Front. Agr. Sci. Eng. 2025, 12(2): 355–366 https://doi.org/10.15302/J-FASE-2024591

RESEARCH ARTICLE

Graph-based pan-genome analysis reveals diversity of structural variations in native and commercial chicken

Yiming WANG, Zijia NI, Yinhua HUANG (✉)

State Key Laboratory for Farm Animal Biotech Breeding, College of Biology Sciences, China Agricultural University, Beijing 100193, China.

KEYWORDS

Graph-based pan-genome, chicken, next-generation sequencing, structural variations

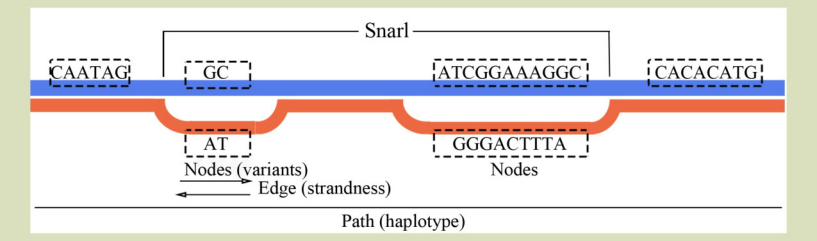
HIGHLIGHTS

- A graph-based pan-genome of native chicken was constructed.
- Structural variations related to egg production were identified in Leghorn.
- Structural variations related to highland adaption were identified in Tibetan chicken.
- A methodology for structural variations calling is proposed.

Received March 16, 2024;
Accepted September 11, 2024.

Correspondence: cauhyh@cau.edu.cn

GRAPHICAL ABSTRACT



ABSTRACT

Chickens are one of the most important domesticated animals, serving as an important protein source. Studying genetic variations in chickens to enhance their production performance is of great potential value. The emergence of next-generation sequencing has enabled precise analysis of single nucleotide polymorphisms and insertions/deletions in chicken, while third-generation sequencing achieves the accurate structural variant identification. However, the high cost of third-generation sequencing technology limits its application in population studies. The graph-based pan-genome strategy can overcome this challenge by enabling the detection of structural variations using cost-effective next-generation sequencing data. This study constructed a graph-based pan-genome for chickens using 12 high-quality genomes. This pan-genome used linear genome GRCg6a as the reference genome, containing variant information from two commercial and nine native chicken breeds. Compared to the linear genome, the pan-genome provided significant improvements in the efficiency of structural variation identification. On the basis of the graph-based pan-genome, high-frequency structural variations related to high egg production in Leghorn chicken were predicted. Additionally, it was discovered that potential structural variations was associated with highland adaptation in Tibetan chickens according to next-generation sequencing and transcriptomics data. Using the pan-genome graph, a new strategy to identify structural variations related to traits of interest in chickens is presented.

1 Introduction

Chicken (*Gallus gallus*) is one of the most important domesticate animals in the world. It originated from the southern Yunnan subspecies of the red jungle fowl, primarily in regions such as south-western China, northern Thailand and Myanmar[1]. As the most widely domesticated farm animal, chicken is extensively raised for both commercial and backyard farming purposes, surpassing the number of large animals such as pigs, sheep and cattle[2].

Chicken production is widely recognized as an excellent source of protein, with low fat and high nutritional value[3]. While most market chicken production is derived from commercial strains, native chicken is an alternative source for chicken production, which is highly regarded as a delicacy, particularly in Asian food culture[4–6]. In addition to being a specialty food source, the study of native chickens possess significant value in enhancing comprehension of avian environmental adaptation and chicken breeding practices. Silkies are a native chicken breed with the rare fibromelanosis trait and high melanin content in their organs[7,8]. Tibetan chickens inhabit high-altitude areas and has strong adaptability to low-oxygen environments. It has become an important avian model in high-altitude adaption[9,10]. Naked Neck game fowls, with no feathers on their neck, were once widely used in cock fighting in North America. The naked neck trait is considered a valuable trait to resist heat stress for commercial chickens[11,12].

A high-quality reference genome is of great importance to accurately identify genetic variations in different chicken breeds. Previous studies on chicken primarily relied on linear reference genomes, such as GRCg6a, GRCg7b, and GRCg7w. However, as the availability of NGS (next-generation sequencing) data and genome assemblies increases, it is hard for a single linear reference genome to fully represent the whole chicken species[13]. Also, traditional methods based on linear reference genomes and NGS data have difficulty accurately identifying SVs (structural variations with lengths of > 50 bp)[14,15]. The limitation and demand of SV detection prompted the development of pan-genomics, which incorporates genetic information from multiple individuals to create a reference genome with a more comprehensive representation of the species[16]. In the past, pan-genomes were initially constructed using the iterative assembly method. However, this approach cannot accurately represent the original genomic sequences of individuals despite containing genetic information from multiple genomes[17]. On the basis of a graph methodology strategy, the emergence of graph-based pan-genomes has addressed this issue and made a crucial contribution to recent genomics studies[18,19]. In a graph-based pan-genome, genetic variants are stored as nodes with edges encoding the strandedness, and the regions where variations appear are defined as snarls in the graph[20]. In this strategy, the genome becomes nonlinear and different haplotypes are represented as multiple paths. Unlike to the iterative method, the structure of graph-based pan-genome will avoid overwriting of variants from different individuals.

We constructed a graph-based pan-genome for worldwide chicken using 12 high-quality assemblies from 12 chicken breeds. Compared to the linear genome-based method, our pan-genome graph outperformed in genome NGS data alignment and SV detection. The graph-based pan-genome enabled us to identify breed-specific SVs in the chicken population using NGS data. This effort revealed the potential SVs related to high egg production and identified breed-specific SVs related to environmental adaption in Leghorn and Tibetan chicken. In summary, our study presents a reliable methodology for graph-based SV calling for agricultural animals.

2 Materials and methods

2.1 Data resource

Two commercial and nine native chicken genomes assembled in previous studies were downloaded from NCBI chicken genome databases[21,22]. The Illumina short reads of NGS used in this study were collected from the NCBI Sequence Read Archive[23]. Transcriptomics data used in this study were collected from the CNCB genome sequence archives and NCBI sequence read archive[24,25].

2.2 Graph genome construction

The graph-based pan-genome was constructed by the Minigraph-Cactus pipeline[26,27]. GRCg6a assembled from red jungle fowl data was used as a reference genome for graph-based pan-genome construction. In addition to GRCg6a, the other 11 assemblies were used in the following steps. First, the *cactus-minigraph* function was used to construct a GFA minigraph based on a series of chicken genome FASTA files. Second, the *cactus-graphmap* function was used to map each input assembly back to the graph. Third, the *cactus-graphmap-split* function was used to split the input assemblies and PAF into chromosomes to reduce the computing cost according to rGFA (reference graphical fragment assembly) tags in the GFA (graphical fragment assembly), whose format is suitable for

both Bruijn graph- and string graph-based assemblers. Fourth, the *cactus-align* function was used to complete the assembly to graph minigraph alignment of Cactus multiple genomes. Finally, the *cactus-graphmap-join* function was used to join the chromosome graph and produced the final graph genome.

2.3 Structural variation calling by graph-based pan-genome and linear genome

VG is a complementary tool for SV calling based on graph-based pan-genomes, which was used for NGS data mapping and SV calling in this study[28]. The VG Giraffe was used to align high-depth NGS data to the graph-based pan-genome[29]. NGS data from commercial and native chicken breeds were used to validate the reliability of the Minigraph-Cactus pipeline. VG Pack and VG Call toolkits were used to generate the VCF file after the alignment with VG Giraffe[28]. BCFtools joined the VCF file from the same species, and SVs with an allele frequency ratio of $\geqslant 0.4$ were retained as high-frequency SVs[30]. Lumpy (version 0.3.1) was used for SV calling based on GRCg6a with default parameters[31].

2.4 Identification of structural variations overlapping specific genomic promoter regions

The upstream 3000 bp areas of genes was regarded promoter regions in chicken according to a previous study[32]. The locations of exon and intron were obtained from annotation file of GRCg6a. The coverage of exons/introns/promoters and filtered SV was obtained through the BEDTools intersect with parameter "-wa -wb," which accepts two bed files and outputs the overlapping areas of them[33]. DAVID and KOBAS were used for the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, respectively[34–36]. The species parameter was set to "chicken" in both website tools to ensure the right background genes. P-values were calculated for the results of GO and KEGG enrichment analysis and adjusted p-value calculation was also performed.

2.5 RNA-seq data analysis

RNA-seq data was from an earlier study, including midbrain and cerebral cortex tissues from both Leghorn (from lowland contexts) and Tibetan chicken (from the Tibetan highland), with six biological replicates for each group. RNA-seq reads were mapped to reference genome GRCg6a by HISAT2 (version 2.2.1) with default parameters[37]. RNA-seq read counting and expression level calculation were performed by StringTie (version 1.3.3) with parameters '-e -B'[38]. Differentially expressed genes were identified by the R package "DEseq2"[39]. The output in ballgown format from StringTie was transformed into read count format, which the DEseq2 package can directly identify. DEseq2 uses generalized linear models to estimate the significance of differentially expressed genes and multiple comparisons are made with adjusted p-values.

3 Results

3.1 Construction of the chicken graph-based pan-genome

We constructed a chicken graph-based pan-genome based on the Minigraph-Cactus pipeline. This pan-genome used linear reference genomes GRCg6a (red jungle fowl) as the reference, consisting of 11 genomes from two commercial breeds (White Leghorn and Rhode Island Red) and nine native breeds (Silkie, Tibetan, Asil, Cornish, Houdan, Fayoumi, Huxu, Naked Neck and Thailand game fowl) (Fig. 1(a)). We categorized the variants in the graph into four classes: SNP, indels with a length of < 50 bp, short SVs with a length of 50–1000 bp, and long SVs with a length of > 1000 bp. The pan-genome encompassed 43,710,347 SNPs, 9,139,049 indels with average length 4.16 bp (median length of 3 bp), 317,533 short SVs with an average length of 241 bp and 55,211 long SVs with an average length of 4292 bp (Fig. 1(b)). The majority of these variants were overlapped with intergenic and intron regions, accounting for 80.5%, 80.7%, and 74.5% of the indels, short SVs, and long SVs, respectively (Fig. 1(c)). The graph comprised a total of 4691 paths (haplotypes), 532,22,140 nodes (variants), and 73,118,282 edges (encode strandedness) within the graph-based pan-genome (Fig. 1(d)). On average, there were 476 nodes and 654 edges per 10 kb window.

3.2 Evaluation of the chicken graph-based pan-genome using genome next-generation sequencing data

To evaluate our graph-based pan-genome, we mapped the high-depth genome NGS data (> 40 times depth) to graph-based pan-genome and linear genome GRCg6a, respectively. The high-depth genome NGS data came from the same individual used for genome assembly. We compared the mapping ratio of the graph-based method to the GRCg6a linear genome-based method. The graph-based method had higher alignment ratios (median of > 98%) than that of GRCg6a using the BWA aligner (Fig. 2(a)). We also compared the variants called from NGS data to the variants contained in the

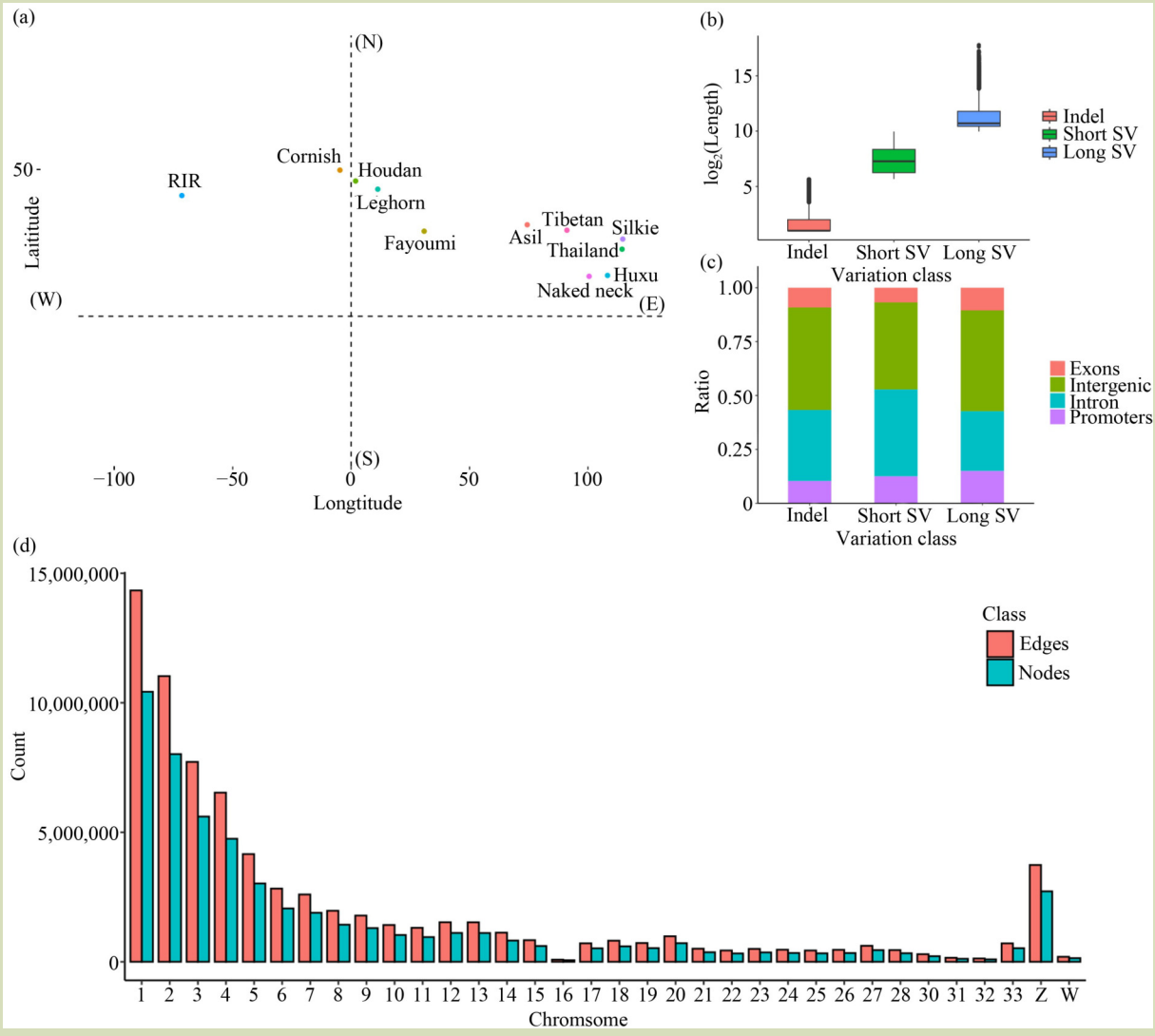


Fig. 1 Graph pan-genome construction and statistics: (a) chicken breeds used in the construction of the graph-based pan-genome; (b) statistics of length for indels and SVs; (c) statistics of overlapping regions for indels and SVs; and (d) number of edges and nodes in the graph.

pan-genome (Fig. 2(b)). The high calling ratio between the two sets of variants validated the sensitivity of graph-based SV calling. Also, we evaluated the graph-based mapping ratio using low-depth NGS data from four chicken breeds (Silkie, White Leghorn, Rhode Island Red and Tibetan chickens with 7–15 times depth). The results demonstrated that the mapping ratio of the VG pipeline for low-depth NGS data was also high, with a median value of > 98.8% in all four population (Fig. 2(c)). To further compare the graph-based and the linear method, we called SVs using genome NGS data with linear genome SV caller Lumpy. Lumpy called 3246, 3690, 3024, and 4916 SVs for Leghorn, Rhode Island Red, Silkie and Tibetan chickens, respectively. In comparison, the graph-based method reported 9944, 8725, 8010, and 11,970 SVs indicating significantly increasing numbers of identified SVs. This highlights the enhanced efficiency of SV calling using the graph-based pan-genome. The graph-based mapping exhibited a slightly lower alignment ratio than the HISAT2 tool (Fig. 2(d)). More complex processes in RNA-seq alignment may restrict the performance of graph-based mapping and further improvements in the tools are needed to address this issue.

3.3 Graph-based method identified the structural variations potentially related to egg production in Leghorn chicken

We then investigated SVs in four chicken populations, including two commercial breeds (White Leghorn for egg

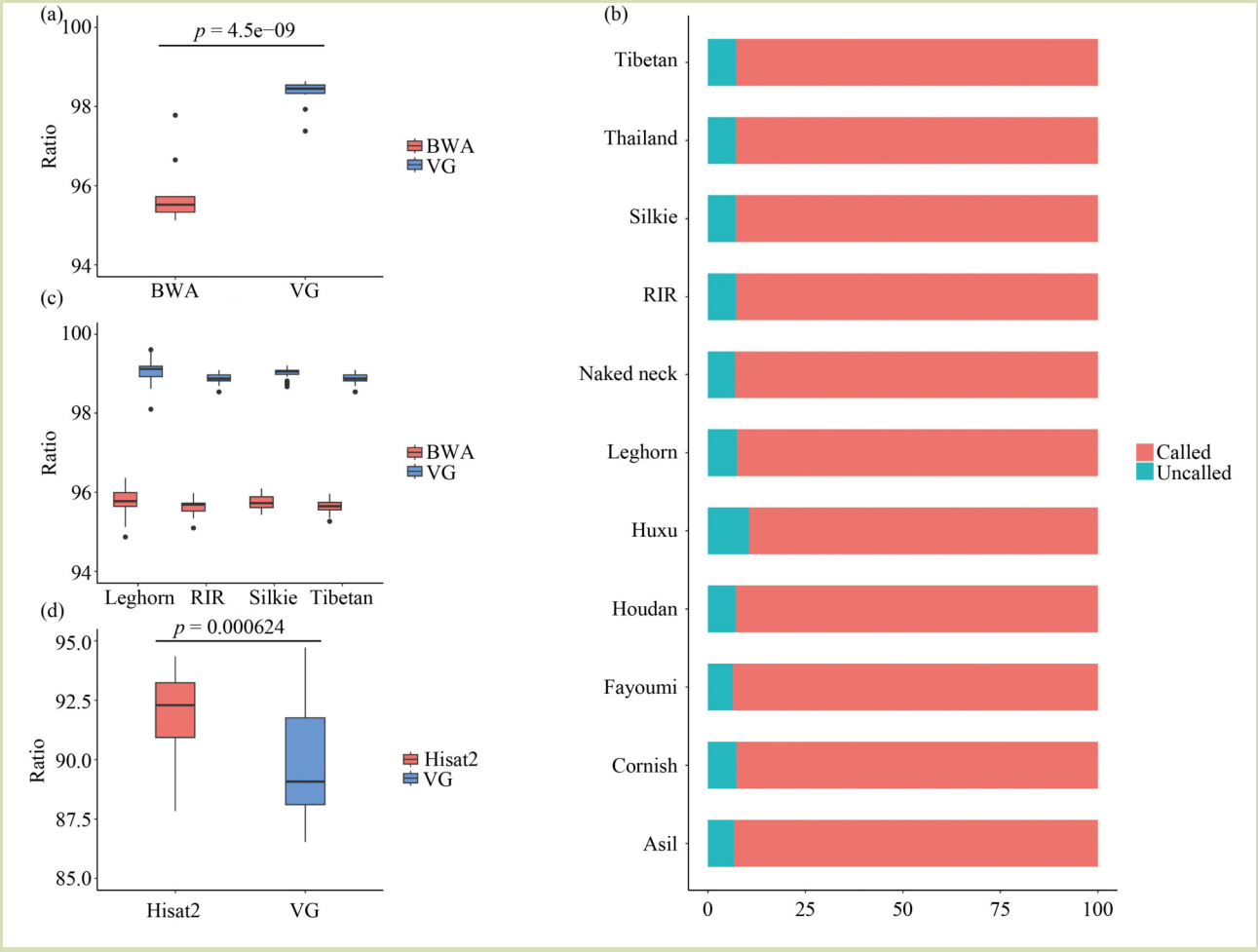


Fig. 2 Evaluation of the chicken graph pan-genome: (a) comparison of the alignment ratio of high-depth sequencing data based on the linear genome and graph-based pan-genome; (b) calling ratio (the number of called SVs/the number of total SVs in graph for each breed) for variants in the graph by VG; (c) alignment ratio of low-depth genome NGS data from four populations of chicken; and (d) comparison of the alignment ratio of RNA-seq data based on the linear genome and graph-based pan-genome.

production and Rhode Island Red for dual purposes) and two native breeds (Silkie and Tibetan). Using genome NGS data from White Leghorn (n = 46), Rhode Island Red (n = 21), Silkie (n = 27), and Tibetan (n = 24) chicken, we mapped the data to our graph-based pan-genome and identified high-frequency SVs (allele frequency of ⩾ 0.4). This effort found 1197, 1059, 918, and 1504 high-frequency SVs in these four populations, respectively. On the basis of high-frequency SV information, we identified 666 SVs unique to Leghorn chicken (Fig. 3(a)) with 52.4% being insertions and 47.6% deletions (Fig. 3(b)). Also, 20.7%, 69.1%, and 10.2% of these Leghorn specific SVs were located in the exon, intron and promoter regions, respectively (Fig. 3(c)).

Although most unique SVs are located in intron regions, it is still necessary to focus on these SVs considering probable intron retention. We first calculated the expression levels of 343 genes with SV overlapping introns using RNA-seq data from ovary granulosa and theca cells (small yellow follicles 6 mm, F1 or F5 stages) (Fig. 4(a)). 218 of these genes were expressed (FPKM > 1) in at least one stage, and 17 of them had expression signals on their SV overlapping introns (coverage of > 3 reads). This finding indicates that SVs in intron regions can also potentially alter the mature transcript sequence. Notably, we identified a 661 bp insertion (chromosome 4: 65, 106, and 861) on *CLOCK* (clock circadian regulator) intron region, which forms a heterodimer with *BMAL1* (basic helix–loop–helix ARNT-like 1). An earlier study showed this heterodimer activated the transcription STAR gene in the goose ovary and has a probable influence on progesterone synthesis[40]. We also found a 61 bp deletion (chromosome 6:

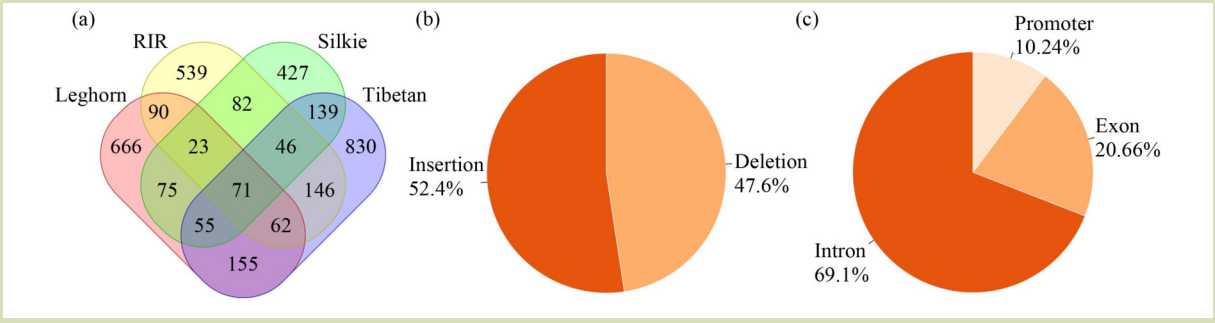


Fig. 3 Analysis of genes overlapped by unique SVs in Leghorn chicken: (a) Venn diagram for high-frequency SVs in four chicken populations; (b) ratio for insertions and deletions of Leghorn specific SVs; and (c) ratio of Leghorn specific SVs overlapping genomic regions.

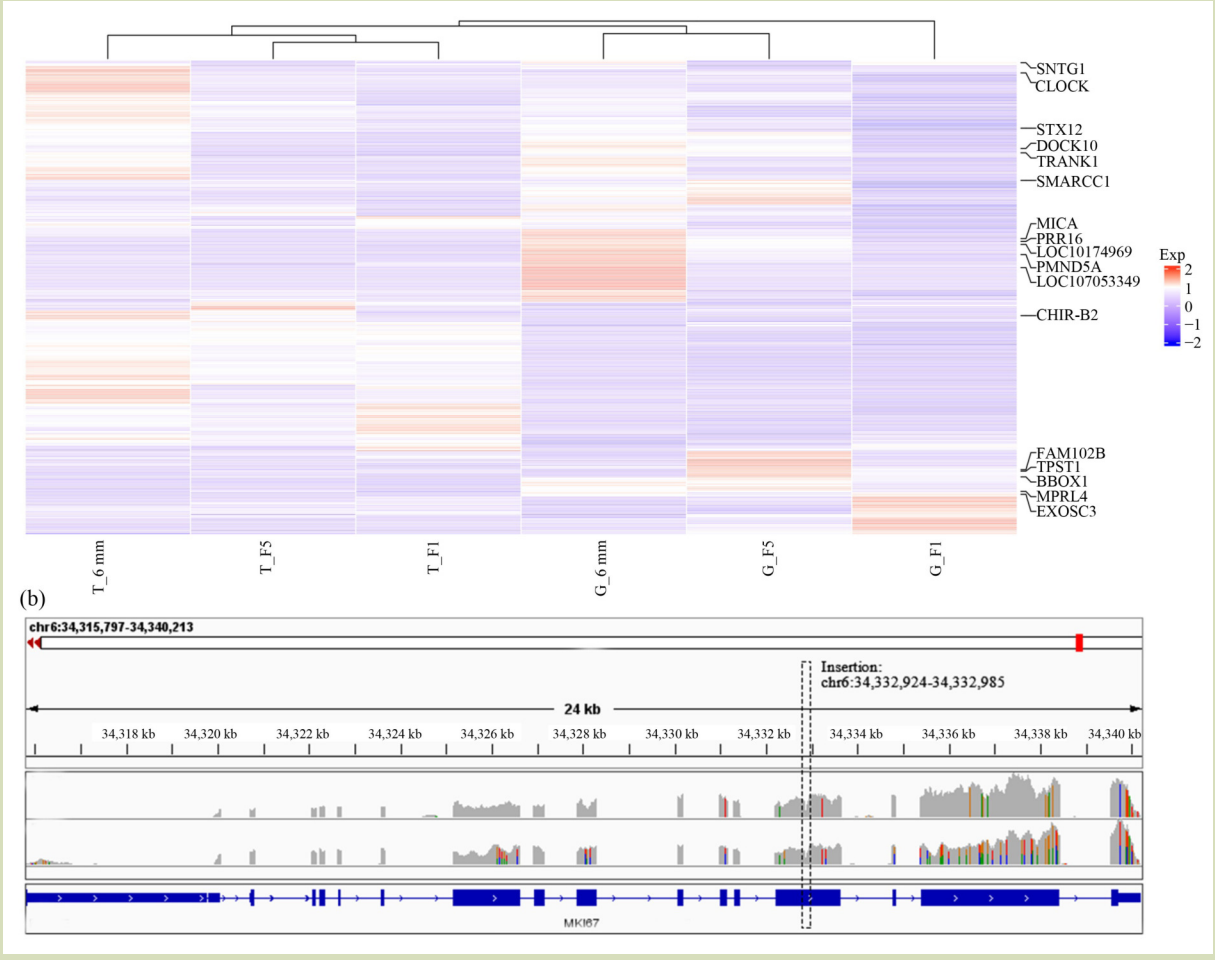


Fig. 4 Identification of SVs in reproduction-related genes in Leghorn chicken: (a) heatmap of expression levels of genes with introns covered by Leghorn specific SVs; and (b) loci of 61 bp insertion on *MKI67* exon and expression signal of *MKI67*.

34, 332, 924–985) on the exon region of *MKI67* (marker of proliferation Ki-67) (Fig. 4(b)). *MKI67* serves as a marker for follicular proliferation[41]. In summary, we identified intronic/exonic region SVs in reproduction-related genes in Leghorn chickens, which require further validation to determine their contribution in reproduction.

3.4 Graph-based method screened unique SVs related to environmental adaptation in Leghorn and Tibetan chickens

To further investigate the potential influence of SVs on the function of nervous tissues, we combined the analysis of SVs and transcriptomics data in White Leghorn and Tibetan chickens. White Leghorn and Tibetan chickens represent chickens produced in quite distinct environments, with the former being primarily produced on modern farms and the latter being raised in household farming in highlands. We conducted the transcriptomic analysis using RNA-seq data from the midbrain and cerebral cortex tissues in these two chickens. This effort separately found 619 and 453 significantly differentially expressed genes (DEGs) in the midbrain and cerebral cortex (p-value < 0.01, |log2FC| > 1). Among the DEGs in the midbrain, 248 were upregulated, and 371 were downregulated in Tibetan chicken (Fig. 5(a)). The top five

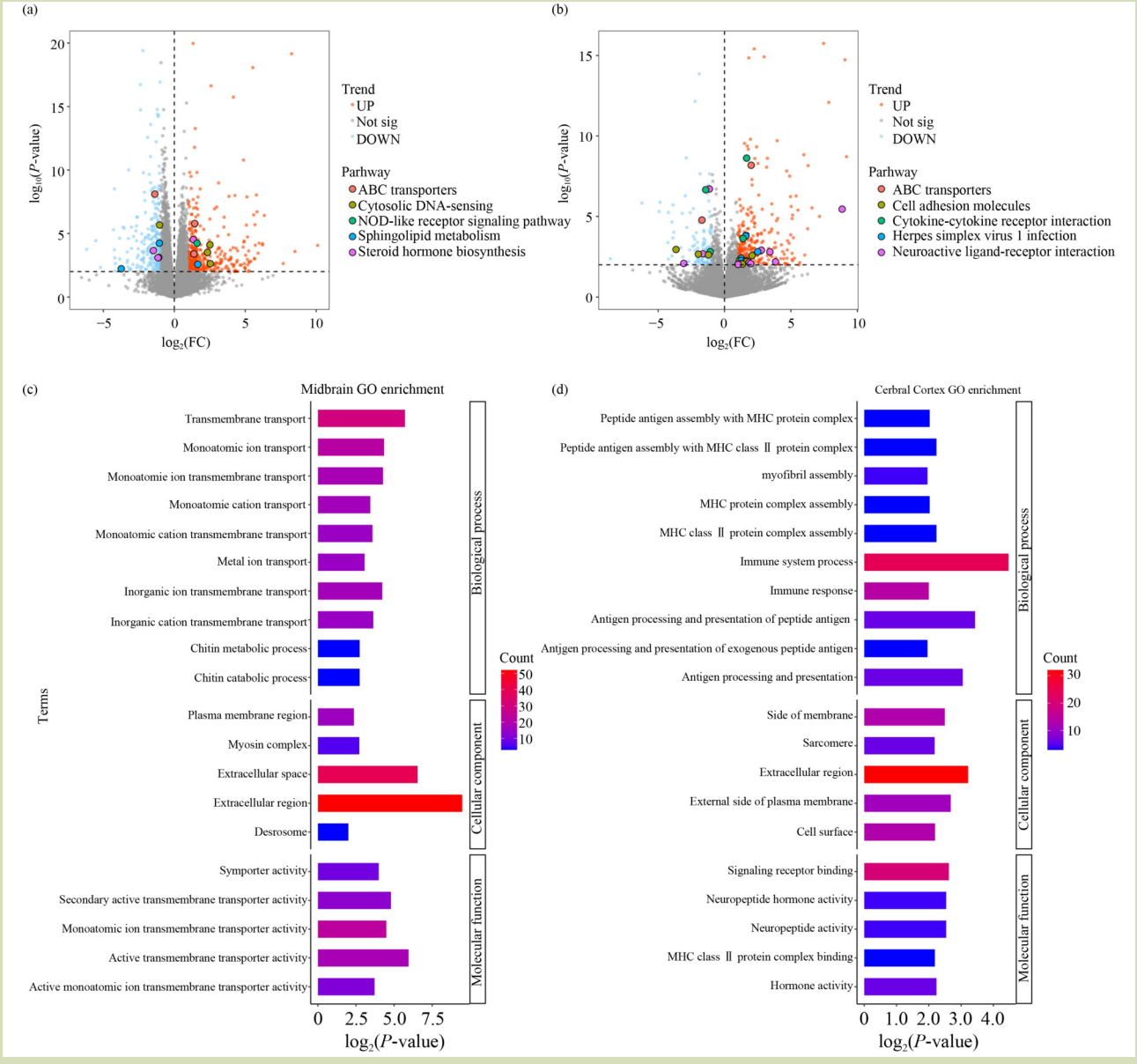


Fig. 5 GO and KEGG analysis for DEGs in midbrain and cerebral cortex tissues: (a) KEGG volcano plot for differentially expressed genes in midbrain between highland-living Tibetan chicken and lowland-living Leghorn chicken; (b) KEGG volcano plot for differentially expressed genes in cerebral cortex highland-living Tibetan chicken and lowland-living Leghorn chicken; (c) GO enrichment for DEGs in the midbrain; and (d) GO enrichment for DEGs in the cerebral cortex.

enriched pathways included ABC transporter, cytosolic DNA-sensing pathway, NOD-like receptor signaling pathway, sphingolipid metabolism, and steroid hormone biosynthesis. In the cerebral cortex, 123 DEGs were upregulated and 330 were downregulated in Tibetan chicken (Fig. 5(b)). Among these DEGs, previous studies have shown that *SULT2B1L1* (sulfotransferase family cytosolic 2B member 1-like) and *HSD11B1b* (hydroxysteroid 11-beta dehydrogenase 1B) are involved in steroid hormone biosynthesis. These genes have homologous members that function in the hormonal regulation of the human endometrium and conceptus elongation in sheep[42,43]. *ACE* (angiotensin I converting enzyme) was also identified, which is associated with improved performance at high altitudes in humans[44]. The presence of these known genes in the DEGs validates the reliability of the transcriptomic data and the pipeline used for identifying the DEGs. GO enrichment analysis also revealed that DEGs in the midbrain are predominantly associated with functions related to ion transport, which plays a vital role in cellular homeostasis, nerve impulse transmission, and muscle function. DEGs in the cerebral cortex were enriched in functions related to the function of major histocompatibility complex, the key member of adaptive immunity. These analyses at the transcriptome level revealed the differences in biological processes between Tibetan chickens from high altitudes and Leghorn chickens from the plains.

We further identified DEGs with promoters overlapped by breed-specific high-frequency SVs in and Leghorn and Tibetan chickens. We found eight DEGs with promoters overlapped by high-frequency unique SVs, including *LOC107051754*, *LOC101747932*, *MRPS24* (mitochondrial ribosomal protein S24), *RAPGEF3* (Rap guanine nucleotide exchange factor 3), *LOC107051637*, *WDR78* (WD repeat domain 78), and *SYTL1* (synaptotagmin-like 1). One notable midbrain DEG is *MRPS24*, which is involved in protein synthesis within the mitochondrion. The presence of the 94 bp insertion in the promoter region may potentially affect the expression level of *MRPS24* and the activity of mitochondria, which could be an adaptation to the hypoxic environment in Tibetan chicken (Fig. 6(a)). Another important gene is *SYTL1*, which has a 260 bp insertion in the promoter region of Leghorn chicken (Fig. 6(b)). *SYTL1*, also known as *JFC1* in humans and *exophilin7* in mice, is involved in the mediation of undocked granule fusion and plays a role in the exocytosis of insulin granules[45]. This insertion in Leghorn chicken may represent an adaptive response to long-term human domestication.

4 Discussion

Our study has shown the capacity of graph-based pan-genome to overcome the limitations of SV detection using genome NGS data. It has provided a comprehensive representation for

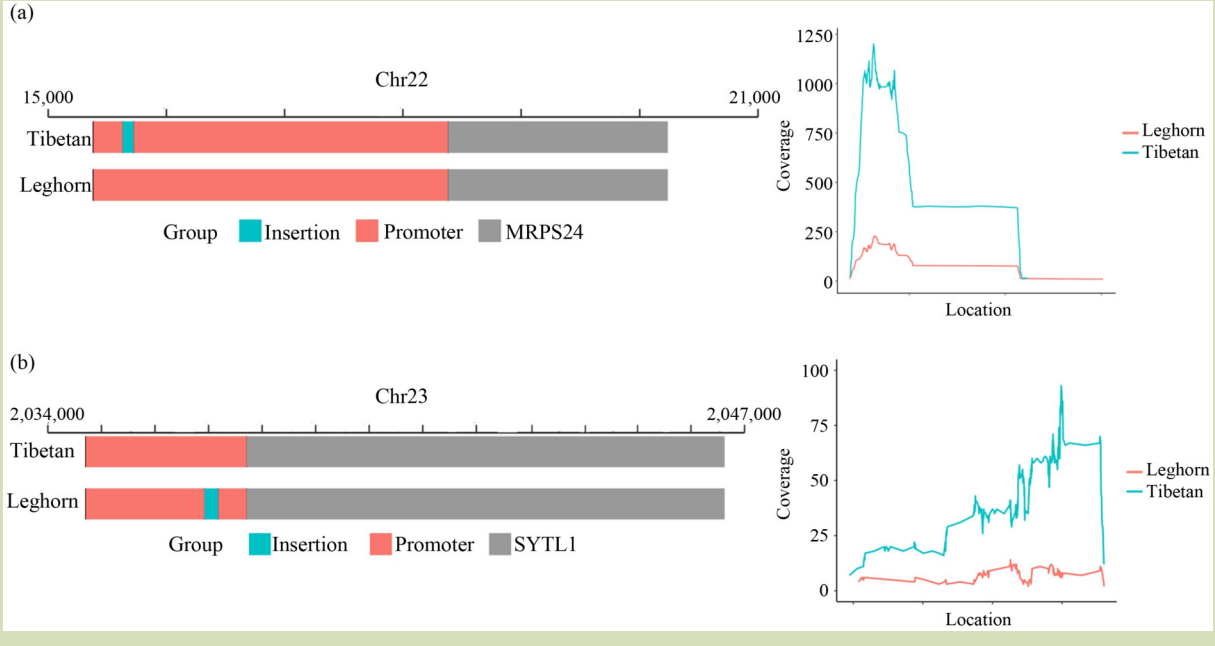


Fig. 6 Predicted functional SVs and expression signal of exon regions; (a) insertion in the promoter of *MRPS24* and the expression signal of *MRPS24* exon regions; and (b) insertion in the promoter of *SYTL1* and expression signal of *SYTL1* exon regions.

chicken and allows for accurate SV detection using genome NGS data. Previous studies performed SV calling using genome NGS data with linear genome callers, such as Delly and Lumpy software[31,46]. However, low representativeness of the reference genome and difficulties in long SV calling are two main challenges for this strategy[13,14,16]. In our study, graph-based pan-genome showed robust representativeness and performed better in the alignment of genome NGS reads and the identification of SVs compared to the linear reference.

However, there are still challenges for graph-based pan-genome that need to be addressed. One challenge is the lower mapping ratio of RNA-seq data. A slightly lower mapping ratio was observed using the graph-based pan-genome compared to the linear genome with the aligner HISAT2[47]. The lower mapping ratio is due to the complexity of the RNA-seq alignment process, which requires considering both the linear reference genome and the variants stored in the graph. Another challenge is the lack of analysis tools specifically designed for graph-based pan-genomes. While several tools have been developed, they may not be fully adapted to the evolving versions of the pan-genome graph[48,49]. The construction method of the pan-genome graph itself is constantly being updated, which can pose challenges for existing analysis tools. Additionally, the quality and completeness of the genomes used for constructing the pan-genome graph can impact its effectiveness. In the case of chicken genomes, the lack of complete assemblies for small chromosomes (chromosomes 29 and 34–39 in chicken) can limit the accuracy and completeness of the graph-based pan-genome. Efforts should be made to improve the assembly quality and coverage of these smaller chromosomes to enhance the overall quality of the pan-genome graph.

Indeed, the use of graph-based pan-genomes in population studies can provide a more efficient and cost-effective approach for identifying and analyzing SVs in agricultural animal populations. Graph-based pan-genomes offer a promising alternative to statistical methods, which are important for studies on avian functional SVs[32,50,51]. Our study presents a series of analyses of SVs from different chicken breeds using a graph-based pan-genome. These examples provide a new pipeline for chicken population research. This pipeline requires only long-read sequencing data from a small number of individuals. Then, we assemble the genome with long reads at the contig level, join and correct the assembly by ragtag and construct a graph-based pan-genome[52]. Finally, VG accurately screened out high-confidence SVs with low-cost genome NGS data.

Overall, despite these challenges, the use of graph-based pan-genomes holds great promise for studying genetic variations in chicken populations. With further advancements in analysis tools and enhanced genome assemblies, the graph-based approach can provide a more comprehensive understanding of genetic variations in chickens.

5 Conclusions

Our study constructed a graph-based pan-genome for chicken and demonstrated its effectiveness in identifying SVs in different chicken breeds using genome NGS data. Through comprehensive analyses, we gained insights into the genetic variations underlying high egg production in Leghorn chicken and highland adaptation in Tibetan chicken. The integration of transcriptomic data further highlighted genes that may contribute to substantively to specific traits. The future challenge will be validating the functions of SVs associated with traits of interest and improving the graph-based pan-genome quality and completeness.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2023YFF1001000).

Compliance with ethics guidelines

Yiming Wang, Zijia Ni, and Yinhua Huang declare that they have no conflicts of interest or financial conflicts to disclose. This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Wang M S, Thakur M, Peng M S, Jiang Y, Frantz L A F, Li M, Zhang J J, Wang S, Peters J, Otecko N O, Suwannapoom C, Guo X, Zheng Z Q, Esmailizadeh A, Hirimuthugoda N Y, Ashari H, Suladari S, Zein M S A, Kusza S, Sohrabi S, Kharrati-Koopaee H, Shen Q K, Zeng L, Yang M M, Wu Y J, Yang X Y, Lu X M, Jia X Z, Nie Q H, Lamont S J, Lasagna E, Ceccobelli S, Gunwardana H G T N, Senasige T M, Feng S H, Si J F, Zhang H, Jin J Q, Li M L, Liu Y H, Chen H M, Ma C, Dai S S, Bhuiyan A K F H, Khan M S, Silva G L L P, Le T T, Mwai O A, Ibrahim M N M, Supple M, Shapiro B, Hanotte O, Zhang G J, Larson G, Han J L, Wu D D, Zhang Y P. 863 genomes reveal the origin and domestication of chicken. *Cell Research*, 2020, **30**(8): 693–701
2. Pollock S L, Stephen C, Skuridina N, Kosatsky T. Raising chickens in city backyards: the public health role. *Journal of Community Health*, 2012, **37**(3): 734–742
3. Jaturasitha S, Srikanchai T, Kreuzer M, Wicke M. Differences in carcass and meat characteristics between chicken indigenous to northern Thailand (Black-Boned and Thai native) and imported extensive breeds (Bresse and Rhode Island Red). *Poultry Science*, 2008, **87**(1): 160–169
4. Wattanachant S, Benjakul S, Ledward D A. Composition, color, and texture of Thai indigenous and broiler chicken muscles. *Poultry Science*, 2004, **83**(1): 123–128
5. Jaturasitha S, Chaiwang N, Kreuzer M. Thai native chicken meat: an option to meet the demands for specific meat quality by certain groups of consumers: a review. *Animal Production Science*, 2017, **57**(8): 1582–1587
6. Guan R F, Lyu F, Chen X Q, Ma J Q, Jiang H, Xiao C G. Meat quality traits of four Chinese indigenous chicken breeds and one commercial broiler stock. *Journal of Zhejiang University. Science. B.*, 2013, **14**(10): 896–902
7. Han D P, Tai Y R, Hua G Y, Yang X, Chen J F, Li J Y, Deng X. Melanocytes in black-boned chicken have immune contribution under infectious bursal disease virus infection. *Poultry Science*, 2021, **100**(12): 101498
8. Tai Y R, Yang X, Han D P, Xu Z H, Cai G X, Hao J Q, Zhang B, Deng X. Transcriptomic diversification of granulosa cells during follicular development between White Leghorn and Silky Fowl hens. *Frontiers in Genetics*, 2022, **13**: 965414
9. Nan J, Yang S, Zhang X, Leng T, Zhuoma J, Zhuoma R, Yuan J, Pi J, Sheng Z, Li S. Identification of candidate genes related to highland adaptation from multiple Chinese local chicken breeds by whole genome sequencing analysis. *Animal Genetics*, 2023, **54**(1): 55–67
10. Li K, Dan Z, Gesang L, Wang H, Zhou Y, Du Y, Ren Y, Shi Y, Nie Y. Comparative analysis of gut microbiota of native Tibetan and Han populations living at different altitudes. *PLoS One*, 2016, **11**(5): e0155863
11. Desta T T. The genetic basis and robustness of naked neck mutation in chicken. *Tropical Animal Health and Production*, 2021, **53**(1): 95
12. Fernandes E, Raymundo A, Martins L L, Lordelo M, de Almeida A M. The naked neck gene in the domestic chicken: a genetic strategy to mitigate the impact of heat stress in poultry production—A review. *Animals*, 2023, **13**(6): 1007
13. Ballouz S, Dobin A, Gillis J A. Is it time to change the reference genome. *Genome Biology*, 2019, **20**(1): 159
14. Merker J D, Wenger A M, Sneddon T, Grove M, Zappala Z, Fresard L, Waggott D, Utiramerur S, Hou Y, Smith K S, Montgomery S B, Wheeler M, Buchan J G, Lambert C C, Eng K S, Hickey L, Korlach J, Ford J, Ashley E A. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genetics in Medicine*, 2018, **20**(1): 159–163
15. Xiao T, Zhou W. The third generation sequencing: the advanced approach to genetic diseases. *Translational Pediatrics*, 2020, **9**(2): 163–173
16. Yang X F, Lee W P, Ye K, Lee C. One reference genome is not enough. *Genome Biology*, 2019, **20**(1): 104
17. Sherman R M, Salzberg S L. Pan-genomics in the human genome era. *Nature Reviews. Genetics*, 2020, **21**(4): 243–254
18. Liu Y C, Du H L, Li P C, Shen Y T, Peng H, Liu S L, Zhou G A, Zhang H, Liu Z, Shi M, Huang X, Li Y, Zhang M, Wang Z, Zhu B, Han B, Liang C, Tian Z. Pan-genome of wild and cultivated soybeans. *Cell*, 2020, **182**(1): 162–176.E13
19. He Q, Tang S, Zhi H, Chen J F, Zhang J, Liang H K, Alam O, Li H, Zhang H, Xing L, Li X, Zhang W, Wang H, Shi J, Du H, Wu H, Wang L, Yang P, Xing L, Yan H, Song Z, Liu J, Wang H, Tian X, Qiao Z, Feng G, Guo R, Zhu W, Ren Y, Hao H, Li M, Zhang A, Guo E, Yan F, Li Q, Liu Y, Tian B, Zhao X, Jia R, Feng B, Zhang J, Wei J, Lai J, Jia G, Purugganan M, Diao X. A graph-based genome and pan-genome variation of the model plant Setaria. *Nature Genetics*, 2023, **55**(7): 1232–1242
20. Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G. De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nature Genetics*, 2012, **44**(2): 226–232
21. Li M, Sun C J, Xu N Y, Bian P P, Tian X M, Wang X H, Wang Y, Jia X, Heller R, Wang M, Wang F, Dai X, Luo R, Guo Y, Wang X, Yang P, Hu D, Liu Z, Fu W, Zhang S, Li X, Wen C, Lan F, Siddiki A Z, Suwannapoom C, Zhao X, Nie Q, Hu X, Jiang Y, Yang N. De novo assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on microchromosomes and subtelomeric regions. *Molecular Biology and Evolution*, 2022, **39**(4): msac066
22. Huang Z, Xu Z X, Bai H, Huang Y J, Kang N, Ding X T, Liu J, Luo H, Yang C, Chen W, Guo Q, Xue L, Zhang X, Xu L, Chen M, Fu H, Chen Y, Yue Z, Fukagawa T, Liu S, Chang G, Xu L. Evolutionary analysis of a complete chicken genome. *Proceedings of the National Academy of Sciences of the United States of America*, 2023, **120**(8): e2216641120
23. Guo Y, Ou J H, Zan Y J, Wang Y Z, Li H F, Zhu C H, Chen K, Zhou X, Hu X, Carlborg Ö. Researching on the fine structure

and admixture of the worldwide chicken population reveal connections between populations and important events in breeding history. *Evolutionary Applications*, 2022, **15**(4): 553–564

24. Zhou J K, Chang Y, Li J Y, Bao H, Wu C. Integrating whole-genome resequencing and RNA requencing data reveals selective sweeps and differentially expressed genes related to nervous system changes in Luxi Gamecocks. *Genes*, 2023, **14**(3): 584
25. Huang Y, Luo W, Luo X, Wu X, Li J, Sun Y, Tang S, Cao J, Gong Y. Comparative analysis among different species reveals that the androgen receptor regulates chicken follicle selection through species-specific genes related to follicle development. *Frontiers in Genetics*, 2022, **12**: 752976
26. Armstrong J, Hickey G, Diekhans M, Fiddes I T, Novak A M, Deran A, Fang Q, Xie D, Feng S, Stiller J, Genereux D, Johnson J, Marinescu V D, Alföldi J, Harris R S, Lindblad-Toh K, Haussler D, Karlsson E, Jarvis E D, Zhang G, Paten B. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, 2020, **587**(7833): 246–251
27. Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics*, 2013, **29**(10): 1341–1342
28. Hickey G, Heller D, Monlong J, Sibbesen J A, Siren J, Eizenga J, Dawson E T, Garrison E, Novak A M, Paten B. Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biology*, 2020, **21**(1): 35
29. Sirén J, Monlong J, Chang X, Novak A M, Eizenga J M, Markello C, Sibbesen J A, Hickey G, Chang P C, Carroll A, Gupta N, Gabriel S, Blackwell T W, Ratan A, Taylor K D, Rich S S, Rotter J I, Haussler D, Garrison E, Paten B. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science*, 2021, **374**(6574): abg8871
30. Danecek P, Bonfield J K, Liddle J, Marshall J, Ohan V, Pollard M O, Whitwham A, Keane T, McCarthy S A, Davies R M, Li H. Twelve years of SAMtools and BCFtools. *GigaScience*, 2021, **10**(2): giab008
31. Layer R M, Chiang C, Quinlan A R, Hall I M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology*, 2014, **15**(6): R84
32. Wang K J, Hu H F, Tian Y D, Li J Y, Scheben A, Zhang C X, Li Y, Wu J, Yang L, Fan X, Sun G, Li D, Zhang Y, Han R, Jiang R, Huang H, Yan F, Wang Y, Li Z, Li G, Liu X, Li W, Edwards D, Kang X. The chicken pan-genome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. *Molecular Biology and Evolution*, 2021, **38**(11): 5066–5081
33. Quinlan A R, Hall I M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 2010, **26**(6): 841–842
34. Bu D C, Luo H T, Huo P P, Wang Z H, Zhang S, He Z H, Wu Y, Zhao L, Liu J, Guo J, Fang S, Cao W, Yi L, Zhao Y, Kong L. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Research*, 2021, **49**(W1): W317–W325
35. Huang D W, Sherman B T, Lempicki R A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 2009, **4**(1): 44–57
36. Sherman B T, Hao M, Qiu J, Jiao X, Baseler M W, Lane H C, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, 2022, **50**(W1): W216–W221
37. Kim D, Paggi J M, Park C, Bennett C, Salzberg S L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 2019, **37**(8): 907–915
38. Pertea M, Pertea G M, Antonescu C M, Chang T C, Mendell J T, Salzberg S L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 2015, **33**(3): 290–295
39. Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 2014, **15**(12): 550
40. Chen R, Qin Y, Du J, Liu J, Dai S, Lei M, Zhu H. Circadian clock gene *BMAL1* regulates *STAR* expression in goose ovarian preovulatory granulosa cells. *Poultry Science*, 2023, **102**(12): 103159
41. Amorim C A, Dolmans M M, David A, Jaeger J, Vanacker J, Camboni A, Donnez J, Van Langendonckt A. Vitrification and xenografting of human ovarian tissue. *Fertility and Sterility*, 2012, **98**(5): 1291–1298
42. Brooks K, Burns G, Spencer T E. Biological roles of hydroxysteroid (11-Beta) dehydrogenase 1 (HSD11B1), HSD11B2, and glucocorticoid receptor (NR3C1) in sheep conceptus elongation. *Biology of Reproduction*, 2015, **93**(2): 38
43. Koizumi M, Momoeda M, Hiroi H, Hosokawa Y, Tsutsumi R, Osuga Y, Yano T, Taketani Y. Expression and regulation of cholesterol sulfotransferase (SULT2B1b) in human endometrium. *Fertility and Sterility*, 2010, **93**(5): 1538–1544
44. Bigham A W, Kiyamu M, Leon-Velarde F, Parra E J, Rivera-Ch M, Shriver M D, Brutsaert T D. Angiotensin-converting enzyme genotype and arterial oxygen saturation at high altitude in Peruvian Quechua. *High Altitude Medicine & Biology*, 2008, **9**(2): 167–178
45. Wang H, Ishizaki R, Xu J, Kasai K, Kobayashi E, Gomi H, Izumi T. The Rab27a effector exophilin7 promotes fusion of secretory granules that have not been docked to the plasma membrane. *Molecular Biology of the Cell*, 2013, **24**(3): 319–330
46. Rausch T, Zichner T, Schlattl A, Stütz A M, Benes V, Korbel J O. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 2012, **28**(18): i333–i339
47. Li H B, Wang S H, Chai S, Yang Z Q, Zhang Q Q, Xin H J, Xu Y, Lin S, Chen X, Yao Z, Yang Q, Fei Z, Huang S, Zhang Z. Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nature Communications*, 2022, **13**(1): 682
48. Monsu M, Comin M. Fast alignment of reads to a variation

graph with application to SNP detection. *Journal of Integrative Bioinformatics*, 2021, **18**(4): 20210032

49. Eggertsson H P, Kristmundsdottir S, Beyter D, Jonsson H, Skuladottir A, Hardarson M T, Gudbjartsson D F, Stefansson K, Halldorsson B V, Melsted P. GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. *Nature Communications*, 2019, **10**(1): 5402
50. Torgasheva A A, Malinovskaya L P, Zadesenets K S, Karamysheva T V, Kizilova E A, Akberdina E A, Pristyazhnyuk I E, Shnaider E P, Volodkina V A, Saifitdinova A F, Galkina S A, Larkin D M, Rubtsov N B, Borodin P M. Germline-restricted chromosome (GRC) is widespread among songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, **116**(24): 11845–11850
51. Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf J B W, Backstrom N, Künstner A, Balakrishnan C N, Heger A, Ponting C P, Clayton D F, Ellegren H. Molecular evolution of genes in avian genomes. *Genome Biology*, 2010, **11**(6): R68
52. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman Z B, Schatz M C, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*, 2022, **23**(1): 258