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 Li1,2,8,*

1Biodiversity Research Center, Academia Sinica, Taipei, Taiwan
2Genomics Research Center, Academia Sinica, Taipei, Taiwan
3Department of Animal Sciences, National Chung Hsing University, Taichung, Taiwan
4Center for the Integrative and Evolutionary Galliformes Genomics (iEGG Center), National Chung Hsing University, Taichung, Taiwan
5Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan
6Molecular Biology of Agricultural Sciences, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan
7Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan
8Department 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 variations 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
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), polyductyly (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 Ix 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/web-site) 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) = 3, 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 (mismatch) = 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 200–1,000-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 PRIMER3 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 MgCl₂, 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 nSNPs 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. 0.2.4p 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 (log2 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’-G CTGCGAGGCAGTTGAACT-3’ and 5’-GCGAGGCGCTTCCCA GGATCA-3’. The primers for the propionylcoenzyme A carboxylase gene (PCCA, chr1) and thyroid hormone-inducible hepatic protein gene (THRP, chr1) are used as references as described (Wang et al. 2010b). The qPCR reaction was done in 10μl 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

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

if $H_{ROH} \leq -1, N_{t,w} \geq 10$ and $N_{het,w} / N_{t,w} \leq 10%$

$N_{t,w}$ = 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 $\leq -1$. 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, 442, 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 genome of the red jungle fowl revealed ~7.6 million SNPs and 635,574 and 633,782 indels (1–73 bp) were found for Silkie and L2, respectively (fig. 1). Comparison of the SNPs and Indel Identification

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
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.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Statistics of Genetic Variants in the Silkie and L2 Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Silkie-RJFa</td>
</tr>
<tr>
<td>All SNPs</td>
<td></td>
</tr>
<tr>
<td>Intergenic</td>
<td>2,828,057</td>
</tr>
<tr>
<td>Intergenic (Upstream w/5-kb)</td>
<td>291,237</td>
</tr>
<tr>
<td>Intergenic (Downstream w/5-kb)</td>
<td>263,964</td>
</tr>
<tr>
<td>Intergenic (Up/Down w/5-kb)</td>
<td>75,295</td>
</tr>
<tr>
<td>Genic</td>
<td></td>
</tr>
<tr>
<td>Intergenic</td>
<td>2,463,943</td>
</tr>
<tr>
<td>ncRNA</td>
<td>590</td>
</tr>
<tr>
<td>5' UTR</td>
<td>3,827</td>
</tr>
<tr>
<td>3' UTR</td>
<td>24,239</td>
</tr>
<tr>
<td>5'/3' UTR</td>
<td>65</td>
</tr>
<tr>
<td>Exonic splice site</td>
<td>1,395</td>
</tr>
<tr>
<td>Intronic splice site</td>
<td>917</td>
</tr>
<tr>
<td>Exonic</td>
<td>67,503</td>
</tr>
<tr>
<td>Stop-gained</td>
<td>48,112</td>
</tr>
<tr>
<td>Stop-loss</td>
<td>232</td>
</tr>
<tr>
<td>Nonframeshift indel</td>
<td>23</td>
</tr>
<tr>
<td>Frameshift indel</td>
<td>948</td>
</tr>
<tr>
<td>Nonsynonymous</td>
<td>19,259</td>
</tr>
</tbody>
</table>

*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 downstream 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' untranscribed regions.
*m*Variants overlap 3' untranscribed 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 one or more nucleotides that cause frameshift changes in protein coding sequence.
*v*Indels of three or multiples of three nucleotides that do not cause frameshift changes in protein coding sequence.
*x*Single nucleotide changes that cause amino acid changes.

**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
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 $\chi^2$ 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,
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

---

**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).
42.49% in the set of all genes in the Silkie and L2 genomes (Pearson’s χ² 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.

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 with 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

![Fig. 3.—Distribution of indel lengths in (A) the Silkie and (B) the L2.](image)

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

Note.—FDR, false discovery rate.
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
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).

**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$_2$ 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.
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 \textit{SOX5} (Wright et al. 2009).

The hyperpigmentation in the Silkie is caused by duplicated regions containing endothelin 3 (\textit{EDN3}) (Dorshorst et al. 2011; Shinomiya et al. 2012). Late feathering is caused by a partial duplication of the \textit{PRLR} and \textit{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 3

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

### Table 4

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

**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.
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,
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. 2012). ESRP2 has been 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


---

**Table 6**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Ensembl ID*</th>
<th>Chr</th>
<th>Breed*</th>
<th>Mutation</th>
<th>Association</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MX</td>
<td>ENS-25999</td>
<td>1</td>
<td>L2</td>
<td>Radical nsSNP</td>
<td>Morbidity, early mortality, viral shedding, and cytokine responses</td>
<td>Livant et al. 2007; Ewald et al. 2011</td>
</tr>
<tr>
<td>POMC</td>
<td>ENS-26793</td>
<td>3</td>
<td>L2</td>
<td>Radical nsSNP</td>
<td>Production traits</td>
<td>Sharma et al. 2008; Bai et al. 2012</td>
</tr>
<tr>
<td>Ovocalyxin-32</td>
<td>ENS-15634</td>
<td>9</td>
<td>B</td>
<td>Radical nsSNP</td>
<td>Eggshell quality traits</td>
<td>Takahashi et al. 2010</td>
</tr>
<tr>
<td>GnRHR</td>
<td>ENS-21200</td>
<td>10</td>
<td>S</td>
<td>Stop gained</td>
<td>Egg-laying traits</td>
<td>Wu et al. 2007</td>
</tr>
<tr>
<td>AGAN</td>
<td>ENS-38569</td>
<td>10</td>
<td>L2</td>
<td>Radical nsSNP</td>
<td>Tibial dyschondroplasia</td>
<td>Ye et al. 2006</td>
</tr>
<tr>
<td>iNOS</td>
<td>ENS-09129</td>
<td>19</td>
<td>L2</td>
<td>Radical nsSNP</td>
<td>Performance</td>
<td>Ray et al. 2006</td>
</tr>
<tr>
<td>CCL4</td>
<td>ENS-35206</td>
<td>19</td>
<td>B</td>
<td>Radical nsSNP</td>
<td>Feather damage</td>
<td>Biscarini et al. 2010</td>
</tr>
<tr>
<td>PRDM16</td>
<td>ENS-01590</td>
<td>21</td>
<td>B</td>
<td>Radical nsSNP</td>
<td>Performance</td>
<td>Han et al. 2012</td>
</tr>
<tr>
<td>MBL2</td>
<td>ENS-04765</td>
<td>21</td>
<td>B</td>
<td>Radical nsSNP</td>
<td>Experimental Pasteurella multocida infection</td>
<td>Schou et al. 2010</td>
</tr>
<tr>
<td>LPL</td>
<td>ENS-24882</td>
<td>2</td>
<td>B</td>
<td>Radical nsSNP</td>
<td>Fat deposition</td>
<td>Liu et al. 2006</td>
</tr>
</tbody>
</table>

*Only the last five digits of the Ensembl chicken gene annotation are shown.

*B: Silkie; L2: Taiwanese L2 chicken; B: Both breeds of S and L2.


Associate editor: Takashi Gojobori