Cytogenetic and Genome Research

Fourth Report on Chicken Genes and Chromosomes 2022

Smith J, Alfieri JM, Anthony N, Arensburger P, Athrey G, Balacco J, Balic A, Bardou P, Barela P, Bigot Y, Blackmon H, Borodin PM, Carroll R, Casono MC, Charles M, Cheng H, Chiodi M, Cigan L, Coghill LM, Crooijmans R, Das N, Davey S, Davidian A, Degalez F, Dekkers JM, Derks M, Diack AB, Djikeng A, Drechsler Y, Dyomin A, Fedrigo O, Fiddaman SR, Formenti G, Frantz LAF, Fulton JE, Gaginskaya E, Galkina S, Gallardo RA, Geibel J, Gheyas A, Godinez CJP, Goodell A, Graves JAM, Griffin DK, Haase B, Han J-L, Hanotte O, Henderson LJ, Hou Z-C, Howe K, Huynh L, Ilatsia E, Jarvis E, Johnson SM, Kaufman J, Kelly T, Kemp S, Kern C, Keroack JH, Klopp C, Lagarrigue S, Lamont SJ, Lange M, Lanke A, Larkin DM, Larson G, Layos JKN, Lebrasseur O, Malinovskaya LP, Martin RJ, Martin Cerezo ML, Mason AS, McCarthy FM, McGrew MJ, Mountcastle J, Muhonja CK, Muir W, Muret K, Murphy T, Ng'ang'a I, Nishibori M, O'Connor RE, Ogugo M, Okimoto R, Ouko O, Patel HR, Perini F, Pigozzi MI, Potter KC, Price PD, Reimer C, Rice ES, Rocos N, Rogers TF, Saelao P, Schauer J, Schnabel R, Schneider V, Simianer H, Smith A, Stevens MP, Stiers K, Tiambo CK, Tixier-Boichard M, Torgasheva AA, Tracey A, Tregaskes CA, Vervelde L, Wang Y, Warren WC, Waters PD, Webb D, Weigend S, Wolc A, Wright AE, Wright D, Wu Z, Yamagata M, Yang C, Yin Z-T, Young M, Zhang G, Zhao B, Zhou H

Fourth Report on Chicken Genes and Chromosomes 2022

Prepared by Jacqueline Smith[1], James M. Alfieri[2,3,4], Nick Anthony[5], Peter Arensburger[6], Giridhar N. Athrey[2,4], Jennifer Balacco[7], Adam Balic[8], Philippe Bardou[9], Paul Barela[10], Yves Bigot[11], Heath Blackmon[2,3], Pavel M. Borodin[12], Rachel Carroll[13], Meya C. Casono[10], Mathieu Charles[14], Hans Cheng[15], Maddie Chiodi[10], Lacey Cigan[10], Lyndon M. Coghill[16], Richard Crooijmans[17], Neelabja Das[10], Sean Davey[10], Asya Davidian[18], Fabien Degalez[19], Jack M. Dekkers[20,21], Martijn Derks[17], Abigail B. Diack[8], Appolinaire Djikeng[22], Yvonne Drechsler[23], Alexander Dyomin[18], Olivier Fedrigo[7], Steven R. Fiddaman[24], Giulio Formenti[7], Laurent A.F. Frantz[25,26], Janet E. Fulton[27], Elena Gaginskaya[18], Svetlana Galkina[18], Rodrigo A. Gallardo[20,28], Johannes Geibel[29,30], Almas Gheyas[1], Cyrill John P. Godinez[31], Ashton Goodell[10], Jennifer A.M. Graves[32,33], Darren K. Griffin[34], Bettina Haase[7], Jian-Lin Han[35,36], Olivier Hanotte[36,37,38], Lindsay J. Henderson[8], Zhuo-Cheng Hou[39], Kerstin Howe[40], Lan Huynh[41], Evans Ilatsia[42], Erich Jarvis[7], Sarah M. Johnson[10], Jim Kaufman[41,43,44], Terra Kelly[20,28], Steve Kemp[45], Colin Kern[46], Jacob H. Keroack[10], Christophe Klopp[47], Sandrine Lagarrigue[19], Susan J. Lamont[20,21], Margaret Lange[48], Anika Lanke[49], Denis M. Larkin[50], Greger Larson[51], John King N. Layos[52], Ophélie Lebrasseur[53,54], Lyubov P. Malinovskaya[55], Rebecca J. Martin[44], Maria Luisa Martin Cerezo[56], Andrew S. Mason[57], Fiona M. McCarthy[10], Michael J. McGrew[8,22], Jacquelyn Mountcastle[7], Christine Kamidi Muhonja[42,45], William Muir[58], Kévin Muret[59], Terence Murphy[60], Ismael Ng'ang'a[25], Masahide Nishibori[61], Rebecca E. O'Connor[34], Moses Ogugo[45], Ron Okimoto[5], Ochieng Ouko[42], Hardip R. Patel[62], Francesco Perini[63], María Inés Pigozzi[64], Krista C. Potter[10], Peter D. Price[65], Christian Reimer[29], Edward S. Rice[66], Nicolas Rocos[41], Thea F. Rogers[67], Perot Saelao[20,46,68], Jens Schauer[29], Robert Schnabel[69], Valerie Schneider[60], Henner Simianer[30], Adrian Smith[24], Mark P. Stevens[8], Kyle Stiers[16], Christian Keambou Tiambo[45], Michele Tixier-Boichard[14], Anna A. Torgasheva[12], Alan Tracey[40], Clive A. Tregaskes[43,44], Lonneke Vervelde[8], Ying Wang[20,46], Wesley C. Warren[66,69], Paul D. Waters[32], David Webb[60], Steffen Weigend[29,30], Anna Wolc[27,21], Alison E. Wright[65], Dominic Wright[56], Zhou Wu[63], Masahito Yamagata[70], Chentao Yang[71], Zhong-Tao Yin[39], Michelle C. Young[10], Guojie Zhang[72], Bingru Zhao[73], Huaijun Zhou[20,46]

[1]Department of Genetics and Genomics, The Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, UK
[2]Interdisciplinary Program in Ecology and Evolutionary Biology, Texas A&M University, College Station, TX, USA
[3]Department of Biology, Texas A&M University, College Station, TX, USA
[4]Department of Poultry Science, Texas A&M University, College Station, TX, USA
[5]Cobb-Vantress, Siloam Springs, AR, USA
[6]Biological Sciences Department, California State Polytechnic University, Pomona, CA, USA
[7]The Rockefeller University, New York, NY, USA
[8]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Edinburgh, UK
[9]Université de Toulouse, INRAE, ENVT, GenPhySE, Sigenae, Castanet Tolosan, France
[10]University of Arizona, Tucson, AZ, USA
[11]PRC, UMR INRAE 0085, CNRS 7247, Centre INRAE Val de Loire, Nouzilly, France
[12]Department of Molecular Genetics, Cell Biology and Bioinformatics, Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
[13]Department of Animal Sciences, Data Science and Informatics Institute, University of Missouri, Columbia, MO, USA
[14]University Paris-Saclay, INRAE, AgroParisTech, GABI, Sigenae, Jouy-en-Josas, France
[15]USDA, ARS, USNPRC, Avian Disease and Oncology Laboratory, East Lansing, MI, USA
[16]Department of Veterinary Pathology, University of Missouri, Columbia, MO, USA
[17]Animal Breeding and Genomics, Wageningen University and Research, Wageningen, The Netherlands
[18]Saint Petersburg State University, Saint Petersburg, Russia
[19]INRAE, INSTITUT AGRO, PEGASE UMR 1348, Saint-Gilles, France
[20]Feed the Future Innovation Lab for Genomics to Improve Poultry, University of California, Davis, CA, USA
[21]Department of Animal Science, Iowa State University, Ames, IA, USA
[22]Centre for Tropical Livestock Genetics and Health (CTLGH) – The Roslin Institute, Edinburgh, UK
[23]College of Veterinary Medicine, Western University of Health Sciences, Pomona, CA, USA
[24]Department of Zoology, University of Oxford, Oxford, UK
[25]Queen Mary University of London, Bethnal Green, London, UK
[26]Palaeogenomics Group, Department of Veterinary Sciences, LMU Munich, Munich, Germany
[27]Hy-Line International, Research and Development, Dallas Center, IA, USA

[28]School of Veterinary Medicine, University of California, Davis, CA, USA

[29]Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut, Neustadt, Germany

[30]Center for Integrated Breeding Research, University of Göttingen, Göttingen, Germany

[31]Department of Animal Science, College of Agriculture and Food Science, Visayas State University, Visca, Baybay City, Philippines

[32]School of Biotechnology and Biomolecular Science, Faculty of Science, UNSW Sydney, Sydney, NSW, Australia

[33]Institute for Applied Ecology, University of Canberra, ACT, Australia

[34]School of Biosciences, University of Kent, Canterbury, UK

[35]CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China

[36]International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia

[37]Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, UK

[38]Centre for Tropical Livestock Genetics and Health, The Roslin Institute, Edinburgh, UK

[39]National Engineering Laboratory for Animal Breeding and Key Laboratory of Animal Genetics, Breeding and Reproduction, MARA; College of Animal Science and Technology, China Agricultural University, Beijing, China

[40]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

[41]Institute for Immunology and Infection Research, University of Edinburgh, Edinburgh, UK

[42]Dairy Research Institute, Kenya Agricultural and Livestock Organization, Naivasha, Kenya

[43]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

[44]Department of Pathology, University of Cambridge, Cambridge, UK

[45]Centre for Tropical Livestock Genetics and Health (CTLGH) – ILRI, Nairobi, Kenya

[46]Department of Animal Science, University of California, Davis, CA, USA

[47]INRAE, MIAT UR875, Sigenae, Castanet Tolosan, France

[48]Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, MO, USA

[49]BASIS Chandler High School, Chandler, AZ, USA

[50]Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London, UK

[51]The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, The University of Oxford, Oxford, UK

[52]College of Agriculture and Forestry, Capiz State University, Burias, Mambusao, Capiz 5807, Philippines

[53]Centre d'Anthropobiologie et de Génomique de Toulouse (CAGT), CNRS UMR 5288, Université Toulouse III Paul Sabatier, Toulouse, France

[54]Instituto Nacional de Antropología y Pensamiento Latinoamericano, Ciudad Autónoma de Buenos Aires, Argentina

[55]Department of Cytology and Genetics, Novosibirsk State University, Novosibirsk, Russia

[56]AVIAN Behavioural Genomics and Physiology, IFM Biology, Linköping University, Linköping, Sweden

[57]Department of Biology, The University of York, York, UK

[58]Department of Animal Sciences, Purdue University, West Lafayette, IN, USA

[59]Université Paris-Saclay, Commissariat à l'Energie Atomique et aux Energies Alternatives, Centre National de Recherche en Génomique Humaine, Evry, France

[60]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

[61]Laboratory of Animal Genetics, Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Japan

[62]The John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia

[63]Department of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Easter Bush Campus, Edinburgh, UK

[64]INBIOMED (CONICET-UBA), Facultad de Medicina, Universidad de Buenos Aires, Buenos Aires, Argentina

[65]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK

[66]Department of Animal Sciences, Bond Life Sciences Center, University of Missouri, Columbia, MO, USA

[67]Department of Molecular Evolution and Development, University of Vienna, Vienna, Austria

[68]Veterinary Pest Genetics Research Unit, USDA, Kerrville, TX, USA

[69]Department of Animal Sciences, University of Missouri, Columbia, MO, USA

[70]Center for Brain Science, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA

[71]BGI-Shenzhen, Shenzhen, China
[72]Center for Evolutionary and Organismal Biology, Zhejiang University School of Medicine, Hangzhou, China
[73]College of Animal Science and Technology, Nanjing Agricultural University, Nanjing, China

Corresponding author:
Jacqueline Smith,
Department of Genetics and Genomics,
The Roslin Institute and R(D)SVS,
University of Edinburgh,
Edinburgh EH25 9RG, UK
jacqueline.smith@roslin.ed.ac.uk

Introduction
(Prepared by J. Smith)
The chicken continues to hold its position as a leading model organism within many areas of research, as well as a being major source of protein for human consumption. The First Report on Chicken Genes and Chromosomes [Schmid et al., 2000], which was published in 2000, was the brainchild of the late, and sadly missed, Prof Michael Schmid of the University of Würzburg. It was a publication bringing together updates on the latest research and resources in chicken genomics and cytogenetics. The success of this First report led to the subsequent publication of the Second [Schmid et al., 2005] and Third [Schmid et al., 2015] reports – each also proving popular references for the research community. It is now our pleasure to be able to introduce publication of the Fourth report. Being seven years since the last report, this publication captures the many advances that have taken place during that time. This includes presentation of the detailed genomic resources that are now available, largely due to increasing capabilities of sequencing technologies and which herald the pangenomic age, allowing for a much richer and more complete knowledge of the avian genome. Ongoing cytogenetic work also allows for examination of chromosomes, specific elements within chromosomes and the evolutionary history and comparison of karyotypes. We also examine chicken research efforts with a much more 'global' outlook with a greater impact on food security and the impact of climate change, and highlight the efforts of international consortia, such as the Chicken Diversity Consortium. We dedicate this Report to Michael.

Multiple Chicken (*Gallus gallus*) Genome References to Advance Genetic Variation Studies
(Prepared by W.C. Warren, O. Fedrigo, A. Tracey, A.S. Mason, G. Formenti, F. Perini, Z. Wu, T. Murphy, V. Schneider, K. Stiers, E.S. Rice, L.M. Coghill, N. Anthony, R. Okimoto, R. Carroll, J. Mountcastle, J. Balacco, B. Haase, C. Yang, G. Zhang, J. Smith, Y. Drechsler, H. Cheng, K. Howe, and E. Jarvis)
We present two phased chromosome-scale assemblies of chicken, a layer (GRCg7w) and broiler (GRCg7b), that better meet research demands to characterize segregating variation important for traits of interest. Annotation with existing long- and short-read RNAseq data improved contiguity, accuracy, and protein-coding and non-coding gene counts, when compared to the existing Red Jungle Fowl reference, GRCg6a. Most striking were the improvements in placed telomeres, corrections for erroneous microchromosome fusions, and gap reduction in these phased assemblies. We add six putative microchromosomes that were previously missing in GRCg6a. Using a pairwise genome comparison of the parental genomes, and two independent cohorts of sequenced chickens, we show small discernable differences in mapping rates of whole genome sequence (WGS) and RNAseq data, gene annotation, and called single nucleotide variants (SNVs) or indels. Structurally, some regional differences suggest future assembly curation will further improve variant ascertainment. These *Gallus* references also enabled a new genome-wide review of endogenous Avian Leukosis Virus (ALVE) integrations, exemplifying the improved representation of chicken genomic diversity by these phased genomes. Our genome references will collectively improve our computational outcomes when testing multiple variant hypotheses that are at the core of our understanding of avian biology.
Today, the poultry industry faces many challenges, perhaps none more than the genetics underlying bird health. A constant balance must be maintained to select on several traits of immense economic impact, such as fast growth in broilers and reproductive success in layers, while not diminishing disease resistance. Genetic studies offer promising avenues to selectively maintain this trait balance with a new accounting of the most important contributing factors: genes and environment (Wolc et al. 2018). More complete and accurate genomic resources

to support their continued discovery of these factors is paramount to generating the robust chicken germlines that can meet a growing demand for this protein food source.

The chicken also supports a vast model organism community that uses a collateral source of scientific data to comparatively inform our understanding of predominantly developmental biology (see review (Cheng and Burt 2018)). The chicken genome is one of the most frequently used resources for comparative genomic studies among vertebrate taxa (Zhang et al. 2014). As the principal avian reference genome, it was used to transfer gene annotation evidence to over 50 bird genomes, which were examined for clade- and species-specific signals of genome evolution (Zhang et al. 2014). Recent research attempts to determine the effect of structural variation (SV) on chicken phenotypic differences, although resolution beyond short-read mapping or hybridization methods must be considered (Rao et al. 2016). The site-directed gene knockouts of chicken *C2EIP* (Zuo et al. 2016) and *PAX* (Gandhi et al. 2017) genes are two functional instances of gained insight into embryonic germ and satellite muscle cell differentiation, respectively.

Since its first iteration in 2004, we have worked to refine the assembly of the chicken genome as technology has enabled over the past two decades (Consortium 2004). We recently summarized these advances (Rhie et al. 2021) like the single haploid phasing of a diploid genome, i.e., trio binning, which sorts and independently assembles divergent parental haplotypes from F1 hybrids (inter- and intraspecies crosses) as a highly efficient method for untangling complex sequence assembly graphs. This phasing strategy exploits the higher heterozygosity observed in some F1 hybrids in order to resolve diploid genomes more precisely and with fewer gaps. Several successful haplotype-resolved de novo assemblies for cats, cattle, zebra finches, and others have been created using this method (Bredemeyer et al. 2021) (Rice et al. 2020) (Korlach et al. 2017), highlighting the astounding improvements in contiguity, with some sequences spanning from telomere to telomere.

Recent characterization of different chicken genomes has demonstrated the necessity for pangenome resources in order to comprehend the comparative evolution of the *Gallus* genus (Li et al. 2022). To date, all chicken genetic studies have relied on the Red Jungle Fowl (RJF) genome, which portrays this diploid genome as a collapsed haploid genome containing a mixture of sequences from the two haplotypes. For the further investigation of trait selection indices, the adoption of additional high-quality chicken references, particularly those that better resemble commercial birds, is highly endorsed by the avian community and has broad applicability. In addition, these resources enable pangenome techniques that will provide higher resolution for discovering SVs that are exclusive to decades of artificial selection. Here, we present two novel haploid *de novo* assemblies for chicken lines with extremely diverse genetic histories: one bred for muscle growth (broiler) and the other for egg production (layer). Large structural adjustments among microchromosomes, overall gap reduction, extension of the W chromosome by adding the pseudo-autosomal region, better chromosome 16 (MHC region) representation, and enhanced telomere sequence placements are notable. Then, we demonstrate these new assemblies' application in determining the extent of alignment, SNV identification, ALVE integration, and structural expansions and contractions in a small sample of chickens.

Sequencing and Assembly

A parent-offspring trio composed of a paternal layer, a maternal broiler, and their female F1 offspring was sequenced to create these assemblies. Briefly, the parents were sequenced with low-coverage Illumina reads (150bp) and the F1 was sequenced with 80x PacBio reads (12kb on average), and all reads were used as input to TrioCanu (see review of methods: (Rhie et al. 2021)). Similar to cross-species trio assembly of cattle and yak, the amount of haplotyped long reads phased from each parental breed source was extremely similar (49.5 and 50.2%) with a low number of unknowns (0.16 percent) (Rice et al. 2020). Broiler (n=676) and layer (n=688) birds had half as many constructed contigs as RJF (n=1,403) birds, indicating that >53 percent of prior gaps have been bridged (currently 878 in RJF). Depending on the descriptive context, we use the assembled GenBank versions (GRCg6a, GRCg7b, and GRCg7w) and their common names (RJF, broiler, and layer) interchangeably throughout the remainder of this report. While contig N50 length was comparable across all assemblies, phasing and revised mapping data led to a 4.5-fold increase in N50 scaffold length and a 2-fold decrease in the number of unplaced sequences in broiler and layer assemblies (Table 1). The paternal layer contributes Z to the ZW sex chromosomes, while the maternal broiler was particularly chosen for her haplotype A mitochondrial genome and W. The female RJF reference is unique with a mitochondrial genome of haplogroup E. The layer Z chromosome is somewhat larger and contains more protein-coding genes than the BAC-curated GRCg6a version of Z (85.2 vs 80Mb; 1,492 vs 1,345 genes) (Bellott et al. 2010). Furthermore, the broiler W chromosome is more complete than the GRCg6a chromosome, which is 7.2Mb in size, due in part to the insertion of the pseudoautosomal region (PAR) that

boosts its comparative utility (online suppl. Material 1, Fig. 1). The initial about 500kb of the W chromosome show diploid coverage. We chose not to join this sequence to the beginning of Z since we lack precise coordinates, and this portion of Z is partially collapsed. In each phased reference, for the sake of completeness, Z and W were incorporated notwithstanding their parental origins. By searching the NCBI assembly archive for 'Gallus gallus', you can find all fully annotated assemblies (see data availability).

Assembly Accuracy Benchmarking

There are inherent assembly artifacts present in all reference genomes, including the human genome. With this knowledge, we wanted to estimate the detected errors in GRCg6a, given its extensive use in chicken genetic studies, and repair them our new phased assemblies using a previously established iterative procedure (Howe et al. 2021). Adding orthogonal evidence including chromatin proximity (HiC) and Bionano optical maps further delineated the error locations for subsequent correction, e.g. 260 and 63 missed joins in GRCg7b and GRCg7w (online suppl. Material 2, Table 1) resulting in an increase in scaffold N50 of 23 and 6%, respectively, when compared to GRCg6a. A greater number of GRCg7b and GRCg7w chromosomes exhibited telomere ends (24 and 13, respectively), than GRCg6a (just 3), demonstrating the much-improved completeness of the new assemblies. Among microchromosomes, we discovered many instances in which GRCg6a chromosomes were wrongly fused into a single chromosome instead of two distinct ones (online suppl. Material 2, Table 1). The first 2 Mb of GRCg6a chr27 is not associated with chr27, but rather a variety of alignments to other chromosomes, including W and chr2, which, upon curation, accurately sizes this chromosome; 8 Mb as opposed to 5.2 Mb in GRCg7b (Fig. 1). Other errors include the fusion of chr31 and chr29 in GRCg6a, which is likely due to repeat sequences identified on the HiC heat map (online suppl. Material 1, Fig. 2).

Avian microchromosomes show more frequent recombination, and thus the positive correlation between recombination and interspecies divergence observed in mammals is not seen in birds, at least at the resolution of whole chromosomes (Consortium 2004). If the origin of the microchromosomes was initiated by a number of random fission events that were channeled towards the present day macro/microchromosome arrangement it was clear from earlier evaluations that there were not sufficiently long compositionally uniform regions in any of the sequenced avian genomes, that would satisfy some classifications, e.g., the classical isochore definition within microchromosomes (Waters et al. 2021). We now have corrected several assembly errors, mostly among the microchromosomes, to test more accurately these and other hypotheses regarding their evolution.

The chicken karyotype has a diploid number of 78 chromosomes, classified as a haploid autosome count of 10 macrochromosomes and 28 microchromosomes (Burt et al. 1999). In earlier chicken assemblies, microchromosomes 29 and 34-38 were absent, primarily due to the absence of linkage groups or physical maps that might assign missing scaffolds to any of these smaller chromosomes (Groenen et al. 2000), as well as difficulty in sequencing through high-GC rich microchromsomes. In both GRCg7b and GRCg7w, using chromatin proximity evidence and long reads that get through GC-rich regions, we identify these missing microchromosomes (online suppl. Material 1, Fig. 3) and an additional microchromosome to a final total of 39 autosomes. Future cytogenetic evaluations or new combinatorial approaches that can yield telomere-to-telomere stepwise assembly of more complete chromosomes (Logsdon et al. 2021) will be necessary to rule out the possibility these nominated microchromosomes are not affiliated with other macro- or microchromosomes. Moreover, the availability of almost complete genome copies of these uniquely selected lines and others will drive reevaluations of all types of segregating variation in a pangenome-dependent manner (Siren et al. 2021).

Structural Differences

To estimate the major structural differences among these phased references, we employed two methods: high resolution alignments to reveal major synteny differences using SyRi (Goel et al. 2019) and the predicted contractions and expansions of deletions, insertions, and repeat elements with different size distributions using Assemblytics (Nattestad and Schatz 2016). Across the chicken genome, differences in local chromosomal synteny were predominately one-to-one; however, when we discover discrepancies, they frequently occur towards chromosome ends, highlighting the difficult nature of placing sequences in these repetitive telomeric regions (Fig. 2). Regardless of their length distribution Assemblytics alignment results show comparable total base size differences (online suppl. Material 2, Table 2), (Fig. 2; online suppl. Material 1, Fig. 4). However, these differences vary by type, such as deletion versus insertion, which may be the result of numerous factors, including genetic diversity and assembly completeness and accuracy of each reference. When employing a phased assembly for pairwise broiler versus layer alignments, the total number and base sizes of discovered deletions and insertions

drop relative to RJF, suggesting the more diversified origins of RJF and its mixed haplotype assembly architecture are the cause (online suppl. Material 2, Table 2). Our initial genome-wide perspective of SVs in RJF, layer, and broiler genomes, including overall alterations in repeat content, will require additional research to validate their patterns of segregation in larger populations of chickens and their accuracy of ascertainment. Overall, we observe structural differences, despite the fact that the percent masked sequence, a measure of all repeat types, is comparable between these references (20.5, 20.2, and 20.3) using default WindowMasker output (Morgulis et al. 2006) but again the biological context is yet unknown.

Gene Annotation
First, protein-coding gene representation was evaluated with BUSCO, which demonstrated an average 54% reduction in the number of missing universal single-copy orthologs in both GRCg7 versions compared to GRCg6a (online suppl. Material 2, Table 3). Automated gene annotation of GRCg7b and GRCg7w using the NCBI workflow (Sayers et al. 2021) reveals an increase in the overall number of protein-coding and non-coding genes (online suppl. Material 2, Table 4). Recent gene annotation of multiple chicken genomes revealed 1,335 more protein-coding genes relative to GRCg6a (Li et al. 2022). The addition of at most 546 genes to the GRCg7 phased assemblies is a modest increase, but not unexpected given the NCBI annotation methods are likely more conservative and do not rely on many varied genome annotations as in Li et al. (Li et al. 2022). We also analyzed the differences in gene set ontology between GRCg7b and GRCg7w given their distinct selection histories. As determined by enrichment analysis, we find 82.8% overlap between GRCg7b and GRCg7w where uniqueness is most often large gene families, e.g., immune genes (online suppl. Material 1, Fig. 5, online suppl. Material 2, Table 5). In the broiler annotated set, there are 154 genes not found in the layer set, but there were only 44 unique to the layer set (online suppl. Material 2, Table 5). This disparity may reflect the slightly higher contiguity of the broiler reference (Table 1). Future study will be required to provide a precise accounting of the genes that are unique to breeds, commercial or research lines, and wild strains of *Gallus gallus*. In addition, solutions to the question of whether avian genes were genuinely lost during the ancient divergence of avian and mammalian lineages will begin with the availability of a full genome, as recently highlighted in humans (Nurk et al. 2022).

WGS Mapping and SNV analysis
It is probable that the choice of broiler, layer or RJF as a reference for alignment of various resequenced chicken populations could contribute to SNV ascertainment bias as has been shown in human (Schneider et al. 2017). We examined the mapping rates of WGS data of six genetically diverse chicken samples (online suppl. Material 2, Table 6): a male and female for layer and broiler chicken, as well as an Ethiopian indigenous chicken breed. For various mapping metrics, regardless of the reference, we find no large differences, indicating that for measures of genetic diversity, all three references have comparable initial abilities to call SNVs or indels (online suppl. Material 2, Table 7).
Despite our conclusion that WGS mapping rates were very similar across references, the optimal SNVs set for the experimental purpose intended is not certain. Next, we mapped WGS data from a separate cohort of solely broilers (n=10) to all three genome assemblies and called SNVs using GATK version 4.2.0. Although we found differences in total SNVs, these were not large, suggesting that SNV detection will be comparable when beginning with any of the three assemblies (online suppl. Material 2, Table 8). However, regional variations may be encountered and must be addressed if certain loci are of great experimental relevance (Fig. 3).

RNAseq mapping
The mapping of RNAseq data to estimate transcriptome changes between samples for biological interpretation is a crucial reference usage. To address this application, we first analyzed a large number of diverse tissues where total percent mapping is available in the NCBI gene annotation report (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Gallus_gallus/106/) and found very little difference between our phased assemblies (GRCg7b and GRCg7w). However, since RNAseq alignment in this application is optimized for verifying gene model predictions, we also tested the STAR aligner, which is typically considered best practice for bulk RNAseq studies (Dobin et al. 2013). Using a small number (n=8) of RNAseq samples from diverse tissue origins, including ileum, bone derived macrophages, and uterus, we observe a small average range of 0.7 to 1.7% differences among six samples in the total percentage of reads uniquely mapping to each of the three references (online suppl. Material 2, Table 9). However, it is unclear why GRCg6a has a somewhat greater percentage of uniquely mapped reads across all samples (online suppl. Material 2, Table 9). We also highlight the

two individual female and male layer muscle RNAseq samples with vastly different average rates of unique mapping between assemblies, 70.6 and 68.5% to each GRCg7 version, respectively, compared to 88% in GRCg6a (online suppl. Material 2, Table 9). In the female layer sample, after analyzing all secondary alignment counts (not read counts), the olfactory receptor 14C36-like gene (>10M) is most abundant in GRCg6a, whereas for GRCg7 versions it is ribosomal RNAs (Fig. 4). Although the number of individual rRNAs in each reference is comparable (~90), the total base size is notably different. The total assembled lengths of rRNA in each genome are GRCg6a (20,568), GRCg7b (113,817), and GRCg7w (55,916) (Fig. 4), suggesting that when evaluating unique versus multi-mapping events for any of these references the type of library sequenced, i.e., ribosome depleted or polyA selected should be considered (Zhao et al. 2018). Overall, for the majority of RNAseq samples studied, we observe minimal differences in unique mapping rates across all references; nonetheless, prior to conducting RNAseq mapping experiments, the reference choice should be considered.

Contrasting ALVE diversity in varied genome assemblies
To show an additional advantage of these phased references, we revisit the question of the RJF reference being unrepresentative of ancestral *Gallus gallus* ALVE diversity. ALVEs are species-specific retroviral integrations which retain the potential for retrotransposition and retroviral expression (Fig. 5A). The previous RJF reference assembly contained two Avian Leukosis Virus subgroup E (ALVE) integrations: ALVE6 (ALVE-JFevA), a truncated ALVE widespread across many breeds; and ALVE-JFevB, an intact integration found in no other chicken to date (Mason, Fulton, and Smith 2020). This new, phased assembly contains a total of eleven ALVEs: five from the maternal broiler and six from the paternal layer (summarized in Fig. 5B; detailed locations in online suppl. Material 2, Table 10). Leghorn layers typically have fewer than six ALVEs (Mason, Lund, et al. 2020) (Mason, Miedzinska, et al. 2020), but the identified ALVE1, ALVE3, ALVE15, ALVE_ros034 and the slow feathering-associated ALVE21 are representative of this Leghorn layer breed. ALVE_ros005, however, has previously only been identified in brown-egg layers and indigenous birds (Mason, Lund, et al. 2020). The ALVEs of the broiler haplotype are widely found across brown-egg commercial layers and broiler lines, representing their recent shared ancestry (Muir et al. 2008), and the presence of ALVE-TYR supports the observed recessive white phenotype (Fox and Smyth 1985) (Chang et al. 2006).
Seven of the ALVEs are full-length (Fig. 5B; online suppl. Material 2, Table 10), and five have completely intact retroviral ORFs, accounting for the -1 ribosomal frameshift between gag and pol (Nikolic et al. 2012). Despite this, ALVE transmission between cells is unlikely, as both parental haplotypes exhibit ALVE-resistance at the TVB receptor (*TNFRSF10B*): maternal Q58* (TVB^R; rs736008824) and paternal P61L (rs318006572). This perhaps represents the effects of selection against P27 expression in commercial birds. Additionally, the ALVE_ros034 gag ORF truncates within P27, and similar mutations have been observed in ALVEs in other commercial backgrounds (Fig. 5B; (Mason, Lund, et al. 2020). Additional high quality chicken genome references of diverse genetic backgrounds interpreted in pangenome visualization modes will continue to resolve the evolution of ALVEs and their role in trait presentation.

Future chicken references
Pangenomic starting points as opposed to single linear representations have been proposed in humans (Siren et al. 2021) to overcome reference bias in genotyping. In Siren et al., the utility of a human pangenome reference in variant ascertainment demonstrates that this is the optimal course of action for future chicken genetic studies, particularly structural analyses (Siren et al. 2021). As a result, we are generating the requisite read-types to follow this same *de novo* assembly process in building multiple telomere-to-telomere single haplotype reference sources. Using these individual linear genome graphs to construct pangenome references will ensure the availability of the next generation of computational resources for optimally estimating segregating variation for significant genotype to phenotype connections in poultry production. The phased assemblies of the broiler and layer genomes as well as the RJF reference provide new insights into their general structure. In addition, we believe a new era in the use of avian genome references has already begun due to the rapid development of methods to build full genome copies.

Bird husbandry
The parent-offspring trio is composed of a male White Leghorn and female broiler, the parents, each raised at the University of Arkansas avian housing facilities. A female F1 offspring was chosen from this cross for sequencing.

DNA for each parent and the F1 was extracted from white blood cells using standard practices for each intended use.

Sequencing and primary genome assembly
We followed the workflow established by the Vertebrate Genomes Project (VGP) to create the haplotype-phased chicken assembly (Rhie et al. 2021). Libraries were sequenced on the PacBio Sequel II instrument with the sequencing kit 2.1 (#101-310-500) and 10 hours movie time to a total of ~98GB. Because sequence coverage is lowered when phasing a diploid genome, we targeted a high read coverage of ~80x, to attempt the accurate assembly of repetitive microchromosomal regions and ZW sex chromosomes. TrioCanu (v1.8+287) was used to bin Consensus Long Reads (PacBio) of the F1 female into maternal and paternal haplotypes using haplotype-specific 21-mer markers derived from the Illumina short reads of the mother and father. Following binning, TrioCanu independently generated contigs for each haplotype (haplotigs). From this point, the maternal and paternal haplotigs independently underwent the same steps. Separately, we assembled the mitochondrial (MT) genome with the mitoVGP pipeline (v2.2) (Formenti 2020) and added it to the haplotigs to keep any raw MT reads from being mapped to nuclear sequences resulting in conversion of possible mitochondrial nuclear integrations into MT sequence during the polishing steps. We used Arrow from smrtlink (v6.0.0.47841) to improve base calling accuracy and purge_dups (v1.0.0) (Guan et al. 2020) in an adapted trio mode to remove erroneous duplications. The median insert sizes of WGS libraries were approximately 400 bp and individual libraries were tagged with unique dual index DNA barcodes to allow pooling and minimize the impact of barcode hopping. Libraries were pooled for sequencing on the NovaSeq 6000 (Illumina) to obtain at least 750 million 151-base pair reads per individual.

Assembly scaffolding and curation
As detailed by Rhie et al., various maps were constructed to facilitate scaffolding of the phased contigs (Rhie et al. 2021). Briefly, long linked read libraries were generated from unfragmented high molecular weight DNA on the 10X Genomics Chromium instrument (Genome Library Kit & Gel Bead Kit v2 PN-120258, Genome Chip Kit v2 PN-120257, i7 Multiplex Kit PN-120262). We sequenced this 10X library on an Illumina HiSeq X instrument with 150bp read length to ~60X coverage. For optical mapping, the extracted DNA (~750ug) was labeled with a direct labeling enzyme (DLE-1) following the BioNano Prep Direct Label and Stain (DLS) Protocol (Document Number 30206). Labelled samples were imaged on the Bionano Saphyr instrument. Finally, HiC map generation was performed by Arima Genomics (https://arimagenomics.com/) using the Arima-HiC kit (P/N: A510008). From size selected fragments, Illumina-compatible libraries were generated using the KAPA Hyper Prep kit (P/N: KK8504). The resulting libraries were sequenced on an Illumina HiSeq X instrument to ~70x coverage.
With 10X long-linked reads, BioNano, and Hi-C maps in hand, the earlier polished and purged haplotigs were scaffolded in three stages according to Rhie et al. (Rhie et al. 2021): first, we used the 10x linked-reads in two rounds of scaff10x (v4.1.0) (https://github.com/wtsi-hpag/Scaff10X) to generate the primary scaffolds. Second, we generated BioNano cmaps and used BioNano Solve (v3.2.1_04122018) (Lam et al. 2012) for hybrid scaffolding and to break mis-assemblies. Third, we used Salsa2 (v2.2) (Ghurye et al. 2019) to generate chromosomal-level scaffolds using the molecular contact information from Hi-C linked reads. Finally, we performed a second round of Arrow polishing on the maternal and paternal scaffolds with the binned long reads. During this round of polishing, gaps between contigs were closed by the gap-filling function of Arrow. The two haplotypes were then combined in a single assembly and underwent two rounds of short read polishing using longranger (v2.2.2) (Bishara et al. 2015) and freebayes (v1.3.3) (Garrison 2017). After separating the scaffolds back into their respective haplotypes and removing the MT genome from each assembly, the two phased assemblies underwent manual curation using gEVAL as described previously (Chow et al. 2016) (Howe et al. 2021), particularly to correct structural assembly errors.

Assembly statistics and evaluation
Following each stage of the assembly, we calculated various metrics of assembly quality, for example, N50 contig length, number of contigs, and quality value (QV) scores for each base call to assess progress. However, primarily we used Merqury (v1.0) for overall assembly evaluations (including k-mer completeness and spectra copy number analysis) as well as phasing assessment with hap-mers. We first generated 21-mer databases (dbs) from the raw F1 10x data and the parental Illumina data using meryl. We then built inherited hap-mer dbs by taking the difference between the maternal and paternal k-mer dbs, filtering according to the filter level used by TrioCanu

for binning, intersecting both with the F1 dbs, and filtering again, as below (steps 1-4). For evaluation of genome completeness and protein-coding gene representation, we ran BUSCO v4.0.2 (Manni et al. 2021) on our phased assemblies to determine the representation of near-universal single-copy orthologs in the vertebrate avian lineage (n=8,338); aves_odb10 (online suppl. Material 2, Table 3).

Genome synteny and structural variation
To estimate sequence structural changes between assemblies for synteny, structural variation, and repeat expansion and contractions we used SyRi (Goel et al. 2019) with default parameters or Assemblytics v1.2.1 (Nattestad and Schatz 2016) with a unique sequence length requirement of 10,000 on nucmer alignments between GRCg6a, GRCg7b, and GRCg7w assemblies.

Gene Annotation
Both assemblies, GRCg7b, and GRCg7w were gene annotated using the standard NCBI pipeline (Pruitt et al. 2014), including masking of repeats prior to ab initio gene predictions, for evidence-supported gene-model building. All annotation processes used publicly available RNA-seq and Iso-Seq data from diverse tissue sources. We relied on the NCBI gene annotation report release 106 to compare the outcomes for each assembly. GRCg6a gene annotation data were reported earlier in NCBI release 104 using the same process as above.

Interspersed repeat estimation
Two independent assessments were made to estimate the percentage of repeats to confirm their similarities between assemblies. RepeatMasker v4.0.9 (Smit A 2013) with *-excln* and *-species chicken* was used to identify and annotate repetitive regions of each genome while ignoring gap sequence then WindowMasker analysis was carried out using default parameters (Morgulis et al. 2006).

WGS and RNAseq mapping
WGS short-read data of six chicken samples; a male and female layer, broiler and Ethiopian indigenous chicken breed were used to compare the mapping rate across the three genome assemblies (online suppl. Material 2, Table 6). WGS were first checked for quality using Fastqc (Andrews 2010). Trimmomatic (Bolger, Lohse, and Usadel 2014) was used to remove the remaining Illumina adapter sequences and low-quality bases with default parameters. Clean reads were mapped to the three reference assemblies (GRCg6a, GRCg7b, and GRCg7w) using bwa-mem with default parameters (Li and Durbin 2009). Picard (http://broadinstitute.github.io/picard/) was used to sort the mapped files and merge files from multiple sequencing runs and to mark duplicate reads. Finally, SAMtools (Li et al. 2009) was used to assess mapping quality.
RNAseq data from eight chicken samples were used to compare the mapping rate across the three genome assemblies. We retrieved all sequence data from the NCBI Sequence Read Archive that included the diverse tissue sources of ileum, bone-derived macrophages, uterus and muscle from a male and female layer (online suppl. Material 2, Table 9). Sequencing quality was checked by FastQC software (v 0.11.7), qualifying reads were mapped using STAR software (v 2.5.3a) with default parameters to all assemblies (GRCg6a, GRCg7b, and GRCg7w), and the percentage of uniquely mapped reads, multiple mapped reads, and reads mapped to too many loci were taken for the comparison. Moreover, the resulting bam files were used for assessing the mapping rate for each sample with the Samtools (v 1.9) 'flagstat' command (Li et al. 2009). Percentages of correctly paired reads were used for comparison.

SNV analysis
To estimate SNV differences in the starting reference alignments we used short-read sequences from a small cohort of broilers (n=10) representing commercial birds generated by Cobb-Vantress (available upon request). All samples attained genome coverage depth greater than 20x and individual reads were aligned to each reference with the Nvidia Clara Parabricks (version 3.6) implementation of the BWA algorithm. Variants were called in GVCF mode with Nvidia Parabricks HaplotypeCaller and GVCF files were loaded into GenomicsDB using GATK 4.2.0 (Poplin 2017), GenomicsDBImport, and joint-genotyped with GATK's GenotypeGVCFs. Hard-filtering was performed on the resulting raw VCF using GATK's current best-practices for filtering.
BCFTools 1.12 was used to extract statistics on SNVs and insertions and deletions (Indels) per chromosome. Variants that did not pass the filtering criteria were removed and mapping data for chromosomes were compared between all assemblies using command-line tools and then plotted in R. Ideograms were generated using

karyoploteR (v1.16.0) in R. Colored regions of the chromosome denote annotated feature regions for that chromosome. Rainfall plots of variants depict where variants were found in the analysis along each chromosome. Each unique color indicates a different type of substitution. We only include variants that passed all filters and were heterozygous in the reference source.

ALVE annotation
Assembled Avian Leukosis Virus subgroup E (ALVE) integrations were identified by BLAST v2.10.0 (Altschul et al. 1990) using the ALVE1 reference sequence (GenBank: AY013303.1) and annotated for ORFs and miR-155 recognition sites (Hu et al. 2016). Analogous GRCg6a locations were identified using flanking sequences, then compared with known ALVE integration sites and target site duplications (TSD) (Mason, Lund, et al. 2020) (Mason, Miedzinska, et al. 2020). ALVE susceptibility was assessed by identifying the TVB receptor (*TNFRSF10B*) genotype.

Karyotypes and Chromosome Level Assemblies in Chickens and Other Dinosaurs
(Prepared by D.K. Griffin, D.M. Larkin, and R.E. O'Connor)

In any article about chicken, or avian generally, genomics is inherently about dinosaur genomics. In the light of recent paleontological evidence, statements such as "birds *evolved from* dinosaurs" and "birds are *related to* dinosaurs" require substantial revision. Rather, it is very clear that birds, including our humble chicken, are in fact extant dinosaurs. Omnipresent in literature, film, television, popular culture and the media since the original fossil discoveries, notions that dinosaurs were obliterated entirely by the latest mass extinction event have undergone radical revision. On the contrary, dinosaurs are the great survivors of mass extinction events and we suggest, this may be due, at least in part, to their unique genome organisation (in other words, their karyotype). Studies of chromosome-level genome assemblies guided us to this conclusion.
Chromosome-level assemblies (CLAs)
Genomic approaches such as array CGH (comparative genomic hybridization) and NGS (next generation sequencing) for cytogenetic application would not be possible for many clinical and veterinary uses were it not for 'chromosome-level' assemblies (CLAs). In other words, one ultimate objective of any genome assembly is that all the sequences are correctly aligned and assigned to their place on the appropriate chromosome, chromosome arm and chromosome band.
Similarly, genome assembly needs cytogenetic analysis. As Lewin et al (2010) put it, "every genome needs a good map." Thankfully, the outline of that map for any given species is provided in the shape of a karyotype. While karyotypes typically are used to detect chromosomal disease in humans, the karyotype can be considered the most basic low-resolution genomic map of an organism. When whole genome assemblies are yet to reach the heights of a CLA (probably the case for most sequenced genomes), their applicability for evolutionary genomics is impeded. To give one of many examples, CLAs are crucial for furthering applied agricultural research in that an established order of sequences is an essential pre-requisite to establish genotype-phenotype associations e.g. in genome wide association studies (GWAS). CLAs have facilitated these in chicken, turkey and duck as well as in

mammals such as pig, cattle and sheep. Other advantages of CLAs include discovery of Mendelian traits, spotting cryptic chromosome translocations, as well as isolating quantitative trait nucleotides (QTNs), expression quantitative trait loci (eQTLs) and long-range regulatory interactions. The ultimate goal of these studies is increased food production efficiency as well as global food security. With the knowledge at our fingertips for a number of species' CLAs, comparative genomics is much more practicable *in silico* and intractable karyotyping of chromosome rearrangements can be detected by fluorescence in situ hybridisation (FISH). Comparative genomics, thereafter, allows us to describe the genome structure of more obscure species (by comparison to a standard, such as chicken) and the identification of chromosome rearrangements that led to each species' unique karyotype. CLAs also facilitate addressing fundamental biological hypotheses pertaining to genome evolution e.g. the mechanisms of chromosome breakage and fusion, as well as the significance and genomic correlates of evolutionary breakpoint regions (EBRs) and homologous synteny blocks (HSBs). In the era of genomics, cytogenetics (or, more specifically, cytogenomics) is not just a descriptive science but it provides a framework for the conceptualization of the structure of any genome. It gives us a starting point through which we can understand genome-phenome correlations.

The particular challenge with establishing genome structure in birds is that while, in most species, correlating genome assembly with karyotype is a bit like identifying cities and towns in the major landmasses of Europe (some big, some small). Doing it in birds is more like doing it for Polynesia where there are a lot of tiny islands, i.e. the microchromosomes. In our research, we have looked at CLAs in a multitude of bird species, providing novel insight into the genome organization of the avian forebears – the extinct dinosaurs.

What is a dinosaur?

The official definition of a dinosaur is "Triceratops, modern birds, their most recent common ancestor and all their descendants." As biologists however, perhaps an easier way to visualise them is as reptiles (in this definition we class birds as reptiles) with hind limbs held erect beneath the body (like mammals). This distinguishes them from most other reptiles such as lizards and crocodilians where the legs are positioned to the side. If we then eliminate the sister group, the pterosaurs, which are easily distinguishable, we can usually spot a dinosaur with relative ease from another organism.

Accepting then that modern paleontological evidence is explicit that birds are dinosaurs, rather than considering them as a group of animals that the Chicxulub meteor wiped out, dinosaurs are, in point of fact, survivors of numerous extinction events including the latest - the K-Pg (Cretaceous-Paleogene) extinction event. This segment describes how groups mostly in Kent and the Royal Veterinary College combined molecular cytogenetics and bioinformatics to establish that this resilience and ability to recover from extinction event may be due, at least partly, to their karyotype.

The Dawn and Dusk and Dawn of the Dinosaurs

As illustrated in Figure 6, around 325 million years ago, the amniote lineage split into Synapsids (ultimately becoming mammals, amongst other groups) and the reptile/bird lineage – Diapsids. There are over 17,500 living diapsid species, majority of which (around 11,000) are birds. Dinosaurs (including birds) pterosaurs, turtles and crocodilians all share a common ancestor that lived 275 million years ago (Shedlock and Edwards 2009; Hedges et al. 2015), with the turtles (testudines) diverging first (around 255 million years ago), the crocodilians around 252 million years ago, the pterosaurs about 245 million years ago and the true dinosaurs about 240 million years ago. For the next 30 million years, dinosaur species were few in number, but amid the Jurassic period, this number, their geographical spread and their body size all dramatically increased (Benton et al. 2014). The proceeding 135 million years of dinosaur evolution is spectacular for being a time when they had an incredible range of species diversity and thus became the dominant vertebrates on the planet. Once human beings evolved scientific investigation, popular culture and media however, then the dinosaurs' legacy was complete. Incredibly, dinosaurs survived the Carnian-Norian and End-Triassic mass extinction events (228 and 201 million years respectively) and now there are >1000 discovered fossil species. About 30 more appear annually, not including birds, in the fossil record (Weishampel 2004). The devastation of the K-Pg extinction event 66 million years ago nearly wiped them out, but they bounced back again as modern birds, with more species than any other terrestrial vertebrate (Fig. 6). The cytogenomic studies of birds is an independent line of enquiry to paleontology, and circumvents some of the problems associated with fossil dating.

The remarkable diversity and species abundance that is observed in dinosaurs is often put down to the fact that competitor species were wiped out, thereby allowing the dinosaurs to thrive. It has nonetheless also been proposed that such impressive levels of abundance and diversity reflect genomic adaptations that are particular to dinosaurs, facilitating their survival over other species in harsh environments. Some examples include

extraordinary bone growth rates and highly evolved respiration systems (Farmer & Sanders 2010), including unidirectional respiration (O'Connor & Claessens 2005). These sorts of evolutionary adaptations could have led to the evolutionary success of avian species, clues of which may be found in their genome structure and organisation.

Dinosaur (Bird) Evolution Just Before, During and After the Last Mass Extinction Event as Revealed by Multiple Genome Sequencing Efforts

As a result of a multiple genome sequencing effort in birds (Zhang et al. 2014a; Jarvis et al. 2014), a revised avian phylogeny based on genome assemblies (some of which were CLAs) corrected the timing of avian diversification. We can now consider the 1st avian evolutionary divergence at about 100 million years ago with the Paleognathae (Ratites/Tinamous) branching from the Neognathae (Galloanseres/Neoaves). The 2nd is where Galloanseres (*Galliformes* and *Anseriformes*) and Neoaves then diverged 80 million years ago, with *Galliformes* (landfowl e.g. chicken, turkey, quail, pheasant) and *Anseriformes* (waterfowl e.g. geese, ducks, swans) diverging about 66 million years ago. Another major divergence of the Neoaves into Columbea (e.g. pigeons) and Passarea (e.g. songbirds) are dated slightly earlier (67-69 million years ago). Data from the Jarvis et al. (2014) analysis plus that of Prum and colleagues (Prum et al. 2015) suggest that, following the K-Pg mass extinction event (Schulte et al. 2010), around the same time of these two divergences, there was a rapid period of diversification, with 36 lineages evolving in a very short evolutionary period of 10-15 million years (Jarvis et al. 2014). Genomics has thus updated our understanding of dinosaurs through comparative studies, proving intriguing insights into relationships with diversity and phenotype (Zhang et al. 2014a; Jarvis et al. 2014). The karyotype of dinosaurs was something therefore that warranted deeper investigation.

Dinosaur Karyotype Evolution

The extraction of intact DNA from Jurassic blood-sucking insects is an outstanding plot device for novelists and film-makers but, alas, not a feasible means of facilitating the making of metaphase preparations. We can, however, with enough avian CLAs, glean insight into extinct dinosaur karyotypes by inference. Analysis of (near) chromosome-level assemblies from six living birds plus an *Anolis* lizard outgroup allowed us to infer the most likely ancestral karyotype of all birds (Romanov et al. 2014). Generally speaking, the common avian ancestor was probably a bipedal, terrestrial, chicken-sized small Jurassic dinosaur with some flying ability (Witmer 2002) and our studies established that its karyotype was very similar to that of a chicken or a Ratite bird. We then went on to reconstruct the most parsimonius sequence of events that led to the karyotypes we typically observe in avian species. The humble chicken (*Gallus gallus*) is in fact the closest, karyotypically, to the reconstructed ancestral pattern among the birds that we studied, whereas zebra finch (*Taeniopygia guttata)* and budgerigar (*Melopsittacus undulatus)* appear to have undergone the most intra- and inter- chromosomal changes respectively (Romanov et al. 2014). We later reconstructed the ancestral avian karyotype using an algorithmic approach, DESCHRAMBLER, on fragmented genome assemblies. There, (Damas et al. 2018) we performed large-scale analysis of ancestral avian chromosome structure around 14 key nodes of bird evolution. Our results provided insight into the variability in the rates of rearrangement that occurred during avian evolution. It also allowed us to detect patterns related to chromosome distribution of EBRs and microchromosomes.

In the same year, we applied a comparable approach to recreate the most likely ancestral karyotype of diapsids (O'Connor et al. 2018c). We developed a universally hybridising bacterial artificial chromosome (BAC) FISH probe set that was able to hybridise directly across species that diverged hundreds of millions of years ago (Damas et al. 2017). The BAC probes employed in FISH experiments gave strong signals on lizard (*Anolis carolinensis)* chromosomes and more so on chromosomes of the red-eared slider (*Trachemys scripta*) and spiny soft-shelled turtle (*Apalone spinifera*). FISH experiments then allowed us to anchor the series of events from the perspective of an archelosaur (bird-turtle) ancestor*.* By combining molecular cytogenetics and bioinformatics, we then recreated the cytogenetic changes that occurred from the ancestral diapsid ancestor, through to the archelosaur ancestor (Benton et al. 2015), through the theropod lineage, through to birds - including of course chicken. Screenshots of these events are depicted in Figure 7 in the panels that represent an animation (online suppl. Material 3).

Hybridization of BACs to *Trachemys scripta* (2n=50) and *Anolis carolinensis* metaphases (2n=36) also unveiled larger chromosomes with microchromosomal homologues attached, giving clues to the ancestral pattern of the diapsid ancestor (see Fig. 6). Figure 7 thus depicts a diapsid ancestral karyotype (275 million years ago) with 2n=36-46 (50:50 macro- and micro-chromosomes) (Beçak et al. 1964; Alföldi et al. 2011). This underwent rapid change over ~20 million years and we established that most of the major features associated with a typical bird

karyotype were already laid down in the archelosaur ancestor 255 million years ago. This is because most chicken (for this, read ancestral avian) chromosomes (numbered 1–28 + Z) are precisely syntenic to those of *Apalone spinifera* (2*n*=66). Analogous studies using chicken chromosome painting on Chinese soft-shelled turtle (*Pelodiscus sinensis*) (2n=66) (Matsuda et al. 2005), *Trachemys scripta* (Kasai et al. 2012) and painted turtle (*Chrysemys picta*) chromosomes (both 2n=50) (Badenhorst et al. 2015) further point to the notion that bird and turtle macrochromosomes are precise counterparts of one another. From this basic pattern that was laid down 255 million years ago in the archelosaur ancestor, only about 7 fissions would be needed to establish the familiar karyotype pattern directly observed in most of the major groups of birds including the Ratites, Galliformes, Anseriformes, Columbaea, Passeriformes and others. Determining the precise timings that these "final" chromosome fissions occurred cannot be easily achieved with the available evidence. However, if the same fission rate that had occurred for the previous 20 million years continued for another 15 million years, a complete bird-like karyotype would have emerged before the appearance of the earliest dinosaurs and pterosaurs 240 million years ago (Baron et al. 2017).

To paraphrase, our evidence strongly suggests that *the chromosomal pattern that we see in the majority of bird species that we choose to karyotype has remained mostly unchanged, not only in most birds, but with a reasonable degree of certainty, in many, if not most, extinct dinosaurs too* (O'Connor et al. 2018b and unpublished results). We thus go so far as suggesting that if we had the opportunity to make chromosome preparations "a la Jurassic Park/World" from the blood of extinct dinosaurs then karyotype analysis and zoo-FISH results would differ very little from that of a modern chicken (Figure 8). Those most closely related to modern birds such as the theropods (Spielberg favourites *Tyrannosaurus rex* and *Velociraptor* are representatives) are the most likely to have a very avian-like karyotype.

Karyotype, the Reduction in Genome Size and the Evolution of Flight
Dave Burt (Burt 2002) suggested that some avian microchromosomes were present in the avian ancestor >80 mya (Cracraft et al. 2015), purporting that it maybe had a karyotype of around 2*n*=60. A very worthy summation of the evidence at the time or writing. As detailed above however, on the basis of our new evidence over 15 years later, we challenged this notion. We also challenged the notion that this fragmented genome organization (i.e. a karyotype with 2n=80 chromosomes accompanied the genome size reduction in birds that has sometimes been associated with the evolution of flight (O'Connor et al. 2018c). That is, it had previously been suggested that there was some evidence of a correlation between genomes with fewer chromosomes (and no microchromosomes) and larger genome sizes (2.5-3Gb) such as is seen in mammals and crocodilians (Kapusta et al. 2017, St John et al. 2012). We however suggest that the avian karyotype was in place first, and that this was followed by a reduction in genome size, which was then followed by the evolution of flight.

Why Not Change?
Why then is a near identical karyotype pattern present in most birds, and why has it been in place for around 255 million years? Conceptually there could be two reasons: that is, either there is little opportunity for change and/or the configuration is so evolutionary successful that there is no pressure to change. In the case of the former, repetitive elements provide substrates for interchromosomal rearrangement, often seen in mammals but hardly even seen in avian species, suggesting that the bird karyotypes provide fewer opportunities for interchromosomal rearrangement because there are fewer recombination hotspots (Kawakami et al. 2014; Smeds et al. 2016), repeat structures (Mason et al. 2016; Warren et al. 2017; Gao et al. 2017), or endogenous retroviruses (Romanov et al. 2014; Cui et al. 2014; Farré et al. 2016). We also provided evidence of purifying selection acting on some of the smallest microchromosomes (Damas et al. 2018). A karyotype that hardly changes for 255 million years also however suggests that it is an evolutionarily successful one. The large chromosome number (especially microchromosomes) with high rates of recombination, could, we hypothesise, be the cause of the great variation we see in dinosaurs (including birds) mediated through random chromosome segregation and increased genetic recombination. Phenotypic variation is, of course, the driver of evolution and, although having many chromosomes is not the only mechanism through which variation can be generated, it may nonetheless explain this apparent paradox in dinosaurs of remarkable phenotypic diversity but very little karyotypic diversity. We should nonetheless recognise that it is possible (indeed likely) that some dinosaurs underwent a lot of interchromosomal change. Kingfishers (Christidis 1990) (many fissions), parrots (Nanda et al. 2007; O'Connor et al. 2018a) and falcons (Damas et al. 2017; Joseph et al. 2018) (many fusions) are modern examples of where this has occurred. Which specific extinct dinosaur groups did this, may however always remain a mystery.

Chromosome Inversion, the Role of Gene Ontology (GO) Analysis and Dinosaur Phenotype

In the absence of interchromosomal change, the principal mechanism for chromosomal change in dinosaur genome evolution was probably chromosome inversion (also depicted in Figure 7). Using the ancestral genome reconstruction tool Multiple Genome Rearrangement and Analysis (MGRA, (Avdeyev et al. 2016)), we generated contiguous ancestral regions (CARs) likely to represent the chromosomes of the diapsid ancestor. Compared to extant birds, we identified inversions along the path from the diapsid ancestor to the modern chicken, probably under-estimating this number of inversions at 49 chromosome inversions. We believe that the rate of intrachromosomal change increased in modern times, even in the chicken (Romanov et al. 2014). An even greater degree of change was seen however in some bird clades, particularly the songbirds (Skinner and Griffin 2011; Zhang et al. 2014b; Farré et al. 2016), the group with the most species. It seems reasonable to hypothesise therefore that periods of faster speciation may have also been accompanied by increased chromosome inversion rates in other dinosaur groups (Skinner and Griffin, 2012; Romanov et al 2014; O'Connor et al 2018c).

We (O'Connor et al. 2018c), identified around 400 HSBs delineated by EBRs that characterize dinosaur genome evolution. Other genomic studies in other species (mostly mammals) established that EBRs commonly occur in gene-dense loci, with genes related to lineage-specific biology, transposable elements and other repetitive sequences (Pevzner and Tesler 2003; Hillier et al. 2004; Larkin et al. 2009, Rao et al. 2012). HSBs on the other hand contain more developmental genes and regulatory elements (Larkin et al. 2009; Warren et al. 2017). Regions more likely to break, e.g. open chromatin areas or recombination hotspots, and chromosome breaks that do not disrupt key genes or provide a selective advantage, are more likely to be fixed in populations (Farré et al. 2016).

HSB analysis (O'Connor et al. 2018c) using gene ontology (GO) tools, established significant enrichments relevant to amino acid transmembrane transport and signalling plus synapse/neurotransmitter transport, nucleoside metabolism, cell morphogenesis and cytoskeleton and sensory organ development. Former studies established that HSBs are enriched for GO terms related to evolutionary constant phenotypic features (Larkin et al. 2009) and our dinosaur results support this hypothesis. EBRs on the other hand, are often proposed to be where the "action" in genome evolution occurs (Sankoff 2009). We initially found GO terms in avian EBRs that were associated with specific adaptive features, e.g. enrichment for forebrain development in the budgerigar EBRs (consistent with vocal-learning) (Farré et al. 2016). Later, we identified significant enrichments in genes and single GO terms pertaining to chromatin modification, chromosome organization and proteasome/signalosome structure (O'Connor et al. 2018c).

The discovery that the avian karyotype likely dates back before the dawn of the dinosaurs complements paleontological research that demonstrates that features such as feathers and pneumatised skeletons arose first among more ancient dinosaur or archosaurian ancestors (Zhou 2004; Baron et al. 2017). Dinosaurs were the dominant group of animals for around two hundred million years, with significant radiations occurring in response to two mass extinction events and, despite being almost wiped out by a third (K-Pg), their resilience as a highly diverse and specious clade (extant birds) (Barrowclough et al. 2016) is evident.

Conclusions

Far from being simply a curating exercise (and a speculative one at that), the study of likely dinosaur karyotype sheds new light into genome evolution, with distinct clues about phenotype and an alternative line of enquiry compared to more established methods. A highly fragmented genome that appears like it has been hit by a meteor is ironic since its pattern was most likely established around 200 million years before Chicxulub hit the earth. Our studies reveal an apparent paradox of an organization that is remarkably unchanging karyotypically during evolution, yet, quite possibly, the driver of such phenotypic evolutionary change. Did the fictional creators of Jurassic Park ever attempt karyotyping? Well, apparently, they did: Check out the last of the Jurassic World movies: at one point, an historical recording of the (now deceased) character Charlotte Lockwood appears, observed by her young daughter Maisie. In the clip, there is a screen with chromosomes on them in the background. There are however no microchromosomes! Unless Spielberg knows something that we don't, these chromosomes were unlikely to be dinosaur in origin.

Chicken as a Model for Bird Genome Alignments

(Prepared by P.D. Waters, J.A.M. Graves, and H.R. Patel)

Chicken has long been a model for bird genomes (International Chicken Genome Sequencing 2004), but it is only in recent years that a chromosome level assembly has been available that includes the gene dense microchromosomes.

The chicken karyotype is characterised by nine macrochromosomes (defined as >35Mb in the genome assembly), which includes the Z, and 30 microchromosomes (defined as <35Mb) that are characteristic of bird genomes (Mendonça et al. 2016). Like all birds, chicken has a ZW female ZZ male sex chromosome system, in which sex is determined by dosage of the gene *DMRT1* on the Z chromosome (Smith et al. 2009).

Both the macro- and microchromosomes are conserved across the most distantly related bird lineages (Waters et al. 2021), with the exception of some clades that have undergone rearrangement (Huang et al. 2022; Nanda et al. 2006) (Figure 9). In fact, strong homologies are conserved with reptiles (lizards, snakes, turtles and alligators), with differences mostly explained by chromosome fusions (less often fissions) so that large regions of synteny are retained (Waters et al. 2021).

Strikingly, it was demonstrated that almost all amphioxus chromosomes shared homology with one or two bird microchromosomes, in addition to two or three regions of macrochromosomes (Waters et al. 2021). These multiple regions of amphioxus homology to the bird genome are likely to represent the four copies resulting from two rounds of whole genome duplication in vertebrates (Simakov et al. 2020). The fact that each amphioxus chromosome shared striking homology with one or two bird microchromosomes suggests that each bird microchromosome represents one copy of an ancestral chromosome.

Microchromosomes are CG rich and gene dense (Waters et al. 2021), as are amphioxus chromosomes. Additionally, microchromosomes cluster together in the nucleus and have many more interactions with each other than with the macrochromosomes (Liu et al. 2021; Waters et al. 2021). Therefore, they are physically isolated from macrochromosomes, so are less likely to fuse with them.

It is simplest to envisage that after genome duplication one complete copy was rapidly packaged up in the interior of the nucleus and isolated from the other genome duplicate, rather than a mix of chromosomes from duplications being isolated. Given their tight association in the nucleus, it is somewhat surprising microchromosomes have not all fused with each other in birds, as they have in crocodiles and mammals. This may have resulted from the selective advantage of higher recombination due to obligate cross overs on each chromosome and their independent segregation at meiosis.

Here we report whole genome alignments of the chicken genome with those of other bird representatives with chromosome level assemblies. The chicken genome aligns very well with the emu genome, which is considered to represent the ancestral bird genome (Liu et al. 2021). The only whole chromosome difference is the fusion of a microchromosome to chromosome 4, uniquely in chicken (Burt 2002; Shetty et al. 1999). There have been some small internal rearrangements at the termini of macrochromosomes, two that are chick-specific and three that are shared with other carinate clades so presumably occurred in the ancestor of all carinates. The Z chromosome has two large internal rearrangements with respect to the emu Z, but these are shared by other carinates.

As a gold standard genome, we propose that the chicken assembly is the best model for the bird genome. It has few autosomal rearrangements relative to the ancestral genome, and four of these are shared by other carinates. All microchromosomes are represented in the most recent assembly. Therefore, whole genome alignments to it can be used to assess assembly errors and/or real rearrangements in new bird assemblies for which karyotype information is available. This presents a more robust comparison than another popular bird model, the zebra finch, which has undergone macrochromosome fission and internal rearrangement (Figure 9). It is also more useful for aligning the genomes of birds with very rearranged chromosomes, since the multiple rearrangements are independent in different clades (e.g. golden eagle and falcon/parrot, Figure 9).

The Avian W Chromosome: Simply a Y with a Different Name?
(Prepared by P.D. Price, T.F. Rogers, and A.E. Wright)

Sex chromosomes have long fascinated biologists due to their unique gene content and evolutionary trajectories relative to the rest of the genome (Furman et al., 2020). In particular, the halting of recombination between sex chromosome pairs has resulted in the evolution of highly-degenerate sex-limited W and Y chromosomes in many species (Charlesworth 1991). Identifying the function of these chromosomes and understanding if and how they can resist the degenerative forces arising from reduced recombination has been the focus of numerous studies (Bachtrog, et al. 2011).

We now know a considerable amount about Y chromosomes, despite the difficulties in sequencing highly heterochromatic and repetitive genomic regions (Tomaszkiewicz, et al. 2017). Their evolution is typically characterised by the accumulation of genes with male-specific functions, large-scale gene amplification, and rapid turnover of gene content across lineages (Bachtrog 2013; Subrini and Turner 2021). In contrast, our understanding of the W chromosome has lagged. However, the last decade has seen an explosion in the number of W-linked genes sequenced across birds (Zhou, et al. 2014; Bravo, et al. 2021), ranging from songbirds (Smeds, et al. 2015; Xu, et al. 2019; Sigeman, et al. 2021; Huang, et al. 2022; Warmuth, et al. 2022) to fowl (Moghadam, et al. 2012; Ayers, et al. 2013; Wright, et al. 2014) to paleognaths (Xu and Zhou 2020; Liu, et al. 2021), and our first sight of a reference assembly of the chicken W chromosome (Warren, et al. 2017). In theory, W chromosomes are in many ways comparable to Y chromosomes, as both are sex-limited and often don't recombine, and so they might be expected to share similar evolutionary fates. However, there are key differences, most notably that the W chromosome is limited to females whereas the Y chromosome is only present in males (Bachtrog, et al. 2011; Mank 2012). Below, we outline new insights into avian W chromosomal evolution and ask whether W and Y chromosomes are really that different.

What Are the Evolutionary Dynamics of W Chromosomes across Birds?

It has been known for decades that the chicken W chromosome is a degenerated version of the Z, with the most recent build of the W reference (GRCg7b) identifying only ~80 protein-coding genes across ~9Mb (Warren, et al. 2017). However, establishing whether the chicken W chromosome is representative of the avian W more generally has only recently been possible due to the plethora of W-linked sequences now available across the avian phylogeny (Wang, et al. 2014; Zhou, et al. 2014; Xu, et al. 2019; Sigeman, et al. 2020).

Sex chromosomes diverge as recombination is suppressed between them, typically assumed to occur in a stepwise process through sequential inversions (Charlesworth, et al. 2005). Consistent with this, 'strata' of different ages can be detected on avian Z and W chromosomes (Wang, et al. 2014; Wright, et al. 2014). These strata are thought to reflect the halting of recombination through large-scale chromosomal rearrangements, such as inversions (Wright, et al. 2016). However, recent evidence suggests that recombination suppression at the earliest stages of avian sex chromosome divergence is a more mosaic and gradual process (Sigeman, et al. 2021). This cessation of recombination has occurred independently in different avian lineages (Zhou, et al. 2014) and many species exhibit a heavily degraded W chromosome, similar to the chicken.

Despite degeneration proceeding independently across birds, the set of ancestral genes retained on the W chromosome is remarkably conserved, suggesting that decay is non-random (Xu and Zhou 2020). For instance, over 80% of W-linked genes in the oldest stratum are conserved across chicken, songbirds and tinamous (Xu and Zhou 2020). This is in stark contrast to the Y chromosome, where frequent gene movement onto and off the chromosome is common (Hughes, et al. 2015; Mahajan and Bachtrog 2017). This has been interpreted as a product of differing selective pressures acting on Y versus W chromosomes, with the W chromosome subject to stronger purifying selection compared to the Y due to its higher effective population size relative to the autosomes (Wright and Mank 2013).

However, there is still remarkable variation in the extent of Z-W divergence across birds (Zhou, et al. 2014). In contrast to the chicken, the paleognath W chromosome recombines along a large proportion of its length and so has experienced limited decay, although this varies across species (Zhou, et al. 2014; Yazdi and Ellegren 2018; Liu, et al. 2021) with greater recombination suppression in ostrich and emu than tinamou. The growing amount of long-read sequencing data for birds has also revealed that fusions between sex chromosomes and autosomes to create neo-sex chromosomes are not uncommon, with two independent origins across Psittaciformes (Huang, et al. 2022), four across Sylvioidea (Pala, et al. 2012; Sigeman, et al. 2020; Sigeman, et al. 2021), one in the eastern yellow robin (*Eopsaltria australis*) (Gan, et al. 2019), cuckoo (*Crotophaga ani*) (Kretschmer, et al. 2020) and Raso lark (*Alauda razae*) (Dierickx, et al. 2020) identified to date. No doubt this number will increase as more bird genomes are probed for sex chromosomes, making it possible to test the evolutionary pressures responsible for driving these fusions. Together, this challenges the traditional view that the avian W chromosome is genetically inert and highly conserved across species.

Is the Avian W Chromosome Selected for Female-Specific Functions?

Given the sex-limited inheritance pattern of Y and W chromosomes, theory predicts that they should be subject to sex-specific selection and accumulate genes with sex-specific functions (Rice 1984). Indeed, the Y chromosome in many species is enriched with genes predominantly expressed in testes that function in spermatogenesis (Bachtrog 2013; Subrini and Turner 2021), although there is a growing awareness of its role in non-reproductive traits (Cīrulis, et al. 2022). It follows that we might expect the W to be subject to female-specific selection to retain genes with female fitness benefits.

The avian W chromosome lacks a candidate sex-determining gene (Schmid, et al. 2015). Instead, sex in birds is determined by dosage of the Z-linked gene, *DMRT1* (Hirst, et al. 2017). However, there are several lines of evidence implicating the avian W chromosome in female fertility, although the precise functions of genes on the W have yet to be defined and expression patterns quantified. First, W-linked genes are highly expressed in developing chicken ovaries (Moghadam, et al. 2012; Ayers, et al. 2013; Xu and Zhou 2020). This is consistent with feminization of the W (Mank, et al. 2010), as key stages of oogenesis are restricted to embryogenesis unlike spermatogenesis, which is a continuous process throughout adult life. Second, W-linked genes expressed during late female development are convergently upregulated in the ovaries of chicken layer breeds subject to artificial selection for fecundity relative to their modern ancestor, the Red Jungle Fowl, and other chicken breeds (Moghadam, et al. 2012).

However, unlike most Y-linked genes, which typically exhibit testes-specific expression (Subrini and Turner 2021), expression of genes on the avian W chromosome is not limited to the ovary. Instead, many are active in both somatic and reproductive tissue during chicken embryogenesis (Xu and Zhou 2020) as well as across adult chicken (Bellott, et al. 2017) and collared flycatcher tissue (Smeds, et al. 2015). While this does not preclude a specific role of the W chromosome in oogenesis, it has led to suggestions that this chromosome has instead been selected to maintain gene dosage and ancestral expression levels of essential genes. Consistent with this, many avian W-linked genes are subject to purifying selection (Wright, et al. 2014; Sigeman, et al. 2021), exhibit a high degree of sequence conservation as well as similar expression patterns to their Z-linked partner (Ayers, et al. 2013; Smeds, et al. 2015; Xu and Zhou 2020), and have human orthologs that exhibit detrimental effects when haploid (Bellott, et al. 2017; Xu, et al. 2019; Bellott and Page 2021; Sigeman, et al. 2021).

It is plausible that apparent differences in the function of Y and W chromosomes could arise from their contrasting inheritance patterns. For instance, W-linked genes, which only pass through the female germ line, are not exposed to sperm competition and so might be subject to weaker sex-specific selection than genes on the Y chromosome. However, it is worth noting that our understanding of the function of the avian W is based on expression data from a limited number of species (chicken, flycatcher) taken across whole, heterogeneous, adult tissue. This precludes accurate contrasts of expression between Z and W orthologs (Price, et al. 2022) and so could lead to false inferences of selection to maintain gene dosage between gametologs. Further expression analyses, incorporating a broader taxonomic range and data for individual cell types throughout female development, are essential to ascertain why specific genes have been retained on the avian W chromosome.

How Do Multi-Copy Gene Families Evolve on the W Chromosome?

Y chromosome degeneration is frequently characterised by massive gene amplification where many remaining Y-linked genes persist as members of multi-copy gene families (Skaletsky, et al. 2003; Soh, et al. 2014; Bachtrog, et al. 2019; Vegesna, et al. 2020). However, until recently, gene amplification on the W chromosome has received comparatively less attention and it remained unclear whether large-scale gene amplification is a general feature of sex chromosome evolution or a peculiar quirk of the Y.

A handful of W-linked multi-copy gene families have been identified in a limited number of avian species (Hori, et al. 2000; Backström, et al. 2005; Davis, et al. 2010; Smeds, et al. 2015; Rogers, et al. 2021). The most comprehensively studied of these is histidine triad nucleotide-binding protein (HINTW), an ampliconic gene family that is hypothesized to play a role in female reproduction and oogenesis (O'Neill, et al. 2000; Ceplitis and Ellegren 2004). At present, ~10 different copies of *HINTW* are annotated in the most recent chicken W chromosome assembly (GRCg7b), however, this is likely an underestimation with previous studies estimating over 40 copies (Hori, et al. 2000; Backström, et al. 2005). Furthermore, large-scale amplification of *HINTW* is conserved across avian non-ratites (Hori, et al. 2000). Currently, evidence for the functionality of HINTW is lacking. However, it is known that HINT can form a heterodimer and the amino acid residues that form the dimer binding site are conserved in many HINTW copies, although many copies are nonfunctional (Hori, et al. 2000; O'Neill, et al. 2000). Therefore, HINTW might act to disrupt the function of its Z-linked ortholog (*HINTZ*) by forming a heterodimer. Interestingly, the size of the *HINTW* gene family varies between chicken layer breeds subject to artificial selection

for fecundity relative to other chicken breeds (Rogers, et al. 2021), potentially suggesting a role of female-specific selection in driving gene amplification, although this relationship was absent across duck breeds.

The paucity of multi-copy gene families on the avian W chromosome is in stark contrast to the abundance of ampliconic genes often present on the Y chromosome. Several mechanisms have been proposed to drive the evolution of multi-copy gene families on the Y, including meiotic drive, sperm competition, genetic drift, and gene conversion (Skaletsky, et al. 2003; Ellis, et al. 2011; Cocquet, et al. 2012; Larson, et al. 2018; Bachtrog, et al. 2019; Vegesna, et al. 2020). In theory, the strength of these processes might differ between the Y and W due to their contrasting inheritance patterns (Wright and Mank 2013). Notably, the Y chromosome is exposed to spermatogenesis, whereas the W is subject to oogenesis, and this likely leads to differences in the potential for antagonistic co-evolution between the X and Y versus the Z and W. Antagonistic co-evolution is predicted to drive the co-amplification of genes on sex chromosomes but should be weaker during oogenesis than spermatogenesis, potentially explaining the limited number of W-linked multi-copy gene families (Bachtrog 2020). Targeted avian gene knockouts (Ioannidis, et al. 2021) provide an exciting opportunity to elucidate the functionality of HINTW copies, whether this varies across avian species, and the potential for antagonism between W and Z orthologs.

Is There a "Toxic W" Effect?

There appears to be a cost for males to carrying a degenerated Y chromosome (Brown, et al. 2020; Xirocostas, et al. 2020; Nguyen and Bachtrog 2021; Connallon, et al. 2022), where males in species with XY chromosomes tend to die earlier (Xirocostas, et al. 2020). Several hypotheses have been put forward to explain this phenomenon, including the presence of deleterious recessive mutations on the single X in males that would otherwise be shielded in females ('unguarded X') or the accumulation of mutations and repetitive elements on the Y chromosome ('toxic Y'). There is also evidence that the Y chromosome acts as a heterochromatin sink, reducing the efficiency of heterochromatin maintenance across the rest of the male genome (Francisco and Lemos 2014; Brown, et al. 2020).

Similar processes may operate on the W chromosome, where females exhibit a shorter lifespan than males across a range of species (Xirocostas, et al. 2020). Consistent with a 'toxic W' effect, the avian W chromsome is a haven for repetitive material and transposable elements in several species. For instance, females in species with a degenerate W carry between 20-90% more endogenous retroviruses than males (Peona, et al. 2021). Furthermore, transposable element suppression is less effective on the crow W chromosome than the rest of the genome, leading to higher expression of transposable elements in females (Warmuth, et al. 2022). Although transposable elements can facilitate adaptive evolution, they also have the potential to reduce fitness through the disruption of gene activity and the promotion of chromosomal rearrangements (McDonald 1993). In theory, they may also contribute to an increased chance of female sterility in hybrids, where mechanistic mismatches between transposable repressor mechanisms and the W chromosome leads to reduced female fertility. This would provide further support for Haldane's rule where the heterogametic sex is more likely sterile in hybrids (Haldane 1922).

Final Remarks

Recent studies have provided new insight into avian W chromosome evolution, challenging the traditional view that the avian W is genetically inert and highly conserved across species. There are clear parallels with Y chromosomal evolution but also key differences, primarily regarding the relative importance of the W in reproduction and fertility traits. Recent technological advances offer new potential to resolve this uncertainty, for instance by using single-cell RNA-seq to establish fine-scale expression patterns of Z and W-linked genes through development and across species (e.g. Estermann, et al. 2020) and targeted gene knockouts to test gene function (e.g. Ioannidis, et al. 2021). Therefore, the next couple of years hold much promise for disentangling the function and evolution of the W chromosome in birds.

Recovering the "Missing" Avian Genes Using Multi-omics Data

(Prepared by Z.-T. Yin, J. Smith, and Z.-C. Hou)

Gene gain and loss are common events in the evolution of species, especially for birds, which have evolved many unique characteristics such as feathers, wings, and flight capabilities, strong and lightweight skeletons, toothless beaks, high metabolic rates, and heat absorption sex, and unique respiratory and excretory systems [Kennedy and

Vevers, 1976, Blomme et al., 2006]. The release of the first chicken genome provided the basis for systematic analysis of the similarities and differences between vertebrate and avian genomes [International Chicken Genome Sequencing et al., 2004]. In comparison with other amniotes, bird genomes are more compact, and this difference may be related to the overall smaller cell size [Hughes and Hughes, 1995, Hughes and Friedman, 2008]. The reductions in genome size may be the result of the loss of noncoding DNA sequences, with bird genomes having less repetitive DNA, fewer pseudogenes, and shorter introns than mammalian genomes [Hillier et al., 2004, Hughes and Piontkivska, 2005]. Importantly, the evolution of avian genomes also appears to involve the loss of protein-coding genes, as the total number of uniquely identified avian-coding genes is much smaller than in other tetrapods (i.e., 23,294 in humans, GRCh38.p14; 19,404 in lizards, AnoCar2.0; 17,007 in chickens, GRCg7b). Paralog analysis revealed a higher overall incidence of gene families with fewer members in birds compared to other vertebrates [Hughes and Friedman, 2008]. Likewise, birds have a high rate of chromosomal rearrangements compared to other organisms, all of which may result in the deletion of protein-coding genes [Backstrom et al., 2010]. In recent years, the genomes of a large number of birds and lizards have been assembled and annotated, including zebra finches [Warren et al., 2010], chickens [Hillier et al., 2004], turkeys [Dalloul et al., 2010] and duck [Zhu et al., 2021]. Moreover, large-scale bird genome projects [Jarvis et al., 2014, Zhang et al., 2014], and chicken pan-genomes [Wang et al., 2021, Li et al., 2022] have also generated considerable genomic data. These large comparative genomic datasets identified hundreds of lost genomic-blocks in the bird genomes, and also suggested that hundreds of genes are missing in birds [Lovell et al., 2014, Zhang et al., 2014].

The missing genes seem to be directly related to the unique physiological phenomena of birds. Several functionally important genes in mammals are supposed 'missing' in chickens and have caused long-debated questions in bird biology. Spurious discovery of the missing/hidden genes in the bird genome has continued for decades. Previously, *BGN* [Blaschke et al., 1996], *CORO1A* [Xavier et al., 2008], *MAPK3* [Lemoine et al., 2009]*, MMP14* [Simsa et al., 2007]*, TBX6* [Lardelli et al., 2003, Ahn et al., 2012]*, TSSK4* [Shang et al., 2013] and five adipokine genes [Dakovic et al., 2014] were reported to be missing in birds, however, several long-debated genes including *TNF-alpha*, and *leptin* have been cloned in birds [Prokop et al., 2014, Seroussi et al., 2016, Rohde et al., 2018]. This hide-and-seek game still continues, and does not appear to be ending anytime soon [Elleder and Kaspers, 2019]. Here we summarize recent efforts using multi-omics data to probe those genes missing/hidden in avian genomes.

Reconstruction of Missing Genes in the Chicken Genome
While the hypothesis of missing genes in birds has been proposed for decades, researchers have found that some of the missing genes were, in fact, present in chickens or other birds. In the presence of large gaps and imperfect gene annotation in the genome, the *de novo* assembly of gene sequences using RNA-seq is considered to be an efficient way to identify unannotated genes in the genome. Attempts that only used a few tissues/organs have identified many missing genes in birds [Hron et al., 2015, Bornelov et al., 2017, Botero-Castro et al., 2017]. Recently, we used the raw data from 26 chicken tissues downloaded from the GenBank database to assemble and obtain 2,048,631 transcripts and identified 589 missing genes in birds [Yin et al., 2019].

At the same time, the continuity and integrity of chicken genome assemblies have been rapidly improving. The chicken genome released in 2017 was assembled by third-generation sequencing technology, and the number of annotated genes increased significantly (2,768 noncoding and 1,911 protein-coding genes) [Warren et al., 2017]. In the Gallus_gallus-5.0 genome, 442 (77.41%, from a total of 571) genes thought to be missing in chickens (in Lovell et al., 2014 see Table S1 and Table S6, plus select entries in Table S4 and Table S18) were annotated, indicating that there is no systematic deletion of genes in birds. With the development of sequencing and hybrid assembly technology, the genomes of different chicken breeds continue to be assembled and another 136 missing genes were further annotated in our recently assembled *Silkie* genome (unpublished). To date, it has now been shown that 528 (92.47%) genes that were thought to be missing, actually exist in chickens. This has been made possible by exploiting a large amount of multi-omics data available in chicken and has led to the revelation of genes with important functions such as *TNF-α* and *Leptin* [Seroussi et al., 2016, Rohde et al., 2018]. Recent large-scale chicken pan-genome data have also identified thousands of genes that are not presented in the current chicken reference genome [Li et al., 2022].

Reconstruction of Missing Genes from Other Birds
In addition to chicken, researchers have reconstructed many genes thought to be missing from other birds. We collected data from various important tissues from duck (24), pigeon (11), goose (8), and zebra finch (22) [Yin et

al., 2019], and an avian transcriptomic database containing a total of 9,296,247 transcripts was constructed by *de novo* transcriptome assembly. From this, we identified several genes in duck (583), pigeon (558), goose (537), and zebra finch (543) from 806 genes that were thought to be missing in birds (in Lovell et al., 2014 see Table S1 and Zhang et al., 2014 see Table S10). Only 135 genes were not found in this bird transcriptome database. The number of missing genes reconstructed in different birds by *de novo* assembly of large transcriptome data is similar, indicating that these genes thought to be missing exist across different bird species.

In recent years, duck functional genomics has developed rapidly. We have assembled the Mallard, Pekin duck, and Shaoxing laying duck genomes using a combination of third-generation sequencing, Bionano, and Hi-C sequencing technologies. These have proved to be a rich source of genetic information [Zhu et al., 2021]. In the Mallard duck the CAU_wild 1.0 genome has 1,872 more protein-coding genes annotated than the previous CAU 1.0 genome, including 89 genes previously thought to be missing in birds. Among these 89 genes, 5 genes have become pseudogenes, losing part of their gene function, 3 genes have been annotated as lncRNAs, and the remaining 81 genes remain as protein-coding genes. In addition, 240 genes were annotated as paralogous genes and 108 genes had similar segments in the genome. Mining large multi-omics data assemblies and annotations now reveals that only 10 genes (from a total of 806 missing genes), to date, have not been reconstructed in birds, with the rest of the genes thought to be missing in birds having been shown to actually exist. The recovered gene list is shown in online supplementary Material 4.

Development of New Methods to Identify More Missing Genes
Summarizing the characteristics of these reconstructed missing genes in birds and the reasons why they are thought to be missing can provide insights and methods for us to identify more missing genes. First, these reconstructed gene sequences have high GC content and length in many birds. The GC content of most of these 'missing' genes is more than 60%, and few genes even have over 80% (the median GC content of the chicken genome is 42.22% and the median GC content of the duck genome is 41.99%) [Hron et al., 2015, Bornelov et al., 2017, Botero-Castro et al., 2017, Yin et al., 2019]. At the same time, the multi-tissue transcriptome expression profiles of birds showed that most of the reconstructed genes usually have strong tissue-specific expression. These genes are generally expressed predominantly in one tissue and are rarely expressed in the other tissues [Yin et al., 2019]. High-throughput transcriptome-based assembly approaches have limitations for fully recovering missing genes due to technical factors such as the PCR amplification bias against GC-rich fragments [Beauclair et al., 2019]. Expression patterns, i.e. tissue-specific expression patterns, and low expression, also limit the ability for full transcriptome assembly. Now, the third-generation sequencing technologies, which have less GC bias, such as single-molecule real-time (SMRT) and nanopore sequencing technologies, can obtain full-length transcripts directly, without assembly [Yin et al., 2019; Kuo et al. 2020]. The missing genes will continue to be discovered with the accumulation of full-length transcriptome data from more avian tissues from different physiological conditions.

Furthermore, the missing genes annotated in the chicken and duck genomes are mainly distributed on the micro-chromosomes, the ends of the chromosomes, and within regions showing a high content of tandem repeats clustering with non-canonical DNA structures. [Zhu et al., 2021, Li et al., 2022]. Long repetitive regions [Treangen and Salzberg, 2011], regions of high GC content [Chen et al., 2013], telomeric regions, fragmented micro-chromosomes [O'Connor et al., 2019], and adaptive assembly strategies have always proved problematic for enabling complete bird genome assembly. To fully resolve the whole chicken gene sets, a Telomere-to-Telomere (T2T) genome is necessary. The recently completed human T2T genome has now paved the way for the finished bird genome assembly [Miga et al., 2020, Hoyt et al., 2022, Mao and Zhang, 2022, Nurk et al., 2022]. Ultra-long ONT sequencing, high-precision HiFi sequencing data, multi-type auxiliary assembly data, and hybrid assembly using multiple strategies will greatly promote the quality of bird genome assembly [Sohn and Nam, 2018]. For large presence/absence variations within species, we can enrich genomic information by constructing high-quality multi-breed pan-genomes [Vernikos et al., 2015]. The Bird 10,000 Genomes (B10K) Project [Zhang et al., 2015] has generated insightful results and the future bird T2T genome and pan-genome will undoubtedly reveal more genes. This complete gene map of birds will be critical for the further understanding of the biology and evolution of birds.

Finally, precise genome annotation will also provide the necessary sequence and structural information for mining more genes in birds. Annotation errors are unavoidable in genome annotation using automated processes, especially for some protein-coding genes that cannot be annotated in complex and high GC regions [Salzberg et al., 2019]. While applying full-length transcriptomic data for genome annotation [Nudelman et al., 2018, Wang et

al., 2019; Kuo et al., 2020], the use of novel annotation methods developed based on machine learning can further improve the accuracy of annotation [Mahood et al., 2020, Stiehler et al., 2020]. More accurate manual annotation of important genome regions is also necessary for novel gene identification [Dunn et al., 2019]. It can be seen that, with the continuous development of omics technology and analysis methods, the genome information will be more complete, the annotation will be more accurate, and the genes that were previously thought to be missing in birds will continue to be discovered.

Statement of Ethics. All experiments with birds were performed under the guidance of ethical regulations from the Animal Care and Use Committee of China Agricultural University, Beijing, China.

Conflict of Interest Statement. The authors declare no competing interests.

Data Availability Statement. Data have been submitted to the public databases under the following accession numbers: The raw data for assembling the transcriptome database of chicken, duck, goose, pigeon, and zebra finch were deposited in Sequence Read Archive (SRA) database under the accession number SRP141084. The Mallard genome is stored in NCBI under accession number PRJNA554956. The raw data for the Silkie genome assembly can be found in the SRA database under the accession number PRJNA805080 (Unpublished data).

Evolution of Protein-Coding and Long Non-Coding Genes of the Chicken Genome through the Different Genome Assemblies and Their Associated Annotations
(Prepared by F. Degalez, K. Muret, and S. Lagarrigue)

The chicken genome was the first avian genome sequenced [International Chicken Genome Sequencing Consortium 2004] because of its importance in human food production in fundamental biology like the study of development or gene function conservation across evolution [Hillier et al. 2004]. Since its first version in February 2004 (Galgal2/WUGSC1.0), five new genome assemblies have been released, each improving the genome sequence's accuracy. Along with these genome assemblies, numerous genome annotations were released, providing at least models for gene loci and transcripts supporting them. Since the first annotated version (Ensembl v22 - 25/05/2004) associated with the Galgal2 assembly, the number of genes and the diversity of their biotypes have increased, especially in 2015 with the introduction of long non-coding RNAs (lncRNAs), which is concurrent with the first initiatives of lncRNA annotation [Chodroff et al. 2010; Necsulea et al. 2014; Li et al. 2015; Muret et al. 2017].

LncRNAs represent a large and heterogeneous class of genes defined by transcripts longer than 200 nucleotides without coding-potential capabilities [Derrien et al. 2012]. They represent a variety of regulatory elements implied in gene expression and can act at different levels by using diverse biological mechanisms based on DNA, RNA, or protein interactions [Guh et al. 2020]. As illustrated in Figure 10, lncRNAs can interact with DNA, RNA, and proteins and act at different molecular levels: nuclear organization (*e.g.* MALAT1 [Wang et al. 2021b] / NEAT1 [Yamazaki et al. 2018] – Figure 10A), genome integrity (*e.g.* TERRA [Barral and Déjardin 2020] – Figure 10B), histone marks modification for silencing (*e.g.* Fendrr [Grote et al. 2013]) or activating (*e.g.* GATA3-AS1 [Gibbons et al. 2018]) gene transcription (Figure 10C), loop formation to connect enhancers to promoters regions (*e.g.* MYMLR [Kajino et al. 2019] – Figure 10D). LncRNAs can modulate RNA splicing (*e.g.* linc-HELLP [van Dijk et al. 2015] – Figure 10E), miRNA maturation (*e.g.* CCAT2 [Yu et al. 2017] / uc.372 [Guo et al. 2018] – Figure 10F) and protein translation (*e.g.* BC1 [Wang et al. 2002]/ MCM3AP-AS1 [Guo et al. 2020] – Figure 10I) or their activity (*e.g.* NORAD [Munschauer et al. 2018] – Figure 10K). They can also control the stabilization or the degradation of molecules as miRNAs (*e.g.* ROR [Li et al. 2017] / DSCR8 [Wang et al. 2018] – Figure 10G), mRNAs (*e.g.* PTB-AS [Zhu et al. 2019] / TINCR [Xu et al. 2015] – Figure 10H) and proteins (*e.g.* PiHL [Deng et al. 2020] / MALAT1 [Yan et al. 2016] – Figure 10J). LncRNAs can host small ORFs [Choi et al. 2019] which code for peptides (*e.g.* CASIMO1 [Polycarpou-Schwarz et al. 2018] / DWORF [Nelson et al. 2016] – Figure 10M) or host in their introns, small RNAs [Sun et al. 2021] (*e.g.* MCM7 [Agranat-Tamir et al. 2014] / DLEU2 [Morenos et al. 2014] – Figure 10N). They can control protein transfers from cytoplasm to nucleus (*e.g.* NRON [Willingham et al. 2005]) or from nucleus to cytoplasm (*e.g.* Discn [Wang et al. 2021a] – Figure 10L). Finally, they can migrate to other cells with exosomes (*e.g.* ZFAS1 [Pan et al. 2017] / GAS5 [Chen et al. 2017] – Figure 10O).

Through their key roles in gene regulation, lncRNAs are consequently involved in diverse biological and pathophysiological processes [Ponting et al. 2009; Muret et al. 2019; Gil and Ulitsky 2020; Statello et al. 2021]. Moreover, since most of the trait-associated variations identified by genome-wide association studies (GWAS) concerned non-coding intervals of the genome [Manolio et al. 2009; Bouwman et al. 2018], this reinforces the need to characterize the regulatory regions of domesticated species such as lncRNA genes. LncRNA genes have different characteristics compared to protein-coding genes (PCGs). They are less expressed [Derrien et al. 2012; Muret et al. 2017; Le Béguec et al. 2018; Jehl et al. 2020] explaining why they have been detected only recently – *i.e.* this last decade – by high-throughput transcriptome sequencing technologies (RNA-seq). Furthermore, lncRNA expression is more specific to tissues, life stages, and conditions than that of PCG [Cabili et al. 2011; Derrien et al. 2012; Jehl et al. 2020]. The identification of these genic entities is therefore dependent on the variety of RNA-seq data available to detect them.

After presenting the different chicken genome assemblies developed over the last two decades, we discuss the associated genome annotations provided by NCBI's RefSeq and EMBL-EBI's Ensembl, the two reference annotation databases. We characterize them in terms of number of gene and transcript models, variety of biotypes, or in terms of models that are shared by the two reference databases. We show that lncRNA loci are even less well known than PCG ones, although, for the latter, knowledge of their transcripts can be further improved. Finally, we discuss the impacts of these weaknesses and the value of gathering different genome annotation resources, in particular, for a better description of lncRNA loci, and then present two initiatives. The MANE project yet limited to the human genome aims to synergize the NCBI's RefSeq and EMBL-EBI's Ensembl "gene" databases to establish a consensus annotation. The second project, specifically realized for the chicken, is to provide a "gene" database built from various resources including the NCBI's RefSeq and EMBL-EBI's Ensembl databases and other resources such as FAANG multi-tissue resources and NONCODE database. This gene catalog is maintained at each significant update in chicken genome assembly and genome annotation, the last version of June 2022, associated to the GRCg7b assembly, being composed of 23,926 PCGs and 44,428 lncRNA genes (available at http://www.fragencode.org).

Evolution of the Reference Sequence of the Chicken Genome
As illustrated in the Figure 11A, while the overall coverage of Galgal2/WUGSC1.0 was 6.63x [Hillier et al. 2004], this parameter is doubled for the two successive assemblies, *i.e.* Galgal3/WUGSC2.1 [NCBI's RefSeq 2006] and Galgal4 [NCBI's RefSeq 2011], which were released in November 2006 and 2011 respectively. Compared to the first version, these include 33 chromosomes (1-28; 32; W/Z; MT) but the number of scaffolds remained very high (~17,000), with 915 gaps between the scaffolds and a scaffold N50 which was quite low (~12M) showing the incompleteness of the chicken genome sequence. Note that scaffold N50 is defined as the sequence length of the shortest scaffold at 50% of the total genome length. With the release of the Galgal5 assembly in December 2015 and with the improvement of sequencing technologies [NCBI's RefSeq 2015; Warren et al. 2017], the average coverage explodes, and reached a global depth of 70x, leading to a better knowledge of the genome sequence. A few chromosomes were newly defined (addition of chromosomes 29 to 31). However, the quality of this genome sequence remained low with a high number of scaffolds (~24,000) and a lower scaffold N50 (~6.4M) than before. These weak performances are likely due to long-read sequencing, which improved the detection of smaller scaffolds thus decreasing the N50 value [Warren et al. 2017].

In March 2018, a new assembly called GRCg6a was released by the Genome Reference Consortium, which has taken the lead concerning the chicken genome assembly previously managed by the International Chicken Genome Consortium [NCBI's RefSeq 2018]. Tremendous progress — due to the addition of long read sequences, improved *de novo* assembly algorithms, manual annotation of contigs, and integration of finished BAC clone sequences — was made regarding the genome accuracy, as indicated by the drop in the number of scaffolds (from ~24,000 to ~500) with only 68 gaps between the scaffolds and an increase in the scaffold N50 (from ~6,4M to ~20M).

Furthermore, with the latest GRCg7 genome assemblies [NCBI's RefSeq 2021a, 2021b], the knowledge of the chicken genome sequence improved even more in two main ways. First, the accuracy of the genome sequence increased due to improvements in sequencing and especially assembly technologies. The chicken genome is now composed of 42 chromosomes (1-39; W/Z; MT), reaching the number observed in the chicken karyotype. This assembly includes more micro-chromosomes with ~250 scaffolds with no gap between scaffolds, and the scaffold N50 reaching 90M. Second, whereas a Red Jungle Fowl breed (known as RJF #256) was always used in previous assembly versions, a trio of chickens from diverse breeds was used for GRCg7. The new reference sequence was

generated from a female offspring from a cross between a broiler female and a white leghorn laying male. The RJF breed, considered to be the descendant of domestic chickens, was used as a good representation of broiler and layer chicken breeds; however, such a choice has a significant impact on the detection of variants leading to the identification of false positives. Since actual breeds have diverged from the RJF, a variation (*e.g.* SNP) at one position may be detected according to the RJF genome sequence whereas this position was fixed in the population of interest. As an illustration, we have previously shown, using RNA-seq data from the liver of 11 different breeds (≈750 RNA-seq of ≈400 birds) aligned on the Galgal5 assembly, that the SNP number with reliable genotypes was on average 549,634 per population, but this number dropped to 339,539 (–38.2%) with a MAF (Minor Allele Frequency) ≥ 10% [Jehl et al. 2021]. This drop is mainly due to fixed variants in the populations since the number decreased to 438,837 (–20.2%) after only excluding the fixed variations.

Consequently, two genome assemblies were released in January 2021: GRCg7b representing the broiler breed and considered as the new genome reference, and GRCg7w representing the laying breed and considered as an alternative.

Because of the quite frequent change of the genome assembly compared to the time needed to conduct and publish a scientific study, a lot of works are published in outdated versions which can lead to the publication of misleading results and in disagreement with more recent versions (Figure 11B). For example, in 2020, two years after the release of GRCg6a, 79 studies using this genome reference were published against 46, 41, and 5 published with the Galgal5, Galgal4, and Galgal3 versions respectively. Some tools such as Liftoff [Shumate and Salzberg 2021] or LiftOver [Kuhn et al. 2013] can be used to convert coordinates from one version to another. The first tool is based on the alignments of the gene features from one annotation to another whereas the second tool is based on alignments of the best/longest syntenic regions for each region of the genome between assemblies (chain files). However, the use of these tools must be done with caution, especially for remote versions, because of important changes in the genome sequence.

Evolution of the Two Reference Genome Annotations: A Breakthrough in 2015 with the Apparition of the lncRNA Gene Biotype

Genome annotation is not only evolving according to the version of the genome assembly but also to the evolution of annotation bioinformatics pipelines and data resources, mainly composed nowadays of RNA-seq data. In Figure 12A, genome annotations, from 2004 with the Galgal2/WUGSC1.0 assembly through 2022 with the GRC7b assembly, produced by the reference centers, NCBI's RefSeq and EMBL-EBI's Ensembl, have been analyzed. As illustrated, the gene number has increased, particularly due to the apparition in 2015 of lncRNAs with, 5,763 and 4,641 lncRNAs modeled by NCBI's RefSeq (v103) and EMBL-EBI's Ensembl (v94). This increase continues with the last genome annotation v107 (associated with the GRCg7b assembly) provided by the EMBL-EBI's Ensembl with 11,944 lncRNAs compared with 5,504 for v106 (associated with GRCg6a). The number of PCGs remains constant at around 17,000 (See further for more explanation regarding such evolution).

In parallel to the gene number, it is important to make some comments about the transcript models that support these genes. As observed in Figure 12B, the transcript models can still be improved, as illustrated by the numerous changes observed between the two versions v105 and v106 of NCBI's RefSeq [NCBI's RefSeq 2022]. Only 6.4% of 93,980 transcripts identified in the 106 version are identical to that found in version v105. Such results can also be observed between genome annotations of different genome assemblies (not shown), or between genome annotations from the two reference centers, NCBI and EMBL-EBI, as illustrated in the next section.

Differences between the Latest NCBI's RefSeq and EMBL-EBI Ensembl Genome Annotations

For the same genome assembly, the two genome annotation bioinformatics centers, EMBL-EBI and NCBI, do not provide the same annotations, as illustrated in Figure 13. First, EMBL-EBI's Ensembl provides twice the number of lncRNA gene models compared to NCBI's RefSeq (shown in Figure 13A) resulting in a total of 30,108 gene models (associated to 72,689 transcripts) including 17,007 PCGs, 11,944 lncRNAs and 674 miRNAs compared with 25,638 gene models (associated to 85,704 transcripts) including 18,024 PCGs, 5,791 lncRNAs and 799 miRNAs for RefSeq. These differences can be explained by the sample datasets and the annotation pipeline thresholds used specifically by the two bioinformatics centers. For example, NCBI's RefSeq does not consider lncRNAs supported by a single mono-exonic transcript in contrast to EMBL-EBI's Ensembl (with 1,157 lncRNA loci). Second, using the "GffCompare" software [Pertea and Pertea 2020], we observed that most of the transcript models are different between the two genome annotations, as shown in Figure 13B. Among the 72,579 transcripts from

EMBL-EBI's Ensembl considered in the analysis, only 17.8% are strictly equal in the NCBI's RefSeq annotation. More than half (55.9%) are identified as new isoforms of an existing locus and 26.1% (18,922) transcripts are associated with 9,958 new gene loci resulting in more than one-third of the 30,108 gene models from EMBL-EBI's Ensembl not being known in NCBI's RefSeq.

Important differences exist between PCG and lncRNA transcripts. For PCG, most of transcripts from EMBL-EBI's Ensembl (70.8%) are new isoforms of the same gene loci existing in the two databases. These results show that the transcript isoforms are not well described with current RNA-seq resources. Indeed, most of RNA-seq data available in the public database are short read RNA-seq; the long read RNA-seq studies using the new technologies such as ONT or PacBio are still very limited [Kuo et al. 2017; Guan et al. 2022] due to the cost of these technologies and their low sequencing depth. For lncRNA, most of the lncRNA transcripts from EMBL-EBI's Ensembl (77.4%) are considered as new loci compared to NCBI's RefSeq. The main cause of this very low gene overlap between the two genome annotations is the difficulty in capturing and therefore modeling lncRNAs compared to PCGs, due to specific features of lncRNA. First, lncRNAs are characterized by a global low expression; around less than 10% of the total reads of a sample analyzed by common technologies support lncRNA transcripts [Lagarrigue et al. 2021]. Second, they are tissue-, developmental stage- and condition- specific, [Cabili et al. 2011; Derrien et al. 2012; Jehl et al. 2020], conditions which are not covered by the limited number of RNA-seq samples used by the reference genome annotation centers compared to the tens of thousands of short-read RNA-seq generated by the avian scientific community which are available in the public database. Moreover, the transcript models from NCBI's RefSeq are significantly longer than those of EMBL-EBI's Ensembl, as shown in Figure 13C, particularly for lncRNAs (almost twice the length), with nearly two supplementary exons by transcript (*resp.* a median of 5 *vs.* 10 exons/transcript, pval < 10^{-16}) for both PCG and lncRNA models, with median exon sizes which remain similar (~250 bp). Moreover, NCBI's RefSeq provides a higher extreme distribution of transcripts per gene for PCGs compared to EMBL-EBI's Ensembl (*resp.* 5 *vs.* 3 for the third quartile and *resp.* 9 and 5 for the last decile). Note that these numbers are far below what is described in Human (*resp.* 6, 11, 18 transcripts per gene for the median, the third quartile, and the last decile (EMBL-EBI's Ensembl v107 with the GRCh38.13 assembly). This discrepancy can be explained in part by the variety of samples each reference used. EMBL-EBI's Ensembl combines a short-read RNA-seq dataset of 21 tissues from the Roslin Institute (1 stage-condition of a same breed per tissue, 21 samples of individual pool) with a short-read dataset of 7 tissues from the GENE-SWitCH project (3 stages of a same breed, 84 samples) and a long-read dataset of 6 tissues (7 samples) [EMBL EBI's Ensembl 2022]. NCBI's RefSeq integrates data from various projects representing more than 20 tissues, different development stages and breeds for a total of 100 and 89 samples for short-read and long-read RNA-seq respectively, in addition to "Cap Analysis Gene Expression" (CAGE) data including those from the FANTOM project [Lizio et al. 2017] for improving the annotation of transcription start sites [NCBI's RefSeq 2022]. For lncRNAs, the pattern in the distribution of the number of transcripts per gene is inverted between NCBI's RefSeq and EMBL-EBI's Ensembl with respectively 2 *vs.* 3 for the third quartile and 3 and 5 for the last decile. Interestingly, these numbers are of the same order of magnitude in Human (*resp.* 1, 2 and 5 transcripts per gene for the median, the third quartile, and the last decile) highlighting the general difficulty in capturing the transcript models associated with lncRNA genes.

Interest in an Annotation Combining NCBI's RefSeq and EMBL-EBI's Ensembl
In summary, different genome annotations coexist with important differences in transcript models for PCGs and gene models for lncRNAs. Initiatives like the MANE project [Morales et al. 2022] for the human genome aim to synergize the NCBI's RefSeq and EMBL-EBI's Ensembl reference genome annotations to establish a consensus, although, so far, these efforts have focused only on PCGs. Such initiatives have yet to exist for livestock species, especially chicken. So far, most RNA-seq studies have analyzed gene expression and focus only on PCGs, using only one of these two reference annotations. As previously reported, the last two chicken reference genome annotations are quite similar in terms of PCG loci. Indeed, 18,024 and 17,007 PCG loci are respectively annotated for NCBI's RefSeq and EMBL-EBI's Ensembl; 15,711 (87.2%) loci from NCBI's RefSeq are shared with 15,848 (93.8%) loci from EMBL-EBI's Ensembl even if most of the transcript models supporting these PCGs are different. However, these numbers drop for the 5,791 and 11,944 lncRNA loci respectively from NCBI's RefSeq and EMBL-EBI's Ensembl where 2,008 (34.7%) loci from NCBI's RefSeq are shared with only 2,118 (17.7%) loci from EMBL-EBI's Ensembl. Therefore, the use of only one of the two reference annotations enables the investigation of most PCG loci but can bias the study of lncRNA loci. Moreover, even when the expression is quantified at the gene level and not the transcript level, the high difference of transcript models previously reported — even for PCG loci

— can have an impact. Thus, in the context of gene expression studies, results could differ depending on the annotation used [Zhao and Zhang 2015]. Furthermore, the difference between transcript models, especially for PCGs, may have an important impact on variant prediction [McCarthy et al. 2014].

Concerning the recent studies interested in lncRNA gene expression, most of them have not used a reference genome annotation because of the very limited number of lncRNA loci represented in the versions – before the latest EMBL-EBI's Ensembl v107 – and produce a *de novo* annotation from investigators' own samples (for review, [Lagarrigue et al. 2021]). Such an approach is due to the recent democratization of RNA-seq data and the RNA-seq processing, gene modeling, and lncRNA prediction pipelines. However, these genome annotations are specific to one tissue or a set of tissues and characterized by their own gene identifiers, making result comparison difficult from one study to another. As reported in recent reviews [Kosinska-Selbi et al. 2020; Lagarrigue et al. 2021], the number of such publications has been constantly growing since 2015, with most of them focusing on the tissue-specific expression of lncRNAs or their differential expression in a given tissue between breeds or animal groups contrasted for an economically important trait in the species of interest. In most of these studies, a few lncRNAs have been highlighted as associated with the trait or tissue of interest whereas the lncRNA catalogues are not really exploited by the scientific community.

In parallel to these tissue-specific studies, a few multi-tissue studies have been performed in order to provide a more comprehensive annotation of lncRNAs and considering their high tissue-specificity. We can point two studies, that are part of the Functional Annotation of ANimal Genome consortium (FAANG - [Andersson et al. 2015]), which have provided a multispecies lncRNA annotation: the first, [Foissac et al. 2019], used 3 tissues of 4 female and male biological replicates of 4 farm species including the chicken updated to 12 tissues (personal communication); the second one, [Kern et al. 2018], used 8 tissues of 2 female and male biological replicates of chicken, pig and cattle and was recently updated to 19 tissues for the chicken species [Guan et al. 2022]. Nevertheless, for a given species such as chicken, these studies remain limited due to the range of tissues, stage of development, condition that may exist.

In this context, we proposed since 2020 to provide a comprehensive gene catalog for chicken by gathering different resources including EMBL-EBI's Ensembl, NCBI's RefSeq and other multi-tissue databases, catalog that we update at each important changes of chicken genome assembly and annotation. Since the release of the new GRCg7b chicken genome sequence, we have recently updated the gene catalogue of 52,075 genes published in 2020 [Jehl et al. 2020], considering the last NCBI's RefSeq and EMBL-EBI's Ensembl annotations available in June 2022. First, we gathered the two genome annotation references, *i.e.* the v106 of NCBI's RefSeq and the v107 of EMBL-EBI's Ensembl resources. In addition to these two references, we chose to gather the two updated FAANG multi-tissue resources described above [Foissac et al. 2019; Guan et al. 2022], in which lncRNAs have been modeled in parallel with the PCG loci. The NONCODE resource composed only of lncRNA loci has also been used, even if this resource has not been updated since 2014 for the chicken [Zhao et al. 2021]. As a result, the EMBL-EBI's Ensembl and NCBI's RefSeq references grew respectively from 17,007 and 18,024 to 23,926 PCGs and from 11,944 and 5,791 to 44,428 lncRNA genes. This atlas associated to GRCg7b assembly is publicly available at http://www.fragencode.org (Degalez et al, in preparation) as the previous ones published in 2020 and associated to the Galgal5 and GRCg6a genome assemblies. In addition to the gene atlas (i.e. gtf file), a functional annotation of the genes across 40 tissues using different public resources is also provided as well as the lncRNA gene naming according to the official HUGO gene nomenclature committee (HGNC). Briefly, for the lncRNAs with an unknown function (frequent cases), the lncRNA adopts the symbol gene name of the gene harboring it, enriched by a suffix describing its genomic location. For more information on the lncRNA nomenclature, see [Wright 2014] and [Muret et al. 2019]. Online suppl. Material 5

Conclusion

This review provides an overview of the evolution of chicken genome assemblies from 2004 to June 2022 and their genome annotations provided by the two most widely used annotation databases, NCBI's RefSeq and EMBL-EBI's Ensembl. We show a great evolution of the genome assembly through 6 different versions due to various technical and technological advances, the latest GRCg7b offers a genome reference sequence composed of 42 chromosomes (1-39; W/Z; MT), reaching the number observed in the chicken karyotype with more micro-chromosomes than the previous versions and with no gap between the ~250 scaffolds. Moreover, we show that the annotation of the chicken genome is constantly evolving according to the version of the genome assembly, the evolution of bioinformatic annotation pipelines, and the RNA-seq data resources. We can highlight the recent emergence, in 2015, of lncRNA models in genome annotations associated with the Galgal5 genome assembly. Concerning the last GRCg7b genome assembly, the two reference genome annotations are quite

different with 18,024 PCGs and 5,791 lncRNAs reported for NCBI's RefSeq and 17,007 PCGs and 11,944 lncRNAs for EMBL-EBI's Ensembl. The PCG entities mainly differ at the transcript model level whereas lncRNAs differ both at the transcript and gene loci levels. Gene loci display a very low overlap mainly explained by the specific features of lncRNAs (low expression, high tissue-, condition- specificity, …) and the limited number of RNA-seq samples used for generating these catalogs. To facilitate the reconstruction of full-length transcript models and so accurate gene models, annotation centers will benefit in the near future from new technologies such as ONT or PacBio allowing long-read RNA sequencing. However, for properly catching lncRNAs, the low sequencing depths of these long-read technologies compared to short-read RNA-seq require preliminarily capture strategies used to boost the concentration of low-abundance transcripts in cDNA libraries. Such strategies have been applied to 4 human and mouse tissues by the GENCODE consortium [Lagarde et al. 2017]. However, the low sequencing depths and the high cost of these technologies limit for the moment their wider use. The main fuel of the genome annotation databases remains the short-read RNA-seq massively generated by the scientific community. In this context, to increase the completeness of the chicken genome annotation, especially lncRNAs, we highlight the interest to combine the two reference NCBI's RefSeq and EMBL-EBI's Ensembl genome annotation databases and even other databases and present two initiatives. One of them applied to the chicken species and updated at each important change of the reference annotation provides a catalogue of 23,926 PCG and 44,428 lncRNA gene models which includes all the gene loci of the last versions (June 2022) of NCBI's RefSeq and EMBL-EBI's Ensembl. **Conflict of Interest Statement.** The authors have no conflicts of interest to declare.

Chicken Cell Atlas: Science and DeSci
(Prepared by M. Yamagata)

Chickens are not only widely consumed for their eggs and meat but are also used as a model species for biological and medical research [Stern, 2005; Burt, 2007; Haniffa et al., 2021]. To advance our understanding of this species, we need to chart the types and properties of all chicken cells across all organs and tissues, build a reference map of the mature and developing chicken bodies, and provide the resources for studying the biology of this species [Yamagata, 2022]. Similar to the mouse and human cell atlas enterprises in progress (see below), this project will generate single-cell transcriptome data for chickens, characterize each cell type, and provide foundational information integrating molecular, spatial, and temporal modalities. It will facilitate fundamental studies of chickens and other birds, including cell biology, molecular biology, developmental biology, neuroscience, physiology, oncology, virology, behavior, ecology, evolution, and animal husbandry.

The Current State of the Chicken Cell Atlas
Recent advances in single-cell RNA sequencing (scRNA-seq) have had significant effects on the study of complex tissues, leading to the discovery of novel cell types, cell states, and biomarkers [Stuart and Satija, 2019; Luecken and Theis, 2019; Alfieri et al., 2022; Zeng, 2022]. The scRNA-seq technology has opened up a plethora of opportunities to perform novel studies using new and classic model animals, including chickens (*Gallus gallus*) [Yamagata, 2022] and other birds [Chen et al., 2021; Colquitt et al., 2021]. A chicken cell atlas project (aka, Tabula Gallus) has been proposed to create a cell atlas of all tissues in the mature and developing chicken [Yamagata, 2022]. Like several ongoing cell atlas projects (see below), the chicken project will collect scRNA-seq data for chickens, characterize each cell type, and eventually make available information that integrates diverse modalities.

The chicken cell atlas is still in its infancy (Table 2) [Yamagata 2022, Liu et al. 2022]. Nonetheless, a couple of pioneering studies have realized this endeavor, revealing various cell types and their states in the chick limb buds [Feregrino et al., 2019; De Lima et al., 2021], the early primitive streak stage [Vermillion et al., 2018; Guillot et al., 2021], the neural crests [Morrison et al., 2017; Gandhi et al., 2020; Williams et al., 2022], and the neural retina [Hoang et al., 2020; Tegla et al., 2020; Yamagata et al., 2021]. Most studies have used embryonic or juvenile tissues, likely due to accessibility. It is practicable because hatched chicks are highly active and generally considered mature. It takes 21 days, on average, for an egg to hatch once incubation begins. However, embryos are often not consistently developed, thus staged according to the Hamburger and Hamilton series [Hamburger and Hamilton, 1951]. Interestingly, in single-cell analysis, one embryo consists of a series of cell types and their states covering different developmental stages. Therefore, making multiple developmental atlases at close time points is not essential. In contrast, auxiliary atlases must be generated reflecting different factors such as sex [Clinton et al., 2001] and variable genetic background [Núñez-León et al., 2019].

The raw data from and references to scRNA-seq studies are searchable at NCBI's GEO database (https://www.ncbi.nlm.nih.gov/geo/) and several single-cell reference sites (e.g., https://panglaodb.se/papers.html, https://www.nxn.se/single-cell-studies/). To explore the original datasets, some interactive viewers for single-cell data are available (Single Cell Portal, https://singlecell.broadinstitute.org; UCSC Cell Browser, https://cells.ucsc.edu; EMBL-EBI, https://www.ebi.ac.uk/gxa/sc/home). Thus, the first endeavor toward the chick cell atlas project will be to create an armamentarium, incorporate multimodal data, display those datasets and assist users. Tabula Muris, Tabula Sapiens/Human Cell Atlas (HCA) /HuBMAP, Tabula *Drosophilae,* and C elegans atlas are examples of other species [Tabula Muris et al., 2018; HuBMAP Consortium, 2019; Haniffa et al., 2021; Taylor et al., 2021; Lindeboom et al., 2021; Eraslan et al., 2022; Li et al., 2022; Tabula Sapiens et al., 2022]. Species-specific community websites like GEISHA [http://geisha.arizona.edu] and Chickspress [https://geneatlas.arl.arizona.edu/] may provide vital starting points as a resource for chicken as in the cases for other model species. However, in the coming years, emerging crypto-technologies such as peer-to-peer data storage and smart contract protocols could transform data sharing methods (see below).

Next Steps: Multimodal, Spatial, and Temporal Atlases
In addition to scRNA-seq and relevant single-nucleus RNA sequencing, other multimodal single-cell technologies, which simultaneously profile multiple data types in the same cell, represent a frontier for discovering new cell types and characterizing cell states [Stuart and Satija, 2019]. The additional modalities include epigenome, proteome, glycome, metabolome, electrophysiology, morphology, and connectome [Guo et al., 2021, Lee et al., 2021; Saunders et al., 2021; Sun et al., 2021; Mund et al., 2022].
Among those modalities, a series of single-cell sequencing methods for detecting heritable DNA methylation and altered chromatin configurations allow the description of epigenetic changes on a genome-wide scale. In particular, a single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq) is the most commonly used method for studying epigenetic landscapes in single cells [Stuart and Satija, 2019; Armand et al., 2021]. Although scRNA-seq and scATAC-seq are different, both represent the activity of genes. Thus, it is also vital to analyze proteomics and metabolomics to understand actual cell function.
The data gained from scRNA-seq and other dissociated protocols lead to the loss of spatial information. By contrast, spatial biology and spatial transcriptomics include histological, cellular, and subcellular information of transcripts in spatial context and coordinates [Rao et al., 2021; Zhuang, 2021; Palla et al., 2022; Moffitt et al., 2022; Chen et al., 2022; Moses and Pachter, 2022]. The type and function of cells are further designated by cell morphology and cell interactions, including neuronal connectivity. eCHIKIN (electroporation- and CRISPR-mediated Homology-Instructed Knock-IN) is a technique for CRISPR-mediated genome editing in somatic cells to insert GFP or Cre cDNA into genes identified as cell-type specific in scRNA-seq data [Yamagata and Sanes, 2021]. This technique will reveal cell morphology and connectivity and potentially describe the molecular and spatial networks that organize the proteome [Cho et al., 2022]. These post-transcriptional modalities, together with spatial and temporal information, facilitate the resolution of cell types and states and provide more critical information on cell function.

Open Science and DeSci
In order to promote any scientific research, it is a prerequisite to have a supportive community, raise funds, and build facilities. Furthermore, all scientific data should be made openly accessible and maintained permanently. Nonetheless, most of the current "big science" projects have suffered from several drawbacks. For example, not all contributions and data submitted by individuals or institutions are satisfactorily credited. Instead, the management is often exceedingly centralized: only a handful of influential scientists are highly recognized as leaders of successful projects.
Decentralized science (DeSci) is an emerging movement that proposes to build a shared infrastructure for disseminating, assessing, funding, crediting, and storing data and knowledge using blockchain technologies [Hamburg, 2021; https://ethereum.org/en/desci/]. It aims to create an ecosystem where all the researchers are motivated to share their data and receive credit for their effort while allowing everyone open access to research materials using crypto- technologies such as non-fungible tokens (NFTs). More important, scientific organizations can be governed using tokenized incentive structures without powerful leaders by establishing decentralized autonomous organizations (DAOs). DAOs can provide more flexible and agile funds using retroactive or quadratic funding by working together with a consortium of academic, philanthropic, and corporate partners. The scientific data and achievements can be owned and credited using research NFTs (rNFTs) or intellectual property NFTs

(ipNFTs). Peer-to-peer data storage such as Interplanetary File System (IPFS) warrants storing and distributing data enduringly. These cutting-edge crypto-technologies will be able to support a gamut of endeavors in various basic science projects, create an armamentarium, and organize research resources and incentives. Thus, I wish to propose an international collaboration among many researchers by establishing gallusDAO to facilitate this chicken cell atlas project in a decentralized manner.

Conflict of Interest Statement. There are no conflicts of interest.

Standardized Chicken Gene Nomenclature to Support Functional and Comparative Studies
(Prepared by F.M. McCarthy, S. Davey, M.C. Young, K.C. Potter, A. Lanke, P. Barela, M.C. Casono, M. Chiodi, L. Cigan, N. Das, A. Goodell, S.M. Johnson, J.H. Keroack, and D. Webb)

Standardized gene nomenclature facilitates unambiguous communication about genes and allows accurate indexing of associated scientific literature. Genes are assigned a full-length name that succinctly describes what is known about its function and a short gene symbol that is unique, the latter being the nomenclature most often used in scientific communications. In addition, we annotate additional names used in published literature to support literature indexing (referred to as synonyms or aliases). Gene nomenclature efforts were developed for representative species of each of the vertebrate lineages (Tweedie et al. 2020; Bradford et al. 2022; Fortriede et al. 2020; Burt et al. 2009; Kusumi et al. 2011) and more recently these groups coordinate their efforts to ensure that gene nomenclature is consistent across vertebrate species. This consistency expedites comparative studies amongst vertebrate species and enables discoveries about gene evolution in these animals. With support from public resources such as NCBI, standardized nomenclature can be propagated from a representative species such as chick to other species in the same taxonomical class (i.e., birds), further promoting comparative studies and revealing genetic differences between species.

The Chicken Gene Nomenclature Consortium (CGNC) was convened in 2009 to develop and promote standardized gene nomenclature for chicken genes (Burt et al. 2009). The CGNC provides updates following each major annotation update of the chicken genome and has an ongoing manual annotation effort. Nomenclature provided by CGNC is displayed and continually updated NCBI's Gene Database. CGNC cross-references both NCBI Gene IDs (formerly Entrez Gene IDs) and Ensembl Gene IDs. Nomenclature may be assigned automatically via multiple orthology and homology searches (Eyre et al. 2007) or by manually reviewing gene location, synteny and published papers. While we make every effort to assign standardized gene symbols and names that are based upon human nomenclature when there is a clear orthology between genes, there are exceptions to this rule. For example, exceptions are made in instances where the chicken gene names are well established in the literature (e.g., ovalbumin) or when human nomenclature refers to genomic features or physiology not common to birds (e.g., when gene names reference the X or Y chromosomes, the HLA region, or human blood groups). Close collaboration between vertebrate gene nomenclature groups allows curators to come to a consensus and ensure that gene nomenclature for orthologs conserved across a diverse range of vertebrate species can be practically applied to all vertebrate lineages.

The last major update of CGNC was completed in May 2022, and this update coincides with updated gene annotations for the GRCg7 assembly. This genome assembly represents a distinct change from other chicken genome assemblies, as it is based upon modern chicken lines rather than Red Jungle Fowl. Since both broiler and leghorn lines were sequenced, CGNC now provides information about genes from both the maternal broiler assembly and the paternal white leghorn layer. There are currently 22,315 genes automatically assigned nomenclature and 3,602 manually approved genes. While initial manual curation efforts focused on assigning nomenclature for 1:1 orthologs between chicken and human, more recently we are focused on assigning standardized names for genes that are not found in mammals. This latter gene set includes gene family expansions found in chickens or birds, or genes found on other vertebrate lineages but lost or significantly altered in mammals. Several of these projects have only been made possible with assistance from community experts, and some examples of these are described below.

MHC-Y Region Genes
The chicken major histocompatibility complex (MHC) contains two independent regions which segregate independently, MHC-B and MHC-Y (Briles et al. 1993; Miller et al. 1994). The unique nature of MHC regions means that these genes need to be manually reviewed, both to identify the genes and to assign nomenclature in a standardized manner. With assistance from Dr. Miller, we reviewed and provided standardized gene

nomenclature for genes associated with the MHC-Y locus. This includes a set of 107 genes, and encompasses class I and II molecules, c-type lectin molecules, zinc finger proteins and leukocyte receptor cluster members (Goto et al. 2022). Class I genes are denoted with the "MHCY" prefix for gene symbols and the full gene name has the format "major histocompatibility complex Y, class I heavy chain", with individual genes numbered sequentially and pseudogenes indicated with the designation "P" in the symbol and "pseudogene" for the full gene name. Likewise, class II genes are named using the "MHCY2B" gene symbol prefix and standardized gene names starting with "major histocompatibility complex Y, class II beta". The MHC-Y region includes eight leukocyte receptor cluster genes that are most closely related to the human leukocyte receptor cluster member 9 family (LENG9; HGNC:16306). To clearly indicate this similarity while ensuring that these genes are not mistaken for direct homologs, the chicken leukocyte receptor genes are designated as "leukocyte receptor cluster member 9 like 3, MHCY region" and gene symbols have the prefix "LENG9L". Likewise, c-type lectin gene names include the information that these genes are located in the MHC-Y region. We note that 40 of these genes are not annotated on current reference assemblies and may be breed-specific; CGNC has reserved these gene names for future use. This systematic review of MHC-Y genes provides a clear and unambiguous naming system that can be extended to genes found in other chicken lines. We expect to extend this work to provide standard gene nomenclature for genes in the MHC-B regions and the Nucleolus Organizer Region (NOR), although this effort may require assistance for other community experts.

Chicken Immunoglobulin Receptor Genes

Chicken microchromosome 31 contains more than 100 chicken Ig-like receptors, which function as activating (CHIRA), inhibitory (CHIRB), or bifunctional receptors (CHIRAB) (Laun et al. 2006; Nikolaidis et al. 2005). This nomenclature represents a challenge because, although it is well established in scientific literature - including in closely related species (Windau, Viertlboeck, and Göbel 2013), it is desirable to remove references to a specific species to promote transfer across multiple species. To resolve this dichotomy, we propose to retain the "CHIR" symbol for these genes but adjust the gene name to "cluster homolog of immunoglobulin like receptor" and to add appropriate "chicken Ig-like receptor" synonyms used in publications to support searching and indexing. We are currently working with Dr. Brandi Sparling to ensure that these chicken genes are correctly identified as activating (CHIRA), inhibitory (CHIRB), or bifunctional (CHIRAB) receptors and named systematically.

Heat Shock Proteins

Heat shock proteins are an example of functionally related proteins that are expressed under stress conditions (Nakai and Ishikawa 2001). Multiple lines of independent research are actively studying stress responses in chickens (Zhao et al. 2011; Olanrewaju et al. 2010; Sarsour and Persia 2022; Jastrebski, Lamont, and Schmidt 2017), and the GRCg7 assemblies have 16 genes currently named with a temporary "LOC" designator that are assigned as "heat shock transcription factor, X-linked-like". Using gene names and literature searches we identified a more comprehensive group of 55 chicken genes associated with heat shock responses. These genes were manually reviewed, and corrected gene nomenclature assigned where required. This process included assigning new names for the expanded gene set of heat shock transcription factors which had formerly been associated with mammalian X and Y chromosomes (HSFX1-HSFX4 and HSFY1). This project highlights the importance of manual review for gene nomenclature automatically assigned across taxonomic groups and we are currently reviewing additional gene sets with names that reference X and Y chromosomes.

Manually assigning gene nomenclature requires that biocurators have a good working knowledge of how public databases store, integrate and share information, as well as understanding key principles in genetics, molecular and cellular biology, evolutionary biology and physiology. CGNC offers research experiences for undergraduate and high school students who wish to complete a research project. Students are introduced to the concept of standardized gene nomenclature and why it is important. They work collaboratively to review automatically assigned gene nomenclature, including reviewing associated publications and completing their own phylogenetic analyses. Students review each other's work, and all nomenclature is quality checked before data entry to the CGNC database. This allows students the opportunity to see the results of their research reflected in databases which use CGNC information.

The current focus for manual curation efforts is to assign nomenclature for gene families and chicken genes that are currently designated with the "LOC" prefix in the NCBI gene set. Gene families frequently have contractions and expansions between species and homology searches often cannot identify orthologs within a family. Many of these LOC genes are similar to (or '-like') genes found in other vertebrates but have no clear ortholog and are either new members of a gene family or represent lineage specific genes. For example, avian gene families with

notable changes compared to mammals include keratins, histones, ribosomal RNAs, cadherins and olfactory receptors and many of these gene families are included in the current LOC genes. Moreover, this process is complicated by the removal and merging of LOC genes during the process of assembly and annotation updates. However, the manual review of these genes necessarily includes closer inspection of annotation and assembly and can provide useful information about regions of the genome that can be improved.

Distinct hypothalamus and breast muscle transcriptomic response to heat stress under Newcastle disease virus infection

(Prepared by Y. Wang, P. Saelao, C. Kern, B. Zhao, R.A. Gallardo, T. Kelly, J.M. Dekkers, S.J. Lamont, and H. Zhou)

Climate change increases ambient temperatures and the frequency of heat waves. Studies investigating the effects of heat stress on poultry production have observed reduced growth rates and egg production, decreased meat and egg quality, and greater risks to food safety. The objective of the study was to identify genes, gene networks, and signaling pathways associated with heat stress under Newcastle disease virus (NDV) infection in the hypothalamus and the breast muscle of chickens by transcriptome profiling, using two highly genetically distinct inbred chicken lines (Leghorn and Fayoumi). All birds were held in the same environment until 14 days of age. On day 14, half of the birds were exposed to a temperature of 38 °C with 50% relative humidity for 4 hours, and finally 35 °C until the end of the experiment. The remaining birds were kept at 25 °C throughout the experiment. The heat-treated birds were inoculated at 21 days of age with 10^7 EID$_{50}$ NDV La Sota strain to investigate the effects of heat stress under NDV infection. Physiological blood parameters were measured at three stages: acute heat (AH) at 4 hours, chronic heat (CH) at 7 days, and chronic heat combined with NDV infection (CH&NDV) at 10 days post-initiation of heat treatment. Hypothalamus and breast muscle samples were harvested at both AH and CH&NDV times to characterize the transcriptome profiles of these two inbred lines. In general, the hypothalamus had a weak response to the treatments compared to the breast muscle in both genetic lines. For both treatments, fewer differentially expressed genes (DEGs) were identified in the hypothalamus than in the breast muscle. Metabolic and immune-related DEGs were identified in the two tissues in each genetic line. Our results suggest that the two genetic lines responded to treatments similarly in the hypothalamus. In the breast muscle, the Fayoumi birds had earlier and more specific responses to both treatments than Leghorn birds. The Fayoumi line enriched the antiviral signaling pathway with acute heat stress and maintained a better iron balance during the heat stress by regulating the corresponding signaling pathways such as the Phospholipase C signaling and Calcium signaling. A list of driver genes was generated by the weighted correlation network analysis (WGCNA). Profiling the gene expression response in the hypothalamus and the breast muscle provided additional insights to help us understand the host response to heat stress under NDV infection at the transcriptome level. The candidate gene list and pathways significantly enriched by the current study warrant further investigation into the underlying mechanisms of specific responses in heat resilient Fayoumi and susceptible Leghorns.

Global warming has had adverse effects on mammals, birds, and plants (Bellard et al., 2012). Physiological and behavioral disorders resulting from heat stress raise concerns for poultry production, health, and welfare. Decreased feed efficiency, growth rate, and egg production, as well as increased susceptibility to disease, are major challenges associated with heat stress in the poultry industry. Total economic losses due to heat stress amount to approximately $1.69 to $2.36 billion annually for the U.S. livestock production industry, of which $128 to $165 million is lost from the poultry industry (St-Pierre et al., 2003). Currently, in genetic selection and breeding programs, high-performance broilers and layers are superior in productivity; however, they cannot maintain their production in the presence of heat stress compared to non-selected birds (Kumar et al., 2013). In

addition, heat stress suppresses the chicken's immune response, which increases the bird's susceptibility to infectious disease threats (Monson et al., 2018). Therefore, it is essential to better understand the deleterious effects and underlying mechanisms of heat stress on immunity and growth performance.

The hypothalamus in the host plays a critical role as a central regulator of temperature in chickens. It links the nervous system to the endocrine system and primarily mediates thermoregulation, food intake, and stress response (Chen et al., 2015). It functions as a central hub to interact and coordinate downstream metabolism (Bohler et al., 2021). Heat stress activates the hypothalamic-pituitary-adrenal (HPA) axis by initiating an appropriate response in the preoptic area of the hypothalamus (Tzschentke and Basta, 2002), which simulates the secretion of gamma-aminobutyric acid and then inhibits the dorsomedial nucleus (DMN) of the hypothalamus. As a result of this inhibition, circuits to muscle, fat, and the cardiovascular system are deactivated reducing heat production and increasing heat loss (Molinas et al., 2019). Activation of the HPA axis results in the release of both corticotrophin releasing hormone (CRH) and arginine vasopressin (AVP), and secretion of adrenocorticotropic hormone (ACTH) by the pituitary gland, which stimulates the release of glucocorticoids from the adrenal cortex. Circulating glucocorticoids interact with a variety of cells to regulate both metabolism and immune function (Chen et al., 2015). Disruption of the metabolic and immune system function in birds gives rise to lower productivity, such as low breast muscle yield, and an increased risk of disease and mortality.

Genetics plays an important role in the host response to heat in chickens (Wolc et al., 2019). Therefore, genetic selection offers a feasible and sustainable option to improve heat stress resistance and immune response in chickens. Genetic selection of chickens with improved resilience to heat stress requires a better understanding of the genetic contribution and molecular mechanisms of the heat stress response. Previous studies conducted by our group have characterized transcriptome profiles of several chicken tissues under heat stress. Candidate genes differentially expressed in the comparisons of heat-stressed and control groups in chicken lungs, tracheas, Harderian gland, and livers were identified as a first step toward identifying targets for validation to serve as selection markers (Saelao et al., 2018; Wang et al., 2020; Saelao et al., 2021). A series of genome-wide association studies (GWAS) have identified several quantitative trait loci (QTL) associated with heat stress response such as body temperature, growth phenotypes such as body weight and breast muscle yield, and immune-related phenotypic traits such as mortality, viral replication, and antibody levels, highlighting the genetic contribution to heat stress tolerance in different chicken populations (Van Goor et al., 2015; Rowland et al., 2018; Saelao et al., 2019; Wolc et al., 2019). However, underlying mechanisms of genetic resistance and susceptibility to heat stress are still not fully understood. Physiological parameters from two genetically distinct, highly inbred lines (Fayoumi and Leghorn) that were heat stressed under NDV infection, were reported (Wang et al., 2018). The relatively more heat tolerant Fayoumi birds showed distinct responses during acute and chronic heat stress compared to the relatively heat stress susceptible Leghorn birds (Wang et al., 2018). Better hemostasis was maintained in the Fayoumi birds under both abiotic and biotic stressors. Acid-base balance in the host was one of the key factors accounting for the genetic differences between these two lines (Wang et al., 2018). The liver transcriptome profile of these two inbred lines revealed that the relative heat tolerant and disease resistant Fayoumi line activated not only metabolic but also immune regulation with heat stress and viral infection, and the susceptible Leghorn birds were only able to maintain the basic metabolic responses (Wang et al., 2020). Thermoregulation is mediated by both the nervous and endocrine systems with a key regulator of the nervous system being the hypothalamus. A major downstream organ, the breast muscle, deserves intensive study to help better understand different aspects of thermoregulation. Therefore, the current study was designed to survey the transcriptomic profiles of the hypothalamus and the breast muscle of these two inbred lines to characterize the global gene expression response to heat stress in distinct genetic backgrounds. Identifying candidate genes and pathways associated with heat tolerance in a variety of tissues will aid discovery of important molecular markers to target in genetic selection to improve heat resilience in poultry.

Experimental Populations

Two genetically distinct, highly inbred lines, Fayoumi (M 15.2) and Leghorn (GHs 6) having inbreeding coefficient=99.95% (Zhou and Lamont, 1999), from the Iowa State University Poultry Teaching and Research Farm (Ames, IA) were used in the current study. Fifty-five Leghorns and fifty-six Fayoumi birds were housed in two temperature and humidity-controlled isolators, birds were provided with *ad libitum* access to food and water. On day (d) 1 of age, 30 Leghorn and 31 Fayoumi birds were randomly chosen as the treatment groups and housed in one isolator and the rest of the birds were used as the non-treatment groups in another isolator. The two genetic lines were mixed in each isolator. From 14 days of age to the end of the experiment (41 days of age), the heat-

treated groups were exposed to continuous heat stress of 38 °C for the first 4 hours and then decreased to 35 °C, while the non-treatment groups were maintained at 29.4°C for the first week and then 25 °C throughout the whole experiment. On d21, birds in the heat-treatment groups were inoculated with 10^7 EID$_{50}$ (one EID$_{50}$ unit is the amount of virus that will infect 50 percent of inoculated embryos) Newcastle Disease virus (NDV) La Sota strain through both eyes and nares (50 ul per each eye and nostril). The non-treatment birds were given 200 ul phosphate-buffered saline (PBS) via the same routes. The animal experiment was performed according to the guidelines approved by the Institutional Animal Care and Use Committee, University of California, Davis (IACUC #17853).

Blood Parameter Measurements
Physiological blood parameters were measured at three stages: acute heat (AH) at 4 hours, chronic heat (CH) at 7 days, and chronic heat combined with NDV infection (CH&NDV) at 10 days post-initiation of heat treatment. Thirteen parameters including four chemistry/electrolyte parameters (concentrations of sodium (Na^+), potassium (K^+), ionized calcium (Ca^{2+}), and glucose (Glu)); seven blood gas parameters (blood pH, carbon dioxide partial pressure (PCO_2), oxygen partial pressure (PO_2), total carbon dioxide (TCO_2), bicarbonate (HCO_3), BE, and oxygen situation (sO_2)) were measured immediately by using an i-STAT Portable Blood Analyzer as we described in the previous study (Wang et al., 2018).

Tissue Sample Collection and RNA Isolation
   A total of 32 chickens (4 birds per line per treatment at d14 (4-hour post-heat stress treatment, acute phase (AH)) and d23 (9 days post-treatment, chronic phase (CH)) were randomly selected from treatment and non-treatment Leghorn and Fayoumi birds. The birds were euthanized, and the hypothalamus and the breast muscle samples were collected for RNA isolation. Total RNA was isolated using Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. DNase I (Ambion, Austin, TX) digestion was carried out after RNA isolation, and the RNA concentration and purity were determined by measuring absorbance at 260 nm and A260/A280 ratio using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and RNA quality was checked by Agilent Bioanalyzer (Agilent, Santa Clara, CA). The RNA samples were stored at −80°C until further use.

RNA-Seq Library Preparation and Data Analysis
   For each sample, 500-ng of total RNA was used to construct a cDNA library by using the NEBNext®Ultra™ RNA library prep Kit for Illumina® (New England BioLabs, Ipswich, MA). In total, there were 64 RNA-seq libraries. The cDNA libraries were quantified by Qubit (ThermoFisher Scientific, Waltham, MA) and validated by Agilent Bioanalyzer High Sensitivity DNA Assay (Agilent, Santa Clara, CA) and then sequenced on the HiSeq4000 platform (Illumina, San Diego, CA) for 100bp paired end reads (DNA Core Facility, University of California, Davis, Davis, CA). Sequencing data can be accessed at GEO (PRJNA896699).
Raw reads were trimmed to remove adapter sequences and low-quality bases were removed using the Trim Galore program (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The RNA-seq analysis was carried out by using the same method as described in the liver transcriptome study (Wang et al., 2020). The statistical model included the effects of line and condition for each treatment in each tissue, along with the interactions between condition and line. DEGs were declared with a false discovery rate (FDR) < 0.05. Gene expression profiles were compared between each pair of the 8 groups in each tissue: Leghorn non-treated (LC) and Leghorn treated (LT) with AH and CH&NDV; Fayoumi non-treated (FC) and Fayoumi treated with AH and CH&NDV. In each tissue, the comparison groups with the AH and CH&NDV treatment were FTFC (Fayoumi treatment vs. Fayoumi non-treatment), LTLC (Leghorn treatment vs. Leghorn non-treatment), FCLC (Fayoumi non-treatment vs. Leghorn non-treatment) and FTLT (Fayoumi treatment vs. Leghorn treatment).

Gene Ontology
   Statistics related to the overrepresentation of functional categories were performed using DAVID (Huang et al., 2009; Sherman et al., 2022). A fold enrichment >2 and FDR <20% was considered significant.

Pathway Analysis
   Pathway analysis using the DEGs of within-line contrasts was performed by using the Ingenuity Pathway Analysis software (IPA; Qiagen, Redwood City, CA, USA (Kramer et al., 2014). Significant associations (p-value < 0.05) and a Z-score cutoff of |z| > 1.64 were used to identify significantly activated or inhibited pathways.

Gene Co-Expression Network Analysis

The Weighted Gene Co-expression Network Analysis (WGCNA) package in R was used for gene co-expression network analysis (Langfelder and Horvath, 2008; Langfelder and Horvath, 2012). A soft threshold of 13 was set for generating an adjacency matrix based on co-expression and the minimum module size was arbitrarily set at 30. To evaluate associations of co-expressed gene clusters with line and treatment, the Leghorn and Fayoumi lines were given nominal values of 1 and 2 and non-treatment and treatment the nominal values of 0 and 1. For the continuous traits collected from the previous physiological study (Wang et al., 2018), association of co-expressed gene clusters with the continuous traits pH, carbon dioxide partial pressure (PCO_2), bicarbonate (HCO_3), base excess (BE), oxygen partial pressure (PO_2), oxygen situation (sO_2), glucose (Glu), sodium (Na^+), potassium (K^+) and ionized calcium (iCa^{2+}) were also evaluated. The driver genes were identified by high absolute values of gene significance (GS > 0.5) and module membership (MM > 0.5) with a threshold of p-value < 0.05 (Horvath and Dong, 2008).

Summary of RNA-Seq Analysis

Sixty-four chicken cDNA libraries were prepared from hypothalamus and breast muscle samples and sequenced by the Illumina HiSeq 4000, which included four treatment groups: treated and non-treated Leghorn (LT and LC); treated and non-treated Fayoumi (FT and FC), at the acute heat stress (AH) and the chronic heat stress & NDV inoculation (CH&NDV), respectively. Of the 24,357 annotated chicken genes in the chicken Galgal 6.0 database, more than 80% of the annotated genes were identified in both lines. The detailed mapping statistics are listed in Table 3. The two tissues showed distinct transcriptome profiles. Many genes (2,242 at the AH stage and 3,052 at the CH&NDV stage) were specifically identified in the hypothalamus. For the breast muscle, 627 genes at the AH stage and 384 genes at the CH&NDV stage were specifically expressed compared to the hypothalamus, many of which are related to muscle biogenesis (online suppl. Material 6).

Differential Gene Expression in Different Comparisons

Gene expression profiles were compared between each pair of the 4 groups with two treatments in the two tissues, respectively. The comparison groups were the same as previously reported, which were FC vs. LC (FCLC), FT vs. LT (FTLT), LT vs. LC (LTLC) and FT vs. FC (FTFC) at AH and CH&NDV for both hypothalamus and breast muscle (Wang et al., 2020). The number of DEGs is shown in Figure 14 for the between-line comparisons and Figure 15 for the within-line comparisons of the two tissues. No genes were identified by the DEG analysis which demonstrated an interaction effect in each tissue. Detailed gene information and fold changes in each comparison are listed in online supplementary Material 7.

For DEGs between genetic lines, up-regulated genes with higher expression in Fayoumi birds were Fayoumi-favored genes, while down-regulated genes with higher expression in Leghorn birds were Leghorn-favored genes. In the hypothalamus, more Leghorn-favored genes were identified than Fayoumi-favored genes at the AH stage, however, the pattern was switched by having more Fayoumi-favored DEGs than Leghorn's at the CH stage (Fig. 14). Meanwhile, more DEGs were identified in the non-treated comparisons than in the treated comparisons. The numbers of DEGs identified between the treated Leghorn and Fayoumi line dramatically decreased at the CH&NDV stage with 320 Fayoumi-favored genes and 224 Leghorn-favored genes. In the breast muscle, there were always more Fayoumi-favored genes than Leghorn-favored genes across all comparisons (Fig. 14).

There were fewer DEGs within genetic lines than between lines, the DEG numbers were significantly decreased especially in the hypothalamus compared to the between line comparisons (Fig. 15). More DEGs were up-regulated at the AH stage for both lines, as compared to the CH&NDV stage in which more DEGs were down-regulated. There were no DEGs shared by the four within line comparisons in the hypothalamus (Fig. 16). Ten DEGs, including 3 heat shock protein family members (Heat shock protein family B member 9 (*HSPB9*), Heat shock 70kDa protein 2 (*HSPA2*) and Heat shock protein 90 beta family member 1 (*HSP90B1*)), were shared by the Leghorn and Fayoumi birds at the AH stage (Table 4). Four DEGs were identified in both lines with the CH&NDV treatment, three of which were hemoglobin genes (Table 4). The list of DEGs specifically identified in each comparison provided more information about the line-specific gene regulatory response to heat stress under NDV infection. Six heat shock related genes were specifically identified in the LTLCAH contrast, 3 immune related genes (T cell leukemia homeobox 3 (*TLX3*), Interferon alpha inducible protein 6 (*IFI6*) and Hemoglobin subunit epsilon 1 (*HBE1*)) were specifically identified in the LTLCCH group, 2 heat shock related genes (Heat shock protein 30C-like (*HSP30C*) and Heat shock protein 90 alpha family class B member 1 (*HSP90AB1*)) and 3 immune-related genes

(Myxovirus resistance 1, interferon-inducible protein (*MX1*), Radical S-adenosyl methionine domain containing 2 (*RSAD2*) and Interferon induced protein with tetratricopeptide repeats 5 (*IFIT5*)) were only identified in the FTFCAH contrast, and 7 metabolism-associated genes were only identified in the FTFCCH contrast (Table 5). The DEG numbers were higher in the breast muscle than hypothalamus, particularly during the CH&NDV stage, in which there were more upregulated DEGs than down-regulated DEGs for all comparisons (Fig. 15). The heat shock protein family H member 1 (*HSPH1*) gene was identified in all four groups (Fig. 17). Nine DEGs were identified in both lines at the AH stage and 200 DEGs were shared by the two lines at the CH&NDV stage. Of the line-specific DEGs, 53 DEGs for the LTLCAH, 33 DEGs for the FTFCAH, 316 for the LTLCCH&NDV, and 677 DEGs for the FTFCCH&NDV, the majority of DEGs were metabolism-related genes except a few immune-related genes such as Interferon regulatory factor 2 (*IRF2*), Interferon alpha inducible protein 6 (*IFI6*) and Interferon regulatory factor 2 binding protein 2 (*IRFBP2*). Detailed gene information and fold changes for breast muscle line specific DEGs are listed in online supplementary Material 8.

Functional Categories of Differentially Expressed Genes
Gene ontology (GO) was used to evaluate the function of DEGs from different comparisons. Because characterizing the line-specific responses to heat stress under NDV infection were one of our major objectives, the gene function analysis focused on the within line comparisons in the two lines and two tissues. All DEGs in the within line comparisons were performed by functional enrichment analysis through the DAVID program (DAVID Bioinformatics Resources 6.8). The biological process and KEGG pathways were presented as functional clusters. The significant enriched GO terms and KEGG pathways were presented if P < 0.05 and FDR < 20%.
Hypothalamus. In the Leghorn line, even with fewer DEGs for the within line comparisons identified in the hypothalamus, seven GO terms including the Protein folding, and one KEGG pathway (Protein processing in endoplasmic reticulum) were enriched by the up-regulated DEGs in the LTLCAH comparison. Most of them are metabolic associated except one GO term, Peptide antigen assembly with MHC class I protein complex, was immune related (Fig. 18a). More GO terms (20) were significantly enriched by the downregulated DEGs in the Leghorn birds with the AH treatment, which included iron transport and cholesterol and cellular glucose homeostasis functions (Fig. 18b). No GO terms were enriched by the up-regulated DEGs in the Leghorn line with the CH&NDV treatment. Three metabolic functional GO terms were enriched by the downregulated DEGs in the LTLCCH comparison (Fig. 18c).
In the Fayoumi line, 2 metabolic functions were enriched by the up-regulated DEGs with the AH treatment (Fig. 18d), same as the Leghorn line having the Protein folding function enriched. Seven GO terms were enriched by the downregulated DEGs in the FTFCAH comparison, which included Defense response to virus (Fig. 18e). With the CH&NDV treatment, 11 GO terms and 1 KEGG pathway (ABC transporters) were enriched by the up-regulated DEGs, and 2 metabolic functional terms were enriched by the down-regulated DEGs (Fig. 18f, g), and all of them were associated with metabolic and catabolic processes.
Breast muscle. At the AH stage, up-regulated DEGs in Leghorns enriched five GO terms including skeletal muscle cell differentiation and negative regulation of apoptotic signaling pathway (Fig. 19a). Down-regulated Leghorn DEGs enriched one GO term: positive regulation of vasodilation, and one KEGG pathway: Alanine, aspartate, and glutamate metabolism (Fig. 19b). Four GO terms were enriched by the up-regulated DEGs in Fayoumi birds, which included both metabolic-related functions such as responding to cold, and immune-related functions such as positive regulation of T cell activation (Fig. 19c). No GO terms were enriched by the downregulated Fayoumi DEGs.
At the CH&NDV stage, many GO terms and pathways were enriched by DEGs identified in the within-line comparisons due to the larger number of DEGs. Seventeen GO terms and two KEGG pathways were enriched by the up-regulated Leghorn DEG genes (Fig. 19d). The innate immune response is one of the enriched functions. Down-regulated Leghorn DEGs enriched seven GO terms and eight KEGG pathways, all of which are metabolic-associated functions, except the biosynthesis of antibiotics enriched by the down-regulated Leghorn DEGs (Fig. 19e). In the Fayoumi line, both up and down-regulated DEGs are involved in metabolism functions such as skeletal muscle cell differentiation by up-regulated DEGs and fatty acid beta-oxidation using acyl-CoA by the down-regulated DEGs (Fig. 19f, g). Two immune related KEGG pathways were enriched by Fayoumi DEGs at this stage. Adipocytokine signaling pathway was enriched by up-regulated Fayoumi DEGs and biosynthesis of antibiotics was enriched by down-regulated Fayoumi DEGs (Fig. 19g).

Gene Network and Signaling Pathway Analysis

In addition to the GO analysis, the Ingenuity Pathway Analysis was used to identify gene networks and signaling pathways by using significant DEGs. Because of the limited DEG numbers in the hypothalamus, there was only one canonical pathway, unfolded protein response, identified in the Leghorn line with the AH treatment. In the breast muscle, no canonical pathways were identified in the Leghorn and Fayoumi lines at the AH stage. Twenty-five canonical pathways were identified in the two genetic lines with the CH&NDV treatment (Fig. 20). Thirteen out of twenty-five had similar activation and inhibition patterns. Five pathways were only activated in Leghorn birds, and three pathways were only identified in Fayoumi birds. Four pathways, Phospholipase C signaling, Calcium signaling, Androgen signaling and Synaptic long-term potential, were activated in Leghorns and inhibited in the Fayoumi line with the CH&NDV treatment.

Weighted Gene Co-expression Network Analysis (WGCNA)
To overcome the limitation of DEG numbers specifically identified in the hypothalamus, a correlation based WGCNA analysis was conducted. Seven gene modules were identified in total by using all the hypothalamus and breast muscle data (Fig. 21). Four out of seven gene modules, black, turquoise, blue, and yellow modules, showed significant correlations with tissue, genetic lines, the CH&NDV treatment, and phenotypic traits. Four trait correlated modules were further analyzed and the results are described in Table 6. The black module was the only one positively correlated with Tissue, which means genes from the black module had higher expression levels in the hypothalamus. The turquoise module was only negatively correlated with Tissue. Genes in the turquoise module had higher expression levels in the breast muscle. The blue module was positively correlated with Line and Na^+ and negatively correlated with pH, HCO_3, TCO_2, BE, PO_2, sO_2 and glucose levels in chicken blood. Genes in the blue module are highly expressed in Leghorn birds and correlated with higher sodium and lower blood gas levels. The yellow module was negatively correlated with the CH&NDV treatment and the blood PCO_2 level. Furthermore, genes driving the biology within these trait-correlated modules were identified to understand the potential biological processes these two genetic lines underwent during treatments. Driver genes were not discovered in all correlated gene modules. Driver genes were only identified from Tissue correlated black and turquoise modules, the CH&NDV treatment correlated yellow module, and the Line, pH, PO2, sO2 and Glu correlated blue module. The top five driver genes with the highest absolute gene significance (GS) and module membership (MM) were listed in Table 7. All the detailed driver gene information is presented in online supplementary Material 9 and 10.

For example, driver genes, Glutamate ionotropic receptor NMDA type subunit1 (*GRIN1*), Tropomyosin 1 (*TMP1*), Solute carrier family 17 member 6 (*SLC17A6*), and Calcium voltage-gated channel auxiliary subunit gamma 1 (*CACNG1*) were identified in the turquoise module which was negatively correlated with Tissue and had higher expression levels in the breast muscle. Some immune-related genes such as RAD52 motif containing 1 (*RMD1*), Class I histocompatibility antigen, F10 alpha chain-like (*HA1F*), and MHC B-G antigen genes were identified as key driver genes in the blue module and highly expressed in the Leghorn line. The Apoptosis inducing factor, mitochondria associated 2 (AIFM2) gene was identified as a driver gene for both pH and sO_2 levels. Only one driver gene, NADH-ubiquinone oxidoreductase chain 2 (MT-ND2) was identified associated with Glucose levels. There were 245 genes in the blue module, in which 145 genes were identified as driver genes correlated with Line, 11 driver genes with pH levels, 20 driver genes with PO_2, and 40 driver genes with sO_2.

GO terms from the biological process and KEGG pathways were further analyzed to understand the biological functions of driver genes in significantly correlated gene modules. Genes in the black module, which had higher expression levels in the hypothalamus, enriched one GO term: Inter-kinetic nuclear migration. Seventy-seven GO terms and fifteen KEGG pathways were enriched by the top 3000 genes in the turquoise module that are highly expressed in the breast muscle. The top 5 enriched GO terms and KEGG pathways are listed in Figure 22. Biological functions enriched by most of these genes were related to muscle contraction and the glycolytic process in the metabolism function, which was not a surprise. Two GO terms, DNA-templated transcription, and initiation and Nucleosome assembly were enriched by genes in the blue module and highly expressed in the Leghorn line. No GO terms and KEGG pathways were enriched by driver genes correlated with other traits which may be due to the lower number of genes.

The blue module was the most interesting one because it correlated with the genetic line as well as multiple phenotypic traits. A gene network of all genes in this module was generated (Fig. 23). More metabolic genes served as node genes in the network. Metabolic genes and immune related genes were relatively clustered in the module, respectively.

The reported study is a part of the Innovation Lab for Genomics to Improve Poultry (GIP: http://gip.ucdavis.edu), which aims to genetically enhance resistance to heat stress and NDV infection in African poultry. Effects of the combination of both biotic (NDV inoculation) and abiotic (heat stress) stressors on the same two inbred lines (Leghorn: relatively susceptible and Fayoumi: relatively resistant used as experimental lines) have been investigated in previous studies (Deist et al., 2017; Deist et al., 2018; Deist and Lamont, 2018; Saelao et al., 2018; Wang et al., 2018; Zhang et al., 2018; Wang et al., 2020) as part of this program, with the Fayoumi line (originating in Egypt) as a representative of local African type chickens.

As demonstrated by the physiological responses of these two lines with the AH and CH&NDV treatments, the relatively heat tolerant Fayoumi birds were able to maintain electrolyte levels, respiratory alkalosis, and metabolic acidosis. Lower levels of TCO_2, HCO_3, and BE, higher levels of PO_2 and sO_2 in Fayoumi birds demonstrated heat resistance characteristics, while higher levels of iCa^{2+} and glucose and lower levels of sO_2 in Leghorn birds indicated heat stress susceptibility (Wang et al., 2018) . Global transcriptome profile surveys of the host response to heat stress under NDV infection, or NDV infection alone, have been studied not only in organs where viral replication occurs: Harderian gland, Lung, and Trachea (Deist et al., 2017; Deist et al., 2018; Deist and Lamont, 2018; Saelao et al., 2018; Saelao et al., 2021) but also of the liver, as a representative of a highly metabolic organ (Wang et al., 2020). In the liver transcriptome study, Leghorn birds responded to heat stress and NDV infection mostly by regulating metabolism function. However, Fayoumi birds recruited many immune-related genes, which may regulate both metabolism and immune function to respond to heat stress and viral infection. Results from the liver transcriptome analysis provided insights into how Fayoumi birds are heat and viral infection resilience by activating immune functions even during acute heat stress without viral infection. The surveillance of the Fayoumi immune system was much more sensitive and active than the Leghorns, which could be one of the reasons why Fayoumi birds are more resilient to both heat stress and viral infection.

When birds are exposed to heat stress, the neuroendocrine system is the first responder, in which the hypothalamus is one of the key regulators for temperature regulation (Chen et al., 2015). The hypothalamus belongs to the hypothalamic-pituitary-adrenal (HPA) axis, which is one of the two stress axes and can release glucocorticoids to circulate in the peripheral system. In circulation, glucocorticoids interact with a wide variety of cells to regulate both metabolic and immune functions (Nawaz et al., 2021). Importantly in poultry, glucocorticoids promote proteolysis by damaging myofibrils in skeletal muscles through the ubiquitin-proteasome system which has negative effects on muscle metabolism and meat quality (Bell et al., 2016). We observed gene expression regulation differences in the liver of heat resistant or susceptible lines. A more comprehensive understanding of the heat stress response from the upstream sensing by the hypothalamus to the downstream effectors such as the skeletal muscle can be obtained by profiling the transcriptomes of both hypothalamus and breast muscle from the same birds. The current study is the first to investigate transcriptome response in hypothalamus and breast muscle, two important organs related to heat stress, under both acute heat stress and chronic heat stress combined with NDV infection in the two inbred lines, which can further elucidate the specific molecular mechanism for heat stress resilience in poultry.

Relatively Mild Response in the Hypothalamus
In general, the hypothalamus had a mild response under both AH and CH&NDV treatments in the two genetic lines with very limited DEGs. Fewer DEGs (41 in LTLCAH, 35 in FTFCAH, 10 in LTLCCH, and 45 in FTFCCH) was what we expected based on other chicken hypothalamus transcriptome studies (Sun et al., 2015; Tu et al., 2016). The two genetic lines showed minor differences in the hypothalamic gene regulation with the treatments, which suggests gene regulation in the hypothalamus during heat stress may undergo only fine-tuning or that the action of the hypothalamus is in a non-genetic role, with effects seen in other target tissues.

As a thermo-regulator, the hypothalamus regulates metabolic and immunological functions by secreting glucocorticoids through the HPA axis. Target genes would participate in the amino acid, glycerol, lipo-biogenesis, muscle biogenesis, and glucose metabolism in the metabolic category (Goel et al., 2021). Meanwhile, they would also involve in inflammatory expression and homeostasis of T lymphocytes (Honda et al., 2015). Even with small numbers of DEGs, many heat shock protein genes were identified under both treatments and a few immune-related genes were identified in the two lines (Tables 4 and 5). This is consistent with previous studies that heat shock family genes are mainly up regulated during heat stress (Hasan Siddiqui et al., 2020; Shehata et al., 2020). For the genetic line-specific DEGs, immune-related genes were identified in both lines with the AH treatment. MX1, IFIT5, and RSAD2 genes were down-regulated in the Fayoumi line and NR1D1 was down-regulated in the Leghorn line. With the combination of heat stress and NDV infection, two immune-related genes, TLX3 and IFI6,

were up-regulated in Leghorn birds (Table 4). These findings contrast with our earlier liver transcriptome results for Fayoumi birds, which had an earlier immune response to heat stress than Leghorns (Wang et al., 2020). That both Fayoumi and Leghorn birds hypothalamus regulates immune gene expression to deliver an immunological response is further supported by the GO analysis.

GO terms enriched by the up-regulation DEGs in the two lines were still quite similar with acute heat stress, which contributed to protein folding and protein processing functions. Leghorn down-regulated DEGs were involved in more aspects of metabolic functions, such as cholesterol, cellular glucose, circadian temperature homeostasis, glycogen biosynthesis process, and iron transportation, than Fayoumi down-regulated DEGs. Interestingly, downregulated DEGs from both lines enriched immune-related functions such as the negative regulation of NF-kappaB signaling and negative regulation of toll-like receptor 4 signaling in the Leghorn line and defense response to the virus in the Fayoumi line. The NR1D1 gene, downregulated in the Leghorn birds, was a contributor to the two functions. NR1D1 is a ligand-sensitive transcription factor that negatively regulates gene expression in metabolic and inflammatory processes (Pivovarova et al., 2016). Down-regulation of NR1D1 may be one of the regulatory mechanisms used by the Leghorn birds to respond to heat stress. Three immune-related genes, MX1, IFIT5 and RSAD2, a part of biological function "Defense response to virus" were identified in Fayoumi birds when they are exposed to acute heat stress. The immune response triggered in the Fayoumi birds is more specific than in Leghorns.

Breast Muscle Responds to Heat Stress under NDV Infection Differently in the Two Genetic Lines
For differential gene regulation, the breast muscle had more dramatic responses to both treatments, especially with chronic heat under NDV infection. Many heat shock protein family genes were upregulated in breast muscle. In breast muscle, upregulation of heat shock protein genes can result in skeletal muscle remodeling to protect the muscle cell from damage (Abdelnour et al., 2019). Functional GO terms, which were metabolically related and enriched by DEGs at the AH stage, were similar between the two lines, except for one immune-related GO term: Positive regulation of T cell activation in the Fayoumi line. Two upregulated DEGs in Fayoumi with the AH treatment, Thy-1 cell surface antigen (*THY1*) and Heat shock protein family D member 1 (*HSPD1*), contributed the most to this function. Therefore, Fayoumi birds had earlier and stronger immune responses than Leghorn birds, mediated by activating T cells.

More metabolic and immune functions were activated with the CH&NDV treatment in breast muscle of both genetic lines. Thermoregulation from the hypothalamus to downstream organs is modulated by glucocorticoids. Negative regulation of glucocorticoid receptor signaling pathway was enriched by DEGs in the Leghorn line. Two DEGs, Cryptochrome circadian regulators 1 and 2 (CRY1 and 2), play important roles in the activation of this pathway and both of them were upregulated with the CH&NDV infection in Leghorns. We speculate that Leghorn birds may increase the gene expression levels of CRY1 and 2, subsequently reduce glucocorticoid receptor signaling (Lamia et al., 2011). Meanwhile, the innate immune response was enriched by the upregulated Leghorn DEGs with both heat and viral stressors. Five DEGs, Catelicidin-B1-like (*CATHB1*), Tyrosine kinase non receptor 2 (*TKNR2*), Joining chain of multimeric IgA and IgM (*JCHAIN*), Tripartite motif containing 25 (*TRIM25*), and MX1, worked together to activate this innate immune response.

The canonical pathway analysis by IPA provided additional insight about the differential response of these two genetic lines. The phospholipase C signaling was activated in Leghorn birds and inhibited in the Fayoumi line with the CH&NDV treatment (Fig. 20). This pathway belongs to the intracellular and second messenger signaling and is associated with cell signaling, molecular transport, and vitamin and mineral metabolism (Putney and Tomita, 2012). In Leghorn birds, u-regulation of the Protein kinase C genes (*PKCs*) activated many down-stream genes such as Nuclear factor of activated T cell family (*NFAT*), Histon deacetylase 3 (*HDAC*), cAMP responsive element binding protein 1 (*CREB*) and Nuclear factor kappa-B (*NFkB*) complex and then promoted down-stream gene expression. With the inhibition of this pathway, Fayoumi birds decreased the expression levels of above genes and then suppressed gene expression (Fig. 24). Calcium is an important secondary messenger in the phospholipase C signaling pathway and calcium signaling plays a critical role in muscle contraction (Zhu et al., 2019). The calcium signaling pathway was also activated in Leghorns and inhibited in Fayoumi birds with the CH&NDV treatment (Fig. 25). Down-regulation of Fayoumi DEGs may prevent Ca^{2+} transportation and then slow down cell growth and development. On the other hand, without the inhibition of Ca^{2+} transportation, Leghorn birds might be focusing on rapid cell growth, development, and inflammations. This is consistent with the GO term, Potassium ion transport, enriched by the up-regulated Fayoumi DEGs (Fig. 19f). Collectively, regulating

mineral transportation might be one of the key differences between the two genetic lines in responding to chronic heat stress and NDV infection in the breast muscle.

WGCNA Revealed Gene Modules and Driver Genes Important in the Response to Heat Stress
WGCNA has been applied to many transcriptome studies, especially for complex traits such as diseases in farm animals (Kogelman et al., 2014; Deist et al., 2017; Monson et al., 2018; Fan et al., 2021; Farhadian et al., 2021). Genes that share a similar function are clustered in gene modules and genes associated with interesting traits can be identified (Langfelder and Horvath, 2008). In the current study, Tissue, Line, Treatment, and physiological parameters were used to identify potential important driver genes.

Four gene modules showed significant correlation with traits, with the blue module having the most correlated traits (9 traits). Genes highly expressed in the blue module are Leghorn-favored genes. These genes were also correlated with higher levels of Na^+ and lower levels of pH, HCO_3, TCO_2, BE, PO_2, sO_2, and the glucose level in the blood. Heat resilient Fayoumi birds had higher oxygen-related parameters and susceptible Leghorn birds had higher glucose, iron, and lower sO2 levels (Wang et al., 2018). Gene expression patterns in the blue module partially explained these physiological phenotypes with transcriptome data, in which genes highly expressed in Leghorn birds correlated with lower sO_2, and higher Na^+ levels.

Potentially important genes were selected from the top driver genes for each significantly correlated gene module with different traits. TNF receptor superfamily member 8 (TNFRSF8) gene was the top driver gene negatively correlated with the pH levels. As a member of the TNF-receptor superfamily, the TNFRSF8 gene is expressed by active T and B cells and leads to the activation of NFkB (Lee et al., 1996; Morais-Perdigao et al., 2022). Subsequently, the lower pH level could be due to the activation of apoptosis by the TNFRSF8 gene (Wang et al., 2008). The pH level is an important parameter for heat stress treatments, which affects several other blood gas parameters. Further investigation of TNFRSF8's effect on heat tolerance is desired. RDM1, identified by our previous liver transcriptome study, was also identified in this study in these two tissues that are highly expressed in Leghorn birds. It is potentially one of the Leghorn signature genes for response to heat stress. The AIFM2 gene was negatively correlated with both pH and sO_2. This gene can be induced by cold stress and contributes to apoptosis in the presence of bacterial and viral DNA with oxidoreductase and NADH dehydrogenase activities (Nguyen et al., 2020). It correlates with thermoregulation and needs further validation for its multiple effects on both antiviral and metabolic functions. Only one driver gene, the MT-ND2 gene, was correlated with glucose. MT-ND2 is the core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase which can catalyze electron transfer from NADH through the respiratory chain and is essential for the catalytic activity (Rhooms et al., 2020). The variant on this gene was reported to be associated with glucose metabolism in skeletal muscle in rats (Houstek et al., 2012). This gene requires further investigation into the molecular mechanisms involved in glucose metabolism during heat stress in chickens. Most of these driver genes were also the node genes on the gene network generated by the blue module genes (Fig. 23). TNFRSF8 is on the boundary of the immune and metabolic gene clusters. MT-ND2 is distant from the immune gene cluster and close to the Solute carrier family 9 member B2 (SLC9B2) gene which contributes to the regulation of intracellular pH and sodium homeostasis (Anderegg et al., 2022). Gene network analysis here demonstrated the interactions of genes of interest and provided more hypotheses for potential future studies.

To understand the distinct physiological responses during acute heat stress or chronic heat stress combined with NDV infection, transcriptome profiles of two metabolically associated organs, the hypothalamus, and the breast muscle, were surveyed in the current study to elucidate the molecular mechanisms of host responses in relatively heat stress resistant Fayoumi and susceptible Leghorn chicken lines. Both lines responded to heat stress and NDV infection by stimulating metabolic and immune functions. The heat and NDV-resistant Fayoumi line had earlier, more active, and specific immune regulation in the breast muscle than the Leghorn line with both treatments. Genes highly expressed in Leghorns correlated with heat-susceptible physiological phenotypes. Important driver genes, gene modules, and interactive networks identified in the current study provide valuable information for future validation of molecular mechanisms of resistance and for developing novel breeding programs to improve heat and disease resistance in chickens.

Statement of Ethics. The animal experiment was performed according to the guidelines approved by the Institutional Animal Care and Use Committee, University of California, Davis (IACUC #17853).
Conflict of Interest Statement. The authors have no conflicts of interest to declare.

Chicken Repeatome
(Prepared by Y. Bigot and P. Arensburger)

The repeatome gathers all repeated sequences found either in tandem or as interspersed sequences in the genome. Studies of the repeatome in bird genomes have so far been limited, despite its demonstrated importance in other vertebrate genomes. Indeed, the repeatome is a source of regulation and signalling for gene transcription as well as for the networks controlling some epigenetic marks, and for drivers of ectopic recombination events such as deletions, origination of new genes, functioning of telomeres, etc. This lack of interest in the bird repeatome was likely due to the suggestion that the reduced size of bird genomes was the result of depletions of useless sequences (i.e. non-genic repeated sequences) and the absence of activity of those repeats that were able to amplify by transposition [Wicker et al., 2005; Ellengren 2010; Gao et al., 2017]. Since publication of the third report on Chicken Genes and Chromosomes in 2015 [Schmid et al., 2015] new data and analyses have dramatically changed our view of bird genomes, specifically with respect to transposable elements (TEs), GC-rich tandem repeats and their connections to the organization of bird genomes. Unfortunately, repeat annotation of bird genomes continues to be an understudied field. For example, of the four chicken genome models available in 2022 (galGal4-6, Ogye1.0, bGalGal1 GRCg7b and GRCg7w), only galGal4 and 5 has had both interspersed and tandem repeats annotated.

Annotation of Repeats in Avian Genomes
For most bird genomes repeats have been annotated automatically using a repeat database (most often Repbase [Kojima 2020]) and a library-based annotation tool (usually RepeatMasker). The resulting annotations include both full-length and fragmented copies of interspersed repeats. These consist mainly of TEs, tandemly repeated units of simple repeats ($(RY)_n$ and homonucleotidic motifs), low complexity repeats (minisatellite/variable number tandem repeat and microsatellite/short tandem repeat), and telomeric and centromeric macrosatellite DNA (large stretches, from thousand to millions of bps, of repeated units ranging in size from ~10 bps to several hundred bps). Some repeated genes encoding small non-coding RNAs such as tRNAs, 7SLRNAs, and ribosomal RNA are also annotated.
The main limitation of such library-based approaches is that their annotations depend heavily on the quality of the reference database used, including completeness and accuracy of consensus sequences. Furthermore, because these methods are primarily based on sequence similarity, they tend to overestimate the diversity of interspersed repeats by flagging very small non-overlapping hits that often display low sequence complexity. Finally, these methods are inappropriate for annotating satellite DNAs in newly sequenced species because those repeated units are absent from reference databases.
Some studies use signature-based methods instead, focusing on traits that are unique to certain TEs or repeats. For example, to detect retrotransposons with large terminal repeats (LTR) at both ends, programs such as LTR Finder, LocaTR, LtrHarvest, or ReroTector can detect specific DNA organization patterns and signatures (motifs) that are specific to retroviruses [Bolisetty et al., 2012; Mason et al., 2016; Ji and DeWoody 2016]. Tandem repeat finder (TRF) [Benson 1999], another signature method tool, is dedicated to detecting all types of uncomplicated tandem repeats such as simple repeats, microsatellites, minisatellites and satellite DNAs.
The final method for repeat annotation are DNA de novo consensus methods that combine a range of detection tools. The REPET pipeline [Permal et al., 2012] includes the TEdenovo module which uses both de novo (RECON, GROUPER and PILER) and signature-based tools such as TRF (it can also be set to include a library-based step). A second REPET module, TEannot, uses the output of the first module to annotate the genome. Such de novo consensus methods have historically been limited by the need for powerful calculation and storage resources, which has restricted their application to small eukaryotic genomes (~10 Mbp to 500 Mbp). However, advances in computing clusters and a recent REPET update have opened the way for the use of this software package with larger genomes such as those of vertebrates. A second de novo package, the RepeatModeler2 pipeline [Flyn et al.,

2020], has an architecture that is similar, but simpler than that of REPET. It employs two discovery algorithms, RepeatScout and RECON, followed by consensus building and classification steps. In addition, RepeatModeler2 includes two signature-based tools LTRharvest and LTR_retriever. The de novo repeat library produced by RepeatModeller2 is then used to annotate genomes using the RepeatMasker program. The main weakness of these de novo methods is that they are unable to detect repeats with very few copies, such as DNA transposable elements or endogenous viral elements (EVEs).

The method(s) used to estimate the amounts of repeats in a chicken genome model has a strong influence on the final result. Physico-chemical approaches, such as reassociation kinetics, indicate that the chicken genome is composed of ~30% repeats. However, annotations of galGal4 through 6 report repeat content of 8% to 21%, depending on the bioinformatic pipeline used (for review see [Guizard et al., 2016]). Compared to other vertebrate species the TE diversity of chickens is lower (33 TE species [Guizard et al., 2016]) compared to other vertebrate lineages.

Interspersed Repeats

Interspersed repeats are primarily composed of TEs. These are DNA segments with the potential to move and/or duplicate from one chromosomal location to another (i.e. transposition). The most common way to classify TEs is based on whether or not they use an RNA molecule as a transposition intermediate. Class I elements (a.k.a. retrotransposons) reverse transcribe into a DNA molecule in the process of integration, while class II elements, DNA transposons, do not [Kojima 2020]. The advantage of classifying elements this way is that it is relatively simple to understand and widely used. However, a significant failing of this classification method is that it gathers many sequences into subclasses and families that are evolutionarily unrelated (see [Arensburger et al., 2016] for a more in-depth discussion). An alternative classification scheme divides TEs into at least ten classes based on the enzymatic machinery used to transpose between loci [Arensburger et al., 2016]. There are two primary advantages to this alternative classification. First, it is easily amenable to the addition of new classes, as new TE transposition mechanisms are discovered. Second, its organization is compatible with the classification scheme adopted by the International Committee for the Taxonomy of Viruses (ICTV) with which it shares a number of retrotransposon, viruses [Katzourakis and Gifford, 2010] and DNA transposon families [Koonin and Krupovic 2017].

The chicken genome repeatome consists mostly of four kinds of interspersed repeats, making up 15.5% of the genome. Most abundant are non-LTR elements (a.k.a. long interspersed elements or LINEs) belonging to the CR1 group of retrotransposons (>410,000 annotated fragments, 11.8% in the galGal4 and galGal5 model chromosomes). CR1 elements contain two open reading frames (ORFs) coding for the ORF1 protein that binds to CR1 mRNA to assemble a ribonucleic particle, and a reverse transcriptase (RT) protein with an endonuclease domain (fig. 26a). In bird genomes at least 22 CR1 subfamilies have been described [Liu et al., 2009], 8 of which are currently annotated in chicken genomes. These do not show signs of recent mobility, but full-length intact CR1s are present (fewer than 20) [Galbraith et al., 2021], suggesting that these might be mobilized in vitro under the right experimental conditions. In addition to the CR1s seven other LINE group elements have been described [Kojima 2020], only two of which are found in bird genomes, R2 and RTE. Short interspersed elements (SINEs) are non-autonomous non-LTR retrotransposons derived from nuclear RNA (tRNA, rRNA, 7SL RNA, snRNA, etc.) that parasitize the enzymatic machinery of LINEs (fig. 26b). There is some controversy regarding the presence of AmnSINE1 elements in the chicken genome [Nishihara et al., 2006]. This is an element that uses the transposition machinery of specific L1 LINE elements for its own transposition, but L1 elements are absent from all sauropsida except Lepidopsauria. We have failed to find these elements in the chicken genome [Guizard et al., 2016]. Other SINEs, AviRTE [Suh et al., 2016] and TguSINE1 [Suh et al., 2017], that use the transposition machinery of specific CR1 and RTE elements respectively, were found in other bird lineages.

The second most abundant interspersed repeats found in the chicken genome are LTR retrotransposons. These TEs have LTRs directly repeated at both ends of the element and contain ORFs encoding a group-specific antigen (Gag), a reverse transcriptase (RT), and in some cases a retroviral envelope protein (Env) (fig. 26c). From an evolutionary and enzymatic standpoint, their mobility mechanism is strikingly different from that of LINEs and SINEs. At least 21 "species" of LTR-retrotransposons were found in galGal4 and galGal5 [Guizard et al., 2016]. These have either two LTRs or are annotated as solo LTRs, likely resulting from the loss of the inner part of the LTR retrotransposon by recombination between the LTRs of each inserted element. We found no copies corresponding to complete, internally deleted, or partly truncated elements of six models of solo LTRs (Birddawg,putative_LTR_group 4, 9, 12, 22, 28 and 30). Among the 10 elements that could reliably be classified as

LTR retrotransposons, 8 belonged to the endogenous retrovirus (ERV) superfamily (EAV, EAV-HP, ERV2, ERv7, ERv11, Kronos, Soprano, and RetroTux) and 2 were related to the *Gypsy-Ty3* superfamily (retroCalimero, retroSaturnin). No element was found to belong to the *Copia-Ty1* superfamily. A peculiarity of the LTR retrotransposons is that they are dramatically enriched on the chicken W chromosome.

DNA transposons, the third kind of interspersed repeat, covers about 1.8% of chromosome sequences in the chicken genome. These TEs typically contain a single ORF encoding a transposase enzyme that is generally sufficient to catalyze all the steps of its mobility. This ORF is flanked at both ends by terminal inverted repeats that are used as binding sites by the transposase to excise and reinsert the transposon. The diversity of DNA transposons in the chicken genome is extremely low, with only one member of the following superfamilies: *IS630/Tc1/mariner*, *Mariner1_GG*, *hobo/Ac/Tam*, and *Charlie*. Two other DNA transposons are amplified in the chicken genome, but one is an internally deleted derivative of *Mariner1_GG* (*Galluhop*) and one a chimeric element consisting of a *Charlie* copy with a *Galluhop* copy inserted (fig. 26d). This DNA transposon profile is shared by most bird genome model species, even by the woodpecker genomes which display elevated TE abundances (17-30% of the genome) [Manthey et al., 2018]. Sequences related to DNA transposons superfamilies *Ginger1*, *Ginger2*, *hAT*, *IS630/Tc1/mariner*, *P*, *piggyBac*, *Polinton*, *Transib*, *Crypton*, and *Zisupton* are also present as single or as a few repeated copies in various bird genomes. Most of these sequences are domesticated genes, derived from ORFs encoding transposases (for an inventory see additional file 12 in [Guizard et al., 2016]). Finally, there are also 7 interspersed repeat sequence types called Hitchcock, and undetermined_group_1 through 6 that do not appear to correspond to any existing classifications [Kojima 2020; Arensburger et al., 2016]. Their annotation covers about 0.8% of chicken chromosomes. So far, their main outstanding feature is that are enriched in the microchromosomes [Guizard et al., 2016].

Tandem Repeats

Non-coding tandem repeats are mainly composed of DNA minisatellites, microsatellites and macrosatellites and account for approximately 4% of the chicken genome. In the case of DNA mini and microsatellites, the percent coverage of the genome is unlikely to be severely biased by issues related to genome completion. These sequences were the subject of a number of studies in chickens because single-nucleotide polymorphisms (SNPs) were long used as genetic markers. Telomeric repeats, which contain large numbers of macro and microsatellites, are actively maintained by the telomerase enzyme. These telomere ends display very different numbers of repeats, depending on which chromosome was examined, animal age, and whether tissues or cell lines were examined [Taylor and Delany 2000; Rodrigue et al., 2005; O'Hare and Delany 2005].

Unlike the mini and microsatellites above, macrosatellites, which are located in centromeres, telomeres and in non-coding regions of sex chromosomes, are likely underestimated. The most studied macrosatellites are those located in the Z chromosome (Z_rep in Repbase and [Guizard et al., 2016]) and in the W chromosome (for review see [Komissarov et al., 2018]), and between the MHC and rRNA genes repeated in tandem on chromosome 16 [Miller et al., 2014]. Those present in the inner regions juxtaposed to telomeres of autosomes remains unknown. In chickens, and other bird species, these regions are highly GC-rich [Federico et al., 2005], G-quadruplex (G4) rich, and contain genes with flanking regions and introns mainly composed of satellite DNA sequences. There are differences in the estimates of repeat content between those based on genome assemblies, and those based on cytologic and physico-chemicals approaches. Resolving these discrepancies may require elucidating the sequence composition of centromeric and telomeric regions.

Figure 27

In addition to the sequences above, tandemly repeated genes represent between 1.5% and 2% of the chicken genome. They encode large ribosomal RNAs (18S, 5.2S, 28S) [Piégu et al., 2020], as well as account for a number of gene duplications in some bird genomes [Warren et al., 2010]. Unfortunately, when genome models are annotated, duplicated genes with sequences that are more than 95-97% identical are not reported as separate genes. In some cases, such as the PHF7 genes, annotations between genome models vary from 0 to 68 gene copies (likely paralogs plus segmental duplication). In the galGal6 release 105 model, these genes are distributed among 4 loci located on two different chromosomes [Fouchécourt et al., 2022]; while in the white leghorn breed, 39 PHF7-like gene copies are annotated at 10 different loci and distributed among 5 chromosomes. Therefore, much work remains to be done to fix the number of repeated genes present in wild fowls and domesticated lines.

Balance between Genome Size and Repeats

To our knowledge only two publications have examined the issue genome and repeats in birds. The first one [Piégu et al., 2020] showed that the genome of numerous domesticated chicken lines was smaller than that of the red jungle fowl. Both bioinformatics and molecular investigations showed that the genome coverage of various tandem repeat types found in the red jungle fowl genome (rDNA, telomeric repeats, macrosatellite DNA and segmental duplications) were lower in domesticated lines, but that novel segmental duplications were present in these lines. This supports the hypothesis that domestic lines have been significantly reshaped during domestication and subsequently by human-mediated selection.

The second publication [Kapusta et al., 2017], focused on genome size variations on the scale of bird evolution. Prior to this publication, it was thought that bird genomes were small due to selection pressure related to metabolic constraints linked to flight, and to a dearth of active TEs able to expand their numbers. This view was supported by research showing that piwi RNA (piRNA) did not control TE activity in birds and which would not therefore be committed to an arm race with TEs [Lee et al., 2009]. However, Kapusta et al. (2017) showed that TE expansions in birds were counteracted by DNA losses, mainly through large segmental deletions (>10 kbp). This new view, where TEs have remained active in bird genomes, has been strengthened by reports that 1) the chicken genome contains full-length and intact CR1 elements that are putatively active in transposition [Nishihara et al., 2006], 2) it contains recently active LTR retrotransposon [Wang et al., 2013], and 3) that W chromosomes may act refugia for active ERVs [Peona et al., 2021]. This new view was further supported by reports that the chicken genome uses the piRNA system to control CR1 elements and at least some ERVs [Sun et al., 2017; Chang et al., 2018]. Together, these findings support the "accordion" model of genome size evolution [Kapusta et al., 2017] and have changed our understanding of genome plasticity in birds, from a static [Wicker et al., 2005; Ellengren 2010; Gao et al., 2017] to a dynamic view where all forms of recombination are involved. However, it is important to note that the analyses above all used genome assemblies made from somatic DNA. We currently do not know the organization of the germinal genome and therefore are unable to determine if variations in genome organization observed in somatic cells are either due to differences in recombination events occurring in germ cells over generations or to differences in programmed DNA rearrangements or aging during development at each generation, or both.

Future Challenges and Technical Perspectives about Repeats in Avian Genomes
Thanks to the latest versions of Pacific Biosciences and Oxford Nanopore MinION sequencing technologies, the Telomere-to-Telomere (T2T) consortium was able to resolve telomeric and centromeric regions of the human genome, adding another 8% of new sequences (~200 million bp) including 1,956 new genes, 99 of which are predicted to be protein coding [Nurk et al., 2022]. Such a T2T project for avian genomes would be very helpful for resolving highly GC-rich genomes such as those of ratites, but also the GC-rich telomeric regions of macro and micro-chromosomes of numerous bird species. However, one should not be too optimistic about the ability of these new reads to resolve the telomeric and subtelomeric regions of birds. Indeed, studies of Pacific Biosciences sequences of mRNA and genomic DNA have demonstrated that these sequences are high in G4 motifs that may lead to formation of G-quadruplexes [Beauclair et al., 2019]. These structures interfere with sequence library fabrication for both Pacific Biosciences and Oxford Nanopore MinION technologies, as well as with the sequencing of these libraries. HiFi reads from Pacific Biosciences do not solve this issue because G4 motifs block DNA polymerization and no sequences are produced [Zhu et al., 2016].

In theory there are technical solutions to circumvent these problems. Indeed, the formation of G-quadruplexes requires the presence of Na^+ or K^+ cations [Guiblet et al., 2021]. While these cations are normally used during library preparation by reverse transcriptases and by DNA polymerase enzymes, it is possible to substitute the Na^+ or K^+ cations with Li^+ or Cs^+ cations using alkali chloride salts in the buffers [Ramos-Aleman et al., 2018]. However, as Pacific Biosciences and Oxford Nanopore do not make the content of their solutions accessible to the public, users cannot modify them to replace the salts in the buffers [Flores-Juárez et al., 2016].

The elucidation of these enigmatic GC-rich telomeric, subtelomeric and centromeric regions of bird genomes is an exciting challenge. Indeed, these regions contain genes where all non-coding segments are filled with GC-DNA satellites, they are G4 motifs repeated in tandem [Beauclair et al., 2019]. In mammalian genomes G-quadruplexes are detected in vivo [Zheng et al., 2020] and have been shown to be important as transcription factor binding hubs [Bochman et al., 2012; Hall et al., 2017; Spiegel et al., 2021] and as epigenetic modulators of chromatin [Guilbaud et al., 2017]. Birds have significantly expanded these motifs in telomeric, subtelomeric and centromeric regions likely to control the expression. From an evolutionary standpoint, it would be important to determine if this characteristic is specific to birds or evolved early in the Sauropsida and expanded in bird genomes.

Ribosomal DNA Repeats in Chicken and Guinea Fowl Genomes
(Prepared by A. Dyomin, S. Galkina, A. Davidian, and E. Gaginskaya)

The ribosomal DNA (rDNA) is one of the key elements in the cellular genome. It consists of multiple tandemly arranged repeats that bear coding sequences for ribosomal RNAs (rRNAs). rRNA molecules play an essential role in the ribosome functioning. They define the size and shape of ribosomal subunits forming a structural scaffold for the specific placement of proteins inside the ribosome. rRNAs are involved in all events of translation, including mRNA initial binding [Martin et al., 2016], codon recognition by aminoacyl-tRNAs [Ogle et al., 2001; Demeshkina et al., 2012], peptide bond formation [Nissen et al., 2000], and tRNA/mRNA translocation [Mohan and Noller, 2017; Noller et al., 2017; Djumagulov et al., 2021]. These processes are based on conformational rearrangements of the rRNA molecules which constitute two ribosomal subunits. In the cytoplasmic ribosome of a eukaryotic cell, 28S, 5.8S, and 5S rRNA make up the large subunit, while 18S rRNA is the core molecule of the small one. Due to the role of rRNA in the mechanism of protein synthesis in the cell, rDNA is sometimes referred to as a separate subgenome [rDNAome: Symonova, 2019].

Most eukaryotes feature two main types of rDNA loci. The first type includes the 5S rRNA genes. The other type, nucleolus organiser regions (NOR), is specific for encoding the 18S, 5.8S and 28S rRNAs. rDNA repeating units are organised the same way in both types of loci: a transcriptional unit (gene) is followed by a spacer sequence. The 5S rRNA gene (~120 bp) is followed by the so-called non-transcribed spacer (NTS), which can be of varying length. 5S rDNA is transcribed by RNA-pol III, regulator sites being located within the coding sequence [Pieler et al., 1987; Hall, 2005; Paule and White, 2000]. In the eukaryotic NOR, rDNA repeats are organised in a more complex way [Singer and Berg, 1991; Shaw and Brown, 2012; Hori et al., 2021]. Each of them consists of a cluster of 18S, 5.8S, and 28S rRNA genes followed by an intergenic spacer (IGS) sequence. The 18S, 5.8S and 28S rRNAs coding sequences are separated by two internal transcribed spacers (ITS1 and ITS2) and flanked by two external transcribed spacers (5'ETS and 3'ETS). The rDNA cluster is transcribed by RNA-pol I into the primary 45/47S rRNA molecule (pre-rRNA) as a single transcriptional unit. The coding sequences for each rRNA are highly conserved in length and nucleotide composition across taxa. However, rRNA inter- and intra-individual genomic polymorphisms have been described in several species [Pillet et al., 2012; Locati et al., 2017; Kim et al., 2018, 2021; Parks et al., 2018; Ding et al., 2022]. The spacer regions containing splicing cleavage sites evolve more rapidly. Their high variability can be detected even within the same genome [Kim et al., 2018, 2021].

The IGS, which plays important regulatory roles in the cell function, is one of the most interesting regions in the rDNA repeat. In many animals, IGS have been found to contain such functional elements as pre-rRNA promoters [Haltiner et al., 1986; Caudy and Pikaard, 2002; Massin et al., 2005; Agrawal and Ganley, 2018], several transcription termination sites (Sal box) [Pfleiderer et al., 1990; Agrawal and Ganley, 2018], non-coding RNA binding sites involved in cell stress response and regulation of rDNA transcription [Audas et al., 2012; Agrawal and Ganley, 2018], cdc27 pseudogene [Grandori et al., 2005; Agrawal and Ganley, 2018] and putative c-Myc and p53 binding sites Gonzalez et al., 1993; Zentner et al., 2011; Agrawal and Ganley, 2018]. Moreover, the formation of R loops (RNA-DNA duplex) at certain IGS loci prevents RNA-pol I from reading sense ncRNAs, which can disrupt rRNA expression in the human nucleolus [Abraham et al., 2020]. Recently, we have shown a localization of the functional 5S gene within the IGS in turtles and crocodiles that is unique for the vertebrates [Davidian et al., 2022]. This NOR-5S rRNA gene is only active in oocytes and apparently plays a role in producing a maternal pool of extra ribosomes during NOR amplification in oogenesis. All this supports the importance of studying the IGS structure. However, the number of complete ribosomal repeats annotated with their IGS constituents remains extremely low among the available genome assemblies. So far few publications on the identification of structural and functional blocks within IGS in Xenopus [Caudy and Pikaard, 2002], mice [Grozdanov et al., 2003], human and other Apes [Gonzalez and Sylvester, 1995; Agrawal and Ganley, 2018], chicken [Dyomin et al., 2019], and some reptiles [Davidian et al., 2022] exist. This under-investigation of spacer regions is a significant obstacle to understanding their function and evolution.

NOR rDNA Repeat Sequences

Information on the genomic location and cytogenetic features of NOR rRNA genes, the rDNA copy number and repeat size variability in chicken has long been available [Delany and Krupkin, 1999; Schmid et al., 2000, 2005].

However, complete sequence details and organisation of the rDNA repeat unit remained unknown until very recently. The 45S rDNA cluster sequence of *Gallus gallus* was completely assembled in 2016 [Dyomin et al., 2016]. A complete sequence of the rDNA repeat including IGS was annotated in 2019 [Dyomin et al., 2019]. We also introduce new data on the rDNA repeat sequence of the guinea fowl *Numida meleagris* so that we can compare complete rDNA repeats in representatives of two galliform families, Phasianidae and Numididae.

Integrated data on the chicken demonstrate the single NOR location on microchromosome 16 [Bloom and Bacon, 1985; Delany et al., 2009; Solinhac et al., 2010], the rDNA repeat copy number intra- and inter-individual variability to be 150–250 per haploid genome, and average rDNA array sizes in various chicken populations having a range from 5 to 7 Mb [Delany and Krupkin, 1999; Delany, 2000; Schmid et al., 2005]. The latter authors also showed that rDNA repetitive units in the chicken NORs vary from 11 to 50 kb. This difference depends on the IGS size, which is significantly larger in broiler breeds [Delany and Krupkin, 1999; Schmid et al., 2005]. Four chicken rDNA repeats identified to date are of 34,497 bp, 26,999 bp, 27,055 bp and 25,865 bp, the latter detected in the red jungle fowl [Dyomin et al., 2019]. A sample of the complete chicken rRNA gene cluster sequence was first assembled from raw reads and annotated as being 11,863 bp long [Dyomin et al., 2016; NCBI accession number KT445934]. This was later verified by data from sequencing BAC clone WAG137G04 containing three complete rDNA clusters using PacBio RSII [Dyomin et al., 2019]. The results obtained were very similar to the previously assembled sequence, both in sequence homology and cluster size (11,871 bp, 11,830 bp, and 11,855 bp).

The recently released *N. meleagris* genome assembly [Vignal et al., 2019] lacks rDNA data. We identified two complete rDNA clusters and three IGSs of the guinea fowl in NCBI JABXER010000123 contig (online suppl. Material 11, Table S1). The sequence lengths and GC content in the elements that make up the rRNA gene clusters have been determined for both chicken and guinea fowl (Fig. 28; Table 8; online suppl. Material 11, Table S1). Their 18S, 5.8S and 28S rRNA gene sequences are typical of vertebrates but ITS1 and ITS2 sequences are more extended in size, and show higher GC content compared to the majority of other Deuterostomia [Dyomin et al., 2017]. The secondary structure of rDNA cluster sequences should therefore be complicated and infusible while containing multiple hairpins, as was shown for the chicken ITS1 [Dyomin et al., 2016]. The internal transcribed spacers may be the source of species-specific microRNAs, as was shown for human ITS1 (miRNA-663) [Chak et al., 2015] and for mouse ITS2 (miRNA-712) [Son et al., 2013].

Sequence Complexity and GC Content in Intergenic Spacers

When deciphering the guinea fowl NOR rDNA repeats, we also found that the repeat length difference between chicken (~27-34 kb) and guinea fowl (~18-20 kb) is due to the great difference in the IGS lengths (Fig. 28, Table 9). Even within a single NOR, IGS sequences may differ in length (Table 9) and, as a consequence, in their nucleotide composition. The difference in IGS lengths correlates with the difference in the amount of internal repeats (Fig. 29), which may be caused by unequal crossing-over [Erickson and Schmickel, 1985; Smirnov et al., 2016]. It is noteworthy that the red jungle fowl IGS is the shortest, which suggests the possibility of the repeat number increasing in the course of domestication [Dyomin et al., 2019]. We should remember the splendid data by M. Delany and coworkers [Delany and Krupkin, 1999; Delany, 2000; Schmid et al., 2005] of broiler chicken breeds having the longest rDNA repeats (up to 50 kb) due to the longer IGS sequences. Interestingly, despite the strongly differing lengths of IGS, the GC content in IGS remains particularly high in chicken and guinea fowl (Fig. 28; Table 9).

Compared to mammals [Agrawal and Ganley, 2018], the IGS of Galliformes representatives feature a strict hierarchy and order: they contain internal tandem repeats conserved in the unit lengths and are arranged in long blocks and oriented in the same direction. According to Dyomin et al. [2019], the chicken IGS contains three internal repeat blocks: 5' SV-AL block, central EL block, and 3' VAL block (Fig. 29a). The 5' block consists of GC-rich AL repeats (~250 bp) alternating with AT-rich SV repeats (~150 bp). The central block is the longest (9,297 to 14,414 bp) and consists of short GC-enriched tandem EL repeats (~93 bp). The 3' block is separated from the central block by a poly-A motif and consists of VAL repeats (~85 bp). The guinea fowl IGS also contains internal repeat blocks of three types: the conserved SV-AL and VAL blocks and apparently species-specific Nme block (Fig. 29b). Both chicken and guinea fowl IGS have a poly-T motif (~10-49 bp) at the 5' end of the IGS (Fig. 29), which is usually considered the transcription termination motif [Mason et al., 1997]. However, according to the results of transcriptome analysis, the SV-AL repeat block following the poly-T motif is transcribed in all analysed chicken tissues [Dyomin et al., 2019]. Chicken IGS contains one long unique highly conserved sequence (~1,940 bp) between SV-AL and EL repeat blocks and one short unique region (~191 bp) after the 3' VAL repeat block. At the same time, the guinea fowl IGS has at least three unique regions: between Nme and VAL repeat block (~1,350 bp), between VAL blocks (~735 bp) and after the 3' VAL block (169 bp) (Fig. 29; online suppl. Material 11, Data S1).

Each unique region contains a single sequence of interspersed repeat (IR, 93 bp), which is enriched with adenine (Fig. 29). (CT)n repeats (32-36 bp) have also been found in both chicken and guinea fowl unique sequence regions (Fig. 29). It is noteworthy that the size of the third guinea fowl IGS is enlarged due to an insertion into the VAL block between VAL_C and VAL_B units (Fig. 29b, dotted rectangle). This insertion duplicates the IGS fragment containing the unique area and a part of the VAL block. Thus, the longest IGS in the contig contains three VAL repeat blocks and four unique regions, with three of these regions carrying a copy of the IR.

Neither within chicken IGS [Dyomin et al., 2019] nor within the IGS of guinea fowl (this research) were functionally significant sites detected.

As it turned out, the variants of EL and VAL repeats are non-randomly scattered over the corresponding blocks. They are strictly organised into groups and form high-order repeats (HORs) (Fig. 30). This pattern is common to domestic and red jungle chicken in case of EL repeats (Fig. 29a, 30) and to guinea fowl in case of VAL repeats (Fig. 29b, 30). A comparison of different IGS copies demonstrates that internal repeats play the key role in the variability of IGS lengths at the individual and species levels (Fig. 29). In both chicken and guinea fowl, the IGSs are very rich in GC and CpG. This brings them closer to the turtle IGS [Dyomin et al., 2019; Davidian et al., 2022]. However, it is still unclear whether the IGS internal repeats perform any function. The IGS internal repeats and unique sequence regions described herein are not homologous to mammalian and amphibian IGS elements. This may reflect a separate evolutionary pathway for avian rDNA regulatory sequences, which are still to be discovered.

5S rDNA in Galliformes

The previous report on chicken genes [Schmid et al., 2005] has estimated the copy number of 5S rRNA genes in chicken as 35–41 copies with a predominant repeat (5Sα) of 2.2 kb [Daniels and Delany, 2003; Schmid et al., 2005]. Genetic linkage analysis and cytogenetic localization assigned the 5S rDNA to chromosome 9 [Daniels and Delany, 2003]. This is fully consistent with what can be found in the latest chicken genome assembly GRCg7w: one 5S rDNA locus (partly annotated) is situated at ~0,5 Mb on chromosome 9 (NC_052540.1). It comprises 37 full and 1 incomplete tandemly repeated unit made up of a 5S rRNA gene followed by a spacer sequence (NTS) that can be of 2,038-2,268 bp. Such a polymorphism in the length is mainly due to the (G)n microsatellite present in the middle of each NTS. Traces of endogenous retroviruses (ERV-LTR) are also present in the NTSs. Two 5S rRNA gene copies seem to be non-functional; they comprise 117 and 118 bp due to nucleotide deletions. Some copies contain one or two nucleotide substitutions, allowing insights into the intra-individual sequence variability (online suppl. Material 11, Fig. S3a). A variant 5Sß rRNA gene repeat of 0.6 kb [Daniels and Delany, 2003] is not present in the Red Jungle fowl genome, but can be found in many WGS contigs of the Rhode Island, Liyang, Houdan, White Leghorn, Cornish, and Naked Neck chicken genomes. The GRCg7w assembly contains three single 5S rRNA gene copies not followed by the NTS (online suppl. Material 11, Fig. S3b). Two of them at chr2:44,384,176 bp (LOC112531980 at NC_052533.1) and at chr6:9,273,471 bp (NC_052537.1) are slightly degenerated (7 and 8 SNPs correspondingly), whereas the gene situated at chr9:3,007,209 bp is completely identical to the reference 5S rRNA gene sequence (NCBI X01309.1). We could assume the existence of a single functional 5S rDNA site in the chicken genome on chromosome 9. However, this assumption would conflict with the earlier descriptions of two [Krol et al., 1981; Lazar et al., 1983] and even three [Keith et al., 1986] 5S rRNA types from various somatic tissues of the chicken. The alignment of 5S rDNA sequences with the available 5S rRNA sequences of the two types [Lazar et al., 1983] shows that 5S rRNA genes located on chromosome 9 correspond to 5S rRNA type I (NCBI M13920, online suppl. Material 11, Fig. S3b). A variant gene on chromosome 2 is identical to 5S rRNA type II (NCBI M13919, online suppl. Material 11, Fig. S3b). Both types of 5S rRNAs were found in the cytoplasm polysome fractions of chicken liver and brain [Lazar et al., 1983]. The 5S rRNA type II coding sequence is not associated with a common NTS.

In the *N. meleagris* genome assembly no 5S rRNA genes have been annotated so far. The corresponding array consisting of 18 full repeats is present in NCBI WGS contig JABXER010000004. The predicted NTSs are much shorter (937-961 bp) than in chicken. In general, the guinea fowl 5S rDNA repeats are highly homogeneous as only one nucleotide substitution in the genic regions (online suppl. Material 11, Fig. S3b) and a dozen mutations in the NTS were found. Neither guinea fowl NTSs nor chicken repeats show the complex organisation as described above for the IGS, but they both feature a high GC content (~70 %). When aligned, NTSs of both species share no related significant features. Only simple oligonucleotide motifs can be found (*e.g.* CCCGC, GGGTCG, GCGTG, GGAGCAG etc.). The biological meaning of these features is still to be elucidated.

Conclusion

Structural features and variability of chicken rDNA arrays are well studied [Delany and Krupkin, 1999; Schmid et al., 2000, 2005; Daniels and Delany, 2003; Dyomin et al., 2016, 2017, 2019]. This report additionally introduces the first decoding and analysis of the guinea fowl rDNA repeat. The 5S rRNA genes are variable in their structure, copy number and location in the chicken genome but appear to be homogeneous in the guinea fowl. Both studied species, and presumably all representatives of Galliformes, do not have a special oocyte type of 5S rRNA genes, in contrast to fish, amphibians and archelosaur reptiles. This perfectly coincides with the absence of rRNA gene amplification in avian oogenesis.

The complete sequences of the ribosomal repeat from the NORs in chicken and guinea fowl, representatives of two closely related galliform families Phasianidae and Numididae, are compared. On the whole, the rDNA clusters were found to be quite similar in both species. The 18S, 5.8S and 28S rRNA coding sequences are typical of higher eukaryotes. The internal transcribed spacers are longer and more enriched in GC than in other Deuterostomia [Dyomin et al., 2017]. The similarity on high GC and CpG content in the ITS1 and ITS2 sequences of both species may indicate the existence of a general evolutionary mechanism that maintains the same proportions of nucleotides in avian rDNA spacers. Avian IGS seem to differ essentially from the IGS in Mammalia. In both studied species of birds, the IGS sequences separating rRNA gene clusters contain several blocks of internal tandem repeats of the same type. The internal tandem repeats of chicken IGS are highly ordered repeats (HORs) enriched in GC and CpG. This brings the IGS closer to turtles and distinguishes them from fish, amphibians and mammals [Dyomin et al., 2019]. The same is true for the guinea fowl IGS. Some of the internal repeats, such as SV-AL, (CT)n, VAL, were found to be common to the IGS of both species and perhaps to IGS of all Galliformes. At the same time, specific internal IGS repeats exist, such as the Nme repeats for the guinea fowl IGS and the EL repeats for the chicken IGS. The existence of specific repeat types within relatively close taxa of birds, the high degree of degeneracy of SV-AL and VAL, and the low interspecies homology of unique sequences indicate a high rate of evolution of both the entire IGS structure and of its elements. No functionally significant sites associated with RNA-polymerase activity were found in the IGS sequences of the chicken and guinea fowl using conventional transcription databases. In the identification of internal regions of IGS with high functional significance, it could be quite useful to study the IGS structure in representatives of all major taxa of birds. Such a study could reveal the main evolutionarily conserved sequences in the genomes of Galliformes. It could also clarify the mechanisms that determine the structural diversity and rapid evolution of IGS sequences compared to other spacer sequences of the avian rDNA arrays, even at the intraspecific level.

Data Availability Statement. One full rDNA repeat annotation of *N. meleagris* in the JABXER010000123 contig, the alignments of the 5S rRNA sequences found in the chicken genome assembly bGalGal1.mat.broiler.GRCg7b and the guinea fowl JABXER010000004 NCBI WGS contig are available in epy Supplementary material.

Endogenous Avian Leukosis Viruses: The Chicken-Specific Endogenous Retrovirus
(Prepared by A.S. Mason)

As with all birds, the chicken genome is repeat sparse, extending across all repetitive element classes (comprehensively described by Bigot & Arensburger in this report). In mammals, approximately 10% of the genome consists of endogenous retroviruses (ERVs), a subclass of Long Terminal Repeat (LTR) retrotransposons which reflect ancestral retroviral infections and integrations into the germline [Bromham, 2002]. However, and consistent with other Neognathae birds, ERVs only constitute approximately 3% of the chicken genome, though this may change with greater assembly contiguity [Mason et al., 2016; Warren et al., 2017; Kapusta & Suh, 2017]. Chicken ERVs include full-length and degraded examples of spumaviruses, beta- and gamma-retroviruses, but the endogenous alpha-retroviruses are of particular interest as this retroviral group is endemic to birds. Chicken genomes consist of two major groups of endogenous alpha-retroviruses: Endogenous Avian Viruses (EAVs) and Avian Leukosis Virus subgroup E (ALV-E) integrations, previously known as "ev" loci [Payne & Nair, 2012; Sacco & Nair, 2014].

EAVs are ancestral, found broadly across Galliformes with element divergence matching host co-speciation patterns. EAVs are further subdivided into three phylogenetically-distinct clades (EAV-0, EAV-HP and EAV-E51/E33), although inter-clade recombination has been described, resulting in ART-CH elements (avian retrotransposon in chicken; EAV-HP/EAV-51 recombinant) [Gudkov et al., 1992; Sacco & Nair, 2014]. Most chicken EAVs are replication-incompetent, but many are polymorphic with some retaining transcriptional or regulatory potential, associated with phenotypic traits such as blue-green eggshell colour [Wang et al., 2013; Wragg et al., 2013]. Most prominently, recombination between an intact EAV-HP *envelope* and exogenous ALV-A produced the emergent ALV-J, with the recombinant virion enabling altered haemopoietic tropism, inducing myelocytomas rather than the typical B cell lymphoma [Payne et al., 1991; Benson et al., 1998; Sacco et al., 2004]. Even these limited examples highlight the potential impact of ERVs when they retain high structural integrity.

ALV-E integrations are evolutionarily recent and recurrent additions to the chicken genome, as these ERVs are endemic only to *Gallus gallus* and exemplify an evolutionarily-brief period in retroviral life history where exogenous and endogenous forms co-exist [Payne & Nair, 2012; Kanda et al., 2013]. Consequently, ALV-Es are highly polymorphic and are present at low copy number, but typically retain high structural integrity. Replication-competent and transcriptionally active ALV-Es are common, even in commercial flocks, where *gag* expression has been associated with reductions in muscle mass and egg number, size and shell thickness [Crittenden et al., 1984; Fox & Smyth, 1985; Kuhnlein et al., 1989; Gavora et al., 1991]. Furthermore, ALV-E expression has complex impacts on the immunological status of flocks. ALV-Es can be shed and transmitted horizontally through a flock, high antigen titre can lead to persistent viremia and immune exhaustion, and yet expression of endogenous *envelope* can provide protection against exogenous ALV by receptor interference [Robinson et al., 1981; Smith et al., 1990]; a form of ERV-derived immunity (EDI) [Aswad & Katzourakis, 2012; Hurst & Magiorkinis, 2015]. Modulated infection dynamics have also been observed with other viruses [Mays et al., 2019]. Efforts to eradicate ALV-Es in commercial flocks have been hindered by their close association with desirable traits such as recessive white [Chang et al., 2006], henny feathering [Li et al., 2019] and sex-linked slow feathering [Bacon et al., 1988; Elferink et al., 2008].

ALV-E Structure, Retrotransposition and Expression

Intact ALV-E integrations have a typical length of 7,524 bp. This includes LTRs of 274 bp which are identical upon integration, terminally flanked by 6 bp target site duplications. Due to their recent integration, most ALV-E LTRs remain identical, and many ALV-Es retain full structural integrity (*e.g.* ALVE21, ALVE-TYR). Degraded ALV-Es, including terminal (*e.g.* ALVE6, ALVE9) or internal (*e.g.* ALVE3) truncations, and solo LTRs (*e.g.* ALVE15, ALVE_ros005) are common, but less frequent than in more ancient ERV groups [Stoye, 2001].

ALV-E LTRs are shorter than exogenous ALV, limiting endogenous promoter activity and transformation capability [Ruddell, 1995; Benachenhou et al., 2013]. The phenotypic impact of ALV-E integrations is therefore dependent on the specific integration site. Like all alpha-retroviruses and lentiviruses, the presence of a nuclear localisation signal in the retroviral integrase leads to an enrichment of ALV-E integrations within open chromatin encompassing protein-coding genes [Narezkina et al., 2004; Justice & Beemon, 2013]. This distribution holds across all identified ALV-E loci, even after the impact of selection [Mason et al., 2020c].

ALV-Es exhibit the canonical retroviral structure without accessory genes, expressed by host RNA polymerase II from two frame-separated open reading frames: *gag-pol* and *env*. Like other retroviruses, translation of the *gag-pol* transcript is regulated by ribosomal frameshifting. Frameshifting is successful 5-10% of the time, resulting in far lower abundance of *pol* proteins compared to those encoded by *gag*, including alpha-retrovirus-specific inclusion of the *protease* domain [Arad et al., 1995]. *env* translation is inhibited by host miR-155 binding within the surface receptor gp85 domain [Hu et al., 2016], although some integrations, such as ALVE6 [Mason et al., 2020a], have escaped regulation by mutation, resulting in high envelope titres [Robinson et al., 1981].

Novel horizontal transmissions of ALV-Es within flocks, and between cells of the host, is dependent on the cell entry receptor genotype [Hunt et al., 2008]. ALV-E, as well as exogenous ALV subgroups B and D, use tumour virus cell entry receptor B (TVB), encoded by *tumor necrosis factor receptor superfamily member 10b* (*TNFRSF10B*). Wildtype TVB (*TVB*S1*) is susceptible to infection, but alleles which result in either truncation before the transmembrane domain (*e.g.* Q58*, Q100*), or direct or indirect disruption of disulphide bridges in cysteine-rich domains (*e.g.* P61L, C62S, C101R, C125S), have been observed to provide resistance to ALV-E infection in commercial flocks [Adkins et al., 2000, 2001; Klucking & Young, 2004; Reinisová et al., 2008]. As mentioned above, production of ALV-E envelope gp85 protein can also infer resistance by receptor interference, as gp85 and TVB complex together in the Golgi apparatus before presentation on the cell surface [Cosset & Lavillette, 2011].

ALV-E Representation in Reference Genomes

Early work characterising ALV-E diversity focused on White Leghorn lines (typically 1-3 loci) due to the well-described detrimental effects on egg-laying success [Gavora et al., 1991]. Expansion into the more genetically diverse brown-egg layers (typically 5-10 loci) and broilers (genotypes rarely published from commercially-relevant lines) was inhibited by available technologies [Iraqi et al., 1991; Sabour et al., 1992; Grunder et al., 1995; Muir et al., 2008]. Restriction fragment length polymorphisms (RFLPs) became harder to interpret with a greater number of ALV-Es, especially following the discovery that some RFLPs varied between breeds for the same loci [Aarts et al., 1991; Boulliou et al., 1991]. Taken together, these studies showed that whilst some ALV-Es were shared between white- and brown-egg layers and broilers (notably ALVE3 and ALVE6), many loci were novel, and there was no clear indication of an ancestral ALV-E complement.

By the time of the publication of the draft red junglefowl (RJF) reference genome [International Chicken Genome Sequencing Consortium, 2004], almost fifty different ALVE loci had been identified, many with diagnostic and commercially-utilised PCR genotyping assays [Benkel, 1998]. It was therefore surprising that the RJF individual used for the reference contained only two ALVEs: ALVE6 (ALVE-JFevA), widespread but polymorphic in commercial layers and broilers yet not found in other RJF, and the intact ALVE-JFevB, which is, so far, unique to the reference genome individual [International Chicken Genome Sequencing Consortium, 2004; Benkel & Rutherford, 2014; Mason et al., 2020a]. Even the study of just these two ALV-E elements has not been straightforward, as ALVE6 is near the chromosome 1 p-arm telomere, and was only fully assembled in GRCg6a [Mason et al., 2020a]. High-throughput sequencing studies, first with bait-capture enrichment [Rutherford et al., 2016] and more recently utilising unenriched short read whole genome sequencing data [Mason et al., 2020b; Mason et al., 2020c], have now characterised almost 1,300 different ALV-E loci. This better reflects the vast diversity across non-commercial chicken populations, including 'wild-caught' RJF, which often have over 20, individual-specific loci.

Not only does the reference RJF assembly poorly represent the diversity of ALV-Es in commercially-relevant chickens, it also poorly represents RJF ALV-E diversity [Mason, 2021]. This corresponds with the well-described White Leghorn introgression into this RJF individual, as well as the general issues studying polymorphic repetitive elements [Ulfah et al., 2016]. Unfortunately, despite the literature on ALV-E polymorphic diversity, and the clear absence of a conserved, ancestral ALV-E complement, the RJF genome is still used to represent the pre-domesticated state [Hu et al., 2017; Sun et al., 2017; Chen et al., 2019], leading to ambiguous or overreaching results, both with ALV-Es and more broadly.

Pangenomes will more comprehensively document ALV-E diversity. The recently derived Chinese and South-East Asian chicken pangenome consisting of 664 individuals [Wang et al., 2021], many of which were previously analysed for their ALV-E content [Mason et al., 2020b], was certainly a good start, however even these resources can never be 'complete'. Researchers studying ALV-Es need to be sure which specific integrations are present in their study system. As high-throughput sequencing is the only unambiguous method for novel integration detection, the reality of having a highly contiguous but nearly ALV-E-blank genome, such as GRCg6a, is actually quite appealing, as the first step of many detection algorithms is to mask homologous repeats.

Whilst no individual- or pangenome can fully represent ALV-E diversity, the newly derived, haplotype-phased layer (GRCg7w; paternal) and broiler (GRCg7b; maternal) references are good representatives of their breeds, and highly informative for the study of Western commercial stock [Warren et al., this report]. GRCg7w contains six ALVEs which, while more than usual for a White Leghorn, contain the common ALVE1, ALVE3, ALVE15 and ALVE21. GRCg7b has five ALVEs commonly observed in brown-egg layers and broilers, including ALVE-TYR, responsible for the recessive white phenotype of the maternal bird. Both haplotypes contain TVB resistance alleles.

The direct benefit of long read-scaffolded assemblies was the full characterisation of ALV-E integrity in these haplotype references. This remains an outstanding issue with ALV-E identification from short read technologies alone, which can optimally identify integration sites, but are unable to uniquely resolve internal integrity; crucial for predicting expression and retrotransposition potential.

Future Relevance of ALV-Es in Disease

Much of the impetus for studying ALV-Es was based on the detrimental impact of these ERVs on productivity traits [Crittenden et al., 1984; Fox & Smyth, 1985; Kuhnlein et al., 1989; Gavora et al., 1991]. However, direct and indirect selection against these elements, particularly in commercial layers, has largely eradicated these effects. A recent association analysis suggested only linkage disequilibrium in ALV-E/trait associations, even for integrations such as ALVE3 [Fulton et al., 2021]. The full picture is less simple, as ALV-Es fixed in a population cannot be measured in such studies. ALVE-TYR, for example, is structurally intact, common in layers and broilers, and its

impact on growth rate has been appreciated since the 1980s [Fox & Smyth, 1985]. Further work is needed to characterise the phenotypic effects of individual ALV-Es to prioritise their eradication from flocks, particularly given the poorly understood role of ALV-E loci in spontaneous lymphoid leukosis [Cao et al., 2015; Mays et al., 2019].

Outside intensive selection, ALV-E diversity and abundance appears to support a more natural role through ERV-derived immunity [Mason et al., 2020c]. Whilst this may improve the host response to novel ALV infections, the deleterious productivity associations impact the food and economic security of subsistence and small-holder poultry farmers, and will introduce unwanted deleterious loci in the generation of 'localised' commercial birds in areas such as sub-Saharan Africa.

Unregulated flocks also enable the more nebulous possibility of novel, emergent recombinant retroviruses. This has been documented extensively in China with exogenous ALV subgroups A, J and K [Chesters et al., 2001; Liu et al., 2011; Dong et al., 2015; Přikryl et al., 2019], but co-infection could facilitate recombination with other retroviruses, leading to novel tropisms or host expansion.

Recombination Frequency and Distribution along the Chicken Chromosomes
*(*Prepared by P.M. Borodin, L.P. Malinovskaya, and A.A. Torgasheva)

What Is the Recombination Rate and Why Is It Important?
Recombination is essential to orderly chromosome segregation and generation of new allele combinations. The efficiency of artificial selection is critically dependent on the recombination rate; that is, the number of recombination events per whole genome, chromosome and chromosome region. Populations with higher recombination rates demonstrate a higher response to selection [Martin et al., 2006; Dapper, and Payseur, 2017; Gonen et al., 2017]. The distribution of recombination events along chromosomes is another important variable affecting the efficiency of selection. A position of two crossing overs too close to each other does not affect the linkage phase [Gorlov, and Gorlova, 2001; Berchowitz, and Copenhaver, 2010]. Similarly, crossing overs located too close to the centromere of an acrocentric chromosome or to the telomere do not produce new allele combinations. Thus, estimates of genome-wide and chromosome- and region-specific recombination rates in livestock are important for breeding programs.

Recombination is a stochastic but tightly controlled process occurring in the prophase of the first meiotic division[Zickler, and Kleckner, 2015; Gray, and Cohen, 2016]. It starts with chromatin remodeling and the scheduled generation of multiple DNA double-strand breaks (DSBs), followed by a RAD51-mediated search for homologous DNA sequences and formation of heteroduplexes involving DNA strands of homologous chromosomes. Polymerization of the synaptonemal complex, a meiosis-specific proteinaceous structure, stabilizes homologous chromosome synapsis. A small percentage of DSBs are repaired in a crossover manner, while the majority are repaired in a non-crossover manner. Most of the crossover sites are located in recombination hotspots, regions 1-2 kb long, usually flanked by longer cold regions with lower than average recombination frequency [Paul et al., 2016]. The sites of crossing over can be visualized at the pachytene stage as recombination nodules containing MLH1 (mismatch repair protein), and at the diplotene-diakinesis stage as chiasmata. Sister chromatid cohesion beyond the chiasmata holds homologues together at metaphase-I, ensuring proper orientation and orderly segregation. Crossover and non-crossover chromatids segregate at the second meiotic division.

How to Measure the Recombination Rate
There are genetic and cytological methods of assessing the recombination rate. Genetic assessment is based on linkage analysis. It requires large sets of well-controlled crosses or well-characterized pedigree records. This approach provides precise estimates of recombination rate even between closely linked markers. However, it is expensive and time- and labor-consuming. Its efficiency is critically dependent on the number and distribution of the markers. Chiasma count at diplotene and diakinesis provides an unbiased estimate of the genome-wide and chromosome-specific rate of recombination. However, the efficiency of this approach is restricted by difficulties

in obtaining the cells at the particular stages of meiosis and by the accuracy of ascertaining chiasma position. The cytological methods of recombination mapping based on the electron microscopic visualization of the recombination nodules or immunolocalization of MLH1 protein in pachytene spermatocytes and oocytes provide highly reliable estimates of the total recombination rate, as well as the frequency and distribution of recombination events in individual chromosomes[Anderson et al., 1999]. The material for this analysis is easily available in the testis of adult males during the breeding season and in the ovaries of juvenile females during the first week after hatching [Pigozzi, 2016]. Figure 31 shows examples of the application of this method to chicken spermatocytes (a) and oocytes (b).

Genetic and Cytological Estimates of the Chicken Recombination Rate

The first genetic map of the chicken chromosomes was constructed by Serevbrovski and Petrov in 1930 [Serebrovsky, and Petrov, 1930]. It contained 12 markers at four linkage groups and four more unlinked markers. The total length of the map was 252 cM. The most recent high-density consensus linkage map is based on the analysis of the segregation of 9268 SNPs and other markers in three different mapping populations. The total length of the linkage map is 3098 cM for female meiosis and 3145 cM for male meiosis [Groenen et al., 2009]. The first cytological estimate of the recombination rate has been carried out by chiasmata count in the cockerel spermatocytes at the diakinesis-metaphase I stage[Pollock, and Fechheimer, 1978]. There was a significant inter-individual variation: from 56 to 66 chiasmata per spermatocyte. Analysis of the number and distribution of chiasmata along the lampbrush chromosomes in chicken diplotene oocytes gave similar interval of variation (59–64)[Rodionov et al., 1992]. The average numbers of recombination nodules estimated by electron microscopic visualization [Rahn, and Solari, 1986] and immunolocalization of MLH1 protein in pachytene oocytes [Pigozzi, 2001] were rather close to each other (57.5 and 65.0, correspondingly) and to the chiasma count in the spermatocytes. Because each chiasma or recombination nodule represents one crossingover (50 cM), the total length of the sex averaged chicken genetic map is estimated to be 2800-3300 cM. There is a reasonably good correspondence between the genetic and cytological estimates of the recombination rate in chicken. It has, for a long time, been considered the highest among birds. It was suggested that such a high recombination rate could have resulted from domestication and strong artificial selection[Groenen et al., 2009; Backström et al., 2010]. Further studies clarified the intermediate position of the chicken recombination rate between those of the white wagtail (3805 cM) and the black tern (2155 cM) Semenov et al. [2018].

The Distribution of the Recombination Nodules along the Macrochromosomes

Mapping of recombination nodules at the synaptonemal complexes made it possible to visualize the recombination landscapes of chicken macrochromosomes. They showed a highly positive correlation between the length of synaptonemal complex and the number of recombination nodules[Pigozzi, 2001; Rahn, and Solari, 1986], which is typical for vertebrate chromosomes. The macrochromosomes of most birds examined show a highly polarized distribution of the recombination nodules with steep peaks near the telomeres and deep valleys near the centromeres[Pigozzi, and del Priore, 2016]. The recombination landscapes of the chicken macrochromosomes are slightly more flatter than those of other Galloanserae [Calderón, and Pigozzi, 2006; Pigozzi, and del Priore, 2016; del Priore, and Pigozzi, 2015].

The recombination landscape at ZW bivalent is of special interest. In all Neoaves examined, it contains a single crossover located in a very small pseudoautosomal region (PAR). Using the limits of MLH1 foci distribution, Torgasheva et al.[2021] estimated the size of PAR in domestic chicken as 5% of the completely paired ZW bivalent. Interestingly, in one oocyte they detected the second MLH1 foci located at 21% of the ZW length from the telomere, far beyond the most distant MLH1 foci found in other chicken oocytes. Although this exceptional MLH1 focus could be an artifact, it might indicate that rare recombination events are possible in the dispersed regions of residual homology between Z and W detected by Zhou et al.[2014].

Interbreed Variation in the Recombination Rate

The recombination rate shows substantial interbreed and individual variation. Malinovskaya et al. [2019] analyzed the number and distribution of MLH1 foci in spermatocytes of the roosters of six chicken breeds. They revealed significant effects of breed ($R^2 = 0.17$; $p < 0.001$) and individual ($R^2 = 0.28$; $p < 0.001$) on variation in this trait, determined mainly by variation in recombination density on macrochromosomes. There was an interesting correspondence between the age of the breed and its recombination rate. Those with high recombination rates were the breeds created during the last century by crossing several local breeds. The breeds with a low recombination rate were the ancient local breeds.

The linkage experiment also revealed significant inter-population differences in recombination rates [Groenen et al., 2009]. F1 male and female hybrids between Red Jungle fowl and White Leghorn showed significantly lower

recombination rates than purebred broiler populations. The authors suggested that a higher recombination rate in purebred domestic animals was the result of strong artificial selection. Alternatively, a lower recombination rate in the F1 hybrids might be the result of negative heterosis. Malinovskaya et al.[2021] found that F1 hybrids between two purebred breeds had a significantly lower recombination rate (2950 cM), than the cockerels of both parental breeds, Russian Crested and Pervomai (3150 cM and 3350 cM, respectively). The authors explained the negative heterosis for the recombination rate by difficulties in homology matching between the DNA sequences of genetically divergent breeds.

Future Prospects

Controlling recombination frequency and, more importantly, its distribution, is a necessary future step in order to increase the efficiency of selection and overcome its limitations. Identification of the key molecular regulators of crossingover in plants lead to new technologies for increasing recombination rate, which could potentially benefit plant breeding[Blary, and Jenczewski, 2019]. The first group of methods relies on suppressing the genes limiting meiotic recombination, such as topoisomerase TOP3α, DNA-helicases FANCM and RECQ4 and AAA-ATPase FIGL1[Girard et al., 2015; Séguéla-Arnaud et al., 2015; Séguéla-Arnaud et al., 2016; Mieulet et al., 2018]. Combinations of mutations in these genes resulted in up to an 8-fold increase in recombination frequency in *Arabidopsis thaliana* hybrids[Fernandes et al., 2018].

This promising result, however, leaves doubts about the feasibility of developing such methods and their applicability for livestock. Although simulation studies show that an increase in genome-wide recombination would indeed result in an increased response to selection, a significant effect can only be achieved with a 10-20-fold increase in recombination rate[Battagin et al., 2016; Gonen et al., 2017]. Such a huge increase in genome-wide recombination rate, if ever possible, will break up existing beneficial allele combinations and end up in decreased genomic selection accuracy[Battagin et al., 2016]. This might outweigh the benefits of reshuffling genetic material and a potential increase of genetic gain[Blary, and Jenczewski, 2019]. In addition, due to the relatively high stability of recombination hotspots in birds[Singhal et al., 2015], a genome-wide increase in recombination may not lead to the appearance of new crossover sites. Indeed, a 19% interbreed variation in recombination rate did not affect pattern of crossover localization along the chicken macrochromosomes[Malinovskaya et al., 2019].

By manipulating the distribution of crossover events (for example, by stimulating recombination in cold regions), the negative aspects of increased recombination might be avoided. Another approach for controlling crossingover tested in *Saccharomyces cerevisiae* operates by targeting SPO11, a protein which induces DSBs, to specific regions[Peciña et al., 2002][Sarno et al., 2017]. Simulation study predicts that inducing crossingover in non-recombining regions decreases the loss of genetic variability and increases genetic gain, especially in cases when polymorphisms associated with the trait are clustered[Gonen et al., 2017]. Thus, the development of a technology for local modifying recombination patterns, taking into account the genetic architecture of a trait, might have the potential to improve breeding programs in the future. Considering the relatively high stability of recombination hotspots in birds[Singhal et al., 2015], their redistribution could have potential for chicken breeding.

Conclusion

Being an important trait for artificial selection, recombination rate is actively studied in the domestic chicken by various cytological and molecular genetics methods. Recent estimates are consistent and indicate that chicken is characterized by a high level of recombination rate, which is typical for birds. It also shows substantial interbreed variation and a relatively stable pattern of distribution along chromosomes. Methods for controlling the position of crossovers could potentially be useful for chicken breeding in the future.

Conflict of Interest Statement. The authors have no conflicts of interest to declare.

The Broad-Scale Recombination Landscape of Chicken Macrochromosomes Inferred by *in situ* Mapping of Crossovers
(Prepared by M.I. Pigozzi)

Meiotic crossovers (CO) serve to maintain and generate genetic variability in genomes by breaking associations between alleles at linked loci, resulting in new haplotypes. This has implications for the effect of selection on molecular evolution, as non-recombining genomic regions of sexually reproducing organisms accumulate

deleterious mutations and deteriorate (Reeve et al. 2016). Also, the correct number and placement of meiotic CO have a vital role to ensure faithful segregation, with failures often resulting in aneuploidy and infertility. Recombination rates, or the number of recombination events per Mb every generation, may be explained by both differences in CO positions within a genome (recombination landscape) and genome-wide recombination (how many CO events occur per meiosis). A direct method to record both features in individual meiosis is the immunodetection of the MLH1 protein, a component of mature recombination nodules in pachytene (Figure 32a). The length of the recombination map in centimorgans (cM) can be obtained multiplying the average number of MLH1 foci per cell by 50 map units, as one recombination event is equivalent to 50 cM. MLH1 focus data from chicken oocytes, established that the average female genetic map is between 3150 and 3250 cM, in white layer and broiler lines, respectively (Pigozzi 2001; del Priore and Pigozzi 2020). The use of the same methodology in spermatocytes from 5 different breeds established similar genetic lengths in males with interbreed variations of up to 19% in overall genomic recombination rates (Malinovskaya et al. 2019). These independent studies also demonstrated a close match to the existing chicken linkage maps (Groenen et al. 2009).

From the MLH1 count in a number of domestic and wild bird species, global recombination rates and recombination patterns for macrochromosomes have been determined, offering a source for comparative analysis (Semenov et al. 2018; Pigozzi 2022). The usage of meiotic chromosomes at pachytene is particularly favorable to investigate the distribution of crossovers along chromosome arms, because at this stage of meiosis genomic distances are proportional to physical distances. This relationship is based on the relatively uniform size of the DNA loops attached to the synaptonemal complexes of pachytene bivalents (Veller et al. 2019). The theory of the regular spacing of DNA loops along the meiotic chromosomal axis can also be extended to the chicken where SC/DNA ratios exhibit little fluctuation (Pigozzi 2007; del Priore and Pigozzi 2021). Thus, physical distances measured in micrometers or percentages along pachytene bivalents can be converted into genomic positions (in base pairs) along chromosomes. The resolution of the method has the limitation of the fluorescence microscopy. Figure 32b,c shows that the FISH signal of a BAC located at 1.2 Mb from the sequence start of the assembly of GGA1 can be clearly separated from the end of the synaptonemal complex, indicating that accurate measurements could be made for sequences located in pachytene oocyte spreads that were separated by 2 Mb (Pigozzi 2001; del Priore and Pigozzi 2021).

Taking this resolution into account, the recombination rates along the 8 largest autosomal bivalents were calculated here for 2.5 Mb intervals based on MLH1 focus data in 138 chicken oocytes from previous work of our laboratory (del Priore and Pigozzi 2020). The fluctuation of the recombination rates along the macrobivalents is shown in the form of a heatmap where the colour of the cells represents the range of recombination rates (Figure 33). The mean recombination rate in these intervals was 2 cM/Mb, with 54% of the intervals showing rates between 1.5 and 2.5 cM/Mb. Even though crossovers can be found anywhere along chromosome arms, recombination rates are higher near chromosome ends and lower in the center. Regions without foci were rare (5 in 283 intervals) and they were limited to the 0.5-μm intervals closest to the centromeres of metacentric/submetacentric chromosomes.

An MLH1 recombination map can be used when only the genetic (cM) positions of markers are known to predict their physical positions on a chromosome (Anderson et al. 2004). One of the initial scans for potential quantitative trait loci (QTL) for production-related traits in the chicken identified a region on GGA1 spanning the area of 263 and 285 cM (Hansen et al. 2005). The cumulative cM frequency along GGA1 calculated from the MLH1 recombination map predicts that this region comprises about 20 Mb between 145 and 165 Mb from the end of the short arm (Figure 34a). This estimate could have been done even with the approximate size of GGA1 obtained by image cytometry (Smith and Burt 1998; Mendonça et al. 2016). Although this particular prediction is now outdated, it is an example of how cM maps based on MLH1 foci can help integrate physical and genetic maps at their initial stages. Identification of genomic loci governing complex traits involves time-consuming and expensive procedures such as QTL mapping or genome-wide association studies (GWAS). Linkage disequilibrium (LD) between quantitative trait loci (QTL) or markers plays a central role in gene localization (Pritchard and Przeworski 2001). As recombination is known to shuffle genetic material, leading to decay of LD, prior knowledge of the broad recombination landscape in regions of interest could be valuable before embarking on more complex genetic analyses. The MLH1 recombination maps can be overlaid onto assembled chromosomes to obtain cM distances between markers of interest. For instance, a region of GGA5 contains the majority of QTL for woody breast and white striping myopathies which are located in an 8 Mb-long area on GGA5, between 8 and 16.5 Mb of the chromosome sequence (Lake et al. 2021). The MLH1-cM map shows that this segment is located in an area

with recombination rates higher than the average for that chromosome and is bordered by regions with low recombination rates (Figure 34b).

The recombination rates obtained by MLH1 mapping have been applied directly or indirectly to analyze macroevolutionary processes, the variation in crossing over between populations or species, the existence of sex-specific recombination landscapes or the evolution of genome wide recombination rates (Segura et al. 2013; Semenov et al. 2018; Guo et al. 2020; Peterson and Payseur 2020). The wide range of biological issues analyzed by the immunocytological location of crossovers highlights the importance of this approach to studying meiotic recombination as well as its versatility to comparing data from independent studies.

The BF-BL and BG Regions of the Chicken B Locus Differ in Genome Dynamics
(Prepared by J. Kaufman)

Genomes do not evolve homogeneously, and different regions may evince divergent properties. A classic example is the presence of isochores that separate by Giemsa staining and by density centrifugation. In mammals, GC-rich isochores are considered to replicate earlier in the cell cycle than AT-rich isochores, and also have a greater density of genes which are more compact (Bernardi, 2021; Constantini and Musto 2017). Analyses of the chicken B locus illustrate regions that differ in another way: the BF-BL region (which corresponds to the major histocompatibility complex, MHC) is relatively compact and stable with little obvious recombination, whereas the BG region undergoes frequent expansion and contraction leading to copy number variation (CNV). These two regions are organised differently, with the structure of the BF-BL region differing significantly from the MHC of mammals and having a profound effect on function, while the phenotypic effects of recombination and deletion on the BG genes are at present still a mystery (Kaufman et al., 1999a; Salomonsen et al., 2014).

The long history of the B locus has been extensively reviewed (with many more citations: Afrache et al., 2020; Kaufman, 2021; Miller and Taylor, 2016). It was discovered as a polymorphic alloantigen system on chicken red blood cells, then found to co-segregate with a variety of functions associated with the MHC in mammals, and later found to co-segregate with the nuclear organiser region (NOR) on a microchromosome, now numbered chromosome 16. Two regions became apparent by immunoprecipitation of the antigens recognised by the alloantisera: the BG region encoding large erythrocyte membrane proteins without obvious glycans, and the BF-BL region encoding membrane glycoproteins corresponding to MHC classical class I molecules (BF antigens with wide tissue distribution) and class II molecules (lymphocyte BL antigens). Later, it became clear that both BG and BL molecules are specifically expressed on various other cell types. Different B locus haplotypes were found to determine decisive resistance and susceptibility to various economically-important pathogens, later narrowed down to the BF-BL region (Briles et al., 1983; reviewed in Kaufman, 2021; Miller and Taylor, 2016).

Figure 35

More recently, a variety of approaches have shown that the BF-BL and BG regions flank a region primarily encoding tripartite motif-containing (TRIM) genes, and are separated by high levels of recombination from non-classical MHC-like genes in the Rfp-Y region, the ribosomal RNA genes of the NOR, olfactory receptors and scavenger receptors (Delany et al., 2009; Miller et al., 2014; Ruby et al., 2005; Shiina et al., 2007). Moreover, the BF-BL region was discovered to extend towards the telomere with a few genes found in the MHC class III region and some non-polymorphic non-classical class I genes encoding CD1 molecules, which in mammals are found on other chromosomes in so-called MHC-paralogous regions (Maruoka et al., 2005; Miller et al., 2005; Salomonsen et al., 2005; reviewed in Kaufman 2021). The non-classical class I-like YF genes are apparently polymorphic and likely bind hydrophobic ligands, but they are not found in mammals nor do they determine rapid graft rejection or other classical MHC functions (Goto et al., 2022; Thoraval et al., 2003).

An early description of the BF-BL region was a "minimal essential MHC" (Kaufman et al., 1999a, b), based on the fact that most genes expected from the MHC of typical mammals are missing (including genes encoding the complement components C2 and factor B, and the class II A chain BLA), with those few genes present being critical to classical MHC function, including the genes encoding the TAP1 and TAP2 chains forming the peptide transporter, the class I-dedicated chaperone tapasin and the class II-dedicated chaperones formed by DMA, DMB1 and DMB2. Also present is the *BRD2* (or *RING3*) gene encoding a serine-threonine kinase, somewhat mysteriously present in the MHC of all jawed vertebrates examined down to cartilaginous fish. However, the organisation of the BF-BL region is significantly different from the MHC of typical mammals, particularly with the class III region on the outside of the class I and class II regions, and with the two class I (*BF*) genes flanking the *TAP* genes. Moreover, at one end of the BF-BL region there is an unexpected pair of genes encoding lectin-like membrane glycoproteins, likely a natural killer (NK) receptor and ligand which in mammals are found on a different chromosome in the natural killer complex (NKC) (Rogers et al., 2005; Rogers and Kaufman, 2008), and a single gene resembling the many genes in the BG region, now called *BG1* (Goto et al., 2009; Salomonsen et al., 2014).

Compared to mammalian MHCs, the BF-BL region is notably compact and overall there has been little evidence for ongoing recombination, so that the BF-BL region exists as relatively stable haplotypes with potential co-evolution between the closely-linked polymorphic genes. Unlike typical mammals, the peptide-loading genes (*TAP1*, *TAP2*, *tapasin*, *DMA*, *DMB1* and *DMB2*) are polymorphic and, as an example, the peptide-translocation of the TAP protein allele from a particular BF-BL haplotype correlates with the peptide-binding specificity of the dominantly-expressed class I molecule encoded by the *BF2* gene. A few examples of apparent recombination within the BF-BL region have been described, including one between B15 and B12 haplotypes giving rise to the famous B19 haplotype. In fact, out of 242 haplotypes recently identified by sequence (PCR-NGS) typing of commercial, fancy and African chickens (Tregaskes et al., this volume), 53% could be related by potential recombination between the *BF* and *BLB* genes, suggesting that even low levels of recombination could lead to new combinations given sufficient time. Thus far, deletions have only identified due to short direct repeats in the *BF1* gene, leading to disabling of the promoter or to loss of the gene locus.

The stability of the BF-BL region may also depend on the presence of many pairs of sequence-related genes in opposite transcriptional orientation: *BNK* and *Blec*, *BLB1* and *BLB2*, *BF1* and *BF2*, and *TAP1* and *TAP2* (Kaufman et al., 1999b). Homologous recombination between the members of such pairs should lead to inversion rather than deletion or unequal crossing-over, thus preserving the essential genes in a putative "minimal essential MHC". The presence of BLB1-like sequences in the *BLB2* locus adjacent to the *BRD2* gene might be a result of such inversions, although gene conversion or some other form of segmental exchange might also be responsible (Afrache et al., 2020). The orientations of TAP and tapasin genes in closely-related galliform birds are consistent with historic inversions (He et al., 2021)

In contrast to the several different kinds of genes in the BF region that exist in relatively stable haplotypes, the BG region is relatively homogeneous but very dynamic. Almost all the genes in the BG region have similar structures, with exons encoding a signal peptide, a single extracellular immunoglobulin-like (Ig) domain, a hydrophobic transmembrane (TM) region and a long cytoplasmic tail composed of heptad repeats (Miller et al., 1991; Salomonsen et al., 2014). BG molecules are dimers, so the two chains interact through the Ig and TM regions, and the cytoplasmic tails wind around each other in a coiled-coil. All these genes are in the same transcriptional orientation, which is well-suited for unequal crossing-over and deletion, leading to CNV which has been observed by sequence comparisons and by fibre fluorescent *in situ* hybridisation (FISH) (Salomonsen et al., 2014).

Further evidence for the dynamic recombinational landscape of the BG region comes from sequence comparisons (Salomonsen et al., 2014). The *BG* genes in the BG region are expressed either in haemopoietic cells (such as erythrocytes, lymphocytes and myeloid cells) or in non-haemopoietic cells (generally epithelial cells), and this expression correlates with two kinds of promoter and 5'UTR sequences. The exons encoding for the cytoplasmic tail and the 3'UTR also fall into two classes. Overall, the *BG* genes represent the four combinations of the two classes of promoter/5'UTR and the two classes of cytoplasmic tail/3'UTR, with the exons for the signal peptide and the Ig-like region not following any particular pattern, as through recombination in the middle of the genes creates chimeric genes.

This kind of recombination leads to a conundrum. All the exons of *BG* genes are polymorphic, but determining whether the variation is selected has depended on being able to identify alleles, which is difficult in an expanding and contracting multigene family. A comparison of BG transcripts from different BG haplotypes that are expressed in the same cell type (so-called "functional alleles") showed that selection is apparent only in the

cytoplasmic tails, despite the fact that the B locus was originally identified by serological reactions with the extracellular Ig-like domain (Chattaway et al., 2016; Chen et al., 2018). The conundrum is that recombination will lead to a particular cytoplasmic tail being switched to expression in different cell types, and the rationale for the same sequence functioning in different cell types is unclear. However, the function of *BG* genes is currently a mystery, with the first clue being sequence similarities of the extracellular Ig-like domain to mammalian butyrophilins, which are thought primarily to be involved in negative regulation of αβ T cells and with localisation and function of γδ T cells (Henry et al., 1999; Herrmann and Karunakaran, 2022; Rhodes et al., 2016). However, there is evidence that the cytoplasmic tail is important, forming coiled-coils similar to tropomyosin and affecting actin-myosin interactions (Bikle et al., 1996). Moreover, there is some evidence that suggests that the cytoplasmic tail is important for viral resistance (Goto et al., 2009). Perhaps the apparent conundrum will be resolved when the mysteries of BG function are solved.

This report summarises evidence for the divergent properties of two closely-linked regions of the B locus, the relatively stable BF-BL region and the highly dynamic BG region. In between is the TRIM region, whose properties have not been explored. On the centromeric side of the BG region are various genes which, along with the NOR, olfactory genes and scavenger receptor genes, largely remain to be analysed (Delany et al., 2009; Miller et al., 2014). However, evidence for the dynamic nature of the Rfp-Y region has recently been presented (Goto et al., 2022). Future studies may reveal how all these regions are organised and maintained, and may give clues to how they are reorganised in the transition to the different arrangements in other species.

600 Alleles and 200 Haplotypes Identified for the Chicken BF-BL Region
(Prepared by C.A. Tregaskes, R.J. Martin, L. Huynh, N. Rocos, and J. Kaufman)

The major histocompatibility complex (MHC) of mammals is a large and complex region, with hundreds of genes and much recombination, and encodes a few highly polymorphic classical class I and class II molecules that have central roles in immune responses (Kaufman 2016). The functional equivalent of the mammalian MHC in chickens is the BF-BL region, which is remarkably simple and compact with few genes, most of which are critical to the function of classical MHC molecules, so that this region was originally dubbed a "minimal essential MHC". Moreover, recombination within the BF-BL region was considered to be rare, so that this region could exist as relatively stable haplotypes, with co-evolution between these closely-linked genes leading to functional consequences (Kaufman et al., 1999; Kaufman, 2018; Tregaskes and Kaufman, 2021). However, most of these ideas arose from the analyses of a few "standard haplotypes" dating back to the original descriptions by Briles and co-workers (Miller et al., 2004; Miller and Taylor, 2016; Afrache et al., 2020).

We set out to understand more about the diversity of MHC alleles and haplotypes in different chicken populations, starting with reference strand-mediated conformational analyses (RSCA) followed by cloning and sequencing (Potts et al., 2019). As the need for higher through-put became clear, we developed a polymerase chain reaction-next generation sequencing (PCR-NGS) system to type the classical class II B genes *BLB1* and *BLB2*, and the classical class I genes *BF1* and *BF2*. Taking advantage of the compact nature of chicken MHC genes, we amplified exon 2 through the intron to the end of exon 3 (roughly 750 nucleotides) from genomic DNA, and used the Illumina MiSeq to paired-end sequence both exons from each gene (Fig. 36). Coupled with DNA isolation using relatively cheap reagents (which worked for most samples) and a double-barcoding system (12 pairs of barcoded primers each for BF and BL, and 96 barcoded Illumina adaptors), we were able to analyse up to 1152 samples in a run. We developed a bespoke bioinformatics pipeline that automatically generated sequences for all the alleles present, compared them to known alleles and then assembled them into known haplotypes, leaving the unknown sequences to be analysed by inspection (Martin, 2021; C. A. Tregaskes, R. J. Martin and J. Kaufman, unpublished). The initial samples included DNA, blood cells or tissues primarily from experimental lines, red jungle fowl, commercial flocks, fancy birds and African chickens provided by many collaborators. Altogether, 22 different MiSeq runs were performed covering roughly 20,000 samples. For some populations, we used the microsatellite LEI0258 (Fulton et al., 2006) and BF2-specific PCRs to confirm and extend the assignments (Bertzbach et al., 2022; L. Huynh, C. A. Tregaskes and J. Kaufman, unpublished). For some samples of blood cells, flow cytometry was

performed to determine the expression level of the class I (BF) molecules on erythrocytes that is known to correlate inversely with peptide repertoire (Chappell et al., 2015), which in some cases was determined by immunopeptidomics. Almost all the experimental work is now complete, except for some PacBio sequencing that has become necessary to assign BLB sequences to the appropriate loci (N. Rocos, C. A. Tregaskes and J. Kaufman, unpublished). We identified roughly 600 alleles and found over 200 haplotypes, but there is much analysis to complete, so only an initial overview of some preliminary results will be summarised in this report, with some details expected to change as the analyses are refined.

For understanding alleles and haplotypes, we began by assembling the data known from the scientific literature as well as from nucleotide databases (such as NCBI/GenBank) (Afrache et al., 2020). From the literature, sixteen complete BF-BL haplotypes were known from sequencing cosmids, bacterial artificial chromosomes (BACs) and long-range PCR products. Other haplotypes could be assembled from complete or partial (exon 2-exon 3 for BF, exon 2 and exon 2-exon 3 for BLB) genomic sequences, as well as complete or partial cDNA sequences. However, many gene and cDNA sequences in the databases had to be ignored, even if published, since they were deposited from single studies and differed by only one or two nucleotides, with the associated papers revealing that only one amplification had been carried out, so most of those sequences were not separate alleles but the result of nucleotide misincorporation. Altogether, 17 standard haplotypes seemed secure, and an additional 16 haplotypes were suggested (Afrache et al., 2020). Even these data must be treated with some caution; for example, the B6 and B15 haplotypes reported in the largest haplotype-sequencing project (Hosomichi et al., 2008) have not yet been found in any sample examined by PCR-NGS.

It was very easy to assign the 339 class I sequences found by PCR-NGS (after 22 MiSeq runs) to the *BF1* and *BF2* loci (Martin, 2021) since they were found almost exclusively in different clades by neighbour-joining, maximum likelihood and minimum evolution tree building algorithms (Fig. 37). The exceptions include nine alleles related to BF1*0201 and BF1*0901 which cluster together, as well as a couple of other sequences present in other clades, all in the BF2 part of the tree. The *BF2* locus was more polymorphic, with 247 *BF2* alleles compared to 92 *BF1* alleles. Despite largely being in one clade, *BF1* alleles had much sequence diversity, with deep branches in the BF1 clade tree. However, much of this variation was not obviously in the peptide-binding site, with 74% of the BF1 sequences having His9 and Asp24, which may mean a wide peptide repertoire (as recently found from the structure of BF1*1901). Moreover, 89% of the BF1 sequences had an identical or near-identical sequence in the region of the C1/C2 epitope on the α1 helix, consistent with the suggested role of BF1 molecules as natural killer (NK) ligands (Ewald and Livant, 2004; Kim et al., 2018).

In contrast to class I loci, it became very difficult to assign all 259 BLB sequences to the *BLB1* and *BLB2* loci (Martin, 2021), largely due to finding haplotypes with two new class II B sequences, both of which were most closely-related to known BLB1 sequences. The known BLB1 and BLB2 sequences (as well as many identified by PCR-NGS) were mixed in phylogenetic trees (Fig. 37). Ongoing experiments using "between gene" primers and PacBio sequencing to assign sequences to the *BLB1* locus adjacent to *Blec* and to the *BLB2* locus adjacent to *BRD2* have resolved the ambiguities in 26 of 47 unclear BLB haplotypes (N. Rocos, C. A. Tregaskes and J. Kaufman, unpublished). These experiments have revealed the same sequence in one locus in one haplotype and the other locus in a second haplotype, as well as one haplotype with the same sequence in both loci. It seems likely that most haplotypes will have to be checked to ensure that new sequences most closely-related to known BLB2 sequences are actually located in the *BLB2* (that is *BRD2*-adjacent) locus. Given the facts that the *BLB1* and *BLB2* genes are in opposite transcriptional orientation and that most of the gene sequences are nearly identical, one possibility is that recombination between homologous sequences in the *BLB1* and *BLB2* loci leads to inversion; in the PacBio run done thus far, there has been no convincing evidence for such inversions.

The optimal choice of nomenclature for BF and BLB alleles continues to be unclear. The old accepted nomenclature was based on haplotypes, so the same sequence in two haplotypes would have different names (Miller et al., 2004). Based on the nomenclature system originally described for human MHC alleles and widely used for other species (Ballingall et al., 2018; Robinson et al., 2020; Afrache et al., 2020), the gene designation would be separated from the allele designations by a star or asterisk, with distantly-related alleles of a single locus differing in the first two numbers (eg., BLB1*02 versus BLB1*04), and with closely-related alleles of a single locus having the same first two numbers and differing in the next two numbers after a colon (eg., BLB1*02:01 and BLB2*02:02). Haplotypes would then be constructed by strings of alleles (eg., BLB1*02:01-BLB2*02:02-BF1*02:04-BF2*02:05 or in short 2-2:02-2:04-2:05). This elegant solution ran into trouble from the criteria for close relationship, in that the number of sequence differences within clade of closely-related sequences could exceed the number of sequence differences between two sequences from different clades. Moreover, the same

BLB sequence has now been found experimentally in both the *BLB1* and *BLB2* loci, so how should it be named? At the moment, designations for many sequences are simply *ad hoc*, as we struggle to develop a consistent approach.

Of the "standard haplotypes", seven were exhaustively analysed over some years in our lab, five of which were included among the 14 subsequently analysed by another lab in a single sequencing paper (Wallny et al., 2006; Shaw et al., 2007; Hosomichi et al., 2008). Of these 16 haplotypes, all had different *BF2* alleles except for two B15 haplotypes which differed in *BF1*. The original B15 haplotype described (and almost all subsequently) had no expressed *BF1* allele (as did the B14 haplotype), but the B15 haplotype from a chicken line in Japan had a *BF1* allele present. Among the 242 haplotypes identified by PCR-NGS (although there are a few from published data that we have not found), 27 (11%) have no *BF1* allele amplified by the primers used. Originally Southern blots suggested an insertion in the BF1 loci that were not expressed (Wallny et al., 2006; Shaw et al., 2007), but the latest experiments with primers outside the gene have amplified this region in those haplotypes, and identified a deletion of the whole *BF1* gene which is the result of two short direct (but imperfect) repeats (N. Rocos, F. J. Coulter and J. Kaufman, unpublished).

Of the "standard haplotypes", B19 was identified as a recombinant of B12 and B15 haplotypes, and three haplotypes (B5, B8 and B11) were also found to be recombinants, although by "gene conversion" of long stretches of DNA (Wallny et al., 2006; Shaw et al., 2007; Hosomichi et al., 2008). Among the 242 haplotypes, 128 (53%) could have arisen by recombination between *BLB2* and *BF1* (some with subsequent mutation to produced closely-related alleles); 22 BLB haplotypes are in combination with 96 BF haplotypes (with closely-related *BF* alleles combined, since they might have arisen from mutation subsequent to recombination). The most extreme is the BLB haplotype 5-5, which is associated with 25 different BF haplotypes (with closely-related BF alleles combined). There is also apparent recombination between *BF1* and *BF2*, with 15 different *BLB1-BLB2-BF1* haplotypes in association with 37 *BF2* alleles. As an example, 4-8-4 is found with BF2*24:01, 53:01 and four closely-related 43-type alleles.

The first analyses by RSCA were performed with high level (elite, great-grandfather) lines of commercial breeders, and we were shocked at the low diversity of these populations; some had only a single BF-BL haplotype. In order to better understand the commercial chickens that are actually in the field, we obtained farm-level samples from our collaborators, examining six broiler lines, 15 egg-layer lines and one dual purpose line. The take-home message is that there are typically very few haplotypes, mostly four or five haplotypes above 1% genotype frequency, usually with one haplotype by far the majority. In particular, a haplotype not described before (made up of alleles previously described, provisionally called B31) is present in 33-64% genotypes of the six farm-level broiler flocks. Similarly, a previously undescribed haplotype provisionally called B9:02 dominates nearly all the brown egg layer flocks. If these numbers are representative, then there are billions of chickens in the field that are MHC homozygotes.

How do these commercial chickens survive with such low MHC diversity? Part of the answer may be that most of the high frequency haplotypes are those with BF2 alleles that have low cell surface expression and promiscuous peptide binding (for those with known peptide motifs). Such so-called "promiscuous haplotypes" are known to protect against a variety of economically-important infectious diseases in chickens and have been suggested to act as generalists, in contrast to "fastidious haplotypes" which may act as specialists (Chappel et al., 2015; Kaufman, 2018; Tregaskes and Kaufman, 2021). A few haplotypes with high-expressing *BF2* alleles are found in some populations; these may function as specialists or have some other useful attribute(s).

A wealth of information has already emerged from this PCR-NGS typing, but there is much more to be learned by finishing the detailed analysis of commercial, fancy and African chickens. Moreover, there are many chickens worldwide that have not been examined by this kind of analysis, particularly in South, Southeast and East Asia. As the typing methods benefit from longer reads that cover more genes, many interesting attributes of the chicken MHC are likely to be revealed.

The Chicken Genomes and Their Diversities
(Prepared by J.-L. Han and O. Hanotte)

This review intends to provide a comprehensive summary of the status of chicken genome resources and what we have been learning from the analyses of genomic diversity of our commonest domestic poultry species.

Chicken Genome Assemblies
The chicken was the first bird and agricultural animal species to have its genome sequenced. The first reference genome of 6.6× genome coverage from a female red jungle fowl, known as "RJF #256" (inbred line, UDC 001), was published in December 2004 [International Chicken Genome Sequencing Consortium, 2004]. It was obtained through Sanger sequencing and included around one billion nucleotides and 20,000–23,000 annotated genes. This genome assembly was further improved by adding 14× genome coverage next-generation sequencing (NGS) data and referred to as Galgal3 or Gallus_gallus-2.1 (GenBank RefSeq assembly accession no. GCF_000002315.1 and GCF_000002315.2 released in November 2006) [Groenen et al., 2011]. It was followed by a *de novo* assembly of the "RJF #256", using all available Sanger and NGS reads available at the time and new Illumina short reads at 68.6× genome coverage (Galgal4, Gallus_gallus-4.0 with GCF_000002315.3 released in November 2011) [Ye et al., 2011; Schmid et al., 2015]. Then, another *de novo* assembly, based on third-generation sequencing data (PacBio RSII long reads at 50.6× genome coverage) of the "RJF #256" was merged with all Galgal4 sequences and referred to as Gallus_gallus-5.0 (GCF_000002315.4 released in December 2015) [Warren et al., 2017]. This genome assembly was finally upgraded by adding PacBio RSII long reads to a sequencing depth of around 80× and by generating a high chromatin proximity map to help in the order and orientation of the assembled contigs and scaffolds. It is referred to as GRCg6a (GCF_000002315.6 released in March 2018). This genome has a 1.07 Gb total genome size. It includes 34 chromosomes with 1,402 contigs assembled into 524 scaffolds. The N50 length for the contigs is 17.66 Mb, while the scaffold N50 is 20.79 Mb. The NCBI *Gallus gallus* Annotation Release 104, including 17,477 protein-coding and 6558 noncoding genes (released in May 2018), and the Ensembl release 94, including 16,878 coding and 7166 noncoding genes (released in April 2018) (database version 106.6) complement the GRCg6a assembly.

Development and Application of SNP Arrays for the Assessment of Genomic Variation
Considering the details on gene expression microarrays in an early review (Gheyas and Burt, 2013), we focus on the development of genotyping arrays in this summary. The first Illumina 3K SNP array was developed based on the selection and validation of 3,072 out of the 2.8 million SNPs at one SNP each in 3,072 bins distributed evenly throughout the chicken genome. It was based on the genome sequence of WASHUC1 assembly (Galgal2 released in February 2004 at https://genome.ucsc.edu/cgi-bin/hgGateway) with linkage information to account for the recombination rate of each chromosome [Aerts et al., 2007]. In addition, 34 SNPs in genes of interest were added [Muir et al., 2008a]. Then, a moderate density (60K) Illumina SNP BeadChip was developed using additional SNPs from the Illumina short reads of broiler and layer chicken populations aligned against Gallus_gallus-2.1.0 (released in May 2006), 454-read-based contigs of the "RJF #256", and the mitochondrial genome. It included 60,800 SNPs [Groenen et al., 2011]. A 600K Affymetrix Axiom high-density (HD) genotyping array was independently constructed based on 139 million SNPs identified by aligning the Illumina short reads of 243 chickens from 24 lines of commercial broilers, white-egg layers, and brown-egg layers as well as experimental inbred layers and unselected layer line against the Galgal4. They were scaled down to 1.8 million SNPs as a robust and tractable subset for HD array design, including those built in the Illumina 3K (Muir et al., 2008a) and 60K (Groenen et al., 2011) arrays, and SNPs among the seven million SNPs identified from chicken populations of different lines [Rubin et al., 2010]. The 580,954 SNPs were validated by genotyping an additional 282 chickens, including trio samples from three types of commercial lines and traditional breeds [Kranis et al., 2013]. All segregating SNPs on this Affymetrix 600K array were evenly spaced across the genome, following the genetic map distances, instead of their physical distances, for both broiler and layer lines [Groenen et al., 2009], to account for the difference in recombination rates between macro- and micro-chromosomes [Rodionov, 1996; Groenen et al., 2000, 2009; International Chicken Genome Sequencing Consortium, 2004; Megens et al., 2009; Liu et al., 2021]. These SNP arrays have significantly facilitated genome-wide association studies (GWAS), paving the way for genomic selection, identification of selection signatures, fine mapping of QTLs, and detection of copy number

variations (CNVs). Although the first genome-wide scan based on the WGS data at low genome coverage of four birds failed to identify selective sweeps for adaptive alleles following chicken domestication [International Chicken Polymorphism Map Consortium, 2004], several selective sweeps linked with the genes associated with growth, appetite, and metabolic regulation were subsequently detected using enhanced genome-wide SNPs of broilers, while one of the most striking selective sweeps at *TSHR* locus has been considered as a 'domestication' signature [Rubin et al., 2010; Elferink et al., 2012].

These SNP arrays were also widely used for the characterization of population genomic diversity and genetic structure of many wild, commercial, experimental, indigenous, and fancy chicken breeds/populations [Muir et al., 2008b; Elferink et al., 2012; Zhang et al., 2020; Malomane et al., 2019; Cendron et al., 2021], even though SNP ascertainment bias might be of some concern [Malomane et al., 2018; Geibel et al., 2021]. Also, population stratification can lead to false associations in GWAS. Henceforth, a careful examination of population structure is warranted [Kranis et al., 2013].

The Illumina 3K array was used for genotyping 2,551 informative SNPs in 2,580 chickens from commercial male and female broilers, white- and brown-egg layer pure lines, experimental, and traditional breeds. Analysis of commercial lines showed a loss of 50% or more of the genetic variability present in ancestral breeds, a consequence of founding effects in making these lines. It raises questions about the suitability of these lines to respond to future needs (consumers, societal, and producers' needs), as well as about their genetic repertoire for resistance to infectious disease challenges [Muir et al., 2008b]. The Illumina 60K array was used to genotype 51,076 autosomal SNPs in 67 populations, including Ceylon Junglefowl (*G. lafayetti*), red junglefowl (*G. g. gallus* and *G. g. spadiceus*), eight broiler sire lines, five broiler dam lines, 11 white egg-layer lines, 11 brown egg-layer lines, 19 traditional Dutch breeds, and 10 Chinese indigenous breeds [Elferink et al., 2012]. A phylogenetic tree rooted at the Ceylon Junglefowl separated the red junglefowls from all domestic chickens that were divided into two branches, one including brown-egg layers, broilers, and Chinese breeds, and the other white-egg layers and Dutch breeds. Among the broilers, the sire and dam lines were separated from each other, supporting their distinct origins.

The Affymetrix 600K array was evaluated for its utility in inferring population stratification with samples of known history through principal component analysis (PCA). It was found that birds from the same line/breed clustered together. The broiler lines were closer to the brown-egg layers than to the white-egg layers. In contrast, the two white-egg layer lineages were separated from each other, supporting different origins of the two lines [Kranis et al., 2013].

Also, it was observed that the phylogeny reconstructions using genome-wide SNP arrays mirrored those based on microsatellite data [Eding et al., 2002] and known geographic origins and breeding history of the studied populations [Elferink et al., 2012]. In particular, the broilers and brown-egg layers shared their ancestry. They were initially developed by crossing European and Asian breeds, while the white-egg layers originated from the single-combed White Leghorn of European origin [Crawford, 1990; Muir et al., 2008b; Elferink et al., 2012].

The Affymetrix 600K array had been further applied to several large-scale population genomic studies. For instance, genotyping of 1200 chickens, ranging from 41 to 469 samples of five Chinese breeds of Beijing-You, Hongshan, Shouguang, Taihe Silkie, and Tibetan, along with White Leghorn originating from Italy, Houdan chicken from France, and Rhode Island Red from USA, identified a higher genomic diversity in most Chinese breeds compared to exotic ones, while all breeds carried some unique polymorphisms allowing successful assignment of all samples back to the breeds of origins. Local Tibetan chickens with high-altitude adaptability had a high level of genetic admixture from White Leghorn [Nie et al., 2019]. The addition of 69 samples (including five Chinese game breeds, Cornish chicken, and red junglefowls) to the 1200 chicken dataset strengthened the distinct genomic variability present within each breed, with the highest genomic uniqueness observed in White Leghorn, Houdan, and Rhode Island Red. Cornish breed, known as 'Indian Game' carrying the genetic footprint of Malay and other Oriental chicken blood, showed a close genomic relationship with five Chinese game breeds, all sharing a highly admixed genomic background of likely a single origin. As expected, the red junglefowls showed the highest genomic diversity, followed by Cornish, four of the five Chinese game breeds, other Chinese local breeds, French Houdan, and commercial breeds, based on linkage disequilibrium decay (LD), effective population size, and runs of homozygosity (ROH) estimates [Zhang et al., 2020]. A regional study on seven indigenous breeds in the Jiangxi province of China indicated a similarly higher genomic diversity but smaller ROH in most Chinese indigenous breeds relative to European and commercial breeds. Recent meat and egg production selection have resulted in reduced genomic diversity but increased ROH in improved Chinese local breeds [Chen et al., 2019]. The genotyping of eight Italian local chicken breeds showed unique genomic diversity in most breeds, calling for

breed-specific conservation strategies [Cendron et al., 2021]. Also, genomic variation in three Chinese indigenous breeds (Baier Yellow, Beijing-You, and Langshan) maintained by ongoing *ex-situ* conservation programs were evaluated at three different generations (generations 7, 10, and 15 for Baier Yellow; 7, 10, and 13 for Beijing-You, and 10, 12, and 15 for Langshan). These conserved flocks were managed by keeping 30 males and 300 females per generation and implementing a random mating within families with one son selected one sire family and one daughter from one dam family. There was no differentiation in population genetic structure within the breeds over the three generations [Zhang et al., 2020]. The genomic diversity of a live Norwegian poultry Genebank indicated that inbreeding level was high in all lines, while the relatively more inbred white-egg layers were differentiated from the brown-egg layers that contributed more to total genetic diversity. Though distinct from other commercial populations, the newly developed Norwegian commercial lines were closely related to them. These Norwegian Genebank lines were therefore believed to be of conservation value at national and international levels [Brekke et al., 2020].

Several customized SNP arrays have also been developed in recent years. For example, a 42K Illumina SNP array was developed by private funds from the EW Group (Visbeck, Germany) [Fulton, 2012]. Also, a proprietary Affymetrix 50K SNP array was built based on a subset of SNPs extracted from the Affymetrix 600K array to capture specific genetic diversity in highly selected and pedigreed populations, along with additional SNPs identified in previous studies [Wolc et al., 2020]. It is expected that the genotypes of this 50K SNPs may be imputed into the 600K SNPs using the parent, grandparent, and great-grand-parent 600K SNP data [Herry et al., 2018; Psifidi et al., 2021]. These two arrays were specifically designed for QTL mapping and genomic selection in commercial layer and broiler lines, and they were never made publicly available.

Another Affymetrix 55K genotyping array (IASCHICK) incorporating specific SNPs identified by aligning the WGS reads from several Chinese indigenous chicken breeds, which were poorly represented in early arrays, against the Galgal4 [Liu et al., 2019], and an Illumina 50K BeadChip (PhenoixChip-I) containing unique SNPs from several commercial layer lines important to the Chinese market, based on the 1,846,003 SNPs screened by aligning their WGS reads against the GRCg6a [Liu et al., 2021], were recently developed. Both IASCHICK and PhenoixChip-I arrays also include many candidate SNPs associated with important economic traits in chicken [Liu et al., 2019; Liu et al., 2021]. The IASCHICK array proved to be effective in detecting within-population genetic diversity in nine indigenous and improved Chinese breeds and three commercial layer and broiler lines, with calling rates ranging from 97.0% to 98.7% and polymorphic SNPs ranging from 76.7% to 88.0% across the breeds/lines. But the genotyping results from the PhenoixChip-I array showed some level of ascertainment bias. Still, the genetic variation identified by both IASCHICK and PhenoixChip-I arrays had a sufficiently high resolution to support the assignment of all samples back to their expected breeds/lines of origins, validating the power of 50K genome-wide SNP arrays for population genetic diversity and structure investigations.

SNP genotyping arrays have also been used to analyze copy number variations (CNVs) of genomic structural variants (SVs) in the form of segmental insertion, deletion, and duplication greater than 50 bp. They are a significant source of genomic diversity underlying phenotypic variation. For instance, a duplicated sequence close to the first intron of *SOX5* has been linked to the pea-comb phenotype [Wright et al., 2009], while an inverted duplication including *EDN3* is associated with dermal hyperpigmentation in chicken [Dorshorst et al., 2011]. In addition, partial duplication of *PRLR* has been related to late feathering [Elferink et al., 2008], whereas a CNV including the *HOXB7* and *HOXB8* genome region has been associated with the beard phenotype [Yang, et al., 2020] in chicken.

Using the Illumina 60K array, up to 818 CNVs were identified in 184 White Leghorns and 233 brown-egg dwarf layers, of which 315 were unique. They aggregated into 209 copy number variation regions (CNVRs) on 27 autosomal chromosomes. These CNVRs were distributed proportionately to the chromosomal length, e.g., 14.7 CNVRs on macro-chromosomes but 3.65 CNVRs on micro-chromosomes. The CNVRs shared by the two breeds suggest their relatively ancient origins, while some CNVRs were breed specific [Jia et al., 2013]. Also, 137 CNVRs were reported in 1310 Beijing-You chickens [Zhou et al., 2014]. Genotypes of 475 birds after the 11th generation of divergent selection for abdominal fat content in the same grandsire line of the Arbor Acres broilers revealed 438 and 291 CNVs, of which 271 (176 loss, 68 gain, and 27 loss+gain) and 188 (143 loss, 25 gain, and 20 loss+gain) CNVRs were low- and high-fat line specific, respectively. Differences in genetic drift and selection were thought to have contributed to the variation in the numbers and distribution pattern of both CNVs and CNVRs between the two lines [Zhang et al., 2014]. Genotypes of 554 chickens from an F_2 full-sib population, a cross between Xinghua and White Recessive Rock chickens, identified 1875 CNVs distributed in 209 CNVRs, of which 109 were novel [Rao et al., 2016]. A follow-up analysis of the CNVs in this F_2 population revealed a polymorphic CNV overlapping with

SOX6, with the number of CNVs positively associated with the expression of *SOX6*, which is linked to skeletal muscle development [Lin et al., 2018].

Only large CNVRs were identified in these studies using the Illumina 60K array due to its low marker density and non-uniform marker distribution. To improve the efficiency and reliability of SNV detection, the Affymetrix 600K array has been used in several studies. For example, Chinese indigenous chicken and exotic commercial breeds carried 5.1 and 3.3 CNVs per bird, respectively; both values were much higher than that reported earlier [Jia et al., 2013]. After the calibration of the CNVs between the Galg4 and Galgal3 for comparison with those detected in early studies, 153 CNVs were found to be novel [see Table 2 in Yi et al., 2015]. Genotyping of 48 deformed-beak and 48 normal Beijing-You chickens identified *LRIG2* as a candidate for beak deformity [Bai et al., 2018]. Four SNP chips (Illumina 42K, Affymetrix 600K, and two customized Affymetrix 50K chips) were successfully used to genotype and identify CNVs in 18,719 chickens from four pure lines and one commercial cross. Here, the genome landscape of CNVs was determined not only by the number of samples and genetic background of the lines, but also by the SNP density on the arrays: the higher the density of the arrays, the more the CNVs per sample [Drobik-Czwarno et al., 2018]. As many as 1,003 CNVs in 564 CNVRs were identified in 94 chickens of six local Italian breeds, contributing to their distinctiveness [Strillacci et al., 2017]. Also, 1924 CNVs in 1216 CNVRs were observed in 265 local Mexican chickens [Gorla et al., 2017]. These CNVRs were mapped on 28 autosomes in the Galgal4 [Gorla et al., 2017; Strillacci et al., 2017]. Using 23,214 CNVs and 5042 CNVRs in 1238 chickens of a Brazilian male broiler line, originating from the crossing of White Plymouth Rock and White Cornish breeds, CNV-based GWAS revealed potential candidate genes such as *KCNJ11*, *MyoD1*, and *SOX6* nearby several CNVs associated with growth traits [Fernandes et al., 2021].

In most of these studies, the same parameters implemented in the PENNCNV software, based on Hidden Markov Model (HMM), were applied to call CNVs from genotyping data, including signal intensity (log R ratio, LRR), allelic intensity (B allele frequency, BAF), and marker distance and population frequency of allele B [Wang et al., 2007]. The HMM with default parameters but a cutoff for the standard deviation of LRR < 0.30 were considered [Jia et al., 2012; Zhang et al., 2014; Zhou et al., 2014; Rao et al., 2016]. The CNVs containing at least three consecutive SNPs were specifically chosen [Zhang et al., 2014]. Aggregating overlapping CNVs determined CNVRs among individuals [Redon et al., 200]. A common pattern was that most CNVs and CNVRs ranging 68.8% to 82.9% were singletons and thus segregated among individuals within a breed/population [Yi et al., 2014; Gorla et al., 2017; Strillacci et al., 2017].

To reduce the genotyping cost, low-density SNP chips (e.g., 10K SNPs) have been used in populations previously studied using high-density SNP arrays followed by SNP imputation [Herry et al., 2018; Huang et al., 2018]. Other low-cost methods such as reduced representation sequencing based on restriction enzyme cleavage have also been attempted for SNP discovery, validation, and characterization in chicken populations. Among them, restriction-site-associated DNA sequencing (RAD-seq) [Zhai et al., 2015; Yang et al., 2020], specific locus amplified fragment sequencing (SLAF-seq) [Jin et al., 2015; Wang et al., 2016, 2019; Li F et al., 2018, 2021], genotyping by genome reducing and sequencing (GGRS) [Liao et al., 2015, 2016; Zhao et al., 2018], and genotyping by sequencing (GBS) [Pértille et al., 2016; Zhang et al., 2018; Habimana et al., 2021; Zhang et al., 2021] proved to be effective in identifying and genotyping some novel SNPs in chickens. However, these methods have not yet been widely applied.

De novo Assembly of the New Chicken Reference Genome

Two *de novo* haplotype assemblies using a trio of samples, including an F_1 hybrid female (bGalGal1) from a broiler hen (bGalGal2) mated with a White Leghorn cock (bGalGal3), were recently released: bGalGal1.mat.broiler.GRCg7b (GenBank and RefSeq assembly accession nos. GCA_016699485.1 and GCF_016699485.2, released in January 2021) and bGalGal1.pat.whiteleghornlayer.GRCg7w (GenBank and RefSeq assembly accession nos. GCA_016700215.2 and GCF_016700215.2, released in October 2021). They were assembled using PacBio Sequel I CLR, Illumina NovaSeq, Arima Genomics HiC, and Bionano Genomics DLS reads at 102.01× genome coverage. bGalGal1.mat.broiler.GRCg7b has a total genome length of 1.05 Gb in 677 contigs assembled into 214 scaffolds with N50 contigs of 18.83 Mb and N50 scaffolds of 90.86 Mb, while bGalGal1.pat.whiteleghornlayer.GRCg7w also had 1.05 Gb in total length in 685 contigs assembled into 276 scaffolds with N50 contigs of 17.74 Mb and N50 scaffolds of 90.56 Mb. There are 17,007 coding and 13,040 noncoding genes in GCA_016699485.1 and 16,884 coding and 13,294 noncoding genes in GCA_016700215.2 from the Ensembl release 104 published in January 2022 (Database version 108.7); while 18,023 coding and 7330 noncoding genes in GCF_016699485.2, and 17,981 coding and 7,310 noncoding genes in GCF_016700215.2 are

from the NCBI *Gallus gallus* Annotation Release 106 published in March 2022. It may be expected that these two new *de novo* assembled genomes will be of broad interest to the scientific and industrial communities, considering their highest sequencing depth and coverages, most complete annotations, and relevance to commercial chicken production [See article by Warren et al., this report].

De novo Assemblies of Indigenous and Commercial Chicken Breeds
Following the reduction in sequencing cost and the advent of new sequencing technologies, *de novo* genome sequences of chicken breeds of particular interest are being generated to identify the genetic control of their unique phenotypes. These are warranted for comparative genomic analyses of SNPs, insertions, and deletions (INDELs), structural variations (SVs), and coding and noncoding transcriptomes.

The first *de novo* draft genome of an indigenous chicken, the Korean Yeonsan Ogye, was released in December 2017 (Ogye1.0, GenBank Genome Accession no. GCA_002798355.1). This breed is characterized by having an entirely black external feature and internal organs. Following a hybrid *de novo* assembly pipeline, the genome was assembled into 8,241 pseudo-contigs and 1906 scaffolds, which were further aligned and anchored to the GalGal4 chromosomes, based on high-depth Illumina HiSeq short reads (376.6×) and low-depth PacBio RS II long reads (9.7×). The Ogye1.0 genome size is 1.02 Gb with 7721 contigs at N50 of 639.81 Kb and 1,821 scaffolds at N50 of 90.11 Mb. Compared with the Galgal4, the Ogye1.0 genome has 551 SVs, including the duplication of the *FM* locus related to hyperpigmentation. Moreover, 15,766 coding and 6,900 long noncoding RNA genes were annotated based on transcriptomic data from 20 tissues, of which 946 were novel coding genes. However, 164 functional coding genes reported previously were not identified. The Ogye 1.1 genome was estimated to be 97.6% complete based on assessing single-copy orthologous genes using the Benchmarking Universal Single-Copy Orthologs (BUSCO), a value similar to the Galgal5 (97.4%) [Sohn et al., 2018].

Recently, more *de novo* genome sequences of indigenous chicken breeds have been assembled. They include the genome of a female Huxu chicken (GCA_024206055.1, submitted in July 2022). The genome is around 1.10 Gb with 54 contigs at N50 of 91.36 Mb and 40 scaffolds at N50 of 91.36 Mb. The PacBio RSII and ONT long reads and the Illumina NovaSeq short reads at 80.0× genome coverage were used to assemble the Huxu genome. Also, 20 genomes of 14 breeds were *de novo* assembled. They were all generated using long and short sequence reads. Specifically: PacBio Sequel II long reads (>87×), Illumina HiSeq short reads (>56×), and HiC data (>112×) for six genomes at chromosome level for Houdan chicken (GCA_024653045.1), Rhode Island Red (GCA_024652985.1), White Leghorn (GCA_024652995.1), Cornish (GCA_024653035.1), Silky (GCA_024653025.1), and Tibetan chicken (GCA_025370635.1); PacBio Sequel II long reads (>53×) and Illumina HiSeq short reads (>45×) for three genomes at scaffold level for Asil (GCA_024686355.1), Naked Neck (GCA_024686465.1), and Thailand Gamefowl (GCA_024686285.1); Illumina HiSeq short reads (around 134×) for 10 genomes at scaffold level (GCA_024653045.1 - Silky, GCA_024679355.1 - Daweishan, GCA_024679375.1 - Liyang, GCA_024679395.1 - Tibetan, GCA_024679415.1 - White Plymouth Rock, GCA_024679765.1 - Rhode Island Red, GCA_024679905.1 - White Leghorn, GCA_024686295.1 - Chahua, GCA_024686315.1 - Langshan chicken, and GCA_024687005.1 – Cornish chicken); and one at contig level based on PacBio Sequel II long reads and Illumina HiSeq short reads (GCA_024686275.1 - Fayoumi). The completeness of these 20 genomes ranged from 92.4% to 95.3% (BUSCO analysis), comparable to the GRCg6a (95.4%). The pangenome analysis of these 20 genomes aligned to the GRCg6a identified 1,335 novel coding genes, among which many were housekeeping genes and genes involved in immune pathways. These immune-related genes had a 3-fold elevated substitution rate. There were also 3,011 new long noncoding RNAs absent in the GRCg6a [Li et al., 2022].

WGS-Based Population Genome Analyses
As mentioned above, the SNP arrays were designed for genotyping polymorphisms at hundreds or thousands of specific locations across the chicken genome. They have been used to assess the ancestry of particular breeds/populations/lines and identify signatures of selection and candidate genomic regions associated with complex and multifactorial phenotypes through GWAS, including production and adaptive traits. However, they have some limitations. These SNP arrays largely include SNPs common either across different breeds/populations from a relatively large geographic coverage or within specific breeds/lines of limited ancestry and/or under intensive selection. These SNP arrays will therefore perform poorly for genotyping rare genetic variants. It is well known that genetic diversity is not equally distributed across populations of different ancestries, and some of these rare variants may become common in genetically isolated populations. Together with the challenge of uneven spacing of SNPs across the genome, they contribute to SNP ascertainment bias, which leads to biased

inferences (e.g., proportion of admixture) [Lachance and Tishkoff, 2013; McTavish and Hillis, 2015; Dokan et al., 2021]. There are also limitations associated with the application of the SNP arrays in detecting CNVs (e.g., chromosomal translocations and inversions).

Considering the constraints of SNP arrays and the reduction of DNA sequencing costs, NGS-based WGS approach is now the most popular tool to infer the evolutionary history of chicken genomes, assess the genetic diversity and population structure at regional, continental, and global scales, identify the genetic controls of specific traits of interest, and to screen for the signatures of selection for productive traits and environmental adaptation [Rubin et al., 2010; Mugal et al., 2013; Wang MS, et al., 2015, 2016a, 2016b, 2017, 2020, 2021; Guo X, 2016, 2022; Guo Y, et al., 2016, 2021; Zhang et al., 2016, 2022; Li D, et al., 2017, 2019; Boschiero et al., 2018; Derks et a., 2018; Lawal et al., 2018, 2020; Sohrabi et al., 2018; Sohrabi et al., 2018; Almeida et al., 2019; Talebi et al., 2020; Tiley et al., 2020; Wang Q, et al., 2020; Gheyas et al., 2021, 2022; Li YD, et al., 2021; Liu et al., 2021; Mariadassou et al., 2021; Rostamzadeh Mahdabi et al., 2021; Tan et al., 2021; Yang Z, et al., 2021; Asadollahpour Nanaei et al., 2022; Chen et al., 2022; Gao et al., 2022; Geibel et al., 2022; Li Y, 2022; Liu et al., 2022; Sun et al., 2022; Wang et al., 2022; Xu et al., 2022; Yuan J, et al., 2022; Yuan X, et al., 2022; Zeng et al., 2022].

For example, a novel panel of 64 exonic SNPs screened from WGS data were applied to the genetic characterization of Italian local breeds and proved to be cost-effective for genotyping many samples to aid genetic traceability and breeding programs [Viale et al., 2017]. Also, to further explore the deficit in homozygous carriers for 77 haplotypes in four purebred white- and brown-egg lines and two crossbred lines observed using the Illumina 60K array, the WGS data of 250 white-egg layers were annotated. This analysis identified 4219 putative deleterious variants, including 152 mutations relevant to embryonic lethality, at homozygous state. These deleterious variants present in genomic regions of low recombination rates have been subjected to purifying selection [Derks et al., 2018]. Many deleterious and stop-gain/loss SNPs were also observed in the WGS data of Brazilian meat and white-egg layer chickens [Boschiero et al., 2018].

Homozygous deleterious variants were also identified in the WGS data of red junglefowls and domestic chickens, with 2.95% more mutations in the chicken genomes compared to its wild ancestor, which was interpreted as the result of the "cost of domestication". Around 62.4% of these deleterious variants were in the heterozygote state and believed to be recessive [Wang et al., 2021]. Considering the possible occurrence of harmful mutations linked to favorable variants in positively selected sweeps and the positive relationship between recombination rate and purging efficiency, a genomic marker-assisted selection is highly recommended to minimize the frequency of undesirable functional mutations, while to sustainably improve both indigenous and commercial chicken genetic resources [Derks et al., 2018; Wang et al., 2021].

The WGS-based population landscape of SNPs in 10 Chinese indigenous breeds showed less diversity on the Z chromosome than autosomes that were likely under relatively strong selection pressures. Tibetan chickens were admixed with other breeds, while cockfighting Game chickens were closely related to red junglefowls. Strong signatures of selection were observed at genomic regions linked to the genes likely responsible for the rapid adaptation of Tibetan chickens to high altitudes [Li et al., 2017, 2019]. The WGS data from 863 chickens and junglefowls from America, China, Indonesia, and Europe illustrated their general patterns of variability, ancestry, evolutionary relationship, and breeding history, and convincingly localized the main center of chicken domestication in Southeast Asia. The intensive genomic analyses of domestic chickens also detected several specific genetic admixture events with red junglefowls and other junglefowl species outside the proposed center of domestication [Wang et al., 2020]. Such admixture and/or introgression events have also been reported in another WGS study [Lawal et al., 2020]. Genomic footprints of admixture from indigenous large-size Asian breeds and egg-laying Mediterranean breeds were identified in the gene pool of commercial meat and laying chickens, supporting their contributions to modern commercial meat and egg industries, respectively [Guo et al., 2021]. The first, and largest WGS dataset of African indigenous chickens, including 234 genomes from 24 Ethiopian chicken populations distributed in different types of climates and productive systems, was generated to call up to 15 million SNPs mapped to the GRCg6a reference. High-quality variants have been used to assess population genomic diversity and screen for genomic regions under selection for environmental adaptations. The identified candidate regulatory genes could be epigenetic machineries driving rapid adaptation of African chickens of a relatively recent origin from Asian counterparts (Gheyas et al., 2021, 2022). The WGS analysis of eight indigenous breeds from Guangxi, China revealed genomic diversity similar to the red junglefowl one but limited genetic differentiation and little genetic admixture from commercial chickens [Sun et al., 2022].

The WGS-based phylogeny of all four junglefowl species was reconstructed to achieve a highly confident topology of the genus, which is crucial in inferring the effects of interspecific introgression and genetic admixture on

chicken domestication processes. Following an intensive evaluation of major topological inconsistencies caused by different genetic/genomic datasets and bioinformatic methods, the genetic contamination of some captive junglefowl flocks is verified, while the importance of large number of genetic variants in different genomic components of many reliable samples (e.g., collected from their original ranges in the wild) to phylogenetic inference is substantiated [Tiley et al., 2020; Wang et al., 2020; Mariadassou et al., 2021].

WGS data analysis have also provided new insights on the ROH in chicken genomes. Compared to the red junglefowls with the least number of ROH, white layers carried the largest number of ROH per bird, mostly shorter than 1 Mb, while broilers tended to have relatively more ROH > 2 Mb. Both SNP-based Wright's F_{IS} and ROH-based F_{ROH} metrics suggested higher inbreeding in the white layers compared to the broilers and red junglefowls [Talebi et al., 2020]. Candidate genes in the ROH islands may also be associated with QTLs responsible for productive traits [Talebi et al., 2020] and high-altitude adaptation [Yuan et al., 2022].

As expected, more CNVs and CNVRs were identified from the WGS data compared to the findings of SNP genotyping arrays. For instance, 12,955 CNVs in 5467 CNVRs, accounting for 9.42% of the genome, were found in two Iranian indigenous and commercial chicken breeds, with 34% of these CNVRs overlapping with those identified in SNP-array-based analyses [Sohrabi et al., 2018]. Across 51 WGS datasets of Chinese indigenous breeds (Xinghua, Luxi Game, Beijing-You, and Silkie), commercial lines (Recessive White Rock and White Leghorn), and red junglefowls, 19,329 duplications and 98,736 deletions in 11,123 CNVRs, accounting for 7% of the total autosomal size, and overlapping with 2,636 protein-coding genes were identified [Chen et al., 2022]. Like in all previous SNP-array-based studies, the vast majority of the CNVRs were singletons, with only 152 CNVRs common to all 51 birds. Around 600 CNVRs, including 90 protein-coding genes, were breed specific, suggesting their functional importance in driving chicken phenotypic and adaptive evolution [Chen et al., 2022]. These two analyses clearly demonstrated the power of the WGS data in CNV identification and functional annotation. Last but not least, the possibility of identifying SVs, though challenging for calling accuracy, was demonstrated from the NGS-WGS of commercial chicken genomes, among which deletions were believed to be more accurately detected compared with duplications, inversions, and translocation breakpoints [Geibel et al., 2022].

Chicken Genomic Diversity Consortium: Large-Scale Genomics to Unravel the Origins and Adaptations of Chickens
(Prepared by S.R. Fiddaman, C. Klopp, M. Charles, P. Bardou, O. Lebrasseur, M. Derks, J. Schauer, C. Reimer, J. Geibel, A. Gheyas, A. Smith, R. Schnabel, M.L. Martin Cerezo, M. Nishibori, C.J.P. Godinez, J.K.N. Layos, J.M. Alfieri, H. Blackmon, G.N. Athrey, G. Larson, I. Ng'ang'a, W. Muir, M. Lange, D. Wright, H. Cheng, H. Simianer, S. Weigend, W. Warren, R. Crooijmans, O. Hanotte, J. Smith, M. Tixier-Boichard, and L.A.F. Frantz)

Consortium Description
On October 25-26, 2019, a satellite meeting devoted to the preparation of a Chicken Genome Diversity Consortium was organised after the 11th European Symposium of Poultry Genetics in Prague. Researchers involved in chicken genomics from Europe, Africa and China, discussed the objectives of such a consortium with some presenting their data. However, the technical aspects of how to share and jointly analyse the data were not finalized, nor was the funding model for the cost of data storage and computation. In 2021, an opportunity arose with the call for projects of the SuperMUC computing cluster of the Leibniz-Rechenzentrum in Germany. A new consortium of scientists re-launched the discussion to establish a project with the aim to explore how the high-throughput genomics age can be harnessed to answer evolutionary questions surrounding the chicken. The FARMGENOMIC project (23826) was accepted for funding in autumn 2021, gathering around 20 members from 10 institutions in Europe, North America and Africa. This newly-formed Chicken Genomic Diversity consortium brings together members from a variety of disciplines, including genomics, palaeogenetics, animal breeding, immunology, organismal biology, evolutionary biology, and archaeology. Central to the consortium are the concepts of inclusivity and openness – all data are to be made available to all members of the consortium, and later distributed to the wider community, and collaborations between groups are fostered and actively encouraged. It is hoped this state-of-the-art resource, curated in-house by bioinformaticians, will enable the community to answer previously intractable questions in chicken evolution.

Dataset description

At the core of the consortium is a substantial genomic dataset of chickens and junglefowl. At the time of writing (September 2022), the dataset comprises 4,392 chicken and junglefowl genome sequences, of which 2,307 were derived from public databases, and the remainder provided by consortium members. In addition to domesticated chickens and red junglefowl (comprising all five subspecies: *G. g. gallus*, *G. g. bankiva*, *G. g. jabouillei*, *G. g. murghi*, *G. g. spadiceus*; total n = 291), we also included members of the congeneric *Gallus* species *G. varius* (n = 21), *G. lafayettii* (n=12), and *G. sonneratii* (n=15). Among the domesticated chickens, a wide array of geographical locations are represented (Africa, n = 1,047; East Asia, n = 856; South East Asia, n = 72; South Asia, n = 137; Middle East, n = 219; European fancy breeds, n = 462; North America, n = 835; South America, n = 15; Oceania, n = 24) as well as commercial birds (n = 329) and experimental lines (n = 42). The wide scope of the dataset aims to capture a significant proportion of the extant genetic variation in the chicken genome. Furthermore, the addition of 15 ancient chicken genomes from Europe and the Middle East will provide supplemental time-depth, including a window into past genetic variation following the arrival of chickens into Europe from Southeast Asia (Fig. 38)

Consortium Aims
The aims of the consortium are numerous and varied, reflecting the diverse interests of the contributing groups. Specific scientific aims include:

1. **Deleterious alleles and possible inbreeding**. Breeds with high rates of inbreeding and potential health risks will be identified on the basis of genetic load and deleterious variants in sequence data. We will also investigate loss-of-function variation in relation to pseudogenes and adaptation.

2. **Structural variation**. The impact of structural variants (e.g. deletions and duplications) on trait variation will be assessed. This analysis aims to produce a catalog of structural variation and associated frequency estimates, as well as predicted functional consequences of the variants identified. We also aim to construct a graph genome from a diverse selection of breeds using a combination of long- and short-read sequencing technologies.

3. **Phenotype and trait adaptations**. We aim to identify causal gene variants that underlie adaptive traits. For instance, we are interested in covariation between genotypic variants and agro-pastoral markers to shed light on the genetic basis of adaptation to different environments. We will also investigate adaptation to phenotypic and production traits, such as feather colour and egg shell quality.

4. **Distribution of extant chicken genetic diversity**. Sequence data from such diverse geographical sources permits a detailed investigation of extant chicken diversity with respect to geographical spread. Within this investigation, finer scale analyses of diversity – particularly within the continents of Africa and Asia – are to be conducted.

5. **Evolutionary history of chickens**. The introduction of chickens into Europe (when and how many times) remains unclear. By comprehensively mapping the extant variation in chicken populations, we aim to build a high-quality reference panel for variants, which can be used to phase and impute genomes, including low-coverage ancient genomes from Europe dating to ~2000 years ago (a few centuries following the introduction of chickens in Europe). Using similar approaches, we also plan to decipher the evolutionary history of chickens in Neotropical America, in which chickens underwent a much more recent (~500 years ago) introduction.

6. **Evolutionary history of *Gallus* spp**. Combining data from domesticated chickens and congeneric junglefowl is expected to help answer questions regarding the ultimate origin of domestic chickens and the contribution of junglefowl to modern chicken ancestry.

7. **Evolution and adaptation in the immune system**. A simple prediction is that chickens have had to adapt to cope with (i) exposure to novel pathogens, and (ii) increased intensity of pathogen pressure due to increased flock size and density of rearing. This is likely to have left signatures of adaptation at immune loci of the chicken genome. Genes such as the Toll-like receptors and other pattern-recognition receptors at the front line of defense against pathogens will be investigated for signals of selection. We aim to conduct *in vitro* testing to validate bioinformatic predictions of functional change in immune receptors using methods that are well-established within the consortium.

These lines of investigation will be synthesized into several publications over the course of the consortium, led by different principal investigators depending on expertise. At the outset, the consortium has aimed to be as inclusive as possible, and as such, the studies listed above are neither exhaustive nor limited to current members of the consortium. The consortium welcomes input from any groups wishing to make the best use of this genomic resource.

Processing Pipeline

In order to provide complete consistency of analysis, the entire dataset was re-processed from raw reads using a state-of-the-art mapping and processing pipeline implemented on the SuperMUC computing cluster of the Leibniz-Rechenzentrum, Bavarian Academy of Science, Germany (Fig. 39). All reads underwent pre-processing (quality trimming, adaptor removing) with Fastp (v 0.21.0) then were mapped to the most recent version of the chicken genome (GRCg7b) (GCA_016699485.1) with BWA (v 0.7.17-r1188). The resulting BAM files from the same samples were then merged with samtools (v 1.9). For the variant calling, we generated gvcf using Elprep (v 5.1.1, compiled with go1.17; a reimplementation of GATK in GO langage). These gvcfs were integrated into a GATK genomicDB (gatk v4.2.3.0). To optimize performance, we built 48 databases corresponding to partitioning the genome into 48 intervals of equal size (~20Mb). Variant calling was performed using GATK genotypeGVCFs to obtain a VCF file by interval. We then obtained a global VCF file using GATK GatherVCFs. GATK VariantRecalibrator was then used to recalibrate variants using known SNPs.

Project Timeline

The project began in 2021 and is expected to conclude (at least the first tranche of analyses) in 2023. The first phase of the project (Q3-Q4 2021) involved data gathering from both public and private sources and curation of associated metadata. In Q1 2022, the read files were quality checked to remove low quality samples and to check pre-processing from the variety of sequencing platforms included in the dataset. In Q2 2022, read mapping commenced, soon followed by variant calling. At the time of writing (September 2022), mapping and variant calling have been completed and the VCF will shortly be made available for further analyses.

Data Hosting and Availability

The SuperMUC computing cluster will provide the processing power and storage capability to generate and store raw read files, alignment map files and variant call files (VCF) for the duration of the project. The final VCF will be made available in the first instance to members of the consortium, and will also be provided to the community for wider use. High quality SNPs will be made available to the community on GLOBUS, via sftp, and the European Variation Archive via the European Bioinformatics Institute.

Biobanking and Candidate Genes for Genome Editing to Support the Tropical Poultry Improvement Programmes

(Prepared by C. Keambou Tiambo, C. Kamidi Muhonja, M. Ogugo, O. Ouko, E. Ilatsia, S. Kemp, A. Djikeng, and M. McGrew)

As the world increasingly relies on a handful of chicken breeds in intensified systems, a great diversity of tropical indigenous breeds is kept in backyard systems and on small farms, sometimes undergoing extinction threats through inbreeding and negative selection. If tropically adapted indigenous breeds disappear, the global poultry industry and the science community stand to lose access to adaptive genetic traits that have been nurtured and developed by local communities over millennia. The conservation of animal genetic resources through cryopreservation, referred to as biobanking, is an important component for the conservation and revival of rare or endangered species. While it was not previously possible to completely and efficiently biobank avian genetic material, a scientific innovation using Primordial Germ Cells (PGCs) is changing the landscape and providing a way forward to preserve the biodiversity of tropical poultry breeds, hence also providing sustainable biological material from a more ethical 3Rs - Reduction, Refinement, Replacement - poultry biotechnology research and innovation. The tropical poultry PGC biobanking innovation includes the development and use of chickens that are devoid of their own sperm or ovum as recipients of the biobanked germ cells from a donor. With the introduction of the PGCs into the growing "sterile" embryo, the chick then develops into a surrogate fertile animal, but only produces gametes (ovum or sperm) that are 100% genetically those of the donor chicken breed. A sire-dam crossing of the surrogates, then allows restoration of indigenous chicken breeds from biobanked material in a manner that is animal welfare-friendly. Most importantly, the lab-based techniques developed as part of this innovation are teachable and transferable to partner institutions, thereby enabling countries across the tropics to adopt chicken genetic resource biobanking.

Africa's population is growing and transforming very fast, and will double by 2050, making food security the main challenge for the continent (UNDESA, 2017). The rapid mutations in the demand for animal-source food, derived products and services, coupled with rising concerns of climate change, command a sustainable intensification of livestock production in Africa to accomplish food and nutritional security. Today, Africa must build the foundations to steer livestock on a sustainable development trajectory. This can be a short to medium term perspective if relying on short life cycle and locally adapted genetic resources.

Over 1,600 local chicken breeds have been identified globally (Eda 2021), of which 126 are from Africa (DAGRIS 2021). These breeds contain vast ranges of phenotypic and genetic diversity derived from the varied pathogenic, environmental and selection conditions under which these ecotypes were developed.

Unfortunately, many of these local breeds are considered at risk due to the introduction and adoption of non-local breeds, the acceptance of intensive chicken production systems, changes in environment and disease conditions and adverse development policies.

To minimize this genetic erosion, it is essential to increase knowledge of local breeds and production systems, improve planning and raise awareness of the threat at the policy level. New innovations in genetic preservation technologies for chickens are also needed.

Conservation of poultry breeds and genetic lines pose particular challenges. One means of genetic diversity conservation in livestock species has been biobanking, whereby sperm, eggs or zygotes are preserved for future use. Cryopreservation of eggs and zygotes are not possible in avian species due to the large amount of lipid deposited in the female oocyte (Petitte 2006; Whyte and McGrew 2015). Other genetic preservation and propagation techniques such as cloning using somatic cell nuclear transfer are not possible, as embryo transfer cannot be done in avian species (Kjelland et al. 2014). Recent developments by researchers from the International Livestock Research Institute and the Centre for Tropical Livestock Genetics and Health (CTLGH), at the Roslin Institute, University of Edinburgh in the UK, have shown that the isolation and freezing of PGCs from chicken embryos can provide a new approach for biobanking poultry material. Biobanking of these PGCs will have a significant role in the conservation of African poultry genetic resources; further adding safeguards against the population and diversity losses that could threaten a breed's survival (Hall 2013).

Furthermore, beyond its role as an important source of protein and a valuable model for the study of developmental biology, immunology, physiology and neurology in vertebrates (Yasugi and Nakamura, 2000; Speedy, 2003; Mozdziak and Petitte, 2004; Stern, 2004, 2005), there is an increasing interest in generating genetically modified chickens resistant to specific pathogens, benefiting from the availability of gene manipulation techniques. Hence, the full potential of new chicken breeding tools such as genome editing needs to be exploited in addition to conventional technologies. Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein (CRISPR/Cas)-based genome editing has rapidly become the most prevalent genetic engineering approach for developing improved chicken lines because of its simplicity, efficiency, specificity, and ease of use. Genome editing improves chicken breeds by conferring special attributes including

specific chicken bioreactors, production of knock-in/out chickens for various production and adaptability traits, low-allergenicity eggs, or to serve as disease-resistance models (Chojnacka-Puchta and Sawicka, 2020). In African countries with the most advanced regulations in animal biotechnologies like Kenya and Nigeria, such genome-edited animals with no foreign gene integration are not regulated as genetically modified organisms (GMOs). Researchers from the Centre for Tropical Livestock Genetics and Health (CTLGH) et the Roslin Institute and at the International Livestock Research institute (ILRI) are using CRISPR/Cas-based genome editing for improving chicken productivity and adaptability.

1- *Biobanking Africa Chicken Genetic Resources for the Future*

Primordial Germ Cell Isolation, Conservation, and Recovery for Production of Chimeric Chickens

PGCs are specialized stem cells that can be isolated from chick embryos, which – depending on the sex of the individual embryo – will eventually form a sperm or egg. Following isolation, these PGCs can be cultured and cryopreserved. Chickens are one of the few species from which PGCs can easily be propagated *in vitro* to increase cell numbers up to 100,000 cells from a single embryo in four weeks (Whyte et al. 2015) and these cells can easily be cryopreserved (Glover and McGrew 2012; Glover et al. 2013).

Biobanked chicken material is important in cases where a specific breed might be selected in the future to be scaled up to useful production levels. The genetic material can also be used to support research to isolate certain traits that can then be introduced into existing chicken populations.

When these biobanked chicken genes are needed, the preserved PGCs can be transferred into a two-day-old 'recipient' chick embryo. Part of this PGC biobanking innovation is the development and use of sterile recipient chicks. Because the recipient two-day-old chick embryo is sterile, it lacks its own PGCs, and this eliminates the need to manage the PGCs of the recipient bird. Implanted with biobanked PGCs, the recipient chick grows up into a fertile bird. These adult birds have only had their reproductive cells changed to the genetics of the donor bird and will now act as surrogate parents. The recipient birds will still look like and have the genetic components of their breed, but they will produce sperm or eggs that are genetically identical to the original biobanked donor breed implanted.

The development of the PGC biobanking has revolutionized the ability to preserve tropical chicken genetic resources at ILRI-Nairobi. Table 10 and 11 present the progress in African chicken ecotypes and lines of chicken already cryopreserved.

Along with the active biobanking of African chicken ecotypes and PGC lines, poultry biobanking activities are being scaled up through active knowledge transfer to African scientists in collaboration with the African Union-Inter African Bureau for Animal Resources (AU-IBAR) Training of Trainers (ToT) workshops in Kenya, Cameroon and the Democratic Republic of Congo.

These workshops have led to increased knowledge of the role and importance of locally adapted chicken breeds and fostered greater understanding of the need for indigenous breed conservation. The Kenyan Agricultural and Livestock Research Organization (KALRO) has embraced the technology for the conservation of indigenous chicken ecotypes and plans to use it to sustain the development of the Kinyeji (local) chicken breeding programme. Further collaboration with KALRO is creating opportunities for greater upscaling and uptake by National Research Systems across Africa.

Restoration of Poultry Biodiversity and Dissemination of Potential Elite Lines
Using the Chimeras Conventional Crossbreeding

Germline chimeric chickens were produced at ILRI-Nairobi by transfer of primordial germ cells from indigenous chicken to White Leghorn. An average of 500 primordial germ cells from indigenous chicken were injected into the bloodstream through the dorsal aorta of stage 14-15 White Leghorn recipient embryos which were then incubated until hatching, and the chimera derived offspring were identified based on their feather colour (Fig. 40).

Using the Gene Edited Surrogate Host Technologies
- The DDX4 Knockout (KO) Surrogate Host Technology

The creation of the DDX4 KO line was led by Mike McGrew at the Roslin Institute and originally described in Taylor et al. (2017). Woodcock et al. (2019) demonstrated that the female chicken rendered sterile using genome editing technology can be used as a surrogate host for transplanted cryopreserved germ cells, and only lay eggs of the transplanted rare chicken breed. The DDX4 KO surrogate hosts are genetically sterile female surrogate host

chickens. This sterile female surrogate provides a major advance for the creation of genetically altered chicken lines and the preservation of rare breeds. The DDX4 KO line transmits 100% of its offspring from donor female primordial germ cells (PGCs). Therefore, the DDX4 KO line can produce offspring from genetically altered PGCs or PGCs from other breeds of chicken. Woodcock et al. (2019) have shown that by injecting PGCs from a heritage breed of Vantress chicken into DDX4 KO female embryos, all the offspring produced by these DDX4 females were derived from the donor heritage breed. Subsequent, artificial insemination of the DDX4 KO female surrogate host with frozen Vantress semen produced several pure heritage breed chicks.

- The iCaspase9 Surrogate Host Technology

The creation of the iCaspase9 line was led by Mike McGrew at the Roslin Institute and originally described by Ballantyne et al. (2021). The iCaspase9 surrogate hosts are conditionally sterile male and female surrogate host chickens. In the iCaspase9 surrogate host line the germ cell lineage of both males and females are chemically ablated. These conditionally sterile transgenic chickens provide a major advantage for creating genetically altered (GA) chicken lines. GA primordial germ cells (PGCs) or PGCs from other chicken breeds can be introduced into sterile male and female iCaspase9 embryos. Once the iCaspase9 birds reach sexual maturity, mating between male and female iCaspase9 birds or Sire Dam Surrogate mating results in 100% transmission of the donors' PGCs in the first generation of offspring. The iCaspase9 surrogate hosts cut the time required to create GA chicken lines, and reduce the number of animals required to produce GA chicken lines. The iCaspase9 surrogate host line has been used to create GA chickens with altered feather traits and birds with targeted mutations in the *DMRT1* gene to investigate avian sex determination.

The chicken surrogate host technology has been approved by the National Biosafety Authority of Kenya for the biobanking and revival of the African indigenous poultry biodiversity. It can be used to harness enhanced resilience and productivity, and support future response to new poultry breeding requirements.

2- Primordial Germ Cells (PGCs) for 3Rs Tropical Poultry Research and Gene Editing

PGC technology has brought high hopes for advancing poultry research in breeding, production and veterinary health problems. The stem cell application in restoration and regeneration of tissue, cloning, and transgenic poultry production carries much promise for 1) booster productivity and feed efficiency, 2) disease-free or disease-resistant chickens and 3) production of heat stress resilient chickens. McGrew has stated "*Discovering a way to easily freeze avian reproductive cells and subsequently bring back a genetically diverse flock will help the preservation of endangered breeds of poultry, increase food security from disease outbreaks and reduce numbers of animals used in research.*" The poultry PGC process allows restoration of indigenous chicken breeds from biobanked stem cells in a manner that supports the 3Rs—Reduction, Refinement, Replacement—and is animal welfare-friendly. Most importantly, the lab-based techniques developed as part of this innovation are teachable and transferable to partner institutions, thereby enabling countries across Africa to adopt chicken genetic resource biobanking.

PGCs for Poultry Candidate Gene Screening

Biobanked poultry primordial germ cells (PGCs) as precursors of sperm and egg have the potential to transmit the complete genetic and epigenetic information to the next generation (Matsui et al., 1992; Han, 2009), unlike poultry cryopreserved sperm. PGCs also exhibit unique migration and settling activity which plays a pivotal role in avian genetic resource protection and stem cell research (Burt and Pourquie, 2003; Li et al., 2004; Oishi et al., 2016). According to Taylor et al. (2017), the ability to precisely genetically edit the chicken genome will not only allow the investigation of key developmental signaling pathways in avian species but also the examination of genes involved in egg production, disease susceptibility and resistance with a view to promoting sustainability and biosecurity in both livestock and poultry production (Tizard et al., 2016; Whyte et al., 2016).

Biobanked Chicken PGC lines to support genome editing research and productivity improvement

Chicken is one of the few vertebrate species for which the long-term *in vitro* propagation of primordial germ cells is possible, so performing genetic manipulation of cultured PGCs is becoming a standard practice. as demonstrated in recent studies (Fig. 41A-2). Several researchers have been developing CRISPR/Cas9 and TALENS technologies to investigate multiple genetic variations into pure breeds of chicken (Oishi et al. 2016; Dimitrov et al. 2016). Recently, TALENs were used to target the *DDX4* locus in chicken PGCs (Taylor et al. 2017). *DDX4* is located on the chicken Z sex chromosome and the mRNA is only expressed in the germ cell lineage. Genome editing in chicken is an emerging field and examples of gene editing in bird species other than chicken are currently lacking (Woodcock et al., 2017). According to previous authors, genome editing can provide additional benefit to marker assisted selection in breeding programmes, by either producing novel markers or introducing

new traits to the genome. This is the case for chickens with reduced transmission of avian influenza virus produced by lentiviral transfection of embryos to insert an RNA hairpin molecule into the genome to interfere with viral replication (Lyall et al. 2011). Greater understanding of the pathogenicity of specific diseases could open new avenues for avian disease management, through the application of genome editing. However, most tropical countries still have a long way to go in acquiring and domesticating the genetic manipulation technologies. This is, however, in process through research from the Centre for Tropical Livestock Genetics and Health (CTLGH), and will be facilitated through improvement of poultry PGC line culture and the advances in the adoption of the surrogate host technology at the International Livestock Research Institute (ILRI).

Collectively, rapidly developing genome-editing technology will also accelerate progress in the poultry biotechnology field, opening up new opportunities for poultry to contribute to various industries (Fig. 42)

Candidate Genes for Genome Editing in Tropical Poultry

Poultry performance in tropical regions has been restricted by environmental conditions that cause heat stress and favour the development of parasites and diseases, impairing animal health and productivity. As stated by Professor Appolinaire Djikeng in Poultry World, "Poultry is a key livestock animal for millions of smallholder farmers in low- and middle- income countries. Any gains in efficiency, productivity and health from introducing useful traits from other poultry breeds could significantly improve the lives of these farming families through increased food production and income." (McDougal, 2021).

In a collaborative project between the Roslin Institute and ILRI under the Centre for Tropical Livestock Genetics and Health (CTLGH) and the African Chicken Genetic Gains (ACGG) program, Gheyas et al. (2021) conducted an integrated environmental genome analysis of indigenous chickens from three African countries (Ethiopia, Nigeria and Tanzania) to elucidate the drivers of tropical adaptation in African indigenous chicken populations. The results from the whole-genome sequencing analysis identified some strongly-supported genomic regions under selection for environmental challenges related to altitude, temperature, water scarcity, and food availability. These regions harbour several gene clusters including regulatory genes, suggesting a predominantly oligogenic control of environmental adaptation. Few candidate genes detected in relation to heat-stress, indicates likely epigenetic regulation of thermo-tolerance for a domestic species originating from a tropical Asian wild ancestor. These results provide possible explanations for the rapid past adaptation of chickens to diverse African agro-ecologies, while also representing new landmarks for sustainable breeding improvement for climate resilience. On the other hands, Marchesi et al., (2021) also highlighted relevant candidate genes such as *ATRNL1, PIK3C2A, PTPRN2, SORCS3* and *gga-mir-1759* that could help to elucidate the genetics of feed efficiency traits, hence providing new insights on the mechanisms underlying the consumption and utilization of food in chickens. Harnessing these genes revealed by numerous researchers, and the potential of genome editing could be a powerful tool to use tropical poultry as a key driver for global food security and poverty alleviation.

3- Scaling Up of Poultry Biobanking for Better Livelihoods

The scaling potential of poultry conservation is illustrated by the ongoing close collaboration between CTLGH, the Tropical Poultry Genetic Solutions (TPGS) programme, African Union – InterAfrican Bureau for Animal resources (AU-IBAR) and the African National Agricultural Research Systems (NARS) like the Kenyan Agricultural Research Organization (KALRO). The collaboration is intensifying at various levels on capacity development and effective biobanking across Africa and Southeast Asia. In Kenya, this is mostly supported by the regulatory authorities (the National Biosafety Authority- NBA) that facilitates the adoption of the surrogate host technology for rapid recovery of the biobanked indigenous chicken ecotypes and the intensive dissemination of the elite local and locally adapted and improved breed for the betterment of the livelihood in local communities.

Conclusion

An extensive PGC biobank for indigenous avian breeds will not only support research and development to prevent problems with inbreeding and preserve at-risk poultry breeds, but will also reduce the large number of live animals needed to be kept for research across the world. It could also have an important role within poultry breeding companies to maintain important parental lines of mainstream poultry breeds used in commercial poultry production, without the need to keep large populations of live birds. Having cracked the difficult problem of chicken biobanking, this innovation of PGC preservation, coupled with sterile surrogate use, will revolutionize the preservation and future use of diverse chicken genetics.

Stem cell technologies coupled with genome editing have a prominent role to play in improving poultry production in Africa. There is a need for creating an enabling environment in Africa in general with science-based

regulatory guidelines for the release and adoption of chickens developed using CRISPR/Cas9-mediated genome editing. Some progress has been made in this regard.

Considering the high demand coming from various African national agricultural research systems (NARS), the Centre for Tropical Livestock Genetics and Health (CTLGH) and the Tropical Poultry Genetics Solutions (TPGS) project have projected to scale out activities in Africa in collaboration with the Animal seed centres of Excellence of African Union-Inter African Bureau for Animal Resources (AU-IBAR), and to southeast Asia. Several countries in Central and West Africa have already identified the poultry value chain as their priority for food security and poverty alleviation in rural communities. It is further recommended that the technology be used by the Animal Seed Working group of the African Seed and Biotechnology Partnership Platform; a continental program led by the African Union Commission that frames the development of the seed sector in Africa through improved decision making and policy formulation, supporting evidence-based advocacy, and enhanced knowledge sharing. The critical factors of success of this initiative will be the effective operationalization of the African regional animal seed centres of excellence, continuous investment from national, bilateral and multilateral partners for development of spearheading research, and a harmonized legal framework for access to poultry genetic resources and knowledge sharing between African countries and their scientific collaborators.

Prime beneficiaries of tropical chicken biobanking will be the local National Agricultural Systems (NARS), which will have the capability and technology to cryopreserve all avian genetic resources. Similarly, this avian genetic biobanking could benefit other stakeholders in poultry value chains worldwide. Scientific and research communities, as well as public and private breeding organizations will have access to genetic material from biorepositories across Africa. However, further work is needed to support countries to enforce regulations outlined in the Nagoya Protocol related to the fair and equitable sharing of benefits arising from the utilization of genetic resources.

Application of Genomic Information in Layer Breeding Programs

(Prepared by A. Wolc and J.E. Fulton)

As we are approaching two decades since the first draft sequence of the chicken genome was produced (International Chicken Genome Sequencing Consortium, 2004a), we can reflect on the profound impact of this project on poultry breeding. That reference sequence, plus the simultaneous release of 2.8 million SNP (single nucleotide polymorphisms) identified in multiple chicken breeds (International Chicken Genome Sequencing Consortium, 2004b) provided the essential resources needed to initiate genomic selection in poultry breeding programs. This information has been utilized in multiple ways with the most impactful being the development of high-, medium- and low-density SNP chips that can provide SNP information for hundreds of thousands of SNP that cover the entire genome and individual assays concentrated in specific locations of interest. Combination of low- and high-density SNP panels with imputation (Wang et al 2013) or the use of medium density panels has enabled implementation of genomic selection in poultry breeding (Wolc et al 2016). Genomic selection has been used in layers to increase genetic gain through all parts of the "breeder's equation":

- Increasing accuracy of estimated breeding values for better choice of parents to reproduce the next generation,
- Increasing selection intensity through the ability to hatch more and preselect male candidates using genotype information
- Preserving genetic variation by utilizing Mendelian sampling information captured by an individual's genotypes as opposed to relying purely on family information
- Shortening generation interval by reducing the need to wait for own phenotypes or those of the closest relatives at the point of selection

The increase in accuracy and shortened generation time has been particularly valuable for selection in egg layer programs. The important traits of egg production and shell quality cannot be measured in males, and extended production cycles require 100-plus weeks of trait measurement for the females. The ability to apply genomic

selection for these traits to the males, and to younger females allows the selection program to make large gains. Furthermore, genomic selection can be applied to improvement in many difficult or expensive-to-measure traits including health and resilience under challenging conditions, crossbred performance, and behaviour. While the analysis methods are still evolving, the genomic and phenotyping information is accumulating to provide expanded training sets. Studies report increases of accuracy from addition of genomic information in a range of 20% to over 100% depending on trait, population, training size and methodology (Wolc et al 2011, Alemu et al., 2016, Druet et al 2020). In addition to genomic prediction, genotyping enables parentage verification and assignment (for example in identification of hens laying floor eggs, Wolc et. al 2021) or in the use of cross-classified mating with mixed semen (Hsu et al., 2015, Wolc et al., 2015) and improved product traceability (verification of correctness of line crosses throughout the breeding pyramid). With decreasing costs of genotyping, multiple "Genome Wide Association Studies" (GWAS) have been performed, identifying 16,656 QTL across 370 traits as curated in the www.animalgenome.org database (as of 05/11/22). These studies, combined with additional -omics data are starting to give insights into the complex biology of chickens, their health, response to environment, feed utilization and production traits.

All these developments were made possible by the availability of SNP chips whose contents were based on genomic sequence obtained from a small number of individuals or from DNA pools. These sequence and SNP chips have also been utilized to identify CNV (copy number variants) (Kranis et al, 2013; Boschiero et al, 2015; Drobik-Czwarno et al 2018). This information also has 'non-genomic' uses including line characterization, identification of variation within specific genes of interest, development of SNP sets to identify MHC haplotypes (Fulton et al., 2016), identification of retroviral inserts in the chicken genome (Mason et al., 2020) and their effects on phenotypes (Fulton et al 2021). Until recently, deep sequencing had limited use in applied breeding due to its prohibitive cost. Sequence information from high quality references combined with bioinformatic analysis of conserved genomic regions and homology can also be used for fine mapping of important genes such a blood types (Fulton et al. submitted). Moreover, the development of low-pass sequencing methods with significant cost reduction relative to deep sequencing enables wider use of sequencing for basic and applied projects. Thousands of birds have been sequenced with data used for GWAS with increased resolution (Li et al 2022) which has allowed the exploration of low-pass sequencing to potentially improve the accuracy of genomic prediction (Wolc et al 2022).

The initial chicken reference genome was from a Red Jungle Fowl bird, which is one of the progenitors of the modern chicken. Recently two additional genomes have been produced which better represent the modern chicken; one genome (GRCg7b) was from a commercial meat production bird (broiler) and the other (GRCg7w) was from a commercial egg production bird (White Leghorn) (see article by Warren et al., this report). Additional genome sequences obtained from a wide variety of other chicken breeds (pan-genomes) will provide further insights into variation that exists within the chicken and how this variation can impact health, performance, and behaviour of the birds.

The use of increasingly accurate and affordable genomic data is expected to be critical for breeding modern layers to improve bird welfare, production and optimal use of resources for the sustainable future of egg production.
Conflict of Interest Statement. The authors have no conflicts of interest to declare.

The National Avian Research Facility
(Prepared by L.J. Henderson, A.B. Diack, L. Vervelde, M.P. Stevens, A. Balic, and M.J. McGrew

The National Avian Research Facility (NARF) provides a range of resources and expertise for the avian research community in the UK and internationally. Areas of research supported by the NARF include avian immunology, host-pathogen interactions, physiology and behaviour, developmental biology, and genetics. The NARF was founded in 2013, and is based at The Roslin Institute, on the University of Edinburgh's Easter Bush Campus, UK. The NARF was established via capital support from The Roslin Foundation and the University of Edinburgh, in addition to funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and the Wellcome Trust. The facility continues to be supported by the University of Edinburgh and the BBSRC.
The facility consists of two units; the Greenwood building, a conventional biosecure facility and the Bumstead building that has specified pathogen-free (SPF) status. Both facilities include accommodation for the maintenance and breeding of poultry flocks for research purposes. The NARF is at the forefront of genome engineering technologies in poultry and is one of the few resources globally able to produce genetically altered (GA) chicken

lines under both conventional and SPF conditions. Currently the NARF provides resources and expertise in three main areas; 1) Curation of unique poultry lines: The NARF maintains a wide range of transgenic chicken lines, wild-type layer lines, a broiler line, Japanese quail, and chicken lines with defined genetic characteristics. 2) Genome modification in chicken; the production and maintenance of genome-edited and transgenic chicken lines, and 3) the cryopreservation of research chicken lines, and rare or endangered chicken breeds.

Curator of Research Poultry Flocks
The NARF is home to one of the largest collections of transgenic chicken lines, including multiple ubiquitous and gene-specific fluorescent reporter lines (Davey et al. 2018) and sterile surrogate host chicken lines (Taylor et al. 2017; Ballantyne et al. 2021a). In the Greenwood facility, the NARF also maintains a range of wild-type poultry lines, such as commercially relevant layer lines, a broiler line, and Japanese quail. In 2014, the inbred chicken lines that were maintained at The Pirbright Institute, and formerly held at the Institute for Animal Health (IAH) Compton, were relocated to the Bumstead SPF facility. Nine inbred White Leghorn lines (Lines 6_1, 7_2, 15l, C.B4, C.B12, N, 0, P2a, and W) and two closed outbred chicken lines (Rhode Island Red and Light Sussex) are housed within the Bumstead facility (Kaspers and Schat 2022). These lines have been studied for their susceptibility and resistance to various avian pathogens based on their known MHC I haplotypes (Bacon et al. 2000; Alber et al. 2019; Chintoan-Uta et al. 2020; Bremner et al. 2021; Russell et al. 2021; Mountford et al. 2022). Some of the inbred lines represent a considerable investment by BBSRC and other governmental departments over many years, with some tracing their origins to the late 1920s. Information regarding available resources and poultry lines can be found on the NARF website (www.narf.ac.uk).

Genome Modification in Chickens
The NARF provides world leading resources and expertise in genome modification in chickens. The production of GA birds has been more challenging compared with other model organisms. This was in part due to the complex structure of the avian zygote and the organisation of the avian embryo (Love et al. 1994; Sang 2004). However, methods like lentiviral vectors enabled the creation of a number of transgenic chicken lines that are valuable tools for developmental biology (McGrew et al. 2004; Davey et al. 2018), avian immunology (Balic et al. 2014) and biotechnology (Herron et al. 2018). These methods have some limitations, for example to produce a stable GA chicken line, in which all offspring carry the introduced gene, requires multiple crosses and generations of chickens. This process is time-consuming and requires many animals. In addition to this, the size of the transgene introduced is limited by the vector capacity, and precise modification of genes is not possible with these methods. Refinement of the culture conditions for chicken primordial germ cells (PGCs) (Van De Lavoir et al. 2006; Whyte et al. 2015), and the development of chicken sterile surrogate hosts (Taylor et al. 2017; Ballantyne et al. 2021a), combined with CRISPR/Cas9 methods that enable precise and targeted edits (Ballantyne et al. 2021a; Ioannidis et al. 2021), have improved the efficiency of genome editing in the chicken and consequently the creation of GA chicken lines. PGCs are stem cells that give rise to sperm or eggs and can be harvested from the blood supply of early-stage chicken embryos. PGCs can be grown in culture (Whyte et al. 2015) and modified via transgenesis (Macdonald et al. 2012) or gene-editing (Taylor et al. 2017). These modified PGCs can be re-introduced into the blood supply of a wild-type chicken embryo, where they migrate to the testes or ovaries, and give rise to sperm or eggs once the animal is sexually mature. However, because the introduced PGCs must compete with endogenous PGCs, this results in variable and low rates of chicks produced from the modified PGCs. To avoid this, sterile surrogate host chickens have been developed, where the germ cell lineage (PGCs) of both males and females can be ablated after the introduction of a chemical compound using the iCaspase 9 system (Ballantyne et al. 2021a). When modified PGCs and the chemical compound are microinjected into the blood supply of a developing sterile surrogate host embryo (Ballantyne et al. 2021a), the endogenous PGCs are ablated and only the modified PGCs migrate to the testes or ovaries. Mating of male and female sterile surrogate hosts that have received modified PGCs results in 100% transmission of introduced PGCs in the first generation of offspring (Ballantyne et al. 2021a). Compared to previous methods, this vastly reduces the number of animals used to create a GA chicken line (Panda and McGrew 2022) and reduces the time required to create a GA chicken line from over two years to less than one year.

The NARF has sterile surrogate host chicken lines available for use by researchers, including the transgenic iCaspase9 chicken line that is conditionally sterile (Ballantyne et al. 2021a), and the *DDX4* knockout chicken line, in which all females are genetically sterile (Taylor et al. 2017). Sterile surrogate host chicken lines are also maintained in the SPF Bumstead facility, providing the capability to generate transgenic or gene-edited chickens

under SPF conditions. Recently, the NARF's genome editing technologies have been used to investigate avian sex determination using targeted mutations in the *DMRT1* gene (Ioannidis et al. 2021), and the modification of feather traits (Ballantyne et al. 2021a). Precision editing and transgenesis has been greatly facilitated by advances in the quality of avian reference genomes and functional annotations, including transcriptome atlases spanning cells, tissues and developmental stages and the mapping of regulatory elements. For example, having identified genes and transcripts specific to chicken conventional dendritic cells (cDCs) (Wu et al. 2022), the NARF recently produced a novel fluorescent reporter chicken line that reports on cDCs via *XCR1* gene expression. In addition, using the iCaspase 9 system (Straathof et al. 2005), this line also enables inducible ablation of cDCs. This transgenic chicken line will provide insights into the role of cDCs in natural and vaccine-mediated immunity, which is poorly understood in avian species.

Cryopreservation of Avian Genetic Resources
The NARF is in the process of creating a biobank to maintain avian genetic resources. This resource will be a repository for cryopreserved genetic material from valuable research chicken lines and rare or endangered poultry breeds. A biobank allows for the storage and protection of this genetic diversity, by mitigating against loss of resources in the event of disease outbreak or genetic bottlenecks in closed populations. Furthermore, cryopreserving chicken lines that are not currently being used for research will reduce the number of live animals that need to be maintained annually for research purposes, thereby addressing the principles of Replacement, Refinement and Reduction (The 3Rs) of animal use in research.
Advances in PGC culture and the development of sterile surrogate host chickens supports the NARF's ability to cryopreserve avian germplasm and re-derive cryopreserved lines. In concert with these technological advances, the NARF is in the process of designing robust and time efficient procedures to cryopreserve avian germplasm via PGCs (Nandi et al. 2016; Ballantyne et al. 2021b) or embryonic gonads (Hu et al. 2022). As the biobank develops it will be important to create information sharing systems, to itemise which chicken lines have been cryopreserved, so that information is shared with the scientific community, thus preventing effort and resources being used to replicate GA lines that are already available. This is in line with similar resources that exist for other model organisms, such as mice, for which there are centralised repository systems (www.findmice.org) (Jackson et al. 2021).

Change in Focus
Previously the NARF aimed to collate and share biological tools such as antibodies, primers, and reagents for avian species, and build annotation of, and informatics-based resources for avian genomes. As the NARF's research focus has evolved, these areas are no longer directly supported by the NARF. However, The Roslin Institute continues to support the development of avian antibodies, primers, and reagents through "The Immunological Toolbox" (www.immunologicaltoolbox.co.uk), in addition to a wide array of poultry research. Avian genomics and bio-informatics resources are now supported by projects undertaken by the "Functional Annotation of Animal Genomes" (FAANG) initiative (www.faang.org). Moving forward, NARF intends to build on its own genome engineering technologies, the production of novel GA lines, and the development of an avian biobank. We are keen to hear from researchers in both academic and industry sectors to explore how NARF can further meet the needs of the avian research community.

Conflict of Interest Statement. The authors have no conflicts of interest to declare.

References
Aarts HJ, Van Der Hulst-Van Arkel MC, Beuving G, Leenstra FR. Variations in endogenous viral gene patterns in White Leghorn, medium heavy, White Plymouth Rock, and Cornish Chickens. Poult Sci. 1991;70(6):1281–6.
Abdelnour SA, Abd El-Hack ME, Khafaga AF, Arif M, Taha AE, Noreldin AE. Stress biomarkers and proteomics alteration to thermal stress in ruminants: A review. J Therm Biol. 2019;79:120–34.
Abraham KJ, Khosraviani N, Chan JNY, Gorthi A, Samman A, Zhao DY, et al. Nucleolar RNA polymerase II drives ribosome biogenesis. Nature. 2020;585:298–302.

Adkins HB, Brojatsch J, Young JA. Identification and characterization of a shared TNFR-related receptor for subgroup B, D, and E avian leukosis viruses reveal cysteine residues required specifically for subgroup E viral entry. J Virol. 2000;74(8):3572–8.

Adkins HB, Blacklow SC, Young JA. Two functionally distinct forms of a retroviral receptor explain the nonreciprocal receptor interference among subgroups B, D, and E avian leukosis viruses. J Virol. 2001;75(8):3520–6.

Aerts J, Megens HJ, Veenendaal T, Ovcharenko I, Crooijmans R, Gordon L, et al. Extent of linkage disequilibrium in chicken. Cytogenet Genome Res. 2007;117(1–4):338–345.

Afrache H, Tregaskes CA, Kaufman J. A potential nomenclature for the Immuno Polymorphism Database (IPD) of chicken MHC genes: progress and problems. Immunogenetics. 2020;72(1-2):9–24.

Agranat-Tamir L, Shomron N, Sperling J, Sperling R. Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome. Nucleic Acids Res. 2014;42(7):4640–51.

Agrawal S, Ganley ARD. The conservation landscape of the human ribosomal RNA gene repeats. PLoS One. 2018;13(12):e0207531.

Ahn D, You KH, Kim CH. Evolution of the tbx6/16 subfamily genes in vertebrates: insights from zebrafish. Mol Biol Evol. 2012;29(12):3959–83.

Alber A, Costa T, Chintoan-Uta C, Bryson KJ, Kaiser P, Stevens MP, et al. Dose-dependent differential resistance of inbred chicken lines to avian pathogenic Escherichia coli challenge. Avian Pathol. 2019;48(2):157–67.

Alemu SW, Calus MPL, Muir WM, Peeters K, Vereijken A, Bijma P. Genomic prediction of survival time in a population of brown laying hens showing cannibalistic behavior. Genet Sel Evol. 2016;48(1):68.

Alfieri JM, Wang G, Jonika MM, Gill CA, Blackmon H, Athrey GN. A primer for single-cell sequencing in non-model organisms. Genes (Basel). 2022;13(2):380.

Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 2011;477:587–91.

Almeida OAC, Moreira GCM, Rezende FM, Boschiero C, de Oliveira Peixoto J, Ibelli AMG, et al. Identification of selection signatures involved in performance traits in a paternal broiler line. BMC Genomics. 2019;20(1):449.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

Anderegg MA, Gyimesi G, Ho TM, Hediger MA, Fuster DG. The less well-known little brothers: The SLC9B/NHA sodium proton exchanger subfamily-structure, function, regulation and potential drug-target approaches. Front Physiol. 2022;13:898508.

Anderson LK, Reeves A, Webb LM, Ashley T. Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. Genetics 1999;151:1569–79.

Anderson LK, Salameh N, Bass HW, Harper LC, Cande WZ, Weber G, Stack SM. Integrating genetic linkage maps with pachytene chromosome structure in maize. Genetics. 2004;166:1923–33.

Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol. 2015;16(1):57.

Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arad G, Bar-Meir R, Kotler M. Ribosomal frameshifting at the Gag-Pol junction in avian leukemia sarcoma virus forms a novel cleavage site. FEBS Lett. 1995;364(1):1–4.

Arensburger P, Piégu B, Bigot Y. The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine? Mob Genet Elements. 2016;6:e1256852.

Armand EJ, Li J, Xie F, Luo C, Mukamel EA. Single-cell sequencing of brain cell transcriptomes and epigenomes. Neuron. 2021;109(1):11–26.

Asadollahpour Nanaei H, Kharrati-Koopaee H, Esmailizadeh A. Genetic diversity and signatures of selection for heat tolerance and immune response in Iranian native chickens. BMC Genomics. 2022;23(1):224. Erratum in: BMC Genomics. 2022;23(1):395.

Aswad A, Katzourakis A. Paleovirology and virally derived immunity. Trends Ecol Evol. 2012;27(11):627–36.

Audas TE, Jacob MD, Lee S. Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. Mol Cell. 2012;45:147–57.

Avdeyev P, Jiang S, Aganezov S, Hu F, Alekseyev MA. Reconstruction of ancestral genomes in presence of gene gain and loss. J Comput Biol. 2016;23(3):150–64.

Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH, et al. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. Genome Biol. 2013;14:1–17.

Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. 2013;4:113–24.

Bachtrog D. The Y chromosome as a battleground for intragenomic conflict. Trends Genet. 2020;36:510–22.

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, et al. Are all sex chromosomes created equal? Trends Genet. 2011;27:350–7.

Bachtrog D, Mahajan S, Bracewell R. Massive gene amplification on a recently formed Drosophila Y chromosome. Nat Ecol Evol. 2019;3:1587–97.

Backström N, Ceplitis H, Berlin S, Ellegren H. Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. Mol Biol Evol. 2005;22:1992–1999.

Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, et al.: The recombination landscape of the zebra finch Taeniopygia guttata genome. Genome Res 2010;20:485–95.

Bacon LD, Smith E, Crittenden LB, Havenstein GB. Association of the slow feathering (K) and an endogenous viral (ev21) gene on the Z chromosome of chickens. Poult Sci. 1988;67(2):191–197.

Bacon LD, Hunt HD, Cheng HH. A review of the development of chicken lines to resolve genes determining resistance to diseases. Poult Sci. 2000;79(8):1082–1093.

Badenhorst D, Hillier LW, Literman R, Montiel EE, Radhakrishnan S, Shen Y, et al. Physical mapping and refinement of the painted turtle genome (Chrysemys picta) inform amniote genome evolution and challenge turtle-bird chromosomal conservation. Genome Biol Evol. 2015;7:2038–50.

Bai H, Sun Y, Liu N, Liu Y, Xue F, Li Y, et al. Genome-wide detection of CNVs associated with beak deformity in chickens using high-density 600K SNP arrays. Anim Genet. 2018;49(3):226–236.

Balic A, Garcia-Morales C, Vervelde L, Gilhooley H, Sherman A, Garceau V, et al. Visualisation of chicken macrophages using transgenic reporter genes: Insights into the development of the avian macrophage lineage. Development. 2014;141(16):3255–3265.

Ballantyne M, Woodcock M, Doddamani D, Hu T, Taylor L, Hawken RJ, et al. Direct allele introgression into pure chicken breeds using Sire Dam Surrogate (SDS) mating. Nat Commun. 2021a;12(1):659.

Ballantyne M, Taylor L, Hu T, Meunier D, Nandi S, Sherman A, et al. Avian primordial germ cells are bipotent for male or female gametogenesis. Front Cell Dev Biol. 2021b;9:2596.

Ballingall KT, Bontrop RE, Ellis SA, Grimholt U, Hammond JA, Ho CS, et al. Comparative MHC nomenclature: report from the ISAG/IUIS-VIC committee 2018. Immunogenetics. 2018;70(10):625–32.

Baron MG, Norman DB, Barrett PM. A new hypothesis of dinosaur relationships and early dinosaur evolution. Nature. 2017;543:501–506.

Barral A, Déjardin J. Telomeric Chromatin and TERRA. J Mol Biol. 2020;432(15):4244–56.

Barrowclough GF, Cracraft J, Klicka J, Zink RM. How many kinds of birds are there and why does it matter? PLoS One 2016;11:e0166307.

Battagin M, Gorjanc G, Faux AM, Johnston SE, Hickey JM. Effect of manipulating recombination rates on response to selection in livestock breeding programs. Genet Sel Evol 2016;48:44.

Beauclair L, Rame C, Arensburger P, Piegu B, Guillou F, Dupont J, et al. Sequence properties of certain GC-rich avian genes, their origins and absence from genome assemblies: case studies. BMC Genomics. 2019;20(1):734.

Beçak W, Beçak ML, Nazareth HRS, Ohno S. Close karyological kinship between the reptilian suborder serpentes and the class aves. Chromosoma. 1964;15:606–17.

Bell RA, Al-Khalaf M, Megeney LA. The beneficial role of proteolysis in skeletal muscle growth and stress adaptation. Skelet Muscle. 2016;6:16.

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F. Impacts of climate change on the future of biodiversity. Ecol Lett. 2012;15:365–77.

Bellott DW, Page DC. Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. Genome Res. 2021;31:198–210.

Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, et al. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature. 2010;466:612–6.

Bellott DW, Skaletsky H, Cho TJ, Brown L, Locke D, Chen N, et al. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. Nat Genet 2017;49:387–94.

Benachenhou F, Sperber GO, Bongcam-Rudloff E, Andersson G, Boeke JD, Blomberg J. Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). Mob DNA. 2013;4(1):5.

Benkel BF. Locus-specific diagnostic tests for endogenous avian leukosis-type viral loci in chickens. Poult Sci. 1998;77(7):1027–35.

Benkel B, Rutherford K. Endogenous avian leukosis viral loci in the Red Jungle Fowl genome assembly. Poult Sci. 2014;93(12):2988–90.

Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.

Benson SJ, Ruis BL, Fadly AM, Conklin KF. The unique envelope gene of the subgroup J avian leukosis virus derives from ev/J proviruses, a novel family of avian endogenous viruses. J Virol. 1998;72(12):10157–64.

Benton MJ, Forth J, Langer MC. Models for the rise of the dinosaurs. Curr Biol 2014;24:R87–95.

Benton MJ, Donoghue PCJ, Asher RJ, Friedman M, Near TJ, Vinther J. Constraints on the timescale of animal evolutionary history. Palaeontol Electron 2015;18:1–106.

Berchowitz LE, Copenhaver GP: Genetic interference: don't stand so close to me. Curr Genomics 2010;11:91.

Bernardi G. The "Genomic Code": DNA pervasively moulds chromatin structures leaving no room for "junk". Life (Basel). 2021;11(4):342.

Bertzbach LD, Tregaskes CA, Martin RJ, Deumer US, Huynh L, Kheimar AM, et al. The diverse major histocompatibility complex haplotypes of a common commercial chicken line and their effect on Marek's disease virus pathogenesis and tumorigenesis. Front Immunol. 2022;13:908305.

Berv JS, Field DJ. Genomic signature of an avian lilliput effect across the K-Pg extinction. Syst Biol 2018;67:1–13.

Bikle DD, Munson S, Komuves L. Zipper protein, a B-G protein with the ability to regulate actin/myosin 1 interactions in the intestinal brush border. J Biol Chem. 1996;271(15):9075–83.

Bishara A, Liu Y, Weng Z, Kashef-Haghighi D, Newburger DE, West R, et al. Read clouds uncover variation in complex regions of the human genome. Genome Res. 2015;25:1570–80.

Blary A, Jenczewski E: Manipulation of crossover frequency and distribution for plant breeding. Theor Appl Genet 2019;132:575–92.

Blaschke UK, Hedbom E, Bruckner P. Distinct isoforms of chicken decorin contain either one or two dermatan sulfate chains. J Biol Chem. 1996;271(48):30347–53.

Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y. The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol. 2006;7(5):R43.

Bloom SE, Bacon LD. Linkage of the major histocompatibility (B) complex and the nucleolar organiser in the chicken. Assignment to a microchromosome. J Hered. 1985;76:146–54.

Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet. 2012;13:770–80.

Bohler MW, Chowdhury VS, Cline MA, Gilbert ER. Heat stress responses in birds: a review of the neural components. Biology (Basel). 2021;10(11):1095.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

Bolisetty M, Blomberg J, Benachenhou F, Sperber G, Beemon K. Unexpected diversity and expression of avian endogenous retroviruses. mBio. 2012;3(5):e00344–12.

Bornelov S, Seroussi E, Yosefi S, Pendavis K, Burgess SC, Grabherr M, et al. Correspondence on Lovell et al.: identification of chicken genes previously assumed to be evolutionarily lost. Genome Biol. 2017;18(1):112.

Boschiero C, Gheyas AA, Ralph HK, Eory L, Paton B, Kuo R, et al. Detection and characterization of small insertion and deletion genetic variants in modern layer chicken genomes. BMC Genomics. 2015;16:562.

Boschiero C, Moreira GCM, Gheyas AA, Godoy TF, Gasparin G, Mariani PDSC, et al. Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. BMC Genomics. 2018;19(1):83.

Botero-Castro F, Figuet E, Tilak MK, Nabholz B, Galtier N. Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. Mol Biol Evol. 2017;34(12):3123–31.

Boulliou A, Le Pennec JP, Hubert G, Donal R, Smiley M. Restriction fragment length polymorphism analysis of endogenous avian leukosis viral loci: determination of frequencies in commercial broiler lines. Poult Sci. 1991;70(6):1287–96.

Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet. 2018;50(3):362–7.

Bradford YM, Van Slyke CE, Ruzicka L, Singer A, Eagle A, Fashena D, et al. Zebrafish information network, the knowledgebase for Danio rerio research. Genetics. 2022;220(4):iyac016.

Bravo GA, Schmitt CJ, Edwards SV. What have we learned from the first 500 avian genomes. Ann Rev Ecol Evol Syst 2021;52:611–39.

Bredemeyer KR, Harris AJ, Li G, Zhao L, Foley NM, Roelke-Parker M, et al. Ultracontinuous single haplotype genome assemblies for the domestic cat (Felis catus) and Asian leopard cat (Prionailurus bengalensis). J Hered. 2021;112:165–73.

Brekke C, Groeneveld LF, Meuwissen THE, Sæther N, Weigend S, Berg P. Assessing the genetic diversity conserved in the Norwegian live poultry genebank. Acta Agric Scand A Anim Sci. 2020;69(1–2):68–80.

Bremner A, Kim S, Morris K, Nolan MJ, Borowska D, Wu Z, et al. Kinetics of the immune response to Eimeria maxima in relatively resistant and susceptible chicken lines. Front Immunol. 2021;12:653085.

Briles WE, Briles RW, Taffs RE, Stone HA. Resistance to a malignant lymphoma in chickens is mapped to subregion of major histocompatibility (B) complex. Science. 1983;219(4587):977–9.

Briles WE, Goto RM, Auffray C, Miller MM. A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. Immunogenetics. 1993;37(6):408–14.

Bromham L. The human zoo: endogenous retroviruses in the human genome. Trends Ecol Evol. 2002;17(2):91–7.

Brown EJ, Nguyen AH, Bachtrog D. The Y chromosome may contribute to sex-specific ageing in Drosophila. Nat Ecol Evol. 2020;4:853–62.

Burt DW. Origin and evolution of avian microchromosomes. Cytogenet Genome Res. 2002;96:97–112.

Burt DW. Emergence of the chicken as a model organism: implications for agriculture and biology. Poult Sci. 2007;86(7):1460–71.

Burt D, Pourquie O. Chicken genome–science nuggets to come soon. Science 2003;300:1669.

Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, et al. The dynamics of chromosome evolution in birds and mammals. Nature. 1999;402:411–3.

Burt DW, Carrë W, Fell M, Law AS, Antin PB, Maglott DR, et al. The chicken gene nomenclature committee report. BMC Genomics. 2009;10(Suppl 2):S5.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25(18):1915–27.

Calderón PL, Pigozzi MI: MLH1-focus mapping in birds shows equal recombination between sexes and diversity of crossover patterns. Chromosome Res 2006;14:605–612.

Cao W, Mays J, Kulkarni G, Dunn J, Fulton RM, Fadly A. Further observations on serotype 2 Marek's disease virus-induced enhancement of spontaneous avian leukosis virus-like bursal lymphomas in ALVA6 transgenic chickens. Avian Pathol. 2015;44(1):23–27.

Caudy AA, Pikaard CS. Xenopus ribosomal RNA gene intergenic spacer elements conferring transcriptional enhancement and nucleolar dominance-like competition in oocytes. J Biol Chem. 2002;277:31577–84.

Cendron F, Mastrangelo S, Tolone M, Perini F, Lasagna E, Cassandro M. Genome-wide analysis reveals the patterns of genetic diversity and population structure of 8 Italian local chicken breeds. Poult Sci. 2021;100(2):441–451.

Ceplitis H, Ellegren H. Adaptive molecular evolution of HINTW, a female-specific gene in birds. Mol Biol Evol. 2004;21:249–54.

Chak LL, Mohammed J, Lai EC, Tucker-Kellogg G, Okamura K. A deeply conserved, noncanonical miRNA hosted by ribosomal DNA. RNA. 2015;21(3):375–384.

Chang CM, Coville JL, Coquerelle G, Gourichon D, Oulmouden A, Tixier-Boichard M. Complete association between a retroviral insertion in the tyrosinase gene and the recessive white mutation in chickens. BMC Genomics. 2006;7:19.

Chang KW, Tseng YT, Chen YC, Yu CY, Liao HF, Chen YC, et al. Stage-dependent piRNAs in chicken implicated roles in modulating male germ cell development. BMC Genomics. 19:425 (2018).

Chappell P, Meziane el K, Harrison M, Magiera Ł, Hermann C, Mears L, et al. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. Elife. 2015;4:e05345.

Charlesworth B. The evolution of sex chromosomes. Science. 1991;251:1030–3.

Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. Heredity 2005;95:118–28.

Chattaway J, Ramirez-Valdez RA, Chappell PE, Caesar JJ, Lea SM, Kaufman J. Different modes of variation for each BG lineage suggest different functions. Open Biol. 2016;6(9):160188.

Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell. 2022;185(10):1777-1792.e21.

Chen L, Yang W, Guo Y, Chen W, Zheng P, Zeng J, et al. Exosomal lncRNA GAS5 regulates the apoptosis of macrophages and vascular endothelial cells in atherosclerosis. PLoS One. 2017;12(9):e0185406.

Chen L, Fakiola M, Staines K, Butter C, Kaufman J. Functional alleles of chicken BG genes, members of the butyrophilin gene family, in peripheral T cells. Front Immunol. 2018;9:930.

Chen L, Wang X, Cheng D, Chen K, Fan Y, Wu G, et al. Population genetic analyses of seven Chinese indigenous chicken breeds in a context of global breeds. Anim Genet. 2019;50(1):82–86.

Chen S, Hu X, Cui IH, Wu S, Dou C, Liu Y et al. An endogenous retroviral element exerts an antiviral innate immune function via the derived lncRNA lnc-ALVE1-AS1. Antivir Res. 2019;170:104571.

Chen X, Bai X, Liu H, Zhao B, Yan Z, Hou Y, et al. Population genomic sequencing delineates global landscape of copy number variations that drive domestication and breed formation of in chicken. Front Genet. 2022;13:830393.

Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. PLoS One. 2013;8(4):e62856.

Chen Y, Arsenault R, Napper S, Griebel P. Models and Methods to Investigate Acute Stress Responses in Cattle. Animals (Basel) 2015;5:1268–95.

Cheng Y, Burt DW. Chicken genomics. Int J Dev Biol. 2018;62:265–71.

Chesters PM, Howes K, McKay JC, Payne LN, Venugopal K. Acutely transforming avian leukosis virus subgroup J strain 966: defective genome encodes a 72-kilodalton Gag-Myc fusion protein. J Virol. 2001;75(9):4219–25.

Chintoan-Uta C, Wisedchanwet T, Glendinning L, Bremner A, Psifidi A, Vervelde L, et al. Role of caecal microbiota in the differential resistance of inbred chicken lines to colonization by Campylobacter jejuni. Appl Environ Microbiol. 2020;86(7):e02607–19.

Cho NH, Cheveralls KC, Brunner AD, Kim K, Michaelis AC, Raghavan P, et al. OpenCell: Endogenous tagging for the cartography of human cellular organization. Science. 2022;375(6585):eabi6983.

Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, et al. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol. 2010;11(7):R72.

Choi SW, Kim HW, Nam JW. The small peptide world in long noncoding RNAs. Brief Bioinform. 2019;20(5):1853–64.

Chojnacka-Puchta L, Sawicka D. CRISPR/Cas9 gene editing in a chicken model: current approaches and applications. J Appl Genetics. 2020;61:221–9.

Chow W, Brugger K, Caccamo M, Sealy I, Torrance J, Howe K. gEVAL - a web-based browser for evaluating genome assemblies. Bioinformatics. 2016;32:2508–10.

Christidis L. Animal Cytogenetics 4: Chordata 3 B: Aves. Berlin: Gebrüder Borntraeger; 1990.

Cīrulis A, Hansson B, Abbott JK. Sex-limited chromosomes and non-reproductive traits. BMC Biology 2022;20:156.

Clinton M, Haines L, Belloir B, McBride D. Sexing chick embryos: a rapid and simple protocol. Br Poult Sci. 2001;42(1):134-138.

Cocquet J, Ellis PJ, Mahadevaiah SK, Affara NA, Vaiman D, Burgoyne PS. A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. PLoS Genet. 2012;8(9): e1002900.

Colquitt BM, Merullo DP, Konopka G, Roberts TF, Brainard MS. Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. Science. 2021;371(6530):eabd9704.

Connallon T, Beasley IJ, McDonough Y, Ruzicka F. How much does the unguarded X contribute to sex differences in life span? Evol Lett. 2022;6(4):319–29.

Cosset FL, Lavillette D. Cell entry of enveloped viruses. Adv Genet. 2011;73:121–183.

Costantini M, Musto H. The Isochores as a Fundamental Level of Genome Structure and Organization: A General Overview. J Mol Evol. 2017;84(2-3):93-103.

Cracraft J, Houde P, Ho SYW, Mindell DP, Fjeldså J, Lindow B, et al. Response to Comment on "Whole-genome analyses resolve early branches in the tree of life of modern birds". Science. 2015;349:1460.

Crawford RD. Poultry Breeding and Genetics. Amsterdam, New York: Elsevier; 1990.

Crittenden LB, Smith EJ, Fadly AM. Influence of endogenous viral (ev) gene expression and strain of exogenous avian leukosis virus (ALV) on mortality and ALV infection and shedding in chickens. Avian Dis. 1984;28(4):1037–1056.

Cui J, Zhao W, Huang Z, Jarvis ED, Gilbert MT, Walker PJ, et al. Low frequency of paleoviral infiltration across the avian phylogeny. Genome Biol. 2014;15:539.

Dakovic N, Terezol M, Pitel F, Maillard V, Elis S, Leroux S, et al. The loss of adipokine genes in the chicken genome and implications for insulin metabolism. Mol Biol Evol. 2014;31(10):2637-46.

Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg Le A, et al. Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis. PLoS Biol. 2010;8(9).

Damas J, O'Connor R, Farré M, Lenis VPE, Martell HJ, Mandawala A, et al. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. Genome Res. 2017;27:875–84.

Damas J, Kim J, Farré M, Griffin DK, Larkin DM. Reconstruction of avian ancestral karyotypes reveals differences in the evolutionary history of macro- and microchromosomes. Genome Biol. 2018;19:155.

Daniels LM, Delany ME. Molecular and cytogenetic organization of the 5S ribosomal DNA in chicken (Gallus gallus). Chromosome Res. 2003;11:305–317.

Dapper AL, Payseur BA: Connecting theory and data to understand recombination rate evolution. Philos Trans R Soc B Biol Sci 2017;372:20160469.

Davey MG, Balic A, Rainger J, Sang HM, McGrew MJ. Illuminating the chicken model through genetic modification. Int J Dev Biol. 2018;62(1-2-3):257–264.

Davidian AG, Dyomin AG, Galkina SA, Makarova NE, Dmitriev SE, Gaginskaya ER. 45S rDNA repeats of turtles and crocodiles harbour a functional 5S rRNA gene specifically expressed in oocytes. Mol Biol Evol. 2022;39(1):msab324.

Davis JK, Thomas PJ, Thomas JW. A W-linked palindrome and gene conversion in New World sparrows and blackbirds. Chromosome Res. 2010;18:543–53.

Deist MS, Lamont SJ. What makes the harderian gland transcriptome different from other chicken immune tissues? A gene expression comparative analysis. Front Physiol. 2018;9:492.

Deist MS, Gallardo RA, Bunn DA, Dekkers JCM, Zhou HJ, Lamont SJ. Resistant and susceptible chicken lines show distinctive responses to Newcastle disease virus infection in the lung transcriptome. BMC Genomics 2017;18(1):989.

Deist, M.S., Gallardo, R.A., Bunn, D.A., Kelly, T.R., Dekkers, J.C.M., Zhou, H.J. and Lamont, S.J., 2018. Novel analysis of the Harderian gland transcriptome response to Newcastle disease virus in two inbred chicken lines. Sci Rep. 2018;8(1):6558.

Delany ME. Patterns of ribosomal gene variation in elite commercial chicken pure line populations. Anim Genet. 2000;31:110–116.

Delany ME, Krupkin AB. Molecular characterization of ribosomal gene variation within and among NORs segregating in specialized populations of chicken. Genome. 1999;42:60–71.

Delany ME, Robinson CM, Goto RM, Miller MM. Architecture and organization of chicken microchromosome 16: order of the NOR, MHC-Y, and MHC-B subregions. J Hered. 2009;100(5):507-14.

De Lima JE, Blavet C, Bonnin MA, Hirsinger E, Comai G, Yvernogeau L, et al. Unexpected contribution of fibroblasts to muscle lineage as a mechanism for limb muscle patterning. Nat. Commun. 2021;12(1):3851.

del Priore L, Pigozzi MI: Sex-specific recombination maps for individual macrochromosomes in the Japanese quail (Coturnix japonica). Chromosome Res 2015;23:199–210.

del Priore L, Pigozzi MI. MLH1 focus mapping in the guinea fowl (Numida meleagris) give insights into the crossover landscapes in birds. PLoS One. (2020) 15:e0240245 doi:10.1371/journal.pone.0240245

del Priore L, Pigozzi MI. DNA Organization along Pachytene Chromosome Axes and Its Relationship with Crossover Frequencies. Int J Mol Sci. 2021;22(5):2414.

Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. A new understanding of the decoding principle on the ribosome. Nature. 2012;484:256–259.

Deng X, Li S, Kong F, Ruan H, Xu X, Zhang X, et al. Long noncoding RNA PiHL regulates p53 protein stability through GRWD1/RPL11/MDM2 axis in colorectal cancer. Theranostics. 2020;10(1):265–80.

Derks MFL, Megens HJ, Bosse M, Visscher J, Peeters K, Bink MCAM, et al. A survey of functional genomic variation in domesticated chickens. Genet Sel Evol. 2018;50(1):17.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775–89.

Dierickx EG, Sin SYW, van Veelen HPJ, Brooke MdL, Liu Y, Edwards SV, et al. Genetic diversity, demographic history and neo-sex chromosomes in the critically endangered Raso lark. Proc Biol Sci. 2020;287(1922):20192613.

Dimitrov L, Pedersen D, Ching KH, Yi H, Collarini EJ, Izquierdo S, et al. Germline Gene Editing in Chickens by Efficient CRISPR-Mediated Homologous Recombination in Primordial Germ Cells. PLoS ONE 2016;11(4):e0154303.

Ding Q, Li R, Ren X, Chan LY, Ho VWS, Xie D, et al. Genomic architecture of 5S rDNA cluster and its variations within and between species. BMC Genomics. 2022;23(1):238.

Djumagulov M, Demeshkina N, Jenner L, Rozov A, Yusupov M, Yusupova G. Accuracy mechanism of eukaryotic ribosome translocation. Nature. 2021;600:543–546.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 2013;29: 15-21.

Dokan K, Kawamura S, Teshima KM. Effects of single nucleotide polymorphism ascertainment on population structure inferences. G3 (Bethesda). 2021;11(9):jkab128.

Domestic Animal Genetic Resources Information System (DAGRIS). http://dagris.ilri.cgiar.org/species. (Accessed 21 August 2022).

Dong X, Zhao P, Xu B, Fan J, Meng F, Sun P et al. Avian leukosis virus in indigenous chicken breeds. Emerg Microbes Infect. 2015;4(1):1-2.

Drobik-Czwarno W, Wolc A, Fulton JE, Dekkers JCM. Detection of copy number variations in brown and white layers based on genotyping panels with different densities. Genet Sel Evol. 2018;50(1):54.

Druet DP, Varenne A, Herry F, Hérault F, Allais S, Burlot T, et al. Reliability of genomic evaluation for egg quality traits in layers. BMC Genet. 21:17 (2020).

Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, et al. Apollo: Democratizing genome annotation. PLoS Comput Biol. 2019;15(2):e1006790.

Dyomin AG, Koshel EI, Kiselev AM, Saifitdinova AF, Galkina SA, Fukagawa T, et al. Chicken rRNA gene cluster structure. PLoS One. 2016;11:e0157464.

Dyomin A, Volodkina V, Koshel E, Galkina S, Saifitdinova A, Gaginskaya E. Evolution of ribosomal internal transcribed spacers in Deuterostomia. Mol Phylogenet Evol. 2017;116:87–96.

Dyomin A, Galkina S, Fillon V, Cauet S, Lopez-Roques C, Rodde N, et al. Structure of the intergenic spacers in chicken ribosomal DNA. Genet Sel Evol. 2019;51:59.

Eda M. Origin of the domestic chicken from modern biological and zooarchaeological approaches. Anim Front. 2021;11:52–61.

Eding H, Crooijmans RPMA, Groenen MAM, Meuwissen THE. Assessing the contribution of breeds to genetic diversity in conservation schemes. Genet Sel Evol. 2002;34(5):613–33.

Elferink MG, Vallée AAA, Jungerius AP, Crooijmans RPMA, Groenen MAM. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. BMC Genomics. 2008;9:391.

Elferink MG, Megens HJ, Vereijken A, Hu X, Crooijmans RPMA, Groenen MA. Signatures of selection in the genomes of commercial and noncommercial chicken breeds. PLoS One. 2012;7(2):e32720.

Elleder D, Kaspers B. After TNF-alpha, still playing hide-and-seek with chicken genes. Poult Sci. 2019;98(10):4373-4.

Ellegren H: Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol. 2010;25:283–291.

Ellis PJ, Bacon J, Affara NA. Association of Sly with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids? Hum Mol Genet. 2011;20:3010-3021.

EMBL EBI's Ensembl. Gallus_gallus - Ensembl genome browser 107 [Internet]. 2022. Available from: http://uswest.ensembl.org/Gallus_gallus/Info/Annotation

Eraslan G, Drokhlyansky E, Anand S, Fiskin E, Subramanian A, Slyper M, et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. Science. 2022;376(6594):eabl4290.

Erickson JM, Schmickel RD. A Molecular Basis for Discrete Size Variation in Human Ribosomal DNA. Am J Hum Genet. 1985;37:311-325.

Estermann MA, Williams S, Hirst CE, Roly ZY, Serralbo O, Adhikari D, et al. Insights into Gonadal Sex Differentiation Provided by Single-Cell Transcriptomics in the Chicken Embryo. Cell Rep. 2020;31(1):107491.

Ewald SJ, Livant EJ. Distinctive polymorphism of chicken B-FI (major histocompatibility complex class I) molecules. Poult Sci. 2004 Apr;83(4):600-5.

Eyre, Tina A., Mathew W. Wright, Michael J. Lush, and Elspeth A. Bruford. 2007. "HCOP: A Searchable Database of Human Orthology Predictions." Brief Bioinform. 8 (1): 2–5.

Fan, Y., Arbab, A.A.I., Zhang, H., Yang, Y., Nazar, M., Han, Z. and Yang, Z., 2021. Lactation Associated Genes Revealed in Holstein Dairy Cows by Weighted Gene Co-Expression Network Analysis (WGCNA). Animals (Basel) 11.

Farhadian, M., Rafat, S.A., Panahi, B. and Mayack, C., 2021. Author Correction: Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. Sci Rep 11, 18245.

Farré M, Narayan J, Slavov GT, Damas J, Auvil L, Li C, Jarvis ED, et al (2016) Novel Insights into Chromosome Evolution in Birds, Archosaurs, and Reptiles. Genome Biol Evol 8:2442–2451.

Federico C, Cantarella CD, Scavo C, Saccone S, Bed'Hom B, et al. Avian genomes: different karyotypes but a similar distribution of the GC-richest chromosome regions at interphase. Chromosome Res. 2005; 13:785-793.

Feregrino C, Sacher F, Parnas O, Tschopp P. A single-cell transcriptomic atlas of the developing chicken limb. BMC Genom. 2019;20(1):401.

Fernandes AC, da Silva VH, Goes CP, Moreira GCM, Godoy TF, Ibelli AMG, et al. Genome-wide detection of CNVs and their association with performance traits in broilers. BMC Genomics. 2021;22(1):354.

Fernandes JB, Duhamel M, Seguéla-Arnaud M, Froger N, Girard C, Choinard S, et al.: FIGL1 and its novel partner FLIP form a conserved complex that regulates homologous recombination. PLoS Genet 2018;14:e1007317.

Fleming DS, Koltes JE, Markey AD, Schmidt CJ, Ashwell CM, Rothschild MF, et al. Genomic analysis of Ugandan and Rwandan chicken ecotypes using a 600 k genotyping array. BMC Genomics. 2016;17:407.

Flores-Juárez CR, González-Jasso E, Antaramian A, Pless RC. PCR amplification of GC-rich DNA regions using the nucleotide analog N4-methyl-2'-deoxycytidine 5'-triphosphate. Biotechniques. 2016; 61:175-182.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020; 117:9451-9457.

Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. BMC Biol. 2019;17(1):108.

Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, et al. 2020. Complete Vertebrate Mitogenomes Reveal Widespread Gene Duplications and Repeats. bioRxiv. https://doi.org/10.1101/2020.06.30.177956

Fortriede JD, Pells TJ, Chu S, Chaturvedi P, Wang D, Fisher ME, et al. 2020. Xenbase: Deep Integration of GEO & SRA RNA-Seq and ChIP-Seq Data in a Model Organism Database. Nucleic Acids Res. 48 (D1): D776–82.

Fouchécourt S, Fillon V, Marrauld C, Callot C, Ronsin S, et al. Expanding duplication of the testis PHD Finger Protein 7 (PHF7) gene 1 in the chicken genome. Genomics. 2022; 114:110411.

Fox W, Smyth JR Jr. The effects of recessive white and dominant white genotypes on early growth rate. Poult Sci. 1985;64(3):429–433.

Francisco FO, Lemos B. 2014. How do Y-chromosomes modulate genome-wide epigenetic states: genome folding, chromatin sinks, and gene expression. J Genomics. 2:94.

Fulton JE. Genomic selection for poultry breeding. Anim Front. 2012;2:30–36.

Fulton JE, Juul-Madsen HR, Ashwell CM, McCarron AM, Arthur JA, O'Sullivan NP, Taylor RL Jr. Molecular genotype identification of the Gallus gallus major histocompatibility complex. Immunogenetics. 2006 Jun;58(5-6):407-21.

Fulton J.E., McCarron A.R., Lund A.R., Pinegar K., Wolc A., Chazara O., Bed'Hom B., Berres M. and Miller M.M. A high density SNP panel reveals extensive diversity, frequent recombination and multiple recombination hotspots within the chicken major histocompatibility complex B region between BG2 to CD1A1. Genetics Selection Evolution 2016. 48:1.

Fulton J.E., Mason A.S., Wolc A., Arango J., Settar P., Lund A.R., Burt D.W. The impact of endogenous Avian Leukosis Viruses (ALVE) on production traits in elite layer lines. Poult Sci. 2021. 100(6):101121.

Furman BL, Metzger DC, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank JE. Sex chromosome evolution: so many exceptions to the rules. Genome Biol Evol 2020. 12:750-763.

Galbraith JD, Kortschak RD, Suh A, Adelson DL. Genome stability is in the eye of the beholder: CR1 retrotransposon activity varies significantly across avian diversity. Genome Biol Evol. 2021; 13:evab259.

Gan HM, Falk S, Morales HE, Austin CM, Sunnucks P, Pavlova A. 2019. Genomic evidence of neo-sex chromosomes in the eastern yellow robin. GigaScience 8.

Gandhi, S., M. L. Piacentino, F. M. Vieceli, and M. E. Bronner. Optimization of CRISPR/Cas9 genome editing for loss-of-function in the early chick embryo. Dev Biol, 2017. 432: 86-97.

Gandhi S, Hutchins EJ, Maruszko K, Park JH, Thomson M, Bronner ME. Bimodal function of chromatin remodeler Hmga1 in neural crest induction and Wnt-dependent emigration. eLife. 2020;9:e57779.

Gao B, Wang S, Wang Y, Shen D, Xue S, et al. Low diversity, activity, and density of transposable elements in five avian genomes. Funct Integr Genomics. 2017; 17:427-439.

Gao J, Xu W, Zeng T, Tian Y, Wu C, Liu S, et al. Genome-wide association study of egg-laying traits and egg quality in LingKun chickens. Front Vet Sci. 2022;9:877739. doi:10.3389/fvets.2022.877739

Garrison, E., Marh, G. 2017. Haplotype-based variant detection from short-read sequencing. arXiv, 1207:3907v2.

Gavora JS, Kuhnlein U, Crittenden LB, Spencer JL, Sabour MP. Endogenous viral genes: association with reduced egg production rate and egg size in White Leghorns. Poult Sci. 1991;70(3):618–623.

Geibel J, Reimer C, Weigend S, Weigend A, Pook T, Simianer H. How array design creates SNP ascertainment bias. PLoS One. 2021;16(3):e0245178.

Geibel J, Praefke NP, Weigend S, Simianer H, Reimer C. Assessment of linkage disequilibrium patterns between structural variants and single nucleotide polymorphisms in three commercial chicken populations. BMC Genomics. 2022;23(1):193.

Gheyas AA, Burt DW. Microarray resources for genetic and genomic studies in chicken: a review. Genesis. 2013;51(5):337–356.

Gheyas AA, Vallejo-Trujillo A, Kebede A, Lozano-Jaramillo M, Dessie T, Smith J, et al. Integrated environmental and genomic analysis reveals the drivers of local adaptation in African indigenous chickens. Mol Biol Evol. 2021;38(10):4268–4285.

Gheyas A, Vallejo-Trujillo A, Kebede A, Dessie T, Hanotte O, Smith J. Whole genome sequences of 234 indigenous African chickens from Ethiopia. Sci Data. 2022;9(1):53.

Gholami M, Erbe M, Gärke C, Preisinger R, Weigend A, Simianer H. Population genomic analyses based on 1 million SNPs in commercial egg layers. PLoS One. 2014;9(4):e94509.

Ghurye, J., A. Rhie, B. P. Walenz, A. Schmitt, S. Selvaraj, M. Pop, A. M. Phillippy, and S. Koren. 2019. 'Integrating Hi-C links with assembly graphs for chromosome-scale assembly', PLoS Comput Biol, 15: e1007273.

Gibbons HR, Shaginurova G, Kim LC, Chapman N, Spurlock CF, Aune TM. Divergent lncRNA GATA3-AS1 Regulates GATA3 Transcription in T-Helper 2 Cells. Front Immunol. 2018;9:2512.

Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. Nat Rev Genet. 2020;21(2):102–17.

Girard C, Chelysheva L, Choinard S, Froger N, Macaisne N, Lehmemdi A, et al. AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms. PLoS Genet 2015;11:e1005369.

Glover JD, McGrew MJ. Primordial germ cell technologies for avian germplasm cryopreservation and investigating germ cell development. Journal of Poultry Science 2012; 49 (3):155–162.

Glover JD, Taylor L, Sherman A, Zeiger-Poli C, Sang HM, and McGrew MJ. A novel piggybac transposon inducible expression system identifies a role for Akt signaling in primordial germ cell migration. PLoS One 2013;8(11):e77222.

Goel, M., H. Sun, W. B. Jiao, K. Schneeberger. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20: 277.

Gonen S, Battagin M, Johnston SE, Gorjanc G, Hickey JM: The potential of shifting recombination hotspots to increase genetic gain in livestock breeding. Genet Sel Evol 2017;49:55.

Gonzalez IL, Sylvester JE. Complete Sequence of the 43-Kb Human Ribosomal DNA Repeat—Analysis of the Intergenic Spacer. Genomics. 1995;27:320–328.

Gorla E, Cozzi MC, Román-Ponce SI, Ruiz López FJR, Vega-Murillo VE, Cerolini S, et al. Genomic variability in Mexican chicken population using copy number variants. BMC Genetics. 2017;18(1):61.

Gorlov IP, Gorlova OY: Cost-benefit analysis of recombination and its application for understanding of chiasma interference. J Theor Biol 2001;213:1–8.

Goto RM, Wang Y, Taylor RL Jr, Wakenell PS, Hosomichi K, Shiina T, et al. BG1 has a major role in MHC-linked resistance to malignant lymphoma in the chicken. Proc Natl Acad Sci U S A. 2009;106(39):16740-5.

Goto RM, Warden CD, Shiina T, Hosomichi K, Zhang J, Kang TH, et al. The Gallus gallus RJF reference genome reveals an MHCY haplotype organized in gene blocks that contain 107 loci including 45 specialized, polymorphic MHC class I loci, 41 C-type lectin-like loci, and other loci amid hundreds of transposable elements. G3 (Bethesda). 2022 Aug 23:jkac218.

Grandori C, Gomez-Roman N, Felton-Edkins ZA, Ngouenet C, Galloway DA, Eisenman RN. c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. Nat Cell Biol. 2005;7(3):311–318.

Gray S, Cohen PE: Control of meiotic crossovers: from double-strand break formation to designation. Annu Rev Genet 2016;50:175–210.

Groenen MAM, Cheng HH, Bumstead N, Benkel BF, Briles WE, Burke T, et al. A consensus linkage map of the chicken genome. Genome Res. 2000;10(1):137–147.

Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens HJ, Crooijmans RPMA, et al.: A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res 2009;19:510–519.

Groenen MAM, Megens HJ, Zare Y, Warren WC, Miller, LDW, Crooijmans RPMA, et al. The development and characterization of a 60K SNP chip for chicken. BMC Genomics. 2011;12(1):274.

Grote P, Wittler L, Währisch S, Hendrix D, Beisaw A, Macura K, et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev Cell. 2013 Jan 28;24(2):206–14.

Grozdanov P, Georgiev O, Karagyozov L. Complete sequence of the 45-kb mouse ribosomal DNA repeat: analysis of the intergenic spacer. Genomics. 2003;82(6):637–643.

Grunder AA, Benkel BF, Chambers JR, Sabour MP, Gavora JS. Characterization of Eight Endogenous Viral (ev) Genes of Meat Chickens in Semi-Congenic Lines. Poult Sci. 1995;74(9):1506-1514.

Guan D, S. A. McCarthy, J. Wood, K. Howe, Y. Wang, and R. Durbin. 'Identifying and removing haplotypic duplication in primary genome assemblies', Bioinformatics, 2020. 36: 2896-98.

Guan D, Halstead MM, Islas-Trejo AD, Goszczynski DE, Ross P, Zhou H. A comprehensive prediction of transcript isoforms in 19 chicken tissues by Oxford Nanopore long-read sequencing. 2022 Jul 19;21.

Gudkov AV, Komarova EA, Nikiforov MA, Zaitsevskaya TE. ART-CH, a new chicken retroviruslike element. J Virol. 1992;66(3):1726–1736.

Guh CY, Hsieh YH, Chu HP. Functions and properties of nuclear lncRNAs—from systematically mapping the interactomes of lncRNAs. J Biomed Sci. 2020 Mar 17;27:44.

Guiblet WM, Cremona MA, Harris RS, Chen D, Eckert KA, et al. Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. Nucleic Acids Res. 2021; 49:1497-1516.

Guilbaud G, Murat P, Recolin B, Campbell BC, Maiter A, et al. Local epigenetic reprogramming induced by G-quadruplex ligands. Nat Chem. 2017; 9:1110-1117.

Guillot C, Djeffal Y, Michaut A, Rabe B, Pourquié O. Dynamics of primitive streak regression controls the fate of neuromesodermal progenitors in the chicken embryo. eLife. 2021;10:e64819.

Guizard S, Piégu B, Arensburger P, Guillou F, Bigot Y. Deep landscape update of dispersed and tandem repeats in the genome model of the red jungle fowl, Gallus gallus, using a series of de novo investigating tools. BMC Genomics. 2016; 17:659.

Guo C, Gong M, Li Z. Knockdown of lncRNA MCM3AP-AS1 Attenuates Chemoresistance of Burkitt Lymphoma to Doxorubicin Treatment via Targeting the miR-15a/EIF4E Axis. Cancer Manag Res. 2020 Jul 16;12:5845–55.

Guo J, Fang W, Sun L, Lu Y, Dou L, Huang X, et al. Ultraconserved element uc.372 drives hepatic lipid accumulation by suppressing miR-195/miR4668 maturation. Nat Commun. 2018 Feb 9;9(1):612.

Guo S, Zhang C, Le A. The limitless applications of single-cell metabolomics. Curr Opin Biotechnol. 2021;71:115-122.

Guo X, Fang Q, Ma C, Zhou B, Wan Y, Jiang R. Whole-genome resequencing of Xishuangbanna fighting chicken to identify signatures of selection. Genet Sel Evol. 2016;48(1):62.

Guo X et al. Genetic signature of hybridization between Chinese spot-billed ducks and domesticated ducks Anim Genet (2020) 51:866-875.

Guo X, Xing CH, Wei W, Zhang XF, Wei ZY, Ren LL, et al. Genome-wide scan for selection signatures and genes related to heat tolerance in domestic chickens in the tropical and temperate regions in Asia. Poult Sci. 2022;101(7):101821.

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

Guo Y, Ou JH, Zan Y, Wang Y, Li H, Zhu C, et al. Researching on the fine structure and admixture of the worldwide chicken population reveal connections between populations and important events in breeding history. Evol Appl. 2021;15(4):553–564.

Habimana R, Ngeno K, Okeno TO, Hirwa CA, Keambou Tiambo C, Yao NK. Genome-wide association study of growth performance and immune response to Newcastle disease virus of indigenous chicken in Rwanda. Front Genet. 2021;12:723980.

Haldane JB. 1922. Sex ratio and unisexual sterility in hybrid animals. J Genet 12:101-109.

Hall HC, Ostrowski LA, Pietrobon V, Mekhail K. Repetitive DNA loci and their modulation by the non-canonical nucleic acid structures R-loops and G-quadruplexes. Nucleus. 2017; 8:162-181.

Hall SJG. Defra Project Report GC0146: Development of coordinated in situ and ex situ UK Farm Animal Genetic Resources conservation strategy and implementation guidance. 2013; http://eprints.lincoln.ac.uk/id/eprint/22909/1/GC0146%20Conservation%20Strategies%2024%20April%202013-pdf.pdf

Hall TM. Multiple modes of RNA recognition by zinc finger proteins. Curr Opin Struct Biol. 2005;15(3):367–373.

Haltiner MM, Smale ST, Tjian R. Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. Mol Cell Biol. 1986;6:227-235.

Hamburg S. Call to join the decentralized science movement. Nature. 2021; 600(7888):221.

Hamburger V, Hamilton HL. A series of normal stages in the development of the chick embryo. J. Morphol. 1951; 88(1), 49–92.

Han JY. Germ cells and transgenesis in chickens. Comp. Immunol. Microbiol. Infect. Dis. 2009; 32, 61–80.

Haniffa M, Taylor, D, Linnarsson S, Aronow BJ, Bader GD, Barker RA, et al. A roadmap for the Human Developmental Cell Atlas. Nature 2021; 597(7875):196–205.

Hansen C, Yi N, Zhang YM, Xu S, Gavora J, Cheng HH (2005) Identification of QTL for production traits in chickens Animal biotechnology 16:67-79.

Hasan Siddiqui, S., Kang, D., Park, J., Choi, H.W. and Shim, K., 2020. Acute Heat Stress Induces the Differential Expression of Heat Shock Proteins in Different Sections of the Small Intestine of Chickens Based on Exposure Duration. Animals (Basel) 10.

He C, Zhao L, Xiao L, Xu K, Ding J, Zhou H, et al. Chromosome level assembly reveals a unique immune gene organization and signatures of evolution in the common pheasant. Mol Ecol Resour. 2021 Apr;21(3):897-911.

Hedges SB, Marin J, Suleski M, et al (2015) Tree of life reveals clock-like speciation and diversification. Mol Biol Evol 32:835–45.

Henry J, Miller MM, Pontarotti P. Structure and evolution of the extended B7 family. Immunol Today. 1999 Jun;20(6):285-8.

Herrmann T, Karunakaran MM. Butyrophilins: γδ T Cell Receptor Ligands, Immunomodulators and More. Front Immunol. 2022 Mar 17;13:876493.

Herron LR, Pridans C, Turnbull ML, Smith N, Lillico S, Sherman A, et al. A chicken bioreactor for efficient production of functional cytokines. BMC Biotechnol. 2018;18(1):82.

Herry F, Hérault F, Picard Druet D, Varenne A, Burlot T, Le Roy P, et al. Design of low-density SNP chips for genotype imputation in layer chicken. BMC Genetics. 2018;19(1):108.

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004 Dec;432(7018):695–716.

Hirst CE, Major AT, Ayers KL, Brown RJ, Mariette M, Sackton TB, et al. 2017. Sex reversal and comparative data undermine the W chromosome and support Z-linked DMRT1 as the regulator of gonadal sex differentiation in birds. Endocrinology 158:2970-2987.

Hoang T, Wang J, Boyd P, Wang F, Santiago C, Jiang L, et al. Gene regulatory networks controlling vertebrate retinal regeneration. Science. 2020;370(6519):eabb8598.

Honda, B.T., Calefi, A.S., Costola-de-Souza, C., Quinteiro-Filho, W.M., da Silva Fonseca, J.G., de Paula, V.F. and Palermo-Neto, J., 2015. Effects of heat stress on peripheral T and B lymphocyte profiles and IgG and IgM serum levels in broiler chickens vaccinated for Newcastle disease virus. Poult Sci 94, 2375-81.

Hori T, Asakawa S, Itoh Y, Shimizu N, Mizuno S. 2000. Wpkci, encoding an altered form of PKCI, is conserved widely on the avian W chromosome and expressed in early female embryos: implication of its role in female sex determination. Molecular Biology of the Cell 11:3645-3660.

Hori Yu, Shimamoto A, Kobayashi T. The human ribosomal DNA array is composed of highly homogenized tandem clusters. Genome Res. 2021;31:1971–1982.

Horvath, S. and Dong, J., 2008. Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol 4, e1000117.

Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, et al. Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. J Immunol. 2008 Sep 1;181(5):3393-9. doi: 10.4049/jimmunol.181.5.3393. Erratum in: J Immunol. 2010 May 1;184(9):5415. PMID: 18714011.

Houstek, J., Hejzlarova, K., Vrbacky, M., Drahota, Z., Landa, V., Zidek, V., et al. 2012. Nonsynonymous variants in mt-Nd2, mt-Nd4, and mt-Nd5 are linked to effects on oxidative phosphorylation and insulin sensitivity in rat conplastic strains. Physiol Genomics 44, 487-94.

Howe, K., W. Chow, J. Collins, S. Pelan, D. L. Pointon, Y. Sims, J. Torrance, A. Tracey, and J. Wood. 2021. 'Significantly improving the quality of genome assemblies through curation', Gigascience, 10.

Hoyt SJ, Storer JM, Hartley GA, Grady PGS, Gershman A, de Lima LG, et al. From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. Science. 2022;376(6588):eabk3112.

Hron T, Pajer P, Paces J, Bartunek P, Elleder D. Hidden genes in birds. Genome Biol. 2015;16:164.

Hsu W.-L., Fernando R.L., Dekkers J.C.M., Arango J., Settar P., Fulton J.E., et al. 2015. A simulation study on the effect of nested vs factorial mating on response to pedigree and genomic selection. J. Anim. Sci. Vol. 93, Suppl. s3: 242.

Hu T, Taylor L, Sherman A, Tiambo CK, Kemp SJ, Whitelaw BW, et al. Direct cryopreservation of poultry/avian embryonic reproductive cells: A low-tech, cost-effective and efficient method for safeguarding genetic diversity. eLife. 2022;11:e74036.

Hu X, Zhu W, Chen S, Liu Y, Sun Z, Geng T et al. Expression of the env gene from the avian endogenous retrovirus ALVE and regulation by miR-155. Arch Virol. 2016;161(6):1623–1632.

Hu X, Zhu W, Chen S, Liu Y, Sun Z, Geng T et al. Expression patterns of endogenous avian retrovirus ALVE1 and its response to infection with exogenous avian tumour viruses. Arch Virol. 2017;162(1):89–101.

Huang, D.W., Sherman, B.T. and Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4, 44-57.

Huang S, He Y, Ye S, Wang J, Yuan X, Zhang H, et al. Genome-wide association study on chicken carcass traits using sequence data imputed from SNP array. J Appl Genet. 2018;59(3):335–344.

Huang Z, De OFI, Liu J, Peona V, Gomes AJB, Cen W, Huang H, Zhang Y, Chen D, Xue T, Zhang Q, Yue Z, Wang Q, Yu L, Chen Y, Suh A, de Oliveira EHC, Xu L: Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. Nat Commun 13:944 (2022).

HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. Nature. 2019;574(7777):187-192.

Hughes AL, Friedman R. Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. Mol Biol Evol. 2008;25(12):2681-8.

Hughes AL, Hughes MK. Small genomes for better flyers. Nature. 1995;377(6548):391.

Hughes AL, Piontkivska H. DNA repeat arrays in chicken and human genomes and the adaptive evolution of avian genome size. BMC Evol Biol. 2005;5:12.

Hughes JF, Skaletsky H, Koutseva N, Pyntikova T, Page DC. 2015. Sex chromosome-to-autosome transposition events counter Y-chromosome gene loss in mammals. Genome Biology 16:104.

Hunt H, Fadly A, Silva R, Zhang H. Survey of endogenous virus and TVB* receptor status of commercial chicken stocks supplying specific-pathogen-free eggs. Avian Diseases. 2008;52(3):433–440.

Hurst TP, Magiorkinis G. Activation of the innate immune response by endogenous retroviruses. J Gen Virol. 2005;96(6):1207–1218.

International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004;432:695-716.

International Chicken Polymorphism Map Consortium. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature. 2004;432(7018):717–722. Doi:10.1038/nature03156

Ioannidis J, Taylor G, Zhao D, Liu L, Idoko-Akoh A, Gong D, et al. Primary sex determination in birds depends on DMRT1 dosage, but gonadal sex does not determine adult secondary sex characteristics. Proc Natl Acad Sci. 2021;118(10):e2020909118.

Iraqi F, Soller M, Beckmann JS. Distribution of Endogenous Viruses in Some Commercial Chicken Layer Populations. Poult Sci. 1991;70(4):665–679

Jackson I, Doe B, Forty E, Fray M, Hart-Johnson S, Moncaut N, et al. Sharing and archiving of genetically altered mice. Opportunities for reduction and refinement. NC3Rs. 2021. https://www.nc3rs.org.uk/3rs-resources/breeding-and-colony-management/sharing-and-archiving-ga-mice.

Janesick A., Scheibinger M., Benkafadar N., Kirti S., Ellwanger D.C., Heller S. Cell-type identity of the avian cochlea. Cell Rep. 2021;34(12):108900.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 2014;346(6215):1320-31.

Jastrebski, Sara F., Susan J. Lamont, and Carl J. Schmidt. 2017. "Chicken Hepatic Response to Chronic Heat Stress Using Integrated Transcriptome and Metabolome Analysis." PLoS ONE 12 (7): e0181900.

Jehl F, Muret K, Bernard M, Boutin M, Lagoutte L, Désert C, et al. An integrative atlas of chicken long non-coding genes and their annotations across 25 tissues. Scientific Reports. 2020 Nov 24;10(1):20457.

Jehl F, Degalez F, Bernard M, Lecerf F, Lagoutte L, Désert C, et al. RNA-Seq Data for Reliable SNP Detection and Genotype Calling: Interest for Coding Variant Characterization and Cis-Regulation Analysis by Allele-Specific Expression in Livestock Species. Front Genet. 2021;12:1104.

Ji Y, DeWoody JA. Genomic landscape of long terminal repeat retrotransposons (LTR-RTs) and solo LTRs as shaped by ectopic recombination in chicken and zebra finch. J Mol Evol. 2016; 82:251-263.

Jia X, Chen S, Zhou H, Li D, Liu W, Yang N. Copy number variations identified in the chicken using a 60K SNP BeadChip. Anim Genet. 2013;44(3):276–284.

Jin CF, Chen YJ, Yang ZQ, Shi K, Chen CK. A genome-wide association study of growth trait-related single nucleotide polymorphisms in Chinese Yancheng chickens. Genet Mol Res. 2015;14(4):15783–15792.

Joseph S, O'Connor RE, Al Mutery AF, et al (2018) Chromosome Level Genome Assembly and Comparative Genomics between Three Falcon Species Reveals an Unusual Pattern of Genome Organisation. Diversity 10:113.

Justice J IV, Beemon KL. Avian retroviral replication. Curr Opin Virol. 2013;3(6):664–669.

Kajino T, Shimamura T, Gong S, Yanagisawa K, Ida L, Nakatochi M, et al. Divergent lncRNA MYMLR regulates MYC by eliciting DNA looping and promoter-enhancer interaction. EMBO J. 2019;38(17):e98441.

Kanda RK, Tristem M, Coulson T. Exploring the effects of immunity and life history on the dynamics of an endogenous retrovirus. Philos Trans R Soc B: Biol Sci. 2013;368(1626):20120505.

Kapusta A, Suh A. Evolution of bird genomes-a transposon's-eye view. Ann N Y Acad Sci. 2017;1389(1):164–185.

Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci U S A. 2017; 114:E1460-E1469.

Kasai F, O'Brien PCM, Martin S, Ferguson-Smith MA (2012) Extensive homology of chicken macrochromosomes in the karyotypes of Trachemys scripta elegans and Crocodylus niloticus revealed by chromosome painting despite long divergence times. Cytogenet Genome Res 136:303–7.

Kaspers B, Schat KA. Appendix 1 - Genetic stocks for immunological research. In: Kaspers B, Schat KA, Göbel TW, Vervelde L, editors. Avian Immunology (Third Edition). Boston: Academic Press. 2022;573–581. https://www.sciencedirect.com/science/article/pii/B9780128187081000166.

Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. PLoS Genet. 2010; 6:e1001191.

Kaufman, J. (2016) "Genetics and genomic organisation of the major histocompatibility complex (MHC)". In: Encyclopedia of Immunobiology, vol. 2, pp 166-173. M. J. H. Ratcliffe, editor in chief. Academic Press, Oxford.

Kaufman J. Generalists and Specialists: A New View of How MHC Class I Molecules Fight Infectious Pathogens. Trends Immunol. 2018 May;39(5):367-379. doi: 10.1016/j.it.2018.01.001. PMID: 29396014.

Kaufman J. The Avian MHC (Chapter 7). In: Avian Immunology, third edition, 2021; K. A. Schat, L. Vervelde and B. Kaspers, editors. Elsevier, Ltd.

Kaufman J, Jacob J, Shaw I, Walker B, Milne S, Beck S, et al. Gene organisation determines evolution of function in the chicken MHC. Immunol Rev. 1999 Feb;167:101-17.

Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, et al. The chicken B locus is a minimal essential major histocompatibility complex. Nature. 1999 Oct 28;401(6756):923-5.

Kawakami T, Smeds L, Backström N, et al (2014) A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol Ecol 23:4035–4058.

Keith G, Nys Y, Fix C, Heyman T. Accumulation and specific cleavage of 5S RNA in the isthmus of laying hen oviduct. Evidence for three chicken 5S RNA. Biochem Biophys Res Commun. 1986;138:1405–1410.

Kennedy GY, Vevers HG. A survey of avian eggshell pigments. Comp Biochem Physiol B. 1976;55(1):117-23.

Kern C, Wang Y, Chitwood J, Korf I, Delany M, Cheng H, et al. Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. BMC Genomics. 2018 Sep 18;19(1):684.

Kim DW, Place E, Chinnaiya K, Manning E, Sun C, Dai W, et al. Single-cell analysis of early chick hypothalamic development reveals that hypothalamic cells are induced from prethalamic-like progenitors. Cell Rep. 2022;38(3):110251.

Kim JH, Dilthey AT, Nagaraja R, Lee HS, Koren S, Dudekula D, et al. Variation in human chromosome 21 ribosomal RNA genes characterized by TAR cloning and long-read sequencing. Nucleic Acids Res. 2018;46:6712–25.

Kim JH, Noskov VN, Ogurtsov AY, Nagaraja R, Petrov N, Liskovykh M, et al. The genomic structure of a human chromosome 22 nucleolar organizer region determined by TAR cloning. Sci Rep. 2021;11(1):2997.

Kim T, Hunt HD, Parcells MS, van Santen V, Ewald SJ. Two class I genes of the chicken MHC have different functions: BF1 is recognized by NK cells while BF2 is recognized by CTLs. Immunogenetics. 2018 Sep;70(9):599-611.

Kjelland ME, Romo S and Kraemer D. Avian Cloning: Adaptation of a Technique for Enucleation of the Avian Ovum. Avian Biology Research 2014; 7 (3): 131-138.

Klucking S, Young JAT. Amino acid residues Tyr-67, Asn-72, and Asp-73 of the TVB receptor are important for subgroup E avian sarcoma and leukosis virus interaction. Virology. 2004;318(1):371–380.

Kogelman, L.J., Cirera, S., Zhernakova, D.V., Fredholm, M., Franke, L. and Kadarmideen, H.N., 2014. Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. BMC Med Genomics 7, 57.

Kojima KK. Structural and sequence diversity of eukaryotic transposable elements. Genes Genet Syst. 2020; 94:233-252.

Komissarov AS, Galkina SA, Koshel EI, Kulak MM, Dyomin AG, et al. New high copy tandem repeat in the content of the chicken W chromosome. Chromosoma. 2018; 127:73-83.

Koonin EV, Krupovic M. Polintons, virophages and transpovirons: a tangled web linking viruses, transposons and immunity. Curr Opin Virol. 2017; 25:7-15.

Korlach, J., G. Gedman, S. B. Kingan, C. S. Chin, J. T. Howard, J. N. Audet, et al. 2017. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. Gigascience, 6: 1-16.

Kosinska-Selbi B, Mielczarek M, Szyda J. Review: Long non-coding RNA in livestock. Animal. 2020 Oct;14(10):2003–13.

Kramer, A., Green, J., Pollard, J., Jr. and Tugendreich, S., 2014. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 30, 523-30.

Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics. 2013;14:59. doi:10.1186/147121641459

Kretschmer R, Gunski RJ, Garnero AdV, de Freitas TRO, Toma GA, Cioffi MdB, et al. 2020. Chromosomal analysis in Crotophaga ani (Aves, Cuculiformes) reveals extensive genomic reorganization and an unusual Z-autosome Robertsonian translocation. Cells 10:4.

Krol A, Gallinaro H, Lazar E, Jacob M, Branlant C. The nuclear 5S RNAs from chicken, rat and man. US RNAs are encoded by multiple genes. Nucleic Acids Res. 1981;9:769-787.

Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Brief Bioinform. 2013 Mar;14(2):144–61.

Kuhnlein U, Sabour M, Gavora JS, Fairfull RW, Bernon DE. Influence of selection for egg production and Marek's disease resistance on the incidence of endogenous viral genes in White Leghorns. Poult Sci. 1989;68(9):1161–1167.

Kumar, R., Kirubaharan, J.J., Chandran, N.D. and Gnanapriya, N., 2013. Transcriptional response of chicken embryo cells to Newcastle disease virus (D58 strain) infection. Indian J Virol 24, 278-83.

Kuo RI, Tseng E, Eory L, Paton IR, Archibald AL, Burt DW. Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. BMC Genomics. 2017;18(1):323.

Kuo RI, Cheng Y, Zhang R, Brown JWS, Smith J, Archibald AL, et al. Illuminating the dark side of the human transcriptome with long read transcript sequencing. BMC Genomics. 2020;21(1):751.

Kusumi, Kenro, Rob J Kulathinal, Arhat Abzhanov, Stephane Boissinot, Nicholas G Crawford, Brant C Faircloth, Travis C Glenn, et al. 2011. Developing a Community-Based Genetic Nomenclature for Anole Lizards. BMC Genomics 12: 554.

Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. Bioessays. 2013;35(9):780-786.

Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, Davis C, et al. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. Nat Genet. 2017 Dec;49(12):1731–40.

Lagarrigue S, Lorthiois M, Degalez F, Gilot D, Derrien T. LncRNAs in domesticated animals: from dog to livestock species. Mamm Genome [Internet]. 2021 Nov 13; Available from: https://doi.org/10.1007/s00335-021-09928-7

Lake JA, Dekkers JCM, Abasht B (2021) Genetic basis and identification of candidate genes for wooden breast and white striping in commercial broiler chickens. Scientific Reports 11:6785

Lam, E. T., A. Hastie, C. Lin, D. Ehrlich, S. K. Das, M. D. Austin, et al. 2012. 'Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly', Nat Biotechnol, 30: 771-6.

Lamia, K.A., Papp, S.J., Yu, R.T., Barish, G.D., Uhlenhaut, N.H., Jonker, J.W., et al. 2011. Cryptochromes mediate rhythmic repression of the glucocorticoid receptor. Nature 480, 552-6.

Langfelder, P. and Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Langfelder, P. and Horvath, S., 2012. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw 46.

Lardelli M. The evolutionary relationships of zebrafish genes tbx6, tbx16/spadetail and mga. Dev Genes Evol. 2003;213(10):519-22.

Larkin DM, Wind AE der, Rebeiz M, et al (2003) A Cattle–Human Comparative Map Built with Cattle BAC-Ends and Human Genome Sequence. Genome Res 13:1966–1972.

Larkin DM, Pape G, Donthu R, et al (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res 19:770–777.

Larson EL, Kopania EE, Good JM. 2018. Spermatogenesis and the evolution of mammalian sex chromosomes. Trends in Genetics 34:722-732.

Laun Katja, Penny Coggill, Sophie Palmer, Sarah Sims, Zemin Ning, Jiannis Ragoussis, et al. 2006. "The Leukocyte Receptor Complex in Chicken Is Characterized by Massive Expansion and Diversification of Immunoglobulin-Like Loci." PLoS Genetics 2 (5): e73.

Lawal RA, Hanotte O. Domestic chicken diversity: Origin, distribution, and adaptation. Anim Genet. 2021;52(4):385–394.

Lawal RA, Al-Atiyat RM, Aljumaah RS, Silva P, Mwacharo JM, Hanotte O. Whole-genome resequencing of red junglefowl and indigenous village chicken reveal new insights on the genome dynamics of the species. Front Genet. 2018;9:264.

Lawal RA, Martin SH, Vanmechelen K, Vereijken A, Silva P, Al-Atiyat RM, et al. The wild species genome ancestry of domestic chickens. BMC Biol. 2020;18(1):13.

Lazar E, Haendler B, Jacob M. Two 5S genes are expressed in chicken somatic cells. Nucleic Acids Res. 1983;11(22):7735-7741.

Le Béguec C, Wucher V, Lagoutte L, Cadieu E, Botherel N, Hédan B, et al. Characterisation and functional predictions of canine long non-coding RNAs. Sci Rep. 2018 Sep 7;8(1):13444.

Lee BR, Budzillo A, Hadley K, Miller JA, Jarsky T, Baker K, et al. Scaled, high fidelity electrophysiological, morphological, and transcriptomic cell characterization. eLife. 2021;10:e65482.

Lee SH, Eldi P, Cho SY, Rangasamy D. Control of chicken CR1 retrotransposons is independent of Dicer-mediated RNA interference pathway. BMC Biol. 2009; 7:53.

Lee, S.Y., Park, C.G. and Choi, Y., 1996. T cell receptor-dependent cell death of T cell hybridomas mediated by the CD30 cytoplasmic domain in association with tumor necrosis factor receptor-associated factors. J Exp Med 183, 669-74.

Lemoine M, Dupont J, Guillory V, Tesseraud S, Blesbois E. Potential involvement of several signaling pathways in initiation of the chicken acrosome reaction. Biol Reprod. 2009;81(4):657-65.

Li A, Zhang J, Zhou Z, Wang L, Liu Y, Liu Y. ALDB: a domestic-animal long noncoding RNA database. PLoS One. 2015;10(4):e0124003.

Li BC, Chen GH, Xiao XJ, Qin J, Wu SX, Xie KZ, et al. Relationship between PGCs settle and gonad development in the early chicken embryo. Asian Aust. J. Anim. Sci. 2004; 17, 453–459.

Li C, Lu L, Feng B, Zhang K, Han S, Hou D, et al. The lincRNA-ROR/miR-145 axis promotes invasion and metastasis in hepatocellular carcinoma via induction of epithelial-mesenchymal transition by targeting ZEB2. Sci Rep. 2017 Jul 5;7:4637.

Li D, Che T, Chen B, Tian S, Zhou X, Zhang G, et al. Genomic data for 78 chickens from 14 populations. Gigascience. 2017;6(6):1–5.

Li D, Li Y, Li M, Che T, Tian S, Chen B, et al. Population genomics identifies patterns of genetic diversity and selection in chicken. BMC Genomics. 2019;20(1):263.

Li F, Han H, Lei Q, Gao J, Liu J, Liu W, et al. Genome-wide association study of body weight in Wenshang Barred chicken based on the SLAF-seq technology. J Appl Genet. 2018;59(3):305–312.

Li F, Liu J, Liu W, Gao J, Lei Q, Han H, et al. Genome-wide association study of body size traits in Wenshang Barred chickens based on the specific-locus amplified fragment sequencing technology. Anim Sci J. 2021;92(1):e13506.

Li H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25: 1754-60.

Li H, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25: 2078-9.

Li H, Janssens J, De Waegeneer M, Kolluru SS, Davie K, Gardeux V, et al. Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. Science. 2022;375(6584):eabk2432.

Li J, Davis BW, Jern P, Dorshorst BJ, Siegel PB, Andersson L. Characterization of the endogenous retrovirus insertion in associated with henny feathering in chicken. Mob DNA. 2019;10:38.

Li J, Xing S, Zhao G, Zheng M, Yang X, Sun J, Wen J, Liu R. Identification of diverse cell populations in skeletal muscles and biomarkers for intramuscular fat of chicken by single-cell RNA sequencing. BMC Genom. 2020;21(1):752.

Li J, Wang Z., Lubritz D., Arango J., Fulton J., Settar P., Rowland K., Cheng H., Wolc A. Genome-wide association studies for egg quality traits in White Leghorn layers using low-pass sequencing and SNP chip data. J Anim Breed Genet 2022.

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

Li M, Sun C, Xu N, Bian P, Tian X, Wang X, et al. De Novo Assembly of 20 Chicken Genomes Reveals the Undetectable Phenomenon for Thousands of Core Genes on Microchromosomes and Subtelomeric Regions. Mol Biol Evol. 2022;39(4).

Li Y, Liu X, Bai X, Wang Y, Leng L, Zhang H, et al. Genetic parameters estimation and genome-wide association studies for internal organ traits in an F2 chicken population. J Anim Breed Genet. 2022;139(4):434–446.

Li YD, Liu X, Li ZW, Wang WJ, Li YM, Cao ZP, et al. A combination of genome-wide association study and selection signature analysis dissects the genetic architecture underlying bone traits in chickens. Animal. 2021;15(8):100322.

Liao R, Wang Z, Chen Q, Tu Y, Chen Z, Wang Q, et al. An efficient genotyping method in chicken based on genome reducing and sequencing. PLoS ONE. 2015;10(8):e0137010.

Liao R, Zhang X, Chen Q, Wang Z, Wang Q, Yang C, et al. Genome-wide association study reveals novel variants for growth and egg traits in Dongxiang blue-shelled and White Leghorn chickens. Anim Genet. 2016;47(5):588–596.

Lin S, Lin X, Zhang Z, Jiang M, Rao Y, Nie Q, et al. Copy number variation in SOX6 contributes to chicken muscle development. Genes (Basel). 2018;9(1):42.

Lindeboom RGH, Regev A, Teichmann SA. Towards a Human Cell Atlas: Taking Notes from the Past. Trends Genet. 2021;37(7):625-630.

Liu C, Zheng S, Wang Y, Jing L, Gao H, Gao Y et al. Detection and molecular characterization of recombinant avian leukosis viruses in commercial egg-type chickens in China. Avian Pathol. 2011;40(3):269-275.

Liu GE, Jiang L, Tian F, Zhu B, Song J. Calibration of mutation rates reveals diverse subfamily structure of galliform CR1 repeats. Genome Biol Evol. 2009; 1:119-130.

Liu J, Shen Q, Bao H. Comparison of seven SNP calling pipelines for the next-generation sequencing data of chickens. PLoS One. 2022;17(1):e0262574.

Liu J, Wang Z, Li J, Xu L, Liu J, Feng S, et al. A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. Genome Res 31:497-511 (2021).

Liu R, Xing, S., Wang, J. Zheng M, Cui H, Crooijmans RPMA, et al. A new chicken 55K SNP genotyping array. BMC Genomics. 2019;20(1):410. doi:10.1186/s1286401957368

Liu Z, Sun C, Yan Y, Li G, Li XC, Wu G, et al. Design and evaluation of a custom 50K Infinium SNP array for egg-type chickens. Poult Sci. 2021;100(5):101044.

Lizio M, Deviatiiarov R, Nagai H, Galan L, Arner E, Itoh M, et al. Systematic analysis of transcription start sites in avian development. PLoS Biol. 2017 Sep;15(9):e2002887.

Locati MD, Pagano JFB, Girard G, Ensink WA, van Olst M, van Leeuwen S, et al. Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. RNA. 2017;23(8):1188–1199.

Logsdon, G. A., M. R. Vollger, P. Hsieh, Y. Mao, M. A. Liskovykh, S. Koren, et al. The structure, function and evolution of a complete human chromosome 8. Nature, 2021. 593: 101-07.

Love J, Gribbin C, Mather C, Sang H. Transgenic birds by DNA microinjection. Bio/Technology. 1994;12(1):60–63.

Lovell PV, Wirthlin M, Wilhelm L, Minx P, Lazar NH, Carbone L, et al. Conserved syntenic clusters of protein coding genes are missing in birds. Genome Biol. 2014;15(12):565.

Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;15(6):e8746.

Lyall J, Irvine RM, Sherman A, Trevelyan JM, Nuñez A, Purdie A, et al. Suppression of avian influenza transmission in genetically modified chickens. Science 2011; 331:223–226.

Macdonald J, Taylor L, Sherman A, Kawakami K, Takahashi Y, Sang HM, et al. Efficient genetic modification and germ-line transmission of primordial germ cells using piggyBac and Tol2 transposons. Proc Natl Acad Sci U S A. 2012;109(23):8803.

Machete JB, Kgwatalala PM, Nsoso SJ, Hlongwane NL, Moreki JC. Genetic diversity and population structure of three strains of indigenous Tswana chickens and commercial broiler using single nucleotide polymormophic (SNP) markers. Open J Anim Sci. 2021;11(4):515–531.

Mahajan S, Bachtrog D. 2017. Convergent evolution of Y chromosome gene content in flies. Nature Communications 8:785.

Mahood EH, Kruse LH, Moghe GD. Machine learning: A powerful tool for gene function prediction in plants. Appl Plant Sci. 2020;8(7):e11376.

Malinovskaya LP, Tishakova K V., Volkova NA, Torgasheva AA, Tsepilov YA, Borodin PM: Interbreed variation in meiotic recombination rate and distribution in the domestic chicken Gallus gallus. Arch Anim Breed 2019;62:403–411.

Malinovskaya LP, Tishakova K V., Bikchurina TI, Slobodchikova AY, Torgunakov NY, Torgasheva AA, et al. Negative heterosis for meiotic recombination rate in spermatocytes of the domestic chicken Gallus gallus. Vavilov J Genet Breed 2021;25:661–668.

Malomane DK, Reimer C, Weigend S, Weigend A, Sharifi AR, Simianer H. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. BMC Genomics. 2018;19(1):22.

Malomane DK, Simianer H, Weigend A, Reimer C, Schmitt AO Weigend S. The SYNBREED chicken diversity panel: a global resource to assess chicken diversity at high genomic resolution. BMC Genomics. 2019;20(1):345.

Mank JE. 2012. Small but mighty: the evolutionary dynamics of W and Y sex chromosomes. Chromosome Research 20:21-33.

Mank JE, Nam K, Brunström B, Ellegren H. 2010. Ontogenetic Complexity of Sexual Dimorphism and Sex-Specific Selection. Molecular Biology and Evolution 27:1570-1578.

Manni, M., M. R. Berkeley, M. Seppey, and E. M. Zdobnov. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. Curr Protoc, 1: e323.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct;461(7265):747–53.

Manthey JD, Moyle G, Boissinot S. Multiple and independent phases of Ttansposable element amplification in the genomes of piciformes (Woodpeckers and Allies). Genome Biol Evol. 2018; 10: 1445-1456.

Mantri M, Scuderi GJ, Abedini-Nassab R, Wang MFZ, McKellar D, Shi H, et al. Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis. Nat Commun. 2021;12(1):1771.

Mao Y, Zhang G. A complete, telomere-to-telomere human genome sequence presents new opportunities for evolutionary genomics. Nat Methods. 2022;19(6):635-8.

Marchesi JAP, Ono RK, Cantão ME et al. Exploring the genetic architecture of feed efficiency traits in chickens. Sci Rep 2021; 11, 4622.

Mariadassou M, Suez M, Sathyakumar S, Vignal A, Arca M, Nicolas P, et al. Unraveling the history of the genus Gallus through whole genome sequencing. Mol Phylogenet Evol. 2021;158:107044.

Martin F, Ménétret J-F, Simonetti A, Myasnikov AG, Vicens Q, Prongidi-Fix L, et al. Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. Nat Commun. 2016;7:12622.

Martin G, Otto SP, Lenormand T: Selection for recombination in structured populations. Genetics 2006;172:593–609.

Martin RJ. Diversity and Evolution in the Avian Major Histocompatibility Complex. PhD thesis, University of Cambridge. 2021.

Maruoka T, Tanabe H, Chiba M, Kasahara M. Chicken CD1 genes are located in the MHC: CD1 and endothelial protein C receptor genes constitute a distinct subfamily of class-I-like genes that predates the emergence of mammals. Immunogenetics. 2005 Sep;57(8):590-600.

Mason AS. Falling fowl of the chicken reference genome: pitfalls of studying polymorphic endogenous retroviruses. Retrovirology. 2021;18(1):10.

Mason AS, Fulton JE, Hocking PM, Burt DW. A new look at the LTR retrotransposon content of the chicken genome. BMC Genomics. 2016;17:688.

Mason AS, Fulton JE, Smith J. Endogenous avian leukosis virus subgroup E elements of the chicken reference genome. Poult Sci. 2020a;99(6):2911–2915.

Mason AS, Lund AR, Hocking PM, Fulton JE, Burt DW. Identification and characterisation of endogenous Avian Leukosis Virus subgroup E (ALVE) insertions in chicken whole genome sequencing data. Mob DNA. 2020b;11:22.

Mason AS, Miedzinska K, Kebede A, Bamidele O, Al-Jumaili AS, Dessie T et al. (2020). Diversity of endogenous avian leukosis virus subgroup E (ALVE) insertions in indigenous chickens. GSE. 2020c;52(1):29.

Mason SW, Sander EE, Grummt I. Identification of a transcript release activity acting on ternary transcription complexes containing murine RNA polymerase I. EMBO J. 1997;16:163-172.

Massin P, Rodrigues P, Marasescu M, van der Werf S, Naffakh N. Cloning of the chicken RNA polymerase I promoter and use for reverse genetics of influenza A viruses in avian cells. J Virol. 2005;79:13811–13816.

Matsuda Y, Nishida-Umehara C, Tarui H, et al (2005) Highly conserved linkage homology between birds and turtles: bird and turtle chromosomes are precise counterparts of each other. Chromosom Res 13:601–615

Matsui Y, Zsebo K, and Hogan BLM. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell 1992; 70, 841–847.

Mays JK, Black-Pyrkosz A, Mansour T, Schutte BC, Chang S, Dong, K et al. Endogenous Avian Leukosis Virus in Combination with Serotype 2 Marek's Disease Virus Significantly Boosted the Incidence of Lymphoid Leukosis-Like Bursal Lymphomas in Susceptible Chickens. J Virol. 2019;93(23):e00861-19.

McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Cazier JB, et al. Choice of transcripts and software has a large effect on variant annotation. Genome Medicine. 2014 Mar 31;6(3):26.

McDonald JF. 1993. Evolution and consequences of transposable elements. Current Opinion in Genetics & Development 3:855-864.

McDougal T, 2021. Gene editing to enhance production in developing nations. https://www.poultryworld.net/poultry/gene-editing-to-enhance-production-in-developing-nations/ consulted 26th August 2022.

McGrew MJ, Sherman A, Ellard FM, Lillico SG, Gilhooley HJ, Kingsman AJ, et al. Efficient production of germline transgenic chickens using lentiviral vectors. EMBO Rep. 2004; 5(7):728–733.

McTavish EJ, Hillis DM. How do SNP ascertainment schemes and population demographics affect inferences about population history? BMC Genomics. 2015;16: 266.

Mead D, Ogden R, Meredith A, Peniche G, Smith M, Corton C, et al. The genome sequence of the European golden eagle, Aquila chrysaetos chrysaetos Linnaeus 1758. Wellcome Open Res 6:112 (2021).

Megens HJ, Crooijmans RPMA, Bastiaansen JWM, Kerstens HHD, Coster A, Jalving R, et al. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. BMC Genetics. 2009;10:86.

Mendonça MAC, Carvalho CR, Clarindo WR (2016) DNA amount of chicken chromosomes resolved by image cytometry Caryologia 69:201-206.

Mendonça MAC, Carvalho CR, Clarindo WR: DNA amount of chicken chromosomes resolved by image cytometry. Caryologia 69:201-206 (2016).

Mieulet D, Aubert G, Bres C, Klein A, Droc G, Vieille E, et al. Unleashing meiotic crossovers in crops. Nat Plants 2018;4:1010–1016.

Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, et al. Telomere-to-telomere assembly of a complete human X chromosome. Nature. 2020;585(7823):79-84.

Miller MM, Taylor RL Jr. Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. Poult Sci. 2016 Feb;95(2):375-92.

Miller MM, Goto R, Young S, Chirivella J, Hawke D, Miyada CG. Immunoglobulin variable-region-like domains of diverse sequence within the major histocompatibility complex of the chicken. Proc Natl Acad Sci U S A. 1991 May 15;88(10):4377-81.

Miller MM, R Goto, A Bernot, R Zoorob, C Auffray, N Bumstead, and W E Briles. "Two Mhc Class I and Two Mhc Class II Genes Map to the Chicken Rfp-Y System Outside the B Complex." Proc Natl Acad Sci U S A 1994. 91 (10): 4397–4401.

Miller MM, Bacon LD, Hala K, Hunt HD, Ewald SJ, Kaufman J, et al. 2004 Nomenclature for the chicken major histocompatibility (B and Y) complex. Immunogenetics. 2004 Jul;56(4):261-79.

Miller MM, Wang C, Parisini E, Coletta RD, Goto RM, Lee SY, et al. Characterization of two avian MHC-like genes reveals an ancient origin of the CD1 family. Proc Natl Acad Sci U S A. 2005;102(24):8674-9.

Miller MM, Robinson CM, Abernathy J, Goto RM, Hamilton MK, Zhou H, et al. Mapping genes to chicken microchromosome 16 and discovery of olfactory and scavenger receptor genes near the major histocompatibility complex. J Hered. 2014;105(2):203-15.

Moffitt JR, Lundberg E, Heyn H. The emerging landscape of spatial profiling technologies. Nat Rev Genet. 2022 Jul 20.

Moghadam HK, Pointer MA, Wright AE, Berlin S, Mank JE. 2012. W chromosome expression responds to female-specific selection. Proc Natl Acad Sci U S A 109:8207-8211.

Mohan S, Noller HF. Recurring RNA structural motifs underlie the mechanics of L1 stalk movement. Nat Commun. 2017;8:14285.

Molinas ATR, Desmoulins, L.D., Hamling, B.V., Butcher, S.M., Anwar, I.J., Miyata, K., Enix, C.L., et al. 2019. Interaction between TRPV1-expressing neurons in the hypothalamus. Journal of Neurophysiology 121, 140-151.

Monson MS, Van Goor, A.G., Ashwell, C.M., Persia, M.E., Rothschild, M.F., Schmidt, C.J. et al. Immunomodulatory effects of heat stress and lipopolysaccharide on the bursal transcriptome in two distinct chicken lines. BMC Genomics 2018;19:643.

Morais-Perdigao, A.L., Rodrigues-Fernandes, C.I., Araujo, G.R., Soares, C.D., de Andrade, B.A.B., Martins, M.D., et al. CD30 Expression in Oral and Oropharyngeal Diffuse Large B Cell Lymphoma, not Otherwise Specified. Head Neck Pathol. 2022;16, 476-485.

Morales J, Pujar S, Loveland JE, Astashyn A, Bennett R, Berry A, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature. 2022;604(7905):310–5.

Morenos L, Chatterton Z, Ng JL, Halemba MS, Parkinson-Bates M, Mechinaud F, et al. Hypermethylation and down-regulation of DLEU2 in paediatric acute myeloid leukaemia independent of embedded tumour suppressor miR-15a/16-1. Mol Cancer. 2014;13:123.

Morgulis, A., E. M. Gertz, A. A. Schaffer, and R. Agarwala. WindowMasker: window-based masker for sequenced genomes. Bioinformatics. 2006;22:134-41.

Morrison JA, Box AC, McKinney MC, McLennan R, Kulesa PM. Quantitative single cell gene expression profiling in the avian embryo. Dev Dyn. 2015;244(6):774-784.

Morrison JA, McLennan R., Wolfe L.A., Gogol M.M., Meier S., McKinney M.C., et al. Single-cell transcriptome analysis of avian neural crest migration reveals signatures of invasion and molecular transitions. eLife. 2017;6:e28415.

Moses L, Pachter L. Museum of spatial transcriptomics. Nat Methods. 2022;19(5):534-546.

Mountford J, Gheyas A, Vervelde L, Smith J. Genetic variation in chicken interferon signalling pathway genes in research lines showing differential viral resistance. Anim Genet. 2022;53(5):640–656.

Mozdziak PE, Petitte JN. Status of transgenic chicken models for developmental biology. Dev. Dyn. 2004;229:414–421.

Mugal CF, Nabholz B, Ellegren H. Genome-wide analysis in chicken reveals that local levels of genetic diversity are mainly governed by the rate of recombination. BMC Genomics. 2013;14:86.

Muir WM, Wong GK, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, et al. Review of the initial validation and characterization of a 3K chicken SNP array. World's Poult Sci J. 2008a;64(2):219–225.

Muir WM, Wong GKS, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, et al. 2008b Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. Proc Natl Acad Sci USA. 2008b;105(45):17312–17317.

Mund A, Brunner AD, Mann M. Unbiased spatial proteomics with single-cell resolution in tissues. Mol Cell. 2022;82(12):2335-2349.

Munschauer M, Nguyen CT, Sirokman K, Hartigan CR, Hogstrom L, Engreitz JM, et al. The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. Nature. 2018;561(7721):132–6.

Muret K, Klopp C, Wucher V, Esquerré D, Legeai F, Lecerf F, et al. Long noncoding RNA repertoire in chicken liver and adipose tissue. Genet Sel Evol. 2017;49(1):6.

Muret K, Désert C, Lagoutte L, Boutin M, Gondret F, Zerjal T, et al. Long noncoding RNAs in lipid metabolism: literature review and conservation analysis across species. BMC Genomics. 2019;20(1):882.

Nadeau JH, Taylor BA (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. Proc Natl Acad Sci U S A. 81:814–818.

Nakai A, and Ishikawa T. Cell Cycle Transition under Stress Conditions Controlled by Vertebrate Heat Shock Factors. EMBO J. 2001;20(11):2885–95.

Nanda I, Karl E, Volobouev V, Griffin DK, Schartl M, Schmid M: Extensive gross genomic rearrangements between chicken and Old World vultures (Falconiformes: Accipitridae). Cytogenet Genome Res. 112:286-295 (2006).

Nanda I, Karl E, Griffin DK, et al. Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. Cytogenet Genome Res. (2007) 117:43–53.

Nandi S, Whyte J, Taylor L, Sherman A, Nair V, Kaiser P, Mcgrew MJ. Cryopreservation of specialized chicken lines using cultured primordial germ cells. Poult Sci. 2016;95(8):1905–1911.

Narezkina A, Taganov KD, Litwin S, Stoyanova R, Hayashi J, Seeger C, et al. Genome-wide analyses of avian sarcoma virus integration sites. J Virol. 2004;78(21):11656–11663.

Nattestad, M., M. C. Schatz. 2016. Assemblytics: a web analytics tool for the detection of variants from an assembly. Bioinformatics, 32:3021-3.

Nawaz AH, Amoah K, Leng, Q.Y., Zheng, J.H., Zhang, W.L. and Zhang, L., 2021. Poultry Response to Heat Stress: Its Physiological, Metabolic, and Genetic Implications on Meat Production and Quality Including Strategies to Improve Broiler Production in a Warming World. Front Vet Sci 8:699081.

NCBI's RefSeq. Gallus_gallus-2.1 - galGal3 - Genome - Assembly - NCBI [Internet]. 2006. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.1/

NCBI's RefSeq. Gallus_gallus-4.0 - galGal4 - Genome - Assembly - NCBI [Internet]. 2011. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.3/

NCBI's RefSeq. Gallus_gallus-5.0 - Genome - Assembly - NCBI [Internet]. 2015. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.4/

NCBI's RefSeq. GRCg6a - galGal6 - Genome - Assembly - NCBI [Internet]. 2018. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.5/

NCBI's RefSeq. bGalGal1.mat.broiler.GRCg7b - Genome - Assembly - NCBI [Internet]. 2021a. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_016699485.2/

NCBI's RefSeq N. bGalGal1.pat.whiteleghornlayer.GRCg7w_WZ - Genome - Assembly - NCBI [Internet]. 2021b. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_016700215.2

NCBI's RefSeq. Gallus gallus Annotation Report [Internet]. 2022. Available from: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Gallus_gallus/106/

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 2014;505(7485):635–40.

Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science. 2016 Jan 15;351(6270):271–5.

Nguyen AH, Bachtrog D. 2021. Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome. PLoS Genet. 17:e1009438.

Nguyen HP, Yi, D., Lin, F., Viscarra, J.A., Tabuchi, C., Ngo, K., Shin, G., Lee, A.Y., Wang, Y. and Sul, H.S., 2020. Aifm2, a NADH Oxidase, Supports Robust Glycolysis and Is Required for Cold- and Diet-Induced Thermogenesis. Mol Cell. 77, 600-617 e4.

Nie C, Almeida P, Jia Y, Bao H, Ning Z, Qu L. Genome-wide single-nucleotide polymorphism data nnveil admixture of Chinese indigenous chicken breeds with commercial breeds. Genome Biol Evol. 2019;11(7):1847–1856.

Nikolaidis, Nikolas, Izabela Makalowska, Dimitra Chalkia, Wojciech Makalowski, Jan Klein, and Masatoshi Nei. 2005. Origin and Evolution of the Chicken Leukocyte Receptor Complex. Proc Natl Acad Sci U S A. 102 (11): 4057–62.

Nikolic EI, L. M. King, M. Vidakovic, N. Irigoyen, and I. Brierley. 2012. 'Modulation of ribosomal frameshifting frequency and its effect on the replication of Rous sarcoma virus', J Virol, 86: 11581-94.

Nishihara H, Smit AFA, Okada N. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res. 2006; 16:864-874.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. Science. 2000;289:920–930.

Noller HF, Lancaster L, Zhou J, Mohan S. The ribosome moves: RNA mechanics and translocation. Nat Struct Mol Biol. 2017;24(12):1021–1027.

Nudelman G, Frasca A, Kent B, Sadler KC, Sealfon SC, Walsh MJ, et al. High resolution annotation of zebrafish transcriptome using long-read sequencing. Genome Res. 2018;28(9):1415-25.

Núñez-León D, Aguirre-Fernández G, Steiner A, Nagashima H, Jensen P, Stoeckli E, Schneider RA, Sánchez-Villagra MR. Morphological diversity of integumentary traits in fowl domestication: Insights from disparity analysis and embryonic development. Dev Dyn. 2019;248(11):1044-1058.

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376(6588):44-53.

O'Connor RE, Farré M, Joseph S, et al. (2018a) Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. Genome Biol. 19(1):171.

O'Connor RE, Romanov MN, Kiazim LG, et al (2018c) Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. Nat Commun 9:1883.

O'Connor RE, Kiazim L, Skinner B, Fonseka G, Joseph S, Jennings R, et al. Patterns of microchromosome organization remain highly conserved throughout avian evolution. Chromosoma. 2019;128(1):21-9.

Ogle JM, Brodersen DE, Clemons WM Jr, Tarry MJ, Carter AP, Ramakrishnan V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science. 2001;292(5518):897–902.

O'Hare TH, Delany ME. Genetic variation exists for telomeric array organization within and among the genomes of normal, immortalized, and transformed chicken systems. Chromosome Res. 2009; 17:947-964.

Oishi I, Yoshii K, Miyahara D, Kagami H and Tagami T. Targeted mutagenesis in chicken using CRISPR/Cas9 system. Sci Rep. 2016; 6:23980.

Olanrewaju HA, JL Purswell, S. D. Collier, and S. L. Branton. 2010. Effect of Ambient Temperature and Light Intensity on Physiological Reactions of Heavy Broiler Chickens1. Poult Sci. 89 (12): 2668–77.

O'Neill M, Binder M, Smith C, Andrews J, Reed K, Smith M, et al. 2000. ASW: a gene with conserved avian W-linkage and female specific expression in chick embryonic gonad. Dev Genes Evol 210:243-249.

Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B. 2012. Evidence of a neo-sex chromosome in birds. Heredity 108:264-272.

Palla G, Fischer DS, Regev A, Theis FJ. Spatial components of molecular tissue biology. Nat Biotechnol. 2022;40(3):308-318.

Pan L, Liang W, Fu M, Huang Z hua, Li X, Zhang W, et al. Exosomes-mediated transfer of long noncoding RNA ZFAS1 promotes gastric cancer progression. J Cancer Res Clin Oncol. 2017;143(6):991–1004.

Panda SK, McGrew MJ. Genome editing of avian species: implications for animal use and welfare. Lab Anim. 2022;56(1):50–59.

Park JS, Lee KY, Han JY. Precise Genome Editing in Poultry and Its Application to Industries. Genes. 2020; 11(10):1182.

Parks MM, Kurylo CM, Dass RA, Bojmar L, Lyden D, Vincent CT, et al. Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. Sci Adv. 2018;4(2):eaao0665.

Paul P, Nag D, Chakraborty S: Recombination hotspots: Models and tools for detection. DNA Repair (Amst) 2016;40:47–56.

Paule MR, White RJ. Transcription by RNA polymerases I and III. Nucleic Acids Res. 2000;28(6):1283–1298.

Payne LN, Nair V. The long view: 40 years of avian leukosis research. Avian Pathol. 2012;41(1):11-19.

Payne LN, Brown SR, Bumstead N, Howes K, Frazier JA, Thouless ME. A novel subgroup of exogenous avian leukosis virus in chickens. J Gen Virol. 1991;72(4):801–807.

Peciña A, Smith KN, Mézard C, Murakami H, Ohta K, Nicolas A: Targeted Stimulation of Meiotic Recombination. Cell 2002;111:173–184.

Penalba JV, Deng Y, Fang Q, Joseph L, Moritz C, Cockburn A: Genome of an iconic Australian bird: High-quality assembly and linkage map of the superb fairy-wren (Malurus cyaneus). Mol Ecol Resour 20:560-578 (2020).

Peona V, Palacios-Gimenez OM, Blommaert J, Liu J, Haryoko T, Jønsson KA, et al. The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. Philos Trans R Soc Lond B Biol Sci. 2021; 376:20200186.

Permal E, Flutre T, Quesneville H. Roadmap for annotating transposable elements in eukaryote genomes. Methods Mol Biol. 2012; 859:53-68.

Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare [Internet]. F1000Research; 2020. Available from: https://f1000research.com/articles/9-304

Pértille F, Guerrero-Bosagna C, Silva VH, Bosciero C, Nunes JRS, Ledur MC, et al. High-throughput and cost-effective chicken genotyping using next-generation sequencing. Sci Rep. 2016;6:26929.

Peterson AL, Payseur BA. (2021) Sex-specific variation in the genome-wide recombination rate Genetics 217(1):1-11.

Petitte JN. Avian germplasm preservation: Embryonic stem cells or primordial germ cells? J Poult Sci. 2006; 85 (2): 237–242.

Pevzner P, Tesler G (2003) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci U S A 100:7672–7.

Pfleiderer Ch, Smid A, Bartsch I, Grummt I. An undecamer DNA sequence directs termination of human ribosomal gene transcription. Nucleic Acid Res. 1990;18(16):4727–4736.

Piégu B, Arensburger P, Beauclair L, Chabault M, Raynaud E, et al. Variations in genome size between wild and domesticated lineages of fowls belonging to the Gallus gallus species. Genomics. 2020; 112:1660-1673.

Pieler T, Hamm J, Roeder RG. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. Cell. 1987;48(1):91–100.

Pigozzi MI: Distribution of MLH1 foci on the synaptonemal complexes of chicken oocytes. Cytogenet Cell Genet 2001;95:129–133.

Pigozzi MI (2007) Localization of single-copy sequences on chicken synaptonemal complex spreads using fluorescence in situ hybridization (FISH) Cytogenet Genome Res 119:105-112

Pigozzi MI: The chromosomes of birds during meiosis. Cytogenet Genome Res 2016;150:128–138.

Pigozzi MI (2022) A bird's eye view of chromosomes during meiotic prophase I BAG Journal of Basic and Applied Genetics 23:27-47

Pigozzi MI, del Priore L: Meiotic recombination analysis in female ducks (Anas platyrhynchos). Genetica 2016;144:307–312.

Pillet L, Fontaine D, Pawlowski J. Intra-genomic ribosomal RNA polymorphism and morphological variation in Elphidium macellum suggests inter-specific hybridization in Foraminifera. PLoS One. 2012;7(2):e32373.

Pivovarova O, Gogebakan O, Sucher, S., Groth, J., Murahovschi, V., Kessler, K., et al. 2016. Regulation of the clock gene expression in human adipose tissue by weight loss. Int J Obes (Lond) 40, 899-906.

Pollock DL, Fechheimer NS: The chromosomes of cockerels (Gallus domesticus) during meiosis. Cytogenet Genome Res 1978;21:267–281.

Polycarpou-Schwarz M, Groß M, Mestdagh P, Schott J, Grund SE, Hildenbrand C, et al. The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation. Oncogene. 2018;37(34):4750–68.

Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009;136(4):629–41.

Poplin R, Ruano-Rubio, V.; DePristo, M.A.; Fennell, T.J.; Carneiro, M.O.; Van der Auwera, G.A, et al . 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv, 201178.

Potts ND, Bichet C, Merat L, Guitton E, Krupa AP, Burke TA, et al. Development and optimization of a hybridization technique to type the classical class I and class II B genes of the chicken MHC. Immunogenetics. 2019;71(10):647-663.

Price PD, Palmer Droguett DH, Taylor JA, Kim DW, Place ES, Rogers TF, et al. 2022. Detecting signatures of selection on gene expression. Nature Ecol Evol 6:1035-1045.

Přikryl D, Plachý J, Kučerová D, Koslová A, Reinišová M, Šenigl F et al. The Novel Avian Leukosis Virus Subgroup K Shares Its Cellular Receptor with Subgroup A. J Virol. 2019;93(17):e00580-19.

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1-14.

Prokop JW, Schmidt C, Gasper D, Duff RJ, Milsted A, Ohkubo T, et al. Discovery of the elusive leptin in birds: identification of several 'missing links' in the evolution of leptin and its receptor. PLoS One. 2014;9(3):e92751.

Pruitt KD, G. R. Brown, S. M. Hiatt, F. Thibaud-Nissen, A. Astashyn, O. Ermolaeva, et al. 2014. 'RefSeq: an update on mammalian reference sequences', Nucleic Acids Res, 42: D756-63.

Prum RO, Berv JS, Dornburg A, et al (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526:569–573.

Psifidi A, Kranis A, Rothwell LM, Bremner A, Russell K, Robledo D, et al. Quantitative trait loci and transcriptome signatures associated with avian heritable resistance to Campylobacter. Sci Rep. 2021;11(1):1623.

Putney JW, Tomita T. 2012. Phospholipase C signaling and calcium influx. Adv Biol Regul 52, 152-64.

Qu X, Li X, Li Z, Liao M, Dai M. Chicken Peripheral Blood Mononuclear Cells Response to Avian Leukosis Virus Subgroup J Infection Assessed by Single-Cell RNA Sequencing. Front Microbiol. 2022;13:800618.

Rahn MI, Solari AJ: Recombination nodules in the oocytes of the chicken, Gallus domesticus. Cytogenet Genome Res 1986;43:187–193.

Ramos-Alemán F, González-Jasso E, Pless RC. Use of alternative alkali chlorides in RT and PCR of polynucleotides containing G quadruplex structures. Anal Biochem. 2018; 543:43-50.

Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596(7871):211-220.

Rao M, Morisson M, Faraut T, et al. (2012) A duck RH panel and its potential for assisting NGS genome assembly. BMC Genomics 13:513.

Rao YS, Li J, Zhang R, Lin XR, Xu JG, Xie L, et al. Copy number variation identification and analysis of the chicken genome using a 60K SNP BeadChip. Poult Sci. 2016;95(8):1750–1756.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444–454.

Reeve J, Ortiz-Barrientos D, Engelstädter J. (2016) The evolution of recombination rates in finite populations during ecological speciation. Proc Biol Sci. 283(1841):20161243.

Reinisová M, Senigl F, Yin X, Plachy J, Geryk J, Elleder D et al. A single-amino-acid substitution in the TvbS1 receptor results in decreased susceptibility to infection by avian sarcoma and leukosis virus subgroups B and D and resistance to infection by subgroup E in vitro and in vivo. J Virol. 2008;82(5):2097–2105.

Rengaraj D, Cha DG, Lee HJ, Lee KY, Choi YH, Jung KM, et al. Dissecting chicken germ cell dynamics by combining a germ cell tracing transgenic chicken model with single-cell RNA sequencing. Comput Struct Biotechnol J. 2022;20:1654-1669.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature 592:737-746 (2021).

Rhodes DA, Reith W, Trowsdale J. Regulation of Immunity by Butyrophilins. Annu Rev Immunol. 2016 May 20;34:151-72.

Rhooms SK, Murari, A., Goparaju, N.S.V., Vilanueva, M. and Owusu-Ansah, E. 2020. Insights from Drosophila on mitochondrial complex I. Cell Mol Life Sci 77:607-618.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:735-742.

Rice ES, S Koren, A Rhie, M. P. Heaton, T. S. Kalbfleisch, T. Hardy, et al. 2020. Continuous chromosome-scale haplotypes assembled from a single interspecies F1 hybrid of yak and cattle. Gigascience. 9(4):giaa029.

Robinson HL, Astrin SM, Senior AM, Salazar FH. Host Susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. J Virol. 1981;40(3):745–751.

Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. Nucleic Acids Res. 2020 Jan 8;48(D1):D948-D955.

Robinson JA, Bowie RCK, Dudchenko O, Aiden EL, Hendrickson SL, Steiner CC, Ryder OA, Mindell DP, Wall JD: Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. Curr Biol 31:2939-2946 e2935 (2021).

Rodionov AV. Micro vs. macro: Structural-functional organization of avian micro- and macrochromosomes. Genetika. 1996;32(5):597–608.

Rodionov A, Myakoshina Y, Chelysheva LA, Solovei I, Gaginskaya E: Chiasmata in the lambrush chromosomes of Gallus gallus domesticus: The cytogenetic study of recombination frequency and linkage map lengths. Genetika 1992;28:53–63.

Rodrigue KL, May BP, Famula TR, Delany ME. Meiotic instability of chicken ultra-long telomeres and mapping of a 2.8 megabase array to the W-sex chromosome. Chromosome. 2005; Res 13:581-591.

Rogers TF, Pizzari T, Wright AE. 2021. Multi-copy gene family evolution on the avian W chromosome. J Hered. 112:250-259.

Rohde F, Schusser B, Hron T, Farkasova H, Plachy J, Hartle S, et al. Characterization of Chicken Tumor Necrosis Factor-alpha, a Long Missed Cytokine in Birds. Front Immunol. 2018;9:605.

Romanov MN, Farré M, Lithgow PE, et al (2014) Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. BMC Genomics 15:1060.

Rostamzadeh Mahdabi E, Esmailizadeh A, Ayatollahi Mehrgardi A, Asadi Fozi M. A genome-wide scan to identify signatures of selection in two Iranian indigenous chicken ecotypes. Genet Sel Evol. 2021;53(1):72.

Rowland K, Saelao P, Wang, Y., Fulton, J.E., Liebe, G.N., McCarron, A.M., Wolc, A., Gallardo, R.A., Kelly, T., Zhou, H., Dekkers, J.C.M. and Lamont, S.J., 2018. Association of Candidate Genes with Response to Heat and Newcastle Disease Virus. Genes (Basel) 9(11):560.

Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464(7288):587–591.

Ruby T, Bed'Hom B, Wittzell H, Morin V, Oudin A, Zoorob R. Characterisation of a cluster of TRIM-B30.2 genes in the chicken MHC B locus. Immunogenetics. 2005 Apr;57(1-2):116-28.

Ruddell A. Transcription regulatory elements of the avian retroviral long terminal repeat. Virology. 1995;206(1):1–7.

Russell KM, Smith J, Bremner A, Chintoan-Uta C, Vervelde L, Psifidi A, et al. Transcriptomic analysis of caecal tissue in inbred chicken lines that exhibit heritable differences in resistance to Campylobacter jejuni. BMC Genomics. 2021;22(1).

Rutherford K, Meehan CJ, Langille MGI, Tyack SG, McKay JC, McLean NL, et al. Discovery of an expanded set of avian leukosis subgroup E proviruses in chickens using Vermillion, a novel sequence capture and analysis pipeline. Poult Sci. 2016;95(10):2250–2258.

Sabour MP, Chambers JR, Grunder AA, Kuhnlein U, Gavora JS. Endogenous viral gene distribution in populations of meat-type chickens. Poult Sci. 1992;71(8):1259–1270.

Sacco MA, Nair VK. Prototype endogenous avian retroviruses of the genus Gallus. J Gen Virol. 2014;95(9):2060–2070.

Sacco MA, Howes K, Smith LP, Nair VK. Assessing the roles of endogenous retrovirus EAV-HP in avian leukosis virus subgroup J emergence and tolerance. J Virol. 2004;78(19):10525–10535.

Saelao P, Wang Y, Gallardo RA, Lamont SJ, Dekkers JM, Kelly T, et al. 2018. Novel insights into the host immune response of chicken Harderian gland tissue during Newcastle disease virus infection and heat treatment. BMC Vet Res 14, 280.

Saelao P, Wang Y, Chanthavixay G, Gallardo RA, Wolc A, Dekkers JCM, et al. 2019. Genetics and Genomic Regions Affecting Response to Newcastle Disease Virus Infection under Heat Stress in Layer Chickens. Genes (Basel) 10(1):61.

Saelao P, Wang Y, Chanthavixay G, Yu V, Gallardo RA, Dekkers JCM, et al. 2021. Distinct transcriptomic response to Newcastle disease virus infection during heat stress in chicken tracheal epithelial tissue. Sci Rep 11, 7450.

Salomonsen J, Marston D, Avila D, Bumstead N, Johansson B, Juul-Madsen H, et al. The properties of the single chicken MHC classical class II alpha chain (B-LA) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. Immunogenetics. 2003 Dec;55(9):605-14.

Salomonsen J, Sørensen MR, Marston DA, Rogers SL, Collen T, van Hateren A, et al. Two CD1 genes map to the chicken MHC, indicating that CD1 genes are ancient and likely to have been present in the primordial MHC. Proc Natl Acad Sci U S A. 2005 Jun 14;102(24):8668-73.

Salzberg SL. Next-generation genome annotation: we still struggle to get it right. Genome Biol. 2019;20(1):92.

Sang H. Prospects for transgenesis in the chick. Mech Dev. 2004;121(9):1179–1186.

Sankoff D. (2009) The where and wherefore of evolutionary breakpoints. J Biol 8:66.

Sarno R, Vicq Y, Uematsu N, Luka M, Lapierre C, Carroll D, et al.: Programming sites of meiotic crossovers using Spo11 fusion proteins. Nucleic Acids Res 2017;45:e164–e164.

Sarsour AH, Persia ME. 2022. Effects of Sulfur Amino Acid Supplementation on Broiler Chickens Exposed to Acute and Chronic Cyclic Heat Stress. Poult Sci 101 (7): 101952.

Saunders A, Huang KW, Vondrak C, Hughes C, Smolyar K, Sen H, et al. Ascertaining cells' synaptic connections and RNA expression simultaneously with massively barcoded rabies virus libraries. bioRxiv 2021.09.06.459177.

Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. 2021. 'Database resources of the National Center for Biotechnology Information', Nucleic Acids Res, 49: D10-D17.

Schmid M, Nanda I, Guttenbach M, Steinlein C, Hoehn M, Schartl M, et al: First report on chicken genes and chromosomes 2000. Cytogenet Cell Genet. 2000;90:169–218.

Schmid M, Nanda I, Hoehn H, Schartl M, Haaf T, Buerstedde JM, et al: Second report on chicken genes and chromosomes 2005. Cytogenet Genome Res. 2005;109:415–479.

Schmid M, Smith J, Burt DW, Aken BL, Antin PB, Archibald AL, et al. Third report on chicken genes and chromosomes 2015. Cytogenet Genome Res. 2015;145(2):78–179.

Schneider VA, Graves-Lindsay T, K. Howe, N. Bouk, H. C. Chen, P. A. Kitts, et al. 2017. 'Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly', Genome Res. 27: 849-64.

Schulte P, Alegret L, Arenillas I, et al. (2010) The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. Science 327:1214–8.

Séguéla-Arnaud M, Crismani W, Larchevêque C, Mazel J, Froger N, Choinard S, et al. Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM. Proc Natl Acad Sci U S A. 2015;112:4713–4718.

Séguéla-Arnaud M, Choinard S, Larchevêque C, Girard C, Froger N, Crismani W, et al. RMI1 and TOP3α limit meiotic CO formation through their C-terminal domains. Nucleic Acids Res 2016;45:gkw1210.

Segura J, Ferretti L, Ramos-Onsins S, Capilla L, Farré M, Reis F, et al. Evolution of recombination in eutherian mammals: insights into mechanisms that affect recombination rates and crossover interference. Proc Biol Sci. 2013;280(1771):20131945.

Semenov GA, Basheva EA, Borodin PM, Torgasheva AA. High rate of meiotic recombination and its implications for intricate speciation patterns in the white wagtail (Motacilla alba). Biol J Linn Soc 2018;125:600–612.

Serebrovsky A, Petrov S: On the composition of the plan of the chromosomes of the domestic hen. Zhurnal Exp Biol 1930;6:157–180.

Seroussi E, Cinnamon Y, Yosefi S, Genin O, Smith JG, Rafati N, et al. Identification of the Long-Sought Leptin in Chicken and Duck: Expression Pattern of the Highly GC-Rich Avian leptin Fits an Autocrine/Paracrine Rather Than Endocrine Function. Endocrinology. 2016;157(2):737-51.

Shang P, Hoogerbrugge J, Baarends WM, Grootegoed JA. Evolution of testis-specific kinases TSSK1B and TSSK2 in primates. Andrology. 2013;1(1):160-8.

Shaw I, Powell TJ, Marston DA, Baker K, van Hateren A, Riegert P, et al. Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. J Immunol. 2007 May 1;178(9):5744-52.

Shaw P, Brown J. Nucleoli: composition, function, and dynamics. Plant Physiol. 2012;158(1):44–51.

Shedlock AM, Edwards SV. (2009) Amniotes. In: Hedges SB, Kumar S, editors. The Timetree of Life. Oxford: Oxford University Press, p 375–379.

Shehata AM, Saadeldin IM, Tukur, H.A. and Habashy, W.S., 2020. Modulation of Heat-Shock Proteins Mediates Chicken Cell Survival against Thermal Stress. Animals (Basel) 10(12):2407.

Sherman BT, Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T. and Chang, W., 2022. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 50(W1):W216-W221.

Shetty S, Griffin DK, Graves JA: Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. Chromosome Res 7:289-295 (1999).

Shiina T, Briles WE, Goto RM, Hosomichi K, Yanagiya K, Shimizu S, et al. Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. J Immunol. 2007;178(11):7162-72.

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021;37(12):1639–43.

Sigeman H, Ponnikas S, Hansson B. 2020. Whole-genome analysis across 10 songbird families within Sylvioidea reveals a novel autosome–sex chromosome fusion. Biology Letters 16:20200082.

Sigeman H, Strandh M, Proux-Wéra E, Kutschera VE, Ponnikas S, Zhang H, et al. 2021. Avian Neo-Sex Chromosomes Reveal Dynamics of Recombination Suppression and W Degeneration. Mol Biol Evol. 38:5275-5291.

Simakov O, Marletaz F, Yue JX, O'Connell B, Jenkins J, Brandt A, et al. Deeply conserved synteny resolves early events in vertebrate evolution. Nat Ecol Evol 4:820-830 (2020).

Simsa S, Genina O, Ornan EM. Matrix metalloproteinase expression and localization in turkey (Meleagris gallopavo) during the endochondral ossification process. J Anim Sci. 2007;85(6):1393-401.

Singer M, Berg P. Genes and Genomes, a Changing Perspective. Mill Valley: University Science Books; 1991.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, et al. Stable recombination hotspots in birds. Science 2015;350:928–932.

Siren J, J Monlong, X Chang, AM Novak, JM Eizenga, C Markello, et al. 2021. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. Science, 374:abg8871.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837.

Skinner BM, Griffin DK (2011) Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity (Edinb) 108:37–41.

Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, et al. 2015. Evolutionary analysis of the female-specific avian W chromosome. Nat Commun 6:1-10.

Smeds L, Mugal CF, Qvarnström A, Ellegren H. (2016) High-Resolution Mapping of Crossover and Non-crossover Recombination Events by Whole-Genome Re-sequencing of an Avian Pedigree. PLoS Genet 12:e1006044.

Smirnov E, Cmarko D, Mazel T, Hornacek M, Raska I. Nucleolar DNA: The host and the guests. Histochem Cell Biol. 2016;145:359–372.

Smit A, Hubley R, Green P. 2013. RepeatMasker. http://repeatmasker.org.

Smith CA, Roeszler KN, Ohnesorg T, Cummins DM, Farlie PG, Doran TJ, et al: The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature 461:267-271 (2009).

Smith EJ, Fadly AM, Crittenden LB. Interactions between endogenous virus loci ev6 and ev21:1. Immune response to exogenous avian leukosis virus infection. Poult Sci. 1990;69(8):1244–1250.

Smith J, Burt DW. (1998) Parameters of the chicken genome (Gallus gallus). Anim Genet 29:290-294

Soh YS, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, et al. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159:800-813.

Sohn JI, Nam JW. The present and future of de novo whole-genome assembly. Brief Bioinform. 2018;19(1):23-40.

Sohn J, Nam K, Hong H, Kim JM, Lim D, Lee KT, et al. Whole genome and transcriptome maps of the entirely black native Korean chicken breed Yeonsan Ogye. GigaScience. 2018;7(7):giy086.

Sohrabi SS, Mohammadabadi M, Wu DD, Esmailizadeh A. Detection of breed-specific copy number variations in domestic chicken genome. Genome. 2018;61(1):7–14.

Solinhac R, Leroux S, Galkina S, Chazara O, Feve K, Vignoles F, et al. Integrative mapping analysis of chicken microchromosome 16 organization. BMC Genomics. 2010;11:616.

Son DJ, Kumar S, Takabe W, Kim CW, Ni C-W, Alberts-Grill N. The atypical mechanosensitive microRNA-712 derived from pre-ribosomal RNA induces endothelial inflammation and atherosclerosis. Nat Commun. 2013;4:3000.

Speedy AW. Global production and consumption of animal source foods. J Nutrition. 2003; 133, 4048S–4053S.

Spiegel J, Cuesta SM, Adhikari S, Hänsel-Hertsch R, Tannahill D, et al. G-quadruplexes are transcription factor binding hubs in human chromatin. Genome Biol. 2021; 22:117.

Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021 Feb;22(2):96–118.

Stern CD. The Chick Embryo–Past, Present and Future as a Model System in Developmental Biology. Mech Dev. 2004; 121, 1011–1013.

Stern CD. The chick; A great model system becomes even greater. Dev Cell. 2005; 8(1), 9–17.

Stiehler F, Steinborn M, Scholz S, Dey D, Weber APM, Denton AK. Helixer: Cross-species gene annotation of large eukaryotic genomes using deep learning. Bioinformatics. 2020.

St John JA, Braun EL, Isberg SR, et al (2012) Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. Genome Biol 13:415.

Stoye JP. Endogenous retroviruses: still active after all these years? Curr Biol. 2001;11(22):R914–R916.

St-Pierre NR, Cobanov B, Schnitkey G. 2003. Economic Losses from Heat Stress by US Livestock Industries. J Dairy Sci 86, E52-E77.

Straathof KC, Pulè,MA, Yotnda P, Dotti G, Vanin EF, Brenner MK, et al. An inducible caspase 9 safety switch for T-cell therapy. Blood, 2005;105(11):4247–4254.

Strillacci MG, Cozzi MC, Gorla E, Mosca F, Schiavini F, Román-Ponce SI, et al. Genomic and genetic variability of six chicken populations using single nucleotide polymorphism and copy number variants as markers. Animal. 2017;11(5):737–745.

Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019;20(5):257-272.

Subrini J, Turner J. 2021. Y chromosome functions in mammalian spermatogenesis. eLife 10:e67345.

Suh A, Witt CC, Menger J, Sadanandan KR, Podsiadlowski L, et al. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. Nat Commun. 2016; 7:11396.

Suh A, Bachg S, Donnellan S, Joseph L, Brosius J, et al. De-novo emergence of SINE retroposons during the early evolution of passerine birds. Mob DNA. 2017; 8:21.

Sultanova Z, Downing PA, Carazo P. 2020. Genetic sex determination and sex-specific lifespan in tetrapods–evidence of a toxic Y effect. bioRxiv.

Sun H, Jiang R, Xu, S., Zhang, Z., Xu, G., Zheng, J. and Qu, L., 2015. Transcriptome responses to heat stress in hypothalamus of a meat-type chicken. J Anim Sci Biotechnol 6, 6.

Sun J, Chen T, Zhu M, Wang R, Huang Y, Wei Q, et al. Whole-genome sequencing revealed genetic diversity and selection of Guangxi indigenous chickens. PLoS ONE. 2022;17(3):e0250392.

Sun Q, Song YJ, Prasanth KV. One locus with two roles: microRNA-independent functions of microRNA-host-gene locus-encoded long noncoding RNAs. Wiley Interdiscip Rev RNA. 2021 May;12(3):e1625.

Sun SH, Xie LH, Zhuo X, Chen Q, Ghoneim D, et al. Domestic chickens activate a piRNA defense against avian leukosis virus. Elife. 2017; 6:e24695.

Sun YC, Chen X, Fischer S, Lu S, Zhan H, Gillis J, Zador AM. Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. Nat Neurosci. 2021:873-885.

Sun YH, Xie LH, Zhuo X, Chen Q, Ghoneim D, Zhang B et al. Domestic chickens activate a piRNA defense against avian leukosis virus. eLife. 2017;6:e24695.

Symonov R. Integrative rDNAomics—importance of the oldest repetitive fraction of the eukaryote genome. Genes (Basel). 2019;10(5):345.

Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 2018, 562(7727), 367–372.

Tabula Sapiens Consortium. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. Science 2022;376(6594):eabl4896.

Talebi R, Szmatoła T, Mészáros G, Qanbari S. Runs of homozygosity in modern chicken revealed by sequence data. G3 (Bethesda). 2020;10(12):4615–4623.

Tan X, Liu L, Liu X, Cui H, Liu R, Zhao G, et al. Large-scale whole genome sequencing study reveals genetic architecture and key variants for breast muscle weight in native chickens. Genes (Basel). 2021;13(1):3.

Taylor HA, Delany ME. Ontogeny of telomerase in chicken: impact of downregulation on pre- and postnatal telomere length in vivo. Dev Growth Differ. 2000; 42:613-621.

Taylor L, Carlson DF, Nandi S, Sherman A, Fahrenkrug SC, McGrew MJ. Efficient TALEN-mediated gene targeting of chicken primordial germ cells. Development. 2017;144(5):928–934.

Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, et al. Molecular topography of an entire nervous system. Cell. 2021;184(16):4329-4347.

Tegla MGG, Buenaventura DF, Kim DY, Thakurdin C, Gonzalez KC, Emerson MM. OTX2 represses sister cell fate choices in the developing retina to promote photoreceptor specification. eLife. 2020;9:e54279.

Thoraval P, Afanassieff M, Bouret D, Luneau G, Esnault E, Goto RM, et al. Role of nonclassical class I genes of the chicken major histocompatibility complex Rfp-Y locus in transplantation immunity. Immunogenetics. 2003 Dec;55(9):647-51.

Tiley GP, Pandey A, Kimball RT, Braun EL, Burleigh JG. Whole genome phylogeny of Gallus: introgression and data-type effects. Avian Res. 2020;11:7.

Tizard M, Hallerman E, Fahrenkrug S, Newell-McGloughlin M, Gibson J, de Loos F, et al. Strategies to enable the adoption of animal biotechnology to sustainably improve global food safety and security. Transgenic Res. 2016; 25, 575-595.

Tomaszkiewicz M, Medvedev P, Makova KD. 2017. Y and W chromosome assemblies: approaches and discoveries. Trends Genet. 33:266-282.

Torgasheva A, Malinovskaya L, Zadesenets KS, Slobodchikova A, Shnaider E, Rubtsov N, et al. Highly conservative pattern of sex chromosome synapsis and recombination in Neognathae Birds. Genes (Basel) 2021;12:1358.

Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2011;13(1):36-46.

Tregaskes CA, Kaufman J. Chickens as a simple system for scientific discovery: The example of the MHC. Mol Immunol. 2021;135:12-20.

Tregaskes CA, Martin BJ, Huynh L, Rocos N, Kaufman J. 600 alleles and 200 haplotypes identified for the chicken BF-BL region. This volume, in press.

Tu WL, Cheng CY, Wang SH, Tang PC, Chen CF, Chen HH, et al. 2016. Profiling of differential gene expression in the hypothalamus of broiler-type Taiwan country chickens in response to acute heat stress. Theriogenology 85, 483-494 e8.

Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. 2020. Genenames.org: The HGNC and VGNC Resources in 2021. Nucleic Acids Res. 49 (D1): D939–46.

Tzschentke B, Basta D. 2002. Early development of neuronal hypothalamic thermosensitivity in birds: influence of epigenetic temperature adaptation. Comp Biochem Physiol A Mol Integr Physiol 131, 825-32.

Ulfah M, Kawahara-Miki R, Farajalllah A, Muladno M, Dorshorst B, Martin A et al. Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam. BMC Genomics. 2016;17:320.

UNDESA (United Nations Department of Economic and Social Affairs, Population Division). World Population Prospects: The 2017 Revision. 2017; New York: United Nations. Available at: https://www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html.

Uno Y, Nishida C, Tarui H, et al (2012) Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. PLoS One 7:e53027.

Van De Lavoir MC, Diamond JH, Leighton PA, Mather-Love C, Heyer BS, Bradshaw R, et al. Germline transmission of genetically modified primordial germ cells. Nature. 2006;441(7094):766–769.

van Dijk M, Visser A, Buabeng KML, Poutsma A, van der Schors RC, Oudejans CBM. Mutations within the LINC-HELLP non-coding RNA differentially bind ribosomal and RNA splicing complexes and negatively affect trophoblast differentiation. Hum Mol Genet. 2015;24(19):5475–85.

Van Goor A, Bolek KJ, Ashwell, C.M., Persia, M.E., Rothschild, M.F., Schmidt, C.J. et al. 2015. Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress. Genet Sel Evol 47, 96.

Vegesna R, Tomaszkiewicz M, Ryder OA, Campos-Sánchez R, Medvedev P, DeGiorgio M, et al. 2020. Ampliconic genes on the great ape Y chromosomes: Rapid evolution of copy number but conservation of expression levels. Genome Biol Evol 12:842-859.

Veller C, Kleckner N, Nowak MA (2019) A rigorous measure of genome-wide genetic shuffling that takes into account crossover positions and Mendel's second law. Proc Natl Acad Sci U S A. 116:1659-1668.

Vermillion KL, Bacher R, Tannenbaum AP, Swanson S, Jiang P, Chu LF, et al. Spatial patterns of gene expression are unveiled in the chick primitive streak by ordering single-cell transcriptomes. Dev Biol. 2018;439(1):30-41.

Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. Curr Opin Microbiol. 2015;23:148-54.

Viale E, Zanetti E, Özdemir D, Broccanello C, Dalmasso A, De Marchi M, et al. Development and validation of a novel SNP panel for the genetic characterization of Italian chicken breeds by next-generation sequencing discovery and array genotyping. Poult Sci. 2017;96(11):3858–3866.

Vignal A, Boitard S, Thebault N, Dayo G-K, Yapi-Gnaore V, Karim IYA, et al. A guinea fowl genome assembly provides new evidence on evolution following domestication and selection in galliformes. Mol Ecol Resour. 2019; 19:997–1014.

Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, et al. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A. 2006 Jan 31;103(5):1434-9.

Wang C, Habier D, Peiris BL, Wolc A, Kranis A, Watson KA, et al. 2013. Accuracy of genomic prediction using an evenly spaced, low-density single nucleotide polymorphism panel in broiler chickens. Poult Sci 92:1712-1723.

Wang H, Iacoangeli A, Popp S, Muslimov IA, Imataka H, Sonenberg N, et al. Dendritic BC1 RNA: Functional Role in Regulation of Translation Initiation. J Neurosci. 2002;22(23):10232–41.

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17(11):1665–1674.

Wang K, Hu H, Tian Y, Li J, Scheben A, Zhang C, et al. The Chicken Pan-Genome Reveals Gene Content Variation and a Promoter Region Deletion in IGF2BP1 Affecting Body Size. Mol Biol Evol. 2021;38(11):5066–5081.

Wang L, Li J, Zhou H, Zhang W, Gao J, Zheng P. A novel lncRNA Discn fine-tunes replication protein A (RPA) availability to promote genomic stability. Nat Commun. 2021a Sep 22;12(1):5572.

Wang M, Windgassen D, Papoutsakis, E.T., 2008. A global transcriptional view of apoptosis in human T-cell activation. BMC Med Genomics 1, 53.

Wang MS, Li Y, Peng MS, Zhong L, Wang ZJ, Li QY, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Mol Biol Evol. 2015;32(7):1880–1889.

Wang MS, Huo YX, Li Y, Otecko NO, Su LY, Xu HB, et al. Comparative population genomics reveals genetic basis underlying body size of domestic chickens. J Mol Cell Biol. 2016a; 8(6):542-552.

Wang MS, Zhang RW, Su LY, Li Y, Peng MS, Liu HQ, et al. Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. Cell Res. 2016b;26(5):556–573.

Wang MS, Otecko NO, Wang S, Wu DD, Yang MM, Xu YL, et al. An evolutionary genomic perspective on the breeding of dwarf chickens. Mol Biol Evol. 2017;34(12):3081–3088.

Wang MS, Thakur M, Peng MS, Jiang Y, Frantz LAF, Li M, et al. 863 genomes reveal the origin and domestication of chicken. Cell Res. 2020;30(8):693–701.

Wang MS, Zhang JJ, Guo X, Li M, Meyer R, Ashari H, et al. Large-scale genomic analysis reveals the genetic cost of chicken domestication. BMC Biol. 2021;19(1):118.

Wang Q, Li D, Guo A, Li M, Li L, Zhou J, et al. Whole-genome resequencing of Dulong Chicken reveal signatures of selection. Br Poult Sci. 2020;61(6):624–631.

Wang S, Wang Y, Li Y, Xiao F, Guo H, Gao H, et al. Genome-wide association study and selective sweep analysis reveal the genetic architecture of body weights in a chicken F2 resource population. Front Vet Sci. 2022;9:875454.

Wang W, Zhang T, Wang J, Zhang G, Wang Y, Zhang Y, et al. Genome-wide association study of 8 carcass traits in Jinghai Yellow chickens using specific-locus amplified fragment sequencing technology. Poult Sci. 2016;95(3):500–506.

Wang WH, Wang JY, Zhang T, Wang Y, Zhang Y, Han K. Genome-wide association study of growth traits in Jinghai Yellow chicken hens using SLAF-seq technology. Anim Genet. 2019;50(2):175-176.

Wang X, You X, Langer JD, Hou J, Rupprecht F, Vlatkovic I, et al. Full-length transcriptome reconstruction reveals a large diversity of RNA and protein isoforms in rat hippocampus. Nat Commun. 2019;10(1):5009.

Wang X, Liu C, Zhang S, Yan H, Zhang L, Jiang A, et al. N6-methyladenosine modification of MALAT1 promotes metastasis via reshaping nuclear speckles. Developmental Cell. 2021b;56(5):702-715.e8.

Wang Y, Sun L, Wang L, Liu Z, Li Q, Yao B, et al. Long non-coding RNA DSCR8 acts as a molecular sponge for miR-485-5p to activate Wnt/β-catenin signal pathway in hepatocellular carcinoma. Cell Death Dis. 2018;9(9):851.

Wang Y, Saelao P, Chanthavixay, K., Gallardo, R., Bunn, D., Lamont, S.J., Dekkers, J.M., Kelly, T. and Zhou, H., 2018. Physiological responses to heat stress in two genetically distinct chicken inbred lines. Poult Sci. 97, 770-780.

Wang Y, Saelao P, Kern C, Jin, S., Gallardo, R.A., Kelly, T., Dekkers, J.M., Lamont, S.J. and Zhou, H., 2020. Liver Transcriptome Responses to Heat Stress and Newcastle Disease Virus Infection in Genetically Distinct Chicken Inbred Lines. Genes (Basel) 11.

Wang Z, Qu L, Yao J, Yang X, Li G, Zhang Y et al. An EAV-HP insertion in 5' Flanking region of SLCO1B3 causes blue eggshell in the chicken. PLoS Genet. 2013;9(1):e1003183.

Wang Z, Zhang J, Yang W, An N, Zhang P, Zhang G, Zhou Q. 2014. Temporal genomic evolution of bird sex chromosomes. BMC Evol Biol. 14:1-12.

Warmuth VM, Weissensteiner MH, Wolf JB. 2022. Accumulation and ineffective silencing of transposable elements on an avian W Chromosome. Genome Res. 32:671-681.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, et al. The genome of a songbird. Nature. 2010;464(7289):757-62.

Warren WC, Fedrigo O, Tracey A, Mason AS, Formenti G, Perini F et al. Multiple chicken (Gallus gallus) genome references to advance variation studies. Cytogenet Genome Res

Warren WC, Hillier LW, Tomlinson C, Minx P, Kremitzki M, Graves T, et al. A new chicken genome assembly provides insight into avian genome structure. G3 (Bethesda). 2017;7(1):109–117.

Waters PD, Patel HR, Ruiz-Herrera A, Alvarez-Gonzalez L, Lister NC, Simakov O, et al. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. Proc Natl Acad Sci U S A 118(45):e2112494118 (2021).

Weishampel DO (2004) The Dinosauria: Second Edition. University of California Press

Weng Z, Wolc A, Su H, Fernando RL, Dekkers JCM, Arango J, et al. Identification of recombination hotspots and quantitative trait loci for recombination rate in layer chickens. J Anim Sci Biotechnol. 2019;10:20.

Whyte J, Glover JD, Woodcock M, Brzeszczynska J, Taylor L, Sherman A, et al. FGF, insulin, and SMAD signaling cooperate for avian primordial germ cell self-renewal. Stem Cell Rep. 2015;5(6):1171–1182.

Whyte J, Blesbois E, McGrew MJ. Increased sustainability in poultry production: new tools and resources for genetic management. In Sustainable Poultry Production in Europe (ed. E. Burton, J. Gatcliffe, H. Masey O'Neill and D. Scholey), pp. 214-234. 2016; Cambridge, MA, USA: CABI Publishing.

Wicker T, Robertson JS, Schulze SR, Feltus FA, Magrini V, et al. The repetitive landscape of the chicken genome. Genome Res. 2005; 15:126-136.

Williams RM, Lukoseviciute M, Sauka-Spengler T, Bronner ME. Single-cell atlas of early chick development reveals gradual segregation of neural crest lineage from the neural plate border during neurulation. eLife. 2022;11:e74464.

Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, et al. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science. 2005 Sep 2;309(5740):1570–3.

Windau, K, BC. Viertlboeck, TW Göbel. 2013. The Turkey Ig-Like Receptor Family: Identification, Expression and Function. PLoS One 8 (3): e59577.

Witmer LM (2002) The Debate on Avian Ancestry. Mesozoic Birds

Wolc A, C Stricker, J Arango, P. Settar, J.E. Fulton, N. O'Sullivan, et al. 2011. Breeding Value Prediction For Production Traits In Layers Using Pedigree And Marker Based Methods. Genet Sel Evol. 43:5.

Wolc A, Arango J, Jankowski T, Dunn I, Settar P, Fulton JE, et al. Genome-wide association study for egg production and quality in layer chickens. J Anim Breed Genet. 2014;131(3):173–182.

Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. 2015. Response and inbreeding from a genomic selection experiment in layer chickens. Genetics Selection Evolution 47:59.

Wolc A, Kranis A, Arango J, Settar P., Fulton J.E., O'Sullivan N.P., et al. 2016. Implementation of genomic selection in the poultry industry. Animal Frontiers 6(1):23-31.

Wolc A, W Drobik-Czwarno, JE Fulton, J. Arango, T. Jankowski, and J. C. M. Dekkers. 2018. Genomic prediction of avian influenza infection outcome in layer chickens. Genet Sel Evol, 50: 21.

Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Dekkers JCM. 2019. Genome wide association study for heat stress induced mortality in a white egg layer line. Poult Sci 98, 92-96.

Wolc A, Drobik-Czwarno W, Jankowski T, Arango J, Settar P, Fulton JE, et al. Accuracy of genomic prediction of shell quality in a White Leghorn line. Poult Sci. 2020;99(6):2833–2840.

Wolc A, Settar P, Fulton JE, Arango J, Rowland K, Lubritz D, et al. 2021. Heritability of perching behavior and its genetic relationship with incidence of floor eggs in Rhode Island Red chickens. Genetics, selection, evolution: GSE, 53(1), 38.

Wolc A, Li J, Lubritz D, Arango J, Fulton J, Settar P, et al. 2022. Application of low-pass sequencing to genomic prediction of egg quality in laying hens. World Congress on Genetics Applied to Livestock Production 3-8.07.2022 Rotterdam, The Netherlands

Woodcock ME, Idoko-Akoh A, McGrew MJ. Gene editing in birds takes flight. Mamm Genome 2017; 28, 315–323.

Woodcock ME, Gheyas AA, Mason AS, Nandi S, Taylor L, Sherman A, et al. Reviving rare chicken breeds using genetically engineered sterility in surrogate host birds. Proc Natl Acad Sci U S A, 2019; 116, 20930–20937.

Wragg D, Mwacharo JM, Alcalde JA, Wang C, Han JL, Gongora J et al. Endogenous retrovirus EAV-HP linked to blue egg phenotype in Mapuche fowl. PloS One. 2013;8(8):e71393.

Wright A, Mank J. 2013. The scope and strength of sex-specific selection in genome evolution. J Evol Biol. 26:1841-1853.

Wright AE, Harrison PW, Montgomery SH, Pointer MA, Mank JE. 2014. Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. Evolution 68:3281-3295.

Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. Nat Commun. 7:1-8.

Wright MW. A short guide to long non-coding RNA gene nomenclature. Hum Genomics. 2014;8(1):7.

Wu Z, Hu T, Chintoan-Uta C, Macdonald J, Stevens MP, Sang H, et al. Development of novel reagents to chicken FLT3, XCR1 and CSF2R for the identification and characterization of avian conventional dendritic cells. Immunology. 2022;165(2):171–194.

Xavier CP, Eichinger L, Fernandez MP, Morgan RO, Clemen CS. Evolutionary and functional diversity of coronin proteins. Subcell Biochem. 2008;48:98-109.

Xirocostas ZA, Everingham SE, Moles AT. 2020. The sex with the reduced sex chromosome dies earlier: a comparison across the tree of life. Biol Lett. 16:20190867.

Xu L, Zhou Q. 2020. The female-specific W chromosomes of birds have conserved gene contents but are not feminized. Genes 11:1126.

Xu L, Auer G, Peona V, Suh A, Deng Y, Feng S, et al. 2019. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. Nat Ecol Evol. 3:834-844.

Xu NY, Liu ZY, Yang QM, Bian PP, Li M, Zhao X. Genomic analyses for selective signatures and genes involved in hot adaptation among indigenous chickens from different tropical climate regions. Front Genet. 2022;13:906447.

Xu T p, Liu X x, Xia R, Yin L, Kong R, Chen W m, et al. SP1-induced upregulation of the long noncoding RNA TINCR regulates cell proliferation and apoptosis by affecting KLF2 mRNA stability in gastric cancer. Oncogene. 2015 Nov;34(45):5648–61.

Yamagata M. Towards Tabula Gallus. Int J Mol Sci. 2022;23(2):613.

Yamagata M, Sanes JR. CRISPR-mediated Labeling of Cells in Chick Embryos Based on Selectively Expressed Genes. Bio Protoc. 2021;11(15):e4105.

Yamagata M, Yan W, Sanes JR. A cell atlas of the chick retina based on single-cell transcriptomics. eLife. 2021;10:e63907.

Yamazaki T, Souquere S, Chujo T, Kobelke S, Chong YS, Fox AH, et al. Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. Mol Cell. 2018;70(6):1038-1053.e7.

Yan C, Chen J, Chen N. Long noncoding RNA MALAT1 promotes hepatic steatosis and insulin resistance by increasing nuclear SREBP-1c protein stability. Sci Rep. 2016;6:22640.

Yang KX, Zhou H, Ding JM, He C, Niu Q, Gu CJ, et al. Copy number variation in HOXB7 and HOXB8 involves in the formation of beard trait in chickens. Anim Genet. 2020;51(6):958–963.

Yang Z, Deng J, Li D, Sun T, Xia L, Xu W, et al. Analysis of population structure and differentially selected regions in Guangxi native breeds by restriction site associated with DNA sequencing. G3 (Bethesda). 2020;10(1):379–386.

Yang Z, Zou L, Sun T, Xu W, Zeng L, Jia Y, et al. Genome-wide association study using whole-genome sequencing identifies a genomic region on chromosome 6 associated with comb traits in Nandan-Yao chicken. Front Genet. 2021;12:682501.

Yasugi S, Nakamura H. Gene transfer into chicken embryos as an effective system of analysis in developmental biology. Dev. Growth Differ. 2000; 42, 195–197.

Yazdi HP, Ellegren H. 2018. A Genetic Map of Ostrich Z Chromosome and the Role of Inversions in Avian Sex Chromosome Evolution. Genome Biol Evol. 10:2049-2060.

Ye L, Hillier LW, Minx P, Thane N, Locke DP, Martin JC, et al. A vertebrate case study of the quality of assemblies derived from next-generation sequences. Genome Biol. 2011;12(3):R31.

Yi G, Qu L, Chen S, Xu G, Yang N. Genome-wide copy number profiling using high-density SNP array in chickens. Anim Genet. 2015;46(2):148–157.

Yin Z, Zhang F, Smith J, Kuo R, Hou ZC. Full-length transcriptome sequencing from multiple tissues of duck, Anas platyrhynchos. Sci Data. 2019;6(1):275.

Yin ZT, Zhu F, Lin FB, Jia T, Wang Z, Sun DT, et al. Revisiting avian 'missing' genes from de novo assembled transcripts. BMC Genomics. 2019;20(1):4.

Yu Y, Nangia-Makker P, Farhana L, Majumdar APN. A novel mechanism of lncRNA and miRNA interaction: CCAT2 regulates miR-145 expression by suppressing its maturation process in colon cancer cells. Mol Cancer. 2017;16:155.

Yuan J, Li S, Sheng Z, Zhang M, Liu X, Yuan Z, et al. Genome-wide run of homozygosity analysis reveals candidate genomic regions associated with environmental adaptations of Tibetan native chickens. BMC Genomics. 2022;23(1):91.

Yuan X, Cui H, Jin Y, Zhao W, Liu X, Wang Y, et al. Fatty acid metabolism-related genes are associated with flavor-presenting aldehydes in Chinese local chicken. Front Genet. 2022;13:902180.

Zeng H. What is a cell type and how to define it? Cell. 2022;185(15):2739-2755.

Zeng T, Yin J, Feng P, Han F, Tian Y, Wang Y, et al. Analysis of genome and methylation changes in Chinese indigenous chickens over time provides insight into species conservation. Commun Biol. 2022;5(1):952.

Zentner GE, Saiakhova A, Manaenkov P, Adams MD, Scacheri PC. Integrative genomic analysis of human ribosomal DNA. Nucl Acids Res. 2011;39(12):4949–4960.

Zhai Z, Zhao W, He C, Yang K, Tang L, Liu S, et al. SNP discovery and genotyping using restriction-site-associated DNA sequencing in chickens. Anim Genet. 2015;46(2):216–219.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014;346(6215):1311-20.

Zhang G, Rahbek C, Graves GR, Lei F, Jarvis ED, Gilbert MT. Genomics: Bird sequencing project takes off. Nature. 2015;522(7554):34.

Zhang H, Du ZQ, Dong JQ, Wang HX, Shi HY, Wang N, et al. Detection of genome wide copy number variations in two chicken lines divergently selected for abdominal fat content. BMC Genomics. 2014;15:517.

Zhang J, Kaiser MG, Deist, M.S., Gallardo, R.A., Bunn, D.A., Kelly, T.R., et al. Transcriptome Analysis in Spleen Reveals Differential Regulation of Response to Newcastle Disease Virus in Two Chicken Lines. Sci Rep. 2018;8:1278.

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

Zhang J, Lv C, Mo C, Liu M, Wan Y, Li J, Wang Y. Single-cell RNA sequencing analysis of chicken anterior pituitary: A bird's-eye view on vertebrate pituitary. Front Physiol. 2021;12:562817.

Zhang M, Han W, Tang H, Li G, Zhang M, Xu R, et al. Genomic diversity dynamics in conserved chicken populations are revealed by genome-wide SNPs. BMC Genomics. 2018;19(1):598.

Zhang Q, Gou W, Wang X, Zhang Y, Ma J, Zhang H, et al. Genome resequencing identifies unique adaptations of Tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments. Genome Biol Evol. 2016;8(3):765–776.

Zhang Y, Wang Y, Li Y, Wu J, Wang X, Bian C, et al. Genome-wide association study reveals the genetic determinism of growth traits in a Gushi-Anka F2 chicken population. Heredity (Edinb). 2021;126(2):293–307.

Zhang Z, Yang W, Zhu T, Wang L, Zhao X, Zhao G, et al. Genetic parameter estimation and whole sequencing analysis of the genetic architecture of chicken keel bending. Front Genet. 2022;13:833132.

Zhao GP, MJ Han, M. Q. Zheng, J. P. Zhao, J. L. Chen, and J. Wen. 2011. Effects of Dietary Vitamin E on Immunological Stress of Layers and Their Offspring. J Anim Physiol Animal Nutrition 95(3):343–50.

Zhao L, Wang J, Li Y, Song T, Wu Y, Fang S, et al. NONCODEV6: an updated database dedicated to long non-coding RNA annotation in both animals and plants. Nucleic Acids Res. 2021;49(D1):D165–71.

Zhao QB, Liao RR, Sun H, Zhang Z, Wang QS, Yang CS, et al. Identifying genetic differences between Dongxiang Blue-Shelled and White Leghorn chickens using sequencing data. G3 (Bethesda). 2018;8(2):469–476.

Zhao S, Zhang B. A comprehensive evaluation of Ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. BMC Genomics. 2015;16(1):97.

Zhao S, Zhang Y, Gamini R, Zhang B, von Schack D. 2018. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. Sci Rep. 8:4781.

Zheng KW, Zhang JY, He Y, Gong JY, Wen CJ, et al. Detection of genomic G-quadruplexes in living cells using a small artificial protein. Nucleic Acids Res. 2020; 48:11706-11720.

Zhou H, Lamont SJ. 1999. Genetic characterization of biodiversity in highly inbred chicken lines by microsatellite markers. Anim Genet. 30:256-264.

Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 2014;346:1246338.

Zhou W, Liu R, Zhang J, Zheng M, Li P, Chang G, et al. A genome wide detection of copy number variation using SNP genotyping arrays in Beijing-You chickens. Genetica. 2014;142(5):441–450.

Zhou Z. (2004) The origin and early evolution of birds: discoveries, disputes, and perspectives from fossil evidence. Naturwissenschaften 91:455–471.

Zhu F, Yin ZT, Wang Z, Smith J, Zhang F, Martin F, et al. Three chromosome-level duck genome assemblies provide insights into genomic variation during domestication. Nat Commun. 2021;12(1):5932.

Zhu L, Wei Q, Qi Y, Ruan X, Wu F, Li L, et al. PTB-AS, a Novel Natural Antisense Transcript, Promotes Glioma Progression by Improving PTBP1 mRNA Stability with SND1. Mol Ther. 2019;27(9):1621–37.

Zhu XJ, Sun S, Xie B, Hu X, Zhang Z, et al. Guanine-rich sequences inhibit proofreading DNA polymerases. Sci Rep. 2016; 6:28769.

Zhu Y, Qu J, He L, Zhang F, Zhou Z, Yang S, et al. 2019. Calcium in Vascular Smooth Muscle Cell Elasticity and Adhesion: Novel Insights Into the Mechanism of Action. Front Physiol. 10, 852.

Zhuang X. Spatially resolved single-cell genomics and transcriptomics by imaging. Nat. Methods. 2021;18(1):18-22

Zickler D, Kleckner N: Recombination, pairing, and synapsis of homologs during meiosis. Cold Spring Harb Perspect Biol 2015;7:a016626.

Zuo Q, Wang Y, Cheng S, Lian C, Tang B, Wang F, Z. Lu, Y. Ji, R. Zhao, W. Zhang, K. Jin, J. Song, Y. Zhang, and B. Li. 2016. Site-Directed Genome Knockout in Chicken Cell Line and Embryos Can Use CRISPR/Cas Gene Editing Technology. G3 (Bethesda), 6:1787-92.

Figure Legends

Fig. 1. Assembled structural errors detected in RJF compared to broiler for chromosome 27. Genetic linkage map markers (n=125) displayed as green tic marks below the x-axis for the chromosome 27 heat map were mapped to each assembly to validate sequence order and orientation. Generated with BioRender software.

Fig. 2. Sequenced differences in the phased broiler and layer genomes for A) large and B) small autosomes. From the inside out SNV density (red), window size of 500kb, range of 0 to 2.5%, indels <50bp (coral), 500kb window size and 0-0.8%; large indels (blue) per Mb, range of 0 to 60; CNV count per Mb (green); highlighted inversions (black dashes); chicken karyotype (varied color); ideograms of GRCg7b and GRCg7w chromosomes (varied colors). Generated with BioRender software.

Fig. 3. The distribution of called heterozygous SNVs across chicken macrochromosome 7 and microchromosome 20. Rainfall plots of heterozygous variants depict their location, and each unique color indicates a different type of

base substitution. We only include variants that passed all filters and were heterozygous in either reference source.Generated with BioRender software.

Fig. 4. The RNAseq alignment detection of multimapping events and rRNA number and size distributions by reference source.Generated with BioRender software.

Fig. 5. ALVE integration, propagation, and degradation within the chicken genome. A) shows the retroviral genomic lifecycle. Retroviral positive sense, single stranded RNA is reverse transcribed into cDNA and associates with the retroviral integrase integration complex, which primes the cDNA 3' ends and initiates strand transfer with genomic DNA. Integration creates overhangs which are repaired by host machinery, creating target site duplications (TSDs; grey). Following integration, retroviral expression and retrotransposition is possible. Over evolutionary timescales integrated ERVs degrade, either by non-homologous recombination events (I, II) or internal LTR recombination leaving solo LTRs (III). The top schematic in B) indicates an intact ALVE with putative transcripts, with the ribosomal -1 frame slip and recognition site for miR-155 indicated. Phased chicken genome ALVE content and integrity is shown, with likely transcript and regulatory implications. CA: capsid; INT: integrase; LTR: long terminal repeat; MA: matrix; NC: nucleocapsid; PR: protease; RH: RNaseH; RT: reverse transcriptase; SU: surface; TM: transmembrane. Generated with BioRender software.

Fig. 6. Evolution of major groups of reptiles (including birds) with major extinction events noted. Timelines given but scale is not linear.

Fig. 7. Chromosome evolution from the diapsid ancestor, to the archelosaur ancestor, via the theropod dinosaur lineage, to modern birds and finally to chicken. A recognisable avian pattern had evolved just before the dinosaurs emerged 240 million years ago. After this time, chromosome inversions were the principal mechanisms of change. Print and pdf version: a) 275 million years ago; b) 255 million years ago; c) 240 million years ago at the dawn of the dinosaurs; d) snapshot 100-150 years ago; e) modern chicken

Fig. 8. Imagined dinosaur karyotype (deliberately manipulated image based on chicken and spiny soft shelled turtle chromosomes)

Fig. 9. Bird whole genome alignments. A) Alignment of chicken to zebra finch (Rhie et al. 2021), Anna's hummingbird [Rhie et al. 2021], superb fairywren [Penalba et al. 2020], and the ancestral emu genome [Liu et al. 2021]. B) Alignment of the chicken genome to species that have undergone significant chromosome rearrangement: saker falcon, California condor [Robinson et al. 2021] and golden eagle [Mead et al. 2021].

Fig. 10. Different mechanisms of lncRNA roles. Effects at the nuclear and telomere (A-B), transcriptional (C-D), post-transcriptional (E-H), translational (I), and post-translational levels (J-L). Role as small ORF host (M) and small non-coding RNA host (N). Implication in the exosome-mediated transfer (O). In purple, lncRNA; in blue, DNA; in green, other RNAs; in dark red, proteins. For more examples, other genes are presented in Muret et al. 2019 [Muret et al. 2019] specifically for genes involved in the regulation of lipid metabolism and their regulatory mechanisms.

Fig. 11. Assembly versions associated with the chicken genome (A) and the number of publications associated with them (B). (A) RJF: Red Jungle fowl; W-L: White Leghorn; Cov.: coverage; Scaf: scaffold; N50: scaffold N50. A supplementary figure (Supp. Tab. 1) is provided for more details. (B) Blue number: articles published using the corresponding assembly during the year 2020. Identification was made on PubMed Central by searching for the assembly name in different formats (*e.g.* GRCg6a or GRCg6 or Galgal6).

Fig. 12. Gene numbers provided by NCBI's RefSeq and EMBL-EBI's Ensembl according to the genome annotation and genome assembly versions (A) and transcript model changes between two genome annotation versions from NCBI RefSeq for the same assembly – GRCg7b (B). (A) PCG: Protein-coding gene; lncRNA: Long non-coding RNA. (B) Comparison between versions 105 and 106 provided by NCBI [NCBI's RefSeq 2022]. Briefly, a score (between 0 and 1) for current and previous transcript features is calculated based on overlap in exon sequence and matches in exon boundaries. Pairs of current and previous features were categorized based on these scores and considering changes in attributes. New: New transcript models; Deprecated: Transcripts removed or merged in the new version; Major changes: Changes with great impact on the sequence or on the transcript attributes; Minor changes: Minimal change ensuring similarity.

Fig. 13. Features of the current NCBI RefSeq (v106) and EMBL-EBI Ensembl (v107) genome annotations based on the latest GRCg7b genome assembly. (B) The transcript models were compared between the two annotations according to 4 main classes (Equal isoform; New isoform; New loci and Artifacts) according to the software "GffCompare" (options: -S --no-merge) [Pertea and Pertea 2020]. (C) #Tr/gn: Number of transcripts per gene; #Ex/tr: Number of exons per transcript; Tr. Size: Transcripts size considering only exonic regions; Ex. Size: Exons size; In. Size: Introns size. The median transcript sizes between RefSeq & Ensembl are *resp.* 3,465 pb *vs.* 2,317,

p-value < 10^{-16} (Wilcoxon rank sum test); for PCG *resp.* 3,634 pb *vs.* 2,870 pb, *: p-value < 10^{-16}; for lncRNAs *resp.* 2,952 pb *vs.* 1,487, *: p-value < 10^{-16}.

Fig. 14. Numbers of differentially expressed genes identified between genetic lines. A false discovery rate < 0.05 was used to classify genes as differentially expressed. AH: Acute heat stress; CH: Chronic heat stress and NDV infection at 2dpi; FCLC: Fayoumi non-treated vs. Leghorn non-treated; FTLT: Fayoumi treated vs. Leghorn treated.

Fig. 15. Numbers of differentially expressed genes identified within genetic lines. A false discovery rate < 0.05 was used to classify genes as differentially expressed. AH: Acute heat stress; CH: Chronic heat stress and NDV infection at 2dpi; LTLC: Leghorn treated vs. non-treated; FTFC: Fayoumi treated vs. non-treated.

Fig. 16. Venn diagram of differentially expressed genes within genetic lines in the hypothalamus. LTLCAH: Leghorn treated vs. non-treated with acute heat stress; FTFCAH: Fayoumi treated vs. non-treated with acute heat stress; LTLCCH&NDV: Leghorn treated vs. non-treated with chronic heat stress and 2dpi NDV infection; FTFCCH&NDV: Fayoumi treated vs. non-treated with chronic heat stress and 2dpi NDV infection.

Fig. 17. Venn diagram of differentially expressed genes within genetic lines in the breast muscle. LTLCAH: Leghorn treated vs. non-treated with acute heat stress; FTFCAH: Fayoumi treated vs. non-treated with acute heat stress; LTLCCH&NDV: Leghorn treated vs. non-treated with chronic heat stress and 2dpi NDV infection; FTFCCH&NDV: Fayoumi treated vs. non-treated with chronic heat stress and 2dpi NDV infection.

Fig. 18. Gene ontology (GO) biological processes and KEGG pathway overrepresentation (P<0.05 and FDR < 20%) for within-line comparisons in the hypothalamus. (a) GO terms and KEGG pathways significantly enriched by upregulated genes in the LTLCAH comparison. (b) GO terms and KEGG pathways significantly enriched by downregulated genes in the LTLCAH comparison. (c) GO terms and KEGG pathways significantly enriched by downregulated genes in the LTLCCH comparison. (d) GO terms and KEGG pathways significantly enriched by upregulated genes in the FTFCAH comparison. (e) GO terms and KEGG pathways significantly enriched by down-regulated genes in the FTFCAH comparison. (f) GO terms and KEGG pathways significantly enriched by upregulated genes in the FTFCCH comparison. (g) GO terms and KEGG pathways significantly enriched by downregulated genes in the FTFCCH comparison.

Fig. 19. Gene ontology (GO) biological processes and KEGG pathway overrepresentation (P<0.05 and FDR < 20%) for within-line comparisons in the breast muscle. (a) GO terms and KEGG pathways significantly enriched by up-regulated genes in the LTLCAH comparison. (b) GO terms and KEGG pathways significantly enriched by downregulated genes in the LTLCAH comparison. (c) GO terms and KEGG pathways significantly enriched by upregulated genes in the FTFCAH comparison. (d) GO terms and KEGG pathways significantly enriched by upregulated genes in the LTLCCH comparison. (e) GO terms and KEGG pathways significantly enriched by downregulated genes in the LTLCCH comparison (f) GO terms and KEGG pathways significantly enriched by upregulated genes in the FTFCCH comparison. (g) GO terms and KEGG pathways significantly enriched by downregulated genes in the FTFCCH comparison.

Fig. 20. Comparative analysis of significantly enriched canonical pathways through Ingenuity Pathway Analysis among differentially expressed genes by genetic line and treatment in the breast muscle (P<0.05 and z >|1.64|), where (orange positive z-score) refers to predicted activation and blue (negative z-score) predicted of inhibition.

Fig. 21. Module-trait relationships from WGNCA. Each module (y-axis) is correlated with each phenotype (x-axis) the correlation and p-values were reported for each comparison. Strong positive correlations are coloured in red, and strong negative correlations are coloured in green. (TrtAH: Acute heat stress; TrtCH: Chronic heat stress and 2dpi NDV infection). a: correlation coefficient; b: p-value.

Fig. 22. Top GO terms and KEGG pathways enriched by genes highly expressed in the turquoise module in the breast muscle.

Fig. 23. The Blue module gene network. Red highlighted dots: immune related genes; Blue highlighted dots: metabolic genes; Gray highlight dots: all other genes in the Blue module.

Fig. 24. Phopholipase C signaling pathway and gene heat map in within-line comparisons in the breast muscle. A: The molecule activity prediction of the pathway in Leghorn birds with CH&NDV treatment; B: The molecule activity prediction of the pathway in Fayoumi birds with AH; C: The DEG heatmap matching the pathway.

Fig. 25. Calcium signaling pathway and gene heat map in within-line comparisons in the breast muscle. A: The molecule activity prediction of the pathway in Leghorn birds with CH&NDV treatment; B: The molecule activity prediction of the pathway in Fayoumi birds with AH; C: The DEG heat map matching the pathway.

Fig. 26. Sequence organization of transposable elements in avian genomes. (a) CR1 elements are composed of segments that resemble long retro-inserted messenger RNA (mRNA) with an A-rich tail at their 3' end. Within a "species" of CR1s many copies are truncated at their 5' ends. Full length elements contain two open reading

frames (ORFs) whith ORF2 encoding a protein containing an apurinic endonuclease domain fused to a reverse transcriptase. (**b**) SINEs using the CR1 machinery are present in some bird species, but not in chickens. They consist of the fusion of a former tRNA to a 3' CR1 end. (**c**) LTR retrotransposons have all the signatures of an endogenous retrovirus-like element, including long terminal repeats (LTR) at both ends and oORFs coding for a group antigen (Gag), a reverse transcriptase (RT), and in some case an envelope protein. (**d**) DNA transposons that transpose directly from DNA to DNA have short terminal inverted repeats (arrows) at both ends. When these elements are intact, they may contain a gene coding a transposase, an enzyme required for their own transposition. There are also internally deleted form such as *Galluhop* and chimeric elements such as Charlie/Galluhop.

Fig. 27. Hybridization of GC-rich and GC-poor DNA probes on chicken chromosomes. Chicken DNA fractions characterized as having the lowest and the highest GC levels were hybridized to chicken chromosomes. (**a**) The DNA fraction with the highest GC level (red signals) was localized to the microchromosomes and to telomeric bands of the macrochromosomes (see white arrows as examples). Some internal bands of the macrochromosomes also hybridized (see yellow arrows as examples). In contrast, the DNA fraction with the lowest GC levels (green signals) localized to the internal bands of the macrochromosomes. (**b**) The same metaphase shown in the panel is DAPI stained in order to better show microchromosomes. The bar in the upper right is 5 ⬚m long. Figure taken from [Federico et al., 2005].

Fig. 28. Comparison of the chicken *Gallus g. domesticus* and guinea fowl *Numida meleagris* rDNA repeat structure. The structures of chicken rDNA repeat II from the WAG137G04 BAC clone (a) and guinea fowl rDNA repeat II from the NCBI JABXER010000123 contig (b). 18S, 5.8S and 28S rRNA genes are indicated by red blocks, external (5' and 3'ETS) and internal (ITS1 and ITS2) transcribed spacers – by yellow blocks, intergenic transcribed spacers (IGS) – by green blocks. GC pair distribution is shown in the graphs as 'GC%'.

Fig. 29. Chicken and guinea fowl IGS structure. (a) Four aligned chicken IGS sequences. IGS_I, IGS_II, and IGS_III are from the *Gallus g. domesticus* BAC-clone containing rDNA (WAG137G04). IGS_IV belongs to a red jungle fowl (AADN04001305.1). All four IGS have different sizes caused by the difference in repeat blocks of each type (SV-AL, EL, VAL). The unique regions are almost of the same length in all analysed IGS (see also Suppl. Table 1). (b) Three aligned IGS sequences from the guinea fowl JABXER010000123 contig. Two of them are completely identical, IGS_III has an insertion at 5,500 bp (dotted rectangle). Each of the IGS contains no EL repeats, one SV-AL repeat block and at least two VAL repeat blocks, differentiated into six repeat variants (see also Suppl. Fig. S1, S2, Data S1). A species-specific Nme repeat block following SV-AL block is marked in black. Sequence gaps are designated with fine black lines.

Fig. 30. Internal IGS repeats demonstrate a HOR (High Order Repeats) organisation. Contracted IGS_II figure from WAG137G04 contig of *G. g. domesticus* (a) and IGS_I figure from JABXER010000123 contig of *N. meleagris* (b).

Fig. 31. Chicken spermatocyte (a) and oocyte (b) after immunolocalization of SYCP3 (red), centromeric proteins (blue) and MLH1 (green). Arrows point to the synaptonemal complexes of the macrochromosomes identified by their lengths and centromeric indices. Arrowheads indicate MLH1 signals at ZW bivalent. Bar—5 μm. Fig. 1a from Malinovskaya et al.[2019] is licensed under Creative Commons Attribution 4.0 License. Fig. 1b from Torgasheva et al.[2021] is licensed under Creative Commons Attribution 4.0 License.

Fig. 32. Immunolocalization of recombination and FISH mapping of single copy sequences on synaptonemal complexes of the chicken. **a.** immunostained chicken oocyte showing the complete set of synaptonemal complexes labeled with anti-SMC3 and the crossovers detected with anti-MLH1. The eight largest autosomal bivalents have a number next to the centromere signal (red protruding marks). The ZW pair has a single MLH1 focus located near the homologous end of the bivalent. Bar. 10 μm. **b.** Localization of BAC clone 40N14 on the synaptonemal complex of GGA1 (SC1). The BAC insert is at 1.2 Mb from the sequence start. **c.** Graph showing the distribution of distances from the end of SC1 to the FISH signal in eight pachytene nuclei (data from del Priore and Pigozzi 2021). Bar: 1 μm.

Fig. 33. Recombination rates along GGA macrochromosomes 1 to 8. Each chromosome is divided in cells of 2.5 Mb. The recombination rates are represented by the different colours as indicated in the legend. The arrowheads point at the centromere positions calculated from the centromeric indexes in synaptonemal complex spreads.

Fig. 34. MLH1 recombination maps of GGA1 and GGA5 integrated to physical positions. In both graphs the *x* axis is the length of the chromosome in Mb from the sequence start on the short arm (p) to the end on the long arm (q) with *c* indicating the centromere. A schematic representation of the chromosomes is shown below. **a.** The shaded area between the lines represents a chromosome region localized between 263 and 285 cM in a QTL analysis (Hansen et al., 2005). The physical location of this region can be predicted from distribution of the cumulative cM

distances (blue line). **b.** The shaded area near the centromere spans over 8 Mb and contains multiple QTL for two myopathies (Lake et al 2021). The MLH1-cM map shows the recombination pattern in this segment.

Fig. 35. Organisation of regions on chicken chromosome 16, as currently published. A. Depiction of chromosome 16, based on analysis by FISH, radiation hybrids, genetics, southern blotting and sequencing. B, B locus; GC, G+C rich region of PO1 repeats; Y, Rfp-Y region; NOR, nucleolar organiser region; BLA, class II A gene; fB, factor B gene; ORs, olfactory receptor genes; SRCRs, scavenger receptor with cysteine repeat genes. Double-headed arrows indicate recombination frequencies between B and BLA, fB and Rfp-Y, and B and Rfp-Y. B. Region of the B locus currently sequenced, including the BF-BL region, the TRIM region and the BG region. Genes represented by boxes. Rising and falling stripes indicate genes of the classical class I and class II presentation system, respectively; stippled indicate class II region genes; black indicates lectin-like genes and pseudogenes; horizontal stripes indicate TRIM family genes; vertical stripes indicate BG genes. Names of genes above indicate transcription from left to right, below indicate transcription from right to left; note the homologous genes in opposite transcriptional orientation in the BF-BL region but in the same transcriptional orientation in the BG region, strongly affecting the dynamics of evolution based on recombination. (Figure from Kaufman 2021).

Fig. 36. The basis for the PCR-NGS typing of the chicken MHC. Organisation of the BF-BL region and gene names from Kaufman et al 1999 (*RING3* is now known as *BRD2*); primers are designated by lab names. Figure from Martin 2021.

Fig. 37. The chicken classical class I sequences mostly separate into two large clades, while classical class II B sequences are all mixed together in phylogenetic trees. BLB* indicates class II B sequences that could not easily be assigned to the *BLB1* or *BLB2* loci in the haplotypes examined. Figure from Martin 2021, derived from data available in 2020.
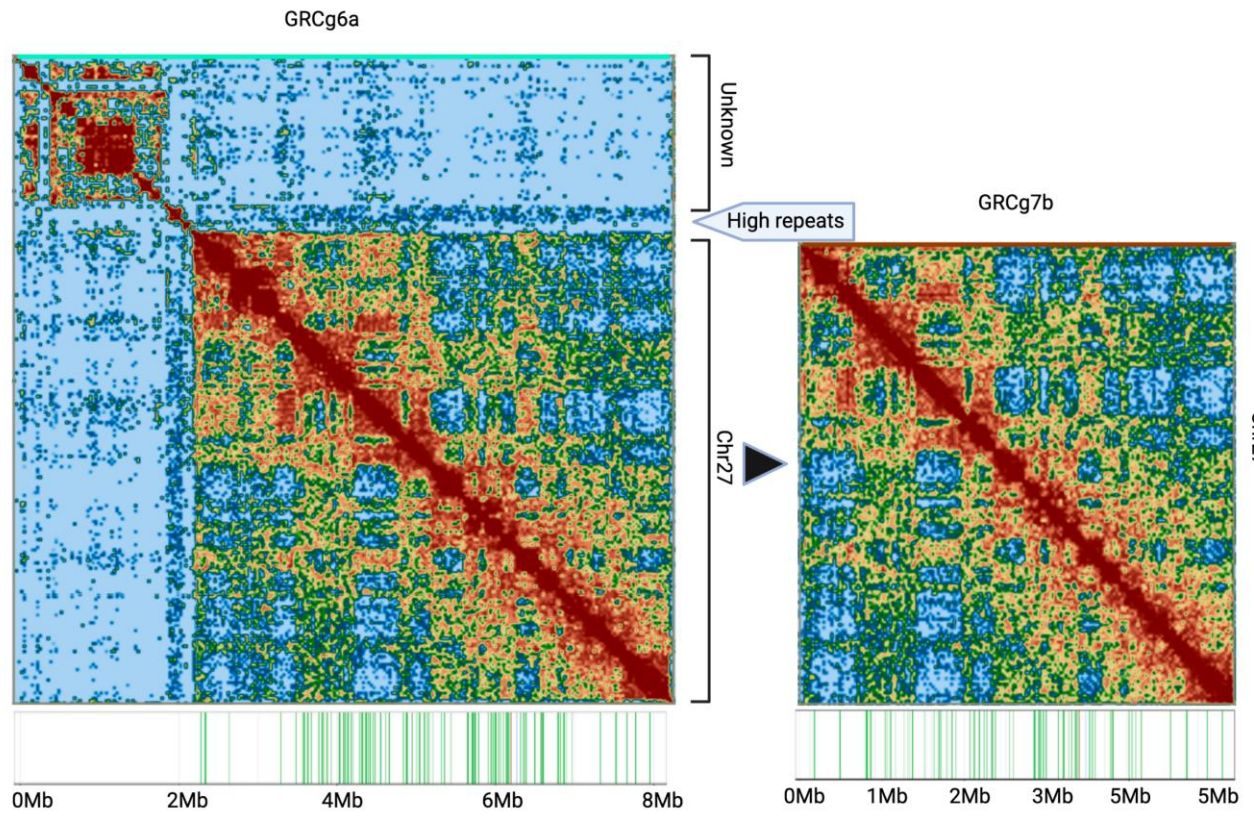
Fig. 38. Sampling of global *Gallus* spp. diversity. The map shows the sampling locations for the 4,392 genomes from domestic chickens and congeneric junglefowl species. To illustrate group size, commercial birds and European fancy breeds are also included on the map, although physical sampling location is not presumed to be important for these birds.

Fig. 39. Processing pipeline. To ensure between-sample consistency, all samples have been re-processed from raw fastq reads. Reads underwent pre-processing and quality control before mapping to the latest version of the chicken genome (GRCg7b), variant calling, and generating a VCF.

Fig. 40. Biobanking PGC and chimera chicken production at CTLGH / ILRI. Primordial germ cells are isolated from the germinal crescent of the blastoderm, from the circulating blood or from embryonic gonads at day 2.5 of their development (1), the isolated cells are immediately transferred into the freezing media (2) and kept overnight in Mr Frosty before transfer for cryopreservation in liquid nitrogen (3). For use in the case of gonadal PGCs, the embryonic gonads removed from the liquid nitrogen are dissociated, characterized, and propagated if necessary (4) or directly injected into a 2 day-old recipient embryo for gonad re-colonization. The injected egg will be incubated up to day 21 to produce the chimeric chick (6) which may be very similar to the recipient breed. Depending on the level of gonad colonization by the donor PGCs, the mature chimeric chicken may display some phenotypic characteristics of the donor breed (7).

Fig. 41. Protocols for the restoration of biobanked tropical poultry genetic resources and the potential dissemination of potential elite lines using chimeras (A1) and gene edited surrogate host technology (A2 and B).

Fig. 42. Schematic illustration for future application of genome-edited poultry to industries. Genome editing in poultry can improve disease resistance and meat productivity. By targeting egg white protein genes, genome edited poultry can economically produce protein drugs with improved biological efficacy. When the reported genes are targeted to the Z chromosome, the male embryo can be screened out before hatching by detecting fluorescence during incubation.

A   Chr1-20

B   Chr21-39

A.

Broiler - GRCg7b

Chr7

Layer - GRCg7w

Chr7

RJF - GRCg6a

Chr7

B.

Broiler - GRCg7b

Chr20

Layer - GRCg7w

Chr20

RJF - GRCg6a

Chr20

A

(+)ssRNA

cDNA

integration
complex

genomic transfer
and repair

integrated
retrovirus

retrotransposition   transcription
(+)ssRNA   (-)ssRNA

degradation

I) truncation

II) internal excision

III) LTR recombination

B

gag   env

5'LTR MA CA NC PR   RT   RH   INT   SU   TM 3'LTR
                        pol   miR-155 site

                                                 gagpol

                                                        env

Intact   ♂ ALVE21   ♀ ALVE-TYR, ALVE_ros010, ALVE_ros171

Internal modifications

♂ ALVE1   RT +1 frameshift truncating *gagpol*

♂ ALVE3   PRO/RT absent in frame

♂ ALVE_ros034   P27 nonsense mutation truncating *gagpol*, possible second ORF

♀ ALVE_ros304   LTR 16bp deletions not impacting promoter; miR-155 site mutated

Solo LTRs   ♂ ALVE15, ALVE_ros005   ♀ ALVE_ros159

Permo-
Triassic
extinction
event

Carnian–
Norian
extinction
event

End-Triassic
mass
extinction
event

Cretaceous-
Paleogene (K-Pg)
mass extinction
event

275 mya >255 mya ≤255 mya 251 mya   245-240 mya 240-235 mya 228 mya   201 mya   170-165 mya 66 mya Present day

Timeline is not to scale in
order to allow more clarity
(mya=million years ago)

Diapsida

Lepidosauromorpha

Squamates
(Snakes & Lizards)

Sphenodontia
(Tuatara)

Archelosauria

Testudines
(Turtles)

Archosauromorpha/
Archosauria

Pseudosuchia
(Crocodilians)

Avemetatarsalia

Pterosauria

DINOSAURIA

Saurischia

Ornithoscelida

Ornithischia

Theropoda

Maniraptora

Avialae

Neornithines
(Modern Birds)

A.

B.

C.

Ancestor to Avian Genome
- via Dinosaur Lineage

A

Superb fairywren
Anna's hummingbird
Zebra finch
Chicken
Emu

B

Golden Eagle
California condor
Chicken
Saker falcon

A. nuclear organization

speckles
paraspeckles

B. telomere stability

C. modulation of chromatin signatures

D. enhancer-promoter loop formation

enh.

E. splicing

F. miRNA maturation

G. miRNA sponge (i.e. ceRNA)

H. mRNA stabilization or degradation

I. modulation of translation

J. protein stabilization or degradation

K. modulation of activity

L. re-localization

M. sORF host

N. miRNA host

O. exosome-mediated tranfer

LncRNAs

can interact with

DNA   RNA   Protein

ZFAS1
GAS5
MALAT1 (s)
NEAT1 (p)
TERRA
GATA3-AS1 ✓
Fendrr ✗
MYMLR
linc-HELLP
CCAT2 ✗
uc.372 ✗
ROR
DSCR8
TINCR ✗
PTB-AS ✓
BC1 ✗
MCM3AP-AS1 ✓
PtHL ✗
MALAT1 ✓
NORAD
NRON
Disc1
CASIMO1
DWORF
MCM7
DLEU2

A.

Genome assembly version

| galgal2 | galgal3 | galgal4 | galgal5 | GRCg6a | GRCg7w | GRCg7b |
| 2004 | 2006 | 2011 | 2015 | 2018 | 2021 | 2021 |

17709   16736   15508 17208   18346 19119   16779 17485   16884 18016   17007 18024

4641 5763   5504 5147   12194 5815   11944 5791

v42   v70   v85   v94 v103   v106 v104   v107 v106   v107 v106

Genome annotation version

Ensembl   NCBI   PCG   lncRNA   others

B.

1.5% (1432)

1.6% (1499)   6.4% (6023)

9% (8456)

NCBI
v105 → v106
#tr = 93980
(GRCg7b)

81.3% (76445)

New
Depreciated
Major changes
Minor changes
Identical

A. ☐ NCBI ▧ Ensembl

#Genes

(10³) — PCG ■ lncRNA ■ miRNA ■ others ■

15 — 17007 18024

10 — 11944

5 — 5791 — 674 799 — 483 1059

#Transcripts

(10³)

60

40 —

20 —

0 —

B.

Total
Ncbi: 85671 *tr*
Ens: 72579 tr
55.9% 26.1% 17.8% (175)

PCG
Ncbi: 72576 tr
Ens: 44790 tr
70.8% 25.6% 3.4% (60)

lncRNA
Ncbi: 10062 tr
Ens: 26656 tr
77.4% 21.4% 1.1% (11)

■ New loci ■ New isoform ☐ Equal Isoform (#Artifact)

C.

#Tr/gn   #Ex/tr   Tr. Size (10³)   Ex. Size (10³)   In. Size (10³)

* *   * *

FTFCAH

LTLCCH&NDV

33
(2.5%)

316
(23.7%)

4
(0.3%)

3
(0.2%)

187
(14%)

53
(4%)

0
(0%)

5
(0.4%)

677
(50.9%)

1
(0.1%)

2
(0.2%)

16
(1.2%)

7
(0.5%)

4
(0.3%)

23
(1.7%)

LTLCAH

FTFCCH&NDV

a

b

c

d

e

f

g

a

b

c

d

e

f

g

BME within line

Activation z−score

−3.606   2.887

Canonical Pathways	BMELTLCAH – 2020–...	BMEFTFCAH – 2020–...	BME LTLCCH – 2020...	BMEFTFCCH – 2020–...
Oxidative Phosphorylation				
Cardiac Hypertrophy Signali...				
TCA Cycle II (Eukaryotic)				
Phospholipase C Signaling				
Wnt/Ca+ pathway				
Insulin Secretion Signaling P...				
Synaptogenesis Signaling Pa...				
Protein Kinase A Signaling				
ERK/MAPK Signaling				
Cell Cycle: G1/S Checkpoint...				
White Adipose Tissue Brown...				
Calcium Signaling				
P2Y Purigenic Receptor Sign...				
Dopamine-DARPP32 Feedb...				
Androgen Signaling				
Factors Promoting Cardioge...				
Cell Cycle Control of Chromo...				
Synaptic Long Term Potentia...				
Estrogen Receptor Signaling				
Melanocyte Development an...				
Regulation Of The Epithelial ...				
Acetyl-CoA Biosynthesis I (P...				
Aldosterone Signaling in Epit...				
Apelin Cardiomyocyte Signal...				
STAT3 Pathway				

HYP&BME AH&CH module−trait relationships

	Tissue	Line	TrtAH	TrtCH	pH	PCO2	PO2	HCO3	BE	sO2	Glu	Na	K	TCO2	iCa
MEturquoise	−1[a] (2e−70)[b]	0.012 (0.9)	0.0069 (1)	−0.0091 (0.9)	−0.011 (0.9)	−0.0016 (1)	−0.0024 (1)	−0.011 (0.9)	−0.014 (0.9)	−0.0061 (1)	0.00015 (1)	−0.0036 (1)	−0.012 (0.9)	−0.012 (0.9)	−0.0049 (1)
MEblue	0.15 (0.2)	0.96 (4e−35)	−0.00055 (1)	−0.029 (0.8)	−0.46 (2e−04)	0.15 (0.2)	−0.5 (2e−05)	−0.31 (0.01)	−0.4 (0.001)	−0.53 (7e−06)	−0.45 (2e−04)	0.25 (0.05)	0.031 (0.8)	−0.28 (0.03)	−0.16 (0.2)
MEgreen	0.53 (7e−06)	0.091 (0.5)	0.13 (0.3)	−0.076 (0.6)	−0.11 (0.4)	−0.076 (0.6)	0.089 (0.5)	−0.19 (0.1)	−0.19 (0.1)	0.02 (0.9)	−0.074 (0.6)	−0.042 (0.7)	−0.12 (0.3)	−0.16 (0.2)	−0.15 (0.2)
MEred	0.47 (9e−05)	0.22 (0.08)	0.16 (0.2)	−0.064 (0.6)	−0.18 (0.2)	0.024 (0.9)	−0.038 (0.8)	−0.16 (0.2)	−0.19 (0.1)	−0.16 (0.2)	−0.16 (0.2)	0.22 (0.08)	−0.12 (0.3)	−0.16 (0.2)	−0.096 (0.4)
MEbrown	0.19 (0.1)	−0.039 (0.8)	−0.15 (0.2)	0.031 (0.8)	0.13 (0.3)	−0.095 (0.5)	0.18 (0.2)	−0.00051 (1)	0.047 (0.7)	0.23 (0.07)	−0.046 (0.7)	0.13 (0.3)	0.12 (0.4)	0.021 (0.9)	−0.011 (0.9)
MEblack	0.6 (2e−07)	−0.029 (0.8)	0.022 (0.9)	−0.2 (0.1)	0.2 (0.1)	−0.24 (0.06)	0.085 (0.5)	−0.066 (0.6)	0.012 (0.9)	0.13 (0.3)	−0.17 (0.2)	−0.063 (0.6)	0.11 (0.4)	−0.051 (0.7)	−0.16 (0.2)
MEyellow	−0.14 (0.3)	−0.00091 (1)	0.066 (0.6)	−0.29 (0.02)	0.23 (0.07)	−0.29 (0.02)	0.084 (0.5)	−0.099 (0.4)	−0.0081 (0.9)	0.13 (0.3)	−0.22 (0.08)	−0.084 (0.5)	0.12 (0.3)	−0.081 (0.5)	−0.21 (0.1)
MEgrey	−0.088 (0.5)	0.62 (4e−08)	−0.058 (0.6)	−0.1 (0.4)	−0.23 (0.07)	−0.081 (0.5)	−0.26 (0.04)	−0.32 (0.01)	−0.34 (0.005)	−0.26 (0.04)	−0.34 (0.006)	0.15 (0.2)	−0.017 (0.9)	−0.3 (0.02)	−0.22 (0.08)

GO terms
- Rgulation of membrane potential
- Muscle contraction
- Glycolytic process
- Rgulation of microtubule polymerization or depolymerization
- Cardiac muscle thin filament assembly

KEGG pathway
- Ribosome
- Cardiac muscle contraction
- Biosynthesis of amino acids
- Fructose and mannose metabolism
- Glycolysis / Gluconeogenesis

Fold enrichment

0   5   10   15   20

a

b

c

Figure 1

a. CR1, a non-LTR retrotransposon, LINE

ORF1 *ORF2 = APE + RT*

$(ATTCTRTG)_n$
$(ATTCTRTG)_n$
$(ATTCTRTG)_n$
$(ATTCTRTG)_n$
$(ATTCTRTG)_n$

b. TguSINE1, a non-LTR retrotransposon using the CR1 transposition machinery

tRNA-Ile CR1
$(ATTCTRTG)_n$
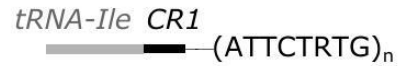
c. LTR-retrotransposons, ERV and Gypsy

LTR *Gag* *RT* *Env* LTR

LTR *Gag* *RT* LTR

d. DNA transposons

Mariner1_GG, Charlie

TIR *Transposase* TIR

Galluhop

TIR TIR

Charlie/Galluhop

TIR TIR TIR TIR

a *G. g. domesticus* (WAG137G4)

b *N. meleagris* (JABXER010000123)

a

Gallus gallus
(after Dyomin et al. 2019)

IGS_I
IGS_II
IGS_III
IGS_IV

Poly-T

- AL repeats
- SV repeats

CT(n)   IR

- EL1
- EL2A
- EL2B
- EL2C
- EL2D
- EL2F
- EL3A
- EL3B
- EL3C
- EL3D

- VAL1
- VAL2
- VAL3
- VAL4
- VALF

SV-AL repeat block   **EL repeat block**   **VAL repeat block**

b

Numida meleagris
(JABXER010000123)

IGS_I
IGS_II
IGS_III

Poly-T

- AL repeats
- SV repeats

- Nme repeats

CT(n)   Nme   CT(n)
         repeat block

IR

- VALF
- VAL_A
- VAL_B
- VAL_FC

IR

- VALF
- VAL_A
- VAL_B
- VAL_FC
- VAL_C
- VAL_D
- VAL_E

SV-AL repeat block   **VAL repeat block**   **VAL repeat block**

GGA1 **GGA5**

A. Chicken chromosome 16

BLA fB ORs, SRCRs

B | GC | Y | NOR q arm p arm

5 cM 12 cM

>50 cM

B. The B locus

BF-BL region

extended class I | class II | class I | class III

B30.2 B30.2 pseudo lectin BNK BLB2 DMB1 DMA DMB2 BF1 TAP2 C4 CenpA CYP21 leukotriene B receptor

tRNA-L BG1 Blec BLB1 Tapasin RING3 TAP1 BF2 tenascin CD1.2 CD1.1

TRIM region

extended class I

Zn-finger Zn-finger L-aa-oxidase TRIM27 TRIM41

TRIM7 Zn-finger unknown TRIM7 HEP21 TRIM39 TRIM27 guanine-nt-binding protein
tRNA-K TRIM39 tRNA-K
tRNA-V

BG region

extended class I

BG13 BG12 BG11 BG10 kinesin LOC BG9 BG8 BG7 BG6 BG5 BG4 BG3 BG2 kinesin

lectin lectin

0 50 100 kB

Harvesting of PGCs

PGCs in germinal crescent · Circulating PGCs in Blood · PGCs in embryonic gonads

Transfer into the freezing media

Overnight in Mr Frosty and Cryopreservation in liquid N2

Dissociate of embryonic gonadal PGC

Injection into the recipient embryo

Donor

Characterization and validation of PGCs lines

- Sex determination
- Alkaline Phosphatase (AKP) and Periodic Acid–Schiff (PAS)
- Immunochemistry & flow cytometry
- PGC specific gene expression
- Proliferation & pluripotency genes essay

Countess II Live/Dead Report

PGC propagation and characterization
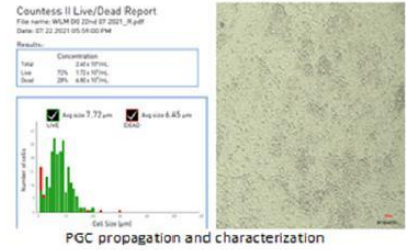
Production of chimeras chicks

Chimeric hen showing some phenotypic characteristics of the local donor

A

- Evaluation of cells viability
- Cryoconservation of PGS lines
- Recipient
- Incubation of recipient (host) egg for injection
- Injection of donor's PGCs into the vascular system of recipient embryo
- Thawing and in vitro culturing of cryopreserved PGC lines
- Genome editing of PGC and selection of gene-edited cells
- Injection of gene-edited cells in recipient embryos
- Production of germline chimeras and test cross
- Production of genome-edited offsprings
- Eg: sterile surrogate host
- Generation of presumptive germline chimeras
- Ovum
- Spermatozoa
- Donors-derived gametes
- Successive Chimeras' crosses
- Donor-derived progeny, regeneration of donor's breed

B

- Production of genome-edited offsprings
- Eg: sterile surrogate host
- Biobanked PGS lines from indigenous chicken (donor)
- Genome edited chickens
- Incubation of surrogate recipient (sterile) egg for injection
- Injection of donor's PGCs into the vascular system of recipient embryo
- Ovum
- Spermatozoa
- 100% donors-derived gametes
- 100% donor-derived progeny, regeneration of indigenous (donor) breed
- 100% restored donor indigenous breed

Improved disease resistance & productivity

Efficient production of protein-drugs with improved efficacy

Genome editing in poultry

Female

Male

Chicken sexing during embryonic stage

Source : Park et al., (2020)

Table 1. Phased assembly comparisons of broiler and layer genomes to RJF for Gallus gallus. Each assembly contains the Z and W sex chromosomes despite there being only one copy of each from the parents.

Common name	Assembly version	N50 contig (Mbp)	Total size (Mbp)	Total contigs	N50 Scaffold Length (Mbp)	Unplaced Sequences (Mbp)
Red Jungle Fowl	GRCg6a	17.6	1,055	1,403	20	14.1
Broiler	GRCg7b	18.8	1,049	677	90	6.6
Layer	GRCg7w	17.7	1,046	685	90	7.1

[1] NCBI assembly metrics.

Table 2 Chicken cell atlas: scRNA-seq data

Focused tissue	Method [a]	Reference	doi	GEO accession
Embryo	SMART-seq, dissociation, FACS, LCM	Morrison et al. (2015)	10.1002/dvdy.24274	-
Neural crest	SMART-seq, dissociation, FACS	Morrison et al. (2017)	10.7554/eLife.28415	GSE108230
Primitive streak	SMART-seq Fluidigm C1	Vermillion et al. (2018)	10.1016/j.ydbio.2018.04.007	GSE89910
Developing limb	10x, dissociation	Feregrino et al. (2019)	10.1186/s12864-019-5802-2	GSE130439
Gonads	10x, dissociation	Estermann et al. (2020)	10.1016/j.celrep.2020.03.055	GSE143337
Neural crest (hindbrain)	10x, dissociation	Gandhi et al. (2020)	10.7554/eLife.57779	PRJNA624258
Retina	10x, dissociation	Hoang et al. (2020)	10.1126/science.abb8598	https://github.com/jiewwwang/Single-cell-retinal-regeneration
Skeletal muscle and fat (breast)	10x, dissociation	Li et al. (2020)	10.1186/s12864-020-07136-2	CRA002353
Developing retina	10x, dissociation	Tegla et al. (2020)	10.7554/eLife.54279	GSE142244
Limb muscle	10x, dissociation	De Lima et al. (2021)	10.1038/s41467-021-24157-x	GSE166981
Primitive streak	inDrops , dissociation	Guillot et al. (2021)	10.7554/eLife.64819	GSE161905
Cochlea	SMART-seq, dissociation	Janesick et al. (2021)	10.1016/j.celrep.2021.108900	https://umgear.org/dataset_explorer.html?search_terms=33761346
Developing heart	10x, dissociation	Mantri et al. (2021)	10.1038/s41467-021-21892-z	GSE149457

Retina	10x, dissociation	Yamagata et al. (2021)	10.7554/eLife.63907	GSE159107
Pituitary	10x, dissociation	Zhang et al. (2021)	10.3389/fphys.2021.562817	CRA003604
Developing hypothalamus	10x, methanol	Kim et al. (2022)	10.1016/j.celrep.2021.110251	GSE171649
Lymphocyte	10x, suspension	Qu et al. (2022)	10.3389/fmicb.2022.800618	PRJNA687808
Germ cell	10x, dissociation, FACS	Rengaraj et al. (2022)	10.1016/j.csbj.2022.03.040	PRJNA761874
Neural crest	10x, dissociation	Williams et al. (2022)	10.7554/eLife.74464	GSE131688, GSE181577

a) 10x, 10x chromium system from 10x Genomics (Pleasanton, CA, USA). Dissociation indicates enzymatic dissociation.

Table 3. Summary Statistics of RNA-Seq output

	Line	Treatment	Time Point	Raw Reads	Aligned Reads	Alignment Rate
Hypothalamus	Leghorn	Non-treated	Acute Heat	90,131,605	82,596,745	91.91%
	Leghorn	Treated	Acute Heat	99,594,536	92,251,909	92.91%
	Fayoumi	Non-treated	Acute Heat	88,123,256	81,628,366	92.85%
	Fayoumi	Treated	Acute Heat	88,021,678	81,552,876	92.88%
	Leghorn	Non-treated	CH&NDV	85,586,252	78,499,724	93.38%
	Leghorn	Treated	CH&NDV	82,230,451	75,662,675	93.28%
	Fayoumi	Non-treated	CH&NDV	85,981,153	80,602,593	93.72%
	Fayoumi	Treated	CH&NDV	84,433,550	78,218,756	93.76%
Breast Muscle	Leghorn	Non-treated	Acute Heat	202,976,814	171,248,763	84.65%
	Leghorn	Treated	Acute Heat	197,916,434	168,158,916	85.16%
	Fayoumi	Non-treated	Acute Heat	199,002,995	172,003,441	86.63%
	Fayoumi	Treated	Acute Heat	167,292,872	145,714,038	86.93%
	Leghorn	Non-treated	CH&NDV	97,856,074	85,109,242	99.91%
	Leghorn	Treated	CH&NDV	101,741,687	87,181,252	99.91%
	Fayoumi	Non-treated	CH&NDV	100,416,625	86,497,655	99.91%
	Fayoumi	Treated	CH&NDV	102,815,215	88,592,510	99.92%

Table 4. DEGs identified in multiple contrasts in the hypothalamus

DEG	Gene description	LTLC AH	FTFC AH	LTLC CH&NDV	FTFC CH&NDV
HSPB9	Heat shock protein family B (small) member 9	2.58[1]	3.06		
HSPA2	Heat shock 70kDa protein 2	1.23	1.07		
CRELD2	Cysteine rich with EGF lik domains 2	1.07	1.04		
HSP90B1	Heat shock protein 90 beta family member 1	0.96	0.80		
GPX1	Glutathione peroxidase	0.77	0.94		
CALR	Calreticulin	0.64	0.90		
SDF2L1	Stromal cell derived factor 2 like 1	0.58	0.59		
MANF	Mesencephalic astrocyte derived neurotrophic factor	0.57	0.89		
HYOU1	Hypoxia up regulated 1	0.37	0.43		
PDIA3	Protein disulfide isomerase family A member 3	0.28	0.38		
HBBA	Hemoglobin beta, subunit A			-1.18[2]	-1.54
HBA1	Hemoglobin subunit alpha 1			-1.26	-1.47
HBAD	Hemoglobin alpha, subunit D			-1.27	-1.59
EPB42	Erythrocyte membrane protein band 4.2			-2.73	-1.77

Note: [1] Log_2 fold change and red highlight means up-regulation; [2] Blue highlight means down-regulation.

Table 5. Contrast specific DEGs identified in hypothalamus

DEG	Gene description	LTLC AH	FTFC AH	LTLC CH&NDV	FTFC CH&NDV
CLIC5	Chloride intracellular channel 5	1.03[1]			
HSPB8	Heat shock protein family B member 8	0.89			
BAG3	BCL2 associated athanogene 3	0.67			
HSPH1	Heat shock protein family H member 1	0.64			
HSPA5	Heat shock 70kDa protein 5	0.43			
DNAJB6	DnaJ heat shock protein family (Hsp40) member B6	0.24			
DDX3X	DEAD-box helicase 3, X-linked	0.22			
DNAJC3	DnaJ heat shock protein family member C3	0.21			
SLC38A2	Solute carrier family 38 member 2	-0.30[2]			
NR1D1	Nuclear receptor subfamily 1 group D member 1	-0.61			
HSP30C	Heat shock protein 30C-like		2.23		
HSP90AB1	Heat shock protein 90 alpha family class B member 1		-0.47		
MX1	Myxovirus resistance 1, interferon-inducible protein		-0.85		
RSAD2	Radical S-adenosyl methionine domain containing 2		-1.50		
IFIT5	Interferon induced protein with tetratricopeptide repeats 5		-1.60		
TLX3	T cell leukemia homeobox 3			1.46	
IFI6	Interferon alpha inducible protein 6			1.17	
HBE1	Hemoglobin subunit epsilon 1			-2.08	
SLC2A12	Solute carrier family 2 member 12				1.15
MRC1	macrophage mannose receptor 1-like				1.05
SLC13A5	Solute carrier family 13 member 5				0.91
TGM2	Transglutaminase 2				0.71
CRCSB	cytokine receptor common subunit beta-like				0.64
HS3ST1	Heparan sulfate-glucosamine 3-sulfotransferase 1				0.43
NKAIN4	Sodium/potassium transporting ATPase interacting 4				-0.22

Note: [1] Log2 fold change and red highlight means up-regulation; [2] Blue highlight means down-regulation.

Table 6.   Significantly correlated traits and gene modules.

Module Traits	Turquoise	Blue	Black	Yellow
Tissue	**-1**	0.15	**0.60**	-0.14
Line	0.01	**0.96**	-0.03	-0
CH&NDV	-0	-0	-0.2	**-0.29**
pH	-0.01	**-0.46**	0.2	0.23
PCO2	-0.01	0.16	0.09	**-0.29**
HCO3	-0.01	**-0.31**	-0.07	-0.1
TCO2	-0.01	**-0.28**	-0.05	-0.08
BE	-0.01	**-0.40**	0.01	-0
PO2	-0.01	**-0.50**	0.09	0.08
sO2	-0.01	**-0.53**	0.13	0.13
Na$^+$	-0.01	**0.25**	-0.06	-0.08
Glu	0	**-0.45**	-0.17	-0.22

Table 7. Top 5 driver genes in each significant trait correlated gene modules

Factor	Correlation	Module color	Gene name (GS[1], MM[2])
Tissue	Positive	Black (Hypothalamus)	Novel (0.82, 0.92), ELP1(0.76, 0.91), HINT2 (0.75, 0.94), BRIX1 (0.72, 0.97), RPP25L (0.71, 0,.93)
	Negative	Turquoise (Breast muscle)	GRIN1(-1, 1)), TPM (-1, 1), VSTM2A (-1, 1), GABRE (-1,1), GRIK2(-1, 1)
Line	Positive	Blue (Leghorn)	KPFL(0.98, 0.94), Novel (0.98. 0.,92), RMD1 (0.97, 0.89), KIFC1 (0.97, 0.92), HA1F (0.97, 0.91)
CH&NDV	Negative	Yellow	AUH (-0.55, 0.68), SKP2 (-0.53, 0.74)
pH	Negative	Blue	TNFRSF8 (-0.61, 0.69), MRERF2(0.60, -071), AIFM2 (0.55, -0.85), MOG (0.51, -0.73), HA1F (-0.53, 0.91)
PO₂	Negative	Blue	BLEC3 (-0.59, 0.78), KIFC1 (-0.56, 0.92), PYROXD2 (0.55, -0.71), ENSGALG00000039964 (-0.54, 0.94), ENSGALG00000054582 (-0.54, 0.92),
sO₂	Negative	Blue	AIFM2 (0.57, -0.85), BLEC2 (0.57, -0.88), ENSGALG00000049158 (0.56, -0.90), ENSGALG00000033498 (0.55, -0.92), Protein Manbal-like (0.54, -0.94)
Glu	Negative	Blue	MT-ND2 (0.50, -0.82)

Table 8. Length and GC content in the rDNA cluster and cluster elements in chicken and guinea fowl.

		5'ETS	18S	ITS1	5.8S	ITS2	28S	3'ETS
Gallus g. domesticus (KT445934)	Cluster length (bp)	11,863						
	Cluster element lengths (bp)	1,836	1,823	2,530	157	733	4,441	343
	GC%	75	54.5	82.2	57.3	82	68	79.9
Numida meleagris (JABXER010000 0123)	Cluster length (bp)	12,133						
	Cluster element lengths (bp)	1,792	1,824	2,837	157	761	4,422	340
	GC%	75.7	54.4	82.7	57.3	82.3	68	83.2

Table 9. Intra-individual variability of IGS lengths in chicken and guinea fowl

		IGS_I	IGS_II	IGS_III
Gallus g. domesticus (WAG137G4)	**Length (bp)**	22,627	15,169	15,241
	GC (%)	68.1	69.2	68.1
Numida meleagris (JABXER010000123)	**Length (bp)**	6,530	6,562	8,223

Table 10. African chicken ecotypes cryopreserved using the blood and blastoderm methods.

Country	Ecotype	Number cell lines	Country	Ecotype	Number cell lines
KENYA	White leghorn	43	**ETHIOPIA**	Arobe	59
	Karen (Nairobi)	7		Horro	29
	Narok	14		Hawassa	32
	Bomet	42	**TANZANIA**	Shinyanga	84
	Siaya	23		Mwanza	11
	Migori	20		Mbeya	43
	Homabay	8		Morogoro	69
	Kakamega	13			

Table 11. Kenyan chicken ecotypes and lines cryopreserved using the embryonic gonad method.

Kenyan Ecotypes	Male lines	Female lines
Laikipia (LP)	44	43
Bungoma (BG)	9	18
Kilifi (KF)	4	1
Kakamega (KK)	14	13
Bomet (BM)	1	1
Homabay (HB)	21	19
Siaya (SY)	13	5
Kwale (KW)	0	1
KALRO chicken Improved lines 1 (KC1)	61	63
KALRO chicken Improved lines 2 (KC2)	75	41
Total	242	205