

# A High-Quality Assembly Reveals Genomic Characteristics, 1 Phylogenetic Status and Causal Genes for White Feather of 2 Indian Peafowl

**shaojuan Liu**

South China Agricultural University College of Animal Science

**hao Chen**

Jiangxi Science and Technology Normal University School of Life Science

**jing Ouyang**

Jiangxi Science and Technology Normal University School of Life Science

**min Huang**

South China Agricultural University College of Animal Science

**hui Zhang**

South China Agricultural University College of Animal Science

**sumei Zheng**

South China Agricultural University College of Animal Science

**suwang Xi**

Jiangxi Agricultural University College of Animal Science and Technology

**hongbo Tang**

Jiangxi Science and Technology Normal University School of Life Science

**yuren Gao**

Jiangxi Science and Technology Normal University School of Life Science

**yanpeng Xiong**

Jiangxi Science and Technology Normal University School of Life Science

**di Cheng**

Jiangxi Agricultural University College of Animal Science and Technology

**kaifeng Chen**

Jiangxi Agricultural University College of Animal Science and Technology

**bingbing Liu**

South China Agricultural University College of Animal Science

**wanbo Li**

Jimei University

**xueming Yan**

Jiangxi Science and Technology Normal University School of Life Science

**huirong Mao** (✉ [maohuirong82@hotmail.com](mailto:maohuirong82@hotmail.com))

Jiangxi Agricultural University <https://orcid.org/0000-0003-2588-1521>

**Jun Ren**

South China Agricultural University College of Animal Science

---

## Research

**Keywords:** Indian peafowl, Genome assembly, Phylogeny, PMEL, White feather

**Posted Date:** July 1st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-650113/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**A high-quality assembly reveals genomic characteristics,  
phylogenetic status and causal genes for white feather of  
Indian peafowl**

Shaojuan Liu<sup>1#</sup>, Hao Chen<sup>3#</sup>, Jing Ouyang<sup>3</sup>, Min Huang<sup>1</sup>, Hui Zhang<sup>1</sup>, Sumei Zheng<sup>1</sup>,  
Suwang Xi<sup>2</sup>, Hongbo Tang<sup>3</sup>, Yuren Gao<sup>3</sup>, Yanpeng Xiong<sup>3</sup>, Di Cheng<sup>2</sup>, Kaifeng Chen<sup>2</sup>,  
Bingbing Liu<sup>1</sup>, Wanbo Li<sup>4</sup>, Xueming Yan<sup>3\*</sup>, Huirong Mao<sup>2\*</sup>, Jun Ren<sup>1\*</sup>

<sup>1</sup> College of Animal Science, South China Agricultural University, Guangzhou 510642,  
China

<sup>2</sup> School of Animal Science and Technology, Jiangxi Agricultural University,  
Nanchang 330045, China

<sup>3</sup> College of Life Science, Jiangxi Science & Technology Normal University, Nanchang  
330013, China

<sup>4</sup> Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of  
Agriculture and Rural Affairs, Jimei University, Xiamen 361021, China

<sup>#</sup> Both authors contribute equally to this paper.

<sup>\*</sup> Corresponding author.

E-mail address: maohuirong82@hotmail.com (H. M); xuemingyan@hotmail.com (X.  
Y)

## Abstract

**Background:** Indian peafowl (*Pavo cristatus*) attracts people's attention because of the exclusively dazzling phenotypic characteristics. However, little is known about the phenotypic evolution and phylogeny of Indian peafowl at the whole-genome level. So far, there has been no report on the genetic mechanism of the formation of white feather in white feather peafowl.

**Results:** This study assembled a draft genome of Indian peafowl with a genome size of 1.05 Gb (the sequencing depth is 362×), and N50 of the contig and scaffold was up to 6.2 Mb and 11.4 Mb, respectively. Compared with other birds, Indian peafowl changed in terms of metabolism, immunity, skeletal development and feather development, which provided a novel insight into the phenotypic evolution of peafowl, such as the large body size and feather morphologies. It was confirmed that the phylogeny of Indian peafowl was closer to that of turkey than that of chicken. Specially, it was identified that *PMEL* was a causal gene leading to the formation of white plumage in blue and white feather peafowl.

**Conclusions:** This study provides a peafowl genome with high-quality as well as a novel understanding in the phenotypic evolution and phylogeny of peafowl among other birds. The results contribute a valuable reference genome to the study of the avian genome evolution. In addition, the discovery of the genetic mechanism of white plumage not only is a breakthrough in the exploration of peafowl plumage, but also provides clues and new ideas for further investigations of the avian plumage coloration and artificial breeding in peafowl.

**Keywords:** Indian peafowl; Genome assembly; Phylogeny; *PMEL*; White feather

## Introduction

*Pavo cristatus*, commonly called Indian peafowl or blue peafowl, is the king of birds, representing elegance, honour, beauty, luck, and romance in many Asian cultures (Figure 1a) (Gadagkar, 2003; Kushwaha & Kumar, 2016). Peafowl, belonging to the Aves, Galliformes, Phasianidae, and *Pavo*, has two species: green peafowl and blue peafowl. Indian peafowl, as the national bird of India, has been widely distributed in Bangladesh, Bhutan, India, Nepal, Pakistan, and Sri Lanka (Kushwaha & Kumar, 2016; Ramesh & McGowan, 2009). Commonly, the male Indian peafowl is more attractive than female because of its larger body size and more glittering plumage. Indian peafowl is one of the largest and most beautiful birds in pheasant as a valuable ornamental display with their fan-shaped crests, brightly blue plumage, and glaring tails. Moreover, Indian peafowl is well known as a protein resource, because their meat, internal organs, and bones have high nutritional values, even with medicinal values (Mushtaq-ul-Hassan, Ali, Arshad, Mahmood, & Research, 2012; Paranjpe & Dange, 2019; Talha, Mia, & Momu, 2018).

With the improvement of whole genome sequencing technology, more and more avian genomes were assembled, such as chicken (*Gallus gallus*) ("Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution," 2004), turkey (*Meleagris gallopavo*) (Dalloul et al., 2010), duck (*Anas platyrhynchos*) (Z. Zhang et al., 2018), and other birds (Jarvis et al., 2014), which provided basic references for the study of phenotypic characteristics, evolution, economic traits, and environmental adaptation of birds. As for peafowl, the first draft genome of peafowl was released in 2018. However, the length of scaffold and contig N50 of the assembling were only 25.6 kb and 19.3 kb, respectively (Shubham K. Jaiswal et al., 2018). Subsequently, Dhar, et al. improved peafowl genome by using Illumina and Oxford Nanopore technology (ONT), and the length of scaffold N50 was up to 0.23 Mb (Dhar et al., 2019). The previous studies of peafowl focused on courtship behaviour (Dakin, McCrossan, Hare, Montgomerie, & Amador Kane, 2016), immunity (Wang, Zhao, Liu, Shao, & Xing, 2019), and productivity (Samour, Naldo, Rahman, & Sakkir, 2010). Additionally, most researches on the phylogenetic relationship of peafowl and the Phasianidae were based on the mitochondrial genomes, DNA transposable factors, and partial DNA nucleotide sequences, and the conclusions of

these researches are still controversial (J. Naseer et al., 2018; Shen et al., 2014; Zhou, Sha, Irwin, & Zhang, 2015). Therefore, a better quality assembly of peafowl genome is needed to provide baseline data for further studies on peafowl. It has been suggested that the grey peafowl pheasant (*Polyplectron bicalcaratum*) could be the ancestor of peafowl (Bush & Strobeck, 2003). However, the phylogenetic relationships of peafowl with other pheasants are still unclear, though there have been several recent attempts to investigate it. Study on Hainan peafowl pheasant provided evidences to argue whether it is a subspecies of the grey peafowl pheasant (Chang et al., 2008). Zhou et al. (2015) analysed the phylogenetic status of Indian peafowl in Phasianidae by using a complete mitochondrial genome, and found out its closest genetic affinity with green peafowl (Zhou et al., 2015). Sun et al. (2014) revealed the close relationships among *Pavo*, *Polyplectron* and *Argusianus* within the Phasianidae (Sun et al., 2014). It is expected that a more clarified taxonomic status of the peafowl in Phasianidae can be investigated via the high-throughput sequencing technology.

Previous studies reported that there were many plumage colour mutants, including white, black, variegated, cameo, and oaten (Ouyang et al., 2009; R. G. Somes & Burger, 1993; R. G. Somes & Burger, 1991), among which, the most ornamental colour was the white plumage, belonging to leucism rather than albinism since the feather was white but the eyes had melanin pigmentation (Figure 1b). The inherited basis of plumage colour has attracted researchers for a long time. The first report on it suggested that the plumage phenotype of peafowl was determined by autosomal genes in a recessive model (R. G. Somes & Burger, 1991). Another study verified that a single autosomal locus was in control of all plumage phenotypes in peacock, where the pied colour appeared in two heterozygous mutant alleles, with the black on recessive mutant allele, and the all-white plumage in homozygous mutant allele as the most dominant (R. G. Somes & Burger, 1993). Nevertheless, further studies on the genetic mechanism of the white plumage in peafowl are needed, i.e., to clarify the causative mutations of this phenotype.

Therefore, a high-quality (near-chromosomal) reference genome of Indian peafowl was constructed by using the third-generation de novo assembly technology. Based on the assembly, it was devoted to investigate the molecular evolution and phylogenetic classification of peafowl in the Phasianidae at the genome-wide level. Subsequently,

comparative genomics analysis was performed to investigate the biological characteristics of evolution through comparing the genome of Indian peafowl with the high-quality genomes of other birds, human and mouse. Furthermore, the transcriptomic and pooled resequencing data were analysed to identify the genetic mechanism of the white plumage in Indian peafowl. This work will provide an updated understanding of and key reference for genomic characteristics, phylogenetic status and genetic mechanism of white feather in Indian peafowl.

## **Materials and Methods**

### **Sample collection**

All procedures used for this study and involved in animals fully complied with guidelines for the care and utility of experimental animals established by the Ministry of Agriculture of China. The ethics committee of South China Agricultural University approved this study. A blood sample was collected from a female Indian peafowl for genome assembling, and 51 blood samples from 35 blue feather peafowls and 16 white feather peafowls for pooled resequencing in Leping Sentai special breeding Co., Ltd in Jiangxi Province, China, under the principles and standards of animal welfare ethics. Meanwhile, two liver and two muscle tissues were sampled from a female Indian peafowl to assist the process of assembling the Indian peafowl genome. Additionally, feather pulps from 8 blue and 8 white peafowls were collected for RNA-seq.

### **DNA and RNA extraction**

Genomic DNA was extracted from blood samples by using a routine phenol-chloroform protocol. The concentration of the extracted DNA was evaluated by using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and diluted to a final concentration of 100 ng/μL. The integrity of DNA was checked via electrophoresis on 0.8% agarose gel. Total RNA of feather pulp was extracted by using TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA). The purity and degradation of RNA was detected by Nanodrop 2000 spectrophotometer and agarose gel electrophoresis.

### **De novo assembling of the Indian peafowl reference genome**

Library Preparation and Sequencing: Genomic DNA was used to make a 350 bp insert fragment libraries by using the Illumina TruSeq Nano method, starting with 100 ng DNA. Mate pair libraries were made by using the Nextera Mate Pair Sample Preparation Kit (Illumina) with the gel plus option, and sequenced by using Illumina NovaSeq 6000 platform. For PacBio sequencing, genomic DNA was sheared by a g-TUBE device (Covaris) with 20 kb settings for further preparing a 20 kb Single-Molecule Real Time (SMRT) bell, and then the single-molecule sequencing was completed on a PacBio RS-II platform. For 10X genomics sequencing, each GEM was amplified by PCR and added P7 sequencing adapters for Illumina sequencing.

Genome Assembly: The genome assembling of Indian peafowl was performed in five steps, which was illustrated in Supplementary Figure S1. The raw reads were generated from two paired-end libraries sequenced on Illumina NovaSeq 6000 platform. The sequencing adapters, contaminated reads, and low-quality reads were removed by using megablast v2.2.26 (Y. Chen, Ye, Zhang, & Xu, 2015). The genome size was calculated by using the formula:  $\text{Genome size} = \text{kmer\_Number} / \text{Peak\_Depth}$ . Secondly, PacBio sequencing was used to control and correct errors. The error corrected data were assembled by falcon software (C.-S. Chin et al., 2016), and the Overlap-Layout-Consensus algorithm was used to obtain the consensus sequences, which were then corrected by quiver software (C. S. Chin et al., 2013). Combined with the second-generation sequencing data, the consensus sequences were recalibrated by using the pilon software (Walker et al., 2014) to improve the accuracy, and high-quality consensus sequences were obtained. Thirdly, the 10X genomics sequencing was used to assist the genome assembly. The 10X genomics library was sequenced to obtain linked-reads, which were aligned to the consensus sequences obtained from the PacBio sequencing assembly, and then linked reads were added to assemble the super-scaffolds by fragScaff software (Adey et al., 2014). Fourthly, similar to the third step, Chicago sequencing data was used to assist assembly to map the draft genome assembly. Finally, the Illumina reads were mapped to the draft genome by using BWA (Burrows-Wheeler Aligner) (H. Li & Durbin, 2009). Then, pilon (version 1.22) was used to correct the assembled errors based on the mapped results.

Consistency and completeness: The consistency and integrity of the assembled peafowl genome were separately assessed by using the BUSCO (Benchmarking



Universal Single-Copy Orthologs) (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) and CEGMA (Core Eukaryotic Genes Mapping Approach) (Parra, Bradnam, & Korf, 2007; Parra, Bradnam, Ning, Keane, & Korf, 2009), based on single-copy orthologues from the AVES (odb9) database. In order to evaluate the accuracy, integrity and sequencing uniformity of the genome assembly, small fragment library reads were selected and aligned to the assembled genome by using BWA software. All the genomic sequences were generated by Novogene Inc, Beijing, China.

Genome Annotation: Genome annotation mainly included three aspects: repetitive sequence annotation, gene annotation (including gene structure prediction and gene function prediction) and non-coding RNA (ncRNA) annotation (Supplementary Figure S2). The repetitive sequence annotation included the annotation through homologous sequence alignment and ab initio prediction. The RepeatMasker and RepeatproteinMask software (Tempel, 2012) were employed to identify known repetitive sequences against the RepBase library (Jurka et al., 2005). In ab initio prediction, LTR\_FINDER (Z. Xu & Wang, 2007), RepeatScout (Price, Jones, & Pevzner, 2005), and RepeatModeler (Flynn et al., 2020) were used to establish the de novo repeat sequence library, and then repetitive sequences were predicted by Repeatmasker software. The Tandem Repeats (TEs) in the genome were found by Tandem Repeat Finder software (Benson, 1999). In gene annotation, it mainly combined three prediction methods: homology-based prediction, de novo prediction, and other evidence-backed predictions. Homology-based prediction used the protein sequences of *Gallus gallus*, *Meleagris gallopavo*, Peking duck, *Struthio camelus*, *Nipponia nippon*, and Eastern Zhejiang white goose, downloaded from Ensembl (release 74), to align to the Indian peafowl genome by using TblastN (Kent, 2002). Genewise (Birney, Clamp, & Durbin, 2004) was then used to align to the matched proteins for a precise gene model.

In addition, Augustus (Stanke et al., 2006), GlimmerHMM (Majoros, Pertea, & Salzberg, 2004), Geneid (Alioto, Blanco, Parra, & Guigó, 2018), GenScan (Burge & Karlin, 1997), and SNAP software (Korf, 2004) were used for the ab initio predictions of gene structures. The above predictions with transcriptome-based data being combined, EVidenceModeler software (Haas et al., 2008) was used to integrate the gene set and generate a non-redundant and more complete gene set. Finally, PASA was used

to correct the annotation results of EVIDENCEModeler for the final gene set. Gene function of the final gene set was annotated by using the protein database of SwissProt (Bairoch & Apweiler, 2000), NR (O'Leary et al., 2016), Pfam (El-Gebali et al., 2019), KEGG (Kanehisa & Goto, 2000), and InterPro (Zdobnov & Apweiler, 2001). tRNAscan-SE software (Lowe & Eddy, 1997) was used to search for the tRNA sequence of genome, with INFERNAL software (<http://infernal.janelia.org/>) (Nawrocki & Eddy, 2013) from Rfam (Griffiths-Jones et al., 2005) to predict miRNA and snRNA of genome.

### **Gene family**

The amino acid sequences of the following were downloaded from NCBI database to identify the gene families and single-copy orthologous genes. They are: Japanese quail (*Coturnix japonica*) (Nishibori, Hayashi, Tsudzuki, Yamamoto, & Yasue, 2001), chicken (*Gallus gallus*) (Bellott et al., 2017), turkey (*Meleagris gallopavo*) (Dalloul et al., 2010), northern bobwhite (*Colinus virginianus*) (Oldeschulte et al., 2017), common mallard (*Anas platyrhynchos*) (Gregory & James, 2014), zebra finch (*Taeniopygia guttata*) (Korlach et al., 2017), collared flycatcher (*Ficedula albicollis*) (Ellegren et al., 2012), medium ground-finch (*Geospiza fortis*) (G. Zhang et al., 2014), tibetan ground-tit (*Pseudopodoces humilis*) (Q. Cai et al., 2013), rock pigeon (*Columba livia*) (Shapiro et al., 2013), peregrine falcon (*Falco peregrinus*) (Zhan et al., 2013), saker falcon (*Falco cherrug*) (Friedman-Einat et al., 2014), human (*Homo sapiens*) (Mohajeri et al., 2016), and mouse (*Mus musculus*) (Church et al., 2011). The longest transcript of each gene was extracted and then the genes with the length of protein sequences shorter than 50 amino acids were filtered. Based on the filtered protein-coding sequences data set, Orthofinder v2.3.7 (Yu, Zavaljevski, Desai, & Reifman, 2011) was used to identify gene families and orthologous gene clusters of 15 species. The single-copy orthologous sequences from the gene families were aligned by using MAFFT v7.450 software (Katoh & Standley, 2013), and then the poorly sequences were removed by using Trimal software with default parameters (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009). The final result was used as a single data set for the subsequent comparative genome analyses.

## **Phylogenetic tree and divergence time**

To determine the phylogenetic relationship of 15 species, IQ-tree v2.1.2 software was first used to find the best model for constructing phylogenetic tree with options “-m MF” and the species tree with bootstrap 1000 based on the concatenated alignment of single-copy orthologues sequences from 15 species (Minh et al., 2020). RAxML software was used to construct phylogenetic tree with parameters “-m PROTGAMMALGX -f a” with bootstrap 1000. Divergence time of 15 species was estimated by using MCMCtree program implemented in PAML packages (Yang, 2007). Five calibration time (human-mouse (85~97Mya), human-zebra finch (294~323Mya), zebra finch-medium ground finch (30.4~46.8Mya), common mallard-zebra finch (93.2~104.6Mya) and saker falcon-peregrine falcon (1.66~3.68Mya)) from TimeTree database (Hedges, Dudley, & Kumar, 2006) were used as constrains in the divergence time estimation. The MCMC process was run to sample 1,000,000 times, sample frequency set to 10, and burn-in 40,000, to finally achieve a convergence by using Tracer v1.7.1 (<http://tree.bio.ed.ac.uk/software/tracer/>).

## **Genome Synteny and Collinearity Analysis**

To compare the genome synteny of peafowl with chicken and turkey, the homologue of the genome was identified by using BLASTp (E-value <  $1e^{-10}$ ). Gene pairs of synteny blocks within the genome were identified by using MCScanX (Wang et al., 2012), and the synteny blocks were showed by circos program from TBtools (C. Chen et al., 2020). To estimate the positively selected genes for peafowl-chicken and peafowl-turkey, the value of Ka/Ks ( $\omega$ ) for each gene pair was calculated by KaKs\_calculator (Z. Zhang et al., 2006), and the density curve of values was visualized by R software. The positively selected genes ( $\omega > 1$ ) were conducted based on functional enrichment analysis.

## **Gene-family expansion and contraction**

To identify the gene family expansion and contraction in peafowl, the gene families in 15 species and phylogenetic tree with divergent times were taken into account to estimate the significance of gene gain and loss in gene family by using the CAFE v4.2.1 with a random birth and death model and significance of P-values < 0.05 (De Bie, Cristianini, Demuth, & Hahn, 2006). The parameter  $\lambda$  represents the probability of gene gain and loss in a divergent time. In order to investigate the evolutionary rates of

different branches of the tree, the argument with “-t” was used to define three different branches for 15 species: the first branch included mouse and human, the second branch was the Phasianidae, and other birds were regarded as the third branch. Then, they were conjunct with the “-s” option to search the optimal  $\lambda$  value for different branches by using the maximum likelihood.

### **Positive Selection Analyses**

To determine the adaptive evolution under the positive selection in peafowl, the single-copy orthologous protein sequences shared among the 11 species (peafowl, chicken, turkey, common mallard, zebra finch, collared flycatcher, medium ground-finch, tibetan ground-tit, rock pigeon, peregrine falcon and saker falcon) were searched, filtered, and then converted to coding gene sequence (CDS) by using EMBOSS backtranseq program (Rice, Longden, & Bleasby, 2000), and the CDS were then aligned to codon by using PRANK with the option “-codon” (Löytynoja, 2014). The above alignments were analysed by using CODEML program of the PAML package 4.9 (Yang, 2007). A branch-site model (TEST-II) (model = 2, NSsites = 2) was conducted to identify the positively selected genes of peafowl. The model assumed that a particular branch (foreground, alternative hypothesis) had a different  $\omega$  value from all the sites compared to all other branches (background, null hypothesis), suggesting that positive selection occurred at only a few sites on a particular branch (foreground) (Yang, 2007). The peafowl was regarded as a foreground branch and other species as a background branch. Additionally, the branch model was used to identify the rapidly evolving genes in peafowl, assuming that the branch of peafowl was as an alternative hypothesis (model = 2) and the branches of other species were as the null hypothesis (model = 0). The dN/dS ( $\omega$ ) values between foreground branch and background branch were estimated by using Likelihood Ratio Test (LRT) values based on chi-square test. When the  $\omega$  value in the foreground branch was greater than that of the background branch, it suggested that the genes of the foreground branch were under positive selection ( $P < 0.05$ ), and the positively selected sites were determined by using the Bayesian Empirical Bayes method. All the positively selected genes were performed in functional enrichment analysis by using KOBAS (Xie et al., 2011).

### **Whole-genome resequencing and variant calling**

The genomic DNA from 35 blue feather peafowls and 16 white feather peafowls were pooled, respectively. Then 1.5 µg DNA per pool was used for constructing the sequencing libraries by using Truseq Nano DNA HT Sample preparation Kit (Illumina, USA) following manufacturer's constructions. Each pooled DNA was fragmented through sonication to a size of 350bp and end repaired, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. PCR-amplified sequencing libraries were purified (AMPure XP system) and analysed for size distribution on Agilent2100 Bioanalyzer, and were quantified by using real-time PCR. These libraries constructed above were sequenced on an Illumina NovaSeq platform and 150bp paired-end reads were generated with insert size around 350 bp. The raw data were filtered by removing reads with  $\geq 10\%$  unidentified nucleotides (N), reads with  $> 50\%$  bases having phred quality  $< 5$ , and reads with  $> 10$  nt aligned to the adapter allowing  $\leq 10\%$  mismatches. The clean reads were mapped to the assembled reference genome by using BWA with parameters "mem -t 4 -k 32 -M -R". Alignment files were converted to BAM files by using SAMtools software (settings: -bS -t) (H. Li et al., 2009). In addition, potential PCR duplications were removed by using SAMtools command "rmdup". Single nucleotide polymorphisms (SNPs) and insertions/ deletions (Indels) ( $< 50$  bp) were detected by using Genome Analysis Toolkit v 4.0 (GATK) pipeline (McKenna et al., 2010).

### **RNA sequencing (RNA-seq) on PacBio platform**

The cDNA of feather was acquired through PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa Bio. Inc, Dalian, China) according to the manufacturer's instructions. The cDNA was performed damage repair, end repair, SMRT (single-molecule, real-time) dumbbell-shaped adapters, and ligation of the adapters to construct a mixed library. Primers and DNA polymerase were then combined to form a complete SMRT bell library. The qualified library was used for sequencing on a PacBio Sequel platform. The clean data were aligned to the reference genome of Indian peafowl by using STAR v2.5.3a (Dobin et al., 2013). The Transcript assembly and gene expression levels were conducted by using StringTie v1.3.3 (Pertea, et al., 2015) and featureCounts (Liao, et al., 2014) in Subread software (Liao, Smyth, & Shi, 2013). Differentially expressed genes (DEGs) between blue and white feather were identified through DESeq2 (Love, et al., 2014) in condition of fold change  $> 2$  and P-value  $< 0.01$ . Subsequently, the

functional enrichment analyses of DEGs were annotated through GO (Gene Ontology) (Ashburner et al., 2000) and KEGG (Kyoto Encyclopedia of Genes and Genomes) database.

### **cDNA amplification**

cDNA of feathers was reversely transcribed with PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa). The reverse transcription quantitative PCR (RT-qPCR) was conducted in a total volume of 10 µl including 5 µl SYBR Taq II kit (TaKaRa), 0.3 µl Rox Reference Dye (50x), 2.7 µl distilled water, 1 µl cDNA and 1 µl primers, and performed on a 7900HT RT-qPCR system (ABI).  $\beta$ -actin was selected as the internal reference gene. All primer sequences were presented in Supplementary Table S20.

## **Results**

### **Genomic characteristics of Indian peafowl**

The third-generation PacBio single-molecule real-time sequencing technology and the second-generation Illumina sequencing technology were used and combined with 10X genomics to assemble the Indian peafowl genome. The distribution of 17-kmer showed a major peak at 154x (Supplementary Figure S3). Based on the total number and corresponding k-mer depth of 154, the Indian peafowl genome size was estimated to 1.05 Gb. A total of two libraries were constructed with a sequencing volume of 164.03 Gb and a coverage depth of 154x performed on Illumina NovaSeq 6000 platform. 10X Genomics sequencing platform yielded 112.57 Gb sequencing data with a coverage depth of 92x, and 110.74 Gb sequencing data was produced by using the PacBio sequencing platform with a coverage depth of 103x (Supplementary Table S1). In total, 387.34 Gb sequencing data and a total coverage 362x were obtained from the three sequencing strategies with the lengths of scaffold N50 and contigs N50 separately up to 6.2 Mb and 11.4 Mb, which exhibited 446-fold and 50-fold improvement in the scaffold N50 compared to previously published Indian blue peafowl genome reported by Jaiswal, et al. (Shubham K. Jaiswal et al., 2018) and Dhar, et al. (Dhar et al., 2019) (Figure 2, Table 1 and Supplementary Table S2). Current peafowl assembly anchored into 726 scaffolds with 1.05 Gb of sequences and guanine-cytosine (GC) content was 42.03% with normally ratio of A, T, G and C (Figure 2 and Supplementary Table S2 - S3). Notably, the GC contents of scaffold 108 and scaffold 31 were more than 50%,

and the gene density of them was relatively conferred (Figure 2). Moreover, it was apparently observed that the density of tandem repeat sequences in scaffold 39 was the higher than any other scaffolds (Figure 2).

To assess the completeness and base accuracy of the assembled Indian peafowl genome above, the short reads were aligned back to the draft genome with high mapping rate (98.05%), high coverage rate (99.87%) and low homozygous SNP rate (0.0002%), generally reflecting the accuracy of genome assembly (Supplementary Table S4-S5). The draft assembly was further evaluated by CEGMA, and the results showed that 88.71% of 248 core genes selected from 6 eukaryotic model organisms could be covered. Additionally, 97.4% complete genes (including 96.8% complete and single-copy genes and 0.6% complete and duplicated genes) were predicted, 1.7% fragmented genes and 0.9% missing genes were identified from 2,586 genes in Aves dataset by using the BUSCO (Supplementary Table S6). Collectively, these important indicators implied relatively high genome coverages and continuity for Indian blue peafowl genome, providing an important resource for molecular breeding and evolutionary studies for peafowl. The Indian peafowl genome with integrated annotations has been deposited in the Genome Warehouse in National Genomics Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number GWHAZTP000000000 publicly accessible at <https://bigd.big.ac.cn/gwh> ("Database Resources of the National Genomics Data Center in 2020," 2020).

Taking the *Gallus gallus*, *Meleagris gallopavo*, Peking duck, *Struthio camelus*, *Nipponia nippon*, and Eastern Zhejiang white goose genome as references, and according to the homologous alignment and ab initio prediction, the peafowl genome comprised 15.20% non-redundant repeat sequences, including 1.27% tandem repeats, 14.12% transposable elements and 7.35% transposable element protein (Supplementary Table S7). A total of 14.56% of transposable elements was identified after combine TEs, 0.70% of which was DNA transposons, 3.93% was long terminal repeats (LTRs), 0.01% was short interspersed nuclear elements (SINE), and 10.68% was the long interspersed nuclear elements (LINE) (Supplementary Figure S4 and Supplementary Table S7-S8). Altogether, 19,465 non-redundant protein-coding genes were predicted,

of which 15,766 (81%) were annotated to function by six public databases, including NR, Swissprot, KEGG, InterPro, GO and Pfam (Table 1 and Supplementary Table S9-S10). Additionally, 354 microRNAs (miRNAs), 308 transfer RNAs (tRNAs), 151 ribosomal RNAs (rRNAs) and 334 small nuclear RNAs (snRNAs) were also identified (Supplementary Table S11). Notably, compared with the other birds such as turkey, chicken, duck or previously reported draft peafowl genome, more non-redundant protein-coding genes and repetitive sequences as well as non-coding RNA in this draft peafowl genome were predicted. Overall, this assembly has more continuity, completeness, and accuracy, and the annotation of core eukaryotic genes and universal single-copy orthologs have also been improved.

#### **Gene families and phylogenetic relationship within the *Phasianidaes***

The protein sequences of 15 species were used to search the orthologues by using the OrthoFinder (Emms & Kelly, 2015). The results showed that a total of 18,038 orthogroups were identified in 15 species, of which 5,999 single-copy orthologues were shared among these species (Figure 3a). In addition, 93 gene families were identified specific to peafowl and 11,447 gene families were shared by peafowl and other Phasianidaes (chicken, turkey and Japanese quail) (Figure 3b). The peafowl species-specific gene families were mainly involved in the immune response and biological process such as *FOXP3*, *FZD3*, and *TP53* participated in many immunological process and played an important role in melanoma and bone homeostasis (Supplementary Table S12) (Fischer et al., 2019; C. Li et al., 2019; Thoenen, Curl, & Iwakuma, 2019). *FOXP3* is necessary for the development of regulatory T lymphocytes and is essential for maintaining the immune homeostasis and immune self-tolerance to the environmental antigens by eliminating natural reactive T cells in the thymus and peripheral organs. Meanwhile, *FOXP3* plays an important role in the bone and hematopoietic homeostasis, inflammatory bone loss diseases and abnormal bone weight, which can affect lymphoid hematopoiesis by acting on the development and function of osteoclasts (Fischer et al., 2019). *TP53* plays an important role in inhibiting the progression of bone and soft tissue sarcoma. The loss of *TP53* activity can promote the osteogenic differentiation of bone marrow stromal cells and the development of osteosarcoma of bone marrow



mesenchymal stem cells, which can prevent the malignant transformation of bone marrow mesenchymal stem cells (Thoenen et al., 2019). In this study, the enrichment of these genes specific to chicken and turkey in peafowl showed that the healthy development and immunity of bones were of great significance in peafowl evolution, which was conducive to achieving people's demand for rapid growth, large size, and strong disease resistance in domestication of peafowl.

To determine the phylogenetic relationship within the Phasianidae family, 5,999 single-copy orthologues were concatenated and aligned to construct phylogenetic tree with a bootstrap value of 1,000 by using the maximum likelihood method (Supplementary Figure S5 and S6). The results showed that the Galliformes order were clustered, of which the Phasianidae family formed a group. Moreover, peafowl was found closer to turkey than chicken in the Phasianidae family that was inconsistent with Jaiswal, et al. reported (Shubham K. Jaiswal et al., 2018), the relationship of chicken and quail was closer than turkey, and duck belonging to the Anseriformes order was close to the Galliformes order (Figure 3c). Additionally, the divergence time of all species was estimated and calibrated through the divergence time between human and mouse, human and zebra finch, zebra finch and medium ground finch, common mallard and zebra finch, and saker falcon and peregrine falcon from the TimeTree database. The divergence between Galliformes and Anseriformes was estimated to be 81.2 million years ago (Mya). The divergence between northern bobwhite and Phasianidae family was represented by the calibration point of northern bobwhite and turkey. The divergence between the peafowl and turkey was about 35.1 Mya, sharing a common ancestor with chicken about 36.9 Mya (Figure 3c). However, divergence between chicken and Japanese quail was estimated to be 34.7 Mya, within the range of divergence (33.2~42.3 Mya) from TimeTree (T. Cai et al., 2017), which suggested that the relationship between the common ancestor of peafowl and turkey, chicken and Japanese quail was very close to the relationship between these four species. The divergence time of pheasant birds is in the Tertiary era, which marks the advent of the modern biological era and is the heyday of the divergence of animals and plants. It presents that the new generation replaces the ancient types, with the increasing of the same species, common and diverse divergence of birds, and rapid evolution of more species during the Tertiary era.

## **Genome synteny and collinearity among the Indian peafowl, chicken, and turkey**

Collinearity analysis can reflect the homology of different species and genetic relationship. Genes with the pairwise ratio of nonsynonymous to synonymous substitutions (dN/dS) could be used to infer the positive selection and contributed to understanding the evolutionary characteristics in species. In this study, the pairwise synteny was compared to peafowl and chicken, peafowl and turkey, and the ratio of dN/dS was calculated. The scaffolds of length greater than the scaffold N70 (5 Mb) in peafowl genome and other collinear scaffolds marked as others (Figure 4a and 4b) were displayed. Moreover, the distribution density of the dN/dS ratio was calculated and showed in Figure 4c. 97 positively selected genes (dN/dS > 1) in peafowl compared to chicken were associated with the biological process and immune-related pathways (*IL4*, *CD3D*, *CD3E* and *HLA-DMB*) ( $P < 0.05$ ), such as, Th1 and Th2 cell differentiation, T cell receptor signaling pathway, and intestinal immune network for IgA production. Besides, compared with turkey, 43 positively selected genes were significantly enriched in the GO terms of organelle (GO:0043226), extracellular space (GO:0005615), and epithelium migration (GO:0090132), and the pathways of glutathione metabolism (*GPX1*, *GPX2* and *GPX4*) and thyroid hormone synthesis (*GPX1*, *DUOXA2* and *GPX2*) ( $P < 0.05$ ) (Supplementary Table S13 and S14), which were involved in gastrointestinal health, anti-stress, growth development and metabolism. Notably, as a common positive selection gene, *EDN1* was reported to participate in many biological process, such as epithelium migration and differentiation, pigmentation and their receptors (EDNRs) widely distributed in various tissues in chicken (H. Liu et al., 2019). These enrichment results indicated that the positively selected genes in peafowl were mainly related to intestinal immunity, anti-stress, growth development and metabolism, and pigmentation compared with turkey and chicken in the evolutionary process, which was beneficial for peafowl to enhance adaptability, improve disease resistance and anti-stress ability, enrich plumage colour, and better adapt to the living environment of long-term artificial breeding.

## **Gene family expansion and contraction across the Indian peafowl genome**

Likelihood analysis could identify the evolutionary rate and the significant expansion and contraction of gene family in species (De Bie et al., 2006). In this study, the changes of gene family in peafowl with likelihood ratio test were examined. Compared to the

gene families in other species, the results suggested that 121 expansions and 2,999 contractions of gene families ( $P < 0.05$ ) were detected in peafowl (Figure 3c), of which, 21 significantly gained genes were mainly involved in energy metabolism and storage (*GIMAP1*, *GIMAP2* and *GIMAP8*) and immune response (*CD244*) ( $P < 0.05$ ), such as the GO terms of natural killer cell activation involved in immune response (GO:0002323), MHC class I protein binding (GO:0042288), positive regulation of interleukin-8 production (GO:0032757), positive regulation of interferon-gamma production (GO:0032729), and lipid droplet (GO:0005811) (Supplementary Table S15). Conversely, 23 significantly contracted genes were mainly relevant to the biological process such as fatty acid degradation (*ALDH3A2*) (GO:0001561), myocardium development (GO:0048739), muscle contraction and cardiac disease (*MYH6*, *MYH7*, and *MYH7b*), olfactory receptor activity (*OR52B2*, *OR52K1* and *OR4S1*) (GO:0004984), and the pathways of olfactory transduction, metabolism and cardiac muscle contraction (Supplementary Table S16). For example, the expression of *MYH6* and *MYH7* directly dictated the slow- or fast-twitch phenotype in skeletal muscle and played a vital role in cardiomyocyte energetics and metabolism (Stuart et al., 2016; Toepfer et al., 2020). The olfactory genes were importantly characteristic during the adaptive evolution in birds (Steiger, Kuryshev, Stensmyr, Kempnaers, & Mueller, 2009). Peafowl is artificially raised and feed-derived food as a result of the declining of finding food and flighting ability in the wild under the long-term domestication, which is likely to cause the contraction of genes related to the sense of smell and the regulation of skeletal muscle movement in peafowl. In addition, it was observed that the branch of Phasianidae had a higher rate of birth and death than that of other two branches, indicating that the family Phasianidae experienced a rapid evolution.

### **Positively selected genes in the Indian peafowl genome**

To reveal the adaptive divergence and evolution of peafowl, positive selection was analysed by using the branch-site model in CODEML program. The significantly positive sites were evaluated by the Bayes Empirical Bayes values ( $\text{BEB} \geq 0.95$ ), which demonstrated that the sites were under positive selection in branch-site model A (foreground). In the branch of peafowl (foreground), 3,417 genes were under significantly positive selection based on the Bayes Empirical Bayes values ( $P < 0.05$ ). These genes were annotated and classified through the analysis of GO ontology and

KEGG pathways in order to further explore the impact of adaptive evolution on peafowl. According to the results of functional enrichment analyses, it was briefly summarized that these positively selective genes were mainly participated in the process of lipid metabolism (i.e., GO:0005811, GO:0030169, and GO:0008289), limb and skeletal development (i.e., GO:0060173, GO:0001503, and GO:0030509), immune response (i.e., GO:0070498, GO:0043123, and GO:1901224), pigmentation (GO:0042470 and GO:0030318), sensory perception (i.e., GO:0008542, GO:0008542, and GO:0007605), and other GO terms (Supplementary Table S17). Additionally, the pathways of positively selected genes were significantly enriched in metabolic pathways, PI3K-Akt signaling pathway, NF-kappa B signaling pathway, pathways in cancer, MAPK signaling pathway, TNF signaling pathway, Jak-STAT signaling pathway, mTOR signaling pathway, FoxO signaling pathway, fatty acid metabolism, IL-17 signaling pathway, cholesterol metabolism, Th17 cell differentiation, and so on (Supplementary Table S18), which were mainly associated with immunity, energy metabolism, and cell growth and differentiation.

Furthermore, the branch model was used to identify a total of 10 rapidly evolving genes in peafowl, including *BCI7A*, *MEF2C*, *MED27*, *COPS7A*, *NMNAT2*, *SLC25A25*, *TNIP2*, *ETS1*, *CCDC6* and *GSGIL*. Functional enrichment showed that the significantly pathways included pathways in cancer, nicotinate and nicotinamide metabolism, thyroid cancer, renal cell carcinoma, parathyroid hormone synthesis, secretion and action, thyroid hormone signaling pathway, apelin signaling pathway, fluid shear stress and atherosclerosis ( $P < 0.05$ ). The significant GO terms were involved in melanocyte differentiation, skeletal muscle and bone development, immunity, and response to stress (Supplementary Table S19). Particularly, it was observed that *MEF2C* participated in most of GO terms and pathways, which might have important implications in the rapid evolution of peafowl. Cartilage formation was a key process in vertebrate bone development and health maintenance that most bones were developed through cartilage ossification. *MEF2C* could regulate muscle and cardiovascular development, and was not only a core component of development in regulating muscle, nerve, cartilage-like, immune and endothelial cells, but also necessary for normal chondrocyte hypertrophy and ossification (Dong et al., 2017; Mackie, Ahmed, Tatarczuch, Chen, & Mirams, 2008). Potthoff, et al. suggested that

*MEF2C* could directly regulate the transcription of myosin gene, and the loss of *MEF2C* in skeletal muscle caused improper sarcomere organization, which revealed the key role of *MEF2C* in maintaining sarcomere integrity and skeletal muscle maturation after birth (Potthoff et al., 2007). Arnold, et al. indicated that the transcription factor *MEF2C* could regulate muscle and cardiovascular development, and control skeletal development by activating the genetic program of chondrocyte hypertrophy (Arnold et al., 2007). In this study, it was observed that *MEF2C* underwent rapid evolution in peafowl, which was conducive to the development and morphology of bones and the maintenance of body shape of peafowl. This may be due to the increase in weight and body size of peafowl in order to meet people's needs during the domestication process, and could be well explain the evolutionary phenotype characteristics of peafowl.

#### **Genes with allele frequency between blue and white feather Indian peafowl**

To localize the genomic region underlying the plumage colour, the allele frequency between blue and white feather peafowl was analysed. The clean data of two pooled resequencing were aligned to the assembled peafowl genome by using the Samtools with option “mpileup”, and filtered to calculate allele frequency differences by using Population2 software (Kofler, Pandey, & Schlötterer, 2011). The significance of allele frequency differences was estimated by Fisher's exact test. The up- and downstream of 50 Kb with  $-\log_{10}$  (P-value) greater than 30 as potential candidate regions was extracted. As a result, it was found that *EDNRB* in scaffold 196 and *PMEL* in scaffold 144 were significantly related to plumage pigmentation (Figure 5a). Additionally, 69 down-regulated genes and 52 up-regulated genes between blue and white peafowl were detected, of which ten up-regulated genes (*TRYPI*, *TYR*, *PMEL*, *EDNRB*, *OCA2*, *SLC24A5*, *SOX10*, *MC1R*, *SLC45A2* and *TRPM1*) were associated with melanin deposition (Figure 5b). The functional enrichment of DEGs showed the most significant pathway was enriched in the process of melanin synthesis based on the criterion of  $P < 0.05$  as significant (Figure 5c). In order to further investigate the allele imbalance difference sites of DEGs, the allele frequency of ten pigmentation-related genes in the blue and white peafowl was calculated, and the function of sites was annotated by using the snpEff software (Cingolani et al., 2012). An observation showed that two differential sites were located in *PMEL* and one in *EDNRB*, but none of the differential

sites were obviously functional mutations, such as missense mutations, splicing mutations, or nonsense mutations (Figure 5d). These results suggested that the formation of white feathers was most likely related to the differential expression of *PMEL* and *EDNRB* in peafowl.

### **Candidate causative gene for the phenotype of white feather in blue and white feather peafowl**

To detect the *PMEL* and *EDNRB* transcripts in blue and white feather peafowl, the RNA-seq data of *PMEL* and *EDNRB* were examined by using the integrative genomics viewer (IGV) application, and it was observed no difference in the transcript of *EDNRB* in the two types of feather pulp (Supplementary Figure S7), suggesting that *EDNRB* was normally expressed in blue and white feather peafowl. When compared the transcript of *PMEL* in blue and white feather peafowl, it was found that *PMEL* was almost not expressed in white feather peafowl (Figure 5e). To further determine the mRNA expression of *PMEL* in white feather peafowl, reverse transcription quantitative PCR (RT-qPCR) of *PMEL* was conducted in blue and white feather peafowl (Supplementary Table S20). RNA samples were extracted from feather pulps and used for subsequent PCR. Surprisingly, it was observed that the mRNA expression of *PMEL* in white feather peafowl significantly reduced in comparison to that in the blue feather peafowl ( $P = 0.013$ ) (Figure 5f), which was consistent with the result of RNA-seq data. Hence, it was argued that *PMEL* was a strong candidate causative gene for the formation of white feathers in blue and white feather peafowl. Further investigations are needed to conduct on the mechanism for the downregulated expression of *PMEL* in white feather peafowl.

### **Discussion**

With the development of sequencing technology, the reduction of sequencing costs and the improvement of assembly methods, the increasing genome sequence maps of various species have been published, making the whole genome sequencing to be an important method for conducting basic genetic research on species. Recently, many avian genomes have been assembled, providing good materials to study the genetic mechanisms of evolution, behaviour and pathology. In this study, three sequencing strategies were combined to construct the India peafowl genome, and totally 1.05 Gb

draft genome sequence was obtained, and even the sequencing depth was up to 362×. Moreover, the lengths of contigs N50 and scaffold N50 were respectively achieved to 6.2 Mb and 11.4 Mb, which was closer to the chromosomal level. Compared with other avian genomes and the draft genomes of peafowl assembled by Jaiswal, et al. and Dhar, et al. (Dhar et al., 2019; Shubham K. Jaiswal et al., 2018), the Indian blue peafowl genome in this study showed a significantly improved assembly quality, including consistency, accuracy and integrality. This draft genome of peafowl was a considerable improvement in terms of the quality of genome assembly and provided a strong support and guarantee for the subsequent comparative genomic analysis.

The phylogeny of birds has always been a matter of great concern and controversy. Generally, researches on phylogenetic relationship of birds are much more based on mitochondrial DNA, cytochrome b gene, nuclear genes or a combination of them (Armstrong, Braun, & Kimball, 2001; Meng, Dai, Ran, Li, & Yue, 2008; Junaid Naseer et al., 2017). Jaiswal's study demonstrated that the phylogenetic trees of peafowl constructed by the data of nuclear intron regions and mitochondrial genomes supported that peafowl and chicken were clustered and closer than turkey (S. K. Jaiswal et al., 2018). Whereas, the phylogenetic tree based on retroposon insertion and the chromosomal showed that peafowl was closer to turkey than to chicken (Shibusawa et al., 2004). Similarly, Kaiser's research suggested that the result by using genome-based retrotransposon data to construct the taxonomic status of Phasianidae including chicken, turkey, Japanese quail and peafowl was consistent with Shibusawa's study, but was disaccord with the observation based on nucleotide sequence that the relationship of chicken and turkey was closer than peafowl (Kaiser, van Tuinen, & Ellegren, 2007). Many studies based on different data types to construct the tree proved that it was uncertain and controversial in the phylogenetic classification of Phasianidae families and may need to search for more evidence to verify. In this study, the single copy homologous amino acid sequence from the whole genome sequencing data was used to construct phylogenetic tree, and it was observed that the position of peafowl was closer to turkey than to chicken, which was disagreed with previous studies (Dhar et al., 2019; Shubham K. Jaiswal et al., 2018). In addition, according to the divergent time, it was

estimated that the divergence between peafowl and turkey was near to the divergence time among chicken and the ancestors of peafowl and turkey, which indicated that the relationship among them was very close, and inconsistent phylogenetic trees were constructed under different strategies.

In recent years, since the rapid development of genomics and the accumulation of genomic data, comparative genomics has become a research hotspot, which could explain biological functions and evolutionary characteristics at the genome-wide level. Especially, avian genomes were favour of people to investigate the adaptive evolution and species-specific biological characteristics by discovering novel genes and gene function through comparative genomics analysis. In this study, comparative genomics analysis was conducted on peafowls and other birds to explore the uniquely biological characteristics of peafowl during the evolution.

Species-specific immune-related genes were always positively selected in the adaptive evolution in many species. In this study, the number of GO terms and pathways related to immunity in peafowl was more than that of others, such as the expansive genes and rapidly evolving genes involved in the process of MHC class I protein binding, TNF signaling pathway, NF-kappaB signaling, IL-17 signaling pathway and Th17 cell differentiation. Likewise, it was observed that many olfactory genes and myosin genes were lost in peafowl. Mysosin was a functional protein and structural protein, which directly regulated muscle contraction, movement and cardiac functions in animals (Harrington & Rodgers, 1984). Olfaction performed a crucial role in life of birds, which contributed to recognise foods, make courtship, or detect danger (Khan et al., 2015; Lu, Wang, Lei, Yu, & Zhao, 2016). It was reported that birds could recognize close relatives to avoid inbreeding and distinguish direction in migration by using their acute sense of olfaction (Holland et al., 2009; Krause, Krüger, Kohlmeier, & Caspers, 2012). However, in this study, peafowl was artificial farmed and the feed-derived food was sufficient, causing the gradual degradation of the ability to find food in the wild and fly, which may explain the loss of myosin family and olfactory family and contribute to inbreeding and reduction of energy expenditure.

Generally, most of the body size of birds was small due to the pressure of body weight and reduction in energy expenditure (Blackburn & Gaston, 1994). However, the peafowl is known to have large body size, huge tail, and beautiful plumage, which is



likely to be gradually evolved due to better adaption to ecological environment. In this study, the enrichment analysis of positive selection genes was mainly involved in skeletal development, bone morphology, energy metabolism and storage, such as the mTOR signaling pathway, MAPK signaling pathway, BMP signaling pathway, limb development, lipid droplet, and lipid binding. mTOR was a central integrator of cellular growth and metabolism, and the mTOR signaling pathway played a vital role in innate and adaptive immune responses and regulating energy balance (Jones & Pearce, 2017; J. Xu, Ji, & Yan, 2012). BMP was an important member of the transforming growth factor- $\beta$  (TGF- $\beta$ ) superfamily through regulating the activity of downstream genes to participate in many important biological processes, such as nervous system differentiation, tooth and bone development, and cancer (Buijs et al., 2007; Huang, Zhong, Post, & Karperien, 2018). MAPK signalling pathway also participated in the regulation of feather growth and development (Fang et al., 2018). Moreover, as the rapid evolving genes, *MEF2C* exhibited many functions such as bone and muscle development, immune, and melanocyte differentiation (C. F. Liu, Samsa, Zhou, & Lefebvre, 2017; Tang et al., 2016; Trajanoska, Rivadeneira, Kiel, & Karasik, 2019). Furthermore, the glittering plumage and long tail also deeply attracted people's attention. Many positively selected genes associated with pigmentation like *TYR*, *SZT2*, *NF1*, *ARCNI*, *KIT*, *HPS5*, *FIG4*, *LYST*, *RACK1*, *USP13*, *Hps6*, *OCA2*, *MITF* and *BCL2* were also identified. All of the above results could be well contributed to understanding the phenotypic characteristics, such as large body size, long tail, and dazzling plumage in peafowl during the adaptive evolutionary process.

To date, a number of researches about the genetic mechanism of plumage colour in avian have been reported (Domyan et al., 2014; Robic et al., 2019). In the present study, the mechanism of the formation of white plumage phenotype was explored in peafowl combined transcriptome analysis and RT-qPCR with resequencing data. As for the result of allele frequency difference, the *PMEL* and *EDNRB* related to melanin were screened. Meanwhile, it was detected that the *PMEL* and *EDNRB* were contained in the DEGs from the transcriptome analysis, and it was found that the enrichment of DEGs was also involved in melanin biosynthetic process, which was in accordance with the result of allele frequency difference. Then, it was identified that the *PMEL* and *EDNRB* were the candidate genes of white plumage through the results of the allele frequency

difference of DEGs sites. Combined with the difference of the *PMEL* and *EDNRB* transcript by the results of IGV visualization and RT-qPCR, it was confirmed that *PMEL* was the causative gene of white plumage in blue and white feather peafowl. The formation and deposition of melanin mainly occurred on amyloid fibres of melanosomes. As a key signal molecule, *PMEL* could directly initiate the formation of melanosomes and promote the synthesis of melanosomes (Watt, van Niel, Raposo, & Marks, 2013). Moreover, it was reported in many studies that mutations of *PMEL* could cause low expression of *PMEL* leading to melanogenesis, further resulting in hypopigmentation phenotypes in animals like silver horse, white chicken and yellowish Japanese quail (Andersson et al., 2013; Ishishita et al., 2018; Kerje et al., 2004). In this study, it was also detected that the low expression of *PMEL* was associated with the white plumage in peafowl. However, in order to further investigate the causal mutations of *PMEL* low-expressed, the mutations of *PMEL* were examined and annotated, but not any functional mutation sites was found. It was supposed that the low expression of *PMEL* transcript was probably caused by the changes of the regulatory elements located in the upstream 5 kb promoter region of *PMEL*, and then the melanin synthesis was impeded. Unfortunately, there were no mutations in the core promoter region and transcription factor binding sites predicted by promoter prediction websites. In addition, the resequencing data were also used to detect the structural variation of the *PMEL* gene and its upstream, and the transcriptome data were used to detect the SNP and Indel variation, as well as PCR amplification of the *PMEL* gene and its upstream 5 kb promoter region for Sanger sequencing, but none possible variations was found. In view of this, it was speculated that the *PMEL* gene was likely to exist as a complex structure so that it couldn't be completely measured through sequencing and the causal sites weren't identified, which was needed to be further explored. Nevertheless, it was firstly identified that *PMEL* was a causal gene of white plumage, providing a novel insight into the formation of white feather phenotype in blue and white feather peafowl. The results revealed the genetic mechanism of white plumage at the whole-genome transcriptome level.

## **Conclusion**

In conclusion, for the first time, the study accomplished the assembly of a higher quality and greater sequencing depth of peafowl genome closer to the chromosomal level.

Firstly, the assembly genome is superior to other two draft genomes of peafowl either in terms of the sequencing depth or the assembly quality. Secondly, the study determined the phylogenetic relationship of peafowl among the Phasianidae, and found that peafowl was closer to turkey than to chicken by the analysis of single-copy orthologues genes based on the above assembly draft genome. Moreover, the comparative genome analysis indicated that Indian peafowl evolved on metabolism, immunity, skeletal development and feather development may approximatively be related to the unique characteristics of peafowl in domestication, which was conducted to provide baseline information about the phenotypic evolution of peafowl. Finally, the study was the first report to perform a combination of resequencing and transcriptome analysis in Indian peafowl, and revealed the molecular mechanism of white plumage formation in peafowl. Altogether, this study work provided a novel reference genome in the systematic evolution of peafowl and other birds, which was helpful to understand the formation of plumage coloration in peafowl and other colorful birds and provided a new idea for artificial breeding in peafowl.

## **Data availability**

The whole genome sequence data reported in this paper have been deposited in the Genome Warehouse in National Genomics Data Center ("Database Resources of the National Genomics Data Center in 2020," 2020), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number GWHAZTP000000000 that is publicly accessible at <https://bigd.big.ac.cn/gwh>. The resequencing raw data has been deposited in the NCBI Sequence Read Archive (SRA) (<https://submit.ncbi.nlm.nih.gov/subs/sra/>) under accession number PRJNA665082. The transcriptomic raw data has been deposited in the NCBI under accession number PRJNA661158.

## **Author Contributions**

X.Y., H.M. and J.R. designed the study and wrote the paper. S.L. and H.C. analyzed the data and wrote the paper. S.L., H.C., H.M. and W.L. revised the paper. S.L., H.Z. and B.L. conducted the validated experiments. S.L., J.O., M.H., S.Z., S.X., H.T., Y.G.,

Y.X., D.C., K.C., H.M. and Y.X. collected samples and performed the sequencing and genotyping experiments. All authors contributed and approved the final manuscript.

## **Funding**

This work was supported by Educational Commission of Jiangxi Province of China (No. GJJ190177) and by the Key Research and Development Program of Jiangxi Province of China (No. 20171BBF60003).

## **Ethics approval and consent to participate**

All procedures used for this study and involved in animals fully complied with guidelines for the care and utility of experimental animals established by the Ministry of Agriculture of China. The Animal Care and Use Committee of the South China Agricultural University approved this study.

## **Consent for publication**

Not applicable.

## **Competing financial interests**

The authors declare that they have no competing financial interests.

## **References**

- International Chicken Genome Sequencing Consortium Adey, A., Kitzman, J. O., Burton, J. N., Daza, R., Kumar, A., Christiansen, L., . . . Shendure, J. (2014). In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Res*, 24(12), 2041-2049. doi:10.1101/gr.178319.114
- Alioto, T., Blanco, E., Parra, G., & Guigó, R. (2018). Using geneid to Identify Genes. *Curr Protoc Bioinformatics*, 64(1), e56. doi:10.1002/cpbi.56
- Andersson, L. S., Wilbe, M., Viluma, A., Cothran, G., Ekesten, B., Ewart, S., & Lindgren, G. (2013). Equine multiple congenital ocular anomalies and silver coat colour result from the pleiotropic effects of mutant PMEL. *PLoS One*, 8(9), e75639. doi:10.1371/journal.pone.0075639
- Armstrong, M. H., Braun, E. L., & Kimball, R. T. (2001). Phylogenetic Utility of Avian Ovomucoid Intron G: A Comparison of Nuclear and Mitochondrial Phylogenies in Galliformes. *118 %J The Auk*(3), 799-804, 796.

808 Arnold, M. A., Kim, Y., Czubryt, M. P., Phan, D., McAnally, J., Qi, X., Olson, E. N. (2007).  
 809 MEF2C transcription factor controls chondrocyte hypertrophy and bone development.  
 810 *Dev Cell*, 12(3), 377-389. doi:10.1016/j.devcel.2007.02.004  
 811 Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Sherlock, G.  
 812 (2000). Gene ontology: tool for the unification of biology. The Gene Ontology  
 813 Consortium. *Nature genetics*, 25(1), 25-29. doi:10.1038/75556  
 814 Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its  
 815 supplement TrEMBL in 2000. *Nucleic Acids Res*, 28(1), 45-48.  
 816 doi:10.1093/nar/28.1.45  
 817 Bellott, D. W., Skaletsky, H., Cho, T. J., Brown, L., Locke, D., Chen, N., Page, D. C. (2017).  
 818 Avian W and mammalian Y chromosomes convergently retained dosage-sensitive  
 819 regulators. *Nature genetics*, 49(3), 387-394. doi:10.1038/ng.3778  
 820 Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic*  
 821 *Acids Res*, 27(2), 573-580. doi:10.1093/nar/27.2.573  
 822 Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Res*, 14(5),  
 823 988-995. doi:10.1101/gr.1865504  
 824 Blackburn, T., & Gaston, K. (1994). The Distribution of Body Sizes of the World's Bird Species.  
 825 *Oikos*, 70, 127-130. doi:10.2307/3545707  
 826 Buijs, J. T., Henriquez, N. V., van Overveld, P. G., van der Horst, G., ten Dijke, P., & van der  
 827 Pluijm, G. (2007). TGF-beta and BMP7 interactions in tumour progression and bone  
 828 metastasis. *Clin Exp Metastasis*, 24(8), 609-617. doi:10.1007/s10585-007-9118-2  
 829 Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA.  
 830 *J Mol Biol*, 268(1), 78-94. doi:10.1006/jmbi.1997.0951  
 831 Bush, K. L., & Strobeck, C. (2003). Phylogenetic relationships of the phasianidae reveals  
 832 possible non-pheasant taxa. *J Hered*, 94(6), 472-489. doi:10.1093/jhered/esg092  
 833 Cai, Q., Qian, X., Lang, Y., Luo, Y., Xu, J., Pan, S., Wang, J. (2013). Genome sequence of  
 834 ground tit *Pseudopodoces humilis* and its adaptation to high altitude. *Genome biology*,  
 835 14(3), R29. doi:10.1186/gb-2013-14-3-r29

836 Cai, T., Fjeldså, J., Wu, Y., Shao, S., Chen, Y., Quan, Q., Lei, F. (2017). What makes the Sino-  
837 Himalayan mountains the major diversity hotspots for pheasants? *Journal of*  
838 *Biogeography*. doi:10.1111/jbi.13156

839 Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for  
840 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*  
841 *(Oxford, England)*, 25(15), 1972-1973. doi:10.1093/bioinformatics/btp348

842 Chang, J., Wang, B., Zhang, Y. Y., Liu, Y., Liang, W., Wang, J. C., Zhang, Z. W. (2008).  
843 Molecular evidence for species status of the endangered Hainan peacock pheasant.  
844 *Zoolog Sci*, 25(1), 30-35. doi:10.2108/zsj.25.30

845 Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools:  
846 An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol*  
847 *Plant*, 13(8), 1194-1202. doi:10.1016/j.molp.2020.06.009

848 Chen, Y., Ye, W., Zhang, Y., & Xu, Y. (2015). High speed BLASTN: an accelerated  
849 MegaBLAST search tool. *Nucleic Acids Res*, 43(16), 7762-7768.  
850 doi:10.1093/nar/gkv784

851 Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Morales-  
852 Cruz, A. J. N. m. (2016). Phased diploid genome assembly with single-molecule real-  
853 time sequencing. *13*(12), 1050-1054.

854 Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Korlach, J.  
855 (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT  
856 sequencing data. *Nat Methods*, 10(6), 563-569. doi:10.1038/nmeth.2474

857 Church, D. M., Schneider, V. A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Hubbard,  
858 T. (2011). Modernizing reference genome assemblies. *PLoS Biol*, 9(7), e1001091.  
859 doi:10.1371/journal.pbio.1001091

860 Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Ruden, D. M. (2012).  
861 A program for annotating and predicting the effects of single nucleotide  
862 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118;  
863 iso-2; iso-3. *Fly (Austin)*, 6(2), 80-92. doi:10.4161/fly.19695

864 Dakin, R., McCrossan, O., Hare, J. F., Montgomerie, R., & Amador Kane, S. (2016).  
865 Biomechanics of the Peacock's Display: How Feather Structure and Resonance  
866 Influence Multimodal Signaling. *PLoS One*, 11(4), e0152759.  
867 doi:10.1371/journal.pone.0152759

868 Dalloul, R. A., Long, J. A., Zimin, A. V., Aslam, L., Beal, K., Blomberg Le, A., Reed, K. M.  
869 (2010). Multi-platform next-generation sequencing of the domestic turkey (*Meleagris*  
870 *gallopavo*): genome assembly and analysis. *PLoS Biol*, 8(9).  
871 doi:10.1371/journal.pbio.1000475

872 Database Resources of the National Genomics Data Center in 2020. (2020). *Nucleic Acids Res*,  
873 48(D1), D24-d33. doi:10.1093/nar/gkz913

874 De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool  
875 for the study of gene family evolution. *Bioinformatics (Oxford, England)*, 22(10),  
876 1269-1271. doi:10.1093/bioinformatics/btl097

877 Dhar, R., Seethy, A., Pethusamy, K., Singh, S., Rohil, V., Purkayastha, K., Karmakar, S. (2019).  
878 De novo assembly of the Indian blue peacock (*Pavo cristatus*) genome using Oxford  
879 Nanopore technology and Illumina sequencing. *GigaScience*, 8(5), giz038.  
880 doi:10.1093/gigascience/giz038

881 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Gingeras, T. R.  
882 (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*,  
883 29(1), 15-21. doi:10.1093/bioinformatics/bts635

884 Domyan, E. T., Guernsey, M. W., Kronenberg, Z., Krishnan, S., Boissy, R. E., Vickrey, A. I.,  
885 Shapiro, M. D. (2014). Epistatic and combinatorial effects of pigmentary gene  
886 mutations in the domestic pigeon. *Curr Biol*, 24(4), 459-464.  
887 doi:10.1016/j.cub.2014.01.020

888 Dong, C., Yang, X. Z., Zhang, C. Y., Liu, Y. Y., Zhou, R. B., Cheng, Q. D., Yin, D. C. (2017).  
889 Myocyte enhancer factor 2C and its directly-interacting proteins: A review. *Prog*  
890 *Biophys Mol Biol*, 126, 22-30. doi:10.1016/j.pbiomolbio.2017.02.002

891 El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Finn, R. D. (2019).  
892 The Pfam protein families database in 2019. *Nucleic Acids Res*, 47(D1), D427-d432.  
893 doi:10.1093/nar/gky995

894 Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Wolf, J. B.  
895 (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature*,  
896 491(7426), 756-760. doi:10.1038/nature11584

897 Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome  
898 comparisons dramatically improves orthogroup inference accuracy. *Genome biology*,  
899 16(1), 157-157. doi:10.1186/s13059-015-0721-2

900 Fang, G., Jia, X., Li, H., Tan, S., Nie, Q., Yu, H., & Yang, Y. (2018). Characterization of  
901 microRNA and mRNA expression profiles in skin tissue between early-feathering and  
902 late-feathering chickens. *BMC Genomics*, 19(1), 399. doi:10.1186/s12864-018-4773-z

903 Fischer, L., Herkner, C., Kitte, R., Dohnke, S., Riewaldt, J., Kretschmer, K., & Garbe, A. I.  
904 (2019). Foxp3(+) Regulatory T Cells in Bone and Hematopoietic Homeostasis. *Front*  
905 *Endocrinol (Lausanne)*, 10, 578. doi:10.3389/fendo.2019.00578

906 Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F.  
907 (2020). RepeatModeler2 for automated genomic discovery of transposable element  
908 families. *Proc Natl Acad Sci U S A*, 117(17), 9451-9457.  
909 doi:10.1073/pnas.1921046117

910 Friedman-Einat, M., Cogburn, L. A., Yosefi, S., Hen, G., Shinder, D., Shirak, A., & Seroussi,  
911 E. (2014). Discovery and characterization of the first genuine avian leptin gene in the  
912 rock dove (*Columba livia*). *Endocrinology*, 155(9), 3376-3384. doi:10.1210/en.2014-  
913 1273

914 Gadagkar, R. (2003). Is the peacock merely beautiful or also honest? *Current Science*, 85(7),  
915 1012-1020.

916 Gregory, M. K., & James, M. J. (2014). Functional characterization of the duck and turkey fatty  
917 acyl elongase enzymes ELOVL5 and ELOVL2. *J Nutr*, 144(8), 1234-1239.  
918 doi:10.3945/jn.114.194159



919 Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., & Bateman, A. (2005).  
920 Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*,  
921 33(Database issue), D121-124. doi:10.1093/nar/gki081

922 Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., Wortman, J. R. (2008).  
923 Automated eukaryotic gene structure annotation using EVidenceModeler and the  
924 Program to Assemble Spliced Alignments. *Genome biology*, 9(1), R7. doi:10.1186/gb-  
925 2008-9-1-r7

926 Harrington, W. F., & Rodgers, M. E. (1984). Myosin. *Annu Rev Biochem*, 53, 35-73.  
927 doi:10.1146/annurev.bi.53.070184.000343

928 Hedges, S. B., Dudley, J., & Kumar, S. (2006). TimeTree: a public knowledge-base of  
929 divergence times among organisms. *Bioinformatics (Oxford, England)*, 22(23), 2971-  
930 2972. doi:10.1093/bioinformatics/btl505

931 Holland, R. A., Thorup, K., Gagliardo, A., Bisson, I. A., Knecht, E., Mizrahi, D., & Wikelski,  
932 M. (2009). Testing the role of sensory systems in the migratory heading of a songbird.  
933 *J Exp Biol*, 212(Pt 24), 4065-4071. doi:10.1242/jeb.034504

934 Huang, X., Zhong, L., Post, J. N., & Karperien, M. (2018). Co-treatment of TGF- $\beta$ 3 and BMP7  
935 is superior in stimulating chondrocyte redifferentiation in both hypoxia and normoxia  
936 compared to single treatments. *Sci Rep*, 8(1), 10251. doi:10.1038/s41598-018-27602-  
937 y

938 Ishishita, S., Takahashi, M., Yamaguchi, K., Kinoshita, K., Nakano, M., Nunome, M., Matsuda,  
939 Y. (2018). Nonsense mutation in PMEL is associated with yellowish plumage colour  
940 phenotype in Japanese quail. *Sci Rep*, 8(1), 16732. doi:10.1038/s41598-018-34827-4

941 Jaiswal, S. K., Gupta, A., Saxena, R., Prasoodanan, V. P. K., Sharma, A. K., Mittal, P., Sharma,  
942 V. K. (2018). Genome Sequence of Peacock Reveals the Peculiar Case of a Glittering  
943 Bird. *Frontiers in genetics*, 9, 392-392. doi:10.3389/fgene.2018.00392

944 Jaiswal, S. K., Gupta, A., Saxena, R., Prasoodanan, V. P. K., Sharma, A. K., Mittal, P., Sharma,  
945 V. K. (2018). Genome Sequence of Peacock Reveals the Peculiar Case of a Glittering  
946 Bird. *Frontiers in genetics*, 9, 392. doi:10.3389/fgene.2018.00392

947 Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Zhang, G. (2014). Whole-  
 948 genome analyses resolve early branches in the tree of life of modern birds. *Science*  
 949 (*New York, N.Y.*), 346(6215), 1320-1331. doi:10.1126/science.1253451

950 Jones, R. G., & Pearce, E. J. (2017). MenTORing Immunity: mTOR Signaling in the  
 951 Development and Function of Tissue-Resident Immune Cells. *Immunity*, 46(5), 730-  
 952 742. doi:10.1016/j.immuni.2017.04.028

953 Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005).  
 954 Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*,  
 955 110(1-4), 462-467. doi:10.1159/000084979

956 Kaiser, V. B., van Tuinen, M., & Ellegren, H. (2007). Insertion events of CR1 retrotransposable  
 957 elements elucidate the phylogenetic branching order in galliform birds. *Mol Biol Evol*,  
 958 24(1), 338-347. doi:10.1093/molbev/msl164

959 Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic*  
 960 *Acids Res*, 28(1), 27-30. doi:10.1093/nar/28.1.27

961 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version  
 962 7: improvements in performance and usability. *Mol Biol Evol*, 30(4), 772-780.  
 963 doi:10.1093/molbev/mst010

964 Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res*, 12(4), 656-664.  
 965 doi:10.1101/gr.229202

966 Kerje, S., Sharma, P., Gunnarsson, U., Kim, H., Bagchi, S., Fredriksson, R., Andersson, L.  
 967 (2004). The Dominant white, Dun and Smoky color variants in chicken are associated  
 968 with insertion/deletion polymorphisms in the PMEL17 gene. *Genetics*, 168(3), 1507-  
 969 1518. doi:10.1534/genetics.104.027995

970 Khan, I., Yang, Z., Maldonado, E., Li, C., Zhang, G., Gilbert, M. T., Antunes, A. (2015).  
 971 Olfactory Receptor Subgenomes Linked with Broad Ecological Adaptations in  
 972 Sauropsida. *Mol Biol Evol*, 32(11), 2832-2843. doi:10.1093/molbev/msv155

973 Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation  
 974 between populations using sequencing of pooled DNA samples (Pool-Seq).

975 *Bioinformatics* (Oxford, England), 27(24), 3435-3436.  
 976 doi:10.1093/bioinformatics/btr589  
 977 Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, 5, 59. doi:10.1186/1471-  
 978 2105-5-59  
 979 Korlach, J., Gedman, G., Kingan, S. B., Chin, C. S., Howard, J. T., Audet, J. N., Jarvis, E. D.  
 980 (2017). De novo PacBio long-read and phased avian genome assemblies correct and  
 981 add to reference genes generated with intermediate and short reads. *GigaScience*, 6(10),  
 982 1-16. doi:10.1093/gigascience/gix085  
 983 Krause, E. T., Krüger, O., Kohlmeier, P., & Caspers, B. A. (2012). Olfactory kin recognition  
 984 in a songbird. *Biol Lett*, 8(3), 327-329. doi:10.1098/rsbl.2011.1093  
 985 Kushwaha, S., & Kumar, A. J. J. W. R. (2016). A review on Indian peafowl (*Pavo cristatus*)  
 986 Linnaeus, 1758. 4(4), 42-59.  
 987 Li, C., Nguyen, V., Clark, K. N., Zahed, T., Sharkas, S., Filipp, F. V., & Boiko, A. D. (2019).  
 988 Down-regulation of FZD3 receptor suppresses growth and metastasis of human  
 989 melanoma independently of canonical WNT signaling. *Proc Natl Acad Sci U S A*,  
 990 116(10), 4548-4557. doi:10.1073/pnas.1813802116  
 991 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler  
 992 transform. *Bioinformatics* (Oxford, England), 25(14), 1754-1760.  
 993 doi:10.1093/bioinformatics/btp324  
 994 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Durbin, R. (2009). The  
 995 Sequence Alignment/Map format and SAMtools. *Bioinformatics* (Oxford, England),  
 996 25(16), 2078-2079. doi:10.1093/bioinformatics/btp352  
 997 Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read  
 998 mapping by seed-and-vote. *Nucleic Acids Res*, 41(10), e108. doi:10.1093/nar/gkt214  
 999 Liu, C. F., Samsa, W. E., Zhou, G., & Lefebvre, V. (2017). Transcriptional control of  
 1000 chondrocyte specification and differentiation. *Semin Cell Dev Biol*, 62, 34-49.  
 1001 doi:10.1016/j.semcdb.2016.10.004  
 1002 Liu, H., Luo, Q., Zhang, J., Mo, C., Wang, Y., & Li, J. (2019). Endothelins (EDN1, EDN2,  
 1003 EDN3) and their receptors (EDNRA, EDNRB, EDNRB2) in chickens: Functional

analysis and tissue distribution. *Gen Comp Endocrinol*, 283, 113231.  
doi:10.1016/j.ygcen.2019.113231

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*, 25(5), 955-964.  
doi:10.1093/nar/25.5.955

Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. *Methods Mol Biol*, 1079, 155-170. doi:10.1007/978-1-62703-646-7\_10

Lu, Q., Wang, K., Lei, F., Yu, D., & Zhao, H. (2016). Penguins reduced olfactory receptor genes common to other waterbirds. *Sci Rep*, 6, 31671. doi:10.1038/srep31671

Mackie, E. J., Ahmed, Y. A., Tatarczuch, L., Chen, K. S., & Mirams, M. (2008). Endochondral ossification: how cartilage is converted into bone in the developing skeleton. *Int J Biochem Cell Biol*, 40(1), 46-62. doi:10.1016/j.biocel.2007.06.009

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics (Oxford, England)*, 20(16), 2878-2879. doi:10.1093/bioinformatics/bth315

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., DePristo, M. A. (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

Meng, Y., Dai, B., Ran, J., Li, J., & Yue, B. (2008). Phylogenetic position of the genus Tetraophaps (Aves, Galliformes, Phasianidae) as inferred from mitochondrial and nuclear sequences. *Biochemical Systematics and Ecology*, 36(8), 626-637.  
doi:https://doi.org/10.1016/j.bse.2008.01.007

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*, 37(5), 1530-1534.  
doi:10.1093/molbev/msaa015

Mohajeri, K., Cantsilieris, S., Huddleston, J., Nelson, B. J., Coe, B. P., Campbell, C. D., Eichler, E. E. (2016). Interchromosomal core duplicons drive both evolutionary instability and

1033 disease susceptibility of the Chromosome 8p23.1 region. *Genome Res*, 26(11), 1453-  
1034 1467. doi:10.1101/gr.211284.116

1035 Mushtaq-ul-Hassan, M., Ali, Z., Arshad, M. I., Mahmood, S., & Research, M. M.-u.-H. J. I. J.  
1036 o. V. (2012). Effects of mating sex ratios in Indian peafowl (*Pavo cristatus*) on  
1037 production performance at Wildlife Research Institute, Faisalabad (Pakistan). *13*(2),  
1038 143-146.

1039 Naseer, J., Anjum, K., Khan, W., Imran, M., Ishaque, M., Hafeez, S., Nazir, M. A. (2018).  
1040 Phylogenetic analysis based studies on genetic variation of cytochrome B gene of  
1041 Indian peafowl (*Pavo cristatus*) in Pakistan. *Indian Journal of Animal Research*, 52,  
1042 343-346. doi:10.18805/ijar.v0iOF.9119

1043 Naseer, J., Anjum, K. M., Khan, W. A., Imran, M., Ishaque, M., Hafeez, S., Nazir, M. A. J. I.  
1044 J. o. A. R. (2017). Phylogenetic analysis based studies on genetic variation of  
1045 Cytochrome b gene of Indian peafowl (*Pavo cristatus*) in Pakistan. *52*(OF), 343-346.

1046 Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches.  
1047 *Bioinformatics* (Oxford, England), 29(22), 2933-2935.  
1048 doi:10.1093/bioinformatics/btt509

1049 Nishibori, M., Hayashi, T., Tsudzuki, M., Yamamoto, Y., & Yasue, H. (2001). Complete  
1050 sequence of the Japanese quail (*Coturnix japonica*) mitochondrial genome and its  
1051 genetic relationship with related species. *Anim Genet*, 32(6), 380-385.  
1052 doi:10.1046/j.1365-2052.2001.00795.x

1053 O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Pruitt, K. D.  
1054 (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic  
1055 expansion, and functional annotation. *Nucleic Acids Res*, 44(D1), D733-745.  
1056 doi:10.1093/nar/gkv1189

1057 Oldeschulte, D. L., Halley, Y. A., Wilson, M. L., Bhattarai, E. K., Brashear, W., Hill, J.,  
1058 Seabury, C. M. (2017). Annotated Draft Genome Assemblies for the Northern  
1059 Bobwhite (*Colinus virginianus*) and the Scaled Quail (*Callipepla squamata*) Reveal  
1060 Disparate Estimates of Modern Genome Diversity and Historic Effective Population  
1061 Size. *G3 (Bethesda)*, 7(9), 3047-3058. doi:10.1534/g3.117.043083

- Ouyang, Y. N., Yang, Z. Y., Da-Lin, L. I., Huo, J. L., Qian, K., & Miao, Y. W. J. J. o. Y. A. U. (2009). Genetic Divergence between *Pavo muticus* and *Pavo cristatus* by Cyt b Gene.
- Paranjpe, D., & Dange, P. J. b. (2019). A tale of two species: human and peafowl interactions in human dominated landscape influence each others behaviour. 412254.
- Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics (Oxford, England)*, 23(9), 1061-1067. doi:10.1093/bioinformatics/btm071
- Parra, G., Bradnam, K., Ning, Z., Keane, T., & Korf, I. (2009). Assessing the gene space in draft genomes. *Nucleic Acids Res*, 37(1), 289-297. doi:10.1093/nar/gkn916
- Potthoff, M. J., Arnold, M. A., McAnally, J., Richardson, J. A., Bassel-Duby, R., & Olson, E. N. (2007). Regulation of skeletal muscle sarcomere integrity and postnatal muscle function by Mef2c. *Mol Cell Biol*, 27(23), 8143-8151. doi:10.1128/mcb.01187-07
- Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics (Oxford, England)*, 21 Suppl 1, i351-358. doi:10.1093/bioinformatics/bti1018
- Ramesh, K., & McGowan, P. (2009). On the current status of Indian Peafowl *Pavo cristatus* (Aves: Galliformes: Phasianidae): keeping the common species common. *Journal of Threatened Taxa*, 1, 106-108. doi:10.11609/JoTT.o1845.106-8
- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, 16(6), 276-277. doi:10.1016/s0168-9525(00)02024-2
- Robic, A., Morisson, M., Leroux, S., Gourichon, D., Vignal, A., Thebault, N., Pitel, F. (2019). Two new structural mutations in the 5' region of the ASIP gene cause diluted feather color phenotypes in Japanese quail. *Genet Sel Evol*, 51(1), 12. doi:10.1186/s12711-019-0458-6
- Samour, J., Naldo, J., Rahman, H., & Sakkir, M. (2010). Hematologic and plasma biochemical reference values in Indian peafowl (*Pavo cristatus*). *J Avian Med Surg*, 24(2), 99-106. doi:10.1647/2008-019.1
- Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. (2004). *Nature*, 432(7018), 695-716. doi:10.1038/nature03154

1091 Shapiro, M. D., Kronenberg, Z., Li, C., Domyan, E. T., Pan, H., Campbell, M., Wang, J. (2013).  
1092 Genomic diversity and evolution of the head crest in the rock pigeon. *Science (New*  
1093 *York, N.Y.)*, 339(6123), 1063-1067. doi:10.1126/science.1230422

1094 Shen, Y. Y., Dai, K., Cao, X., Murphy, R. W., Shen, X. J., & Zhang, Y. P. (2014). The updated  
1095 phylogenies of the phasianidae based on combined data of nuclear and mitochondrial  
1096 DNA. *PLoS One*, 9(4), e95786. doi:10.1371/journal.pone.0095786

1097 Shibusawa, M., Nishibori, M., Nishida-Umehara, C., Tsudzuki, M., Masabanda, J., Griffin, D.  
1098 K., & Matsuda, Y. (2004). Karyotypic evolution in the Galliformes: an examination of  
1099 the process of karyotypic evolution by comparison of the molecular cytogenetic  
1100 findings with the molecular phylogeny. *Cytogenet Genome Res*, 106(1), 111-119.  
1101 doi:10.1159/000078570

1102 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015).  
1103 BUSCO: assessing genome assembly and annotation completeness with single-copy  
1104 orthologs. *Bioinformatics (Oxford, England)*, 31(19), 3210-3212.  
1105 doi:10.1093/bioinformatics/btv351

1106 Somes, R. G., & Burger, R. E. (1993). Inheritance of the White and Pied Plumage Color  
1107 Patterns in the Indian Peafowl (*Pavo cristatus*). *I4*(1), 53-55.

1108 Somes, R. G., & Burger, R. E. J. J. o. H. (1991). Plumage Color Inheritance of the Indian Blue  
1109 Peafowl (*Pavo Cristatus*): Blue, Black-Shouldered, Cameo, and Oaten. (1), 1.

1110 Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006).  
1111 AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*, 34(Web  
1112 Server issue), W435-439. doi:10.1093/nar/gkl200

1113 Steiger, S. S., Kuryshv, V. Y., Stensmyr, M. C., Kempnaers, B., & Mueller, J. C. (2009). A  
1114 comparison of reptilian and avian olfactory receptor gene repertoires: species-specific  
1115 expansion of group gamma genes in birds. *BMC Genomics*, 10, 446. doi:10.1186/1471-  
1116 2164-10-446

1117 Stuart, C. A., Stone, W. L., Howell, M. E., Brannon, M. F., Hall, H. K., Gibson, A. L., & Stone,  
1118 M. H. (2016). Myosin content of individual human muscle fibers isolated by laser

capture microdissection. *Am J Physiol Cell Physiol*, 310(5), C381-389.  
doi:10.1152/ajpcell.00317.2015

Sun, K., Meiklejohn, K. A., Faircloth, B. C., Glenn, T. C., Braun, E. L., & Kimball, R. T. (2014).  
The evolution of peafowl and other taxa with ocelli (eyespot): a phylogenomic  
approach. *Proc Biol Sci*, 281(1790). doi:10.1098/rspb.2014.0823

Talha, M. M. H., Mia, M. M., & Momu, J. M. J. I. J. o. D. R. (2018). Morphometric, productive  
and reproductive traits of Indian peafowl (*Pavo cristatus*) in Bangladesh. 8(02), 19039-  
19043.

Tang, R., Xu, X., Yang, W., Yu, W., Hou, S., Xuan, Y., . . . Deng, W. (2016). MED27 promotes  
melanoma growth by targeting AKT/MAPK and NF- $\kappa$ B/iNOS signaling pathways.  
*Cancer Lett*, 373(1), 77-87. doi:10.1016/j.canlet.2016.01.005

Tempel, S. (2012). Using and understanding RepeatMasker. *Methods Mol Biol*, 859, 29-51.  
doi:10.1007/978-1-61779-603-6\_2

Thoenen, E., Curl, A., & Iwakuma, T. (2019). TP53 in bone and soft tissue sarcomas.  
*Pharmacol Ther*, 202, 149-164. doi:10.1016/j.pharmthera.2019.06.010

Toepfer, C. N., Garfinkel, A. C., Venturini, G., Wakimoto, H., Repetti, G., Alamo, L., Seidman,  
C. E. (2020). Myosin Sequestration Regulates Sarcomere Function, Cardiomyocyte  
Energetics, and Metabolism, Informing the Pathogenesis of Hypertrophic  
Cardiomyopathy. *Circulation*, 141(10), 828-842.  
doi:10.1161/circulationaha.119.042339

Trajanoska, K., Rivadeneira, F., Kiel, D. P., & Karasik, D. (2019). Genetics of Bone and Muscle  
Interactions in Humans. *Curr Osteoporos Rep*, 17(2), 86-95. doi:10.1007/s11914-019-  
00505-1

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Earl, A. M.  
(2014). Pilon: an integrated tool for comprehensive microbial variant detection and  
genome assembly improvement. *PLoS One*, 9(11), e112963.  
doi:10.1371/journal.pone.0112963



1146 Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Paterson, A. H. (2012). MCSanX:  
 1147 a toolkit for detection and evolutionary analysis of gene synteny and collinearity.  
 1148 *Nucleic Acids Res*, 40(7), e49. doi:10.1093/nar/gkr1293

1149 Wang, Y., Zhao, H., Liu, J., Shao, Y., & Xing, M. (2019). Molecular cloning and transcriptional  
 1150 regulation of Indian peafowl (*Pavo cristatus*) IFN- $\alpha$  gene. *Cell Stress Chaperones*,  
 1151 24(2), 323-332. doi:10.1007/s12192-018-00962-0

1152 Watt, B., van Niel, G., Raposo, G., & Marks, M. S. (2013). PMEL: a pigment cell-specific  
 1153 model for functional amyloid formation. *Pigment Cell Melanoma Res*, 26(3), 300-315.  
 1154 doi:10.1111/pcmr.12067

1155 Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Wei, L. (2011). KOBAS 2.0: a web  
 1156 server for annotation and identification of enriched pathways and diseases. *Nucleic*  
 1157 *Acids Res*, 39(Web Server issue), W316-322. doi:10.1093/nar/gkr483

1158 Xu, J., Ji, J., & Yan, X. H. (2012). Cross-talk between AMPK and mTOR in regulating energy  
 1159 balance. *Crit Rev Food Sci Nutr*, 52(5), 373-381. doi:10.1080/10408398.2010.500245

1160 Xu, Z., & Wang, H. (2007). LTR\_FINDER: an efficient tool for the prediction of full-length  
 1161 LTR retrotransposons. *Nucleic Acids Res*, 35(Web Server issue), W265-268.  
 1162 doi:10.1093/nar/gkm286

1163 Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, 24(8),  
 1164 1586-1591. doi:10.1093/molbev/msm088

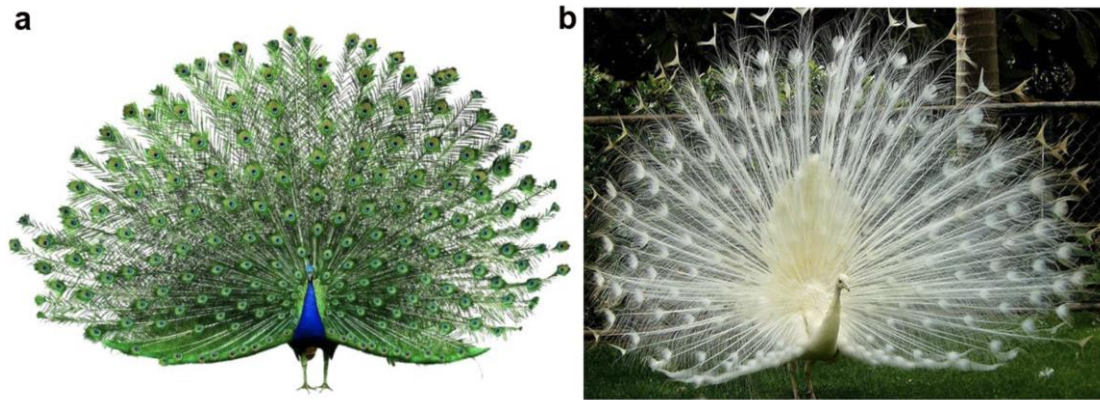
1165 Yu, C., Zavaljevski, N., Desai, V., & Reifman, J. (2011). QuartetS: a fast and accurate algorithm  
 1166 for large-scale orthology detection. *Nucleic Acids Res*, 39(13), e88.  
 1167 doi:10.1093/nar/gkr308

1168 Zdobnov, E. M., & Apweiler, R. (2001). InterProScan--an integration platform for the  
 1169 signature-recognition methods in InterPro. *Bioinformatics (Oxford, England)*, 17(9),  
 1170 847-848. doi:10.1093/bioinformatics/17.9.847

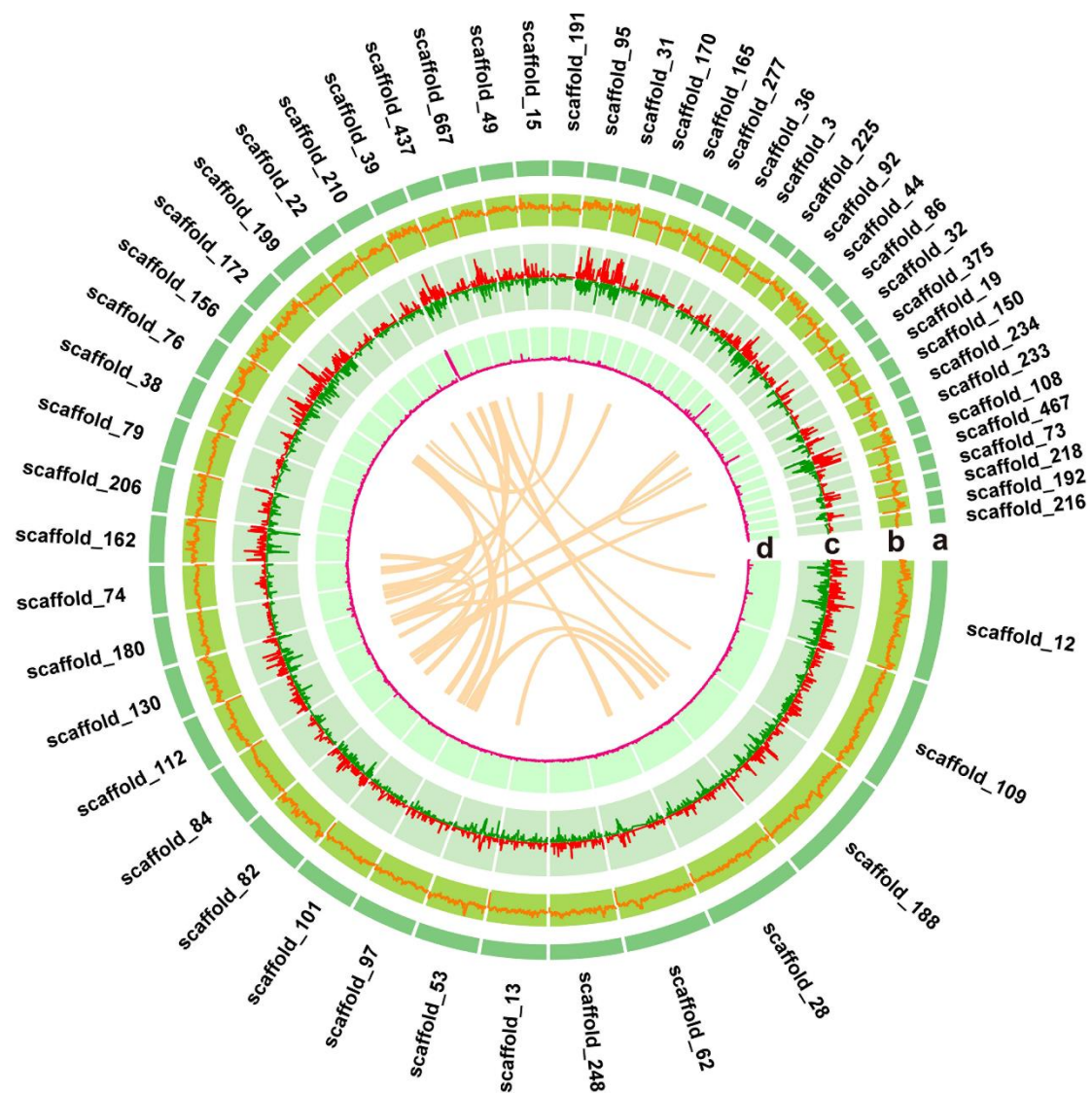
1171 Zhan, X., Pan, S., Wang, J., Dixon, A., He, J., Muller, M. G., Bruford, M. W. (2013). Peregrine  
 1172 and saker falcon genome sequences provide insights into evolution of a predatory  
 1173 lifestyle. *Nature genetics*, 45(5), 563-566. doi:10.1038/ng.2588

- Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Wang, J. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science (New York, N.Y.)*, 346(6215), 1311-1320. doi:10.1126/science.1251385
- Zhang, Z., Jia, Y., Almeida, P., Mank, J. E., van Tuinen, M., Wang, Q., Qu, L. (2018). Whole-genome resequencing reveals signatures of selection and timing of duck domestication. *GigaScience*, 7(4). doi:10.1093/gigascience/giy027
- Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, G. K., & Yu, J. (2006). KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*, 4(4), 259-263. doi:10.1016/s1672-0229(07)60007-2
- Zhou, T. C., Sha, T., Irwin, D. M., & Zhang, Y. P. (2015). Complete mitochondrial genome of the Indian peafowl (*Pavo cristatus*), with phylogenetic analysis in phasianidae. *Mitochondrial DNA*, 26(6), 912-913. doi:10.3109/19401736.2013.863287

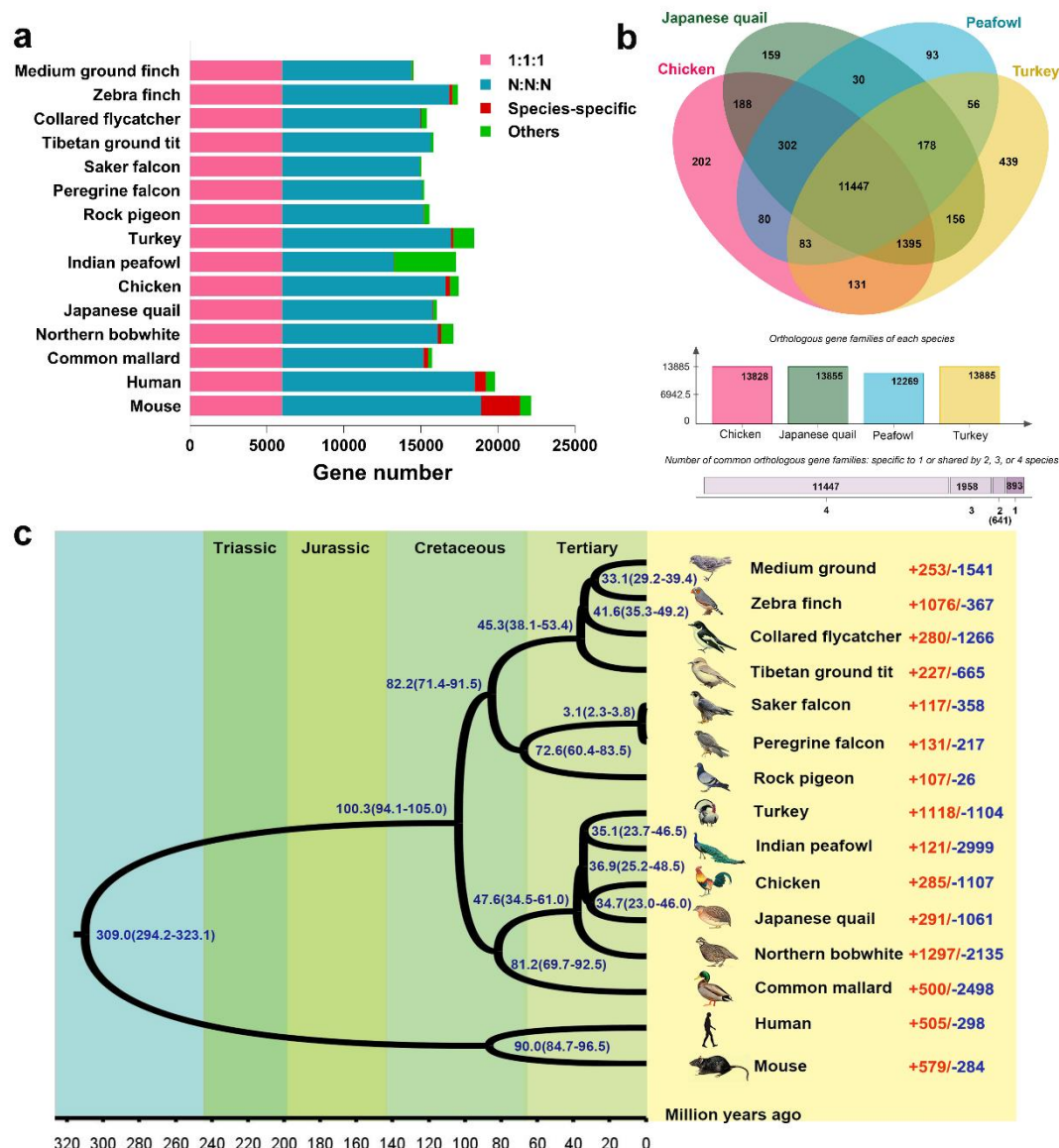
## Figures



**Fig. 1. Photographs of the Indian blue peafowl and white peafowl.** The Indian blue peafowl and white peafowl showed in **a** and **b**, respectively.



**Fig. 2. The global maps of de novo genome assembly of Indian peafowl.** a, 55 scaffolds with the length greater than 5 Mb (scaffolds N70) of the assembled Indian blue peafowl. The perimeter of ring represented the length of scaffolds, and the light orange links in the middle circle indicated the synteny in the peafowl genome. The GC density, gene density and tandem repeat sequence density of peafowl genome were displayed in b, c and d, respectively. And the red and green bars in the gene density diagram represented the positive strand (+) and negative strand (-) in peafowl genome.

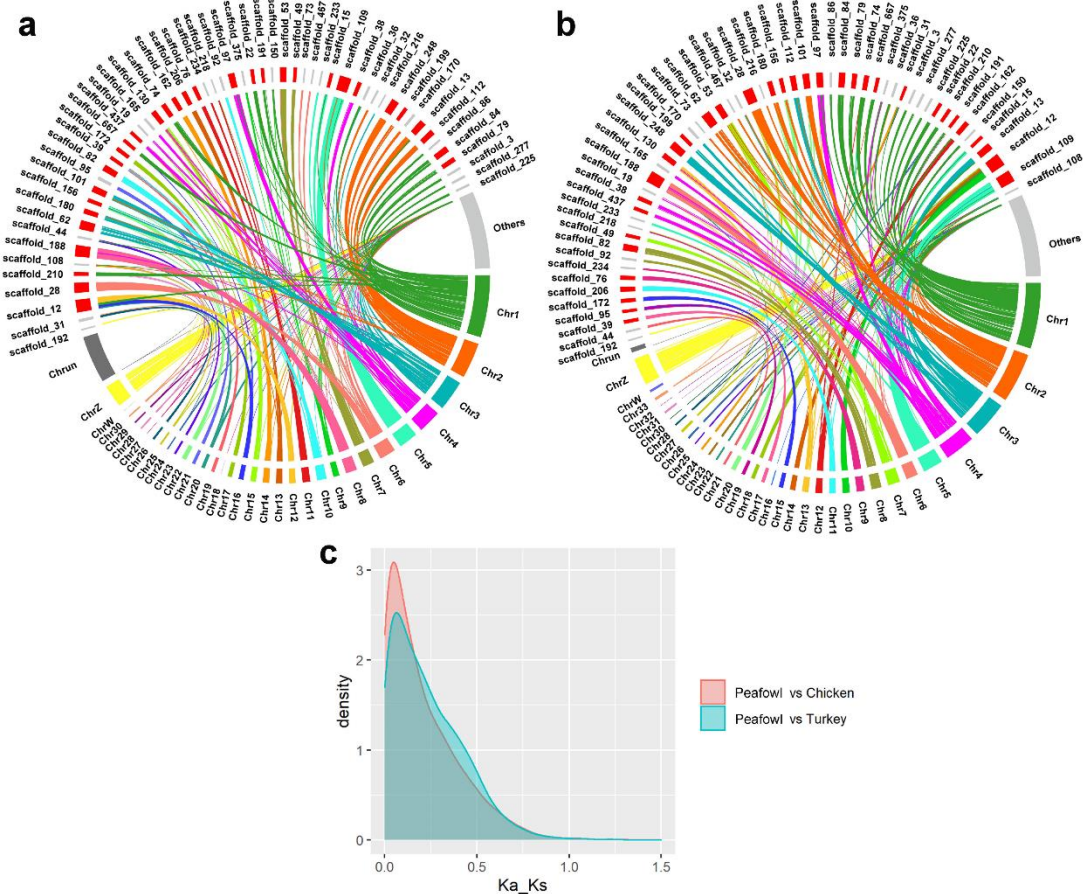


**Fig. 3. Gene family and genome evolution among the peafowl and other 14 species.**

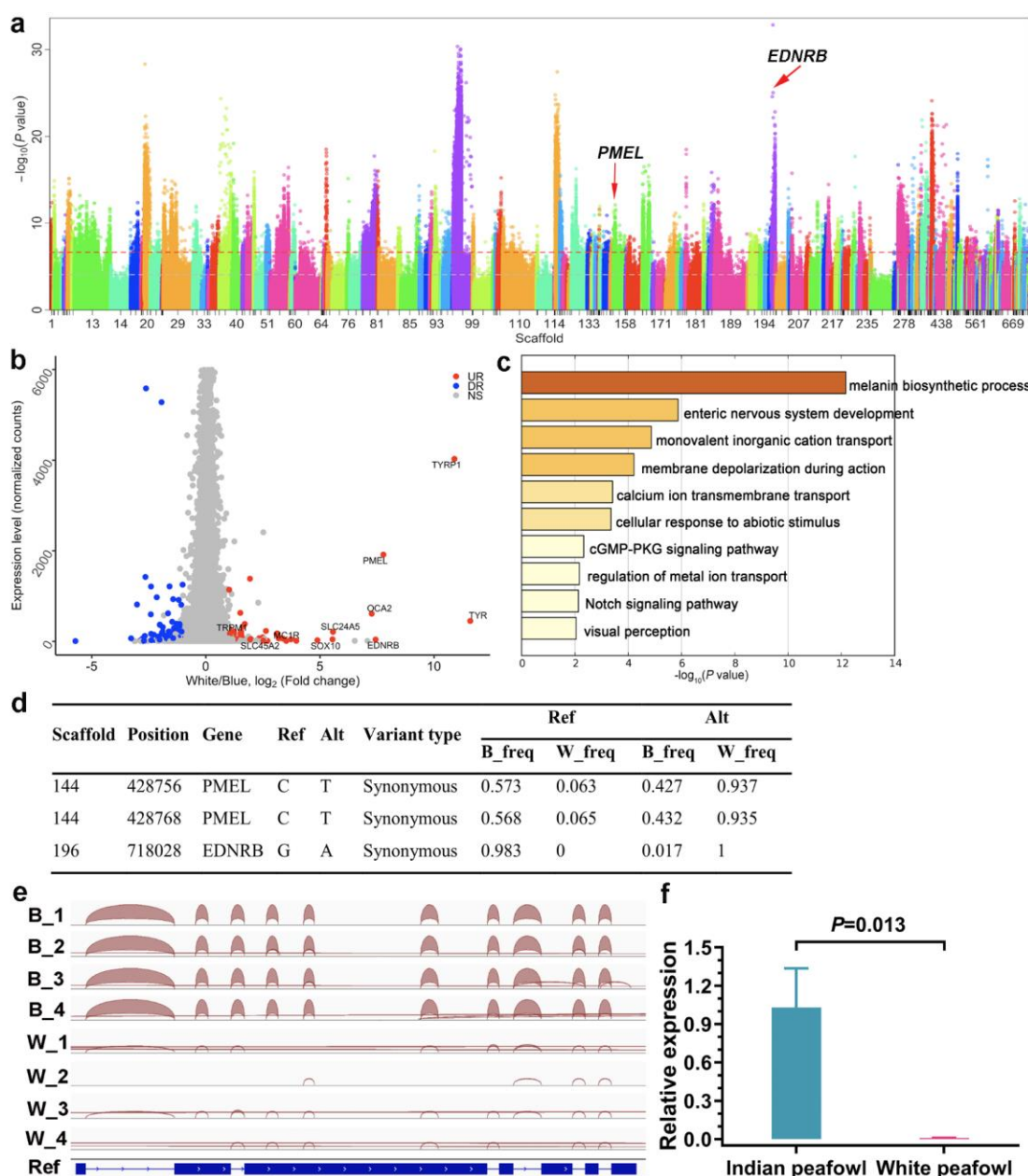
**a**, Statistics of orthologs among 15 species. “1:1:1” indicated the single-copy orthologs were shared by 15 species with one copy. “N:N:N” represented any other orthologous group (missing in one species). Species-specific showed the specific orthologs in each species. Other orthologs were unclustered into gene families. **b**, Venn diagram of the shared orthologous gene families among the Phasianidae species (peafowl, Japanese quail, chicken and turkey). The numbers represented the unique or common gene family among the species. **c**, The phylogenetic relationship tree among 15 species was constructed by maximum likelihood with JTT model based on the single-copy orthologous sequences, human and mouse as outgroups. The divergence time of species was estimated by five calibration time from TimeTree database, including human-

mouse (85~97Mya), human-zebra finch (294~323Mya), zebra finch-medium ground finch (30.4~46.8Mya), common mallard-zebra finch (93.2~104.6Mya) and saker falcon-peregrine falcon (1.66~3.68Mya). Of them, the divergence time of human and mouse was used as a timeline at the bottom of the figure which was divided into four different periods such as tertiary, cretaceous, jurassic and triassic, and showed by different colours. In addition, the expansion and contraction of gene family in 15 species were showed at the right of species name. The red (+) and blue (-) numbers represented the expanded and contracted genes, respectively.





**Fig. 4. Genome synteny and collinearity among the Indian peafowl, chicken, and turkey.** **a**, A syntentic map of the peafowl and turkey genomes. The perimeter of ring represented the length of chromosomes labeled by different colours or scaffolds. It displayed the scaffolds with the length greater than 5 Mb (scaffolds N70) of the assembled Indian blue peafowl, of which, the first 32 scaffolds were shown in red and other scaffolds were marked in gray. **b**, A syntentic map of the peafowl and chicken genomes. The first 35 scaffolds were shown in red, and other scaffolds were marked in gray. **c**, The distribution of Ka/Ks ratio in the genomes of peafowl, turkey, and chicken.



**Fig. 5. Causal genes for white feather in blue and white peafowl.** **a**, Allele frequency differences between blue and white peafowl. Scaffolds were distinguished by different colours. The candidate SNPs along with causal genes were marked by arrows, including *EDNRB* and *PMEL*. **b**, Differentially expressed genes (DEGs) related to the plumage pigmentation. The red and blue dots were used to mark the up- and down-regulated genes in blue and white peafowl, respectively. A total of 69 down-regulated genes and 52 up-regulated genes were identified, of which ten up-regulated genes were associated with the melanin deposition, and so marked out. **c**, KEGG and GO enrichment of DEGs related to the plumage pigmentation in blue and white peafowl. The darker the colour



was, the more significant the difference. The top significant pathway was enriched in the process of the melanin synthesis based on the criterion of  $P < 0.05$  as significant. **d**, Allele frequencies of differentially expressed genes in blue and white peafowl. B: blue peafowl, W: white peafowl. An observation showed that the top two differential sites were located in *PMEL* and *EDNRB*. **e**, *PMEL* transcripts in the feather pulp of blue and white peafowl. B: blue peafowl, W: white peafowl. The RNA sequencing reads of *PMEL* were aligned to the assembly peafowl genome in the feather tissue of blue and white peafowl. The red arc represented the mRNA expression level of *PMEL*. Apparently, *PMEL* was normally expressed in blue peafowl but almost not expressed in white peafowl. **f**, RT-qPCR of *PMEL* transcripts in the feather pulp of blue and white peafowl. The result indicated that the relative expression of *PMEL* was significantly decreased in white peafowl compared to the mRNA expression of *PMEL* in blue peafowl ( $P = 0.013$ ).

## Tables

**Table 1. Quality metrics for the peafowl genome assembly generated in the current work and for other peafowl genome assemblies published in previous studies.**

Items	This study	Shubham et al. (2018)	Ruby et al. (2019)
Sequencing technology	Illumina NovaSeq 6000, PacBio RS-II, 10X Genomics, Chicoga	Illumina NextSeq 500	Illumina HiSeq, ONT
Total sequencing depth	362×	136×	236×
Total scaffolds	726	98,687	179,332
Scaffolds N50 (bp)	11,421,185	25,613	190,304
Contigs N50 (bp)	6,188,159	19,387	103,131
Longest scaffold length (bp)	38,857,732	286,113	2,488,982
Total sequence length (bp)	1,046,718,946	1,137,150,029	1,027,510,962
Total number of predicted protein-coding genes	19,465	15,970	23,153

## **Supplementary materials**

**Supplementary Figure S1.** Pipeline of the draft genome assembly of Indian blue peafowl

**Supplementary Figure S2.** Workflow of the genome annotation of Indian blue peafowl

**Supplementary Figure S3.** 17-kmer frequency distribution of peafowl genome

**Supplementary Figure S4.** Divergence distribution of transposable element of peafowl genome by using RepeatMasker software

**Supplementary Figure S5.** Phylogenetic tree of 15 species constructed with IQ-tree

**Supplementary Figure S6.** Phylogenetic tree of 15 species constructed with RAxML

**Supplementary Figure S7.** *EDNRB* transcripts in the feather tissue of peafowl by IGV visualization

**Supplementary Table S1.** Statistics of genome assembly data of peafowl

**Supplementary Table S2.** Summary of de novo genome assembly of peafowl

**Supplementary Table S3.** Percentage of the base contents of peafowl genome

**Supplementary Table S4.** Statistics of paired-end reads mapping in peafowl genome

**Supplementary Table S5.** Number of SNPs of peafowl genome

**Supplementary Table S6.** Assembly assessment of completeness by using BUSCOs

**Supplementary Table S7.** Whole genome repetitive sequences of Indian peafowl genome predicted by homologous alignment and de novo search

**Supplementary Table S8.** Composition of repetitive sequences in peafowl genome

**Supplementary Table S9.** Prediction of protein-coding genes for peafowl genome

**Supplementary Table S10.** Statistics of functional annotation of protein-coding genes in the peafowl genome assembly

**Supplementary Table S11.** Statistics of non-coding RNAs in the assembly of peafowl

**Supplementary Table S12.** Functional enrichment of species-specific genes in peafowl compared with the Phasianidae (chicken, turkey, and Japanese quail)

**Supplementary Table S13.** Functional categories of positively selected genes ( $dN/dS > 1$ ) between peafowl and chicken

**Supplementary Table S14.** Functional categories of positively selected genes ( $dN/dS > 1$ ) between peafowl and turkey

1361 **Supplementary Table S15.** Functional enrichment of significantly expansive genes in  
1362 peafowl  
1363 **Supplementary Table S16.** Functional enrichment of significantly contractive genes  
1364 in peafowl  
1365 **Supplementary Table S17.** GO terms enrichment of positively selected genes in  
1366 peafowl under branch-site model  
1367 **Supplementary Table S18.** KEGG pathways of positively selected genes in peafowl  
1368 under branch-site model  
1369 **Supplementary Table S19.** Functional categories of positively selected genes in  
1370 peafowl under branch model  
1371 **Supplementary Table S20.** Primer sequences of *PMEL* for RT-qPCR  
1372

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS12TableS19.xlsx](#)
- [Supplementarymaterials.docx](#)