Genome-Wide Patterns of Genetic Variation in Two Domestic Chickens

Wen-Lang Fan[1,2,†], Chen Siang Ng[1,†], Chih-Feng Chen[3,4], Mei-Yeh Jade Lu[1], Yu-Hsiang Chen[1], Chia-Jung Liu[1], Siao-Man Wu[1], Chih-Kuan Chen[1,5], Jiun-Jie Chen[1], Chi-Tang Mao[1,6,7], Yu-Ting Lai[1], Wen-Sui Lo[1], Wei-Hua Chang[3], and Wen-Hsiung Li[1,2,8,*]

[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

[2]Genomics Research Center, Academia Sinica, Taipei, Taiwan

[3]Department of Animal Sciences, National Chung Hsing University, Taichung, Taiwan

[4]Center for the Integrative and Evolutionary Galliformes Genomics (iEGG Center), National Chung Hsing University, Taichung, Taiwan

[5]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan

[6]Molecular Biology of Agricultural Sciences, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan

[7]Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan

[8]Department of Ecology and Evolution, University of Chicago

*Corresponding author: E-mail: whli@sinica.edu.tw.

[†]These authors contributed equally to this work.

Accepted: June 23, 2013

Data deposition: The full data sets have been submitted to NCBI Sequence Read Archive (SRA) under accession nos. SRX286765, SRX286766, SRX286773, SRX286776, SRX286777, SRX286779-SRX286781, SRX286798, and SRX286799.

Abstract

Domestic chickens are excellent models for investigating the genetic basis of phenotypic diversity, as numerous phenotypic changes in physiology, morphology, and behavior in chickens have been artificially selected. Genomic study is required to study genome-wide patterns of DNA variation for dissecting the genetic basis of phenotypic traits. We sequenced the genomes of the Silkie and the Taiwanese native chicken L2 at ~23- and 25-fold average coverage depth, respectively, using Illumina sequencing. The reads were mapped onto the chicken reference genome (including 5.1% Ns) to 92.32% genome coverage for the two breeds. Using a stringent filter, we identified ~7.6 million single-nucleotide polymorphisms (SNPs) and 8,839 copy number variations (CNVs) in the mapped regions; 42% of the SNPs have not found in other chickens before. Among the 68,906 SNPs annotated in the chicken sequence assembly, 27,852 were nonsynonymous SNPs located in 13,537 genes. We also identified hundreds of shared and divergent structural and copy number variants in intronic and intergenic regions and in coding regions in the two breeds. Functional enrichments of identified genetic variants were discussed. Radical nsSNP-containing immunity genes were enriched in the QTL regions associated with some economic traits for both breeds. Moreover, genetic changes involved in selective sweeps were detected. From the selective sweeps identified in our two breeds, several genes associated with growth, appetite, and metabolic regulation were identified. Our study provides a framework for genetic and genomic research of domestic chickens and facilitates the domestic chicken as an avian model for genomic, biomedical, and evolutionary studies.

Key words: single nucleotide polymorphism, whole genome resequencing, genetic variation, CNVs, chicken.

Introduction

The domestic chicken (*Gallus gallus domesticus*) was the first bird domesticated and has been deeply integrated into the human culture for more than 8,000 years (Dohner 2001; Price 2002). Chickens and their eggs have been the most important supply of animal protein for human populations. All chicken breeds are considered to share a common ancestor in the red jungle fowl (*G. gallus gallus*), which is still wildly distributed in parts of India and Southeast Asia (Crawford 1990; Miao et al. 2013), although some recent genetic evidence suggested that several other jungle fowl species (grey junglefowl [*G. sonnerati*], Geylon junglefowl [*G. lafayettei*], and green junglefowl [*G. varius*]) might also have contributed to the ancestral genetic background of domestic chickens (Nishibori et al. 2005; Eriksson et al. 2008; Sawai et al. 2010).

Charles Darwin recognized that phenotypic variations in domesticated plants and animals arise from selective breeding, giving credence to evolution in natural populations by natural selection. Domesticated animals have played a significant role in introducing the ideas of evolution, particularly natural selection, in Darwin's *On the Origin of Species* (Darwin 1859). Darwin further expanded his ideas in the book *The Variation of Animals and Plants under Domestication* (1868), using traits exaggerated by artificial selection to demonstrate the potential power of natural and sexual selection (Darwin 1868).

The vast diversity of phenotypes among breeds that has been created by selective breeding of domesticated animals can compare to those observed among wild species in nature. In addition to agricultural applications, domesticated animals have also contributed to basic and medical biology, as they provide excellent opportunities for unraveling the genetic and molecular basis of phenotypic variations (Andersson 2001; Andersson and Georges 2004). The chicken is the most variable bird (Somes 1988), because there are hundreds of chicken breeds, integrating various mutations affecting body size, reproduction, growth rate, posture, color, feather structure and distribution, comb shape, and behavior.

Domestic animals have some advantages over traditional model organisms. Extreme phenotypes that may not be found in the wild can be maintained by artificial selection for aesthetic or economic demands. Mutations of biologically important traits have a greater chance of appearing and of being selected due to a large population size and higher longevity in domestic animals, providing us an exceptional opportunity to identify novel functions for specific genes. The identification of genetic variants controlling morphology in the chicken population has helped us understand the genetics and genomics of both simple and complex traits of birds (Wright et al. 2009; Dorshorst et al. 2011; Mou et al. 2011; Imsland et al. 2012; Johnsson et al. 2012; Ng et al. 2012; Shinomiya et al. 2012; Wang et al. 2012a).

Among the chicken breeds, the Silkie chicken exhibits several phenotypic variations that are not commonly seen in other chicken breeds. These unique features include elongated feathers on the crown of the head, fluffy plumage, dark blue flesh, viscera, and bones, blue earlobes, feathered legs and feet, and five toes on each foot. Moreover, they have several color variants. Most of these unique traits are controlled by single Mendelian genes. These mutant alleles include crest (*Cr*) and muff (*Mb*), polydactyly (*Po*), ptilopody (*Pt*), hyperpigmentation (*Fm*), and silkiness or hookless (*h*). Genomic regions associated with morphological traits of the Silkie have been published (Dorshorst et al. 2010), giving us an opportunity to perform fine mapping for the causative genes and mutations. Furthermore, the International Chicken Polymorphism Map Consortium has described 2.8 million single-nucleotide polymorphisms (SNPs) among broiler, layer, and Silkie (Wong et al. 2004; Wang et al. 2005), but the coverage was at one-quarter for each of these domestic chicken lines, which is too low for finding causative mutations.

Another breed used in the present study is the Taiwan country chicken (TCC) L2 breed. This breed originated from a single population and were subjected to selection for ~30 years for egg production and body weight/comb size (Chao and Lee 2001). This local breed represents both a heritage and a reservoir of genetic variability that deserves to be explored and properly managed.

Advances in sequencing technologies allow whole genomes to be sequenced more economically and efficiently than ever before, providing an excellent opportunity to discover numerous genetic polymorphisms in a genome and for quantitative trait locus (QTL) analysis and marker-assisted selection. Therefore, we conducted a whole-genome analysis to examine the genetic and genomic features of the Silkie and L2 chickens, providing detailed genetic information of these chicken breeds.

We identified genetic variants, including SNPs, insertion/deletion polymorphisms (indels), structural variations (SVs), and copy number variations (CNVs) in these two chicken genomes. Genetic variants were annotated and their potential impacts on gene structures and functions were examined. The genomic sequences were also explored for mutations associated with QTLs in chickens. Furthermore, we analyzed biological processes and molecular functions enriched for genetic variants. The genome resources presented here are useful for breeding and comparative genomics in chicken and related species.

Materials and Methods

Chickens

We obtained genomic DNA samples from the following two chickens: a male Silkie and a male TCC L2 that were raised at the National Chung Hsing University, Taichung, Taiwan.

DNA Library Construction and Sequencing

Genomic DNAs were extracted from peripheral venous blood using the QIAGEN - Gentra Puregene Cell Kit (Qiagen, Venlo, Netherlands). The purified DNAs were assessed for purity and quality by NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA), Qubit (Invitrogen Corp., Carlsbad, CA, USA), and gel electrophoresis. High quality genomic DNAs were then selected for paired-end (PE) library preparation with the method adopted from manufacturer's protocol (Illumina Inc., San Diego, CA, USA). The genomic DNA was sonicated to the 200–600 bp range with a major peak at 400 bps, followed by end repair, A-tailing, and adaptor ligation. The ligation reaction was then separated on 2% low-range agarose gel (Bio-Rad, Hercules, CA, USA), and four fractions were made at 100-bp intervals on the gel. The DNAs were purified from gel slices using Qiaquick Gel Extraction kit (Qiagen, Venlo, Netherlands) and amplified by 12 cycles of polymerase chain reaction (PCR) using the reagents provided in the Illumina kit. The PCR products were purified using Ampure beads (Beckman Coulter Inc., Brea, CA, USA) and the resulting libraries were determined for quantification and size profiling using Qubit and BioAnalyzer 1000 (Agilent Technologies, Santa Clara, CA, USA). PE sequencing was performed on Illumina GA IIx in the High Throughput Sequencing Core Facility, Biodiversity Research Center, Academia Sinica, Taiwan. High-quality reads (pass-filter rate 78–92%) of PE 2×120 nt were obtained from five lanes for each chicken breed.

Public Data Used

The chicken reference genome, together with annotation of genes and repeats, was downloaded from the Ensembl Genome Browser (http://www.ensembl.org/), which has the same sequence as the NCBI build 2.1. The chicken SNP database (dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/website) were used to compare the putative SNPs we found. The Chicken QTLdb (http://www.animalgenome.org/cgi-bin/QTLdb/GG/index) were used to locate candidate genes. These data were retrieved in March 2011.

Short Reads Alignment

For read alignment and consensus assembly, we used BWA ver. 0.5.8 (http://biobwa.sourceforge.net/) (Li and Durbin 2010). The following parameters were used: maximum edit distance (maxDiff) = 0.04, maximum number of gap opens (maxGapO) = 1, maximum number of gap extensions (maxGapE) = −1 (disabling long gaps), disallow a long deletion within bp (nDelTail) = 10, disallow an indel within bp (nIndelEnd) = 5, take the first subsequence as seed (seedLen) = 32, maximum edit distance in the seed (maxSeedDiff) = 2, number of threads (nThrds) = 8, mismatch penalty (misMsc) = 3, gap open penalty (gapOsc) = 11, gap extension penalty (gapEsc) = 4. The reads mapped to multiple chromosomal positions and unmapped reads were discarded. We only used reads mapped to a unique position on the reference chicken genome for SNP calling.

SNP Calling

To call SNPs, we first filtered out reads with mapping quality score <20. Then, SAMtools (Li et al. 2009) were used and additional filters were applied as follows: minimum read depth = 3, minimum read depth calling the SNP = 2, and a 20% cutoff of percent aligned reads calling the SNP per total mapped reads at the SNP sites. These identified SNPs were also filtered with more stringent parameters (i.e., minimum depth = 4, minimum SNP = 2, and 20% or higher aligned reads calling the SNP; and minimum depth = 5, minimum SNP = 2, and 20% or higher aligned reads calling the SNP). We distinguished heterozygous and homozygous SNPs using an 80% cutoff of percent aligned reads calling the SNP. BWA was also used to estimate the sequence read depth, which influences the coverage and accuracy of SNP calling. After SNP calling, the SNPs were annotated using the Ensembl gene sets (17,934 genes; available from the Ensembl BioMart site [http://www.ensembl.org/biomart/]). The SNPs and indels in gene regions were annotated using the custom software and the ANNOVAR annotation tool (Wang et al. 2010a). Functional annotations of these loci were compared with the complete genome using annotations from the Database for Annotation, Visualization, and Discovery (DAVID) (Huang et al. 2007, 2009), which uses fuzzy clustering to group genes into functionally related classes based on the similarity of their annotations.

SNP and Small Indel Validation

PCR primers to validate SNPs and small indels were designed using PRIMER3 software (http://frodo.wi.mit.edu/primer3) and were used to amplify 200–1000-bp fragments that were positioned within each selected SNP and indel region according to the reference genome sequence of the red jungle fowl. The optimal primer length was set at 20 bp and all other PRIMER 3 defaults were used. PCR was performed on the Silkie and L2 genomic DNA using TaKaRa Ex Taq® DNA Polymerase (Takara Bio Inc., Shiga, Japan). The PCR reaction was done in 20 µl containing 5 ng of genomic DNA, 2.5 mM Mg^{2+}, 0.5 mM dNTPs, 0.2 µM of each primer, 2 µl of 10× PCR buffer, and 0.5 U of Taq DNA polymerase and was conducted using 35 thermal cycles: one cycle of pre-incubation at 94 °C for 3 min and 94 °C for 30 s, 55–60 °C for 30 s, and 72 °C for 2 min, 72 °C for 7 min at the end of the final cycle. The PCR products were run on 1–2% agarose gels containing ethidium bromide and visualized using a UV transilluminator. The approximate sizes of the PCR products were estimated by running molecular weight markers (GeneRuler 100 bp DNA Ladder; Thermo Fisher Scientific, Waltham, MA, USA) on each gel. All PCR products were sequenced directly after treatment with exonuclease I and calf intestinal alkaline phosphatase (New England BioLabs, Ipswich, MA, USA) by standard methods. Each PCR product was sequenced by the BigDye terminator cycle sequencing kit and sequencer, Applied Biosystems ABI3700 (Applied Biosystems, Santa Clara, CA, USA).

In Silico Analysis of Large-Effect Nonsynonymous Variants

A large-effect nsSNP is defined as a homozygous radical nsSNP in a protein domain, which is computationally predicted by InterProScan on the Ensembl peptide (Quevillon et al. 2005). Radical nsSNPs are computationally predicted by SIFT (http://sift.jcvi.org), which is a sequence homology-based tool that distinguishes intolerant from tolerant amino acid changes and predicts whether an amino acid substitution in a protein will be likely to produce a phenotypic effect (Ng and Henikoff 2003; Kumar et al. 2009). A file containing genomic coordinates and base pair substitutions of all identified novel non-synonymous variants was used as an input for the SIFT genome tool.

Structure Variance Identification

To call insertions and deletions, we used Pindel v. 0.2.4 p to search for reads where one end is mapped on the genome and the other end can be mapped with high confidence in two (split) portions (Ye et al. 2009). The following parameters were used: the maximum size of structural variations to be detected (max_range_index = 5) = 32,368, the expected fraction of sequencing errors (sequencing_error_rate) = 0.03, the fraction of mismatches (maximum_allowed_mismatch_rate) = 0.05, the minimum alignment length of read (min_num_matched_bases) = 30, the minimum mapping quality of the reads Pindel uses as anchor (anchor_quality) = 20.

CNV Identification

Putative CNVs were identified using the CNV-seq, which examines the mapped reads from two individual chickens and reports regions that exhibit significant read depth differences (Xie and Tammi 2009). The following parameters were used as the default value ($\log_2$ threshold = 0.6, P value = 0.001 and minimum windows required = 4) to generate a list of CNVs from the best-hit files. A custom script was used to identify overlapping genes between Ensembl and CNVs.

CNV Validation

For the qPCR analysis, primers were designed using PRIMER3 software (http://frodo.wi.mit.edu/primer3) and were used to amplify 150–250-bp fragments that were positioned within each selected copy number variation region (CNVR) locus (Rozen and Skaletsky 2000). The primers for *EDN3* are 5′-GCTGCAGGCAGGTTGGAACT-3′ and 5′-GCAGAGGCTTCCCAGGATCA-3′. The primers for the propionylcoenzyme A carboxylase gene (*PCCA*, chr1) and thyroid hormone-inducible hepatic protein gene (*THRSP*, chr1) are used as references as described (Wang et al. 2010b). The qPCR reaction was done in 10 μl containing 6 ng of genomic DNA, 0.2 μM of each primer, and was conducted using 45 thermal cycles: one cycle of pre-incubation at 95 °C for 3 min, 40 cycles of gain (95 °C for 10 s, 60 °C for 20 s, and 72 °C for 1 s), a dissociation stage (95 °C for 5 s, 65 °C for 1 min, and 97 °C for continuous) for melting curve analysis, and a final stage for cooling (40 °C for 10 s). The SYBR green-based qPCR assays were performed on a Roche LightCycler® 480 Instrument II with a 96-well block (Roche Applied Science, Penzberg, Germany). All test samples for each qPCR were assayed in duplicates. All the data were analyzed by the HTC1 software (Roche Applied Science, Penzberg, Germany). The $2^{-\Delta\Delta Ct}$ method was used to calculate the copy numbers (Livak and Schmittgen 2001).

Selective-Sweep Analysis

Identified SNPs were used to detect signatures of selection in 10-kb sliding windows with a step size of 0.5 kb for the two genome sequences. At each detected SNP position, H_{ROH} values were calculated using the formula

$$H_{ROH} = \log_{10}\left(\frac{N_{het,w}+1}{N_{t,w}}\right)$$

$$\text{if } H_{ROH} \le -1, N_{t,w} \ge 10 \text{ and } N_{het,w}/N_{t,w} \le 10\%$$

$N_{t,w}$ = total SNPs in the window; $N_{het,w}$ = total heterozygous SNPs in the window; $N_{hom,w}$ = total homozygous identical SNPs in the two domestic line and different from the red jungle fowl in the window. We defined selective sweeps by the criterion of the H_{ROH} of the window ≤ -1. The formula and the criterion allow us to omit windows containing very few SNPs and require the identical homozygous SNPs to be in large proportions while effectively detecting a local reduction of heterozygosity.

Data Availability

The full data sets have been submitted to NCBI Sequence Read Archive (SRA) under accession nos. SRX286765, SRX286766, SRX286773, SRX286776, SRX286777, SRX286779-SRX286781, SRX286798, and SRX286799. Bioproject: PRJNA202483. The validated SNPs have been submitted to dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) and the accession nos. are provided in supplementary table S5, Supplementary Material online. The whole set of SNPs and indels will be provided upon request.

Results

Data Production and Short Read Alignment

Genomic libraries were gel size-selected and deep sequencing was carried out on Illumina GA IIx with PE 2 × 120 nt read length. The PE libraries were prepared by gel size selection and the averaged peak sizes of the inserts were 210, 321, 441, 539, and 539 bp for the five Silkie libraries and 202, 312, and 442 bp for the three L2 libraries (supplementary table S1, Supplementary Material online); the pass filtering rate ranged from 85 to 92%. We collected 250 and 319 billion reads for Silkie and L2, respectively (~30 and 38 gigabases [Gb] of sequence; supplementary table S1, Supplementary Material online).

Read mapping to the chicken reference sequence (WUGSC 2.1/galGal 3) was conducted using BWA (Li and Durbin 2010), and 85.70% and 74.83% of the input reads were mapped to unique positions on the chicken reference genome for Silkie and L2, respectively. The total read coverage of the chicken reference genome was 93.22% (including 5.1% Ns, supplementary fig. S1 and S2, supplementary table S2, Supplementary Material online). The reference genome sequence was covered at an average depth of 23.05-fold for Silkie and 24.80-fold for L2. The alignments between the uniquely mapped reads and the reference genome were used to categorize genetic variation, including SNPs, indels, CNVs, and SVs.

SNP and Indel Identification

Mapping of the sequencing reads of Silkie and L2 to the genome of the red jungle fowl revealed ~7.6 million SNPs for the two breeds (table 1), similar to previous findings (Wong et al. 2004; Wang et al. 2005). In total, 6,021,032 and 5,776,404 variants including 5,385,458 and 5,142,622 SNPs and 635,574 and 633,782 indels (1–73 bp) were found for Silkie and L2, respectively (fig. 1). Comparison of the putative SNPs with the chicken SNP database (dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/) revealed 3.28 and 3.33 million sites in "known SNPs" for Silkie and L2, respectively. The remaining 2.12 and 1.82 million SNPs for Silkie and L2 were at positions not previously identified as polymorphic sites. For the genes defined by mRNA transcripts, 11% of the testable SNPs are radical as predicted by SIFT (http://sift.jcvi.org) (Ng and Henikoff 2003; Kumar et al. 2009). As in previous studies, the SNPs are not uniformly distributed on the chromosomes (fig. 1). The genome sequences of other chicken lines (Rubin et al. 2010) had a 44.5-fold genome coverage using ABI SOLiD reads and revealed approximately 7 million known SNPs, in close agreement with our results from Illumina reads of the Silkie and L2 genomes at a lower coverage.

Among the identified SNPs of Silkie, 55% were homozygous and 45% were heterozygous, while the SNPs of L2 showed a slightly higher proportion of heterozygous sites (57%). Among the SNPs of Silkie, 2,949,783 (55%) sites were located in intergenic regions and 142,142 (2.64%) were within the 1-kb gene-flanking regions (fig. 2). The corresponding values for the SNPs of L2 were 2,804,238 (55%) and 141,508 (2.75%).

Among the indels of Silkie, 349,210 (55%) were in intergenic regions, 3,625 (0.57%) were within the 1-kb flanking regions of genes, and 264,954 (42%) were in genic regions (fig. 2). The corresponding numbers for the indels of L2 were 348,983 (55%), 3,812 (0.60%), and 262,235 (41%). The genic indels of Silkie included 1,292 coding, 263,662 intronic, and 3,625 UTR sites, while those of L2 included 1,347 coding, 348,983 intronic, and 3,812 UTR sites. There were 342,259 and 346,192 homozygous indels, and 293,762 and 287,946 heterozygous indels in Silkie and L2, respectively (fig. 3). Insertions accounted for 298,503 and 299,267 events, and deletions for 337,518 and 334,871 events in Silkie and L2, respectively. Indels accounted for only 11% of all events identified, but they involved 28% of all variant bases. The indel sizes ranged from 1 to 73 bp in length and homozygous indels show a wider length distribution than heterozygous ones (fig. 3). The size range over which indels were recognized was limited by the length of the reads. In the analysis of our 120 bp Illumina sequencing reads, the largest identified indel was 41 bp by SAMtools. In total, 968 and 1,005 indels for Silkie and L2, respectively, were found overlapping with coding sequences and can potentially affect protein functions.

To assess the reliability of our data, PCR amplification and Sanger sequencing were applied to 352 SNPs and 38 indels to determine whether they agreed with the deep sequencing results in the same individuals in which they were detected (supplementary table S4, Supplementary Material online). We found that 335 (95.2%) SNPs and 26 (68.4%) indels were consistent with the Illumina sequencing data (supplementary fig. S3 and table S5, Supplementary Material online).

CNV and SV Identification

Putative CNVs (≥2 kb) were detected by identifying genomic regions significantly different in coverage depth between the Silkie and L2 mapped read data sets using the software CNV-Seq (Xie and Tammi 2009). In total, 8,839 CNVs for Silkie relative to L2 were observed, involving ~24.6 Mb of the reference assembly used for mapping (table 2). The CNVs varied in length from 2,081 bp to 45,241 bp and the mean and median were 2,785 bp and 2,081 bp, respectively (fig. 4).

Among the 209 CNVRs larger than 5 kb, 53 (25.4%) variants together completely covered 66 annotated genes, which are enriched for transcription factor activity ($P < 0.001$) (supplementary table S6, Supplementary Material online). Using the Ensembl gene annotations, we detected CNV genes and then assigned a CNV estimate to each gene. The availability of the two genomes helped find the causative mutations associated with interesting traits in Silkie caused by CNV. For instance, a duplicated region on Chr20 containing endothelin 3 (*EDN3*), a gene with a known role in promoting melanoblast proliferation, increases the expression level of *EDN3* and causes dermal and internal organ hyperpigmentation in Silkie (Dorshorst et al. 2011; Shinomiya et al. 2012). This duplication is easily identified and confirmed in our study (fig. 5). This CNV is validated by using quantitative Real-Time PCR assays (supplementary fig. S4, Supplementary Material online).

Using Pindel v. 0.2.4p (Ye et al. 2009), we generated a catalogue of 23,454 structural variants (SVs) (≥50 bp), including 12,068 and 10,778 deletions and 278 and 330 insertions for Silkie and L2, respectively, and also combinations of SVs

Table 1
Statistics of Genetic Variants in the Silkie and L2 Genomes

	Silkie-RJF[a]			L2-RJF[a]			Silkie-L2	
	Total	Specific[b]	Novel[c]	Total	Specific[b]	Novel[c]	Shared[d]	Different[e]
All SNPs								
Intergenic[f]	2,828,057	1,322,844	1,317,281	2,695,771	1,190,558	1,166,362	1,505,213	2,513,402
Intergenic (Upstream w/5-kb)[g]	291,237	133,325	136,942	285,544	127,632	127,049	157,912	260,957
Intergenic (Downstream w/5-kb)[h]	263,964	121,474	123,511	256,170	113,680	112,260	142,490	235,154
Intergenic (Up/Down w/5-kb)[i]	75,295	33,842	36,228	75,621	34,168	35,152	41,453	68,010
Genic								
Intronic[j]	2,463,943	1,153,339	1,096,560	2,364,973	1,054,369	974,496	1,310,604	2,207,708
ncRNA[k]	590	304	320	532	245	280	280	549
5′ UTR[l]	3,827	1,718	1,947	4,110	2,001	2,051	2,109	3,719
3′ UTR[m]	24,239	11,262	10,838	23,649	10,672	9,917	12,977	21,934
5′/3′ UTR[n]	65	27	38	64	26	31	38	53
Exonic splice site[o]	1,395	577	594	1,416	598	557	818	1,175
Intronic splice site[p]	917	390	609	935	408	618	527	798
Exonic[q]	67,503	29,404	25,994	67,619	29,520	24,420	38,099	58,924
Synonymous[r]	48,112	20,828	16,660	48,078	20,794	15,474	27,284	41,622
Stop-gained[s]	232	114	116	228	110	103	118	224
Stop-loss[t]	23	8	14	21	6	15	15	14
Nonframeshift indel[u]	324	174	324	342	192	342	150	366
Frameshift indel[v]	948	385	948	986	423	986	563	808
Nonsynonymous[x]	19,259	8,472	8,527	19,380	8,593	8,057	10,787	17,065

[a]RJF, red jungle fowl.
[b]Variants only exist in one breed.
[c]Variants previously do not exist in the SNPdb.
[d]Total variants shared by two breeds.
[e]Total variants different between two breeds.
[f]Variants in intergenic regions.
[g]Variants overlap 5-kb regions upstream of transcription start site.
[h]Variants overlap 5-kb regions downtream of transcription end site.
[i]Variants located in both downstream and upstream regions (possibly for two different genes).
[j]Variants overlap introns.
[k]Variant overlaps a transcript without coding annotation in the gene definition. It does not mean that the RNA will never be translated and merely means that the gene annotation system did not give a coding sequence annotation.
[l]Variants overlap 5′ untranslated regions.
[m]Variants overlap 3′ untranslated regions.
[n]Variants located in both 5′ UTR and 3′ UTR regions (possibly for two different genes).
[o]Variants within exons but close to exon/intron boundaries.
[p]The 2-bp in introns that are close to exons.
[q]Only coding exonic portions, but not UTR portions.
[r]Single nucleotide changes that do not cause amino acid changes.
[s]Nonsynonymous SNVs, frameshift indels, nonframeshift indels, or block substitutions that lead to the immediate creation of stop codon at the variant site. For frameshift mutations, the stop codon downstream of the variant was not be considered as "stop-gained."
[t]Nonsynonymous SNVs, frameshift indels, nonframeshift indels, or block substitutions that lead to the immediate elimination of stop codon at the variant sites.
[u]Indels of three or multiples of three nucleotides that do not cause frameshift changes in protein coding sequence.
[v]Indels of one or more nucleotides that cause frameshift changes in protein coding sequence.
[x]Single nucleotide changes that cause amino acid changes.

using stringent SV detection constraints (fig. 4, table 3, and supplementary table S7, Supplementary Material online). The vast majority of SVs are located in non-coding regions. Ensembl annotated genes overlapped with SVs are not significantly enriched for any GO categories. We compared the SV affected genes in Silkie to those in L2 and found that only 13.3% of the predicted SVs were shared between them and all of the shared SVs are large deletions, suggesting that most of the gene-affecting SV events occurred after the separation of the two breeds.

Annotation of SNPs and Indels

We conducted a DAVID functional annotation clustering analysis (Huang et al. 2007; Huang et al. 2009) of genes containing variants to identify molecular functions (MF) and biological processes (BP) enriched for these classes of genetic variants. In our data set, nsSNPs were detected in 7,302 and 7,360 genes in Silkie and L2, respectively (supplementary fig. S6 and table S8, Supplementary Material online).

The SNPs in gene regions for Silkie and L2 were annotated using the Ensembl gene set (17,934 genes). We found, for the

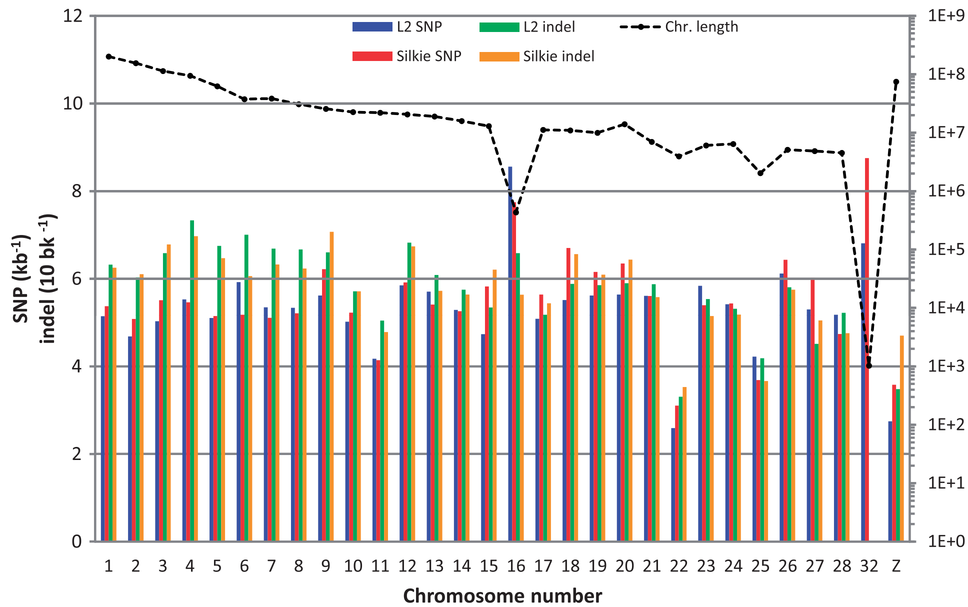

Fig. 1.—Densities of SNPs and indels on individual chromosomes detected between the Silkie and L2 genomes.

two genomes, 2,463,943 and 2,364,973 SNPs in introns, 28,131 and 27,823 SNPs in untranslated regions (UTRs), 2,312 and 2,351 SNPs at splice sites, and 68,898 and 69,035 SNPs in coding regions, leading to 19,259 and 19,380 nonsynonymous nucleotide changes in the two genomes (table 1, fig. 3, supplementary table S8, Supplementary Material online).

We applied SIFT (Ng and Henikoff 2003; Kumar et al. 2009) to predict radical nsSNPs (Supplementary table S9, Supplementary Material online). We defined a large-effect nsSNP as a radical nsSNP located in a protein domain predicted by InterProScan (supplementary table S10, Supplementary Material online) (Quevillon et al. 2005). We found that the large-effect nsSNPs in L2 are significantly enriched in the nitrogen compound biosynthetic process, cofactor binding, and vitamin binding, whereas the nsSNPs in Silkie are not enriched in any group. These results suggest that the phenotypes associated with genes containing these mutations may represent specific characteristics of these breeds.

Mutation and Selection

We investigated two types of loss-of-function mutations: stop-gained (nonsense) mutations and frameshift mutations (table 4). The frameshift mutations of L2 are enriched in the histone-lysine *N*-methyltransferase and amino acid transmembrane transporter activities, whereas those of Silkie are enriched in the nucleotide binding and the modification-dependent protein catabolic process. The stop-gained mutations of L2 are enriched in endopeptidase inhibitor activity, whereas those of Silkie are enriched in proteins involved in muscle contraction, response to radiation, adult behavior, cell death, and regulation of programmed cell death.

To understand the genetic and genomic changes associated with chicken domestication, we searched for local reductions in heterozygosity that might have accompanied selective sweeps in the common ancestor of Silkie and L2. We examined 10-kb sliding windows with at least 10 homozygous SNP sites in every 0.5-kb step. We then computed how often at least 90% of the homozygous SNP sites are identical in the two domestic lines but different in the red jungle fowl. The segments identified contained 509 genes, representing 2.84% of total Ensemble annotated genes (supplementary table S11, Supplementary Material online).

We went further to investigate the functional distribution of loss-of-function mutations (fig. 6), which are predicted to disable gene functions. Loss-of-function mutations have been proposed to be an important consequence of domestication (Olson 1999). However, in agreement with the previous study (Rubin et al. 2010), we found little evidence that selection for loss-of-function mutations played an important role in chicken domestication. Stop-gained mutations accounted for 1.31% and 1.37% of all Ensembl annotated genes for Silkie and L2, respectively, and frameshift mutations accounted for 4.53% and 4.61% of all Ensembl annotated genes, while stop-gained and frameshift mutations accounted for 1.57% and 3.14% of gene located in the putative selective sweeps, respectively (Pearson's χ^2 test, $P > 0.1$). These low proportions suggest that loss-of-function mutations were not enriched in the selective sweeps. We considered the overlap of our CNVRs with putative selective sweeps and found only six Ensembl annotated genes within CNVRs (ENS-12789, ENS-07597,

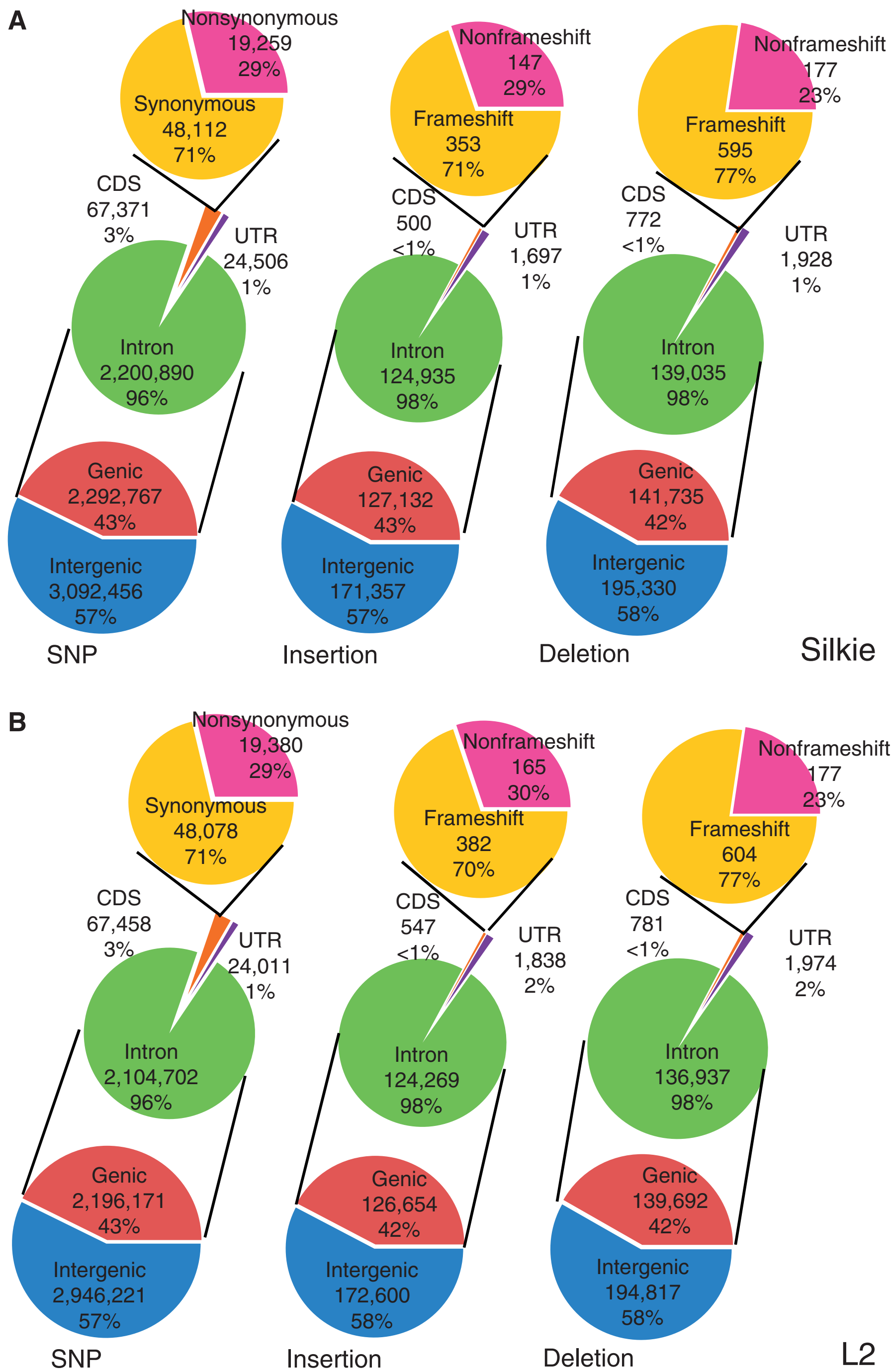

FIG. 2.—Annotation of SNPs and indels and distribution of SNPs. Predicted functional consequences of SNPs and indels of the Silkie (*A*) and L2 (*B*).

ENS-08000, ENS-09357, ENS-10772, and ENS-11056; only the last five digits of the Ensembl chicken gene annotation were shown).

Among the 509 genes in the putative selective sweeps in our study, 46 were also found in the study of Rubin et al. (2010). *TSHR*, *IGF1*, *PMCH*, *TBC1D1*, *ARID4B*, *ROBO2*, *ANK2*, *SLC16A12*, and *OSGIN1* showed the highly selective sweeps in all domestic or commercial breeds (Rubin et al. 2010). In the putative selective sweeps, *IGF1* and *HMGA2* have already been shown to be associated with body weight gains in Silkie (Tang et al. 2010; Song et al. 2011).

It is also interesting to investigate whether protein-coding changes played an important role in evolution under domestication. We found that the protein-coding changes (nsSNP, frameshift, stop-gained) occurred in 50.29% of the genes in the putative selective sweeps compared with the 42.11% and

42.49% in the set of all genes in the Silkie and L2 genomes (Pearson's χ^2 test, $P < 0.001$). The genes in the putative selective sweep regions are functionally enriched for the GTPase regulator activity, the glycoprotein biosynthetic process, nucleoside binding, the regulation of generation of precursor metabolites and energy, the regulation of cell adhesion, the purinergic nucleotide receptor activity, microtubule binding, stem cell development, and cell motion (table 5).

nsSNPs and QTLs

Of the 7,291 and 7,349 nsSNP-containing genes of Silkie and L2, respectively, 4,665 (63.98%) and 4,736 (64.43%) matched to the genes that were found to locate at positions of significant QTLs (data from Chicken QTLdb, http://www.animalgenome.org/cgi-bin/QTLdb/GG/index). However, nsSNP-containing genes are not particularly concentrated in the QTL regions because the proportion of the genome covered by all the QTL regions is 64.11%, which is close to the proportion of nsSNP-containing genes (~64%).

We used DAVID to cluster radical nsSNPs in particular QTL regions and found that some significant enrichments of functionally related genes are associated with the traits (Huang et al. 2007; Huang et al. 2009). For instance, radical nsSNP-containing immunity genes are enriched in the QTL regions associated with several traits such as body weight, early mortality, oxygen saturation, cloacal bacterial burden after a challenge with *Salmonella* E, and *Salmonella* presence in ovary for both Silkie and L2. This may suggest that radical nsSNPs in immunity genes were subjected to human selection because they are associated with these economic traits. However, it is premature to draw a definitive conclusion.

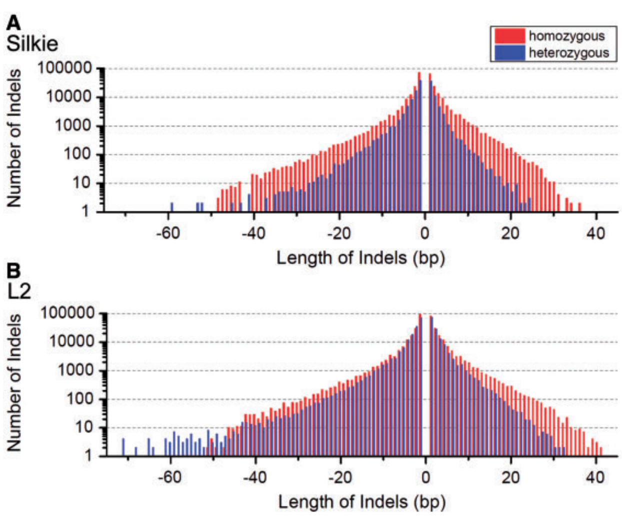

Fig. 3.—Distribution of indel lengths in (*A*) the Silkie and (*B*) the L2.

Associations between genotypes and phenotypes have been recently reported in many studies in chickens. Some SNP sites identified in L2 or/and Silkie have been reported to be associated with phenotypes in other chicken breeds (table 6). For mutations found in a single breed, *MX1* exon 13 polymorphisms in broiler chickens are associated with commercial traits such as body weight, and early and late mortality (Livant et al. 2007), and with morbidity, early mortality, viral shedding, and cytokine responses in chickens infected with an avian influenza virus (Ewald et al. 2011). Pro-opiomelanocortin (*POMC*) was associated with growth and carcass traits in Anak and Gushi chickens (Bai et al. 2012), greater body weight in females of commercial broilers (Sharma et al. 2008), and a mutation in aggrecan (*ACAN*) was associated with the incidence and severity of tibial dyschondroplasia (TD) in chickens (Ray et al. 2006). Inducible nitric oxide synthase (*iNOS*) is associated with general mortality and other performance traits in three elite commercial broiler chicken lines raised in high and low hygiene environments (Ye et al. 2006). Gonadotropin-Releasing Hormone Receptor (*GnRHR*) is associated with egg-laying traits in the Wenchang Chicken (Wu et al. 2007).

For mutations found in both breeds, mannan-binding lectin (*MBL*) plays an important role as the first line of defense against *Pasteurella multocida* by diminishing the infection before the adaptive immune response takes over (Schou et al. 2010). *Ovocalyxin-32* is associated with eggshell traits such as egg weight, short length of the egg, long length of the egg, and non-destructive deformation (Takahashi et al. 2010). C–C motif chemokine 4 (*CCL4*) was found to be associated to plumage condition in laying hens (Biscarini et al. 2010) and also in the selective sweeps we found. PR domain containing 16 gene (*PRDM16*) was found to have positive effects on chicken growth, fatness, and meat quality traits at different stages (Han et al. 2012). Lipoprotein lipase gene (*LPL*) is significantly

Table 2

Over-represented GO Categories Associated with Chicken CNVs

Term	Description	*P* Value	Fold Enrichment	FDR
GO:0003700	Transcription factor activity	1.48E-06	7.33	0.0015
GO:0043565	Sequence-specific DNA binding	2.42E-06	8.57	0.0024
GO:0006355	Regulation of transcription, DNA-dependent	5.32E-06	5.77	0.0065
GO:0051252	Regulation of RNA metabolic process	6.08E-06	5.68	0.0074
GO:0030528	Transcription regulator activity	2.75E-05	5.15	0.0272

Note.—FDR, false discovery rate.

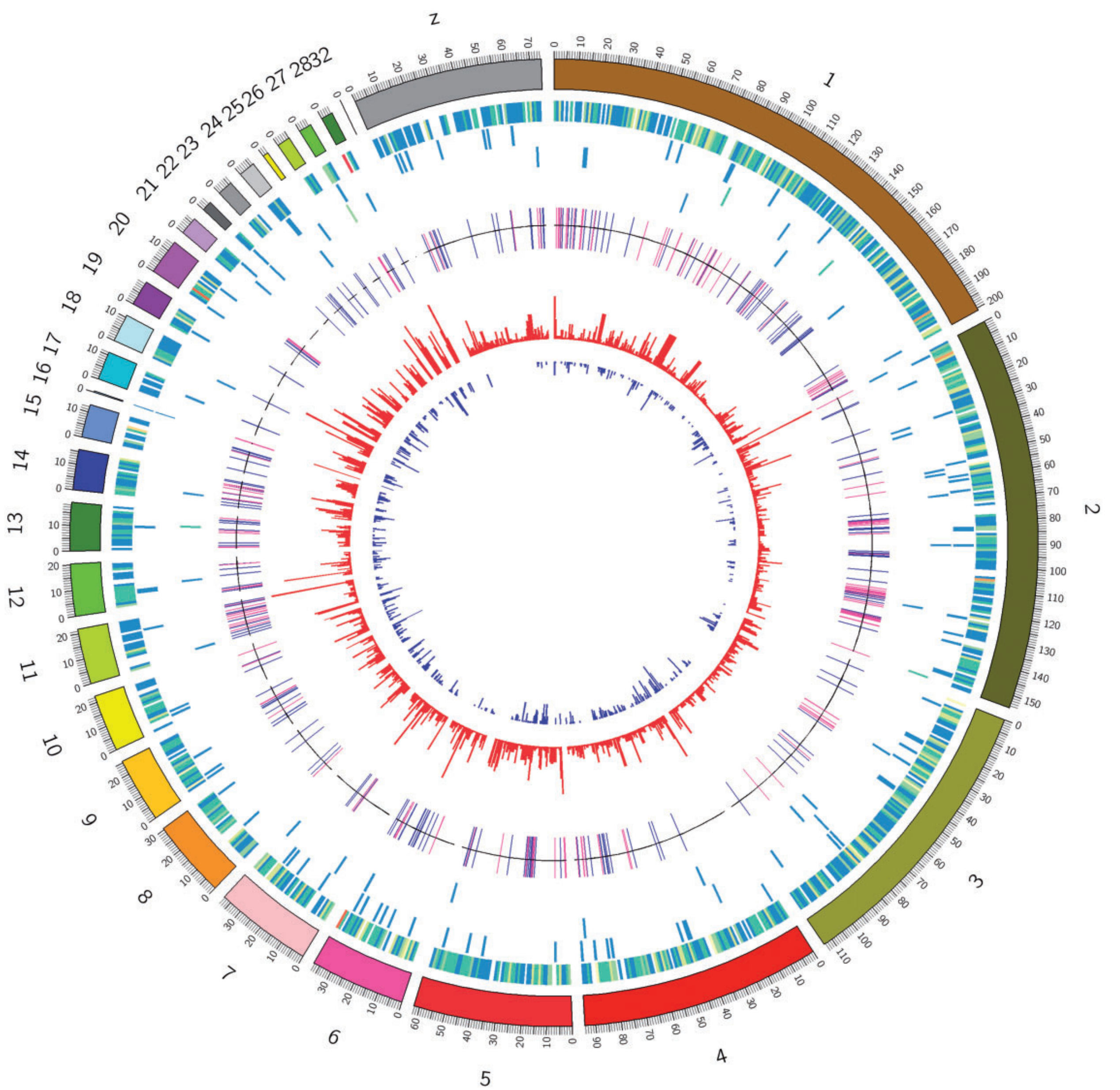

Fig. 4.—Summary of resequencing data of the Silkie and L2 genomes. Distribution of SNPs, indels, and structural variants (≥50 bp) in the Silkie and L2 genomes, relative to the reference red jungle fowl genome. The circular map shows the genomic distributions of different classes of same variations for the Silkie and L2 genomes based on the resolution of 1 Mb bins. Chromosomes are shown in different colors in the outermost circle. The heatmaps from the outermost to the innermost circle show the abundance of four structural variants (length ≥ 50 bps): deletions, insertions, tandem duplications, and inversions. The purple and blue colors in the CNV (≥5 kb) ring represent gain and loss of copy number variation for the Silkie genome relative to the L2 genome. For SNPs, the red color stands for the homozygous SNPs, whereas blue for the heterozygous SNPs in the exonic regions.

associated with intermuscular fat width, abdominal fat weight, and thickness of subcutaneous fat in chickens (Liu et al. 2006).

Discussion

Applying Illumina sequencing, we obtained the first draft genome sequences of Silkie and L2 and identified a total of 7.6 million SNPs in comparison with the genome of their wild ancestor, the red jungle fowl.

DNA Capture Array, exome sequencing, and other target-enrichment strategies for next-generation sequencing have allowed the sequencing of targeted regions in the genome more efficiently and economically (Mamanova et al. 2010; Teer and Mullikin 2010; Mertes et al. 2011), especially with

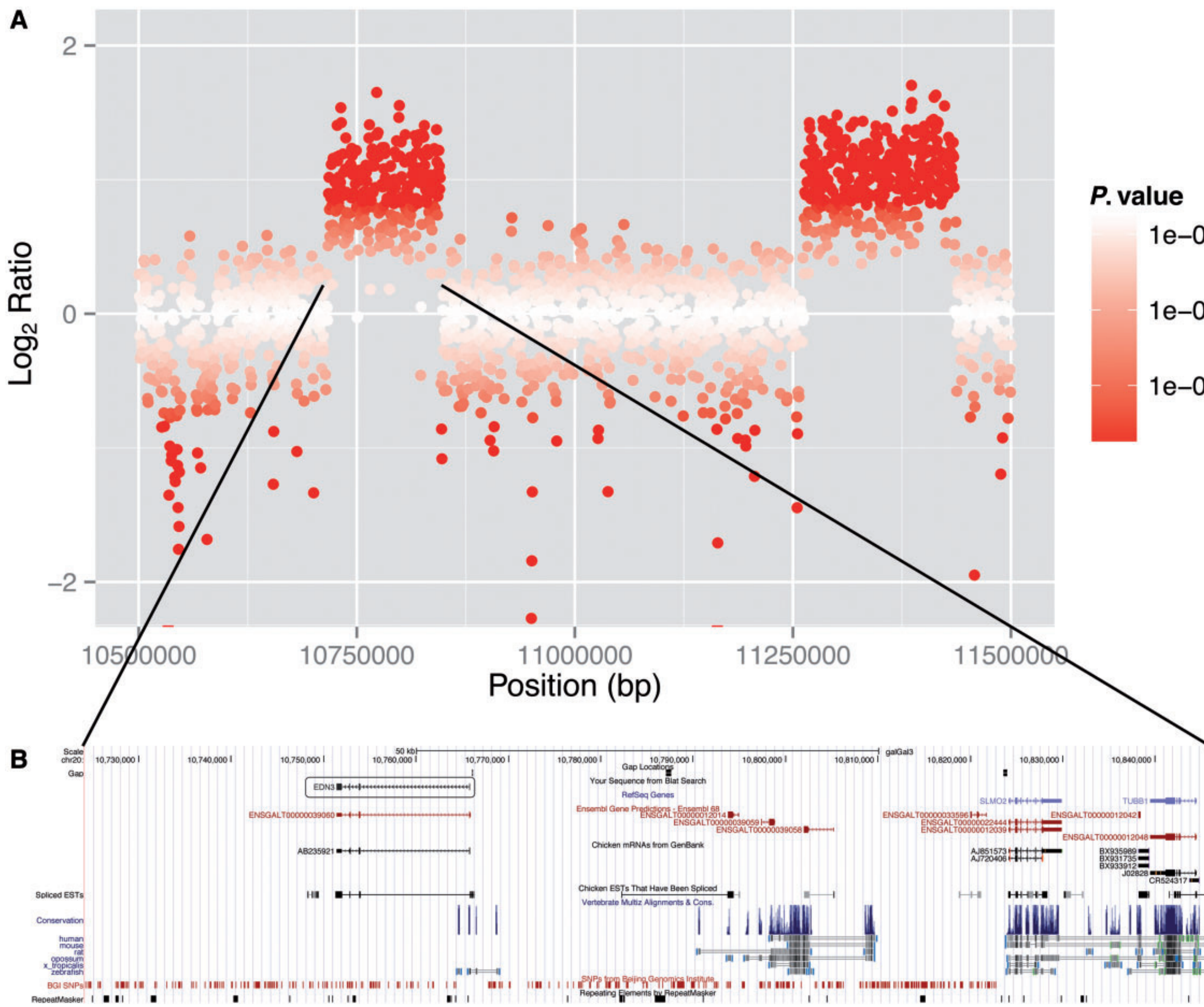

Fig. 5.—CNVs overlapping with the *EDN3* gene region. (*A*) $\log_2$ ratio plot of the *EDN3* (Q3MU75_CHICK; ENSGALT00000039060) gene region. Each point shows the $\log_2$ of the number of Silkie reads mapped to the number of L2 reads mapped. Points are colored based on the $\log_{10}$ *P* value calculated by the CNV-seq software. (*B*) The CNVR containing the *EDN3* gene in chromosome 20 as visualized using the UCSC Genome Browser.

rapidly increasing numbers of samples. These techniques have been used successfully to identify human disease-causing variants in some cases (Bamshad et al. 2011; Gilissen et al. 2011; Ku et al. 2011; Gilissen et al. 2012; Gonzaga-Jauregui et al. 2012; Rabbani et al. 2012) and have been useful for crop improvement in agriculture (Singh et al. 2012). However, exome capture by hybridization can introduce considerable coverage variation (Majewski et al. 2011), affecting comparative analysis. Moreover, targeted sequencing of only specific regions or a specified list of protein-coding sequences could miss DNA variations, especially valuable genetic variants in intronic and intergenic regions which potentially affect gene expression. Furthermore, whole-genome sequencing can also detect CNVs as well as SVs, giving us a more comprehensive view of genetic variation in a genome.

We chose to use several PE libraries for whole-genome resequencing. PE reads can provide rigorous read alignment and enhance the accuracy and coverage of SNP calling and consensus sequence inference. Applying PE reads to study structural variation, we provided the first genome-wide pattern of structural variation and copy number variants in the chicken using stringent SV detection constraints. We found that 96% of the predicted SVs resided in intronic (41%) or intergenic regions (55%), consistent with a previous finding in chickens using PE sequencing of reduced representation libraries (Kerstens et al. 2011). It is possible that some of the SVs identified have contributed to phenotypic differences among domestic chickens and the red jungle fowl. For example, a large insertion affecting the expression of *BMP12* has been shown to be the causative mutation associated with the naked neck trait (Mou et al. 2011).

Table 3
Statistics of Structural Variants in Silkie and L2

SV Type	Breed	
	Silkie	L2
Large deletion (≥50 bps)		
Downstream	149	138
Exonic	54	60
Intergenic	6501	5446
Intronic	4791	4271
ncRNA_exonic	3	5
Splicing	13	14
Upstream	114	145
Upstream; downstream	6	10
UTR3	36	26
UTR5	2	2
Large insertion (≥50 bps)		
Downstream	3	2
Exonic	1	0
Intergenic	145	168
Intronic	104	113
Upstream	4	3
Tandem duplication (≥ 50bps)		
Downstream	4	5
Exonic	2	4
Exonic; splicing	0	1
Intergenic	97	123
Intronic	29	48
Splicing	0	1
Upstream	2	6
Inversion (≥50 bps)		
Downstream	0	2
Exonic	10	12
Intergenic	56	48
Intronic	30	31
Upstream	1	0
Splicing	0	1

Next-generation sequencing technologies and analysis programs can provide an efficient pipeline to characterize CNVs at the genome-wide level. CNVs have been analyzed in several domesticated animals (Clop et al. 2012). In particular, comparative genomic hybridization (CGH) and SNP arrays have been applied to screen for CNVs in chickens (Wang et al. 2010b; Clop et al. 2012; Wang et al. 2012b; Jia et al. 2013) and a few traits have been shown to be associated with CNVs. For instance, the chicken pea comb phenotype was linked to a duplication near the first intron of *SOX5* (Wright et al. 2009). The hyperpigmentation in the Silkie is caused by duplicated regions containing endothelin 3 (*EDN3*) (Dorshorst et al. 2011; Shinomiya et al. 2012). Late feathering is caused by a partial duplication of the *PRLR* and *SPEF2* genes in chickens (Elferink et al. 2008). However, accurate CNV detection using NGS data is still difficult due to the propensity of falsely detecting CNVs by the software CNV-Seq since it does not take local read count variability into account (Klambauer et al. 2012).

The chicken has been used for quantitative genetic studies for decades (Georges and Andersson 2003; Siegel et al. 2006; Georges 2007). Vast numbers of QTLs are being detected in experiments for a large variety of economic traits in chickens, such as growth, carcass composition, reproductive, behaviors, disease resistance, etc. (Andersson 2001; Burt and Pourquie 2003). The chicken may be the only farm animal that can be applied for refining the map position of QTL with relatively low costs (Andersson and Georges 2004). Several candidate genes were successfully identified to be responsible for chicken disease resistance by combining high-resolution QTL mapping, comparative mapping, and functional genomic data (Vallejo et al. 1998; Zhu et al. 2001; Lipkin et al. 2002). The growing genomic resources in addition to a relatively short reproduction time make the chicken even more ideal for unraveling the molecular basis of phenotypic diversity in birds (Burt and Pourquie 2003).

Table 4
Top Annotation Clusters of Loss-of-Function Mutations Identified by the DAVID Functional Annotation Clustering Tool

Breeds	Mutation Types	Representative Annotation Terms	Enrichment Score
Silkie	Stop-gained	Muscle contraction	1.87
		Response to radiation	1.30
		Adult behavior	1.27
		Cell death	1.15
		Regulation of programmed cell death	1.05
	Frameshift	Nucleotide binding	1.33
		Modification-dependent protein catabolic process	1.05
L2	Stop-gained	Endopeptidase inhibitor activity	1.04
	Frameshift	Histone-lysine *N*-methyltransferase activity	1.28
		Amino acid transmembrane transporter activity	1.26
	Large-effect nsSNP	Nitrogen compound biosynthetic process	2.65
		Cofactor binding	1.47
		Vitamin binding	1.01

Note.—The genes were analyzed by the Functional Annotation Clustering Tool. The top annotation clusters have group enrichment scores greater than 1 were listed. The representative biology terms associated with the top annotation clusters are manually summarized.

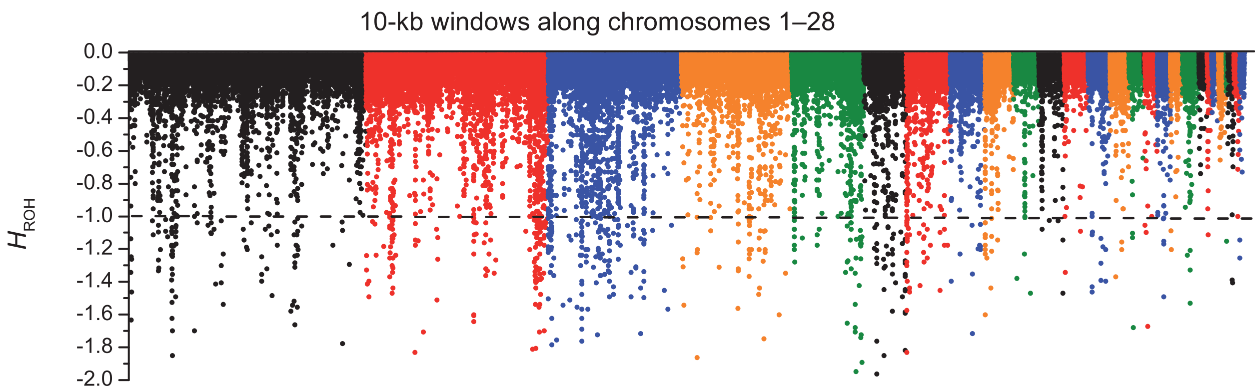

Fig. 6.—Distributions of H_{ROH} values. The horizontal dashed lines indicate the threshold at $H_{ROH} = -1$.

Table 5

Top Annotation Clusters of Gene Identified in the Putative Selective Sweep Regions by the DAVID Functional Annotation Clustering Tool

Representative Annotation Terms	Enrichment Score
GTPase regulator activity	1.75
Glycoprotein biosynthetic process	1.70
Nucleoside binding	1.43
Regulation of generation of precursor metabolites and energy	1.40
Regulation of cell adhesion	1.34
Purinergic nucleotide receptor activity, G-protein coupled	1.25
Microtubule binding	1.23
Stem cell development	1.16
Cell motion	1.05

Note.—The genes were analyzed by the Functional Annotation Clustering Tool. The top annotation clusters that have group enrichment scores >1 are listed. The representative biology terms associated with the top annotation clusters are manually summarized.

With the availability of genome-wide variations of Silkie and L2, we tried to identify candidate genes for important traits in chicken domestication using multiple approaches including SIFT analysis, functional enrichment, and selective sweep detection. We found that the genes involved in nitrogen compound metabolism, the ATP-binding cassette (ABC) transporters, extracellular matrix, and cytoskeletons are enriched in the large-effect containing genes or in the selective sweeps. The ABC transporter superfamily is the largest transporter gene family responsible for transporting specific molecules across cell membranes and essential for regulating organismic homeostasis in all animals (Dean and Annilo 2005); they also might be important for silkworm domestication (Xie et al. 2012). Moreover, genes involved in extracellular matrix have been shown to be differentially expressed between domestic chickens and red jungle fowls, suggesting possible selection pressures on this kind of genes (Li et al. 2012). Similarly, the genes involved in these processes or functions are found to evolve faster in other domesticated animals, such as dogs, cattle, and pigs (Groenen et al. 2012).

Although we only sequenced one individual for each of the two breeds, we can still identify local reductions in heterozygosity due to selective sweeps. We found that coding sequence mutations were slightly enriched in the putative selective sweeps. This observation suggests that physiological traits are likely to be artificially selected since adaptive mutations affecting physiology are more likely to occur in the protein-coding regions than in the cis-regulatory regions of genes (Carroll 2005).

We also found important genes such as *TSHR*, *IGF1*, *PMCH*, and *TBC1D1* that are associated with growth, appetite, and metabolic regulation in broiler in a previous study (Rubin et al. 2010). The thyroid stimulating hormone receptor (*TSHR*) mutation could be involved in regulating photoperiod control of reproduction (Yoshimura et al. 2003; Hanon et al. 2008; Nakao et al. 2008), affecting development, growth, and behavior in the domestic chicken with selective advantages during domestication. Therefore, *TSHR* is suggested to be a domestication gene in chickens, where all individuals of domestic breeds carry the same mutant allele. In fact, Silkie and L2 also carry the G558A mutation, which may drive the residue outwards from the cell membrane and thus influence ligand binding (Rubin et al. 2010). Our results indicate that these genes are also important for lowly selected breeds of chickens.

Within the selective sweeps in all of the domestic chickens used in our and Rubin et al.'s studies, some of the genes have also been detected to be associated with domestication traits in chicken or other farm animals, reinforcing their important roles in chicken domestication. For instance, *ERO1LB* and *ARID4B* have been detected to be associated with residual feed intake in swines (Gorbach 2011). *ARID4B* encodes a subunit of the histone deacetylase-dependant SIN3A transcriptional corepressor complex, which functions in various cellular processes including proliferation, differentiation,

Table 6
Identified nsSNPs Reportedly Associated with Phenotypes in the Other Breeds of Chicken

Gene	Ensembl ID[a]	Chr	Breed[b]	Mutation	Association	Reference
MX	ENS-25999	1	L2	Radical nsSNP	Morbidity, early mortality, viral shedding, and cytokine responses	Livant et al. 2007; Ewald et al. 2011
POMC	ENS-26793	3	L2	Radical nsSNP	Production traits	Sharma et al. 2008; Bai et al. 2012
Ovocalyxin-32	ENS-15634	9	B	Radical nsSNP	Eggshell quality traits	Takahashi et al. 2010
GnRHR	ENS-21020	10	S	Stop gained	Egg-laying traits	Wu et al. 2007
ACAN	ENS-38569	10	L2	Radical nsSNP	Tibial dyschondroplasia	Ray et al. 2006
iNOS	ENS-09129	19	L2	Radical nsSNP	Performance	Ye et al. 2006
CCL4	ENS-35206	19	B	Radical nsSNP	Feather damage	Biscarini et al. 2010
PRDM16	ENS-01590	21	B	Radical nsSNP	Performance traits	Han et al. 2012
MBL2	ENS-04765	21	B	Radical nsSNP	Experimental Pasteurella multocida infection	Schou et al. 2010
LIPL	ENS-24882	Z	B	Radical nsSNP	Fat deposition	Liu et al. 2006

[a]Only the last five digits of the Ensembl chicken gene annotation are shown.
[b]S: Silkie; L2: Taiwanese L2 chicken; B: Both breeds of S and L2.

apoptosis, oncogenesis, and cell fate determination (Wu et al. 2006; Winter et al. 2012). In addition, *NELL1* has been identified in a selective sweep in broilers (Elferink et al. 2012), and *ESRP2* is associated with chicken abdominal fat contents (Zhang et al. 2012). *NELL1* encodes a cytoplasmic protein, which contains epidermal growth factor (EGF)-like repeats and may be involved in cell growth regulation and differentiation in bone and cartilage (Zhang et al. 2010; Chen et al. 2011; Zhang et al. 2011; Zou et al. 2011; Cowan et al. 2012; Siu et al. 2012), and *ESPR2* encodes an epithelial cell-type-specific splicing regulator (Warzecha et al. 2009a, 2009b; Warzecha et al. 2010; Dittmar et al. 2012). *ROBO2*, which was identified in highly selective sweeps in commercial broilers, also appeared in our study. *ROBO2* encodes a protein that is a receptor for SLIT2, which is known to function in axon guidance and cell migration (Anitha et al. 2008). These findings imply that the selection for traits controlled by these genes occurred early and throughout the history of chicken domestication, maintaining a low heterozygosity in these genes. Thus, our approach is successful in identifying some of the important genes in domestication.

ANK2, *SLC16A12*, *ARID4B*, and *OSGIN1*, which were found in the highly selective sweep regions in all domestic lines of chickens in the study of Rubin et al., have also been identified in the selective sweep regions in our study. They are excellent candidate genes for functional studies in animal sciences. Differential expression of *ANK2*, which encodes a member of the ankyrin proteins required for targeting and stability of Na^{+}/Ca^{++} exchanger 1 in cardiomyocytes during cardiac muscle contraction (Mohler 2006; Hashemi et al. 2009), between two layer lines, Lohmann Selected Leghorn (LSL) and Lohmann Brown (LB), has also been found in one study using genome-wide microarray analyses (Habig et al. 2012).

In addition to the above issues, it is important to learn the genetic basis underlying phenotypic differences that are responsible for specific breed characteristics. We observed enrichments in different functional categories in the two chicken breeds we studied, implying differentially selective forces. The observed enrichment in nitrogen compound biosynthetic process, cofactor binding, and vitamin binding genes may be related to selection of L2 for a meat and egg purpose.

In conclusion, we present a whole genome map of SNPs, indels, SVs, and CNVs of two chicken breeds here. Genome-wide comparisons with trait data of chicken breeds using the SNPs, indels, SVs, and CNVs identified here will provide additional clues to the genetic and genomic bases of the interesting traits of domestic chickens and will be a useful resource for future studies of the molecular basis of disease and phenotypic variation in chickens.

Supplementary Material

Supplementary tables S1–S11 and figures S1–S5 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Cheng-Ming Chuong, Ping Wu, Hsu-Chen Cheng, Pin-Chi Tang, and Chua-Ti Ting for helpful comments and Tzi-Yuan Wang for technical help. This work was supported by the National Science Council, Taiwan (99-2321-B-001-041-MY2) and the postdoctoral fellowship from Academia Sinica of Taiwan to C.S.N.

Literature Cited

Andersson L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nat Rev Genet. 2:130–138.
Andersson L, Georges M. 2004. Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet. 5:202–212.
Anitha A, et al. 2008. Genetic analyses of roundabout (ROBO) axon guidance receptors in autism. Am J Med Genet B Neuropsychiatr Genet. 147B:1019–1027.

Bai Y, et al. 2012. Polymorphisms of the pro-opiomelanocortin and agouti-related protein genes and their association with chicken production traits. Mol Biol Rep. 39:7533–7539.

Bamshad MJ, et al. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 12:745–755.

Biscarini F, et al. 2010. Across-line SNP association study for direct and associative effects on feather damage in laying hens. Behav Genet. 40: 715–727.

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

Carroll SB. 2005. Evolution at two levels: on genes and form. PLoS Biol. 3: e245.

Chao CH, Lee YP. 2001. Relationship between reproductive performance and immunity in Taiwan country chickens. Poult Sci. 80:535–540.

Chen F, et al. 2011. NELL-1, an osteoinductive factor, is a direct transcriptional target of Osterix. PLoS One 6:e24638.

Clop A, Vidal O, Amills M. 2012. Copy number variation in the genomes of domestic animals. Anim Genet. 43:503–517.

Cowan CM, et al. 2012. NELL-1 increases pre-osteoblast mineralization using both phosphate transporter Pit1 and Pit2. Biochem Biophys Res Commun. 422:351–357.

Crawford RD. 1990. Poutry genetic resources: evolution, diversity, and conservation. In: Crawford RD, editor. Poutry breeding and genetics. New York (NY): Elsevier. p. 43–60.

Darwin C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London, UK: John Murray.

Darwin C. 1868. The variation of plants and animals under domestication. London, UK: John Murray.

Dean M, Annilo T. 2005. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet. 6:123–142.

Dittmar KA, et al. 2012. Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. Mol Cell Biol. 32:1468–1482.

Dohner JV. 2001. The encyclopedia of endangered livestock and poultry breeds. New Haven (CT): Yale University Press.

Dorshorst B, Okimoto R, Ashwell C. 2010. Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the Silkie chicken. J Hered. 101:339–350.

Dorshorst B, et al. 2011. A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet. 7:e1002412.

Elferink MG, Vallee AA, Jungerius AP, Crooijmans RP, Groenen MA. 2008. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. BMC Genomics 9:391.

Elferink MG, et al. 2012. Signatures of selection in the genomes of commercial and non-commercial chicken breeds. PLoS One 7:e32720.

Eriksson J, et al. 2008. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. PLoS Genet. 4:e1000010.

Ewald SJ, et al. 2011. Association of Mx1 Asn631 variant alleles with reductions in morbidity, early mortality, viral shedding, and cytokine responses in chickens infected with a highly pathogenic avian influenza virus. Immunogenetics 63:363–375.

Georges M. 2007. Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. Annu Rev Genomics Hum Genet. 8:131–162.

Georges M, Andersson L. 2003. Positional identification of structural and regulatory quantitative trait nucleotides in domestic animal species. Cold Spring Harb Symp Quant Biol. 68:179–187.

Gilissen C, Hoischen A, Brunner HG, Veltman JA. 2011. Unlocking Mendelian disease using exome sequencing. Genome Biol. 12:228.

Gilissen C, Hoischen A, Brunner HG, Veltman JA. 2012. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 20:490–497.

Gonzaga-Jauregui C, Lupski JR, Gibbs RA. 2012. Human genome sequencing in health and disease. Annu Rev Med. 63:35–61.

Gorbach DM. 2011. The prediction of single nucleotide polymorphisms and their utilization in mapping traits and determining population structure in production animals [PhD. Dissertation in Animal Science]. Paper 10336. [Ames (IA)]: Iowa State University.

Groenen MA, et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491:393–398.

Habig C, Geffers R, Distl O. 2012. Differential gene expression from genome-wide microarray analyses distinguishes lohmann selected leghorn and lohmann brown layers. PLoS One 7:e46787.

Han R, et al. 2012. Novel SNPs in the PRDM16 gene and their associations with performance traits in chickens. Mol Biol Rep. 39: 3153–3160.

Hanon EA, et al. 2008. Ancestral TSH mechanism signals summer in a photoperiodic mammal. Curr Biol. 18:1147–1152.

Hashemi SM, Hund TJ, Mohler PJ. 2009. Cardiac ankyrins in health and disease. J Mol Cell Cardiol. 47:203–209.

Huang DW, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37:1–13.

Huang DW, et al. 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 8:R183.

Imsland F, et al. 2012. The Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. PLoS Genet. 8:e1002775.

Jia X, et al. 2013. Copy number variations identified in the chicken using a 60K SNP BeadChip. Anim Genet. 44:276–284.

Johnsson M, et al. 2012. A sexual ornament in chickens is affected by pleiotropic alleles at HAO1 and BMP2, selected during domestication. PLoS Genet. 8:e1002914.

Kerstens HH, et al. 2011. Structural variation in the chicken genome identified by paired-end next-generation DNA sequencing of reduced representation libraries. BMC Genomics 12:94.

Klambauer G, et al. 2012. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res. 40:e69.

Ku CS, Naidoo N, Pawitan Y. 2011. Revisiting Mendelian disorders through exome sequencing. Hum Genet. 129:351–370.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 4:1073–1081.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595.

Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Li Q, et al. 2012. Gastrocnemius transcriptome analysis reveals domestication induced gene expression changes between wild and domestic chickens. Genomics 100:314–319.

Lipkin E, Fulton J, Cheng H, Yonash N, Soller M. 2002. Quantitative trait locus mapping in chickens by selective DNA pooling with dinucleotide microsatellite markers by using purified DNA and fresh or frozen red blood cells as applied to marker-assisted selection. Poult Sci. 81: 283–292.

Liu R, Wang YC, Sun DX, Yu Y, Zhang Y. 2006. Association between polymorphisms of lipoprotein lipase gene and chicken fat deposition. Asian-Austral J Anim Sci. 19:1409–1414.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402–408.

Livant EJ, et al. 2007. MX1 exon 13 polymorphisms in broiler breeder chickens and associations with commercial traits. Anim Genet. 38: 177–179.
Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. 2011. What can exome sequencing do for you? J Med Genet. 48: 580–589.
Mamanova L, et al. 2010. Target-enrichment strategies for next-generation sequencing. Nat Methods. 7:111–118.
Mertes F, et al. 2011. Targeted enrichment of genomic DNA regions for next-generation sequencing. Brief Funct Genomics. 10:374–386.
Miao YW, et al. 2013. Chicken domestication: an updated perspective based on mitochondrial genomes. Heredity (Edinb). 110:277–282.
Mohler PJ. 2006. Ankyrins and human disease: what the electrophysiologist should know. J Cardiovasc Electrophysiol. 17:1153–1159.
Mou C, et al. 2011. Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. PLoS Biol. 9:e1001028.
Nakao N, et al. 2008. Thyrotrophin in the pars tuberalis triggers photoperiodic response. Nature 452:317–322.
Ng CS, et al. 2012. The chicken frizzle feather is due to an alpha-keratin (KRT75) mutation that causes a defective rachis. PLoS Genet. 8: e1002748.
Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31:3812–3814.
Nishibori M, Shimogiri T, Hayashi T, Yasue H. 2005. Molecular evidence for hybridization of species in the genus Gallus except for Gallus varius. Anim Genet. 36:367–375.
Olson MV. 1999. When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet. 64:18–23.
Price EO. 2002. Animal domestication and behavior. Wallingford, UK: CABI Publ.
Quevillon E, et al. 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 33:W116–W120.
Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I. 2012. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. J Hum Genet. 57:621–632.
Ray SA, Drummond PB, Shi L, McDaniel GR, Smith EJ. 2006. Mutation analysis of the aggrecan gene in chickens with tibial dyschondroplasia. Poult Sci. 85:1169–1172.
Rozen S, Skaletsky SJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa (NJ): Humana Press. p. 365–386.
Rubin CJ, et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587–591.
Sawai H, et al. 2010. The origin and genetic variation of domestic chickens with special reference to junglefowls Gallus g. gallus and G. varius. PLoS One 5:e10639.
Schou TW, Permin A, Christensen JP, Cu HP, Juul-Madsen HR. 2010. Mannan-binding lectin (MBL) in two chicken breeds and the correlation with experimental Pasteurella multocida infection. Comp Immunol Microbiol Infect Dis. 33:183–195.
Sharma P, Bottje W, Okimoto R. 2008. Polymorphisms in uncoupling protein, melanocortin 3 receptor, melanocortin 4 receptor, and pro-opiomelanocortin genes and association with production traits in a commercial broiler line. Poult Sci. 87: 2073–2086.
Shinomiya A, et al. 2012. Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (Fibromelanosis) in silky chickens. Genetics 190:627–638.
Siegel PB, Dodgson JB, Andersson L. 2006. Progress from chicken genetics to the chicken genome. Poult Sci. 85:2050–2060.
Singh D, Singh PK, Chaudhary S, Mehla K, Kumar S. 2012. Exome sequencing and advances in crop improvement. Adv Genet. 79:87–121.
Siu RK, et al. 2012. NELL-1 promotes cartilage regeneration in an in vivo rabbit model. Tissue Eng Part A. 18:252–261.
Somes RG Jr. 1988. International Registry of Poultry Genetic Stocks. In. Storrs Agr. Exp. Sta Bull. 476. Storrs (CT): The University of Connecticut.
Song C, et al. 2011. Evaluation of SNPs in the chicken HMGA2 gene as markers for body weight gain. Anim Genet. 42:333–336.
Takahashi H, Sasaki O, Nirasawa K, Furukawa T. 2010. Association between ovocalyxin-32 gene haplotypes and eggshell quality traits in an F2 intercross between two chicken lines divergently selected for eggshell strength. Anim Genet. 41:541–544.
Tang S, et al. 2010. Evaluation of the IGFs (IGF1 and IGF2) genes as candidates for growth, body measurement, carcass, and reproduction traits in Beijing You and Silkie chickens. Anim Biotechnol. 21:104–113.
Teer JK, Mullikin JC. 2010. Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet. 19:R145–R151.
Vallejo RL, et al. 1998. Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F2 intercross chickens. Genetics 148:349–360.
Wang J, et al. 2005. ChickVD: a sequence variation database for the chicken genome. Nucleic Acids Res. 33:D438–D441.
Wang K, Li M, Hakonarson H. 2010a. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38:e164.
Wang X, Nahashon S, Feaster TK, Bohannon-Stewart A, Adefope N. 2010b. An initial map of chromosomal segmental copy number variations in the chicken. BMC Genomics 11:351.
Wang Y, et al. 2012a. The crest phenotype in chicken is associated with ectopic expression of HOXC8 in cranial skin. PLoS One 7:e34012.
Wang Y, et al. 2012b. A genome-wide survey of copy number variation regions in various chicken breeds by array comparative genomic hybridization method. Anim Genet. 43:282–289.
Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. 2009a. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. Mol Cell. 33:591–601.
Warzecha CC, Shen S, Xing Y, Carstens RP. 2009b. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. RNA Biol. 6:546–562.
Warzecha CC, et al. 2010. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. EMBO J. 29: 3286–3300.
Winter SF, Lukes L, Walker RC, Welch DR, Hunter KW. 2012. Allelic variation and differential expression of the mSIN3A histone deacetylase complex gene Arid4b promote mammary tumor growth and metastasis. PLoS Genet. 8:e1002735.
Wong GK, et al. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432:717–722.
Wright D, et al. 2009. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS Genet. 5:e1000512.
Wu MY, Tsai TF, Beaudet AL. 2006. Deficiency of Rbbp1/Arid4a and Rbbp1l1/Arid4b alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain. Genes Dev. 20: 2859–2870.
Wu X, et al. 2007. Associations of Gonadotropin-Releasing Hormone Receptor (GnRHR) and Neuropeptide Y (NPY) Genes' Polymorphisms with Egg-Laying Traits in Wenchang Chicken. Agr Sci China. 6: 499–504.
Xie C, Tammi MT. 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 10:80.
Xie X, et al. 2012. Genome-wide analysis of the ATP-binding cassette (ABC) transporter gene family in the silkworm, Bombyx mori. Mol Biol Rep. 39:7281–7291.
Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and

medium sized insertions from paired-end short reads. Bioinformatics 25:2865–2871.
Ye X, Avendano S, Dekkers JC, Lamont SJ. 2006. Association of twelve immune-related genes with performance of three broiler lines in two different hygiene environments. Poult Sci. 85: 1555–1569.
Yoshimura T, et al. 2003. Light-induced hormone conversion of T4 to T3 regulates photoperiodic response of gonads in birds. Nature 426: 178–181.
Zhang H, et al. 2012. Selection signature analysis implicates the PC1/PCSK1 region for chicken abdominal fat content. PLoS One 7: e40736.
Zhang X, Zara J, Siu RK, Ting K, Soo C. 2010. The role of NELL-1, a growth factor associated with craniosynostosis, in promoting bone regeneration. J Dent Res. 89:865–878.
Zhang X, et al. 2011. The Nell-1 growth factor stimulates bone formation by purified human perivascular cells. Tissue Eng Part A. 17:2497–2509.
Zhu JJ, et al. 2001. Screening for highly heterozygous chickens in outbred commercial broiler lines to increase detection power for mapping quantitative trait loci. Poult Sci. 80:6–12.
Zou X, et al. 2011. NELL-1 binds to APR3 affecting human osteoblast proliferation and differentiation. FEBS Lett. 585:2410–2418.

Associate editor: Takashi Gojobori