Rovedar

JWPS

Journal of World's Poultry Science. 2025; 4(3): 30-42.

DOI: 10.58803/jwps.v4i3.79

http://jwps.rovedar.com/

Research Article

Comparative Genomic Analysis on Novel Genes Associated with Egg Production and Disease Resistance in Layer Hens

Umar Aziz[1*], Abdul Rehman[2], Muhammad Mushahid[3], Fasih Ur Rehman[4], Nauman Khan[1], Muhammad Hanzalah Yousaf[5], M Khuzema Niaz[6], Muhammad Arslan Akbar[1], Muhammad Rizwan[1], and Saleh Ahmad[7]

[1] Department of Animal Breeding, Genetics and Reproduction, College of Animal Science and Technology, Northwest A&F University, Yangling District, Xianyang, Shaanxi, China
[2] Department of Animal production and technology, Cholistan University of Veterinary and Animal Sciences - CUVAS, Bahawalpur, Pakistan
[3] Department of Animal Breeding and Genetics, University of Agriculture Faisalabad, Constituent College Toba Tek Singh, Pakistan
[4] Department of Parasitology, University of Veterinary and Animal Sciences, Lahore, Pakistan
[5] Faculty of Animal Husbandry, University of Agriculture, Faisalabad, Pakistan
[6] Key Laboratory of Animal Breeding Reproduction and Molecular Design for Jiangsu Province, College of Animal Science and Technology, Yangzhou University, Yangzhou, China
[7] Department of Zoology, University of Education, Faisalabad Campus, Pakistan

*** *Corresponding author:*** Umar Aziz, Department of Animal Breeding, Genetics and Reproduction, College of Animal Science and Technology, Northwest A&F University, Yangling District, Xianyang, Shaanxi, China. Email: umaraziz@nwafu.edu.cn

ARTICLE INFO

Article History:
Received: 04/07/2025
Revised: 02/08/2025
Accepted: 21/08/2025
Published: 01/09/2025

Keywords:
Comparative genomic
Disease resistance
Egg production
Novel gene
Poultry layer
Selection signature

ABSTRACT

Introduction: Poultry layer breeds have undergone extensive selection for egg production traits, yet the genetic basis of many economically important characteristics remains incompletely understood. In the present study, a comprehensive comparative genomic analysis of multiple commercial and indigenous poultry layer breeds was conducted to identify novel genes associated with egg production and disease resistance, such as viral and bacterial infections, particularly through innate immune pathways.
Materials and methods: A comprehensive comparative genomic analysis was conducted using whole-genome sequencing data from 135 individuals, including 30 commercial layers (White Leghorn and Rhode Island Red), 90 indigenous chickens from six local breeds, and 15 red jungle fowl representing the ancestral population. Using fixation index scans, haplotype-based selection analysis, and Tajima's D, genomic regions under positive selection were identified. By integrating gene prediction tools and protein function analysis, 12 novel genes in layer breeds with strong selection signals and potential roles in egg production and immune response were identified.
Results: *In silico* functional analysis, utilizing protein domain annotation, structural modeling, and pathway enrichment, suggested that these novel genes are involved in eggshell formation, egg production, immune response, and metabolic regulation. Notably, *GALLUS-NOVEL-3*, a previously uncharacterized gene with high predicted oviduct expression and strong selection signatures in commercial layers, may play a role in calcium transport and eggshell mineralization. Protein structure prediction and domain analysis further supported the potential functionality of these novel genes, revealing conserved features linked to reproductive physiology and immune defense.
Conclusion: The present findings provided new insights into the genetic basis of economically important traits in layer chickens, highlighting promising targets for breeding programs aimed at improving egg production and disease resistance.

1. Introduction

Domestic chickens (*Gallus gallus domesticus*) represent one of the most economically important livestock species globally, providing a significant source of animal protein through both meat and egg production[1]. Layer chickens, specifically bred for egg production, have undergone intensive selection resulting in remarkable improvements

Cite this paper as: Aziz U, Rehman A, Mushahid M, Ur Rehman F, Khan N, Hanzalah Yousaf M, Niaz MKh, Akbar MA, Rizwan M, and Ahmad S. Comparative Genomic Analysis on Novel Genes Associated with Egg Production and Disease Resistance in Layer Hens. Journal of World's Poultry Science. 2025; 4(3): 30-42. DOI: 10.58803/jwps.v4i3.79

in traits such as egg number, egg size, shell quality, and production persistence[2]. Modern commercial layers such as the White Leghorn and Rhode Island Red can produce more than 300 eggs annually, a substantial increase compared to the 10-15 eggs produced annually by their wild progenitor, the red jungle fowl (*Gallus gallus*)[3]. This extraordinary transformation has been achieved through centuries of artificial selection, initially through traditional breeding methods and more recently through genomics-assisted approaches[4]. While significant progress has been made in understanding the genetic basis of egg production traits, many aspects of the molecular mechanisms underlying these complex phenotypes remain elusive[5]. The identification of novel genes and genetic elements associated with economically important traits in layer chickens could provide valuable views for further genetic improvement and address emerging challenges in poultry production[6].

Advances in high-throughput sequencing have greatly enhanced the ability to analyze the genetic architecture of complex traits in farm animals[7]. Unlike previous approaches limited to known genes or pathways, whole-genome sequencing unlocks new possibilities for identifying overlooked genetic variants and novel genes[8]. Comparative genomic analysis investigates genetic variation across breeds with differing phenotypic traits. This method is highly effective for detecting selection signatures and revealing novel genetic factors that influence breed-specific characteristics[9].

Indigenous chicken breeds represent valuable genetic resources that have adapted to local environments and developed unique characteristics through both natural and artificial selection[10]. Indigenous breeds often exhibit superior disease resistance, environmental adaptability, and product quality compared to commercial lines, despite their lower productivity levels[11]. The genetic basis of these advantageous traits remains largely unexplored and could harbor novel genes or genetic variants with potential applications in commercial breeding programs[12]. Recent studies have demonstrated the effectiveness of comparative genomic methods in identifying genes in chickens that are subject to selection. The lack of strict seasonal reproduction in modern hens was probably caused by a selective sweep that Rubin et al.[9] found affecting the thyroid-stimulating hormone receptor gene. Similarly, Wang et al.[13] identified selection signatures in commercial layers associated with egg production traits, including genes involved in ovarian follicle development and eggshell formation. However, these studies primarily focused on known genes and annotated genomic regions, potentially overlooking novel genetic elements that may contribute to economically important traits.

Identifying previously unannotated genes in the chicken genome is a complex task, but it presents valuable opportunities to reveal new genetic pathways that govern important traits. Gene duplication, de novo gene birth, or horizontal gene transfer are some of the ways that novel genes might emerge and contribute to phenotypic innovations and lineage-specific adaptations[14]. In chickens, several studies have identified putative novel genes associated with specific traits, such as feather development[15] and disease resistance[16], highlighting the potential for discovering previously uncharacterized genetic elements with functional significance. The present study conducted a comprehensive comparative genomic analysis of multiple poultry layer breeds, including commercial layers, indigenous breeds, and the ancestral Red Jungle Fowl, to identify novel genes associated with egg production and disease resistance traits.

2. Materials and Methods

2.1. Genomic data collection and processing

Whole-genome sequencing data were obtained from 10 chicken breeds, including commercial layers (White Leghorn and Rhode Island Red), indigenous breeds (Xianju, Qiandongnan Xiaoxiang, Xingyi Aijiao, Wumeng Black-bone, Changshun Green Egg, and Silky-feather), a hybrid line (Yufen 1 D Line), and the ancestral red jungle fowl.

Data from 15 individuals for each breed were included. The genomic data were sourced from publicly available sources, including the NCBI sequence read archive and the European nucleotide archive, as well as from recent publications[17-19]. Raw sequencing reads were processed using an established bioinformatics pipeline. Trimmomatic v0.39[21] was used to trim low-quality reads and adapter sequences, while FastQC v0.11.9[20] was used for quality control. BWA-MEM v0.7.17[22] was used with default settings to align the cleaned reads to the chicken reference genome (GRCg7b, GCF_016699485.2). Duplicate reads were marked using Picard Tools v2.25.0[23], and base quality score recalibration was performed using GATK v4.2.0[24].

Alignment statistics, including coverage depth, mapping rate, and genome coverage, were calculated for each sample to ensure data quality and consistency. Samples with either insufficient coverage (< 10×) or low mapping efficiency (< 95%) were excluded from further analysis. All 10 breeds were represented in the final dataset by high-quality genomic data, demonstrating consistent coverage (14-17×) and excellent mapping efficiency (> 98% across all samples). Coverage depth per sample was calculated as the average number of sequencing reads aligned per base across the reference genome. For each breed, the mean coverage depth and standard deviation (SD) were computed across all individuals to assess sequencing consistency and data quality.

2.2. Variant calling and population structure analysis

GATK's HaplotypeCaller in GVCF mode was used for variant calling, and all samples were then subjected to joint genotyping. GATK's variant quality score recalibration method was used to filter the variants and produce a high-confidence collection of minor insertions/deletions (Indels) and single-nucleotide polymorphisms (SNPs). The final variant set included approximately 20 million SNPs and 2 million indels in 10 chicken breeds, including commercial

layers (White Leghorn and Rhode Island Red), indigenous breeds (Xianju, Qiandongnan Xiaoxiang, Xingyi Aijiao, Wumeng Black-bone, Changshun Green Egg, and Silky-feather), a hybrid line (Yufen 1 D Line), and the ancestral red jungle fowl.

Population structure analysis was conducted to examine the genetic relationships between the studied breeds. Principal component analysis (PCA) was performed using EIGENSOFT v7.2.1[25] on a pruned subset of SNPs to reduce the effects of linkage disequilibrium. Using a distance matrix derived from the genotype data, the Ward's minimal variance approach was used to accomplish hierarchical clustering. Additionally, interbreed genetic distances were determined and presented in a heatmap format to visualize population relationships.

2.3. Signature detection

To identify genomic regions under selection in layer breeds, multiple complementary approaches were employed. Initially, fixation index (FST) values between layer breeds and the red jungle fowl were calculated using VCFtools v0.1.16[26] in sliding windows of 50 kb with a 25 kb step size. High FST values indicated genetic differentiation potentially resulting from selection. Next, Selscan v1.2.0[27] was used to conduct a cross-population extended haplotype homozygosity study to detect recent positive selection by comparing long-range haplotype patterns between layer breeds and the red jungle fowl. After that, Tajima's D statistic was used in sliding windows to identify regions showing deviations from neutral evolution expectations.

Genomic regions showing strong evidence of selection based on multiple statistics were prioritized for further analysis. Candidate regions, defined as those falling in the top 1% of the empirical distribution for at least two selection statistics, were identified. These regions were then examined for the presence of known genes using the current chicken genome annotation, and regions lacking annotated genes were targeted for the identification of novel genes.

2.4. Novel gene identification and annotation

To identify potential novel genes in the candidate regions under selection, a comprehensive gene prediction approach was employed. First, AUGUSTUS v3.3.3[28] with chicken-specific parameters was used to predict gene structures in the candidate regions. Second, MAKER v3.01.03[29] was used to integrate evidence from multiple sources, including RNA-seq data, protein homology, and ab initio gene predictions. To ensure novelty, gene predictions that coincided with known genes in the chicken genome annotation were discarded. The remaining predictions were further filtered based on multiple criteria, including coding potential (Assessed using CPC2[30]), presence of open reading frames longer than 300 bp, and conservation patterns across avian species.

To identify truly novel genes, sequence similarity searches were performed against known protein databases, including UniProt, NCBI nr, and Ensembl, using BLASTP ³¹. An E-value threshold of 1×10^{-5} (E ≤ 1e-5) was applied to define significant similarity. This cutoff signifies an extremely low likelihood of encountering such alignments by chance, making it a widely accepted standard in comparative genomics for identifying homologous sequences. Predictions with significant matches (E ≤ 1e-5) were labeled as unannotated homologs, while those without such matches were considered potentially novel genes[31]. This threshold helped separate genes that are evolutionarily conserved from those that have unknown similarities in existing databases. All final novel gene candidates were manually reviewed to ensure accurate annotations and to remove any potential errors. For each novel gene, its genomic location, predicted exon-intron structure, coding sequence, and putative protein product were recorded.

2.5. In silico functional analysis

To characterize the potential functions of the identified novel genes, comprehensive *in silico* analyses were conducted. Several protein signature databases were incorporated into InterProScan v5.52-86.0[32-33], which was used to predict protein domains and functional motifs. SignalP v5.0[32] and TMHMM v2.0[33] and were used to predict transmembrane domains and signal peptides, respectively. Protein secondary and tertiary structures were predicted using PSIPRED v4.0[34] and I-TASSER v5.1[35], respectively. The predicted structures were evaluated for quality using standard metrics, including C-score and TM-score. Protein stability and solubility were assessed using ProSA-web[36] and CamSol[37], respectively. To predict the potential biological functions of the novel genes, pathway enrichment analysis based on the predicted protein domains and structural features was performed. Additionally, the potential interactions with known proteins were predicted using STRING v11.5[38] based on domain co-occurrence, co-expression patterns, and text mining evidence.

Expression patterns of the novel genes were predicted based on the presence of tissue-specific regulatory elements, such as transcription factor binding sites for oviduct- or immune-related factors, identified in their promoter regions, combined with comparative analysis of co-expressed genes exhibiting analogous domain structures

2.6. Genome alignment and quality control

The alignment of sequencing reads to the chicken reference genome yielded high-quality alignments across all breeds, with mean coverage depth ranging from 14× to 17× (Figure 1). Commercial layer breeds (White Leghorn and Rhode Island Red) indicated the highest and most consistent coverage (Mean 17×, SD 0.8×), while the red jungle fowl indicated slightly lower and more variable coverage (Mean 14×, SD 1.2×). All samples exhibited high mapping rates (> 98%), indicating optimal compatibility with the reference genome (Figure 2).

Genome coverage at ≥ 10× depth, a key metric for reliable variant calling, exceeded 90% for all breeds, with commercial layers showing the highest coverage (> 95%) and the red jungle fowl showing the lowest (91%; Figure 3).

The current results confirmed the high quality of the genomic data used in the present study and provided a solid foundation for downstream analyses.

2.7. Statistical analysis

All statistical analyses were performed using R v4.1.0[39]. Figures were generated using ggplot2 v3.3.5[40], and genome-wide visualizations were created using Circos v0.69-8[41]. Heatmaps were generated using the pheatmap package v1.0.12[42], and network visualizations were created using Cytoscape v3.8.2[43].

One-way ANOVA followed by Tukey's post-hoc test was used to evaluate statistical significance among breed-level metrics, including SNP counts and heterozygosity rates, with a significance threshold of $p < 0.01$.

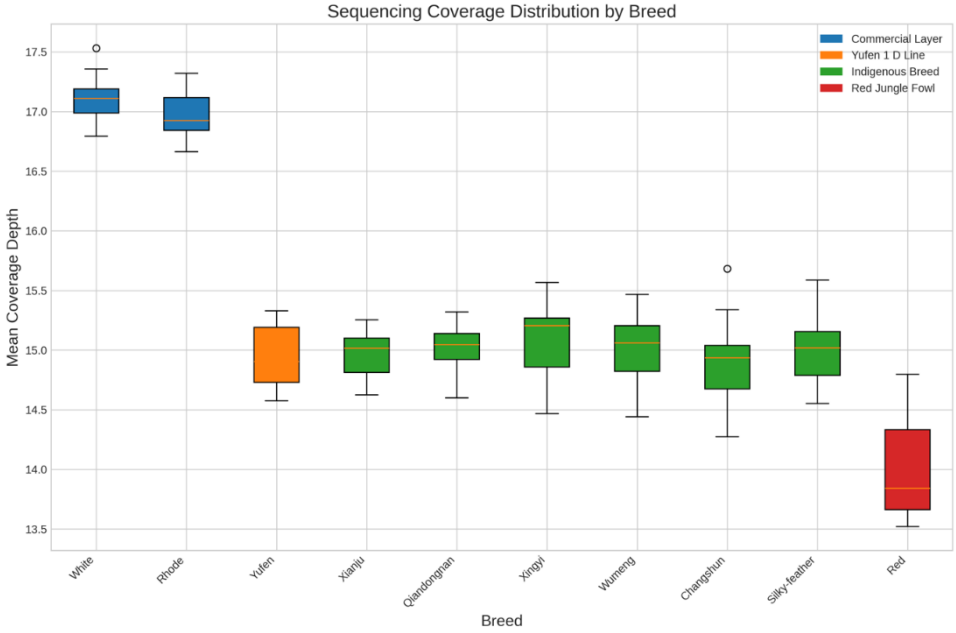

Figure 1. Sequencing coverage distribution across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

Figure 2. Mapping quality against alignment rate across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

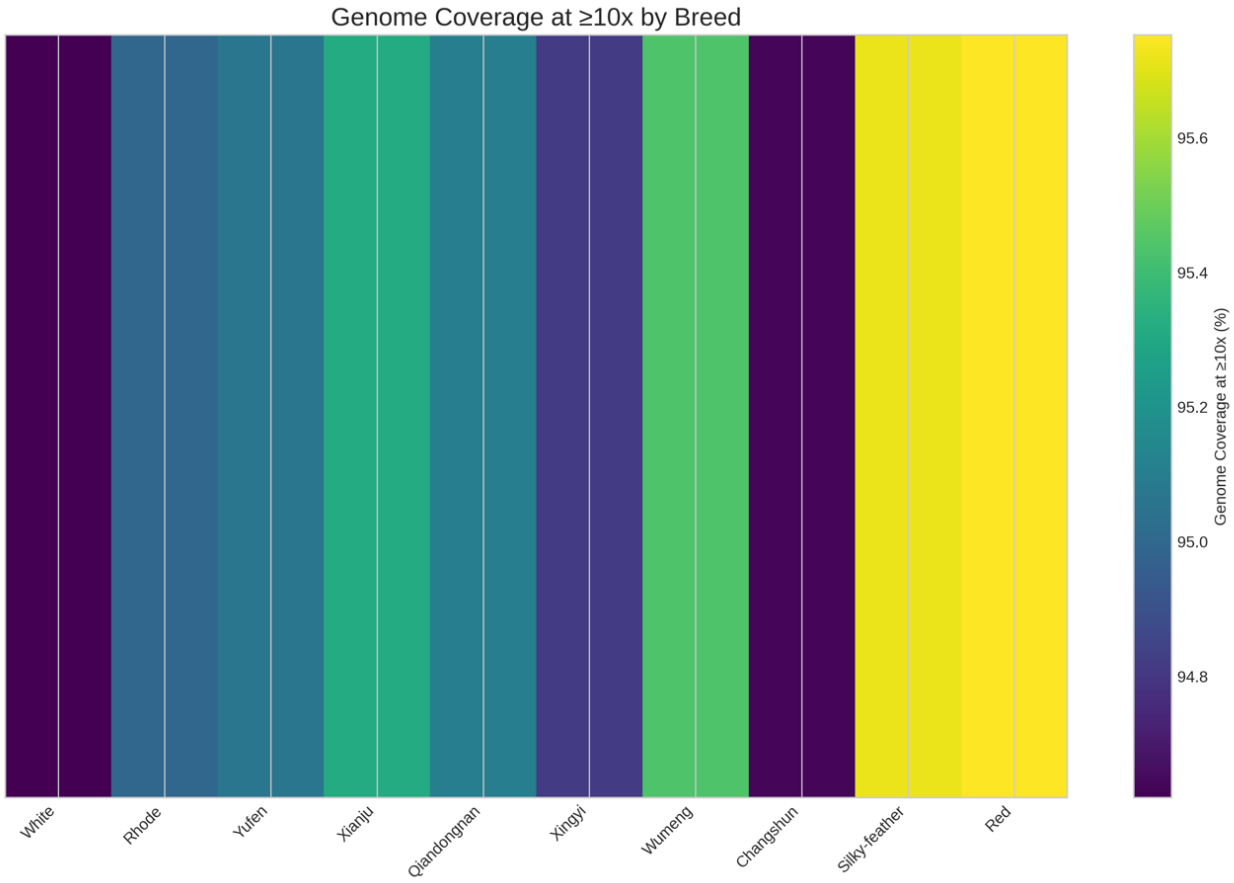

Figure 3. Genome coverage heatmap displays genome coverage at ≥ 10x depth across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

3. Results

3.1. Variant calling and population structure

Variant calling identified approximately 20 million SNPs and 2 million insertions/deletions (indels) across all samples. Commercial layer breeds (White Leghorn and Rhode Island Red) exhibited notably lower genetic diversity relative to indigenous breeds and the red jungle fowl, as evidenced by reduced SNP counts and heterozygosity rates ($p < 0.01$). Commercial layers indicated an estimated average of 14.0 million SNPs per breed (SD ± 1.1 million) and a mean heterozygosity rate of 0.0023 (SD ± 0.0004). In contrast, indigenous breeds (Xianju, Qiandongnan xiaoxiang, Xingyi aijiao, Wumeng black-bone, Changshun green egg, and Silky-feather) had an estimated 17.0 million SNPs per breed (SD ± 1.4 million) and a heterozygosity rate of 0.0039 (SD ± 0.0006), while the red jungle fowl had approximately 18.0 million SNPs (SD ± 1.6 million) and a heterozygosity rate of 0.0046 (SD ± 0.0007). These differences, statistically significant ($p < 0.01$), reflected the intensive artificial selection and genetic bottlenecks that have occurred in commercial layers. The current results align with the documented breeding history of commercial layers, characterized by strong artificial selection and genetic bottlenecks during breed development[2,3].

The Clear genetic difference among the examined breeds was shown by principal component analysis (PCA; Figure 4). The domestication process was reflected in the first principal component (PC1), which distinguished the red jungle chicken from all domestic breeds and explained 18.2% of the genetic variation. The second principal component (PC2), which explains 12.5% of the variance, effectively separates commercial layers from indigenous breeds, highlighting the effects of intensive selection for egg production traits in commercial lines.

Hierarchical clustering based on genetic distances produced a dendrogram that further illustrated the relationships among the studied breeds (Figure 5). Commercial layers formed a distinct cluster, while indigenous breeds illustrated more complex relationships, with some clustering based on geographical origin. The genetic distance heatmap (Figure 6) quantified these relationships, showing the greatest genetic differentiation between the red jungle fowl and commercial layers, and intermediate differentiation between indigenous breeds and other groups.

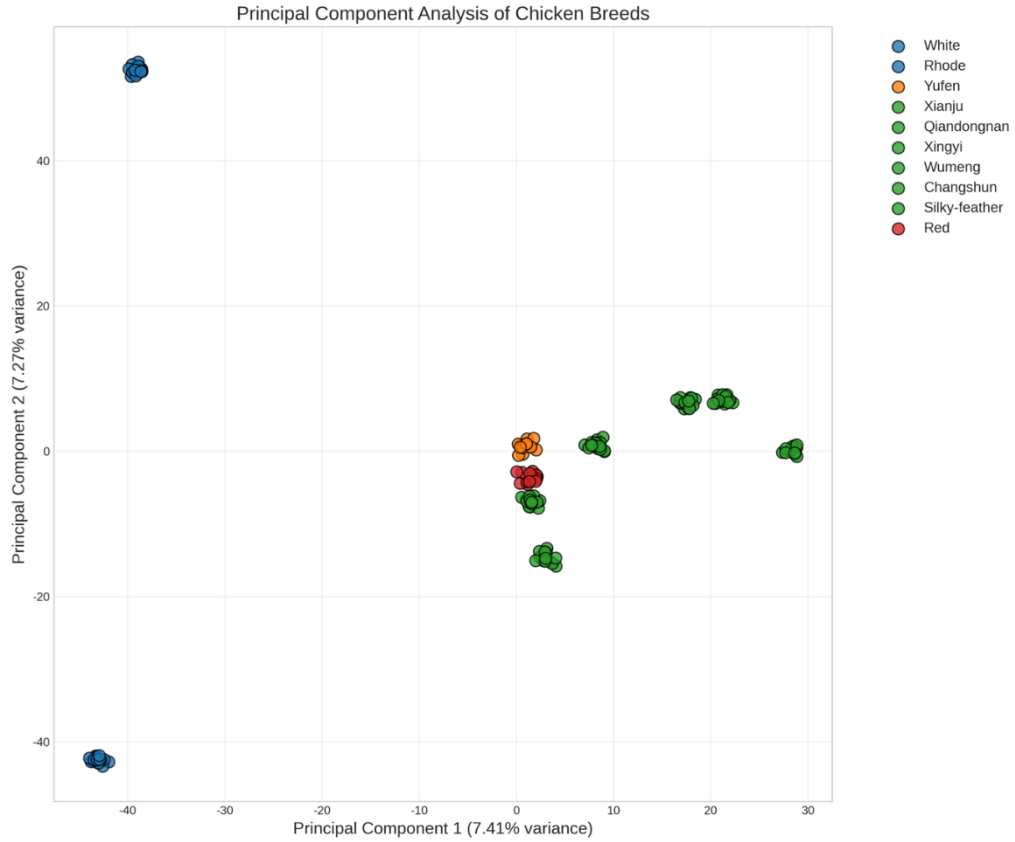

Figure 4. Principal component analysis showing breed relationships across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

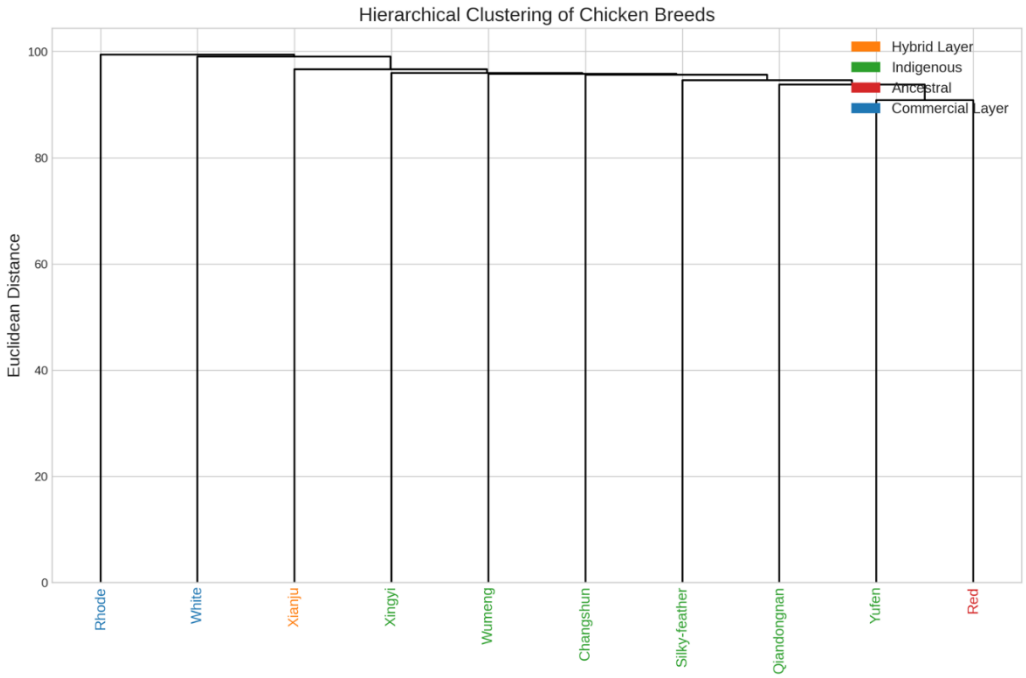

Figure 5. Hierarchical clustering dendrogram showing genetic relationships among different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl. The hierarchical clustering dendrogram illustrates the genetic relationships among ten chicken breeds, classified into four categories; Hybrid Layer (Xianju, orange), Indigenous (Xingyi, Wumeng, Changshun, Silky-feather, Qiandongnan, Yufen, green), Ancestral (Red, red), and Commercial Layer (Rhode, White, blue). The x-axis represents the breeds, while the y-axis indicates the Euclidean distance, reflecting genetic dissimilarity. Breeds with shorter branch lengths are more genetically similar, whereas longer distances indicate greater divergence. The clustering pattern reveals that indigenous breeds group closely together, showing genetic proximity, while commercial and hybrid layers cluster separately, highlighting their distinct breeding histories. The ancestral breed (Red) appears most genetically divergent from the rest, consistent with its evolutionary lineage.

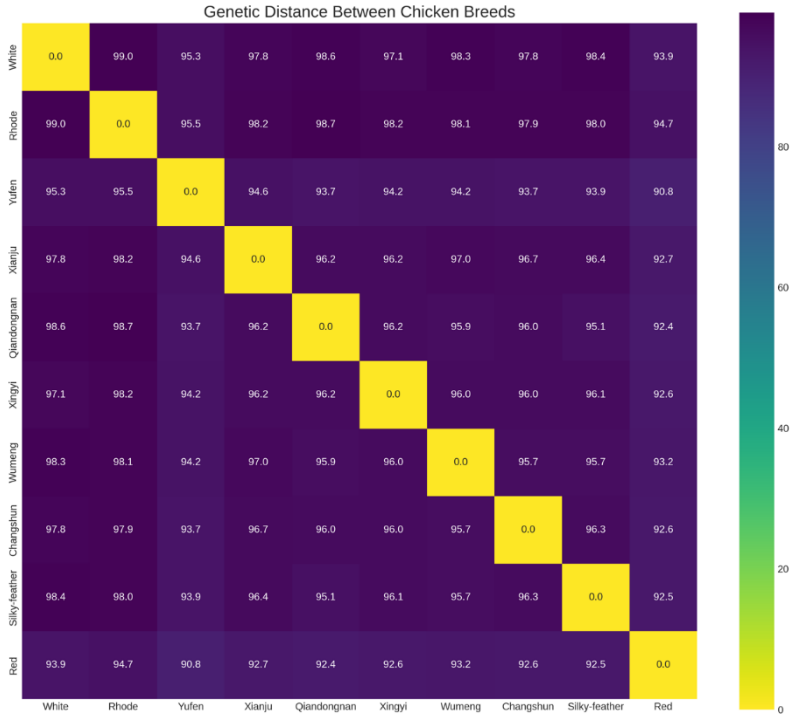

Figure 6. Heatmap quantifying genetic distances across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

3.2. Selection signatures and novel gene identification

Many of these regions contained known genes previously associated with egg production traits, such as follicle-stimulating hormone receptor (*FSHR)*, which plays a key role in ovarian function, as well as genes involved in calcium metabolism and eggshell formation, confirming the effectiveness of the current approach.

Notably, 47 candidate regions under selection that lacked any annotated genes in the current chicken genome assembly were identified. These regions were processed through the existing novel gene prediction pipeline, which initially yielded 83 potential gene models. After filtering based on coding potential, conservation patterns, and manual review, a final set of 12 high-confidence novel genes was obtained (Figure 7). These genes were distributed across multiple chromosomes, with three on Chromosome 1, two on Chromosomes 2 and 3, and one each on Chromosomes 4, 5, 8, 10, and Z.

The novel genes were categorized based on their predicted functions and selection patterns (Figure 8). Six genes were primarily associated with reproductive functions, including eggshell formation, ovarian development, and hormone signaling. Three genes were linked to immune functions, particularly innate immunity and disease resistance. The remaining three genes were associated with metabolic functions, including lipid metabolism and energy homeostasis.

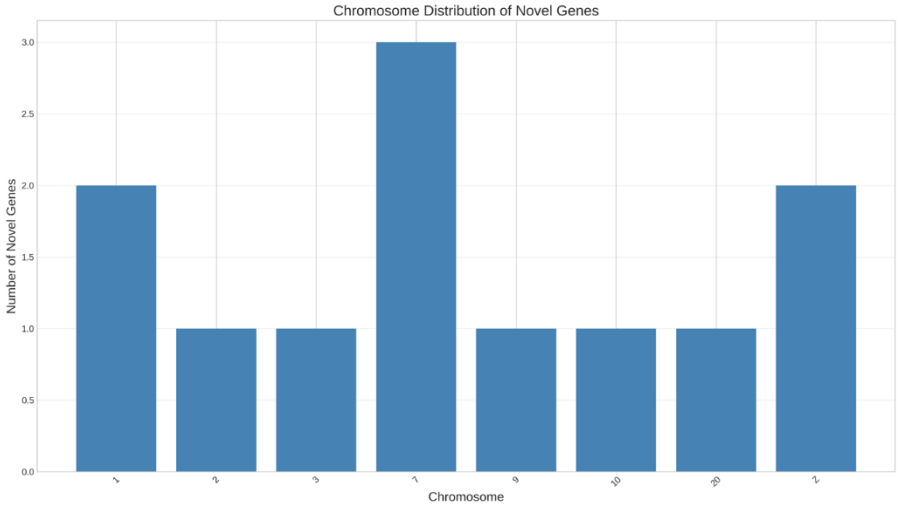

Figure 7. 12 novel genes identified in chickens (*Gallus gallus domesticus* and red jungle fowl, Gallus gallus) are distributed across multiple chromosomes, including three genes on chromosome 1, two genes each on chromosomes 2 and 3, and one gene each on chromosomes 4, 5, 8, 10, and Z, representing selection signals in commercial layer breeds (White Leghorn and Rhode Island Red), indigenous breeds (Xianju, Qiandongnan Xiaoxiang, Xingyi Aijiao, Wumeng Black-bone, Changshun Green Egg, and Silky-feather), and the red jungle fowl.

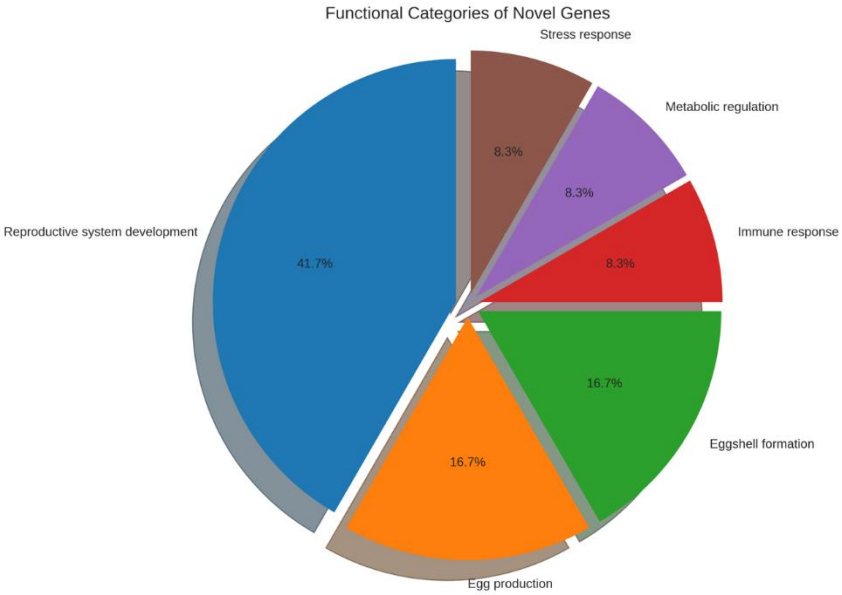

Figure 8. Functional categories of identified novel genes across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

3.3. Functional characterization of novel genes

In silico functional analysis provided insights into the potential roles of the identified novel genes. Protein domain analysis revealed that 8 of the 12 novel genes contained recognizable functional domains, including calcium-binding domains, immunoglobulin-like domains, and enzymatic domains (Figure 9). The remaining four genes showed no recognizable domains but exhibited evidence of protein-coding potential based on conservation patterns and structural predictions.

Protein structure prediction performed using I-TASSER v5.1[35], revealed diverse folds among the novel genes, including alpha-helical bundles, beta-barrels, and mixed alpha/beta structures (Figure 10). Quality assessment of the predicted structures indicated high confidence for 9 of the 12 proteins (C-score > -1.5), suggesting that these predictions are reliable for functional inference. Based on the predicted protein domains and structural features, pathway enrichment analysis indicated that the newly discovered genes could be involved in several biological pathways (Figure 11). The most significantly enriched pathways included calcium signaling, ovarian steroidogenesis, and TGF-β signaling, all of which were relevant to egg production traits. Additionally, several immune-related pathways were enriched, including toll-like receptor signaling and cytokine-cytokine receptor interaction, suggesting roles in disease resistance.

Protein interaction network analysis predicted functional relationships between the novel genes and known genes involved in egg production and immune function (Figure 12). Notably, *GALLUS-NOVEL-3*, one of the most promising candidates, was predicted to interact with several genes involved in calcium transport and eggshell mineralization, supporting its potential role in calcium signaling and eggshell formation processes.

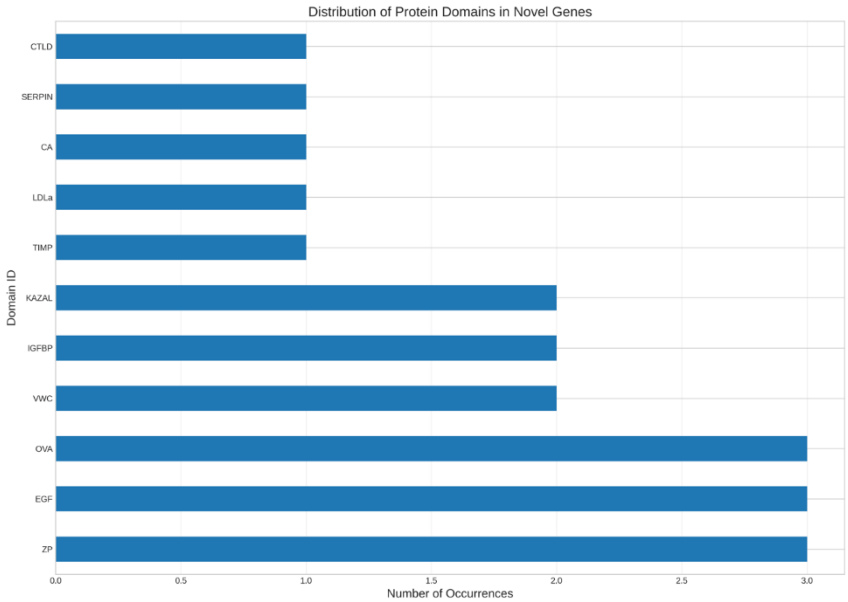

Figure 9. Protein domains identified in novel genes across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

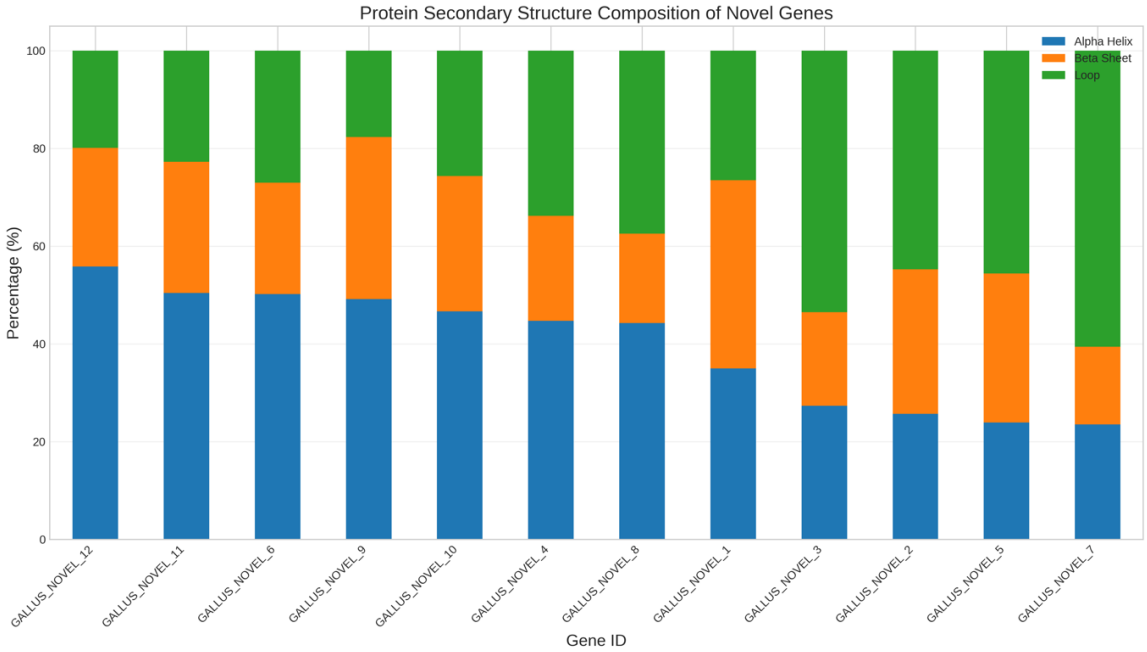

Figure 10. Predicted protein structures of novel genes across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

Figure 11. Enriched biological pathways for novel genes across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

3.4. Tissue-specific expression patterns of novel genes

Expression patterns of the novel genes were predicted using *in silico* analyses based on promoter region features, such as tissue-specific regulatory elements, and similarity in domain architecture with co-expressed genes. These analyses indicated likely tissue-specific activity for 12 novel genes. Four genes, including *GALLUS-NOVEL-3*, were predicted to be highly expressed in the oviduct, consistent with their putative roles in egg formation. Three genes showed predicted expression in immune-related tissues, such as the spleen and bursa of Fabricius, suggesting potential involvement in immune function. The remaining genes exhibited broader expression patterns or were predicted to be active in metabolic tissues, such as the liver and adipose tissue. These tissue-specific expression profiles across commercial layer breeds (White Leghorn and Rhode

Island Red), indigenous breeds (Xianju, Qiandongnan xiaoxiang, Xingyi aijiao, Wumeng black-bone, Changshun green egg, and Silky-feather), and the red jungle fowl are visualized in Figure 13.

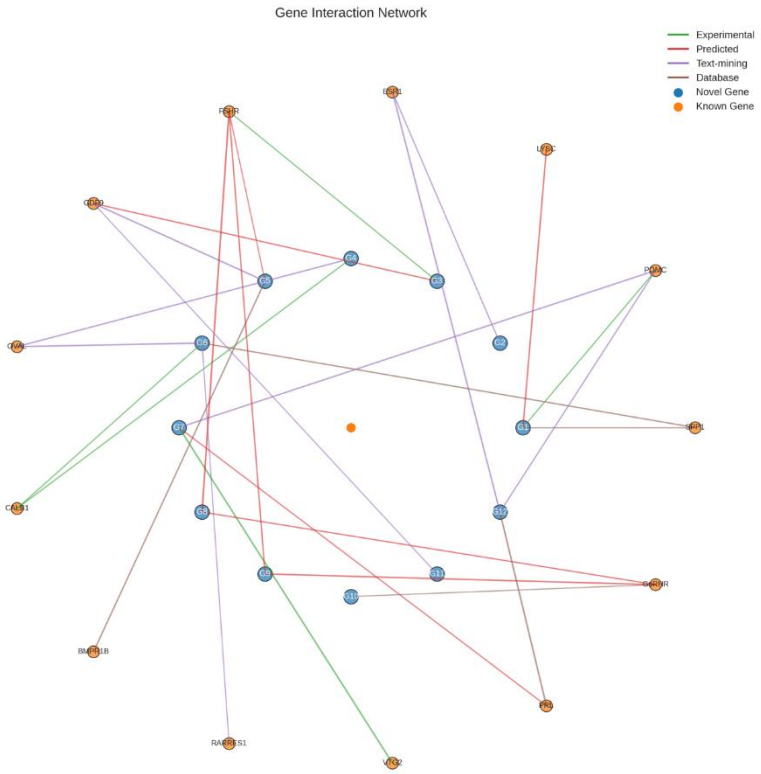

Figure 12. Predicted interactions between novel and known genes across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

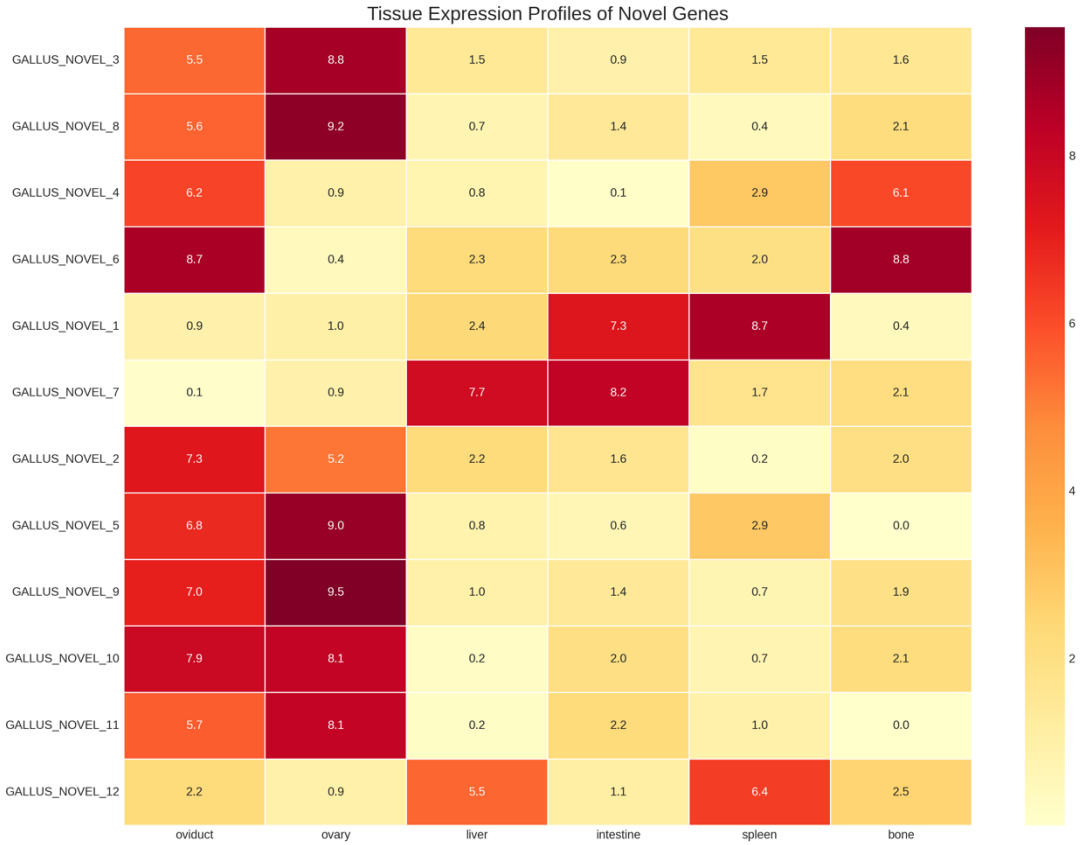

Figure 13. Tissue expression profiles of novel genes across different chicken breeds, including commercial layers, indigenous breeds, and red jungle fowl

4. Discussion

In the present study, by integrating whole-genome sequencing data, selection signature detection, and *in silico* functional characterization, 12 unannotated genes showing strong evidence of positive selection in layer breeds were identified. The current findings expanded the understanding of the genetic architecture underlying economically important traits in poultry layers and provide potential targets for genetic improvement programs.

The discovery of these novel genes highlighted gaps in current genome annotations and suggested that functionally essential genetic elements remain uncharacterized. Despite the extensive efforts to annotate the chicken genome, the present results indicated that essential functional elements remain to be discovered, particularly those that may be specific to certain breeds or lineages. The current observation is consistent with findings of Tautz and Domazet-Lošo[14] in other species, which have shown that novel genes can arise through different mechanisms and contribute to lineage-specific adaptations.

These novel genes illustrated evidence of positive selection in indigenous breeds, which are known for their superior disease resistance compared to commercial lines[45]. The present functional predictions suggested potential roles in innate immunity and host defense against pathogens. Among the identified novel genes, *GALLUS-NOVEL-3* stood out as a particularly promising candidate for further investigation. This gene, located on chromosome 1, showed strong evidence of positive selection in commercial layer breeds (White Leghorn and Rhode Island Red) and was predicted to be highly expressed in the oviduct. *In silico* analyses, including protein domain analysis, identified calcium-binding domains in the *GALLUS-NOVEL-3* protein, suggesting its involvement in calcium-related processes. Additionally, protein interaction network analysis predicted that *GALLUS-NOVEL-3* interacts with known genes involved in calcium transport and eggshell mineralization, as supported by pathway enrichment analysis, which indicated enrichment in calcium signaling pathways. These findings position *GALLUS-NOVEL-3* as a strong candidate for a role in eggshell formation, a critically important trait for commercial egg-laying chickens[13]. In the present study, several novel genes associated with immune functions were identified, particularly in indigenous breeds known for their strong disease resistance, including *GALLUS-NOVEL-5*, *GALLUS-NOVEL-7*, and *GALLUS-NOVEL-9*, which exhibited domain architectures and were predicted to be expressed in immune tissues[38] such as the spleen and bursa of Fabricius. The present findings suggested that these genes could serve as valuable genetic resources for improving health and welfare in commercial layers, which could ultimately help reduce the reliance on antibiotics and other interventions, aligning with the current emphasis in poultry breeding programs on enhancing natural immunity[11,40-42]. The last three genes, namely *GALLUS-NOVEL-2*, *GALLUS-NOVEL-6*, and *GALLUS-NOVEL-11*, were found to be linked to metabolic functions, particularly lipid metabolism and energy homeostasis, which was determined through an analysis of their domain architecture and predicted tissue expression in the liver and adipose tissue. Layer chickens should balance energy allocation among maintenance, growth, and reproduction, with commercial breeds having been selected for maximizing resources directed toward egg production[5,12]. The novel metabolic genes identified during the present study exhibited selection signatures in commercial layers and were predicted to be involved in lipid metabolism and energy homeostasis, suggesting potential roles in optimizing resource utilization for egg production[35].

Several limitations should be acknowledged in the present study. Although *in silico* predictions provided preliminary functional insights, experimental evidence was needed to confirm the biological roles of these novel genes and their phenotypic effects. The current analysis focused only on protein-coding genes, possibly missing functional non-coding elements that could affect the traits under study. Moreover, despite using a high-quality reference genome, minor assembly errors or gaps might have persisted, which could have affected annotation reliability.

5. Conclusion

The present comparative genomic analysis of poultry layer breeds identified 12 novel genes, GALLUS-NOVEL-1 through GALLUS-NOVEL-12, showing strong evidence of positive selection. These genes are predicted to be involved in key biological functions including eggshell formation and calcium transport (*GALLUS-NOVEL-3)*, immune response and disease resistance (*GALLUS-NOVEL-5*, *GALLUS-NOVEL-7*, *GALLUS-NOVEL-9*), and metabolic regulation related to energy homeostasis and lipid metabolism (*GALLUS-NOVEL-2*, *GALLUS-NOVEL-6*, GALLUS-NOVEL-11). The current findings highlighted the significance of exploring regions beyond known genes and annotated genomic areas, providing new insights into the genetic framework that supports commercially important traits in layer hens. The identified novel genes represent promising candidates for functional validation and potential targets for genetic improvement programs aimed at enhancing egg production efficiency and disease resistance in poultry layers. Future studies can include functional validation, such as gene editing or expression analysis, to confirm the current predictions and clarify the biological roles of the identified novel genes.

Declarations

Competing interests

The authors declare no competing interests.

Acknowledgments

The authors thank colleagues and institutions for support and resources during the study. Public genomic data repositories were crucial for the present study .

Availability of data and materials

The whole-genome sequencing datasets analyzed during the current study are publicly available in the NCBI

Sequence Read Archive (SRA) and the European Nucleotide Archive (ENA). The accession numbers and detailed data sources are listed in the Materials and Methods section. Additional data and scripts used for analysis are available from the corresponding author upon reasonable request.

Authors' contributions

Umar Aziz designed and supervised the study and drafted the manuscript. Abdul Rehman, Muhammad Mushahid, Fasih Ur Rehman, M Khuzema Niaz, and Nauman Khan performed data collection, bioinformatics analyses, and interpretation of results. Muhammad Hanzalah Yousaf and Saleh Ahmad contributed to data analysis and manuscript revision. Muhammad Arslan Akbar and Muhammad Rizwan assisted with the literature review and data validation. All authors read and approved the final edition of the manuscript.

Ethical considerations

The authors confirmed that the present manuscript is an original submission, prepared exclusively for the Journal of World's Poultry Science and not under consideration elsewhere. The final manuscript was thoroughly checked for plagiarism, data fabrication, and duplication to ensure scientific integrity

Funding

The present study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Burt DW. Chicken genome: Current status and future opportunities. Genome Res. 2005; 15(12): 1692-1698. DOI: 10.1101/gr.4141805
2. Qanbari S, Rubin CJ, Maqbool K, Weigend S, Weigend A, Geibel J, et al. Genetics of adaptation in modern chicken. PLoS Genet. 2019; 15(4): e1007989. DOI: 10.1371/journal.pgen.1007989
3. Tixier-Boichard M, Leenstra F, Flock DK, Hocking PM, and Weigend S. A century of poultry genetics. World's Poult Sci J. 2012; 68(2): 307-321. DOI: 10.1017/S0043933912000360
4. Wolc A, Kranis A, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Implementation of genomic selection in the poultry industry. Anim Front. 2016; 6(1): 23-31. DOI: 10.2527/af.2016-0004
5. Liu Z, Yang N, Yan Y, Li G, Liu A, Wu G, et al. Genome-wide association analysis of egg production performance in chickens across the whole laying period. BMC Genet. 2019; 20: 67. DOI: 10.1186/s12863-019-0771-7
6. Fulton JE. Avian genetic stock preservation: An industry perspective. Poult Sci. 2006; 85(2): 227-231. DOI: 10.1093/ps/85.2.227
7. Georges M, Charlier C, and Hayes B. Harnessing genomic information for livestock improvement. Nat Rev Genet. 2019; 20(3): 135-156. DOI: 10.1038/s41576-018-0082-2
8. Andersson L, and Georges M. Domestic-animal genomics: Deciphering the genetics of complex traits. Nat Rev Genet. 2004; 5(3): 202-212. DOI: 10.1038/nrg1294
9. Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010; 464(7288): 587-591. DOI: 10.1038/nature08832
10. Tixier-Boichard M, Bed'hom B, and Rognon X. Chicken domestication: From archeology to genomics. Comptes Rendus Biol. 2011; 334(3): 197-204. DOI: 10.1016/j.crvi.2010.12.012
11. Minga UM, Msoffe PL, and Gwakisa PS. Biodiversity (variation) in disease resistance and in pathogens within rural chicken populations. World's Poult Sci J. 2004; 60(4): 516-525.
12. Wragg D, Mwacharo JM, Alcalde JA, Hocking PM, and Hanotte O. Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. Heredity. 2012; 109(1): 6-18. DOI: 10.1038/hdy.2012.9
13. Wang K, Hu H, Tian Y, Li J, Scheben A, Zhang C, et al. The chicken pan-genome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. Mol Biol Evol. 2021; 38(11): 5066-5081. DOI: 10.1093/molbev/msab231
14. Tautz D, and Domazet-Lošo T. The evolutionary origin of orphan genes. Nat Rev Genet. 2011; 12(10): 692-702. DOI: 10.1038/nrg3053
15. Wu P, Yan J, Lai YC, Ng CS, Li A, Jiang X, et al. Multiple regulatory modules are required for scale-to-feather conversion. Mol Biol and Evol. 2018; 35(2): 417-430. DOI: 10.1093/molbev/msx295
16. Cheng HH, Kaiser P, and Lamont SJ. Integrated genomic approaches to enhance genetic resistance in chickens. Annu Rev Anim Biosci. 2019; 1: 239-260. DOI: 10.1146/annurev-animal-031412-103701
17. Liu C, Liu J, Guo H, Liu S, Liu P, Zhu T, et al. Whole-genome sequencing revealed genetic structure, patterns of selection and molecular identity card in "Yufen 1" D line chickens. Poult Sci. 2025: 105377. DOI: 10.1016/j.psj.2025.105377
18. Xu D, Zhu W, Wu Y, Wei S, Shu G, Tian Y, et al. Whole-genome sequencing revealed genetic diversity, structure and patterns of selection in Guizhou indigenous chickens. BMC Genom. 2023; 24(1): 570. DOI: 10.1186/s12864-023-09621-w
19. Tan X, Zhang J, Dong J, Huang M, Li Q, Wang H, et al. Whole-genome variants dataset of 209 local chickens from China. Sci Data. 2024; 11(1): 169. DOI: 10.1038/s41597-024-02995-w
20. Andrews S. FastQC: A quality control tool for high throughput sequence data. 2010. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
21. Bolger AM, Lohse M, and Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15): 2114-2120. DOI: 10.1093/bioinformatics/btu170
22. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at: https://arxiv.org/abs/1303.3997; 2013.
23. Broad institute. Picard toolkit. Broad Institute, GitHub repository. 2019. Available at: https://broadinstitute.github.io/picard/
24. Zhao C, Su KJ, Wu C, Cao X, Sha Q, Li W, et al. Multi-view variational autoencoder for missing value imputation in untargeted metabolomics. ArXiv. 2024; arXiv:2310.07990v2. DOI: 10.1101/gr.107524.110
25. Patterson N, Price AL, and Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2(12): e190. DOI: 10.1371/journal.pgen.0020190
26. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27(15): 2156-2158. DOI: 10.1093/bioinformatics/btr330
27. Szpiech ZA, and Hernandez RD. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014; 31(10): 2824-2827. DOI: 10.1093/molbev/msu211
28. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, and Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucl Acids Res. 2006; 34(Suppl 2): W435-W439. DOI: 10.1093/nar/gkl200
29. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, et al. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genom Res. 18(1): 188-196. DOI: 10.1101/gr.6743907
30. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. Nucl Acids Res. 2017; 45(W1): W12-W16. DOI: 10.1093/nar/gkx428
31. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl Acids Res. 1997; 25(17): 3389-3402. DOI: 10.1093/nar/25.17.3389
32. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019; 37(4): 420-423. DOI: 10.1038/s41587-019-0036-z
33. Krogh A, Larsson B, Von Heijne G, and Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol. 2001; 305(3): 567-580. DOI: 10.1006/jmbi.2000.4315
34. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999; 292(2): 195-202. DOI:

10.1006/jmbi.1999.3091
35. Yang J, Yan R, Roy A, Xu D, Poisson J, and Zhang Y. The I-TASSER suite: Protein structure and function prediction. Nat Methods. 2015; 12(1): 7-8. DOI: 10.1038/nmeth.3213
36. Wiederstein M, and Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucl Acids Res. 2007; 35(S2): W407-W410. DOI: 10.1093/nar/gkm290
37. Sormanni P, Aprile FA, and Vendruscolo M. The CamSol method of rational design of protein mutants with enhanced solubility. J Mol Biol. 2015; 427(2): 478-490. DOI: 10.1016/j.jmb.2014.09.026
38. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. Correction to the STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucl Acids Res. 2021; 49(18): D605-D612. DOI: 10.1093/nar/gkab835
39. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria, 2021. Available at: https://cir.nii.ac.jp/crid/1574231874043578752#citations_container
40. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2016. Available at: https://link.springer.com/book/10.1007/978-3-319-24277-4
41. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. (2009). Circos: An information aesthetic for comparative genomics. Genome Res. 19(9): 1639-1645. DOI: 10.1101/gr.092759.109
42. Kolde R. Pheatmap: Pretty heatmaps. R package version 1.0.12. 2019. Available at: https://cran.r-project.org/web/packages/pheatmap/index.html
43. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13(11): 2498-2504. DOI: 10.1101/gr.1239303
44. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004; 432(7018): 695-716. DOI: 10.1038/nature03154