Received: 12 April 2022 | Accepted: 2 December 2022

DOI: 10.1111/age.13286

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

Estimates of genomic inbreeding and identification of candidate regions in Beijing-You chicken populations

Qiao Wang | **Jin Zhang** | **Hailong Wang** | **Zixuan Wang** | **Qinghe Li** | **Guiping Zhao** | **Maiqing Zheng** | **Jie Wen**

Institute of Animal Sciences of Chinese Academy of Agricultural Sciences, State Key Laboratory of Animal Nutrition, Beijing, China

Correspondence
Jie Wen and Maiqing Zheng, Institute of Animal Sciences of Chinese Academy of Agricultural Sciences, State Key Laboratory of Animal Nutrition, Beijing 100193, China.
Email: wenjie@caas.cn, zhengmaiqing@caas.cn

Funding information
Agricultural Science and Technology Innovation Program, Grant/Award Number: ASTIP-IAS04 and CAASZDRW202005; Central Public-interest Scientific Institution Basal Research Fund, Grant/Award Number: 2020-YWF-YB-01; Earmarked Fund for Modern Agro-industry Technology Research System, Grant/Award Number: CARS-41

Abstract

Runs of homozygosity (ROHs) has become an effective method for analysing inbreeding in livestock populations. Moreover, ROHs is well-suited to detect signatures of selection via ROH islands. This study aimed to investigate the occurrence and distribution of ROHs, compare the genomic inbreeding coefficients and identify the genomic regions with high ROH frequencies in different Beijing-You chicken (BY) populations, including a random conservation population (BY_R), a pedigree conservation population (BY_P), and a commercial population obtained from the market (BY_C). Among them, BY_R in 2010 and 2019 were BY_R1 and BY_R2 respectively. A total of 27916 ROHs were identified. The average number of ROHs per individual across the three BY populations ranged from 213 (BY_P) to 161 (BY_C), and the average length of ROHs ranged from 0.432 Mb (BY_R2) to 0.451 Mb (BY_P). The highest inbreeding coefficient calculated based on ROHs (F_{ROH}) was 0.1 in BY_P, whereas the lowest F_{ROH} was 0.0743 in BY_C. In addition, the inbreeding coefficient of BY_R2 (F_{ROH} = 0.0798) was higher than that of BY_R1 (F_{ROH} = 0.0579). Furthermore, the highest proportion of long ROH fragments (>4 Mb) was observed in BY_P and BY_C. This study showed the top 10 ROH islands of each population, and these ROH islands harboured 53 genes, some of which were related to limb development, body size and immune response. These findings contribute to the understanding of genetic diversity and population demography, and might help improve breeding and conservation strategies for BY populations.

KEYWORDS
Beijing-you chicken, genomic inbreeding coefficient, ROH islands, runs of homozygosity

INTRODUCTION

The intense selection of livestock has alerted the scientific community to the need for strategies to preserve populations. The main goal in conservation programmes is to maintain the most genetic diversity to maximise the survival potential of the population (Bourn, 2003).

Traditionally, the inbreeding coefficient is calculated by pedigree information, and the genotype is determined by microsatellite DNA markers to confirm the genetic relationships within livestock and poultry. However, the inbreeding coefficient calculated based on the pedigree records (F_{PED}) is a statistical expectation that reflects the probability that two alleles are identical by descent (IBD) (Keller et al., 2011). In addition,

Qiao Wang and Jin Zhang contributed equally to this work.

F_{PED} usually ignores historical inbreeding, resulting in underestimated calculated values (Forutan et al., 2018). At present, the common ancestry measurement method based on shared genomic regions attempts to balance the adaptability and genetic diversity of the population in the breeding process. One of the purposes of this method is to avoid the occurrence of runs of homozygosity (ROHs) in the offspring genome. ROHs compose a continuous homozygous fragment of DNA sequence caused by the transmission of homologous haplotypes from parents to offspring, and ROHs exist widely in all populations (Ceballos et al., 2018). The generation of ROHs is an important feature of genetic diversity reduction at the genome level due to selection or genetic bottleneck effects, and it can also indicate a decline in population inbreeding (Szpiech et al., 2013). Because recombination events interrupt long chromosome segments, long ROHs (~10 Mb) arise as a result of recent inbreeding (up to five generations ago), while shorter ROHs (~1 Mb) can indicate more distant ancestral effects (up to 50 generations ago), such as breed founder effects (Howrigan et al., 2011). Therefore, estimating F using ROHs is particularly appealing, as the number of generations of inbreeding and the history of recent selection events can be inferred from the extent and frequency of ROH regions (Purfield et al., 2012). F_{ROH} was first used to calculate the inbreeding coefficient of the human genome (McQuillan et al., 2008). In animal genetics, ROHs is used to identify inbreeding levels and the genetic relationships between individuals, thus estimating the true level of autozygosity at the individual and population levels (Liu et al., 2021).

ROHs are indicative of recent IBD, and the sum of these segments is the exact inbreeding level of an individual (Forutan et al., 2018). The length of IBD segments follows an inverse exponential distribution with a mean of 1/2 t Morgans, where t represents the number of generations from a common ancestor (Ceballos et al., 2018). Therefore, the inbreeding level, genetic diversity, and population history can be revealed by detecting ROHs in the genome (Purfield et al., 2012). ROH patterns are not randomly distributed across genomes (Peripolli et al., 2017). Genomic regions that are selection targets tend to generate ROH hotspots, which are also called ROH islands (Peripolli et al., 2018). Genomic regions sharing ROH patterns potentially contain alleles associated with genetic improvement in livestock (Kim et al., 2013), and ROH islands can be used to identify genes under natural or artificial selection (Peripolli et al., 2017).

The Beijing-You chicken (BY) is a native chicken breed in China that has excellent meat and egg quality characteristics, strong disease resistance, and a distinctive appearance (Ye et al., 2010). As a famous local chicken, BY has long been included in conservation projects, but the protection of BY is still strongly needed due to the lack of an accurate and effective evaluation of conservation efficiency. In this study, genome resequencing data were used to detect and analyse the genomic inbreeding and genetic diversity of three BY populations. This study aimed to provide not only guidance for the effective protection of BY populations but also a valuable reference for the protection and evaluation of local chicken breeds.

MATERIALS AND METHODS

Ethics statement

The animal experiments were carried out in accordance with the Guidelines for Experimental Animals established by the Ministry of Science and Technology (Beijing, China). The study was approved by the Animal Management Committee of the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (Beijing, China). Ethical approval regarding animal survival was given by the animal ethics committee of IASCAAS (approval number: IASCAAS-AE20140615).

Animals

BY is a famous native chicken breed in China that has been listed in the National Key protection list of Livestock and Poultry Species resources by the Ministry of Agriculture and Rural Affairs. At present, the conservation work of BY is mainly undertaken by the National Conservation Farm of Beijing You Chicken. At present, there are two populations in the farm, one is a random conservation population (BY_R) and another is a pedigree conservation population (BY_P). The BY_R refers to the random mating of about 300 hens by artificial insemination with the mixed semen of all the males (about 100) in the breeding process of generations. We collected blood samples from the BY_R population in 2010 and 2019 for DNA extraction and subsequent analysis, which were represented by BY_R1 and BY_R2 respectively. BY_P means that 35 males are selected from each family line and three females are randomly mated to one male in each generation. A commercial population was obtained from the market, which were represented by BY_C.

Sample collection

Blood samples were collected from 40 chickens (40 male) of BY_R1; 40 chickens (17 male and 23 female) of BY_R2; 40 chickens (34 male and six female) of BY_P; and 40 chickens (20 male and 20 female) of BY_C for DNA extraction.

Genome sequence assembly and data analysis

Blood collected from the veins of all animals was extracted using the standard phenol/chloroform extraction

method. The quantity and quality of DNA extracted from whole blood samples were determined using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc.), and agarose gel electrophoresis was used to visually assess DNA integrity. Samples that passed the quality test were sent to the Beijing Compass Biotechnology Company for 10G whole-genome sequencing. After sequencing, paired-end libraries were generated for each eligible sample using standard procedures. The average insert size was 300–500 bp, and the average read length was PE150 bp (150 bp paired-end reads). All libraries were separately sequenced using the Illumina HiSeq X Ten sequencer.

The filtered raw reads were compared with the reference genome (version: ftp://ftp.ensembl.org/pub/release-101/fasta/gallus_gallus/dna/) using the MEM mode of BWA software (version 0.7.12). Picard tools (version 1.119) and Sam tools (version 1.9) were used to obtain the sorted BAM files. Before detecting mutations, the basic quality score recalibration (BQSR) was performed, which involved two steps. In the first step, Base Recalibrator, the Picard tool and stool were used to obtain the classified BAM file for analysis. In the second step, Apply BQSR, the calibration table file obtained in the first step, was used to readjust the basic quality value in the original BAM file. The new quality value was then re-exported to a new BAM file. Genome Analysis ToolKit (version 4.0.2.1) was used to select Genome Analysis ToolKit variants for filtering (Poplin et al., 2017). The hard-filter standards used for single-nucleotide polymorphisms (SNPs) were QualByDepth < 2.0, RMSMappingQuality < 40, FisherStrand > 60.0, StrandOddsRatio > 3.0, and MappingQualityRankSumTest < −12.5.

The data were first filled with Beagle v5.0 software (Ayres et al., 2012) for all SNP loci, and the filled SNP locus was quality controlled using the Plink v1.9 software (Purcell et al., 2007), according to the following criteria: (1) the call rate was higher than 0.9; (2) the minor allele frequency was higher than 0.05; and (3) SNPs were filtered to exclude loci assigned to unmapped contigs and sex chromosomes. After quality control, 160 chickens and 6 252 214 variants were retained.

Genetic characterisation

Different approaches and software were used to disclose the genetic structure of BY populations:

(i) Wright's statistics, including observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficients. Genomic inbreeding based on homozygous SNPs was determined using Plink v1.9 software. The inbreeding coefficient for an individual (F_{HOM}) was computed as follows:

$$F_{HOM} = (O - E)/(L - E),$$

where O is the number of observed homozygotes, E is the number of homozygotes expected by chance, and L is the number of genotyped autosomal SNPs. Genomic SNP-by-SNP inbreeding coefficient (F_{GRM}) estimates were calculated using the genome-wide complex trait analysis (GCTA) software. The F_{GRM} was calculated as follows:

$$F_{GRM} = \frac{1}{m}\sum_{i=1}^{N}\left(\frac{\left[x_i - E(x_i)\right]^2}{2p_i(1-p_i)} - 1\right),$$

where x_i is the number of copies of the reference allele for the ith SNP, m is the number of SNPs, and p_i is the frequency of the reference allele. Genomic inbreeding coefficients were also estimated based on ROHs (F_{ROH}). The F_{ROH} for each animal was calculated as follows:

$$F_{ROH} = \frac{\sum_i L_{ROHi}}{L_{auto}},$$

where L_{ROHi} is the length of ROH_i of individual i, and L_{auto} is the autosomal genome length covered by the SNPs included on the chip. Genomic inbreeding coefficients were estimated based on uniting gametes (F_{UNI}). The F_{UNI} for each animal was calculated as follows:

$$F_{UNI} = \frac{x_i^2 - (1 + 2p_i)x_i + 2p_i^2}{2p_i(1-p_i)},$$

where x_i is the number of copies of the reference allele for the ith SNP, and p_i is the frequency of the reference allele.

(ii) Linkage disequilibrium (LD) within 500 kb was calculated using PopLDdecay software (Zhang et al., 2019), and then the genome-wide LD distribution was mapped based on the marker spacing and r^2 values.

(iii) Principal component analyses (PCAs) were conducted using the GCTA 64 software. The PCA of pairwise individual genetic distances was performed based on the allele frequencies of pruned SNPs.

(iv) ADMIXTURE v.1.3.0 software was used to infer the most probable number of ancestral populations based on the SNP data (Alexander et al., 2009). ADMIXTURE was run from $K = 1$ to $K = 6$, and the optimal number of clusters (K value) was determined as the one having the lowest cross-validation error. Each inferred chicken population structure was visualised using an R script, as suggested in the ADMIXTURE procedure.

(v) ROH analysis was performed for each individual (complete SNP dataset = 6 252 214) using the Plink software. The ROH was defined as follows: (1) a minimum of 200 kb in size and 100 homozygous SNPs; (2) one heterozygous SNP was permitted in the ROH, so that the length of the ROH was not disrupted by an occasional heterozygote; (3) one missing SNP was allowed in the ROH; and (4) the maximum gap between SNPs of 40 kb was predefined to ensure that the SNP density did not affect the ROH. According to the nomenclature reported by other authors (Zavarez et al., 2015), the ROHs were grouped into five classes of length: <0.5 Mb, 0.5–0.1 Mb, 1–2 Mb, 2–4 Mb, and >4 Mb. The number, total length, and average ROH length were calculated across individuals within the chicken populations. In addition, the percentage of the total genome length affected by the ROH was also estimated.

Estimation of effective population size

Historical effective population sizes (*Ne*) of the three chicken populations were computed:

$$N_{T(t)} = \frac{1}{(4f(c_t))}\left(\frac{1}{E\left[r^2_{\text{adj}} \mid c_t\right]} - \alpha\right),$$

where N_T is the effective population size t generations ago, and the expectation is conditional on the markers being the appropriate distance apart given t and the mapping function $f(c)$. A key assumption stated by Hayes et al. (2003) is constant linear growth of *Ne* with t.

Detection of ROH islands

To identify the genomic regions most commonly associated with ROHs, the percentage of occurrence of SNPs in ROH was calculated by counting the number of times that an SNP was detected in an ROH across all the individuals in each population. In this study, the top 1% of the SNPs observed in ROHs was selected as the threshold for identifying the genomic regions most commonly associated with ROH in each population. A series of adjacent SNPs that exceeded this threshold formed a genomic region, which we refer to as an ROH island. In this study, the breed specific thresholds were 27.5%, 37.5%, 42.5%, and 35% of the individuals sharing the overlapping homozygous regions (ROH islands) in BY_R1, BY_R1, BY_P, and BY_C respectively. Gene annotation in ROH islands was performed on the basis of the *Gallus gallus* (chicken) reference genome (RefSeq GCF_000002315.6).

TABLE 1 Genetic diversity measures in BY_R1, BY_R2, BY_P, and BY_C

Population	*Pi*	H_O	H_E
BY_R1	0.002147036	0.2969368	0.3153638
BY_R2	0.002144846	0.3026812	0.3148766
BY_P	0.002102034	0.2664642	0.3083294
BY_C	0.002079772	0.2734101	0.3052186

Note: P_i, H_O, and H_E represent the nucleotide diversity, observed heterozygosity, and expected heterozygosity. BY_R1, BY_R2, BY_P, and BY_C represent a random conservation BY population in 2010, a random conservation BY population in 2019, a pedigree conservation BY population in 2019, and a commercial BY population respectively.

RESULTS

Genetic diversity analysis

The nucleotide diversity (P_i), H_O, and H_E were calculated to reveal the nucleotide polymorphisms of the three populations (Table 1). The P_i of BY_R1 and BY_R2 were very close, 0.002147 and 0.002145 respectively, followed by BY_P (0.002102) and BY_C (0.00208; Table 1). The H_O of BY_R2 was 0.302681, which was the highest among the four populations, followed by BY_R1 (0.296937) and BY_C (0.27341), and BY_P was the lowest, which was 0.266464 (Table 1). H_E was consistent with P_i that BY_R1 was the highest and BY_C was the lowest, which were 0.315364 and 0.305219 respectively (Table 1).

Analysis of genetic characterisation and estimation of effective population size (N_e)

PCA and population structure analysis were performed using GCTA 64 and ADMIXTURE respectively. In the PCA, BY_C was completely separated from the other three populations, although the three groups also had some separation (Figure 1a). LD decay can reflect the genetic diversity of the population. Here, the LD decay map measured by r^2 over a distance between SNPs in four BY populations. The results showed that LD decay lines of BY_R1, BY_R2 and BY_P almost overlapped, and the smallest values of average r^2 were apparent in BY_C, indicating the highest genetic diversity of BY_C. (Figure 1b). Estimates of effective population size (N_e) for the three BY populations are depicted in Figure 1c, a declining trend of N_e across generations was observed. For all generations, BY_C had larger N_e, followed by BY_R2, and the N_e of BY_P was the smallest (Figure 1c). The N_e of the three BY populations in 2019 at 98 generations ago were predicted to be 704 (BY_C), 595 (BY_R2), and 558 (BY_P). The N_e 13 generations ago had reduced to 194 (BY_C), 176 (BY_R2), and 171 (BY_P).

The ADMIXTURE program was run for K values from 2 to 6. The lowest cross-validation error was found at $K = 2$ and represented the number of ancestors in BY

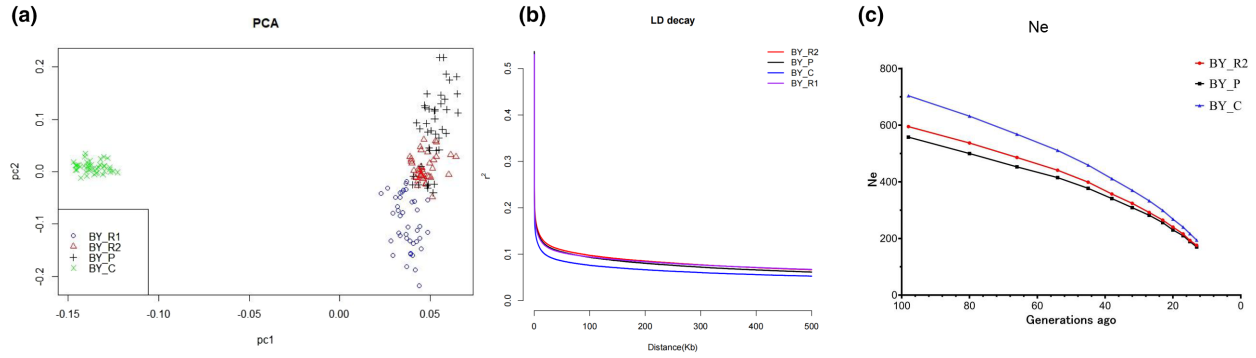

FIGURE 1 Principal component analysis (PCA), linkage disequilibrium (LD) decay analysis, and estimates of effective population sizes (N_e) in BY_R, BY_P, and BY_C. (a) PCA analyses of BY_R, BY_P, and BY_C. (b) LD decay map measured BY r^2 over distance between SNPs in BY_R, BY_P, and BY_C. (c) Estimates of N_e of BY_R, BY_P, and BY_C for 100–10 ancestral generations ago

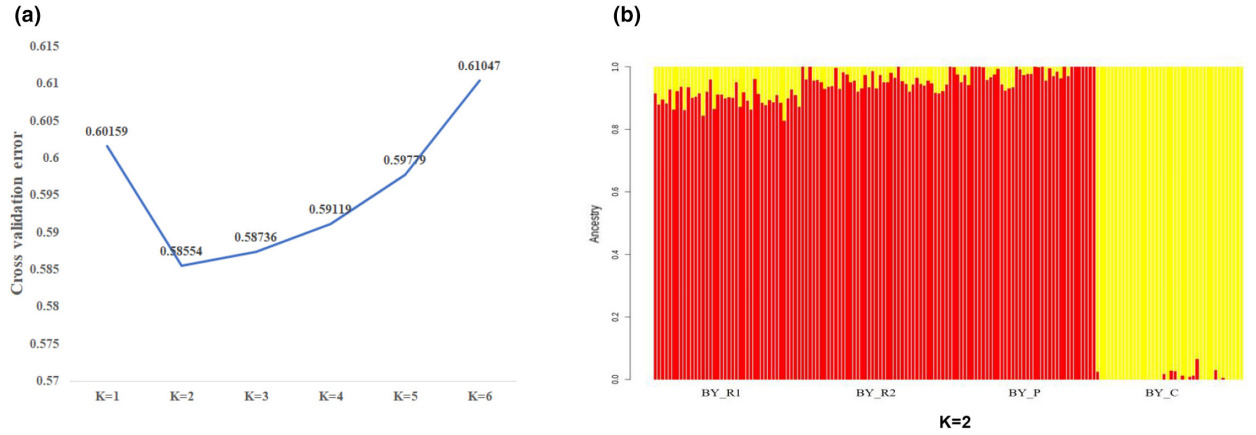

FIGURE 2 Population genetic analysis in BY_R, BY_P, and BY_C. (a) Cross-validation results. (b) Population differentiation analysis in BY_R, BY_P, and BY_C BY ADMIXTURE

populations (Figure 2a). When $K = 2$, an obvious separation was observed between the conservation and commercial BY populations (Figure 2b).

ROH detection

A total of 27916 ROHs were identified across the four BY populations. Figure 3a,b show the number of ROHs per individual based on the mean length and total length of ROHs respectively. Figure 3c,d illustrate the mean and the total length of ROHs of the four BY populations respectively. BY_P was the highest in both the mean and the total length of ROHs in the four populations, followed BY_C, and BY_R1 was the lowest. The average number of ROHs per individual ranged from 213.23 (BY_P) to 145.55 (BY_R1; Table 2). The average length of ROH of each population ranged from 0.451 Mb (BY_P) to 0.382 Mb (BY_R1; Table 2). The results of the percentages of ROHs in the five length categories of 0–0.5, 0.5–1, 1–2, 2–4, and >4 Mb in each population showed that most ROHs were shorter than 4 Mb (Figure 4a). Compared with BY_R1 and BY_R2, BY_C and BY_P had a higher proportion of long ROHs (>4 Mb; Figure 4a). The frequency distribution and average length of ROHs on each chromosome in the four populations were also analysed (Figure 4b,c). In the four populations, the percentages on each chromosome tended to be similar, but the average length of ROHs of some individual chromosomes showed large differences (Figure 4b,c). On chromosome 16, the average length of ROHs in BY_C, BY_R1, and BY_R2 was close to 0.2 Mb, but ROHs were not identified on this chromosome in BY_P (Figure 4c). In addition, the highest percentage was observed on chromosome 1 in all four populations, while the lowest percentage of ROH number was on chromosome 30 in BY_R1 and chromosome 31 in BY_C, BY_P, and BY_R2 (Figure 4b).

Inbreeding coefficient estimates based on genomic data and pedigree

In this study, the genomic inbreeding coefficients (F_{HOM}, F_{GRM}, F_{UNI}, and F_{ROH}) of the four BY populations and pedigree-based inbreeding coefficient (F_{PED}) of the pedigree conservation population were calculated (Table 3 and Figure 5a). The F_{PED} of BY_P was 0.0389. The F_{ROH}

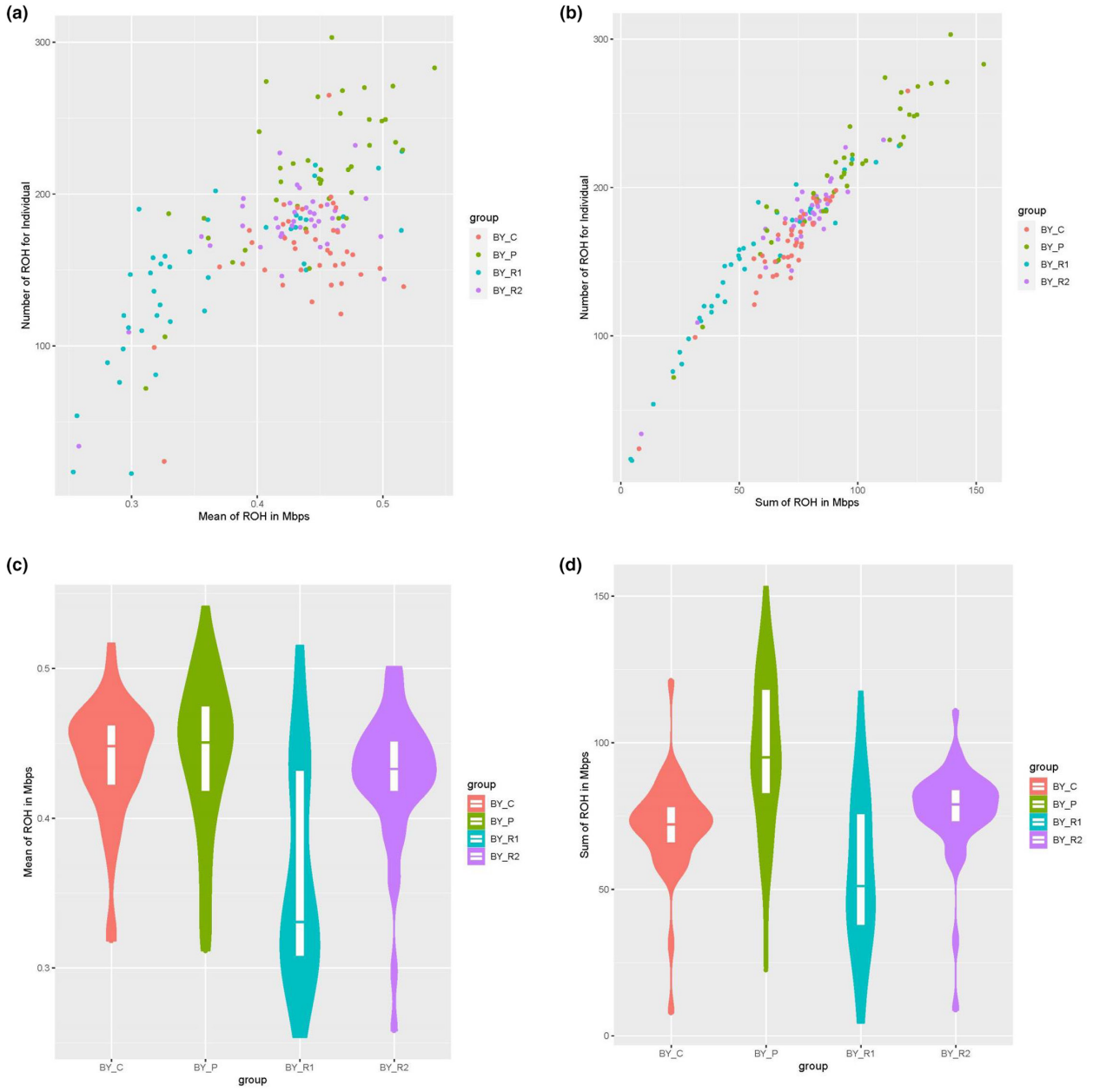

FIGURE 3 Number of runs of homozygosity (ROHs) and the length of genome (mb) covered by ROH segments per individual or population respectively. (a) Number of ROHs per individual based on the mean length of ROH segments. (b) Number of ROHs per individual based on the sum length of ROH segments. (c) Mean length of ROHs of BY_R1, BY_R2, BY_C, and BY_P. (d) Sum length of ROHs of BY_R1, BY_R2, BY_C, and BY_P

TABLE 2 Mean number and mean length of ROHs of BY_R1, BY_R2, BY_C, and BY_P

Population	ROH number mean±SD	ROH length, Mb mean±SD
BY_R1	145.55 ± 50.85	0.382 ± 0.234
BY_R2	177.53 ± 31.05	0.432 ± 0.281
BY_P	213.28 ± 46.66	0.451 ± 0.325
BY_C	161.55 ± 34.96	0.442 ± 0.302

of BY_R1 (0.0579) was the lowest, followed by BY_C (0.0743) and BY_R2 (0.0798), and the highest was BY_P (0.1). The inbreeding coefficient of each chromosome was calculated according to the length of ROH on different chromosomes. The inbreeding degree of chromosome 32 was found to be the highest, and that of chromosomes 16 and 31 was the lowest among the four BY populations (Figure 5b).

The pairwise correlations among five kinds of inbreeding coefficients of BY_P are shown in Figure 6.

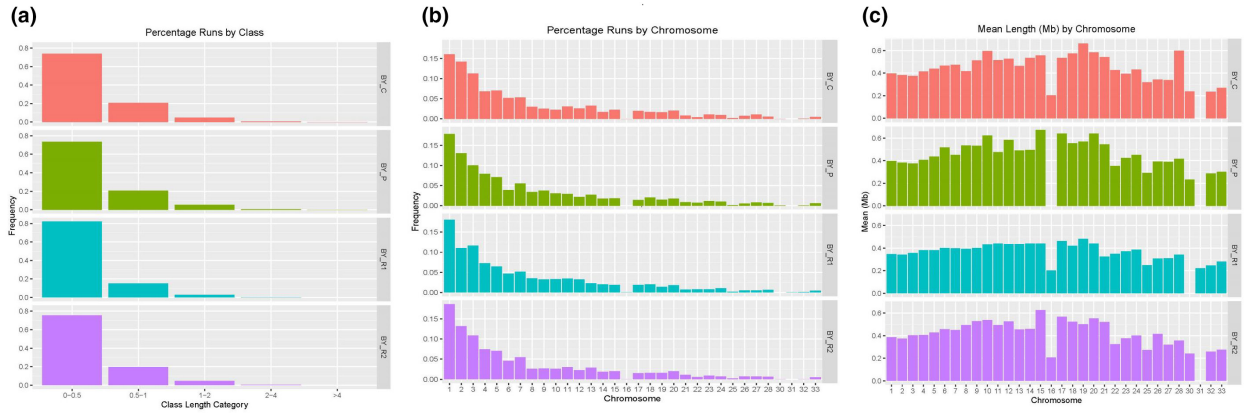

FIGURE 4 Percentages of runs of homozygosity (ROHs) in different length categories and the distribution of ROHs per chromosome in BY_R1, BY_R2, BY_C, and BY_P. (a) Percentages of ROH in different length categories in BY_R1, BY_R2, BY_C, and BY_P. (b) Percentages of ROH per chromosome in BY_R1, BY_R2, BY_C, and BY_P. (c) Mean length (mb) of ROHs per chromosome in BY_R1, BY_R2, BY_C, and BY_P

TABLE 3 Inbreeding coefficients of BY_R1, BY_R2, BY_C, and BY_P

Population	F_{HOM} mean±SD	F_{GRM} mean±SD	F_{UNI} mean±SD	F_{ROH} mean±SD	F_{PED}
BY_R1	0.0784 ±0.0459	0.0859 ±0.0408	0.0806 ±0.0408	0.0579 ±0.0284	–
BY_R2	0.0605 ±0.0398	0.0662 ±0.0347	0.0637 ±0.0354	0.0798 ±0.0171	–
BY_P	0.1729 ±0.0587	0.1368 ±0.0433	0.1587 ±0.0479	0.1000 ±0.0292	0.0389 ±0.0333
BY_C	0.1514 ±0.0282	0.2690 ±0.0315	0.1924 ±0.0274	0.0743 ±0.0177	–

Note: F_{PED} represents the inbreeding coefficient calculated by pedigree records. F_{HOM} represents the inbreeding coefficient for an individual. F_{GRM} represents genomic SNP-by-SNP inbreeding coefficient. F_{UNI} and F_{ROH} represent the genomic inbreeding estimates based on the correlation between uniting gametes and ROHs.

Among all pairwise correlations, the highest correlation was 0.97 between F_{UNI} and F_{HOM}, while the lowest correlation was 0.26 between F_{HOM} and F_{PED}. The correlations of F_{PED} with F_{ROH} and F_{GRM} was 0.57 and 0.58 respectively, which were significantly higher than those with F_{HOM} (0.26) and F_{UNI} (0.39; Figure 6).

Detection of ROH islands and functional annotation of genes

The incidence of ROH fragments across the genome was presented as a Manhattan plot to figure out the difference percentage of occurrence of SNPs in ROH against the position of the SNPs along all the autosomes among the three BY populations (Figure 7). The associated genes involved in the top 10 ROH islands in each population were annotated based on the autosomal chicken gene list (Gallus_gallus.GRCg6a.101.gff3; Table S1). The islands located on Chr8: 12.01–12.32 Mb, which harboured cell division cycle 14A (*CDC14A*), diphthamide biosynthesis 5 (*DPH5*), exostosin-like glycosyltransferase 2 (*EXTL2*), and solute carrier family 30 member 7 (*SLC30A7*), were common to BY_R1, BY_P, and BY_C. Several population-specific ROH islands were found. In BY_R2, the specific ROH islands were located on Chr1: 85.25–85.46 Mb and harboured G protein-coupled receptor 161 (*GPR161*), ST3 β-galactoside α-2,3-sialyltransferase 6 (*ST3GAL6*), and collagen type VIII α 1 chain (*COL8A1*; Table S1). The BY_C population-specific ROH islands were mainly located on Chr2: 8.67–8.76 Mb and Chr27: 7.81–7.92 Mb, which harboured 17 genes (Table S1). The other islands and the genes they harboured are listed in Table S1.

DISCUSSION

BY is a famous local chicken breed in China. It has been included in the national breed conservation system because of its unique and varied characteristics and sharp decline in quantity. Based on genomic information, the population genetic diversity and inbreeding coefficient of two conservation BY populations and one commercial BY population were analysed in this study. At the same time, the ROH information was used to identify genes related to germplasm characteristics and reveal the molecular mechanism of trait formation.

The genomic inbreeding coefficients of three BY conservation populations (BY_R, BY_P, and BY_C) were calculated, and the correlation between genomic inbreeding coefficients (F_{HOM}, F_{ROH}, F_{GRM}, and F_{UNI}) and F_{PED} was analysed in BY_P. The results showed

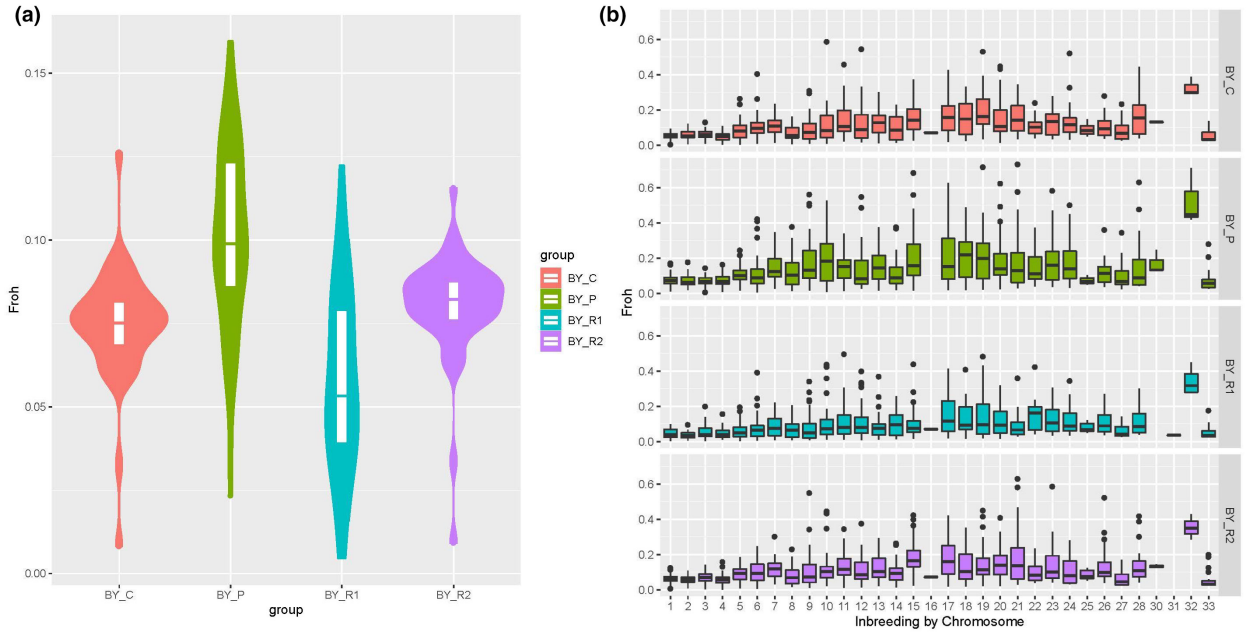

FIGURE 5 F_{ROH} of each population and the distribution of F_{ROH} on each chromosome in BY_R1, BY_R2, BY_C, and BY_P. (a) F_{ROH} of BY_R1, BY_R2, BY_C, and BY_P. (b) Distribution of F_{ROH} on each chromosome in BY_R1, BY_R2, BY_C, and BY_P

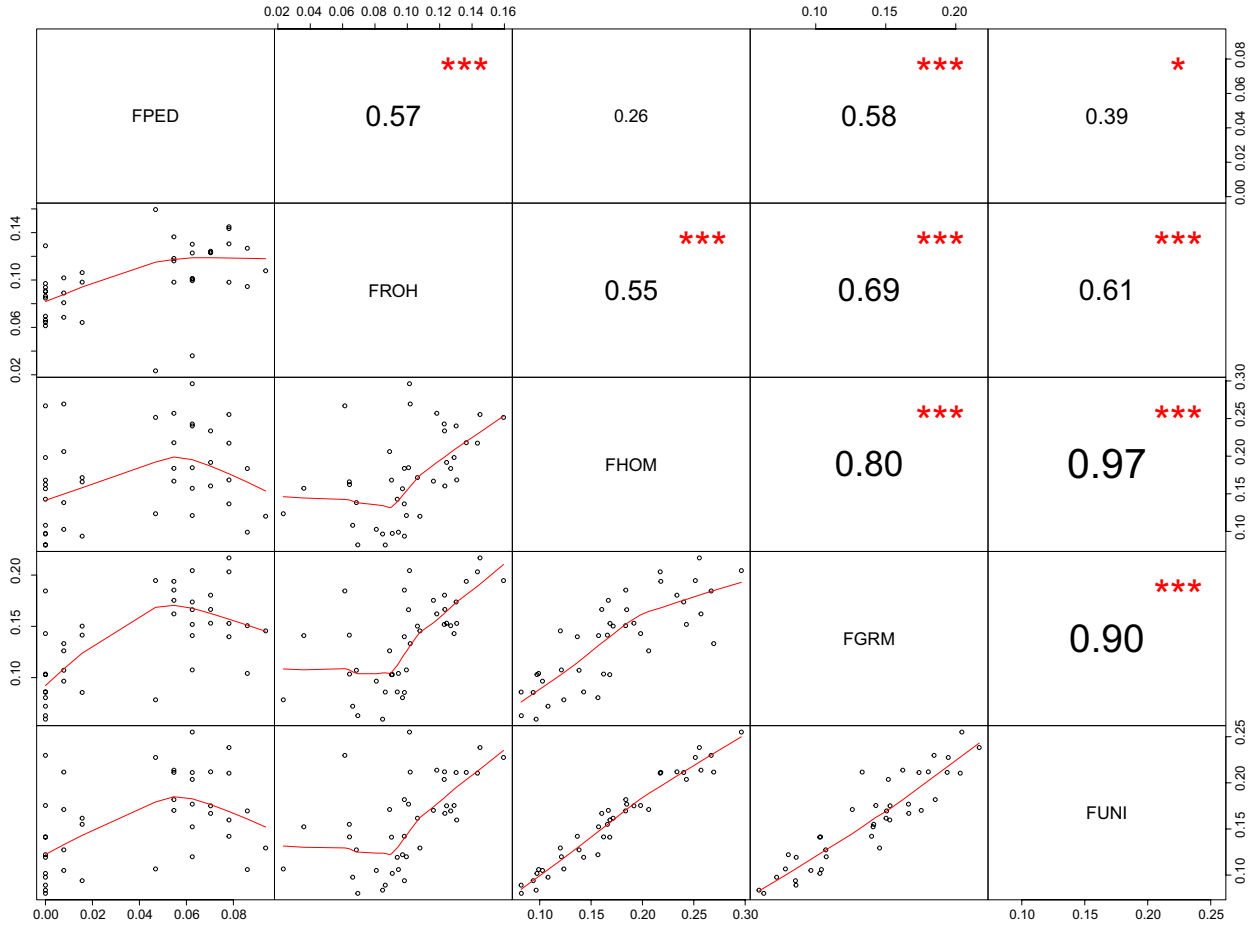

FIGURE 6 Pairwise correlations among five kinds of inbreeding coefficients of BY_P. F_{PED} represents the inbreeding coefficient calculated BY pedigree records. F_{HOM} represents the inbreeding coefficient for an individual. F_{GRM} represents genomic SNP-by-SNP inbreeding coefficient. F_{UNI} and F_{ROH} represent the genomic inbreeding estimates based on the correlation between uniting gametes and ROHs. *$p<0.05$, **$p<0.01$, ***$p<0.001$

that F_{PED} was lower than the genomic inbreeding coefficient, which was consistent with the relevant research results for pigs and cattle (Doekes et al., 2020; Szmatoła et al., 2020) The correlations of F_{PED} with F_{ROH} (0.57) and F_{GRM} (0.58) were higher than those of F_{PED} with F_{HOM} (0.26) and F_{UNI} (0.39; Figure 6), indicating that F_{ROH} and F_{GRM} were relatively accurate for evaluating population inbreeding, at least in chickens. It has been reported that F_{ROH} provides the closest estimates to true inbreeding (Forutan et al., 2018; Keller et al., 2011). Moreover, the correlations between F_{PED} and F_{ROH} were moderate to high (0.62–0.75) (Ferenčaković et al., 2013; Marras et al., 2015), so F_{ROH} was considered an effective approach for evaluating the inbreeding degree of an individual.

The conservation efficiency was affected by the conservation method and the size of the conservation populations. For the BY conservation populations, BY_P maintained 35 families every year, and the ratio of males to females was 1:3. For the BY_R, the number of breeding roosters and hens was approximately 100 and 400 respectively. Therefore, the population size of BY_R was basically three times that of BY_P. The ADMIXTURE program analysis did not find an obvious separation between BY_R and BY_P (Figure 2b), indicating that the two conservation populations maintained basically the same genetic information. However, the *Ne* and F_{ROH} of the BY_R2 and BY_P were quite different (Figures 1c and 5a). The results clearly showed that BY_R2 maintained a larger *Ne* and a lower inbreeding coefficient (Figures 1c and 5a). These results were inconsistent with the conclusion that the pedigree systems were more effective in monitoring inbreeding (Enrique et al., 2016). This might be due to two main reasons. One reason may be that the number of chickens used for BY_P was less than that for BY_R, while the other may be that the ratio of male to female in each family was only 1:3. This suggested that not only the number of families in BY_P should be increased and that the ratio of males to females should be increased to provide more effective protection for BY. Nevertheless, although the *Ne* of BY_R2 was larger than that of BY_P, the reduction rate of the *Ne* of BY_P was slower than that of BY_R2, which reflects the advantage of pedigree conservation in maintaining the content of *Ne*.

The situation of BY populations in the past 10 years was analysed using genomic information to monitor the conservation effect of the random conservation BY population. The results of ROH detection showed that both the

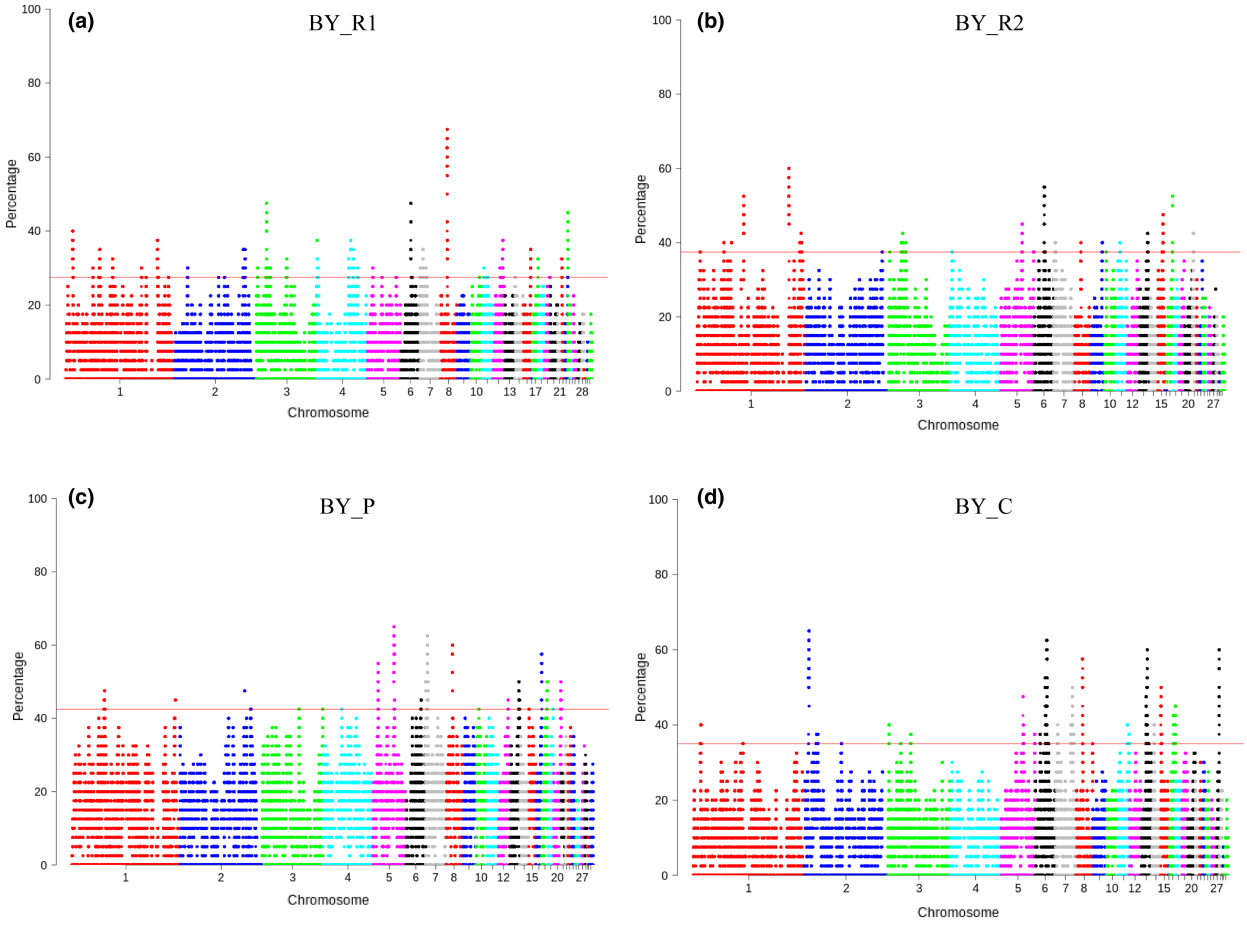

FIGURE 7 Genome-wide frequency of SNP occurrence into runs of homozygosity (ROHs) for each BY population. The red line was the threshold to define the ROH islands

average length and total length of ROHs in BY_R1 were lower than those in BY_R2 (Figure 3c,d), indicating that the inbreeding degree of BY populations was higher than that 10 years ago. The F_{ROH} values of BY_R1 and BY_R2 were calculated, and the F_{ROH} of BY_R1 (0.0579) was found to be lower than that of BY_R2 (0.0798), indicating that the inbreeding degree of the conservation BY population increased during the last decade (Table 3). In addition, BY_R1 and BY_R2 had very similar percentages of ROHs in the five length categories 0–0.5, 0.5–1, 1–2, 2–4, and >4 Mb (Figure 4a). However, differences were found in the average length or distribution frequency of ROHs on chromosomes (Figure 4b,c). These results showed that in addition to the increase in the inbreeding degree, the genetic information of BY populations also changed in the last decade. Long-term feeding in different environments and artificial selection may change genetic information and population genetic structure (Wang, 2014). In this study, the conservation populations (BY_R and BY_P) were raised in the National BY Conservation Farm for more than 30 years. Population genetic analysis showed that BY_R and BY_P were significantly separated from BY_C, indicating potential genetic differentiation (Figure 2). Compared with BY_R, the number of ROHs with the lengths exceeding 4 Mb in BY_C was relatively large, but the difference between BY_C and BY_P was small (Figure 4a). In addition, the average length of ROHs on different chromosomes in BY_C was also different from that in the two conservation populations. The difference between BY_C and BY_P was mainly reflected on chromosome 15 and chromosome 16, and the difference between BY_C and BY_R1 was reflected on chromosome 30 and chromosome 31 (Figure 4c). In addition, the average length of ROHs on different chromosomes of BY_R1 is around 0.4 Mb, while the average size of ROHs on some chromosomes of BY_C, BY_P and BY_R2 is longer, and some are more than 0.6 Mb. These results indicate that BY_R1 has not undergone strong selection compared with other BY populations (Figure 4c).

ROH islands were generated from natural or artificial selection and could be used to identify selection signatures. In this study, ROH islands in different conservation populations and commercial populations were identified to explain the molecular mechanism of trait changes caused by inbreeding or artificial selection in the process of breeding conservation. The percentage of occurrence of SNPs in ROH against the positions of the SNPs was compared along all the autosomes among the four populations (Figure 7). The islands located on Chr8: 12.01–12.32 Mb, which harboured *CDC14A*, *EXTL2*, *DPH5* and *SLC30A7*, were common to BY_R1, BY_P, and BY_C. BY_R1 and BY_R2 shared a common ROH island located on Chr6: 19.19–19.30 Mb, but no genes were found. Several population-specific ROH islands exist. In BY_R2, the specific ROH islands were located on Chr1: 85.25–85.46 Mb, which harboured *GPR161*, *ST3GAL6*, and *COL8A1*. In BY_P, multiple ROH islands were identified in Chr5: 38.43–38.90 Mb, involving *JDP2* and *TGFB3*. BY_C population-specific ROH islands were mainly located on Chr2: 8.67–8.76 Mb and Chr27: 7.81–7.92 Mb, which harboured 17 genes, several of which were related to pathogenic infection and immunity, such as *DNAJB6* (Cao et al., 2019), *CCR10* (Homey et al., 2002), *BECN1* (Cui et al., 2016), and *IFI35* (Xiahou et al., 2017).

In this study, genome resequencing data were used to characterise the pattern of LD, estimate the effective population sizes, and investigate the occurrence and distribution of ROHs in three BY populations. These findings contribute to the understanding of genetic diversity, population demography, and the underlying genetic mechanism of traits and may help improve breeding and conservation strategies for BY populations.

FUNDING INFORMATION

This study was supported by grants from the Central Public-interest Scientific Institution Basal Research Fund (No. 2020-YWF-YB-01), the Agricultural Science and Technology Innovation Program (ASTIP-IAS04 and CAASZDRW202005), and the Earmarked Fund for Modern Agro-industry Technology Research System (CARS-41).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The clean DNA sequencing data reported in this study were deposited in the Genome Sequence Archive in the BIG Data Center under accession number CRA004519 and can be publicly accessed at http://bigd. big.ac.cn/gsa.

ORCID

Qiao Wang https://orcid.org/0000-0002-7144-6064
Qinghe Li https://orcid.org/0000-0002-9698-829X
Guiping Zhao https://orcid.org/0000-0002-1510-6903
Maiqing Zheng https://orcid.org/0000-0002-4205-3595
Jie Wen https://orcid.org/0000-0001-7460-4219

REFERENCES

Alexander, D.H., Novembre, J. & Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.

Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O. et al. (2012) BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Systematic Biology*, 61, 170–173.

Bourn, N. (2004) Book review: Introduction to conservation genetics. Frankham, R., Ballou, J.D. and Briscoe, D.A. *Journal of Insect Conservation*, 8(4). Available from: https://doi.org/10.1007/s10841-004-0978-3

Cao, Y.Q., Yuan, L., Zhao, Q., Yuan, J.L. & Cao, S.J. (2019) Hsp40 protein dnajb6 interacts with viral ns3 and inhibits the replication of the Japanese encephalitis virus. *International Journal of Molecular Sciences*, 20(22), 5719.

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M. & Wilson, J.F. (2018) Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews. Genetics*, 19, 220–234.

Cui, J., Jin, S. & Wang, R.-F. (2016) The BECN1-USP19 axis plays a role in the crosstalk between autophagy and antiviral immune responses. *Autophagy*, 12, 1210–1211.

Doekes, H.P., Bijma, P., Veerkamp, R.F., de Jong, G., Wientjes, Y.C.J. & Windig, J.J. (2020) Inbreeding depression across the genome of Dutch Holstein Friesian dairy cattle. *Genetics, Selection, Evolution*, 52, 64.

Enrique, S.M., Ricardo, P.W. & Georgios, B. (2016) Genomic-based optimum contribution in conservation and genetic improvement programs with antagonistic fitness and productivity traits. *Frontiers in Genetics*, 7, 25.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T.R., Klemetsdal, G., Curik, I. et al. (2013) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics*, 130(4), 286–293.

Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F.S. & Sargolzaei, M. (2018) Inbreeding and runs of homozygosity before and after genomic selection in north American Holstein cattle. *BMC Genomics*, 19, 98.

Hayes, B.J., Visscher, P.M., McPartlan, H.C. & Goddard, M.E. (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, 13, 635–643.

Homey, B., Alenius, H., Müller, A., Soto, H., Bowman, E.P., Yuan, W. et al. (2002) CCL27–CCR10 interactions regulate T cell–mediated skin inflammation. *Nature Medicine*, 8, 157–165.

Howrigan, D.P., Simonson, M.A. & Keller, M.C. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *Bmc Genomics*, 12(1), 460.

Keller, M.C., Visscher, P.M. & Goddard, M.E. (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, 189, 237–249.

Kim, E.-S., Cole, J.B., Huson, H., Wiggans, G.R., Van Tassell, C.P., Crooker, B.A. et al. (2013) Effect of artificial selection on runs of homozygosity in u.s. Holstein cattle. *PLoS One*, 8, e80813.

Liu, J., Shi, L., Li, Y., Chen, L., Garrick, D., Wang, L. et al. (2021) Estimates of genomic inbreeding and identification of candidate regions that differ between Chinese indigenous sheep breeds. *Journal of Animal Science and Biotechnology*, 12, 95.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A. et al. (2015) Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal Genetics*, 46(2), 110–121.

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L. et al. (2008) Runs of homozygosity in European populations. *American Journal of Human Genetics*, 83, 359–372.

Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R. & Baldi, F. (2017) Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics*, 48, 255–271.

Peripolli, E., Metzger, J., de Lemos, M.V.A., Stafuzza, N.B., Kluska, S., Olivieri, B.F. et al. (2018) Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits. *BMC Genomics*, 19, 680.

Poplin, R., Ruano-Rubio, V., Depristo, M.A., Fennell, T.J. & Banks, E. (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559–575.

Purfield, D.C., Berry, D.P., McParland, S. & Bradley, D.G. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics*, 13, 70.

Szmatoła, T., Jasielczuk, I., Semik-Gurgul, E., Szyndler-Nędza, M., Blicharski, T., Szulc, K. et al. (2020) Detection of runs of homozygosity in conserved and commercial pig breeds in Poland. *Journal of Animal Breeding and Genetics*, 137, 571–580.

Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A. et al. (2013) Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics*, 93, 90–102.

Wang, J. (2014) Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of Evolutionary Biology*, 27, 518–530.

Xiahou, Z., Wang, X., Shen, J., Zhu, X., Xu, F., Hu, R. et al. (2017) NMI and IFP35 serve as proinflammatory DAMPs during cellular infection and injury. *Nature Communications*, 8, 950.

Ye, M.H., Chen, J.L., Zhao, G.P., Zheng, M.Q. & Wen, J. (2010) Associations of a-fabp and h-fabp markers with the content of intramuscular fat in beijing-you chicken. *Animal Biotechnology*, 21(1), 14–24.

Zavarez, L.B., Utsunomiya, Y.T., Carmo, A.S., Neves, H.H.R., Carvalheiro, R., Ferenčaković, M. et al. (2015) Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. *Frontiers in Genetics*, 6, 5.

Zhang, C., Dong, S.S., Xu, J.Y., He, W.M. & Yang, T.L. (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, 35, 1786–1788.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang, Q., Zhang, J., Wang, H., Wang, Z., Li, Q. & Zhao, G. et al. (2023) Estimates of genomic inbreeding and identification of candidate regions in Beijing-You chicken populations. *Animal Genetics*, 54, 155–165. Available from: https://doi.org/10.1111/age.13286