNPs located in gene regions. ***Cytogenetic and Genome Research*** 117, 319-326.
63. VAN MARLE-KÖSTER E., HEFER C.A., NEL L.H., GROENEN M.A.M., 2008 – Genetic diversity and population structure of locally adapted South African chicken lines: Implications for conservation. ***South African Journal of Animal Science*** 38, 271-281.
64. WANG D., SUN Y., STANG P., BERLIN J.A., WILCOX M.A., LI Q., 2009 – Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. ***BMC Proceedings*** 3, 1-6.
65. WIMMER V., ALBRECHT T., AUINGER H.J., SCHÖN, C.C., 2012 – Synbreed: a framework for the analysis of genomic prediction data using R. ***Bioinformatics*** 28, 2086-2087.
66. WRAGG D., MWACHARO J.M., ALCALDE J.A., HOCKING P.M., HANOTTE O., 2012 – Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. ***Heredity*** 109(1), 6-18.
67. WRIGHT S., 1951 – The genetical structure of populations. ***Annals of Eugenics*** 15, 323-354.
68. YAMAMOTO Y., OKA T., FARUQUE MO., AMANO T., 2011 – Genetic diversity and genetic relationship of Bangladeshi native chickens. In: Proceedings of the 7h International Poultry Show and Seminar. World's Poultry Science Association. Bangladesh Branch 207215.
69. ZHANG X., LEUNG F.C., CHAN D.K., WU, C., 2002 – Genetic diversity of Chinese native chicken breeds based on protein polymorphism, randomly amplified polymorphic DNA, and Microsatellite polymorphism. ***Poultry Science*** 81, 14631472.
70. ZHANG M., HAN W., TANG H., LI G., ZHANG M., XU R., LIU Y., YANG T., LI W., ZOU J., WU, K., 2018 – Genomic diversity dynamics in conserved chicken populations are revealed by genome-wide SNPs. ***BMC Genomics*** 19(1), 598.
71. ZHAO F., MCPARLAND S., KEARNEY F., DU L., BERRY D.P., 2015 – Detection of selection signatures in dairy and beef cattle using high-density genomic information. ***Genetics Selection Evolution*** 47(1), 49.