Animal 15 (2021) 100171

Contents lists available at ScienceDirect

Animal
The international journal of animal biosciences

Assessing the effects of rare alleles and linkage disequilibrium on estimates of genetic diversity in the chicken populations

N.V. Dementieva [a], O.V. Mitrofanova [a], A.P. Dysin [a], A.A. Kudinov [a], O.I. Stanishevskaya [a], T.A. Larkina [a], K.V. Plemyashov [a], D.K. Griffin [b], M.N. Romanov [b,*], M.G. Smaragdov [a]

[a] Russian Research Institute of Farm Animal Genetics and Breeding — Branch of the L. K. Ernst Federal Science Centre for Animal Husbandry, Pushkin, St Petersburg, Russia
[b] School of Biosciences, University of Kent, Canterbury, Kent, UK

ARTICLE INFO

Article history:
Received 29 September 2020
Received in revised form 16 December 2020
Accepted 18 December 2020
Available online 6 February 2021

Keywords:
Single nucleotide polymorphisms
Fixation index
Heterozygosity
Linkage disequilibrium pruning
Poultry

ABSTRACT

Phenotypic diversity in poultry has been mainly driven by artificial selection and genetic drift. These led to the adaptation to the environment and the development of specific phenotypic traits of chickens in response to their economic use. This study evaluated genetic diversity within and between Russian breeds and populations using Illumina Chicken 60 K SNP iSelect BeadChip by analysing genetic differences between populations with Hudson's fixation index (F_{ST} statistic) and heterozygosity. We estimated the effect of rare alleles and linkage disequilibrium (**LD**) on these measurements. To assess the effect of LD on the genetic diversity population, we carried out the LD-based pruning (LD < 0.5 and LD < 0.1) for seven chicken populations combined (I) or separately (II). LD pruning was specific for different dataset groups. Because of the noticeably large sample size in the Russian White RG population, pruning was substantial for Dataset I, and F_{ST} values were only positive when LD < 0.1 pruning was applied. For Dataset II, the LD pruning results were confirmed by examining heterozygosity and alleles' frequency distribution. LD between single nucleotide polymorphisms was consistent across the seven chicken populations, except the Russian White RG population with the smallest r^2 values and the largest effective population size. Our findings suggest to study variability in each population LD pruning has to be carried separately not after merging to avoid bias in estimates.

Implications

In this research, methodological approaches to evaluating genetic structure in resource and commercial chicken populations representing part of worldwide poultry germplasm were studied using the Illumina Chicken 60 K iSelect BeadChip SNP chip and linkage disequilibrium-based pruning of datasets. This study provides insight into how ignoring the effect of monomorphic and rare SNP alleles can lead to misleading results obtained with F_{ST} statistics. Expected implications of the study include improvement in assessing the genetic diversity of animal populations, especially those derived from a small number of founders and experienced closed breeding, and incorporation of these findings into the poultry breeding and selection programmes.

Introduction

Farm animals used in modern agriculture are products of domestication and selection of their wild ancestors and evolve like wild forms. However, as a consequence of human activity and artificial selection, this evolutionary process is more predictable owing to more stable selection regimes (Darwin, 1875). In addition to the involvement of previously existing gene complexes in their genomes, new alleles occurring due to mutations could be picked up in the evolution process of domestic forms. From this point of view, molecular diversity studies in domestic breeds are aimed at assessing these genetic changes (Andersson, 2001; Weigend and Romanov, 2001; Khanyile et al., 2015; Bortoluzzi et al., 2018).

The breeds are divided into commercial, intensively selected for economic purposes, and local breeds that continue to bear valuable features and genetic diversity (Andersson, 2001; Weigend and Romanov, 2001; Mahammi et al., 2014; Berrezoug et al., 2019). On the basis of genetic diversity evaluation using polymorphic markers, genetic distances between breeds can be estimated, for example, by fixation index (F_{ST}) values and principal component analysis. Breeds produced as a result of strong selection pressure demonstrate a lower genetic diversity and greater genetic distances relative to other breeds. Variability reduction may be due to a small number of founders as well as a limited effective population size (N_e) or the use of inbreeding (e.g., Weigend and Romanov, 2001 and 2002; Hillel et al., 2003; Muir et al., 2008; Granevitze et al., 2009; Eltanany and Distl, 2010). Previously widely used biochemical (protein) markers and, later, mitochondrial,

* Corresponding author.
E-mail address: m.romanov@kent.ac.uk (M.N. Romanov).

https://doi.org/10.1016/j.animal.2021.100171

minisatellite and microsatellite and some other DNA markers (e.g., Weigend and Romanov, 2001 and 2002; Romanov and Weigend, 2001a and 2001b; Moiseyeva et al., 2003; Soller et al., 2006; Tyshchenko et al., 2007; Mahammi et al., 2016; Deniskova et al., 2017; Teinlek et al., 2018) have sparse localization across a genome and not always objectively reflect the genetic features of various breeds at the genome level. Single nucleotide polymorphisms (**SNPs**) markers applied for genotyping DNA samples using arrays allow simultaneously to cover the whole genome and associate the discovered variations at these loci with specific phenotypic breed features. Subsequent genetic analysis makes it possible to reveal if the genotyped individuals belong to a particular breed, and evaluate the molecular architecture of populations, as well (e.g., Weigend and Romanov, 2001; Wragg et al., 2012; Fariello et al., 2013; Mekchay et al., 2014; Khanyile et al., 2015; Smaragdov et al., 2016 and 2018; Dementeva et al., 2017; Verdugo et al., 2019). The extent of divergence between populations depends upon within-population variability. The molecular organization of the genome of Russian White breed (RG) was compared to RS (Table 1) to assess historical changes (Dementeva et al., 2017). The genetic diversity of populations is an important source of information for making decisions on the subsequent conservation of breeds as well as their effective breeding and exploitation. However, some researchers have suggested that with a very limited number of well-distributed SNPs in the genome we can perform the genetic characterization of domestic animal populations, as this has been proven in cattle (Matukumalli et al., 2009).

To analyse genetic variation between populations or within breeds, various statistical tools are implemented, and among them, Wright's F_{ST} statistic is often used for this purpose (Weir and Cockerham, 1984; Weir and Hill, 2002). Many F_{ST} estimators were proposed (Bhatia et al., 2013) including those that are the most widely used and developed by Weir and Cockerham (1984) and Nei (1986). The former statistic depends on the number of animals in the compared populations, while the latter one systematically overestimates F_{ST} (Bhatia et al., 2013). Patterson et al. (2006) suggested the use of a Hudson's F_{ST} statistic as the most suitable in the case of SNP markers. It is not sensitive to the number of animals in the population, does not overestimate F_{ST}, and is less dependent on the assortment bias than other known statistics. Due to these features, we used Hudson's statistic in the present study. Because, to the best of our knowledge, the relationship between F_{ST}, in particular, that of Hudson's statistic, and linkage disequilibrium (**LD**) of SNPs on the chicken have not been evaluated, we decided to explore this, as well.

This study aimed to evaluate the effect of LD and rare alleles on genetic differences between breeds, hybrids, and chicken lines using F_{ST} statistics and LD-based pruning of datasets.

Material and methods

Samples

The material for the study was DNA isolated from the blood of 282 chickens (*Gallus gallus*) from seven Russian populations (Table 1) that have a different history of origin and included only unrelated individuals.

The first three populations, RS, RP and RG, belonged to the Russian White breed developed by crossing local Russian chickens with White Leghorns (Paronyan and Yurchenko, 1989). The initial population of this breed is represented by the RP group. The inbred RS population was under selection pressure for a long time (1953 to 2003). This population had one founder, and the breeding was carried out using strong selection for the tolerance of chicks to cold temperatures (Sokolova, 1999). The RG population is a descendant of the inbred RS population; however, to improve its reproductive traits, an additional crossing with White Leghorns was done in 2006.

The C1 population was composed of the two-line Cornish F_1 hybrids resulted from crossing the commercial parent lines. The populations C2A and C2B were the commercial lines of the C2 cross-bred in Russia. At their breeding, a crossing with a line of the Cornish breed from a related cross was produced.

The RI population comprised the descendants of a paternal line from a commercial cross that has been bred inter se since 1982. In 1988, it was crossed with Rhode Island Reds originated from a club of poultry fanciers in Alma-Ata, Kazakhstan and is at present bred as a closed population.

Genotyping

SNP genotyping of the 282 DNA samples was performed using the Illumina Chicken 60 K SNP iSelect BeadChip (Illumina, USA) and genotyping service of Geneseek/Neogen Corporation. The quality control of genotyped SNPs was performed using the PLINK 1.9 software program (Chang et al., 2015). At the initial dataset processing stage, SNPs were selected using a quality score: QS > 0.8. To eliminate the effect of sex chromosomes on the subsequent estimation, SNP markers located on them were excluded. Then, SNPs were for a total genotype dataset and filtered separately for each population using the PLINK criterion –geno 0.1. As a result, the total genotype dataset and the final genotyping data for the seven populations were produced.

Linkage disequilibrium-based filtering of single nucleotide polymorphisms

Two approaches of LD-based SNP data pruning were tested. The first type of SNP filtering, with LD < 0.1, was performed using the joint genotype dataset and the PLINK command –indep-pairwise 50 5 0.1. This dataset was designated Dataset I. The F_{ST} values before and after the filtering were analysed and compared.

The second type of LD-based SNP pruning was carried out for each population separately, and the resulted dataset was designated Dataset II. In every population SNPs with LD < 0.5 and LD < 0.1 were selected using the PLINK commands: –indep-pairwise 50 5 0.5, and –indep-pairwise 50 5 0.1, respectively (Liu, 2017). Accordingly, the three LD-based subsets were generated: without LD filter, with LD < 0.1 and with LD < 0.5.

Table 1
List of the chicken breeds, lines and crosses used in the study.

Breed/population	Abbreviation	*N*	Sampling year	Source
Russian White	RS	6	2001	RRIFAGB
Russian White	RP	11	2001	All-Russian Poultry Research and Technological Institute
Russian White	RG	170	2016	Bioresource collection, RRIFAGB
White Cornish two-way hybrids, Cross 1	C1	38	2016	Bioresource collection, RRIFAGB
White Cornish, Cross 2, line A	C2A	18	2006	State Poultry Breeder Farm Smena
White Cornish, Cross 2, line B	C2B	20	2006	State Poultry Breeder Farm Smena
Rhode Island Red	RI	19	2016	Bioresource collection, RRIFAGB

Abbreviations: *N* = number of individuals studied; RRIFAGB = Russian Research Institute of Farm Animal Genetics and Breeding.

Within and between populations genetic differences

The determination of genetic diversity between populations was carried out using Hudson's F_{ST} statistic as implemented the EIGENSOFT 6.1.4 software package (Patterson et al., 2006). To calculate heterozygosity and minor allele frequencies (**MAF**), PLINK 1.9 was used.

Effective population size

Linkage disequilibrium and effective population size were analysed separately for each population using PLINK 1.9. The consecutive chromosome intervals (bins) between SNPs were chosen at 4.2–8.4, 8.4–16.6, 16.6–34, 34–65.5, 65.5–140, 140–250, 250–650 and 650–800 Kb window sizes.

The mean LD (r^2) values were calculated for each bin plot (Fig. 1) that was visualized according to the following formula:

$$E\left[r^2_{a\,i}\right] = (\alpha + 4\,N_e \times c)^{-1}$$

where N_e is the effective population size, $\alpha = 1$, c is the distance between the SNP markers in Morgan units, if the recombination rate is averaged over the entire genome and $r^2_{a\,i}$ is the mean value in the interval a i (Corbin et al., 2012).

Results

Within and between populations genetic divergence

The total SNP genotype dataset included variants at 57 636 SNP loci in the seven chicken populations. After completing the quality control of genotyped SNPs, the initial dataset was reduced to 42 586 SNPs. To evaluate the effect of LD, we used two datasets, I and II.

Dataset I consisted of SNP genotypes in the seven chicken populations, which were LD < 0.1 pruned jointly, that is, the populations were merged together before pruning. As a result of this filtering, 39 125 SNPs were removed, and 3 461 SNPs remained in Dataset I. Changes in MAF values after LD-based pruning are shown in Table 2 (see also Supplementary Fig. S1). In Table 3, F_{ST} values and their respective standard errors are shown before and after LD pruning. The most genetically similar to other chicken populations was the RI population, which after LD pruning differed noticeably from other chicken populations (mean pairwise $F_{ST} = 0.164$; $P < 0.001$). Thus, LD pruning had some effect on estimates of mean genetic differences between populations ($P < 0.001$).

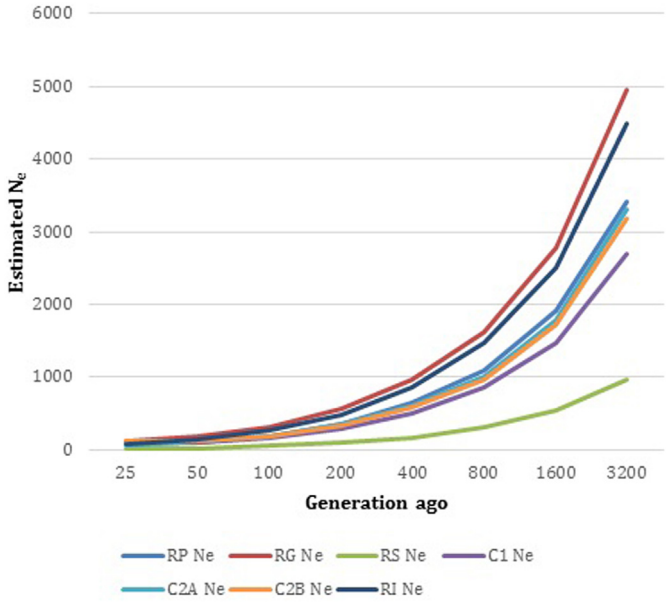

Fig. 1. Estimates of effective population size (N_e) in the experimental and commercial studied chicken populations. See Table 1 for the chicken population descriptions.

Table 2
Frequency distribution of SNP alleles for the studied chicken populations[1] in the initial dataset, and linkage disequilibrium (LD) pruned Datasets I and II. Minor allele frequencies (MAF) intervals corresponded to bins 0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, 0.4–0.5 and 0 (monomorphic alleles).

MAF intervals	Frequency distribution of SNPs alleles in the populations studied						
	RG	RS	RP	C1	C2A	C2B	RI
Initial sample set							
0–0.1	0.27	0.09	0.17	0.15	0.10	0.11	0.13
0.1–0.2	0.19	0.10	0.16	0.15	0.17	0.17	0.19
0.2–0.3	0.18	0.09	0.16	0.19	0.16	0.19	0.16
0.3–0.4	0.17	0.09	0.16	0.15	0.24	0.22	0.23
0.4–0.5	0.17	0.14	0.20	0.17	0.22	0.21	0.20
0	0.03	0.49	0.15	0.18	0.10	0.10	0.10
Dataset I (LD < 0.1 pruned)							
0–0.1	0.25	0.09	0.11	0.14	0.12	0.12	0.13
0.1–0.2	0.06	0.11	0.13	0.13	0.15	0.15	0.18
0.2–0.3	0.06	0.12	0.15	0.17	0.14	0.16	0.14
0.3–0.4	0.20	0.13	0.19	0.15	0.22	0.20	0.22
0.4–0.5	0.40	0.18	0.26	0.16	0.21	0.20	0.21
0	0.03	0.37	0.16	0.25	0.16	0.17	0.13
Dataset II (LD < 0.5 pruned)							
0–0.1	0.24	0.17	0.16	0.20	0.14	0.15	0.15
0.1–0.2	0.19	0.19	0.13	0.17	0.18	0.18	0.20
0.2–0.3	0.18	0.13	0.15	0.20	0.16	0.19	0.15
0.3–0.4	0.18	0.36	0.16	0.16	0.21	0.20	0.21
0.4–0.5	0.22	0.19	0.40	0.27	0.31	0.28	0.29
Dataset II (LD < 0.1 pruned)							
0–0.1	0.27	0.04	0.01	0.05	0.02	0.02	0.01
0.1–0.2	0.03	0.06	0.01	0.02	0.01	0.01	0.01
0.2–0.3	0.03	0.06	0.01	0.04	0.02	0.02	0.01
0.3–0.4	0.06	0.11	0.03	0.08	0.05	0.06	0.05
0.4–0.5	0.61	0.71	0.95	0.81	0.91	0.89	0.92

[1] See Table 1 for the chicken population descriptions.

Owing to the effects of pruning F_{ST} values between the remotely related populations (pairs RG–C1, RG–C2A and RG–C2B) were disproportionately lowered as compared to a slight decrease between the closely related populations (pairs of RG, RS and RP (Table 3). However, LD pruning had a slight effect on the greater F_{ST} difference between the inbreed RS population and other chicken populations ($P < 0.01$), despite the high level of inbreeding in the RS population. It should be borne in mind that there was a significantly larger sample size of RG over the other chicken populations and, accordingly, a major effect of LD-based pruning was observed in this population in the case of Dataset I.

Dataset II was based on SNP genotypes in the seven populations, which were LD pruned separately. Before LD filtering, the populations were genotyped with 32–37 thousand SNPs. After LD < 0.5 pruning, 10–23 thousand SNPs remained, and after LD < 0.1 filtering 1.9–2.9 thousand SNPs were left, meaning that the number of SNPs was noticeably reduced. Changes in MAF values due to LD-based pruning are given in Table 2 (see also Supplementary Fig. S2).

It was also found that the RG population had the lowest number of monomorphic SNPs (Table 2; Supplementary Fig. S1) and the minimum heterozygosity (Fig. 2) among other chicken populations (Table 4). After pruning, heterozygosity in the RS population raised insignificantly. In the population RI, RG, RP and C1A, a significant within population heterozygosity elevation was observed using Dataset II (Table 4). In the RS population, the number of monomorphic SNPs prevailed (Table 2; Supplementary Fig. S1) and, as a result of LD < 0.1 pruning of Dataset II, the number of SNPs with MAF = 0.4–0.5 was the greatest in all cases (Table 2; Supplementary Fig. S2), while the RG population also had a

Table 3
Estimates of fixation index F_{ST}[1] before (above diagonal) and after linkage disequilibrium (LD) < 0.1 pruning (below diagonal) for the joint Dataset I in the studied chicken populations.

Chicken populations[2]	RG	RS	RP	C1	C2A	C2B	RI
RG	×	0.112 (0.003)	0.102 (0.002)	0.271 (0.007)	0.220 (0.005)	0.234 (0.005)	0.175 (0.003)
RS	0.109 (0.004)	×	0.209 (0.004)	0.388 (0.007)	0.335 (0.005)	0.348 (0.006)	0.296 (0.006)
RP	0.099 (0.003)	0.195 (0.007)	×	0.248 (0.007)	0.197 (0.005)	0.212 (0.005)	0.157 (0.005)
C1	0.187 (0.005)	0.302 (0.008)	0.201 (0.006)	×	0.150 (0.004)	0.161 (0.004)	0.202 (0.006)
C2A	0.135 (0.004)	0.247 (0.007)	0.147 (0.005)	0.141 (0.005)	×	0.011 (0.001)	0.151 (0.004)
C2B	0.144 (0.004)	0.257 (0.007)	0.158 (0.005)	0.151 (0.005)	0.012 (0.001)	×	0.164 (0.004)
RI	0.129 (0.004)	0.246 (0.007)	0.136 (0.004)	0.191 (0.007	0.134 (0.005)	0.145 (0.005)	×
Mean F_{ST} (LD < 0.1)[3]	0.134 (0.0017)	0.226 (0.0028)	0.156 (0.0026)	0.196 (0.0025)	0.136 (0.0020)	0.144 (0.0020)	0.164 (0.0021)

[1] SE values are given in parentheses.
[2] See Table 1 for the chicken population descriptions.
[3] For pairwise comparison of mean F_{ST} values, the difference was significant between C2A–C2B and RG–C2B pairs at $P < 0.01$, and in all other cases at $P < 0.001$.

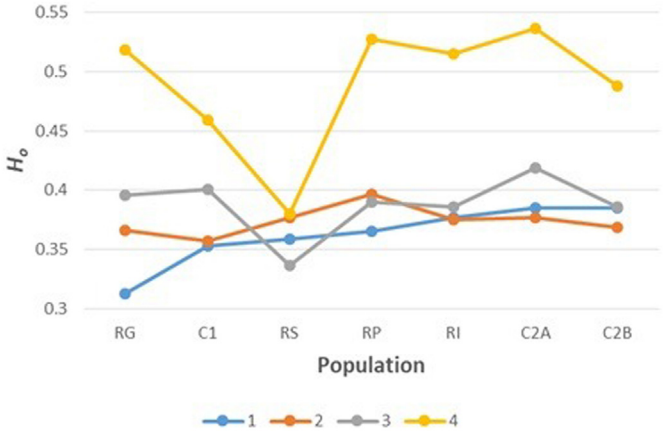

Fig. 2. Heterozygosity (H_o) in the studied chicken populations: 1, initial sample set; 2, LD < 0.5 pruned Dataset II; 3, LD < 0.1 pruned Dataset II; 4, LD < 0.1 Dataset I. See Table 1 for the chicken population descriptions; LD = linkage disequilibrium.

significant proportion of SNPs with MAF < 0.1 after LD < 0.5 pruning (Table 2; Supplementary Fig. S2).

Dependence of linkage disequilibrium on distance between single nucleotide polymorphisms

As seen in Fig. 3, LD is presented as a function of the distance between SNP markers for the seven chicken populations. The distribution of r^2 values in the bins was analogous in the Russian breeds, lines and crosses with the exception of the population RS and RG. For RG, most SNP markers had lower r^2 values. The inbred RS population had a greater LD in all bins, except for the largest interval between the markers (450–500 Kb) where most markers were characterized by $r^2 \leq 0.2$ (Fig. 3E). Thus, there was observed consistency of LD across the studied chicken populations.

Effective population size

As seen in Fig. 1, 25 generations ago the lowest effective population size was found in RS (14), and the greatest one in the RG (124) and C2B (130) populations. On the other hand, 3 200 generations ago effective population size ranged between 962 (RS) and 4 484 (RI) to 4 949 (RG). The commercial Cornish population C1 had an effective population size of 2 694. The remaining chicken populations, C2A, C2B and RP had intermediate values of 3 188 to 3 406.

Discussion

Domestication of the chicken was accompanied by population bottlenecks and subsequent population growth, an admixture of populations, inbreeding, genetic drift and selective breeding. As a consequence of these demographic and selective events, the genetic variation within the domesticated chicken genome might have changed from its ancestral state. Analysis of the samples of animals from populations may face a problem of ascertainment bias. Firstly, obtained marker data does not provide a random SNP polymorphism (due to MAF bias) in a population of interest (Browning et al., 2004; Nielsen, 2004; Heslot et al., 2013). Secondly, the animals may be selected not by random. MAF bias results in the over-representation of polymorphisms with greater MAF values and the under-representation of polymorphisms with lower MAF values. The number of animals in the sample will influence the lower frequency limits of SNPs. Analysis of ascertainment bias in genomic data has largely been conducted on human data (Albrechtsen et al., 2010; Wang and Nielsen, 2012; Lachance and Tishkoff, 2013). To evaluate the effect of monomorphic and rare alleles

Table 4
Heterozygosity in the studied chicken populations.

Chicken populations[2]	H_o[1]			
	Full[3]	Dataset I	Dataset II	
		LD < 0.1	LD < 0.5	LD < 0.1
RS	0.359 (0.009)	0.377 (0.008)	0.337 (0.005)	0.380 (0.004)
RP	0.365 (0.005)	0.397 (0.005)	0.390 (0.003)	0.527 (0.004)
C1	0.353 (0.002)	0.357 (0.002)	0.401 (0.014)	0.459 (0.028)
C2A	0.385 (0.002)	0.377 (0.002)	0.419 (0.007)	0.536 (0.016)
C2B	0.381 (0.002)	0.369 (0.002)	0.386 (0.001)	0.488 (0.003)
RI	0.377 (0.003)	0.375 (0.004)	0.386 (0.001)	0.515 (0.003)
RG	0.313 (0.005)	0.366 (0.004)	0.396 (0.002)	0.518 (0.006)

[1] H_o = mean observed heterozygosity and its SE (given in parentheses).
[2] See Table 1 for the chicken population descriptions.
[3] Initial dataset before linkage disequilibrium (LD) pruning.

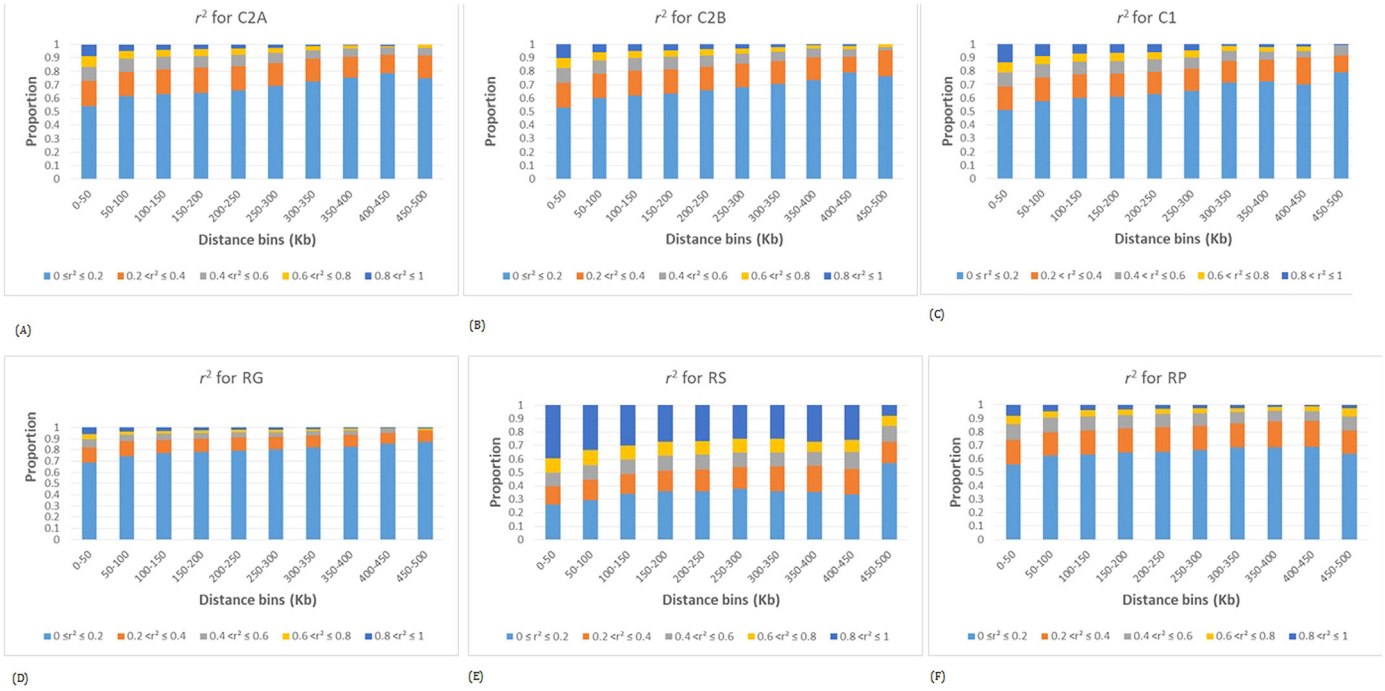

Fig. 3. Comparison of the frequency distribution of linkage disequilibrium (LD) by mean r^2 values in distance bins (0–50, 50–100, 100–150, 150–200, 200–250, 250–300, 300–350, 350–400, 400–450 and 450–500 Kb) in the seven chicken populations: (A) C2A, (B) C2B, (C) C1, (D) RG, (E) RS, (F) RP and (G) RI. See Table 1 for the chicken population descriptions.

in chicken populations, we examined two datasets. Dataset I included the seven chicken populations, which were LD pruned jointly, and in Dataset II they were LD pruned separately.

Distribution of minor allele frequencies

Distribution of MAF value spectrum depends on the SNP array used or sequence data obtained (Malomane et al., 2018) as well as LD-based pruning. The almost uniform distribution across frequency classes in the case of SNP array relies on the ascertainment bias, as discussed by Muir et al. (2008). In consequence of designing an SNPs array, a significant part of rare alleles is lost.

When the seven chicken populations were LD pruned jointly in Dataset I, the RG population (Table 2; Supplementary Fig. S1) significantly differed from other populations in the number of monomorphic SNPs. The other chicken populations were fairly uniform in terms of frequencies of SNP alleles. Only for the RG population, this pattern changed significantly as a result of LD < 0.1 pruning (Table 2; Supplementary Fig. S1B). This phenomenon can be explained by the largest sample size of RG in the total dataset. Basically, different results were obtained for Dataset II where the seven chicken populations were LD pruned separately. SNPs with frequencies 0.1–0.4 were fairly uniformly represented in different populations, except for MAF 0.4–0.5, the proportion of which greatly increased (Table 2; Supplementary Fig. S2). However, in the RG population, many SNPs with MAF ≤ 0.1 have remained. The excess of SNPs with MAF ≤ 0.1 was due to the partial introgression of Leghorn alleles into the RG population genome from 2004 to 2006. As a result, rare SNPs were fixed. This fact may evidence that the genetic differences in the studied chicken populations are determined primarily by monomorphic and rare SNPs. Thus, with LD pruning it is possible to reveal recent genetic introgression from other breeds or populations. The distribution of MAF depends essentially on how the chicken populations were LD-based pruned, i.e., jointly or separately. As a result, loss of rare SNPs and ascertainment bias of F_{ST} may occur. Removal of SNPs in disequilibrium resulted in a decrease of rare and monomorphic alleles and excess of the common alleles (with MAF = 0.4–0.5) for all the populations (Table 2; Supplementary Fig. S2). A distinctive property of the RG population is the retention of rare SNPs alleles as a result of LD pruning (Table 2; Supplementary Fig. S2).

Effect of linkage disequilibrium on fixation index

The LD depends on drift, the intensity of selection and the effective population size. In this study, we used the seven chicken populations to find out the dependence of F_{ST} on LD by analysing the relationship between LD-based pruning and F_{ST}. The basic idea of LD-based pruning is to remove markers, which are highly correlated with other markers within a given window, keeping markers with lower LD to each other in the dataset. This is an efficient way to reduce the multicollinearity effect, which may result in overestimation of effects of SNPs due to highly correlated SNPs (Malomane et al., 2018).

To tackle this problem, the effect of LD pruning on F_{ST} was studied when LD pruning of samples was conducted jointly or separately for each population. It turned out that the use of these two approaches led to substantially different F_{ST} values. If the total dataset was LD < 0.1 pruned jointly, the effect of decreasing F_{ST} values was not so noticeable (3). This was caused by the difference in the haplotypes across the chicken populations. As a result, recombinant active SNPs, which are more common between populations, remained, whereas population-specific SNPs were removed. This was a trend for all the seven chicken populations examined.

Analysis of heterozygosity revealed its considerable decrease when applying LD filter in separate chicken populations (Fig. 2). Changes in the level of heterozygosity due to pruning may be related to the history of populations. In populations where the introductory cross was applied, a significant rise in heterozygosity was observed after the pruning of Dataset II. Moreover, it was shown that F_{ST} values could depend heavily on the level of variation present in a sample, and the frequency of the most frequent allele (Jakobsson et al., 2013). Indeed, Jost (2008) argued that F_{ST} might be so affected by genetic diversity that it should not be used as a measure of population differentiation, gene flow or relatedness (Jost, 2008).

Mathieson and McVean (2012) show that populations can display spatial structure in rare variants even when F_{ST} is low, but that allele frequency-dependent metrics of allele sharing can reveal localized stratification. As described in Albrechtsen et al. (2010), the selection of loci that are polymorphic within populations decreases the estimates of F_{ST} between populations. This decrease in measured F_{ST} suggests lower differentiation between populations than it would be estimated from unbiased data. However, subpopulation-biased ascertainment can inflate F_{ST}, as well (Albrechtsen et al., 2010).

Multiple studies have shown inflated F_{ST} values calculated from ascertained SNPs as compared to whole-genome sequence data (Malomane et al., 2018). Thus, an increase of LD pruning for joint filtered populations leads to a decrease in the SNP share that determines the ancestry of the population and increases of the common SNP. Consequently, only when populations have a related origin, the use of LD pruning of joint populations may be appropriate. Since in our study the chicken populations largely differed in monomorphic and rare alleles and LD pruning could remove them, application of LD pruning might have led to an incorrect evaluation of inter-population differences.

It should be borne in mind that rare alleles are either recent mutations or ancient alleles are driven to lower allele frequencies through time due to drift or natural and artificial selection. These alleles have a higher risk for disappearing after a few generations; thus, in the framework of genetic diversity conservation, it may be desirable to put a higher priority on rare alleles as compared to common alleles to balance the potential loss of genetic diversity (Eynard et al., 2015). Ignoring these circumstances and issues can lead to ascertainment bias. Missing rare SNPs can lead to loss of valuable information and lessen the ability to detect those rare SNPs in association studies (Gorlov et al., 2008). This effect is confirmed through evaluation of variance in deregressed estimated breeding values or proofs in dairy cattle (Zhang et al., 2017) and effect of rare alleles on estimated genomic relationships from whole genome sequence data (Eynard et al., 2015).

Dependence of linkage disequilibrium on distance between single nucleotide polymorphisms

The extent of LD magnitude depends on demography genetics, drift and selection within populations and populations' admixture. In an ideal population model, mutability and recombination frequency determine LD. Since SNP mutability in chickens is ~10^{-8} per site/generation, the main factor should be the recombination frequency. The probability of recombination is known to lower with decreasing distance between markers; therefore, LD should be similar at short distances in different chicken populations. In the RS population with the lowest effective population size, there was a uniform decrease of LD at all distances between SNPs, except for a maximum distance of 450–500 Kb where r^2 value was similar to that in other chicken populations.

Of great importance for the analysis of LD consistency between the chicken populations are the values of LD (r^2) as a function of the distance between markers. Average LD was calculated for bins at distances from 0 to 500 Kb for all samples. Pure lines should have the largest LD at distances greater than 500 Kb (Qanbari et al., 2010). In our case, LD values do not differ significantly among tested populations including the C2A and C2B lines as well as the C1 hybrid. At short distances, up to 100 Kb, LD should have similar large values for commercial lines of laying crosses (Qanbari et al., 2010). For meat cross lines (Andreescu et al., 2007), $r^2 > 0.8$ at short distances is lower than for laying crosses. In the examined chicken populations, LD at short distances is even lower for almost all populations.

As found in our study, there was an unusual distribution of LD relative to the distance between SNPs markers in the RS population. This population is characterized by greater LD in all intervals between the markers; this is explained by the presence of a large number of extended haploblock due to a small number of unrelated individuals. Variation in LD between populations may be a result of the animal origin as it was shown for chickens (Khanyile et al., 2015; Dementeva et al., 2017) and sheep (Beynon et al., 2015). Overall, LD-based consistency is a good indicator of the chicken population relationships.

Effective population size

The origin of the breeds has a great effect on the effective population size. Populations with a single founder have a lower level of variability and greater LD level. The dynamics of the change in the effective population size can serve as an effective indicator of the historical effect of the number of the founders. The minimum population size was found for the RS population, which for more than 30 generations was bred using strong selection for the tolerance of the Russian White chickens to cold growing conditions (Fig. 1). Crossing them with the Leghorn breed was aimed at improving their reproductive traits and caused an increase of the effective population size of RG to the greatest value among the compared populations. The crossbred RI population demonstrated an intermediate to maximum N_e value. Intermediate N_e values were also observed in the populations C2A, C2B and RP (Fig. 1). For the commercial hybrid C1, the effective population size was close to that in the RS population, which might have caused by the effect of commercial inter-crossing and the resulted loss of genetic diversity of commercial birds (Andreescu et al., 2007; Qanbari et al., 2010; Seo et al., 2018).

Conclusion

In conclusion, an important evolutionary role of rare alleles should be once more noted, especially in terms of accelerating the selection of domestic animals. This study provided insights into how ignoring the effect of monomorphic and rare SNPs alleles might lead to misleading results obtained with F_{ST} statistics. Whole genome-wide SNP array analysis with F_{ST} statistic revealed the effect of monomorphic and rare alleles on genetic differences between the Russian chicken breeds, lines and crosses. As a recommendation, the poultry samples should be used without LD pruning in studying the evolutionary mechanisms of the recent selection history, especially in small populations. Not taking into account rare alleles may cause an ascertainment bias of F_{ST} values. Estimates of F_{ST} as a result of appropriate LD-based pruning might be useful for assessing the genetic diversity of populations derived from a small number of founders and experienced closed breeding leading to inbreeding and genetic drift.

Supplementary materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.animal.2021.100171.

Ethics approval

Sampling was approved by the Russian Research Institute of Farm Animal Genetics and Breeding (RRIFAGB) – Branch of the L. K. Ernst Federal Science Centre for Animal Husbandry and performed in accordance with the appropriate ethical guidelines. The authors declare that chicken blood samples were collected by trained personnel under strict veterinary requirements and minimizing animal suffering.

Data and model availability statement

None of the data was deposited in an official repository.

Author ORCIDs

N.V. Dementieva, 0000-0003-0210-9344; O.V. Mitrofanova, 0000-0003-4702-2736; A.P. Dysin, 0000-0002-4468-0365; A.A. Kudinov,

0000-0002-7811-576X; O.I. Stanishevskaya, 0000-0001-9504-3916; T. A. Larkina, 0000-0002-7764-1338; K.V. Plemyashov, 0000-0002-3658-5886; D.K. Griffin, 0000-0001-7595-3226; M.N. Romanov, 0000-0003-3584-4644; M.G. Smaragdov, 0000-0002-5087-6444.

Author contributions

N.V. Dementieva and M.G. Smaragdov: Conceptualization. O.V. Mitrofanova, A.P. Dysin, A.A. Kudinov, O.I. Stanishevskaya, T.A. Larkina and K.V. Plemyashov: Investigation. M.G. Smaragdov, A.A. Kudinov and N.V. Dementieva: Formal analysis. N.V. Dementieva, M.G. Smaragdov, D.K. Griffin and M.N. Romanov: Writing – Original Draft, Review & Editing. D.K. Griffin: Supervision. M.N. Romanov: Project Administration, Funding Acquisition.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicial to the impartiality of the reported research.

Acknowledgements

None.

Financial support statement

The study was supported by the Ministry of Science and Higher Education of the Russian Federation (State Assignment Program AAAA-A18-118021590138-1).

References

Albrechtsen, A., Nielsen, F.C., Nielsen, R., 2010. Ascertainment biases in SNP chips affect measures of population divergence. Molecular Biology and Evolution 27, 2534–2547.

Andersson, L., 2001. Genetic dissection of phenotypic diversity in farm animals. Nature Reviews. Genetics 2, 130–138.

Andreescu, C., Avendano, S., Brown, S.R., Hassen, A., Lamont, S.J., Dekkers, J.C., 2007. Linkage disequilibrium in related breeding lines of chickens. Genetics 177, 2161–2169.

Berrezoug, L., Bouchiba, I., Hedeili, N., Metahri, R., Ameur Ameur, A., Gaouar, S.B.S., 2019. Phenotypic and morphometric diversity of local chickens (BRAHMA) from wilaya of Tlemcen, Northwestern of Algeria. Genetics and Biodiversity Journal 3, 70–79.

Beynon, S.E., Slavov, G.T., Farré, M., Sunduimijid, B., Waddams, K., Davies, B., Haresign, W., Kijas, J., MacLeod, I.M., Newbold, C.J., Davies, L., Larkin, D.M., 2015. Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. BMC Genetics 16, 65.

Bhatia, G., Patterson, N., Sankararaman, S., Price, A.L., 2013. Estimating and interpreting F_{ST}: the impact of rare variants. Genome Research 23, 1514–1521.

Bortoluzzi, C., Crooijmans, R., Bosse, M., Hiemstra, S.J., Groenen, M., Megens, H.J., 2018. The effects of recent changes in breeding preferences on maintaining traditional Dutch chicken genomic diversity. Heredity 121, 564–578.

Browning, B.L., Brashear, D.L., Butler, A.A., Cyr, D.D., Harris, E.C., Nelsen, A.J., Yarnall, D.P., Ehm, M.G., Wagner, M.J., 2004. Linkage analysis using single nucleotide polymorphisms. Human Heredity 57, 220–227.

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, 7.

Corbin, L.J., Liu, A.Y., Bishop, S.C., Woolliams, J.A., 2012. Estimation of historical effective population size using linkage disequilibria with marker data. Journal of Animal Breeding and Genetics 129, 257–270.

Darwin, C.R., 1875. The Variation of Animals and Plants under Domestication. 2nd edition. volume 2. John Murray, London, UK.

Dementeva, N.V., Romanov, M.N., Kudinov, A.A., Mitrofanova, O.V., Stanishevskaya, O.I., Terletsky, V.P., Fedorova, E.S., Nikitkina, E.V., Plemyashov, K.V., 2017. Studying the structure of a gene pool population of the Russian White chicken breed by genome-wide SNP scan. Selskokhoziaĭstvennaia Biologiia [Agricultural Biology] 52, 1166–1174.

Deniskova, T.E., Dotsev, A.V., Bagirov, V.A., Wimmers, K., Reyer, H., Brem, G., Zinovieva, N. A., 2017. Biodiversity assessment in interspecies hybrids of the genus Ovis using STR and SNP markers. Selskokhoziaĭstvennaia Biologiia [Agricultural Biology] 52, 251–260.

Eltanany, M., Distl, O., 2010. Genetic diversity and genealogical origins of domestic chicken. World's Poultry Science Journal 66, 715–726.

Eynard, S.E., Windig, J.J., Leroy, G., van Binsbergen, R., Calus, M.P., 2015. The effect of rare alleles on estimated genomic relationships from whole genome sequence data. BMC Genetics 16, 24.

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M., Servin, B., 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics 193, 929–941.

Gorlov, I.P., Gorlova, O.Y., Sunyaev, S.R., Spitz, M.R., Amos, C.I., 2008. Shifting paradigm of association studies: value of rare single nucleotide polymorphism. American Journal of Human Genetics 82, 100–112.

Granevitze, Z., Hillel, J., Feldman, M., Six, A., Eding, H., Weigend, S., 2009. Genetic structure of a wide-spectrum chicken gene pool. Animal Genetics 40, 686–693.

Heslot, N., Rutkoski, J., Poland, J., Jannink, J.L., Sorrells, M.E., 2013. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLoS ONE 8, e74612.

Hillel, J., Groenen, M.A., Tixier-Boichard, M., Korol, A.B., David, L., Kirzhner, V.M., Burke, T., Barre-Dirie, A., Crooijmans, R.P., Elo, K., Feldman, M.W., Freidlin, P.J., Mäki-Tanila, A., Oortwijn, M., Thomson, P., Vignal, A., Wimmers, K., Weigend, S., 2003. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. Genetics, Selection, Evolution 35, 533–557.

Jakobsson, M., Edge, M.D., Rosenberg, N.A., 2013. The relationship between F_{ST} and the frequency of the most frequent allele. Genetics 193, 515–528.

Jost, L., 2008. G_{ST} and its relatives do not measure differentiation. Molecular Ecology 17, 4015–4026.

Khanyile, K.S., Dzomba, E.F., Muchadeyi, F.C., 2015. Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa. Frontiers in Genetics 6, 13.

Lachance, J., Tishkoff, S.A., 2013. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. Bioessays 35, 780–786.

Liu, Z.J. (Ed.), 2017. Bioinformatics in Aquaculture: Principles and Methods. John Wiley & Sons Ltd, Hoboken, NJ, USA.

Mahammi, F.Z., Gaouar, S.B.S., Tabet-Aoul, N., Tixier-Boichard, M., Saidi-Mehtar, N., 2014. Morpho-biometric characteristics and breeding systems of local chickens in the Oranie region (West Algeria). Cahiers Agricultures 23, 382–392.

Mahammi, F.Z., Gaouar, S.B., Laloë, D., Faugeras, R., Tabet-Aoul, N., Rognon, X., Tixier-Boichard, M., Saidi-Mehtar, N., 2016. A molecular analysis of the patterns of genetic diversity in local chickens from western Algeria in comparison with commercial lines and wild jungle fowls. Journal of Animal Breeding and Genetics 133, 59–70.

Malomane, D.K., Reimer, C., Weigend, S., Weigend, A., Sharifi, A.R., Simianer, H., 2018. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. BMC Genomics 19, 22.

Mathieson, I., McVean, G., 2012. Differential confounding of rare and common variants in spatially structured populations. Nature Genetics 44, 243–246.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P., Sonstegard, T.S., Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE 4, e5350.

Mekchay, S., Supakankul, P., Assawamakin, A., Wilantho, A., Chareanchim, W., Tongsima, S., 2014. Population structure of four Thai indigenous chicken breeds. BMC Genetics 15, 40.

Moiseyeva, I.G., Romanov, M.N., Nikiforov, A.A., Sevastyanova, A.A., Semyenova, S.K., 2003. Evolutionary relationships of Red Jungle Fowl and chicken breeds. Genetics, Selection, Evolution 35, 403–423.

Muir, W.M., Wong, G.K., Zhang, Y., Wang, J., Groenen, M.A., Crooijmans, R.P., Megens, H.J., Zhang, H., Okimoto, R., Vereijken, A., Jungerius, A., Albers, G.A., Lawley, C.T., Delany, M.E., MacEachern, S., Cheng, H.H., 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.

Nei, M., 1986. Definition and estimation of fixation indices. Evolution 40, 643–645.

Nielsen, R., 2004. Population genetic analysis of ascertained SNP data. Human Genomics 1, 218–224.

Paronyan, I.A., Yurchenko, O.P., 1989. Domestic fowl. In: Dmitriev, N.G., Ernst, L.K. (Eds.), Animal Genetic Resources of the USSR. Food and Agriculture Organization of the United Nations, Rome, Italy, pp. 437–468.

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genetics 2, e190.

Qanbari, S., Hansen, M., Weigend, S., Preisinger, R., Simianer, H., 2010. Linkage disequilibrium reveals different demographic history in egg laying chickens. BMC Genetics 11, 103.

Romanov, M.N., Weigend, S., 2001a. Using RAPD markers for assessment of genetic diversity in chickens. Archiv für Geflügelkunde 65, 145–148.

Romanov, M.N., Weigend, S., 2001b. Analysis of genetic relationships between various populations of domestic and jungle fowl using microsatellite markers. Poultry Science 80, 1057–1063.

Seo, D., Lee, D.H., Choi, N., Sudrajad, P., Lee, S.H., Lee, J.H., 2018. Estimation of linkage disequilibrium and analysis of genetic diversity in Korean chicken lines. PLoS ONE 13, e0192063.

Smaragdov, M.G., Saksa, E.I., Kudinov, A.A., Dementieva, N.V., Mitrofanova, O.V., Plemyashov, K.V., 2016. Genome-wide analysis of across herd F_{st} heterogeneity in holsteinized cattle. Russian Journal of Genetics 52, 173–179.

Smaragdov, M.G., Kudinov, A.A., Uimari, P., 2018. Assessing the genetic differentiation of Holstein cattle herds in the Leningrad region using F_{st} statistics. Agricultural and Food Science 27, 96–101.

Sokolova, A.N., 1999. Genetic and selection methods of creation of a chicken population with an increased resistance to neoplasms. Author's abstract, dissertation of the Doctor of Agricultural Sciences. RRIFAGB, Pushkin, St Petersburg, Russia.

Soller, M., Weigend, S., Romanov, M.N., Dekkers, J.C., Lamont, S.J., 2006. Strategies to assess structural variation in the chicken genome and its associations with biodiversity and biological performance. Poultry Science 85, 2061–2078.

Teinlek, P., Siripattarapravat, K., Tirawattanawanich, C., 2018. Genetic diversity analysis of Thai indigenous chickens based on complete sequences of mitochondrial DNA D-loop region. Asian-Australasian Journal of Animal Sciences 31, 804–811.

Tyshchenko, V.I., Mitrofanova, O.V., Dement'eva, N.V., Terletskii, V.P., Yakovlev, A.F., 2007. Estimation of genetic variability in the breeds and hen experimental populations by DNA-fingerprinting. Selskokhoziaĭstvennaia Biologiia [Agricultural Biology] 42, 29–33.

Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delser, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R., Hesse, P., Fulton, D., Sauer, E.W., Mohaseb, F.A., Davoudi, H., Khazaeli, R., Lhuillier, J., Rapin, C., Ebrahimi, S., Khasanov, M., Vahidi, S.M. F., MacHugh, D.E., Ertuğrul, O., Koukouli-Chrysanthaki, C., Sampson, A., Kazantzis, G., Kontopoulos, I., Bulatovic, J., Stojanović, I., Mikdad, A., Benecke, N., Linstädter, J., Sablin, M., Bendrey, R., Gourichon, L., Arbuckle, B.S., Mashkour, M., Orton, D., Horwitz, L.K., Teasdale, M.D., Bradley, D.G., 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science 365, 173–176.

Wang, Y., Nielsen, R., 2012. Estimating population divergence time and phylogeny from single-nucleotide polymorphisms data with outgroup ascertainment bias. Molecular Ecology 21, 974–986.

Weigend, S., Romanov, M.N., 2001. Current strategies for the assessment and evaluation of genetic diversity in chicken resources. World's Poultry Science Journal 57, 275–288.

Weigend, S., Romanov, M.N., 2002. The World Watch List for Domestic Animal Diversity in the context of conservation and utilisation of poultry biodiversity. World's Poultry Science Journal 58, 519–538.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Weir, B.S., Hill, W.G., 2002. Estimating F-statistics. The Annual Review of Genetics 36, 721–750.

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, 6–18.

Zhang, Q., Calus, M., Guldbrandtsen, B., Lund, M.S., Sahana, G., 2017. Contribution of rare and low frequency whole genome sequence variants to complex traits variation in dairy cattle. Genetics, Selection, Evolution 49, 60.