



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms

Citation for published version:

Wong, GK-S, Liu, B, Wang, J, Zhang, Y, Yang, X, Zhang, Z, Meng, Q, Zhou, J, Li, D, Zhang, J, Ni, P, Li, S, Ran, L, Li, H, Zhang, J, Li, R, Li, S, Zheng, H, Lin, W, Li, G, Wang, X, Zhao, W, Li, J, Ye, C, Dai, M, Ruan, J, Zhou, Y, Li, Y, He, X, Zhang, Y, Wang, J, Huang, X, Tong, W, Chen, J, Ye, J, Chen, C, Wei, N, Li, G, Dong, L, Lan, F, Sun, Y, Zhang, Z, Yang, Z, Yu, Y, Huang, Y, He, D, Xi, Y, Wei, D, Qi, Q, Li, W, Shi, J, Wang, M, Xie, F, Wang, J, Zhang, X, Wang, P, Zhao, Y, Li, N, Yang, N, Dong, W, Hu, S, Zeng, C, Zheng, W, Hao, B, Hillier, LW, Yang, S-P, Warren, WC, Wilson, RK, Brandström, M, Ellegren, H, Crooijmans, RPMA, van der Poel, JJ, Bovenhuis, H, Groenen, MAM, Ovcharenko, I, Gordon, L, Stubbs, L, Lucas, S, Glavina, T, Aerts, A, Kaiser, P, Rothwell, L, Young, JR, Rogers, S, Walker, BA, van Hateren, A, Kaufman, J, Bumstead, N, Lamont, SJ, Zhou, H, Hocking, PM, Morrice, D, de Koning, D-J, Law, A, Bartley, N, Burt, DW, Hunt, H, Cheng, HH, Gunnarsson, U, Wahlberg, P, Andersson, L, Kindlund, E, Tammi, MT, Andersson, B, Webber, C, Ponting, CP, Overton, IM, Boardman, PE, Tang, H, Hubbard, SJ, Wilson, SA, Yu, J, Wang, J, Yang, H & International Chicken Polymorphism Map Consortium 2004, 'A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms' *Nature*, vol 432, no. 7018, pp. 717-22., 10.1038/nature03156

Digital Object Identifier (DOI):

[10.1038/nature03156](https://doi.org/10.1038/nature03156)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher final version (usually the publisher pdf)

Published In:

Nature

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 20. Feb. 2015



Published in final edited form as:

Nature. 2004 December 9; 432(7018): 717–722. doi:10.1038/nature03156.

A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms

International Chicken Polymorphism Map Consortium

Summary

We describe a genetic variation map for the chicken genome containing 2.8 million single nucleotide polymorphisms (SNPs), based on a comparison of the sequences of 3 domestic chickens (broiler, layer, Silkie) to their wild ancestor Red Jungle Fowl (RJF). Subsequent experiments indicate that at least 90% are true SNPs, and at least 70% are common SNPs that segregate in many domestic breeds. Mean nucleotide diversity is about 5 SNP/kb for almost every possible comparison between RJF and domestic lines, between two different domestic lines, and within domestic lines - contrary to the idea that domestic animals are highly inbred relative to their wild ancestors. In fact, most of the SNPs originated prior to domestication, and there is little to no evidence of selective sweeps for adaptive alleles on length scales of greater than 100 kb.

Keywords

chicken; polymorphism; domestication; selection

Introduction

The generation of a high quality draft sequence for the genome of chicken (*Gallus gallus*) is an important advance¹. Chickens are good models for studying the genetic basis of phenotypic traits, because of the extensive diversity among domestic chickens selected for different purposes. Monogenic traits are well-studied²⁻⁴, but many interesting traits are complex and determined by an unknown number of genes. Quantitative trait loci (QTLs) have been mapped for a range of traits, including ones for growth, body composition, egg production, antibody response, disease resistance, and behaviour⁵. Determining causative genes is difficult, since each locus controls only a fraction of the phenotypic variance. We will describe a survey of the genetic variation between 3 domestic chickens and their wild ancestor. The 2.8 million single nucleotide polymorphisms (SNPs) that we identified will facilitate mapping of complex traits in many ways. First, improved marker density allows researchers to take advantage of the higher recombination rates in chicken¹, which are 2.5 to 21 cM/Mb depending on the chromosome, as compared to 1 cM/Mb for human and 0.5 cM/Mb for mouse. The previous linkage map used 2000 markers^{6,7}, but only 800 of these were microsatellites or SNPs, which are the most useful⁸. More importantly, our new data allow researchers to construct detailed haplotypes that segregate in different QTL crosses. Because any mutation underlying a QTL must once have originated from a single founder animal, haplotype comparisons will facilitate the fine mapping of QTLs⁹. To this end, we conduct a

The individual SNPs were deposited at GenBank/dbSNP with submitted SNP (ss) number ranges: 24821291 to 24922086, 24922088 to 26161960, 26161962 to 28446123, and 28452569 to 28452598. They may also be found at <http://chicken.genomics.org.cn>⁴⁸, the UCSC genome browser, and the Ensembl genome browser. Access to raw sequencing traces is being provided through the NCBI Trace Archive.

genomewide search for evidence of selection due to domestication, and provide an initial characterization of the expected magnitude of these effects.

Genetic variation and utility

Our experiment is outlined in Figure 1. SNPs are generated by partial sequencing at $\frac{1}{4}$ coverage for each of 3 domestic breeds (a male broiler, a female layer, and a female Silkie), and comparison of the resultant reads to the 6.6x genome for the wild ancestor of domestic chickens, Red Jungle Fowl (RJF). We expect marked heterozygosity within the 3 domestic lines, but not within RJF because the sequenced bird for the genome project is from a highly inbred line that is essentially homozygous.

Comparing the sequence reads for broiler, layer, and Silkie to the genome of RJF, we identified nearly a million SNPs in each instance, at mean rates of about 5 SNP/kb, as shown in Table 1. Notice that all of the “SNP rates” quoted in this paper are computed as *nucleotide diversities* (π), and given in units of $\pi \times 10^3$. After correcting for SNPs detected in more than one line, there are 2,833,578 variant sites, or one potential marker every 374 bp along the 1.06 Gb genome. To assess the reliability of these data, we resequenced 295 SNPs in the same bird in which it was detected (Table S1). As many as 94% of the SNPs were confirmed. However, confirmation rates are sensitive to the functional context (*e.g.*, coding versus non-coding) and SNPs in rare categories are less likely to be confirmed. In fact, only 83% of the non-synonymous SNPs were confirmed. Small indels of a few base pairs in length (mean of 2.3 and median of 1) are detected at rates that are well correlated with the corresponding SNP rates, but smaller by about a factor of 10.

Chicken autosomes are sorted by size into 5 large macrochromosomes (GGA1-5), 5 intermediate chromosomes (GGA6-10), and 28 microchromosomes (GGA11-38). SNP and indel rates are independent of chromosome size, as shown in Figure 2. GGA16 is the sole exception, because it contains the highly variable MHC10. This result is surprising, as recombination rates on microchromosomes are much higher than on macrochromosomes1 and studies in other organisms exhibit a positive correlation between recombination rates and polymorphism rates11-12. We expect that higher gene densities on microchromosomes likely counteract the effect of higher recombination rates.

SNP rates between and within chicken lines can be determined from the overlaps between reads. Table 1 demonstrates that almost every pairwise combination gives a SNP rate of just over 5 SNP/kb, except for broiler-broiler and layer-layer, which show about 4 SNP/kb, as expected since the sequenced broiler and layer are from closed breeding lines. To ensure that there are no confounding factors from the single read nature of our data, or the complexities of the overlap analysis, we used comparisons to 3.8 Mb of finished BAC sequence of a different White Leghorn13 from the same breed but not the same line as the layer sequenced herein. 15 chromosomes were sampled, and the results confirm our rates of 5 SNP/kb. In another study of 15 kb of introns in 25 birds from 10 divergent breeds of domestic chickens14, an autosomal rate of 6.5 SNP/kb was reported.

To quantify SNP and indel rate variation versus functional context, we considered three gene sets representing 3868 confirmed mRNA transcripts, 995 chicken orthologs of human disease genes, and 17,709 Ensembl annotations from the RJF analysis1. Complete details for all 3 lines are tabulated in the supplements (Table S2). An excerpt for broiler is shown in Table 2. Within genes defined by mRNA transcripts, the SNP rates are 3.5, 2.1, 5.7, and 3.4 SNP/kb in 5'-UTR, coding exon, intron, and 3'-UTR regions respectively. In coding regions, indel rates are 43 times smaller than SNP rates. Ka/Ks is 0.098, similar to what is typically seen in vertebrate comparisons. We also studied “conserved non-coding regions” from the RJF analysis1. SNP rates are similar to those of coding exons, but indel rates are

intermediate to those of coding exons and UTRs, which supports the notion that these regions are functional, but may not encode proteins.

Utility of these SNPs depends on their frequency of occurrence in commonly used chicken populations. Hence, we typed 125 SNPs (including coding and non-coding SNPs, randomly distributed across the chicken genome) in 10 unrelated individuals from each of 9 divergent lines representing an assortment of European breeds. This collection includes commercial broiler and layer breeds, standardized breeds selected for their morphological traits, and an unselected breed from Iceland (Table S3). Both alleles segregated in 73% of 1113 successful marker-line combinations (out of 1125 possible combinations). Averaged minor allele frequency is 27%, but it decreases to 20% if marker-line combinations where one of the two alleles is fixed are included. This indicates that a majority of the SNPs are common variants that predate the divergence of modern breeds. Only 12% of the markers had a minor allele frequency of less than 10% in the 90 animals tested.

We demonstrate by example how these data can be used to target specific genome regions. Details of our experiments are in Supplement E (Examples). First, we consider a body weight related QTL on GGA4 that was previously mapped to a 150 cM interval^{15,16}. After a year of effort, where every known microsatellite (>50) was tested, 26 informative markers were developed. Further progress would have required the laborious sequencing of multiple chickens to find additional polymorphisms in this target region. With the SNP map, we selected 47 random broiler-layer SNPs, and ABI SNPLEX assays were developed to type an experimental F₂ cross (n = 466). 28 (60%) of these SNPs were informative, but none had breed specific alleles, confirming that most variations predate domestication. In just one month, we doubled the number of markers, and resolved the initial QTL into two QTLs that affect body weight at 3 and 9 weeks of age.

In addition to providing markers for fine mapping, these SNPs are a rich source of candidate polymorphisms for the causative differences underlying important traits. As an example, candidate genes for disease resistance often include TGF- β ^{17,18}, cytokines¹⁹, and the major histocompatibility complex (MHC). We thus identified 40 SNPs from the SNP map in the coding or promoter regions of 12 cytokine genes. When typed against 8 inbred layer lines, 32 of these SNPs were informative. Cytokine genes on GGA13, including *IL4* and *IL13*, two genes that are expressed in T helper-2 (Th2) cells, drive antibody response. Four of the six SNPs that were polymorphic among lines were in *IL4* and *IL13*, and these SNPs were fixed for different alleles in lines N and 15I, which show differential antibody response to vaccination²⁰. These SNPs therefore allow us to test whether the *IL4* and *IL13* loci directly determine the observed differential antibody response.

Domestication and selection

Domestic animals are useful models of phenotypic evolution under selection. The challenge is to find not only those loci that determine phenotypic differences, but also the causative alleles. We adopt two approaches, searching for evidence of selective sweeps²¹, and for non-synonymous amino acid substitutions at highly conserved sites. One example of a selective sweep is the *IGF2* locus in pigs²². Given the available data, determining the exact haplotype structure is difficult, because blocks of shared alleles can be erroneously disrupted by heterozygosity of the domestic lines and by sequencing errors. However, we can still search for the local reductions in heterozygosity that accompany selective sweep, as long as we are mindful of the sequencing error rate.

We did 3-way comparisons of RJF and all possible combinations of two domestic lines. Given the limited coverage of the latter, we only examined 100 kb segments with at least 10 SNP sites where each qualifying site must have read coverage from every line. In practice,

these segments contained an average of 25 to 28 SNPs. Then, we computed how often 80% or more of the SNP sites are identical in the two domestic lines but different in RJF. In Table S4, we show that 0.4 to 1.5% of the segments qualified. However, when we searched for shared alleles between RJF and one of the domestic lines, 1.2 to 2.6% of the segments qualified. Heterozygosity of the domestic lines is more of a confounding factor in searching for blocks of shared alleles between two domestic lines, versus between RJF and one domestic line. This could explain the difference, but if so, then heterozygosity of the domestic lines is the dominant factor in determining such blocks of shared alleles, not selective sweeps. Hence, selective sweeps that occurred before the divergence of modern domestic breeds must have left behind footprints that are much smaller than 100 kb. This would however be entirely consistent with the historically large effective population size of domestic chickens, and the reported high recombination rates.

For a glimpse of the true haplotype patterns, one can compare the aforementioned 3.8 Mb of finished BAC sequence, from the second layer line (L2), to the genome of RJF. These results are overlaid alongside the primary SNP data set in Figure 3. Short RJF-type fragments can be seen in all 4 lines. Shared domestic-type fragments can also be seen, but at sizes of 5 to 15 kb. This is consistent with our inability to detect footprints of selective sweeps at length scales of 100 kb and suggests that a better choice is 10 kb. However, our data are insufficient for such a genomewide analysis.

It has been hypothesized that loss-of-function (LOF) mutations have accumulated in domestic animals, as the result of relaxed purifying selection and selection for adaptive benefits²³. An example of the latter is the deletion in the myostatin gene in cattle selected for muscularity²⁴. Such deletions are rare, and so we looked for non-synonymous SNPs at highly conserved sites using *SIFT*²⁵. Every substitution is thus classified as being likely to affect function (intolerant) or not (tolerant). For genes defined by mRNA transcripts, 26% of testable SNPs are intolerant, although only 11% are intolerant if we restrict this to high confidence assessments (Table S5). Usually, it is the domestic allele that is intolerant, but we would emphasize that intolerant SNPs are rare, and only 59% were confirmed by PCR resequencing. Given that the domestic allele is represented by a single read, as opposed to 6.6 for the wild allele, much of this effect is likely due to sequencing errors. However, we noticed the same effect in 424 non-synonymous SNPs that we identified from an analysis of 330,000 ESTs, where every allele was seen in two or more ESTs. We conclude that the LOF hypothesis remains intriguing, but any effect is likely to be small.

Some of the experimentally confirmed *SIFT* intolerant SNPs could be functionally important. We show one example in Figure 4, from the ornithine transcarbamylase (*OTC*) gene. It substitutes glycine in RJF to arginine in layer and broiler. This SNP is identical to the G188R substitution associated with hyperammonaemia in humans²⁶. Resequencing of additional domestic birds revealed a high frequency for the intolerant allele in both White Leghorns ($p=0.65$, $n=20$) and in broilers ($p=0.75$, $n=6$). In mammals, *OTC* is expressed in the liver and catalyzes the second step of the urea cycle. Chicken *OTC* is expressed in the kidney and exhibits a low enzymatic activity, with substantial variability among breeds²⁷. Preservation and sequence conservation of *OTC*, along with all other enzymes in the urea cycle¹, was unexpected because avian species excrete uric acid (not urea) as their primary component of nitrogenous waste, and were believed to be lacking a functional urea cycle. The deleterious nature of human G188R makes this an attractive candidate for phenotypic studies of avian-specific adaptations in the urea cycle.

Discussion

This analysis has provided the first global assessment of nucleotide diversity for a domestic animal in comparison to a representative of its wild ancestor. The small number of birds sequenced is compensated for by the vast number of sites examined. We detected surprisingly little difference in diversity in comparisons between RJF and domestic lines, between different domestic lines, and within domestic lines. The total rates are typically 5 SNP/kb, with the only exception being a slight reduction to 4 SNP/kb in broiler and layer lines that are maintained as closed breeding populations. In comparison, 5 SNP/kb is 6 to 7-fold larger than humans²⁸ and domestic dogs²⁹, 3-fold larger than gorillas³⁰, but similar to the diversity between different mouse subspecies³¹.

Most of the nucleotide diversity observed between and within domestic lines must have originated prior to the domestication of chickens 5,000 to 10,000 years ago. Given a neutral substitution rate of 1.8×10^{-9} sites per year for galliform birds³², we estimate that a coalescence time of 1.4 million years would be required to account for the observed rates of 5 SNP/kb. Considering that the rates observed between RJF and domestic lines are not much higher than those between domestic lines, it would seem that domestication has not resulted in a substantial genomewide loss of diversity, as would be expected had a severe population bottleneck occurred. This is important, because it contradicts the assertion that animal domestication began from a small number of individuals in a restricted geographic region³³. That is still a possible scenario for the very earliest phases of domestication, but if so, our data imply that subsequent crossing with the wild ancestor (in the first thousand years until more developed breeds were established) restored this diversity. Nevertheless, extensive diversity is consistent with the ongoing improvements in agricultural traits that have been achieved over the last 80 years, in layer and broiler lines³⁴.

The most important application for this SNP map will be in analysis of QTLs and other genetic traits. Although the density of markers far exceeds what is needed for initial mapping, the principal challenge is not in the detection of linkage but in the identification of genes underlying QTLs⁹. By itself, our SNP map is not adequate. It must be combined with novel strategies and novel resources (like mapping populations specifically designed for fine mapping). The essential problem is the lack of a one-to-one relationship between genotype and phenotype, as the latter is influenced by multiple genetic and environmental factors. This can be overcome, in experimental and domestic animals, by progeny testing and segregation analysis, which permit detailed characterization of haplotypes associated with different QTL alleles, and may eventually lead to the identification of the underlying causative mutations²². This SNP map will facilitate fine mapping.

As an example, the major *Growth1* QTL on GGA1 explains about one third of the difference between RJF and White Leghorn in adult body weight and egg weight³⁵. Initial mapping assigned this locus a ~20 cM confidence interval. Selective back-crossing using sires that have recombinant chromosomes, and QTL analysis using subsequent intercross generations, are currently employed to refine the localization to a few cM, expected to be less than ~1 Mb. This also establishes a collection of chromosomes of known QTL status. Our SNP map can then be used for haplotype analysis, assuming that the White Leghorns share a chromosomal segment - identical by descent (IBD) - with the causative mutation. The small haplotype blocks detected in this study underscore the need for a larger number of SNPs to identify such IBD segments. Although these small blocks may require greater marker density and more recombinants to identify the causative haplotype, less effort will be required to resolve the actual QTL alleles once the haplotype is found.

Materials and methods

Our broiler and layer lines are from European breeds with dramatic differences in meat and egg production traits. This specialization started only during the first half of the 20th century³⁶. The sequenced male White Cornish-type broiler is from a closed breeding population commonly used in the production of commercial meat-type hybrids (Aviagen, Newbridge, Scotland); effective population size is about 800. The female White Leghorn layer is from a closed line developed at Swedish University of Agricultural Sciences³⁷; its effective population size has been 60 to 80 birds for the past 30 years. The Chinese Silkie is used in meat/egg production and traditional Chinese medicine³⁸. Selection intensity has been low, and the sequenced female is from a large outbred population.

DNA was extracted from erythrocytes of a single bird, sheared by sonication, and size fractionated via agarose gels. Fragments of 3-kb size were ligated to SmaI-cut blunt-ended pUC18 plasmid vectors. Single colonies were grown overnight, and plasmids were extracted by an alkaline lysis protocol. Sequences were read from both ends of the insert, with vector primers and Amersham MegaBACE 1000 capillary sequencers. Roughly one million reads were generated for each bird. For broiler, layer, and Silkie, we got a total of 841,790, 841,555, and 870,556 successful reads, whose Q20 lengths add to 380,729,199-bp, 372,263,344-bp, and 397,831,117-bp, respectively.

To minimize sequencing errors, we use the *Phred* quality, Q39,40. This is related to the single base error rate by the equation: $-10 \times \log_{10}(Q)$. We use more stringent thresholds than normal⁴¹, with $Q > 25$ for the variant site and $Q > 20$ in both flanking 5-bp regions. For an insertion-deletion (indel), the variant site in the shorter allele is given the quality of its two flanking bases. We originally found many artifactual deletions relative to RJF, which upon a closer examination of the sequence reads were due to doublet peaks that got called as singlet peaks. This is an unavoidable flaw of the base caller software. Hence, we raised the indel thresholds to Q30 and Q25. We must still advise caution, and to that end, indels in simple repeats are flagged and none are counted in our summary tables.

Paralog confusion is detected in the course of the genome level *BlastN* search that determines where the read is supposed to go. Once this is known, the detailed alignments are done within *CrossMatch*⁴². Analysis of the RJF genome1 shows that recent segmental duplications typically agree to 2%. When the best and second best *BlastN* hits were more than 2% apart, and the best hit was not to a known segmental duplication, the best hit was taken. When either rule was violated, clone-end pairs information was used to resolve the ambiguity. Every alignment had to incorporate 80% of the read. Mapped back to the RJF genome, the amount of usable data for broiler, layer, and Silkie covered 190,513,980-bp, 165,154,746-bp, and 210,214,479-bp respectively.

Polymorphism rates are normalized to the length of the sequence on which we can detect SNPs. To correct for heterozygosity within a line, we compute nucleotide diversity using the

approximation⁴³: $\pi = K / \sum_{i=1}^{n-1} \frac{L}{i}$, where K is the number of variant sites found by sequencing n chromosomes in a region of length L . When comparing RJF to one of the 3 domestic lines, n can only be 2 or 3, and it is a stochastic variable, because there is a 50% chance that any two overlapping reads are from the same chromosome. When there are m overlapping reads, the denominator is $\frac{L}{2^{m-1}} \cdot \left(1 + (2^{m-1} - 1) \cdot \left(1 + \frac{1}{2}\right)\right)$. We then sum over all possible regions, with different L and m for each region, to get what we call the “effective length”. Similar

considerations are used to compute SNP rates within a line, except that n is 1 or 2, and as a result, the denominator becomes $\frac{L}{2^{m-1}} \cdot (2^{m-1} - 1)$.

We compute gene context relative to 5 different data sets. The first 3 are based on experimentally derived genes and the last 2 are based on computer annotations. Riken1 is a set of 1758 full-length cDNAs taken from bursal B-cells of a two week old CB inbred44. GenBank refers to 1178 chicken genes with “complete CDS” designation, downloaded as version 2003-12-15. BBSRC is a set of 1184 cDNAs, taken from a larger group of 18,034 cDNAs45, which are full-length using a *TBlastX* mapping to vertebrate Refseq and *BlastX* mapping to SWALL. Merging all 3 data sets, we have 3868 non-redundant genes. For the detailed gene models, we do a genome level search in *BLAT*46 and use *SIM4*47 to compute the exon-intron boundaries. The last two data sets are for 995 chicken orthologs of human disease genes and 17,709 non-redundant Ensembl genes.

Additional details are in Supplement M (Methods).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Beijing Institute of Genomics of Chinese Academy of Sciences Gallus gallus SNP discovery and analysis was supported by Chinese Academy of Sciences (KSCX2-SW-223), State Development Planning Commission, Ministry of Science and Technology (2002AA104250; 2004AA231050; 2001AA231061; 2001AA231101), National Natural Science Foundation of China (30200163; 90208019), Beijing Municipal Government, Zhejiang Provincial Government, Hangzhou Municipal Government, Zhejiang University, and China National Grid. Some equipment and reagents were provided by Wellcome Trust and Sanger Institute of the UK. Recent segmental duplications were analyzed by G. Cheng and E.E. Eichler. Riken1 cDNAs were provided by R. Caldwell and J.M. Buerstedde. Noncoding conserved motifs were analyzed by J. Taylor and W. Miller. **Washington University School of Medicine** Gallus gallus sequence generation was supported by National Human Genome Research Institute. **Uppsala University** HE was supported by Swedish Research Council, Knut and Alice Wallenberg Foundation, and Royal Academy of Sciences. LA was supported by Wallenberg Consortium North, Foundation for Strategic Research, and Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning. **Institute for Animal Health** PK, NB, JRY, and JK were supported by BBSRC. **Iowa State University** SJL was supported by Hatch Act and State of Iowa. Skeletal data for ISU resource population was collected by C. Ashwell and A. Mitchell. **Roslin Institute** PMH, AL, DJK, and DWB were supported by BBSRC. SNP genotyping was partially funded by Cobb-Vantress. **USDA-ARS Avian Disease and Oncology Laboratory** J. Kenyon and N. Evenson provided technical assistance. **University of Oxford** CPP was supported by UK Medical Research Council. **University of Manchester Institute of Science and Technology** SJH was supported by BBSRC. **University of Sheffield** SAW was supported by BBSRC.

We dedicate this paper to Nat Bumstead, who died during preparation of the manuscript. Nat was recognised as a major figure in researching the genetics of disease resistance in poultry. He worked tirelessly to realise the sequence of the chicken genome, which led in part to this consortium.

International Chicken Polymorphism Map Consortium

†: International Chicken Polymorphism Map Consortium

(Group contributions are listed by their order of appearance in the manuscript)

Polymorphism discovery and analysis: *Beijing Institute of Genomics of Chinese Academy of Sciences*: Gane Ka-Shu Wong^{1-3*}, ‡, Bin Liu^{1*}, Jun Wang^{12*}, Yong Zhang^{14*}, Xu

¹Beijing Institute of Genomics of Chinese Academy of Sciences, Beijing Genomics Institute, Beijing Proteomics Institute, Beijing 101300, China

Yang^{1*}, Zengjin Zhang¹, Qingshun Meng¹, Jun Zhou¹, Dawei Li¹, Jingjing Zhang¹, Peixiang Ni¹, Songgang Li¹⁴, Longhua Ran⁵, Heng Li¹⁶, Jianguo Zhang¹, Ruiqiang Li¹, Shengting Li¹, Hongkun Zheng¹, Wei Lin¹, Guangyuan Li¹, Xiaoling Wang¹, Wenming Zhao¹, Jun Li¹, Chen Ye¹, Mingtao Dai¹, Jue Ruan¹, Yan Zhou², Yuanzhe Li¹, Ximiao He¹, Yunze Zhang¹, Jing Wang¹⁴, Xiangang Huang¹, Wei Tong¹, Jie Chen¹, Jia Ye¹², Chen Chen¹, Ning Wei¹, Guoqing Li¹, Le Dong¹, Fengdi Lan¹, Yongqiao Sun¹, Zhenpeng Zhang¹, Zheng Yang¹, Yingpu Yu², Yanqing Huang¹, Dandan He¹, Yan Xi¹, Dong Wei¹, Qiuhui Qi¹, Wenjie Li¹, Jianping Shi¹, Miaoheng Wang¹, Fei Xie¹, Jianjun Wang¹, Xiaowei Zhang¹, Pei Wang¹, Yiqiang Zhao⁷, Ning Li⁷, Ning Yang⁷, Wei Dong¹, Songnian Hu¹, Changqing Zeng¹, Weimou Zheng¹⁶, Bailin Hao¹⁶

Genome sequence of Red Jungle Fowl: *Washington University School of Medicine:* LaDeana W. Hillier⁸, Shiaw-Pyng Yang⁸, Wesley C. Warren⁸, Richard K. Wilson⁸

Molecular evolution: *Uppsala University:* Mikael Brandström⁹, Hans Ellegren⁹

Population genotyping, BAC sequences and haplotypes: *Wageningen University:* Richard P.M.A. Crooijmans¹⁰, Jan J. van der Poel¹⁰, Henk Bovenhuis¹⁰, Martien A.M. Groenen¹⁰; *Lawrence Livermore National Laboratory:* Ivan Ovcharenko¹¹¹², Laurie Gordon¹¹¹³, Lisa Stubbs¹¹; *DOE Joint Genome Institute:* Susan Lucas¹³, Tijana Glavina¹³, Andrea Aerts¹³

Examples of application to complex traits: *Institute for Animal Health:* Pete Kaiser¹⁴, Lisa Rothwell¹⁴, John R. Young¹⁴, Sally Rogers¹⁴, Brian A. Walker¹⁴, Andy van Hateren¹⁴, Jim Kaufman¹⁴, Nat Bumstead¹⁴; *Iowa State University:* Susan J. Lamont¹⁵, Huaijun Zhou¹⁵; *Roslin Institute:* Paul M. Hocking¹⁶, David Morrice¹⁶, Dirk-Jan de Koning¹⁶, Andy Law¹⁶, Neil Bartley¹⁶, David W. Burt¹⁶; *USDA-ARS Avian Disease and Oncology Laboratory:* Henry Hunt¹⁷, Hans H. Cheng¹⁷

Domestication and selection: *Uppsala University:* Ulrika Gunnarsson¹⁸, Per Wahlberg¹⁸, Leif Andersson^{1819,‡}; *Karolinska Institutet:* Ellen Kindlund²⁰, Martti T. Tammi²⁰²¹, Björn Andersson²⁰

Human disease genes: *University of Oxford:* Caleb Webber²², Chris P. Ponting²²

³UW Genome Center, Department of Medicine, University of Washington, Seattle, WA 98195, USA

*These authors contributed equally to this work.

²James D. Watson Institute of Genome Sciences of Zhejiang University, Hangzhou Genomics Institute, Key Laboratory of Bioinformatics of Zhejiang Province, Hangzhou 310007, China

⁴College of Life Sciences, Peking University, Beijing 100871, China

⁵Beijing North Computation Center, Beijing 100091, China

⁶The Institute of Theoretical Physics Chinese Academy of Sciences, Beijing 100080, China

⁷China Agricultural University, Beijing 100094, China

⁸Genome Sequencing Center, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St. Louis, MO 63108, USA

⁹Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 34 Uppsala, Sweden

¹⁰Animal Breeding and Genetics Group, Wageningen University, Marijkeweg 40, 6709 PG Wageningen, The Netherlands

¹¹Genome Biology Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

¹²Energy, Environment, Biology and Institutional Computing, Lawrence Livermore National Laboratory

¹³DOE Joint Genome Institute, Walnut Creek, CA 94598, USA

¹⁴Institute for Animal Health, Compton, Berkshire RG20 7NN, UK

¹⁵Department of Animal Science, Iowa State University, Ames, IA 50011, USA

¹⁶Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, UK

¹⁷USDA-ARS Avian Disease and Oncology Laboratory, 3606 E. Mount Hope Rd., East Lansing, MI 48823, USA

¹⁸Department of Medical Biochemistry and Microbiology, Uppsala University, Box 597, SE-751 24 Uppsala, Sweden

¹⁹Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-751 24 Uppsala, Sweden

²⁰Center for Genomics and Bioinformatics, Karolinska Institutet, SE-171 77 Stockholm, Sweden

²¹Departments of Biological Sciences and Biochemistry, National University of Singapore, Singapore

EST-based SNP data: *University of Manchester Institute of Science and Technology:* Ian M. Overton²³, Paul E Boardman²³, Haizhou Tang²³, Simon J. Hubbard²³; *University of Sheffield:* Stuart A Wilson²⁴

Scientific management: *Beijing Institute of Genomics of Chinese Academy of Sciences:* Jun Yu¹², Jian Wang¹², HuanMing Yang^{12,‡}

¹Beijing Institute of Genomics of Chinese Academy of Sciences, Beijing Genomics Institute, Beijing Proteomics Institute, Beijing 101300, China

²James D. Watson Institute of Genome Sciences of Zhejiang University, Hangzhou Genomics Institute, Key Laboratory of Bioinformatics of Zhejiang Province, Hangzhou 310007, China

³UW Genome Center, Department of Medicine, University of Washington, Seattle, WA 98195, USA

⁴College of Life Sciences, Peking University, Beijing 100871, China

⁵Beijing North Computation Center, Beijing 100091, China

⁶The Institute of Theoretical Physics Chinese Academy of Sciences, Beijing 100080, China

⁷China Agricultural University, Beijing 100094, China

⁸Genome Sequencing Center, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St. Louis, MO 63108, USA

⁹Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 34 Uppsala, Sweden

¹⁰Animal Breeding and Genetics Group, Wageningen University, Marijkewg 40, 6709 PG Wageningen, The Netherlands

¹¹Genome Biology Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

¹²Energy, Environment, Biology and Institutional Computing, Lawrence Livermore National Laboratory

¹³DOE Joint Genome Institute, Walnut Creek, CA 94598, USA

¹⁴Institute for Animal Health, Compton, Berkshire RG20 7NN, UK

¹⁵Department of Animal Science, Iowa State University, Ames, IA 50011, USA

¹⁶Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, UK

¹⁷USDA-ARS Avian Disease and Oncology Laboratory, 3606 E. Mount Hope Rd., East Lansing, MI 48823, USA

¹⁸Department of Medical Biochemistry and Microbiology, Uppsala University, Box 597, SE-751 24 Uppsala, Sweden

¹⁹Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-751 24 Uppsala, Sweden

²⁰Center for Genomics and Bioinformatics, Karolinska Institutet, SE-171 77 Stockholm, Sweden

²¹Departments of Biological Sciences and Biochemistry, National University of Singapore, Singapore

²²MRC Functional Genetics Unit, University of Oxford, Department of Human Anatomy and Genetics, South Parks Road, Oxford OX1 3QX, UK

²³Department of Biomolecular Sciences, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK

²²MRC Functional Genetics Unit, University of Oxford, Department of Human Anatomy and Genetics, South Parks Road, Oxford OX1 3QX, UK

²³Department of Biomolecular Sciences, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK

²⁴Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK

²⁴Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK

†**Corresponding authors:** Gane Ka-Shu Wong gksw@genomics.org.cn, Leif Andersson leif.andersson@imbim.uu.se, HuanMing Yang hyang@genomics.org.cn.

*These authors contributed equally to this work.

References

1. The Chicken Genome Sequencing Consortium. The sequence of the chicken genome, *Gallus gallus*. Nature. (companion paper).
2. Pisenti JM, et al. Avian genetic resources at risk: an assessment and proposal for conservation of genetic stocks in the USA and Canada. Avian Poult. Biol. Rev. 2001; 12:1–102.
3. Dodgson, JB.; Romanov, MN. Use of chicken models for the analysis of human disease. In: Dracopoli, NC., et al., editors. Current Protocols in Human Genetics. Hoboken: John Wiley & Sons; 2004. p. 15.5.1–11.
4. Nicholas FW. Online Mendelian Inheritance in Animals (OMIA): a comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. Nucleic Acids Res. 2003; 31:275–277. <http://www.angis.org.au/Databases/BIRX/omia>. [PubMed: 12520001]
5. ChickAce database from the Animal Science Group of the Wageningen University and Research Center; <https://acedb.asg.wur.nl>
6. Groenen MA, et al. A consensus linkage map of the chicken genome. Genome Res. 2000; 10:137–147. [PubMed: 10645958]
7. Groenen, MA.; Crooijmans, RP. Structural genomics: integrating linkage, physical and sequence maps. In: Muir, WM.; Aggrey, SE., editors. Poultry Genetics, Breeding and Biotechnology. Wallingford: CABI Publishing; 2003. p. 497–536.
8. Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 2002; 34:275–305. [PubMed: 12081799]
9. Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. Nat. Rev. Genet. 2004; 5:202–212. [PubMed: 14970822]
10. There are only 20-kb of aligned sequence on GGA16, and if we were to remove it, the total SNP rate would only change by 0.02%
11. Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature. 1992; 356:519–520. [PubMed: 1560824]
12. Nachman MW. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 2001; (9):481–485. [PubMed: 11525814]
13. Crooijmans RP, Vrebalov J, Dijkhof RJ, van der Poel JJ, Groenen MA. Two-dimensional screening of the Wageningen chicken BAC library. Mamm. Genome. 2000; 11:360–363. [PubMed: 10790534]
14. Sundstrom H, Webster MT, Ellegren H. Reduced variation on the chicken Z chromosome. Genetics. 2004; 167:377–385. [PubMed: 15166162]
15. Ikeobi CO, et al. Quantitative trait loci for muscling in a broiler layer cross. Livest. Prod. Sci. 2004; 87:143–151.
16. Sewalem A, et al. Mapping of quantitative trait loci for body weight at three, six, and nine weeks of age in a broiler layer cross. Poult. Sci. 2002; 81:1775–1781. [PubMed: 12512565]
17. Li H, et al. Chicken quantitative trait loci for growth and body composition associated with transforming growth factor- β genes. Poult. Sci. 2003; 82:347–356. [PubMed: 12705392]
18. Zhou H, Li H, Lamont SJ. Genetic markers associated with antibody response kinetics in adult chickens. Poult. Sci. 2003; 82:699–708. [PubMed: 12762390]
19. Gallagher, G.; Eskdale, J.; Bidwell, JL. Cytokine genetics - polymorphisms, functional variations and disease associations. In: Thomson, AW.; Lotze, MT., editors. The Cytokine Handbook. 4th Edition. London: Academic Press; 2003. p. 19–55.
20. Bumstead, N., et al. EU Project FAIR3 PL96-1502 New Molecular Approaches for Improved Poultry Vaccines. Compton: Institute for Animal Health; 2000.

21. Maynard-Smith J, Haigh J. The hitch-hiking effect of a favourable gene. *Genet. Res.* 1974; 23:23–35. [PubMed: 4407212]
22. Van Laere AS, et al. regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature.* 2003; 425:832–836. [PubMed: 14574411]
23. Olson MV. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* 1999; 64:18–23. [PubMed: 9915938]
24. Grobet L, et al. A deletion in the bovine myostatin gene causes the double- muscled phenotype in cattle. *Nat. Genet.* 1997; 17:71–74. [PubMed: 9288100]
25. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res.* 2001; 11:863–874. <http://blocks.fhcrc.org/sift/SIFT.html>. [PubMed: 11337480]
26. Gilbert-Dussardier B, et al. Partial duplication [dup. TCAC (178)] and novel point mutations (T125M, G188R, A209V, and H302L) of the ornithine transcarbamylase gene in congenital hyperammonemia. *Hum. Mutat.* 1996; 8:74–76. [PubMed: 8807340]
27. Tamir H, Ratner S. Enzymes of arginine metabolism in chicks. *Arch. Biochem. Biophys.* 1963; 102:249–258. [PubMed: 14064979]
28. Sachidanandam R, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* 2001; 409:928–933. [PubMed: 11237013]
29. Parker HG, et al. Genetic structure of the purebred domestic dog. *Science.* 2004; 304:1160–1164. [PubMed: 15155949]
30. Yu N, Jensen-Seaman MI, Chemnick L, Ryder O, Li WH. Nucleotide diversity in gorillas. *Genetics.* 2004; 166:1375–1383. [PubMed: 15082556]
31. Lindblad-Toh K, et al. Large-scale discovery and genotyping of single- nucleotide polymorphisms in the mouse. *Nat. Genet.* 2000; 24:381–386. [PubMed: 10742102]
32. Axelsson E, Smith NG, Sundstrom H, Berlin S, Ellegren H. Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Mol. Biol. Evol.* 2004; 21:1538–1547. [PubMed: 15140948]
33. Mason, IL., editor. *Evolution of Domesticated Animals*. New York: Longman, Inc.; 1984.
34. Arthur, JA.; Albers, GA. Industrial perspective on problems and issues associated with poultry breeding. In: Muir, WM.; Aggrey, SE., editors. *Poultry Genetics, Breeding and Biotechnology*. Wallingford: CABI Publishing; 2003. p. 1-12.
35. Kerje S, et al. The twofold difference in adult size between the red junglefowl and White Leghorn chickens is largely explained by a limited number of QTLs. *Anim. Genet.* 2003; 34:264–274. [PubMed: 12873214]
36. Crawford, RD., editor. *Poultry Breeding and Genetics*. New York: Elsevier Science; 1990.
37. Liljedahl LE, Kolstad N, Sorensen P, Maijala K. Scandinavian selection and cross-breeding experiment with laying hens. I. Background and general outline. *Acta Agricult. Scand.* 1979; 29:273–285.
38. Niu D, et al. The origin and genetic diversity of Chinese native chicken breeds. *Biochem. Genet.* 2002; 40:163–174. [PubMed: 12137331]
39. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 1998; 8:175–185. [PubMed: 9521921]
40. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 1998; 8:186–194. [PubMed: 9521922]
41. Altshuler D, et al. A SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature.* 2000; 407:513–516. [PubMed: 11029002]
42. Green P. CrossMatch is the underlying alignment tool for the Phrap assembly software. at <http://www.phrap.org>
43. Cargill M, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 1999; 22:231–238. [PubMed: 10391209]
44. Caldwell R, et al. A large collection of bursal full-length cDNA sequences to facilitate gene function analysis. *Genome Biol.* (companion issue).
45. Hubbard SJ, et al. Transcriptome analysis for the chicken based on 19,626 finished cDNA sequences and 485,337 expressed sequence tags. *Genome Res.* (companion issue).

46. Kent WJ. BLAT - the BLAST-like alignment tool. *Genome Res.* 2002; 12:656–664. <http://www.genome.ucsc.edu/cgi-bin/hgBlat>. [PubMed: 11932250]
47. Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* 1998; 8:967–974. <http://globin.cse.psu.edu/html/docs/sim4.html>. [PubMed: 9750195]
48. Wang J, et al. ChickVD: a sequence variation database for the chicken genome. *Nucleic Acids Res.* Jan.2005



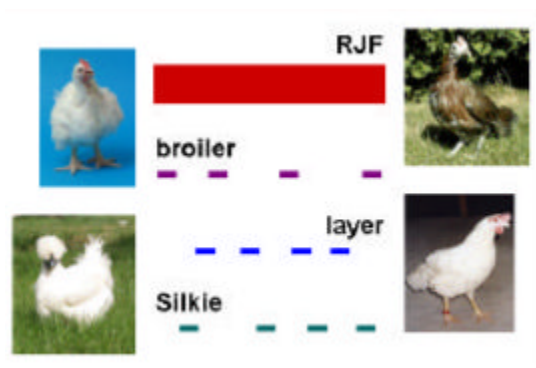


Figure 1. SNP discovery experiment. We sampled 3 domestic chickens at 1/4 coverage each and compared the resultant sequence to the 6.6x draft genome of Red Jungle Fowl (RJF). Chicken photographs shown here are provided by Bill Payne (RJF), Paul Hocking (broiler), Leif Andersson (layer), and Ning Yang (Silkie).

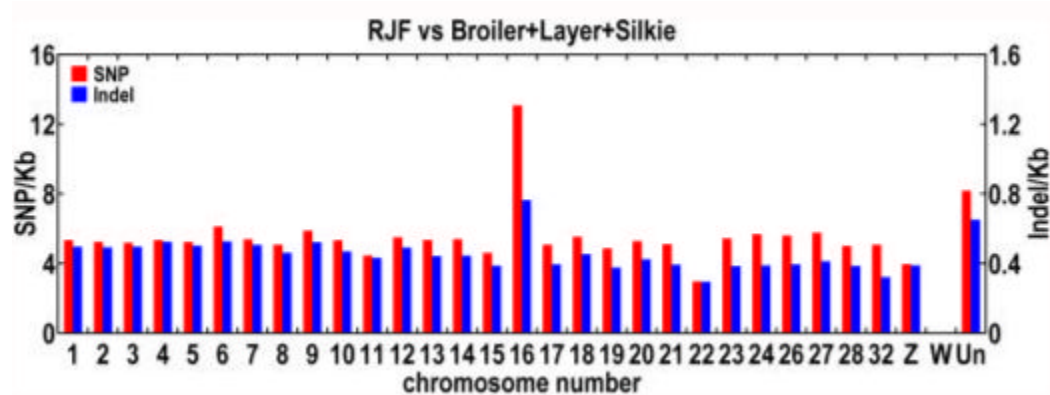


Figure 2. SNP and indel rates versus chromosome number. We excluded all sequences with “random” chromosome positions. Because of the assembly problems on W, it is not shown. The rates are computed as an average of all 3 domestic lines.

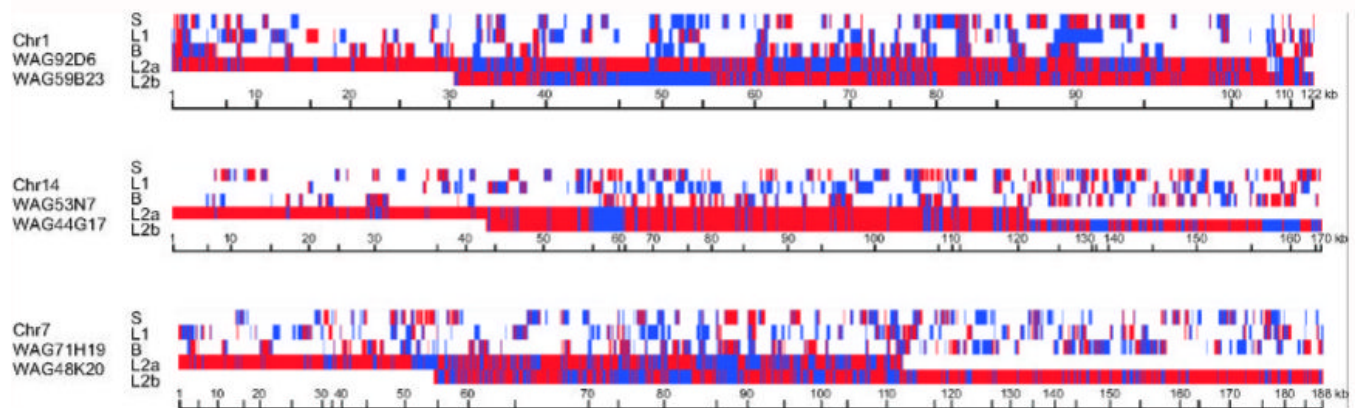


Figure 3.

Detailed haplotype patterns in 3 regions, each covered by 2 overlapping BACs from the second layer line (L2). The primary SNP data are labeled B (broiler), L1 (layer), and S (Silkie). All comparisons are to RJF, and we show only those sites where a SNP is identified in at least one of the 4 lines. Hence, the horizontal scale is linear in the number of SNP sites, but non-linear for size. BLUE colors indicate where a particular line agrees with RJF, while RED colors indicate where it does not. Overlapping BACs on GGA1 and GGA7, but not GGA14, are clearly from different haplotypes.

Ornithine transcarbamylase (*OTC*)
 188

Human	HYSSLK G LTLSWIGDGN
Pig	--GA-----
Mouse	--G-----
Rat	--G-----
Chicken, WT	--GG-N---IA-----
Chicken, MUT	--GG-N R --IA-----

Figure 4. Multi-species alignments for ornithine transcarbamylase (*OTC*), indicating non-synonymous substitutions relative to human protein. *SIFT* intolerant position is indicated by site number and bold-faced lettering. WT=wild type. MUT=mutant.

Table 1

Frequency of SNPs in different comparisons of RJF and the 3 domestic chicken lines. In addition, we show comparisons involving 3.8-Mb of finished BAC sequence from another line of the layer (White Leghorn) breed. SNP rates are an estimate of nucleotide diversity (π), as embodied by the effective length, which considers how much of the data is of sufficiently good quality to actually detect SNPs and the probability that overlapping reads might be derived from homologous chromosomes

	# of SNPs	L(effective)	SNP/kb
Wild versus domestic			
RJF-Broiler	1,041,948	197,431,517	5.28
RJF-Layer	889,377	170,586,544	5.21
RJF-Silkie	1,217,817	217,841,171	5.59
Between domestic lines			
Broiler-Layer	194,605	37,506,800	5.19
Broiler-Silkie	257,849	47,554,311	5.42
Layer-Silkie	246,954	42,682,304	5.79
Within domestic lines			
Broiler-Broiler	59,227	13,835,075	4.28
Layer-Layer	40,412	10,863,595	3.72
Silkie-Silkie	83,630	15,253,383	5.48
Compare to layer BACs			
RJF-to-BAC	20,925	3,809,567	5.49
BAC-Broiler	4,404	847,456	5.20
BAC-Layer	3,904	740,392	5.27
BAC-Silkie	5,089	925,738	5.50



Table 2
RJF-Broiler polymorphisms

Frequency of sequence polymorphisms between RJF and broiler, decomposed by functional context based on three non-redundant gene sets of 3868 confirmed mRNA transcripts, 995 chicken orthologs of known human disease genes, and 17,709 Ensembl annotations. Human-chicken motifs are conserved sequences that exhibit no evidence of being genic in origin. Gene regions are subdivided into 5'-UTR, coding exon, intron, and 3'-UTR. Ka and Ks indicate non-synonymous and synonymous rates

	SNP/kb	Indel/kb	# of SNP	# of Indel
Confirmed mRNA transcripts				
5'-UTR	3.45	0.46	203	27
coding region	2.11	0.05	1,772	41
non-synonymous (Ka)	0.73			
synonymous (Ks)	7.44			
introns	5.70	0.52	86,586	7,915
3'-UTR	3.40	0.42	1,946	243
Human disease genes				
coding region	2.74	0.04	1,005	15
non-synonymous (Ka)	1.10			
synonymous (Ks)	9.40			
introns	5.36	0.49	27,768	2,553
Ensembl (final version 040427)				
5'-UTR	4.22	0.37	616	54
coding region	2.71	0.06	12,229	276
non-synonymous (Ka)	1.17			
synonymous (Ks)	8.28			
introns	5.64	0.52	367,361	33,869
3'-UTR	3.92	0.43	2,130	236
Human-chicken motifs	2.41	0.25	3,636	379
Genomewide average	5.28	0.48	1,041,948	94,578