Comprehensive analysis of structural variants in chickens using PacBio sequencing

Jinxin Zhang[1], Changsheng Nie[1], Xinghua Li[1], Xiurong Zhao[1], Yaxiong Jia[2], Jianlin Han[2], Yu Chen[3], Liang Wang[3], Xueze Lv[3], Weifang Yang[3], Kaiyang Li[3], Jianwei Zhang[3], Zhonghua Ning[1], Haigang Bao[1], Chunjiang Zhao[1], Junying Li[1] and Lujiang Qu[1]*

[1]Department of Animal Genetics and Breeding, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China, [2]Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [3]Beijing Municipal General Station of Animal Science, Beijing, China

Structural variants (SVs) are one of the main sources of genetic variants and have a greater impact on phenotype evolution, disease susceptibility, and environmental adaptations than single nucleotide polymorphisms (SNPs). However, SVs remain challenging to accurately type, with several detection methods showing different limitations. Here, we explored SVs from 10 different chickens using PacBio technology and detected 49,501 high-confidence SVs. The results showed that the PacBio long-read detected more SVs than Illumina short-read technology genomes owing to some SV sites on chromosomes, which are related to chicken growth and development. During chicken domestication, some SVs beneficial to the breed or without any effect on the genomic function of the breed were retained, whereas deleterious SVs were generally eliminated. This study could facilitate the analysis of the genetic characteristics of different chickens and provide a better understanding of their phenotypic characteristics at the SV level, based on the long-read sequencing method. This study enriches our knowledge of SVs in chickens and improves our understanding of chicken genomic diversity.

KEYWORDS

chicken, structural variants, whole-genome sequencing, pacbio data, genetic diversity

Introduction

Structural variants (SVs) are rearrangements larger than 50 bp in chromosomes (Alkan et al., 2011). Several SVs have been associated with phenotypic variation (Wright et al., 2009; Imsland et al., 2012; Dorshorst et al., 2015; Guo et al., 2016; Chen et al., 2017; Bertolotti et al., 2020; Wu et al., 2021), productive traits (Liu et al., 2010; Zhao et al., 2016; Liu et al., 2019), immune function, and disease resistance (Luo et al., 2013; Bickhart and Liu, 2014) in animals.

Chicken is an important protein source for humans. According to the Food and Agricultural Organization (https://www.fao.org/home/en), between 2000 and 2020, the

proportion of poultry meat traded on the international market has doubled. The poultry meat Index reached 130,39 points in June 2022, an increase of 28.9% more than last year. Therefore, genetic research on chicken is highly valuable. Previous studies have shown that structural variations in domestic chicken play a vital role. For example, the pea comb is caused by duplication of the first intron of the *SOX5* gene (Wright et al., 2009), and late feathering is caused by duplication of the K locus (Elferink et al., 2008). Although there is much evidence that SVs are important, they have been largely understudied in comparison to single-nucleotide variants (SNPs) because they can be difficult to detect. An SV likely remains unknown unless the sequence reads cover its entire length (Sedlazeck et al., 2018).

Substantial progress has been made in SV detection over the last decade using short-read sequencing data; however, the structural variation of some complex traits and repetitive regions remains unsolved (English et al., 2015; Chaisson et al., 2019). Nevertheless, long-read sequencing technology has introduced new possibilities for identifying complex variants. Pacific Biosystems (PacBio) is one of the leaders in this field, and its sequencers can generate reads over 10 kb, which may span the entire variation region (Merker et al., 2018; van Dijk et al., 2018). Studies have shown that long-read sequencing improves the detection accuracy and sensitivity of SVs (Kosugi et al., 2019; Mahmoud et al., 2019; Liu et al., 2020b).

Here, we took advantage of whole-genome re-sequencing data, including long- and short-read data, to investigate SVs in 10 chickens with different genetic backgrounds and significant phenotypic differences. We attempted to detect SVs related to chicken phenotypic traits and mined more phenotype-related candidate genes. These results may provide a reference for future research on structural variations in chickens.

Materials and methods

Sample collection and DNA extraction

A total of 10 chickens were collected from different breeds for genomic sequencing. This group consisted of four commercial chickens (White Leghorn, WL; Rhode Island Red, RIR; Cornish, COR; White Plymouth Rock, WR), five indigenous chickens (Silkies, SK; Beijing You, BY; Tibetan, TB; Piao, P; Dong Tao, DT), and one wild chicken (red jungle fowl, RJF). All samples were obtained from experimental farms, including WL, WR, TB, RIR, COR, SK, and BY from poultry genetic resources and breeding experimental bases of China Agricultural University, Piao and DT chicken from Yunnan and Vietnam, and RJF from Indonesia. (Supplementary Table S1; Supplementary Figure S1).

Blood samples were collected from the wing vein and stored at −20°C for DNA extraction. For PacBio continuous long read (CLR) sequencing and short-read sequencing (Illumina sequencing), genomic DNA was extracted using a TIANamp Blood DNA Kit DP348 (Tiangen Biotech Co. Ltd., Beijing, P.R. China).

PacBio and illumina library construction and sequencing

For short-read sequencing, at least 3 μg of genomic DNA was used to construct a paired-end sequencing library with an insert size of approximately 350 bp, which was then sequenced on the Illumina HiSeq X Ten and HiSeq 2000 platforms (Illumina, San Diego, CA, United States) following the manufacturer's instructions at Novogene. Clean reads were obtained by removing reads containing adapters, poly-N, and low-quality reads from raw data (average 10X coverage of the chicken genome).

For PacBio continuous long reads, libraries with an average insert size of 20 kb were constructed using the SMRTbell Template Prep Kit. PacBio sequencing was performed on the Pacific Bioscience Sequel II platform. Finally, Smrtlink was used to filter low-quality raw data (minLength = 50, minReadScore = 0.8).

Alignment and structural variants detection

Different sequencing techniques were used for read alignment and SV calling. For short-read sequencing, the filtered reads were mapped onto the GRCg6a reference genome using the Burrows-Wheeler Aligner (BWA-V0.7.17) with default parameters (Li and Durbin, 2010). Mapping results were then converted into BAM format and sorted using SAMtools version 1.9 (Li et al., 2009). Duplicate reads were removed using Picard version 2.3 (https://broadinstitute.github.io/picard/). GATK-V4.1.9.0 (McKenna et al., 2010) (default settings) was then used to call the raw single nucleotide polymorphisms (SNPs).

A consensus-based approach was used to call the SVs, which involved parallel SV calls by Delly_v0.8.7 (Rausch et al., 2012) and Manta_v1.6.0 (Chen et al., 2016) to obtain a high-confidence set of SVs. In this study, we used the default parameters recommended by the software for structural variation detection. The final SVs were filtered using the following criteria: SVs passing the quality filters (flag PASS) with a length of ≥50 bp. Finally, SV events were merged into a VCF file using SURVIVOR-V0.1.7 for downstream analysis.

For long-read sequencing, Smrtlink was used to filter the low-quality raw data. Alignment of subreads to the GRCg6a chicken reference was undertaken using aligner NGMLR-V0.2.7 (Sedlazeck et al., 2018) and called for genome-wide SV using the SV-calling algorithms Sniffles-V1.0.12 and SVIM-V1.2.0 (Sedlazeck et al., 2018; Heller and Vingron, 2019). To reduce the number of false positives, a strict filtering protocol was followed.

Uncertain and low-quality SVs (flag: IMPRECISE/UNRESOLVED) were removed and only SV calls longer than 50 bp with more than four supporting reads were retained. In the same position, only the SVs that were most supported were

FIGURE 1
Overview of the workflow of this study.
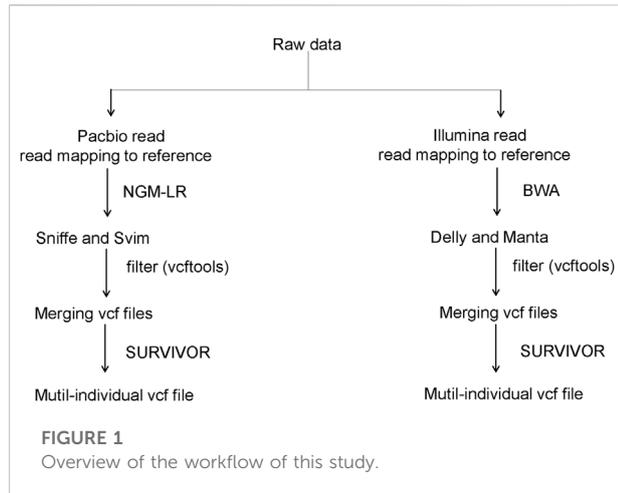
retained. Finally, the SV events predicted for each individual were merged into a VCF file using SURVIVOR (Jeffares et al., 2017) for downstream analysis (Figure 1 illustrates the operation process). Intersection sets for SVs across each chicken were created using UpSetPlot (http://bioconductor.org/biocLite.R).

Determination of structural variants hotspots

We divided the chicken genome into 2,117 non-overlapping 500 kb intervals. We then mapped 46,825 SVs into 2,117 intervals with four types of variations (INS, INV, DUP, and DEL). We assumed that the number of SVs mapped to each interval would follow a Poisson probability distribution if the SVs were distributed randomly across the genome. We generated an expected Poisson distribution, which was used to determine the criteria for SV hotspots. The intervals containing an empirical SV number equal to or higher than the 99th percentile of the expected Poisson probability distribution were classified as SV hotspots. The genes were annotated using the VEP of ENSEMBL (http://ensembl.org/Tools/VEP). The gene ontology (GO) enrichment analysis was conducted using DAVID (https://david.ncifcrf.gov/). Functional clustering included three aspects: molecular function (MF), cell component (CC), and biological progress (BP). The GO terms with an adjusted value smaller than 0.05 were considered significant.

Genetic structure analysis

The obtained SV dataset was used to construct phylogenetic trees using the NJ method for the SNP data. VCF tools (v0.1.13) were used to convert VCF files to a PLINK format, and then in-house bash scripts were used to convert these into a PHYLIP

format for input into IQTree (Nguyen et al., 2015). The tree was plotted using iTOL (https://itol.embl.de/). We performed a principal component analysis for the SNP genotypes and visualized the results in R using the package SNPrelate. (Zheng et al., 2012).

Results

Sequencing and structural variants detection

In long-read sequencing, the average sequencing depth of 10 chickens of different breeds was approximately 39x, and in short-read sequencing, the average sequencing depth was approximately 19x (Table1). Our sequencing data with sufficient coverage ensured high-sensitivity SV calls. The SVs in this study were larger than 50 bases, and there was no limit for the maximum length. Long sequencing technology can cover substantial chromosomal regions and directly reveal long deletions, insertions, duplications, inversions, and translocations, allowing the detection of structural differences between genomes in each chicken. The 49,501 SVs were obtained only from PacBio data, representing 23,817 deletions, 3,292 duplications, 20,847 insertions, 407 inversions, and 1,138 translocations using two detection tools. For short-read sequencing, 22,348 SVs were detected. The proportion of deletions was the highest among the different SV types. In this study, long-read sequencing identified more SVs than short-read sequencing did. We found that the stability and reproducibility of SV results from the long-read sequencing software were better than those of SV results from the short-read sequencing software. The number of SVs detected by Sniffles and SVIM was similar, and the number of overlapping SVs was higher, whereas the number of structural variants detected by Delly and Manta was different, and the number of overlapping SVs was lower. We compared the results of the SVs generated by the two platforms to identify overlapping variants. The results showed that short-read and long-read sequencing overlapped 51.87% of the SVs (Table2; Figure 2).

Genome-wide structural variants distribution and structural variants hotspot analysis by PacBio data

Genomic SVs are not evenly distributed on chromosomes (Perry et al., 2006; Gokcumen et al., 2011). Analyzing the bias of different types of SVs on chromosomes facilitates the understanding of the distribution characteristics of genetic variation and the impact of SVs on the genome. The longer the chromosome, the more SVs it contains (Figure 3A). For example, Chr1 to Chr5 account for 21.7%, 16.2%, 9.8%, 7.6%, and

TABLE 1 Summary of sequencing information for the 10 chicken samples.

Sample	Type	Sex	Pacbio data				Illumina data	
			Average Depth	Mean Reads	N50 for Raw reads	Mapping Ratio	Average Depth	Mapping Ratio
SK	indigenous chicken	F	41.14X	9.5 kb	15.54 kb	87.38%	45.87	99.13%
TB		F	25.44X	10.1 kb	11.4 kb	96.27%	13.73	99.12%
BY		F	52.87X	17.6 kb	20.5 kb	94.73%	14.04	99.12%
P		M	47.89X	13.4 kb	14.3 kb	94.36%	9.60	97.00%
DT		M	82.11X	17.9 kb	20.7 kb	94.42%	10.74	98.66%
Cor	Commercial chicken	F	23.71X	9.3 kb	10.4 kb	96.64%	12.53	99.21%
WR		F	22.40X	9.9 kb	11.2 kb	96.29%	12.72	99.14%
WL		F	41.23X	10.1 kb	15.55 kb	87.76%	48.04	99.29%
RIR		F	25.17X	10.7 kb	12.2 kb	96.29%	12.10	99.15%
RJF	Wild Breed	M	23.69X	6 kb	7.4 kb	93.1%	8.13	99.15%

Silkies, SK; tibetan, TB; beijing you, BY; piao, P; dong tao, DT; cornish, Cor; White Plymouth Rock, WR; white leghorn, WL; rhode island red, RIR; red jungle fowl, RJF.

TABLE 2 Number of SVs detected by Illumina short reads and PacBio long reads in 10 chicken samples.

Sample	Long-read sequencing SV number		Short-read sequencing SV number			
	sniffle	svim	final	delly	Manta	final
TB	12995	14136	11109	11003	20142	4472
BY	20160	21662	15127	11246	20984	4714
P	21155	23185	15080	8206	21339	2877
DT	24990	26585	17410	8113	23983	3161
SK	17357	19702	14823	19236	37169	7681
WL	15663	17331	13219	18275	33956	7507
Cor	11052	11829	9486	9136	17104	3733
RIR	12311	13306	10465	9550	17470	3862
WR	11253	12162	9668	10211	18708	4239
RJF	15665	16153	10238	6543	12168	2603
Total		49501			22348	

4.3% of the total SVs, respectively. It is interesting to note that micro-chromosomes contain a greater density of variation than longer chromosomes. For instance, chromosomes 30 and 32 have a density of 1 SV every 2.94 kb (620 SVs in 1.8 Mb) and 2.06 kb (353 SVs in 0.7 Mb), respectively. For comparison, large chromosomes show much less structural variation (e.g. One every 20 Kb on chromosome 1); the detailed results are presented in Supplementary Table S2. The distribution of SVs had several hotspots on the chromosomes. We further investigated the size of the identified SVs across the chicken genomes, and the majority of smaller-sized SVs were less than 1 kb. Most inversions were ~680 Mb, spanning the largest region among all types of SVs. To observe the distribution of all the SVs throughout the chicken genome, we used SVs to generate a genome-wide variation distribution map. An investigation of chicken genome SV density, and a better understanding of SV

length and number chromosomal distribution. Figure 3B shows the density and distribution of all SVs we used for analysis on chicken chromosome ideogram. The SV density of micro-chromosomes is higher than that of large chromosomes in some regions.

We identified SV hotspots in the chicken genome by comparing the expected and empirical distributions of the SVs (Figure 4). A total of 113 intervals (interval size of 500 kb, 7% of all intervals) containing ≥33 SVs in the empirical distribution were characterized as SV hotspots. These intervals fall within the 99th percentile (or higher) of the expected distributions. All intervals included 6,180 SVs with 2,957 deletions, 791 duplications, 2,396 insertions, and 35 inversions. In total, 768 genes were identified in the defined SV hotspot regions. Among these, the largest number of SV sites (n = 517) was detected on chromosome 2. The SV hotspots are presented in

FIGURE 2
Comparison of structural variations in PacBio and Illumina sequencing platform.
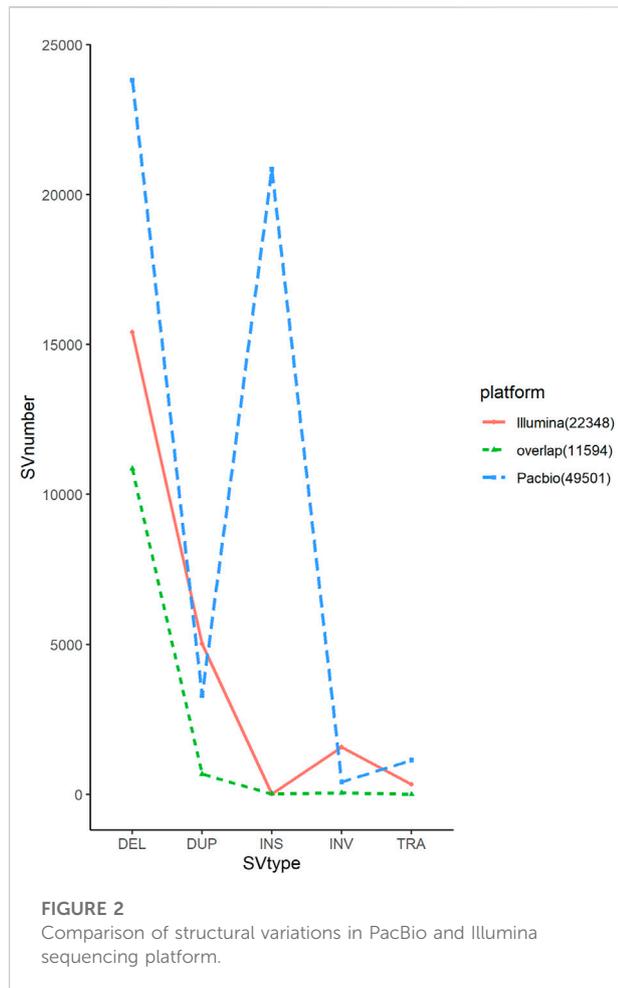
Supplementary Table S3. To functionally infer the SV hotspot data for chickens, we conducted functional enrichment analysis for genes positioned within the SV hotspots. Gene enrichment and ontology analysis results showed that the genes were related to terms such as regulation of stem cell population maintenance, propanoate metabolism, cell proliferation, and calcium channel regulator activity (Supplementary Table S3).
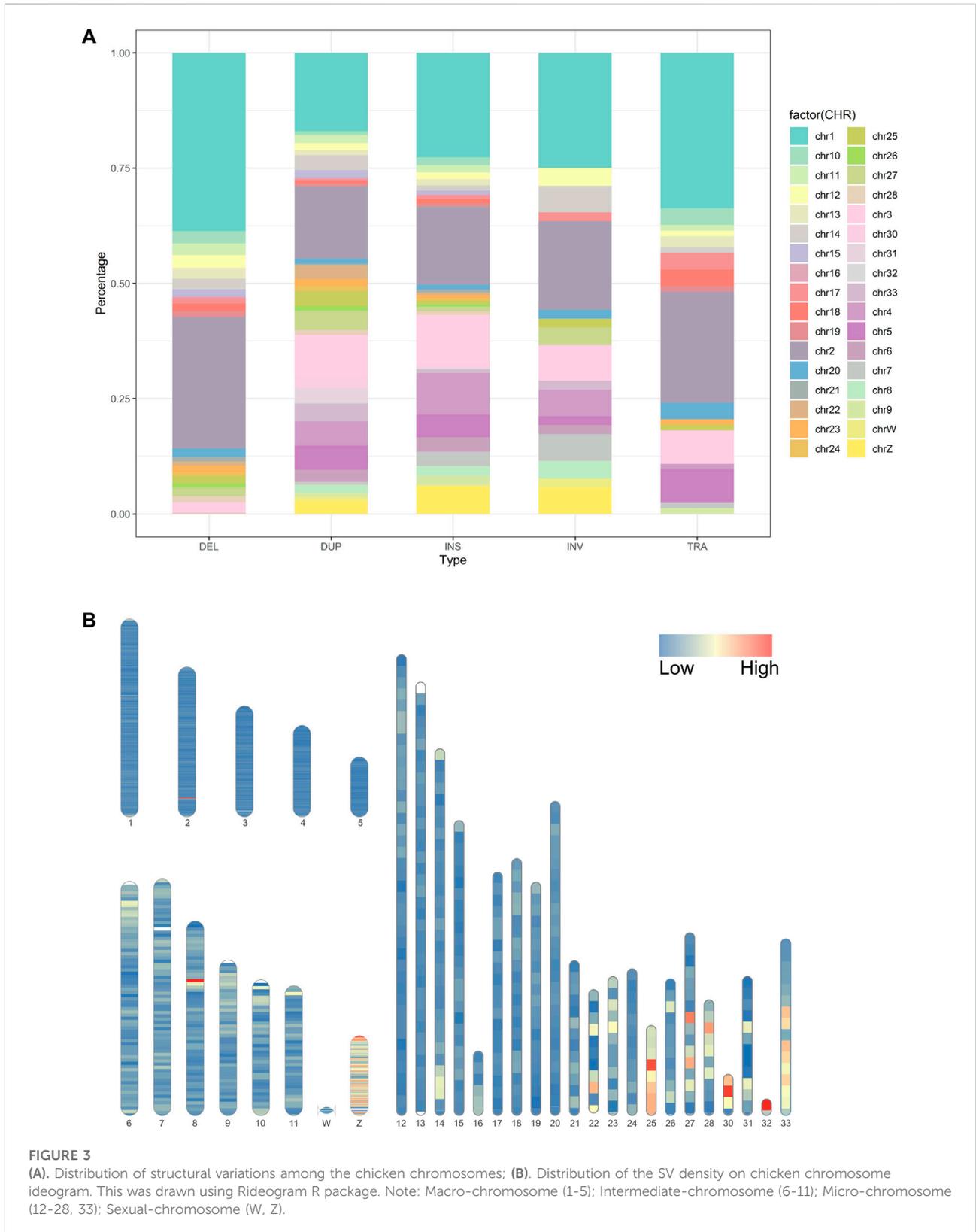
Special structural variants in chickens

To date, little is known about the SVs of some unique indigenous Chinese chickens. To further explore the relationship between these specific SVs and chicken phenotypes, we analyzed SVs in 10 chickens with different phenotypes. There were 654 SVs in all 10 individuals and 27,454 SVs were observed in only one individual. Among them, DT (5,186) and RJF (4,401) chickens had more specific SVs, and highly selective commercial chickens (COR, WR, RIR, and WL) had 2,522; 2,714; 3,224; and 5,207 specific SVs, respectively. A total of 3,559; 4,308; 2,349; and 3,570 specific

SVs for BY, P, TB, and SK, respectively, were detected (Supplementary Table S4). These results showed that more intensive artificial selection resulted in a reduction in genetic variation. To investigate the common SVs among chickens, we analyzed the SVs numbers among chickens of similar backgrounds (Asian and European). Different chickens exhibit varying degrees of common SVs. We showed the intersections of ten chickens for SV numbers using an UpSet plot, where interactions were ranked by SV number. A total of 1023 intersections are included. In Asian chickens, 63 intersections share among themselves, while only 15 intersections share among themselves in European chickens (Supplementary Table S5). The first 50 intersections are shown in Figure 5. The plot showed that most SVs are private to each individual, followed by SVs shared in all chickens. It implies that the reference genome lacks portions relevant to most animals. Additionally, there is also a great deal of SVs (15/50) that are shared privately in Europe or Asian populations. There could be some phenotypic and growth differences in the chickens affected by these SVs (Supplementary Table S6). These SVs may affect the phenotypes and growth characteristics of chickens. Previous studies have shown that the *SOX5* gene is associated with pea comb in chickens (Wright et al., 2009). A *SOX5* mutation was found in the genome of Dongtao chickens, consistent with previous observations. A mutation in *SH3RF2* was found in WR chickens, which was previously reported to influence chicken growth. We also identified several MHC-related genes (*BZFP1*, *TAP1*, and *IL4I1*), all of which are located on chromosome 16. In addition, *AKAP8L*, *ASCC3*, *BG8*, *OPN5L1*, and *SH3RF2* are involved in the immune response. Several genes, including *AKAP8L*, *ASCC3*, *BG8*, *OPN5L1,* and *SH3RF2*, were also found to be involved in the immune response. It is possible that these SVs may affect indigenous chickens, resulting in stronger disease resistance than those in commercial chickens.

Genetic structure of chickens

We used SVs to infer the genetic structure of all 10 chickens (Figure 6). PCA indicated that each chicken was distinctly different from the others. The first principal component (PC1) distinguished wild RJF chickens from domestic chickens. Commercial chickens were distributed in the upper half of the second principal component (PC2), while local chickens were mostly distributed in the lower half of PC2, indicating that indigenous chickens were genetically different from commercial chickens based on SVs (Figure 6A). We constructed phylogenetic trees for chickens (Figure 6B), where all chickens were divided into three major branches: red jungle fowl (RJF), commercial chickens (WR, WL, RIR, and COR), and indigenous chickens (TB, BY, SK, DT, and P). In brief, the phylogenetic tree and PCA analyses support the division of

FIGURE 3
(A). Distribution of structural variations among the chicken chromosomes; **(B)**. Distribution of the SV density on chicken chromosome ideogram. This was drawn using Rideogram R package. Note: Macro-chromosome (1-5); Intermediate-chromosome (6-11); Micro-chromosome (12-28, 33); Sexual-chromosome (W, Z).
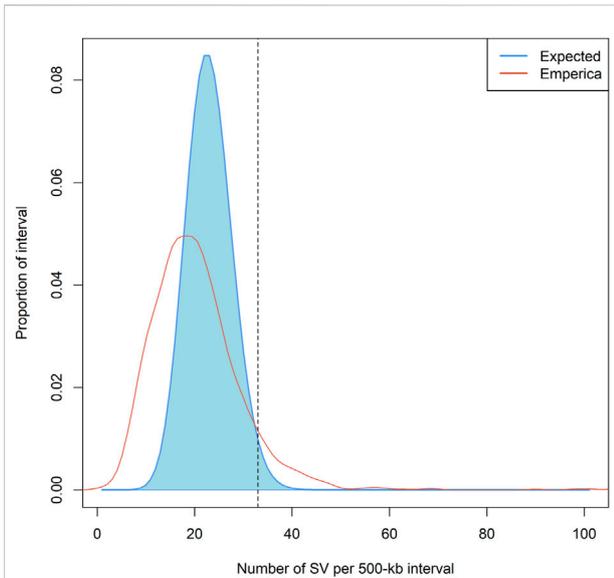
FIGURE 4
Non-random distribution of SVs across chicken genomes.
Note: The density plot showed the empirical and the expected
probability distributions of the number of SV (s) within a 500 kb
long interval of the chicken reference genome (gal6). The
empirical probability distribution was estimated by the actual
number of SVs within each of the 2117 intervals. The expected
probability distribution was estimated assuming a Poisson
distribution.

the panel into commercial chickens, indigenous chickens, and
wild chickens, and the results of SVs could help in understanding
their population genetics.

Discussion

SVs have been recognized as important sources of genetic
variation, and are the main contributors to phenotypic diversity
and evolutionary adaptation in chickens (Wang and Byers, 2014;
Seol et al., 2019; Fernandes et al., 2021). Compared with SNPs,
SVs tend to cause changes in the gene structure, which could
change their function and cluster within SV hotspots, which are
beneficial to organisms during the evolution process (Liu et al.,
2020a). In this study, we used long-read sequencing technology
to detect SVs in 10 chickens, analyzed structural variation in the
chicken genome, and identified several SVs that may be related to
chicken growth and reproduction, which can provide further
information for future studies.

PacBio sequencing appears to be more effective than
Illumina sequencing in detecting SVs, which is consistent
with Mahmoud's findings. However, some evidence suggests
that short-read data with a high sequencing depth may be able
to detect structural variations of the same or greater length than
long-read data (Geibel et al.). Several factors may lead to
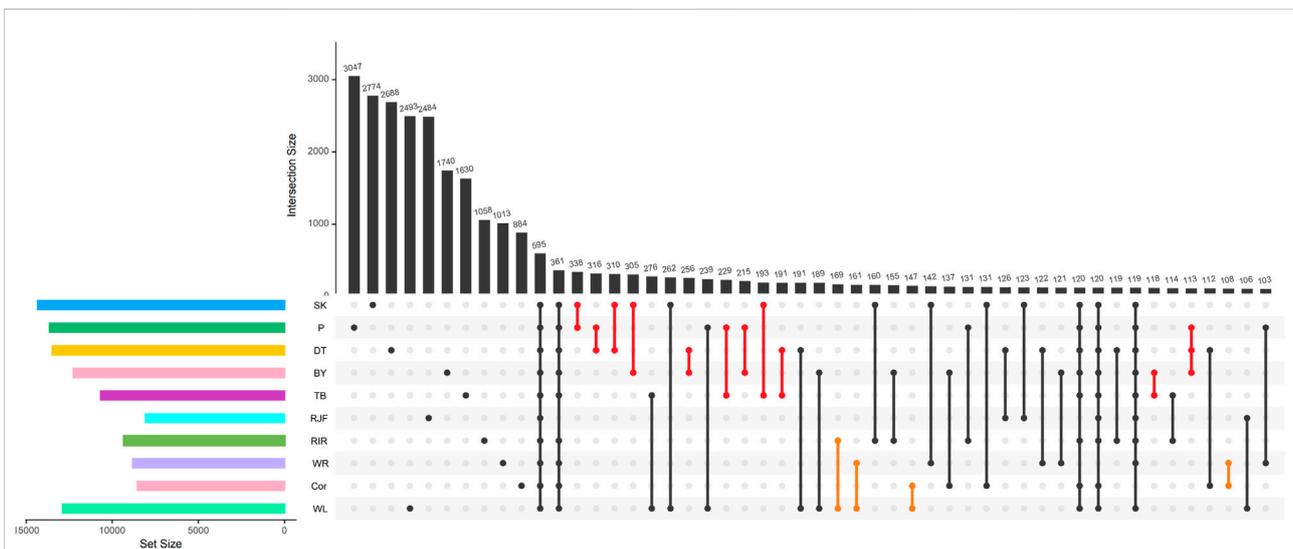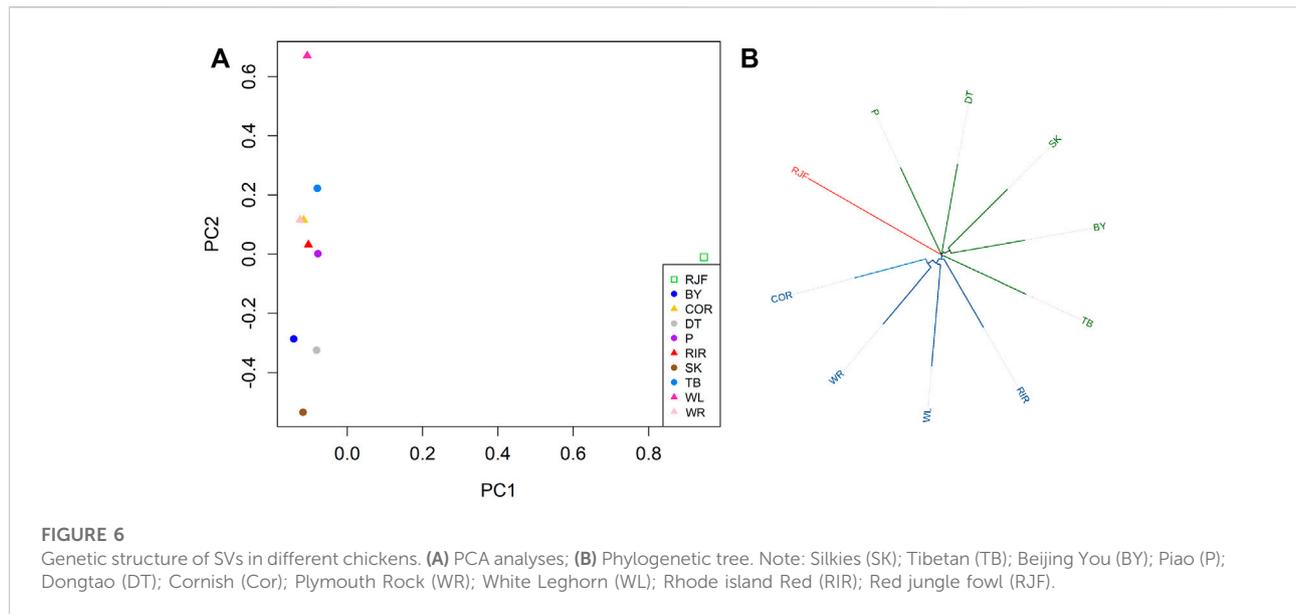inconsistent results, including species, variation complexity,

FIGURE 5
UpSet plot of SVs results detected in each chicken. The intersection between Asian chickens is represented by a vertical red line; the intersection
between European chickens is represented by a vertical orange line, the others is represented by a vertical black line. Note: In the upper left, Venn
diagram of the same genetic background. In the bottom left panel, horizontal bars represent the number of SVs detected by each chicken, vertical
bars represent the size of SVs detected in each chicken, black dots represent the sample set, and the intersection between chickens is
represented by a vertical black line.

FIGURE 6
Genetic structure of SVs in different chickens. **(A)** PCA analyses; **(B)** Phylogenetic tree. Note: Silkies (SK); Tibetan (TB); Beijing You (BY); Piao (P); Dongtao (DT); Cornish (Cor); Plymouth Rock (WR); White Leghorn (WL); Rhode island Red (RIR); Red jungle fowl (RJF).

data accuracy, and software detection efficiency. As long-read sequencing technology matures and becomes more efficient, the detection errors caused by sequencing methods will decrease. PacBio has launched its CCS sequencings, which greatly improves detection accuracy. The emergence of a number of excellent mutation detection software programs has also helped promote mutation research, identifying SVs based on machine learning, like SVFX (Kumar et al., 2020) and dysgu (Cleal and Duncan, 2022). The dysgu software supports the merging of SVs from different callers using different sequencing technologies. It found that the combination of low coverage paired-end and long-reads performs as well as long-reads with higher coverage.

SVs are not distributed uniformly across the genome, as observed in other studies (Liao et al., 2021). The number in each chromosome ranged from 10,148 to 64. The large reference genome tends to have a larger number of SVs, suggesting that each chromosome may have been subjected to different selection pressures during the selection process. Among the different types of SVs, deletions and insertions accounted for the highest proportion, whereas duplications, inversions, and translocations accounted for a smaller proportion (Figure 3). The insertion detection rate was significantly different between long-read and short-read sequencing in this study (Figure 2). Several novel genes have been identified in domestic chickens in a recently published pan-genomic study (Li et al., 2022). To investigate the relationship between insertional SVs in this study and novel genes in chickens, we compared their coordinates with the positions of the identified insertional SVs and found that 4,874 SVs were located in 173 newly discovered genes (Supplementary Table S7). The significance

of the new gene for domestic chickens remains to be determined.

Potential candidate genes affected by structural variants

We identified several genes that were affected by structural variation. They are associated with phenotypes, economic traits, and disease resistance in domestic chicken. Crest is an incomplete dominance mutation that replaces small feathers on the head with dorsal skin feathers (Wang et al., 2012). Li et al. (2021) used short-read re-sequencing to determine that *HOXC10* is related to the crest phenotype in Silkie chickens. Our study also identified this gene mutation in chickens with SK. In contrast to the duplication of SV reported by Li et al., this one is located in an intron with a 19-bp inserted fragment. It is necessary to confirm the results of the mutations affecting this trait. Previously, researchers reported that the expression of *IRX1* and *IRX2* differed between normal and rumpless chickens (Nowlan et al., 2013). Piao chicken is an indigenous Chinese chicken breed lacking a pygostyle, caudal vertebra, uropygial gland, and tail feathers. Our study found a deletion variant (chr2:86914914-86919099) and a duplication variant (2:80868684-87846701) on chromosome two of Piao chicken, the region perches *IRX1* and *IRX2* genes which have been reported to be related to rumplessness in Araucana chickens (Nowlan et al., 2013). This duplication occurred at the lncRNA gene, which is adjacent to *IRX2*. Studies have shown that lncRNAs can regulate the expression of downstream genes by inhibiting the aggregation of ENA polymerase II and inducing chromatin remodeling (Wilusz et al., 2009). In addition, a deletion occurs near *IRX1* in the non-coding region. Non-coding regions in

biological genomes account for the majority, and these non-coding regions may have indirect regulatory effects on genes (Alexander et al., 2010). VEP annotation showed that both the SVs were high-impact effector mutations. Hence, we hypothesized that this region might be related to rumpless Piao chickens. We performed PCR to confirm the authenticity of the variant in the subsequent rumplessness of the Piao chicken study. There is still work to be done on the genetic mechanisms of rumplessness in Piao chickens. Our structural variation results also revealed some candidate genes that have been reported to be associated with chicken feather color (*SOX10, CDKN2A*) (Gunnarsson et al., 2011; Schwochow Thalmann et al., 2017), growth (*MYF5, IGF2BP1, CRY1; SH3RF2*) (Jing et al., 2020; Niessner et al., 2011; Krause et al., 2016; Zhang et al., 2020; Wang et al., 2021).

Application of long-read sequencing in structure variation detection

High-throughput sequencing technology is the mainstream method for SV detection and is considered the gold standard in SV research (Mahmoud et al., 2019; Ho et al., 2020). An increasing number of animal studies have employed long-read sequencing (Long et al., 2018; Bertolotti et al., 2020; Luan et al., 2020). In our study, 49,501 and 22,348 SVs were detected using PacBio and Illumina data, respectively. PacBio sequencing is more sensitive than Illumina data; long-read sequencing can reveal large and complex SV events that are often neglected by short-read sequencing, corroborating the results of a previous study (Liu et al., 2020b). The emergence of long-read sequencing has introduced opportunities and challenges to SVs. More and more studies have showed that SV has the ability to capture the genetic structure differences in breeds and can be used to study population genetic structure. In 2020, researchers carried out structural variation detection in tomatoes using long-read sequencing, and a high-quality pan-genome of SV was constructed, revealing the effects of structural variations on tomato fruit flavor, size, and yield (Alonge, 2020). Weissensteiner (2020) used long-read sequencing to perform SV detection and population evolution in the songbird genera. These results suggest the wealth and evolutionary significance of SV segregation in natural populations. SV combined with selective signal analysis and GWAS analysis is increasing. Because of the limited number of third-generation sequencing samples in this study, it is impossible to carry out population genetic studies based on third-generation sequencing; however, it is believed that SV as a molecular marker will be more widely studied in population genetics in the future.

Conclusion

SVs are the main source of genomic complexity. We identified chicken SVs using PacBio sequencing. In total, 49,501 SVs were identified across ten chicken breeds. SVs were not evenly distributed in the genomes, and there were several SV-hot sites. During selective breeding of chickens, some SVs that were beneficial to the breed or SVs that had no effect on the genome function of the breed were retained, whereas deleterious SVs were eliminated. Finally, we identified SV-related genes associated with growth, reproduction, and phenotypic appearance that could be artificially selected during chicken domestication.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi. nlm.nih.gov/search/all/?term=PRJNA807790.

Ethics statement

The animal study was reviewed and approved by Animal Welfare Committee of China Agricultural University.

Author contributions

LQ and ZN conceived the study and designed the project; JZ performed bioinformatics analyses; HB, CZ, CN, XL, and JinZ interpreted the result; JL, XZ, and JH collected samples and measured data; YC, YJ, LW, XL, WY, KL, and JianZ contributed to reagents and materials. All authors reviewed the manuscript.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene. 2022.971588/full#supplementary-material

References

Alexander, P., Fang, G., Rozowsky, J., Snyder, M., and Gerstein, B. (2010). Annotating non-coding regions of the genome. *Nat. Rev. Genet.* 11 (8), 559–571. doi:10.1038/nrg2814

Alkan, C., Coe, B. P., and Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12 (5), 363–376. doi:10.1038/nrg2958

Bertolotti, A. C., Layer, R. M., Gundappa, M. K., Gallagher, M. D., Pehlivanoglu, E., Nome, T., et al. (2020). The structural variation landscape in 492 Atlantic salmon genomes. *Nat. Commun.* 11 (1), 5176. doi:10.1038/s41467-020-18972-x

Bickhart, D. M., and Liu, G. E. (2014). The challenges and importance of structural variation detection in livestock. *Front. Genet.* 5, 37. doi:10.3389/fgene. 2014.00037

Chaisson, M. J. P., Sanders, A. D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* 10 (1), 1784. doi:10. 1038/s41467-018-08148-z

Chen, L., Chamberlain, A. J., Reich, C. M., Daetwyler, H. D., and Hayes, B. J. (2017). Detection and validation of structural variations in bovine whole-genome sequence data. *Genet. Sel. Evol.* 49 (1), 13. doi:10.1186/s12711-017-0286-5

Cleal, K., and Baird, D. (2022). Dysgu: Efficient structural variant calling using short or long reads. *Nucleic Acids Res.* 9 (50), 53. doi:10.1093/nar/gkac039

Dorshorst, B., Harun-Or-Rashid, M., Bagherpoor, A. J., Rubin, C.-J., Ashwell, C., Gourichon, D., et al. (2015). A genomic duplication is associated with ectopic eomesodermin expression in the embryonic chicken comb and two duplex-comb phenotypes. *PLoS Genet.* 11 (3), e1004947. [Online]Available at: http://europepmc. org/abstract/MED/25789773. doi:10.1371/journal.pgen.1004947

Elferink, M. G., Vallée, A. A. A., Jungerius, A. P., Crooijmans, R. P. M. A., and Groenen, M. A. M. (2008). Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BMC genomics* 9, 391. doi:10.1186/1471-2164-9-391

English, A. C., Salerno, W. J., Hampton, O. A., Gonzaga-Jauregui, C., Ambreth, S., Ritter, D. I., et al. (2015). Assessing structural variation in a personal genome—Towards a human reference diploid genome. *BMC Genomics* 16 (1), 286. doi:10.1186/s12864-015-1479-3

Fernandes, A. C., da Silva, V. H., Goes, C. P., Moreira, G. C. M., Godoy, T. F., Ibelli, A. M. G., et al. (2021). Genome-wide detection of CNVs and their association with performance traits in broilers. *BMC genomics* 22 (1), 354. doi:10.1186/s12864-021-07676-1

Gokcumen, O., Babb, P. L., Iskow, R. C., Zhu, Q., Shi, X., Mills, R. E., et al. (2011). Refinement of primate copy number variation hotspots identifies candidate genomic regions evolving under positive selection. *Genome Biol.* 12 (5), R52. doi:10.1186/gb-2011-12-5-r52

Gunnarsson, U., Kerje, S., Bed'hom, B., Sahlqvist, A.-S., Ekwall, O., Tixier-Boichard, M., et al. (2011). The Dark Brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. *Pigment. Cell Melanoma Res.* 24 (2), 268–274. doi:10.1111/j.1755-148X.2011.00825.x

Guo, Y., Gu, X., Sheng, Z., Wang, Y., Luo, C., Liu, R., et al. (2016). A complex structural variation on chromosome 27 leads to the ectopic expression of HOXB8 and the muffs and beard phenotype in chickens. *PLoS Genet.* 12 (6), e1006071. doi:10.1371/journal.pgen.1006071

Heller, D., and Vingron, M. (2019). SVIM: Structural variant identification using mapped long reads. *Bioinforma. Oxf. Engl.* 35 (17), 2907–2915. doi:10.1093/ bioinformatics/btz041

Ho, S. S., Urban, A. E., and Mills, R. E. (2020). Structural variation in the sequencing era. *Nat. Rev. Genet.* 21 (3), 171–189. doi:10.1038/s41576-019-0180-9

Imsland, F., Feng, C., Boije, H., Bertrand, B., Fillon, V., Dorshorst, B., et al. (2012). Bed'homThe Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet.* 8 (6), e1002775. doi:10.1371/journal.pgen.1002775

Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., et al. (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* 8 (1), 14061. doi:10.1038/ncomms14061

Jing, Z., Wang, X., Cheng, Y., Wei, C., Hou, D., Li, T., et al. (2020). Detection of CNV in the SH3RF2 gene and its effects on growth and carcass traits in chickens. *BMC Genet.* 21 (1), 22. doi:10.1186/s12863-020-0831-z

Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., and Kamatani, Y. (2019). Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol.* 20 (1), 117. doi:10.1186/s13059-019-1720-5

Krause, E. T., Schrader, L., and Caspers, B. A. (2016). Olfaction in chicken (Gallus gallus): A neglected mode of social communication. *Front. Ecol. Evol.* 4 (94). doi:10. 3389/fevo.2016.00094

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 26 (5), 589–595. doi:10. 1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinforma. Oxf. Engl.* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, J., Lee, M.-O., Davis, B. W., Wu, P., Hsieh Li, S.-M., Chuong, C.-M., et al. (2021). The crest phenotype in domestic chicken is caused by a 197 bp duplication in the intron of HOXC10. *G3* 11 (2), jkaa048. doi:10.1093/g3journal/jkaa048

Li, M., Su, C., Xu, N., Bian, P., Tian, X., Wang, X., et al. (2022). De novo assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on microchromosomes and subtelomeric regions. *Mol. Biol. Evol.* 39 (4), msac066. doi:10.1093/molbev/msac066

Liao, Z., Zhang, X., Zhang, S., Lin, Z., Zhang, X., and Ming, R. (2021). Structural variations in papaya genomes. *BMC Genomics* 22 (1), 335. doi:10.1186/s12864-021-07665-4

Liu, C., Ran, X., Yu, C., Xu, Q., Niu, X., Zhao, P., et al. (2019). Whole-genome analysis of structural variations between Xiang pigs with larger litter sizes and those with smaller litter sizes. *Genomics* 111 (3), 310–319. doi:10.1016/j.ygeno.2018.02.005

Liu, G. E., Hou, Y., Zhu, B., Cardone, M. F., Jiang, L., Cellamare, A., et al. (2010). Analysis of copy number variations among diverse cattle breeds. *Genome Res.* 20 (5), 693–703. doi:10.1101/gr.105403.110

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., et al. (2020a). Pan-genome of wild and cultivated soybeans. *Cell* 182 (1), 162–176. e113. doi:10.1016/j.cell. 2020.05.023

Liu, Y., Zhang, M., Sun, J., Chang, W., Sun, M., Zhang, S., et al. (2020b). Comparison of multiple algorithms to reliably detect structural variants in pears. *BMC genomics* 21 (1), 61. doi:10.1186/s12864-020-6455-x

Long, E., Evans, C., Chaston, J., and Udall, J. A. (2018). Genomic structural variations within five continental populations of *Drosophila melanogaster*. *G3* 8 (10), 3247–3253. doi:10.1534/g3.118.200631

Luan, M.-W., Zhang, X.-M., Zhu, Z.-B., Chen, Y., and Xie, S.-Q. (2020). Evaluating structural variation detection tools for long-read sequencing datasets in *Saccharomyces cerevisiae*. *Front. Genet.* 11, 159. doi:10.3389/fgene.2020.00159

Luo, J., Yu, Y., Mitra, A., Chang, S., Zhang, H., Liu, G., et al. (2013). Genome-wide copy number variant analysis in inbred chickens lines with different susceptibility to Marek's disease. *G3* 3 (2), 217–223. doi:10.1534/g3.112.005132

Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., and Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biol.* 20 (1), 246. doi:10.1186/s13059-019-1828-7

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

Merker, J. D., Wenger, A. M., Sneddon, T., Grove, M., Zappala, Z., Fresard, L., et al. (2018). Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet. Med.* 20 (1), 159–163. doi:10.1038/gim.2017.86

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32 (1), 268–274. doi:10.1093/molbev/msu300

Niessner, C., Denzau, S., Gross, J. C., Peichl, L., Bischof, H.-J., Fleissner, G., et al. (2011). Avian ultraviolet/violet cones identified as probable magnetoreceptors. *PloS one* 6 (5), e20091. doi:10.1371/journal.pone.0020091

Perry, G. H., Tchinda, J., McGrath, S. D., Zhang, J., Picker, S. R., Cáceres, A. M., et al. (2006). Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl. Acad. Sci. U. S. A.* 103 (21), 8006–8011. doi:10.1073/pnas.0602318103

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464 (7288), 587–591. doi:10.1038/nature08832

Schwochow Thalmann, D., Ring, H., Sundström, E., Cao, X., Larsson, M., Kerje, S., et al. (2017). The evolution of Sex-linked barring alleles in chickens involves both regulatory and coding changes in CDKN2A. *PLoS Genet.* 13 (4), e1006665. doi:10.1371/journal.pgen.1006665

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., et al. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* 15 (6), 461–468. doi:10.1038/s41592-018-0001-7

Seol, D., Ko, B. J., Kim, B., Chai, H.-H., Lim, D., and Kim, H. (2019). Identification of copy number variation in domestic chicken using whole-genome sequencing reveals evidence of selection in the genome. *Animals.* 9 (10), 809. doi:10.3390/ani9100809

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet.* 34 (9), 666–681. doi:10.1016/j.tig.2018.05.008

Wang, K., Hu, H., Tian, Y., Li, J., Scheben, A., Zhang, C., et al. (2021). The chicken pan-genome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. *Mol. Biol. Evol.* 38 (11), 5066–5081. doi:10.1093/molbev/msab231

Wang, X., and Byers, S. (2014). Copy number variation in chickens: A review and future prospects. *Microarrays (Basel, Switz.* 3 (1), 24–38. doi:10.3390/microarrays3010024

Wang, Y., Gao, Y., Imsland, F., Gu, X., Feng, C., Liu, R., et al. (2012). The crest phenotype in chicken is associated with ectopic expression of HOXC8 in cranial skin. *PLoS One* 7, e34012. doi:10.1371/journal.pone.0034012

Wilusz, J., Sunwoo, H., and Spector, D. (2009). Long noncoding RNAs: Functional surprises from the RNA world. *Genes Dev.* 23 (13), 1494–1504. doi:10.1101/gad.1800909

Wright, D., Boije, H., Meadows, J. R. S., Bed'hom, B., Gourichon, D., Vieaud, A., et al. (2009). Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. *PLoS Genet.* 5 (6), e1000512. doi:10.1371/journal.pgen.1000512

Wu, Z., Jiang, Z., Li, T., Xie, C., Zhao, L., Yang, J., et al. (2021). Structural variants in Chinese population and their impact on phenotypes, diseases and population adaptation. bioRxiv, 2021.2002.2009.430378. doi:10.1101/2021.02.09.430378

Zhang, J., Nie, C., Li, X., Ning, Z., Chen, Y., Jia, Y., et al. (2020). Genome-wide population genetic analysis of commercial, indigenous, game, and wild chickens using 600K SNP microarray data. *Front. Genet.* 11, 543294. doi:10.3389/fgene.2020.543294

Zhao, P., Li, J., Kang, H., Wang, H., Fan, Z., Yin, Z., et al. (2016). Structural variant detection by large-scale sequencing reveals new evolutionary evidence on breed divergence between Chinese and European pigs. *Sci. Rep.* 6 (1), 18501. doi:10.1038/srep18501

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinforma. Oxf. Engl.* 28 (24), 3326–3328. doi:10.1093/bioinformatics/bts606