# Domestication process of Indonesian Cemani chicken: Genetic causes for phenotypic traits

Anik Budhi Dharmayanthi

Doctor of Philosophy

Department of Evolutionary Studies of Biosystems
School of Advanced Sciences
SOKENDAI (The Graduate University for Advanced Studies)

# Domestication process of Indonesian Cemani chicken: Genetic causes for phenotypic traits

**Anik Budhi Dharmayanthi**

**DOCTOR OF PHILOSOPHY**

**Department of Evolutionary Studies of Biosystems**

**School of Advanced Sciences**

**The Graduate University for Advanced Studies**

**2018**

## ACKNOWLEDGMENT

First and foremost, I owe my deepest gratitude to my advisor Dr. Yoko Satta. I am so grateful that you gave me an opportunity to be your PhD student. Thanks for your generosity, encouragement, respect, thoughtful guidance, patience and never-ending enthusiasm throughout all these years.

Dr. Yohey Terai, You have helped me so much during my experiments. Thank you for generously showing techniques in experiment, for patiently explaining the details of procedure of doing experiment. You are so knowledgeable and kind teacher. You have helped me in many other things.

Dr. Toyoko Akiyama, thank you for chicken samples, valuable lesson about pigmentation in chicken and an opportunity to present my research in the conference of pigmentation.

Dr.Quintin: thank you very much for your kindness, for always being ready help me for checking my English and giving valuable comments. It is indeed improve content of my thesis. Lastly, thank you for providing delicious dinner for my family and me.

Dr. Yonezawa Takahiro: Thank you for introducing me to Satta sensei and teaching me DNA technique and analysis. Thank you for your help to obtain Silkie's chicken sample and your effort to involve me in Japanese poultry and animal science society.

Dr. Sri Sulandari (deceased) dan M.Syamsul Arifin Zein, M.Sc, thank you for your support. Dr. Sri, you are one of my motivator that inspiring me a lot to be a researcher.

I would like to thank Dr. Wen-Hsiung Li for allowing me has a short-study in Academia Sinica and Dr. Wen-Lang Fang for teaching me to analyze whole genome sequence data.

Dr. Kenta Yashima and Ken Nagata, I am deeply indebted to you for all your help in analyzing my data using Python.

I would also thank to Pei-Ju, for delicious food from Taiwan. You have helped me so much during my stay in Taiwan.

I indebted to all my colleagues in SOKENDAI, especially Satta's Lab members (Gojobori-San, Fujito-San, Nami-chan, Shiho-San, Kumi-chan, Ri-chan, Ayaka-San, Mieko-San, Wanjing-San and Seiko-Kun), for your support since my first time arrived in Japan and during the entire duration of my doctoral studies.

I am incredibly grateful to my parents, for their unconditional love and support also for endless prayer to her daughter who is struggling to finish her doctoral study. I would also like to thanks to my lovely husband and baby Andra, thank you for your everyday support and smile, for always patiently accepting me and my endeavors, I can't imagine having done all this without them by my side.

Lastly, I would like to thanks to the MEXT for a scholarship and SOKENDAI for my short-study scholarship in Taiwan.

**Domestication process of Indonesian Cemani chicken: genetic causes for phenotypic traits**

## Abstract

Understanding how phenotypes diverge and what genetic factors control phenotypic changes in domestic animals is one of the focuses when studying evolutionary biology. Researchers so far have studied genetic associations of morphological or physiological traits that contribute to diversification of domestic animals. Nevertheless, examples to find out the signature of artificial selection associated with these phenotypes are few. With development of advanced sequencing technology, complete genomes of organisms have been able to be sequenced. In addition, the development of bioinformatics has aided in the increasing ease for handling such big data sequences.

This thesis focuses on specific traits of Cemani chicken, like the fibromelanosis (Fm) phenotype, to understand the domestication process through artificial selection, and how mutations can contribute to phenotypic differences in domestic chicken. In addition, this thesis investigates genetic regions under selective sweeps and predicted candidate genes associated to Cemani traits. Moreover, I elucidate the genetic relationship between Cemani and other black chickens. Overall, this thesis provides a better understanding of the genetic basis of complex traits and the evolutionary history of domestic chicken, particularly in Indonesian *Ayam* Cemani. The specific summary from each chapter is presented below:

**Chapter 2**

This chapter predominantly focused on the fibromelanosis (Fm) phenotype in Indonesian *Ayam* Cemani and Chinese Silkie chicken. I proposed the evolutionary history of the Fm phenotype in Cemani and Silkie chicken by analyzing the Fm region including segmental duplications on chromosome 20 that involve the Endothelin 3 gene, *EDN3*.

Examination of the Fm region included four major components. (i) Detection of duplication boundaries of Fm chickens (Cemani and Silkie) and other domesticated chickens as control showed that duplicated boundaries were detected in Fm chickens but not wild type chicken. (ii) qPCR analysis of *EDN3* of Cemani, Silkie and other domesticated chickens and copy number variation analysis using whole sequence of duplication segment of Fm type (Cemani and Silkie) and wild type (Taiwanese) chickens concluded that Cemani and Silkie have identical genetic rearrangement of Fm phenotype due to duplication segment containing *EDN3*, indicating a single origin of the genetic cause of the Fm phenotype. (iii) Sequence analysis of 1kb of *EDN3* revealed that the duplication arose by unequal crossing-over between alleles with 0.3 MYR divergences in the ancestral Red Jungle Fowl population. (iv) Identification of selective sweeps in the Fm region (including *EDN3*) as a target region of Cemani and Silkie revealed different lengths of heterozygosity reduction in surrounding duplicated regions which suggests the region was artificially selected independently in Cemani and Silkie breeds. Furthermore, I estimated that the two breeds have diverged around 6600 ~ 9100 years ago, suggesting that the divergence of these breeds is consistent with the beginning of domestication of chicken in China.

**Chapter 3**

Homozygosity approach was used in this chapter to analyze a single whole genome sequence of Cemani chicken for detecting signatures of selection in the Cemani genome and identify candidate genes within these regions of putative selective sweeps. I calculated the homozygosity in every 100 kb window width of whole genome sequences of Cemani, Silkie and L2 Taiwanese (single individual each) and extracted the region with homozygosity ratio ≥0.95 (referred to as high homozygosity region, HHR). I compared HHRs among Cemani, Silkie and L2 Taiwanese and identified the genes

located within HHRs shared between the three breeds as well as in HHRs specific to Cemani chicken. I then validated the monomorphism in Cemani-specific HHRs and found that *EGFR* on chromosome 20, as well as *NT5C1A* and LOC419677 on chromosome 23 were monomorphic, indicating that these genes were under selective sweeps. This was supported by further examination in the region surrounding the genes that were also identified as monomorphic. Investigation of the function of *EGFR* revealed that the gene might have two different roles (cell pigmentation and cell growth controller), supporting that this gene may have pleiotropic effects on phenotypic traits in Cemani and commercial chickens. In addition, investigation of *NT5C1A* and LOC419677 function identified that these genes are related to fecundity traits of Kauai chickens and are positively selected in commercial chickens. Taken together, the findings in this chapter suggest that Cemani chickens are a breed of Indonesian local chickens with qualities and genetic attributes that are worthy to be developed as a commercial chicken.

**Chapter 4**

This chapter aimed to elucidate the origin of Cemani chicken and reveal the genetic relationship between black chickens in Indonesia (Black Kedu, Cemani and Black Sumatra), America (Black Java and Sumatra), and China (Silkie, Muchuan, Jiuyuan, Emei, Tianfu). This study used in total 60 whole genome sequence (WGS) data from 15 breeds of chickens: 10 breeds of black chickens, White Leghorn, L2 Taiwanese, Pengxian, Red and Green Jungle Fowl. Principle component analysis (PCA) using SNPs from chromosome 20 of chicken breeds revealed distinct clusters and distribution patterns of Indonesian, Chinese, and American chickens. This suggests that the different geographical distribution of Indonesian, American and Chinese chickens causes limited contact or crossbreeding between the breeds, thus limiting gene flow between the chickens and influencing genetic variation among them. In contrast with my study in

chapter 2, which revealed close relatedness between Cemani and Silkie in the Fm region, these two breeds were in a distinct cluster based on PCA on chromosome 20. Silkie was clustered together with Chinese chickens and Cemani clustered together with Indonesian chickens, indicating that selection for the Fm phenotype in Cemani and Silkie arose recently. Similarly, in BK and Cemani, these breeds shared genetic information in the Fm region but were distantly related based on microsatellite and mitochondrial DNA analysis from published studies [1,2,3]. Finally, I concluded that Cemani might be an independent breed that was brought to Kedu village and experienced interbreeding and selection with BK chickens, resulting in genetic introgression in Fm region between the two breeds.

**Reference**


1. Ismoyowati, Sukardi and Agus Susanto. Genetic Diversity of Kedu Chicken Based on Phenotypic Characteristics and Microsatellite Loci. International Journal of Poultry Science. 2012;11(9):605-610, ISSN 1682-8356
2. Ashari H, Sulandari S, Zein MSA, Han JL. Genetic Diversity and Relationship of Indonesian Indigenous Chickens Inferred from Microsatellite DNA Markers. Tropical Agricultural Research. 2015; 26(4): 642–654.
3. Sulandari S, Zein, MSA, Sartika T. Molecular Characterization of Indonesian Indigenous Chickens based on Mitochondrial DNA Displacement (D)-loop Sequences. HAYATI Journal of Biosciences. 2008;15(4):145-154. https://doi.org/10.4308/hjb.15.4.145

## Table of content

**Acknowledgment** ii

**Abstract** iii

**Table of content** vii

**Chapter 1 Research Background**

**1.1 Domestication** 1

**1.2 Domestication of chicken** 1

**1.3 Variation and gene associated traits in domestic animal** 2

**1.4 Variation and gene associated traits in domestic chicken** 3

**1.5 Cemani chicken : Morphology, traits and history** 6

**1.6 Selective sweep analyses and its application to domestic chicken** 7

**1.7 Purpose of This Study** 9

**1.8 Reference** 10

**1.9 Figure and Table** 14

**Chapter 2 The Origin and Evolution of Fibromelamosis in Domesticated chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds**

**2.1 Introduction** 17

**2.2 Materials and Methods** 20

*2.2.1 Chicken breeds and DNA samples* 20

*2.2.2 Construction of Cemani genomic sequence library* 20

*2.2.3 Public data used* 21

*2.2.4 Amplification of duplication boundaries* 21

*2.2.5 Quantification of gene copy numbers* 21

*2.2.6 Amplification and sequencing of EDN3* 22

*2.2.7 Reads of data from the chicken WGS and nine 3 kb regions, and CNV analysis of the Fm region* 22

*2.2.8 Statistical and population genetics analyses* 23

*2.2.9 Data deposition* 24

**2.3 Results** 24

*2.3.1 Single origin of the Fm phenotype* 24

*2.3.2 Haplotype diversity and duplication of EDN3* 28

*2.3.3 High- homozygosity tracts (HHTs) in Cemani and Silkie* 30

**2.4 Discussion** 32

*2.4.1 Genomic configuration of the* Fm *region* 32

*2.4.2 DR1 duplication and emergence of the Fm phenotype* 33

*2.4.3 Ayam Cemani and Ayam Kedu in Indonesia* 35

*2.4.4 History and strength of artificial selection* 35

**2.5 Conclusions** 37

**2.6 References** 38

**2.7 Figure and Table** 41

**2.8 Supporting Information** 54

**Chapter 3 Analysis of Cemani chicken genome reveals candidate genes under selective sweeps**

**3.1 Introduction** 74

**3.2 Materials and Methods** 75

*3.2.1 SNVs (Single Nucleotide Variations) data* 75

*3.2.2 DNA samples* 75

*3.2.3 Comparison of high homozygosity region (HHR) in Cemani, Silkie and L2 Taiwanese chicken* 76

*3.2.4 Identification and validation of genes in HHRs specific to Cemani.* 77

**3.3 Results** 77

*3.3.1 High homozygosity regions (HHRs) in Cemani, Silkie and L2 Taiwanese chicken* 77

*3.3.2 Identification of High Homozygosity Regions (HHRs) shared in Cemani, Silkie and L2 Taiwanese* 78

*3.3.3 High Homozygosity Region (HHRs) specific to Cemani* 78

*3.3.4 Identification and validation of genes in Cemani-specific HHRs* 79

**3.4 Discussion** 80

*3.4.1 Genes located in High homozygosity regions (HHRs) specific to Cemani* 82

*3.4.2 Epidermal growth factor receptor (EGFR) as a candidate gene under selection* 83

*3.4.3 NT5C1A and LOC419677 as a candidate locus under selection* 85

**3.5 Conclusion** 86

**3.6 Reference** 86

**3.7 Figure and Table** 89

**3.8 Supporting Information** 95

**Chapter 4 The origin of Ayam Cemani and its relationship with other black chicken breeds from Indonesia, America and China.**

**4.1 Introduction** 126

**4.2 Materials and Methods** 128

*4.2.1 Genomic DNA samples* 128

*4.2.2 Whole Genome Sequence of chicken breeds* 128

*4.2.3 Analysis of variation in duplication boundaries* 129

*4.2.4 Mapping and SNP calling* 129

*4.2.5 Population structure and evolutionary history analysis* 130

**4.3 Results** 130

*4.3.1 Variation in duplication boundary* 130

*4.3.2 Mapped reads, SNP calling and SNP identification* 132

*4.3.3 PCA analysis of chromosome 20* 132

**4.4 Discussion** 132

**4.5 Conclusion** 137

**4.6 Reference** 138

**4.7 Figure and Table** 140

**Chapter 5 Conclusion and Future perspective** 148

# Chapter 1

# Research Background

## 1.1 Domestication

Domestication is a process when plants or animals change their traits due to the alteration of environment and/or human selection. Domestication significantly contributes to the evolutionary process of domesticated taxa [1]. The first publication in Charles Darwin's epos "The Variation of Animals and Plants under Domestication" explained that the domestication process influences the variation in plants and animals, in which this variation is not only visible traits through differences between wild domestic plants and animals in phenotype but also behavior [2]. Nowadays, we can see that domestication process has resulted in producing huge diversification of phenotypes between wild and domestic organisms, and this variation in domestic animals has been considered as a model for both rapid and drastic evolution.

Most domestic animals have domesticated not long after people began farming and living in permanent settlements. However, exactly when and how the domestication process of animals began still remains an open question. The earliest evidence of domestication process was found in dog around 15,000 years ago [3]. However, new ancient DNA studies revealed that dogs were domesticated much earlier from the ancestor of modern wolves around 20,000 and 40,000 years ago [4]. After the domestication of dogs, the domestication of other animals used as a source of food such as sheep, goats, cattle and pigs, followed.

## 1.2 Domestication of chicken

Domestication in chicken has begun more recent than other animals like dog, cattle, sheep and pig. Studies of chicken fossils found in several places such as in Yellow river in Northern China [5], China [6], and Harappa and Mohenjo-Daro site in Pakistan

[7], were dated to be around 2000-7500 years old. This indicates that the domestication process in chicken is a recent event.

Since the invention of DNA technology, several studies using molecular approaches were conducted to reveal the origin of domestic chicken. Using a part of mitochondrial DNA (mtDNA), Fumihito, et al (1994, 1996) suggested that domesticated chicken originated directly from a single ancestor, the red jungle fowl (RJF, *Gallus gallus gallus*) from Southeast Asia. This was further supported by a micro-satellite marker study providing evidence that the origin of domestic fowls is monophyletic [10]. In contrast, other studies do not support a sole ancestor of domestic chicken. Examination of whole mtDNA and two regions of nuclear genome for the species in the genus *Gallus* implied that interspecies hybridization occurred between RJF and grey jungle fowls (GJF, *Gallus sonneratii*)), and between GJF and Ceylon jungle fowls [11]. Corresponding to Nishibori et al (2005), analysis of the *BCDO2* gene suggests the genetic introduction of yellow skin color in domestic chicken from GJF and not RJF [12]. Moreover, DNA analysis of fossil chicken in Northern China revealed that the earliest domestication process of chicken started around 10,000 years ago and also provided evidence of introgression of GJF to modern domestic chicken [13].

### 1.3 Variation and gene associated traits in domestic animal

Human intervention has been a large contributor to the domestication process of animals. The interventions of humans to wild animals have triggered differentiation of its phenotypes and behaviors. For example, selection of tame behavior eventually resulted in numerous phenotypic traits such as piebald coat color and lop ears in the fox [14], and captivity under human control corresponded to tooth shape variation between wild and domestic *Sus scrofa* pigs [15].

The phenotypic and behavioral variations in domestic animals have also

encouraged evolutionary biologists and geneticists to become involved in addressing the relationship between phenotypic traits with genetic variation in particular genes. One example is the investigation of genes associated with coat color variation in domestic animals. Most mutations in color-associated genes of domestic animals was caused by artificial selection in which the mutations were increased in frequency and fixed throughout the population [16]. One such color-associated gene is the melanocortin receptor 1 (*MC1R*) gene.

Studies of the *MC1R* associated with coat color variation in pig and wild boar revealed different evolutionary processes of this gene. Mutation in *MC1R* gene indicates strong positive selection for coat color variation in domestic pigs; this variation was found between Asian and European pigs [17] and within Chinese domestic pigs [18]. In contrast, coat color in the wild boar is maintained by purifying selection, as identified from studies in Asian and European wild boar [17] and Chinese wild boar [18].

Moreover, alteration of coat color in Tibetan pigs was correlated with adaptations in high intensive solar ultraviolet (UV) radiation in the environment. This alteration in Tibetan pig phenotypes corresponds with positive selection on *MC1R* causing dark coat color in Tibetan pigs. Meanwhile in landrace pig with white coat color, positive selection is due to intentionally selected by humans after domestication [19]. Apart from pigs, positive selection in *MC1R* has also acted in indigenous goat populations [20] and black Chinese sheep [21]. In addition, it was also reported that polymorphism in *MC1R* is associated with plumage color variation in domestic chicken [22].

### 1.4 Variation and gene associated traits in domestic chicken

Besides plumage color, divergence on body size, plumage type, comb shape, shank type and skin color has also appeared in domestic chickens during the domestication process. Genes associated with such phenotypic traits in chicken have been

well studied. In addition to *MC1R* gene associated to plumage color variation in chicken, several genes were also reported to have correlation with plumage color variation, such as *SLC45A2* on chromosome Z that is associated with plumage color variation in chicken and Japanese quail [23], and polymorphism in *PMEL17* linked to the Dominant white, Dun and Smoky color variants in chicken plumage [24]. In contrast, a more recent study using genome wide scan (GWAS) analysis suggested that polygenic effects have an impact on pigmentation variation instead of a single gene [25].

Several studies have identified the genetic basis of a number of phenotypic variations in chicken feather. (i) Frizzle type is caused by mutation on the α-Keratin gene (*KRT75*) [26]. (ii) Crest type is due to abnormal gene expression on *HOXC8* (Homeobox gene) impacting on development of cranial feathers [27]. (iii) Feather loss in chicken neck (naked neck trait) is due to the *BMP12* gene that modifies the distribution of feathers on the neck [28]. (iv) Silky-feather is correlated with spontaneous mutation in the promoter region of prenyl (decaprenyl) diphosphate synthase, subunit 2 gene (*PDSS2*), which leads to lower expression of *PDSS2* [29].

As well as plumage and feather studies, the genetic cause of diversification in chicken comb types has also been studied. Studies in comb types revealed that (i) duplex comb [V-shaped (D*V) and Buttercup (D*C)] are both associated with tandem duplication located in the upstream region of the eomesodermin gene (*EOMES*) underlying ectopic expression of *EOMES* [30], (ii) rose-comb is caused by a 7.4 Mb unequal crossing over on chromosome 7 [31], and (iii) pea comb is correlated with high expression of *SOX5* during the differentiation of cells essential for the evolvement of comb and wattles [32].

Another phenotype that has been of concern by researchers is hereditary limb malformation development in the chicken shank (polydactyl-type). Study of this

phenotype showed a similar causal mutation for polydactyly across different breeds. A single nucleotide polymorphism (SNP) in a cis-regulatory region of Sonic Hedgehog (Ssh) was significantly associated with polydactyly (Po) in Silkie and Beijing fatty chicken from China and local chicken from Europe [33,34,35].

Lastly, molecular studies related to skin pigmentation in domestic chicken have been conducted. These include (i) analysis of the cause of yellow skin color common in domestic chicken which revealed that cis-acting and tissue-specific regulatory mutations inhibit expression of *BCDO2* allows deposition of yellow carotenoids in the skin tissue [36]; and (ii) investigation of rare black skin color in dermis [Fibromelanosis (Fm) phenotype]] using qualitative trait loci (QTL) and qPCR analyseis which strongly suggested that the duplication segment containing endothelin 3 gene (*EDN3*) leads to higher expression of *EDN3*, and this duplication is correlated with the Fm phenotype in domestic chicken [37,38]. This black pigmentation can be found in several chicken breeds in China (Silkie, Dongxiang, Jianghan, Jiuyuan, Emei, and Muchuan), and those in other countries such as India (Kadaknath), Vietnam (Black Hmong chicken), Sweden (Svart Höna), Indonesia (Ayam Cemani, Sumatra), Korea (Ogol chicken), and Argentina (Argentinean Tuzo type). Even though the previous study has revealed the genetic mechanism that causes such a fascinating Fm phenotype, how and when this occurred across domestic chicken breeds scattered over the world has not yet been elucidated.

Among local chicken breeds, Silkie chicken is well-studied chicken with Fm phenotype as well as other special traits likes polydactyly, Silkie-feather feathered legs (Pti), vulture hock (V), rose comb (R), and duplex comb (D) [33]. Interestingly, molecular studies related to the unique traits in Silky have shown a correlation between phenotype-genotype changes in Silkie chicken. It is likely that there are more genetic loci that may have been targeted by selection and correlated to phenotypic changes in other

chicken breeds, driving the main objective of this thesis for investigating such unique traits in Indonesian Cemani chicken breed.

## 1.5 Cemani chicken : Morphology, traits and history

Apart from Silkie chicken, another breed with unique phenotypes is the Indonesian Cemani chicken. As mentioned before, Cemani also presents deep pigmentation in the skin like in Silkie. However, unlike white Silkie, where hyperpigmentation is restricted only to the skin, flesh and internal organs, Cemani show completely black pigmentation covering all of its body including plumage, comb, shank, tongue, and eye (Fig 1.1). In addition to black skin color, Cemani have relatively large body size with adult male weighing around 1.6 – 3.3 kg and females about 1.4 – 2.5 kg [39]. Moreover, the advantages of this breed are high egg production. This breed is also sometimes used for cultural rituals for purposes that difficult to explain scientifically.

The origin of the Cemani chicken is still unclear since there is no historical record of the bird. However, based on well-known stories, which were passed down from generation to generation, it is known that Cemani originated from Kedu village, Temanggung city, Central Java. Cemani has been classified into Kedu type together with ayam Kedu merah (red Kedu), Kedu hitam (black Kedu) and kedu putih (white Kedu) [40].

Studies about Cemani chicken are very limited and have to-date only focused on Cemani's productivity in poultry or cross-breeding of Cemani chicken with other local chicken breeds. Therefore, in this thesis, I aimed to discover potential genetic factors of Cemani chicken linked to the breed's specific traits, such as Fm phenotype, for better understanding the importance of selection in shaping Cemani phenotypes and genomes.

## 1.6 Selective sweep analyses and its application to domestic chicken

Detecting whether selective sweeps occur in target genes, as well as determining the function of target genes, is important for understanding the links between selection and gene-associated traits. In domestic animals, such analyses will address the questions regarding how selected genes contributed to dramatic phenotypic changes for a number of traits such as reproduction, body composition, and coat color during the domestication process. The principle of selective sweeps is indicated by a reduction of diversity in the target gene and the surrounding gene region, resulting in a specific allele prevailing throughout the population (Fig 1.2).

Several methods are developed for detecting positive selection at the genomic level and provide the high resolution for identifying genes that may contribute to phenotypic variation. In my thesis, some of the statistical tests used to reveal positive selection in Cemani genome include:

1. Linkage disequilibrium (LD)

The term linkage disequilibrium (LD) refers to the non-random association between neighboring loci on the same chromosome, indicating that two loci are linked to each other. Naturally, the combination of alleles at each locus is in equilibrium proportion; however the occurrence of mutations (i.e positive or negative selection) will result in a disequilibrium distribution of alleles. The difference between observed frequency of a combination of alleles and expected frequency of random association can be estimated and defined as LD. Reduction of the level of linkage disequilibrium is known as a signature of strong selective sweeps [41]. Therefore, LD is important for detecting signatures of selective sweeps, as has been supported by a number of theoretical and empirical studies [42,43,44].

2. Ewens-Watterson test (frequency distribution based test)

The Ewens-Watterson (EW) test is based on allelic distribution as well as Tajima's D and Fay & Wu's H test. The difference is that the EW test in general calculates the expected homozygosity (a comparison between the observed levels of homozygosity to the measured allele frequencies under Hardy-Weinberg equilibrium (HWE) in a population) of polymorphic sites. Higher homozygosity is indicative of low genetic diversity in a population.

3. Hudson-Kreitman-Aguade (HKA) test (heterogeneity based test)

The neutral theory is commonly used as a null hypothesis in order to predict the mode of selection. Under neutral theory, the levels of intraspecific polymorphism and interspecific divergence should be positively correlated [44]. If a sequence is subjected to selective pressure, then the null hypothesis must be rejected. The Hudson-Kreitman-Aguade (HKA) test is used to measure the level of polymorphism within species and divergence between species; a significant value corresponds to non-neutrality.

4. Population structure analysis

The program that is widely used for admixture analysis is STRUCTURE, which was developed by Pritchard et al. (2000). This method is widely used for determining genetic structure composed in a population, inferring the source population contributing to sample individuals, and inferring the distribution pattern between populations (i.e. admixture or migration) [47].

5. Haplotype based test

There are several methods used to estimate the genetic diversity based on haplotypes: (i) number of segregating sites (S)- S refers to the polymorphic sites in the sequence sample data; (ii) number of haplotype (K)- K provides a varying number of alleles in the sample data; and (iii) Haplotype diversity (H)- H represents the probability that two randomly sampled alleles are different [48].

Since the emergence of next generation sequencing approaches, there is ability for generating large data sequences that cover the whole genome sequence of an individual. In addition, SNP array tools also can identify known SNPs in the genome specific to a population. For analysis of such big data, more statistical tests have also been invented, for example, a statistical test used for scanning entire genomes to find regions that have experienced recent positive selection. Statistical tests for such purposes have been used in several studies in domestic chicken to identify regions and genes that under selective sweeps (Table 1.1).

## 1.7 Purpose of This Study

This study lays the ground work for studying: 1) the evolutionary origin of Fibromelanosis phenotype in Cemani and Silkie chicken; 2) selective sweeps on genes related to Cemani-specific traits in Cemani chicken genome; 3) history of Cemani chicken based on molecular approaches using Kedu chicken and other black plumage chicken as a comparison. Overall, this study will significantly increase our understanding about the history of chicken domestication processes especially in Cemani breeds, and provide a foundation for future studies on the distribution of Fibromelanosis chicken across the world.

## 1.8 Reference


1. Larson G, Fuller DQ. The evolution of animal domestication. Annual Review of Ecology, Evolution and Systematics. 2014; 45: 115–136.
2. Darwin C. The Variation of Animals and Plants Under Domestication. London: John Murray; 1868; Vol I.
3. Shannon LM, Boyko RH, Castelhano M, et al. Genetic structure in village dogs reveals a Central Asian domestication origin. Proc Natl Acad Sci U S A. 2015; 112(44): 13639-13644.
4. Botigué LR, Song S, Scheu A, et al. Ancient European dog genomes reveal continuity since the Early Neolithic. Nature Comm. 2017; 8:16082.
5. West B, Zhou B-X. Did chickens go North? New evidence for domestication. J Arch Sci. 1988; 15: 515–533.
6. Yuan J. Zooarchaeological study on the domestic animals in ancient China. Quaternary Sci. 2010; 30: 298–306.
7. Zeuner FE. A History of Domesticated Animals. Harper & Row New York; 1963
8. Fumihito A, Miyake T, Sumi S, Takada M, Ohno S, Kondo N. One subspecies of the red junglefowl (Gallus gallus gallus) suffices as the matriarchic ancestor of all domestic breeds. Proc Natl Acad Sci U S A. 1994; 91(26): 12505-12509.
9. Fumihito A, Miyake T, Takada M, et al. Monophyletic origin and unique dispersal patterns of domestic fowls. Proc Natl Acad Sci U S A. 1996; 93(13): 6792-6795.
10. Hillel J, Groenen MA, Tixier-Boichard M, et al. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. Genet Select Evol. 2003; 35(6): 533-557
11. Nishibori M, Shimogiri T, Hayashi T, Yasue H. Molecular evidence for hybridization of species in the genus Gallus except for Gallus varius. Anim Genet. 2005; 36: 367-75.
12. Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, et al. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. PLoS Genet. 2008; 4: e1000010.
13. Xiang H, Gao J, Yu B, Zhou H, Cai D, Zhang Y, et al. Early Holocene chicken domestication in northern China. Proc Natl Acad Sci U S A. 2014; 111(49): 17564–17569.
14. Trut L, Oskina I, Kharlamova A. Animal evolution during domestication: the domesticated fox as a model. BioEssays. 2009; 31(3): 349-360.
15. Evin A, Dobney K, Schafberg R, et al. Phenotype and animal domestication: A study of dental variation between domestic, wild, captive, hybrid and insular Sus scrofa. BMC Evol Biol. 2015; 15(1): 6.
16. Cieslak M, Reissmann M, Hofreiter M, Ludwig A. BioEssays. 2011; 86(4): 885-899.
17. Fang M, Larson G, Ribeiro HS, Li N, Andersson L. Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs. Barsh GS, ed. PLoS Genetics. 2009; 5(1): e1000341.
18. Li J, Yang H, Li J-R, Li H-P, Ning T. Artificial selection of the melanocortin receptor 1 gene in Chinese domestic pigs during domestication. Heredity. 2010; 105(3): 274-81.
19. Liu R, Jin L, Long K, Chai J, Tang Q, Tian S, et al. Detection of genetic diversity and selection at the coding region of the melanocortin receptor 1 (MC1R) gene in Tibetan pigs and Landrace pigs. Gene. 2015; 575(2).

20. Ahmad HI, Liu G, Jiang X, Liu C, Chong Y, Huarong H. Adaptive molecular evolution of MC1R gene reveals the evidence for positive diversifying selection in indigenous goat populations. Ecology and Evolution. 2017; 7(14): 5170-5180.
21. Yang G-L, Fu D-L, Lang X, Wang YT, Cheng SR, Fang SL, et al. Mutations in MC1R Gene Determine Black Coat Color Phenotype in Chinese Sheep. ScientificWorldJournal. 2013; 2013:675382.
22. Kerje S, Lind J, Schütz K, Jensen P, Andersson L. Melanocortin 1-receptor (MC1R) mutations are associated with plumage colour in chicken. Anim Genet. 2003; 34(4): 241-248.
23. Gunnarsson U, Hellström AR, Tixier-Boichard M, Minvielle F, Bed'hom B, Ito S, et al. Mutations in SLC45A2 Cause Plumage Color Variation in Chicken and Japanese Quail. Genetics. 2007; 175(2): 867-877.
24. Kerje S, Sharma P, Gunnarsson U, Kim H, Bagchi S, et al. The Dominant white, Dun and Smoky color variants in chicken are associated with insertion deletion polymorphisms in the PMEL17 gene. Genetics. 2004; 168: 1507–1518.
25. Park MN, Choi JA, Lee K-T, Lee H-J, Choi B-H, Kim H, et al. Genome-wide Association Study of Chicken Plumage Pigmentation. Asian-Australas J Anim Sci. 2013; 26(11): 1523-1528.
26. Ng CS, Wu P, Foley J, Foley A, McDonald M-L, Juan W-T, et al. The Chicken Frizzle Feather Is Due to an α-Keratin (KRT75) Mutation That Causes a Defective Rachis. Roop D, ed. PLoS Genetics. 2012; 8(7): e1002748.
27. Wang Y, Gao Y, Imsland F, Gu X, Feng C, Liu R, et al. The Crest Phenotype in Chicken Is Associated with Ectopic Expression of *HOXC8* in Cranial Skin. PLoS ONE. 2012; *7*(4): e34012.
28. Mou C, Pitel F, Gourichon D, Vignoles F, Tzika A, et al. Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. PLoS Biol. 2011; 9: e1001028.
29. Feng C, Gao Y, Dorshorst B, Song C, Gu X, Li Q, et al. A *cis*-Regulatory Mutation of *PDSS2* Causes Silky-Feather in Chickens. Leeb T, ed. PLoS Genetics. 2014; 10(8): e1004576.
30. Dorshorst B, Harun-Or-Rashid M, Bagherpoor AJ, Rubin C-J, Ashwell C, Gourichon D, et al. A Genomic Duplication is Associated with Ectopic Eomesodermin Expression in the Embryonic Chicken Comb and Two Duplex-comb Phenotypes. Leeb T, ed. PLoS Genet. 2015; 11(3): e1004947.
31. Imsland F, Feng C, Boije H, Bed'hom B, Fillon V, Dorshorst B, et al. The *Rose-comb* Mutation in Chickens Constitutes a Structural Rearrangement Causing Both Altered Comb Morphology and Defective Sperm Motility. PLoS Genet. 2012; 8(6): e1002775.
32. Wright D., Boije H., Meadows J.R., Bed'Hom B., Gourichon D., Vieaud A., Tixier-Boichard M., Rubin C.J., Imsland F., Hallbook F., et al. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS Genet. 2009; 5: e1000512
33. Dorshorst B, Okimoto R, Ashwell C. Genomic Regions Associated with Dermal Hyperpigmentation, Polydactyly and Other Morphological Traits in the Silkie Chicken. J Hered. 2010; 101(3): 339–350,
34. He C, Chen Y, Yang K, Zhai Z, Zhao W, Liu S, et al. Genetic pattern and gene localization of polydactyly in Beijing fatty chicken. Xu P, ed. PLoS ONE. 2017; 12(5): e0176113.
35. Zhang Z, Nie C, Jia Y, Jiang R, Xia H, Lv X, et al. Parallel evolution of polydactyly traits in Chinese and European chickens. PLoS ONE. 2016; 11(2): e0149010.

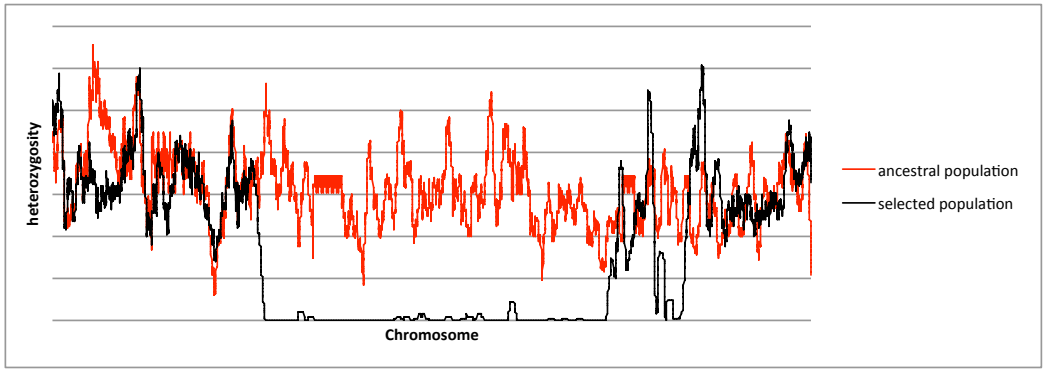
36. Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Strömstedt L, et al. Identification of the Yellow SkinGene Reveals a Hybrid Origin of the Domestic Chicken. Georges M, ed. PLoS Genet. 2008; 4(2): e1000010.
37. Dorshorst B, Molin AM, Rubin CJ, Johansson AM, Strömstedt L, Pham M-H, et al. A complex genomic rearrangement involving the Endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet. 2011; 7: e1002412.
38. Shinomiya A, Kayashima Y, Kinoshita K, Mizutani M, Namikawa T, Matsuda Y, et al. Gene duplication of the endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (Fibromelanosis) in Silkie chickens. Genetics. 2011; 111.136705.
39. Sulandari S, Zein MSA, Paryanti S, Sartika T, Astuti M, Widjastuti T, et al. Sumberdaya genetik ayam lokal Indonesia. In: Keanekaragaman Sumber Daya Hayati Ayam Lokal Indonesia: Manfaat dan Potensi. (Diwyanto, K and S. N. Prijono, Eds). Jakarta: 1st Ed. LIPI Press; 2007; 49-50.
40. Ismoyowati, Sukardi, Susanto A. Genetic Diversity of Kedu Chicken Based on Phenotypic Characteristic and Microsatellite Loci. Int J Poult Sci. 2012; 11(9): 605-610.
41. Kim Y, Nielsen R. Linkage disequilibrium as a signature of selective sweeps. Genetics. 2004; 167(3): 1513-1524.
42. Parsch J, Meiklejohn CD, and Hartl DL. Patterns of DNA sequence variation suggest the recent action of positive selection in the janus-ocnus region of Drosophila simulans. Genetics. 2001; 159: 647–657.
43. Sabeti PC, Reich DE, Higgins JM, Levine HC, Richter DJ, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002; 419: 832–837.
44. Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, et al. Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. Nature. 2002; 418: 320–323
45. Kimura, M. Evolutionary rate at the molecular level. Nature. 1968; 217: 624–626
46. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155(2): 945-959.
47. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo A, Lareu M. An overview of STRUCTURE: applications, parameter settings and supporting software. Frontiers in Genetics. 2013; 4: 98.
48. Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press; 1987.
49. Rubin CJ, Zody MC, Erikson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010; 464: 587-591.
50. Elferink MG, Megens H-J, Vereijken A, Hu X, Crooijmans RPMA, Groenen MAM. Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken Breeds. Shioda T, ed. PLoS ONE. 2012; 7(2): e32720.
51. Gholami M, Reimer C, Erbe M, et al. Genome Scan for Selection in Structured Layer Chicken Populations Exploiting Linkage Disequilibrium Information. Cao Y, ed. PLoS ONE. 2015; 10(7): e0130497.
52. Zhang H, Wang S-Z, Wang Z-P, et al. A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. BMC Genomics. 2012; 13: 704.
53. Fu W, Lee WR, Abasht B. Detection of genomic signatures of recent selection in commercial broiler chickens. BMC Genetics. 2016; 17(1): 122.

54. Boschiero C, Moreira GCM, Gheyas AA, Godoy TF, Gasparin G, Mariani PDSC et al. Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. BMC Genomics. 2018; 19: 83.
55. Zhang H, Hu X, Wang Z, Zhang Y, Wang S, Wang N, et al. Selection Signature Analysis Implicates the PC1/PCSK1 Region for Chicken Abdominal Fat Content. Androulakis IP, ed. PLoS ONE. 2012; 7(7): e40736.
56. Liu Z, Sun C, Qu L, Wang K, Yang N. Genome-Wide Detection of Selective Signatures in Chicken through High Density SNPs. Chaubey G, ed. PLoS ONE. 2016; 11(11): e0166146.

## 1.9 Figure and Table

**Fig 1.1.** Morphology of Cemani chicken and its internal organ; black pigmentation in (a) plumage, comb, and beak; (b) skin of whole body; (c) shank; (d) throat; (e) intestines; (f) ovary; (g) meat and bone

**Fig 1.2.** Signature of selective sweeps (a). Heterozygosity value (He) in ancestral and selected population; He in ancestral population (red line) shows similar value throughout chromosome; He in selected population (black line) shows reduction in and surrounding target region. (b). Schematic describing the impact of selective sweeps on pattern of genetic variation. Each row represents an allele. Dot with various colors depicts a neutral mutation except green dot is an adaptive mutation. The adaptive mutation (green dot) increases its frequency resulting in reduction in level of nearby genetic diversity in a population.



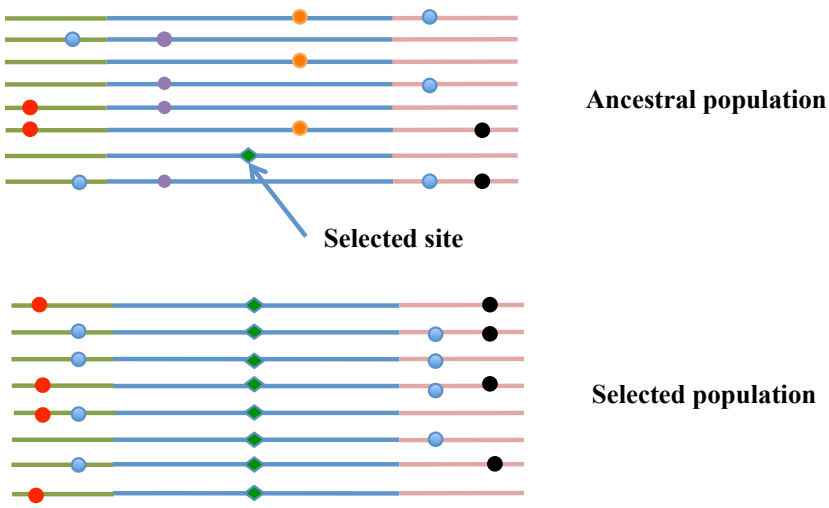



**Table 1.1.** Statistical test used for analyzing selection in domestic chicken.

| Statistical test | Candidate gene under selection | Material | Reference |
|---|---|---|---|
| Hp | Thyroid stimulating hormone receptor (*TSHR*) and 12 other genes | Pool of genomic DNA | [49] |
| Hp | 13 genes previously reported (Rubin et.al, 2010) and 13 new candidate genes (i.e *HNF4G*, NEL-like 1 (*NELL1*)) | 58K SNP of 67 breeds | [50] |
| Fst, FLK and hapFLK | *IGF-1R*, *AGRP* and *STAT5B* gene associated with growth and carcass traits.<br>*SOX10* associated with brown feather color | 1 million SNPs of 70 chicken from three breeds | [51] |
| LD, EHH | *RB1*, *BBS7*, *MAOA*, *MAOB*, *EHBP1*, *LRP2BP*, *LRP1B*, *MYO7 A*,*MYO9A* and *PRPSAP1* associated with abdominal fat in broiler chicken | 60K SNPs of broiler chicken | [52] |
| (XP-EHH) and (XP-CLR) | 91 genes include SOX6 (Sex Determining Region Y-Box 6) and cTR (Thyroid hormone receptor beta) that may have been under recent selection due to their essential roles in growth, development and reproduction in chickens. | 60K SNP | [53] |
| Fst | > 250 genes were detected with a signature of selection associated with important phenotypic traits in chicken, such as lipid metabolism, growth, reproduction, and cardiac development. | 13.93 million SNPs of 28 chickens from meat and white egg-type chicken line | [54] |
| AFD, LD, EHH | PC1/PCSK1 correlated with abdominal fat content in chicken | 60K SNP | [55] |
| LD, EHH | *AASDHPPT*, *GDPD5*, *PAR3*, *SOX6*, *GPC1* and a signal pathway of *AKT1* associated with immune function, sensory organ development and neurogenesis | 600K SNP | [56] |

Heterozygosity pool: Hp; LD: Linkage Disequilibrum; EHH: Extended Haplotype Homozygosity; XP-EHH: cross-population extended haplotype homozygosity; XP-CLR: cross-population composite likelihood ratio; AFD: allellic frequency difference

# Chapter 2

## The Origin and Evolution of Fibromelanosis in Domesticated Chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds

### 2.1 Introduction

With conspicuously diversified phenotypes with respect to morphological, physiological, and behavioral traits, domesticated animals are excellent model organisms for investigating underlying genetic changes as well as for elucidating the underlying evolutionary mechanisms. It is widely accepted that phypes currently observed in domesticated organisms are usually selected from variations that arose spontaneously in wild, ancestral populations. There have been several attempts to identify the genetic bases of such phenotypes and compare them with those of wild ancestors [1–5]. One of the common underlying ideas is that artificial selection reduces the level of genetic variability at linked neutral sites when a selected allele rapidly increases in frequency toward fixation (selective sweep) and/or is kept fixed in a breeding population for a relatively short period. On the basis of 2.8 million SNPs, the International Chicken Polymorphism Map Consortium [6] scanned the genomes of three chicken breeds (Broiler, Layer, and Silkie) and Red Jungle Fowl (RJF), the wild ancestor of domestic poultry. It was found that relatively high levels of genetic variation with nucleotide diversity $\pi = 0.5\%$ are maintained within chicken breeds; however, little evidence is provided for selective sweeps by adaptive (favored) alleles on length scales greater than 100 kb. One reason for the lack of such long-stretch signals could be the rather high recombination rate in chickens [7]. In a small-scale study of 32 introns randomly selected in two chicken breeds (Silkie and Koshamo or fighting cock), RJF and Green Jungle Fowl (GJF), Sawai et al. [8] also showed that domesticated chickens usually maintain nearly the same level of nucleotide diversity as their ancestral RJF population. The authors further argued that

genomic regions that respond to domestication might be rather limited. However, re-sequencing of genomic DNA pools representing eight different populations of domesticated chickens and RJF demonstrated a number of regions under selective sweeps [9]. Another selective sweep analysis of feral Kauai chickens derived from domesticated populations identified genomic regions that are associated with comb mass, maternal brooding behavior and fecundity [10]. Unfortunately, however, these studies did not cover all interesting phenotypes of the various chicken breeds, including Silkie (S1 Figure).

In contrast to these genome-wide scan approaches, I took a candidate gene approach and focused on a particular phenotype known as fibromelanosis (Fm) or dermal hyperpigmentation [11]. Mutations of the Fm gene result in excessive accumulation of black pigment in the skin and several other tissues or organs such as blood vessels, muscles, gonads and tracheas. Chinese Silkie is one of the domesticated chicken breeds with the Fm. The phenotype is inherited in a Mendelian fashion with semidominance [12]. Recombinant analysis using Silkie and Black Minorca (a homozygote for the wild-type chromosome regarding Fm) located the genomic region of Fm between 10.2 Mb and 11.7 Mb on chromosome 20 [12, see also 13,14].

It has been established that the Fm mutation is positively correlated with the duplication of a segment that contains the EDN3 gene encoding endothelin 3 [12,14,15]. EDN3 is a major controller of neural crest cell movement and proliferation. Neural crest cells are pluripotent and thus can develop into several cell types, such as melanoblasts [16–19]. Melanocytes, which differentiated from melanoblasts, produce eumelanin (black and dark pigment melanin) and phaeomelanin (colored melanin) in the skin, comb and other organs [20]. The amount of EDN3 mRNA in whole Silkie embryos at 18 days is approximately twice as high as that in wild-type chicken embryos [12,14]. Thus, *EDN3* is

the most likely candidate gene for such coloring phenotypes in Silkie as well as other domesticated animals, including cats [21] and cattle [22]. Indeed, PCR and next-generation sequencing (NGS) analyses of the Silkie genome unveiled segmental duplication in the Fm region [14,23]. Previously, Dorshorst et al. [14] showed that two regions (DR1 and DR2), separated by a 417 kb spacer, underwent inverted duplication. In the reference genome (Gallus_gallus_4.0, http://www.ncbi.nlm.nih.gov), DR1 is located at nt positions 11,111,559 to 11,238,796 and DR2 at positions 11,651,876 to 11,822,527 on chromosome 20. Each of the duplicated DR1s is 127 kb long, and contains not only *EDN3* but also *HIVEP3, SLMO2, ATP5E,* and *TUBB1*, whereas each of the duplicated DR2s is 171 kb long and is devoid of genes.

Dermal hyperpigmentation is found in other domesticated chicken breeds, such as Ayam Cemani in Indonesia (S2.1 Figure), Kadakhnath in India, Black H'Mong in Vietnam, Argentinean Tuzo type in Argentina, and Svarthöna in Sweden. While they all show excessive melanin accumulation, the overall phenotypes of Cemani and other black chickens differ greatly from those of Silkie [13,14]. For instance, unlike White Silkie, which shows fluffy feathering, Cemani shows black plumage and non-fluffy feathering. Moreover, comb shape in Cemani males is very different from that in Silkie males with rose combs. In light of these similarities and differences between Cemani and Silkie, Shinomiya et al. [12] and Dorshorst et al. [14] examined whether the Fm region in Cemani is functionally and structurally the same as that in Silkie. Shinomiya et al. [12] analyzed the progenies of sib-crosses of F1 hybrids between Cemani and Ayam Arab (a wild-type domesticated breed in Indonesia). Based on the copy number variation (CNV) observed in EDN3 by quantitative PCR (qPCR), they suggested EDN3 duplication in Cemani, but not in Ayam Arab. Similarly, using a PCR-based diagnostic test, Dorshorst et al. [14] found that the complex arrangement of DR1 and DR2 is shared among Silkie,

Cemani, Black H'Mong and Svarthöna. However, because the two copies of DR1 and DR2 cannot be easily distinguished from each other by PCR or NGS, the precise genomic arrangement of these four regions has not fully been elucidated even in Silkie.

In this study, I compared the genomic structure, the pattern and level of DNA variation, and the evolutionary history of the Fm region between Cemani and Silkie. I paid particular attention to the genomic signature of artificial selection on a target gene, *EDN3*, and used it to estimate the duration and strength of artificial phenotypic selection.

## 2.2 Materials and Methods

### *2.2.1 Chicken breeds and DNA samples*

Most domestic chickens of Cemani, Silkie and other breeds as well as the two jungle fowl species (RJF and GJF) were collected at various locations in Indonesia (Table 2.1). Chickens for DNA isolation were collected at farms in rural areas of Java, Sumatra, Sulawesi and Nusa Tenggara, Indonesia, mainly in 2005–2010, although some were obtained in more recent years.

### *2.2.2 Construction of Cemani genomic sequence library*

Total DNA of one Cemani chicken (sample "Cemani 41") was sheared into ~500-bp fragments using an M220 focused ultrasonicator™ (Covaris) and a genomic library NGS was prepared in accordance with the True Seq DNA PCR-free sample preparation protocol. Another genomic DNA library was prepared in accordance with protocols provided with the Illumina Nextera X, for nine regions, each of about 3 kb in length with ~200-kb intervals. Each 3-kb region was amplified by primers designed in-house (S1 Table) in a 30 μl reaction mixture (see S2.2 Table for the reaction conditions of PCR1). PCR products of each sample (5 μl) were pooled and purified using Agencourt AMpure

XP (Beckman Coulter). The libraries were qualitatively and quantitatively verified using an Agilent Bioanalyzer and sequenced on the Illumina HiSeq2000 platform (Illumina).

### *2.2.3 Public data used*

The chicken reference genome was downloaded in September 2015 from the UCSC Genome Browser (https://genome.ucsc.edu/) and has the same sequence as that deposited in the GenBank database (Gallus_gallus_4.0). Additionally, full data sets of Silkie (accession numbers: SRX286765, SRX286766, SRX286773, SRX286776, and SRX286777, see ref. 23) and Taiwanese L2 (accession numbers: SRX286779-SRX286781, SRX286798, and SRX286799, see ref. 23) were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/).

### *2.2.4 Amplification of duplication boundaries*

The presence or absence of duplication boundaries was examined by PCR with two previously published primer pairs, A2 and B2 [14], on a 96-Well GeneAmp® PCR System 9700 from Applied Biosystems (see S2.2 Table for the detailed reaction conditions of PCR2).

### *2.2.5 Quantification of gene copy numbers*

PCR products for *EDN3* (113 bp, primer set qAS044) and *uridine-cytidine kinase 1-like 1* (*UCKL1*) (124 bp, primer set q46) were ligated to the pMD20 vector (TaKaRa Japan). Plasmid DNA was extracted using alkaline lysis [24] and the concentration was determined using NanoVue spectrophotometer (GE Healthcare). Plasmid DNAs were diluted to $10^{-1}$ to $10^{-6}$ ng/μl in distilled water and were used to draw a standard curve for quantification. qPCR for absolute quantitative analysis was carried out with the SYBR® Premix Ex Taq™ II (Tli RNaseH Plus; TaKaRa). All reactions were run in triplicate on a

Thermal Cycler Dice (Applied Biosystems), and the thermal cycling conditions were as indicated under "PCR3" in S2.2 Table.

### *2.2.6 Amplification and sequencing of* EDN3

A ~1-kb genomic fragment encompassing exons 4 to 5 of *EDN3* was amplified and sequenced with the previously reported primer pair AS044F and AS044R [12]. PCR was conducted under reaction conditions listed under "PCR4" in S2.2 Table. Amplified PCR products were purified by isopropanol precipitation and directly sequenced. For heterozygous sequences, the PCR products were cloned into pMD20, and eight clones for each product were sequenced using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) with M4 and Rv universal primers on an ABI 3130xl sequencer (Applied Biosystems).

### 2.2.7 *Reads of data from the chicken WGS and nine 3 kb regions, and CNV analysis of the* Fm *region*

The CLC Genomic Workbench 8.0.3 (www.qiagenbioinformatics.com) was used to map the 3-kb region reads to the reference genome with 90% length and similarity fractions.

To analyze the WGS of Cemani, Silkie, and Taiwanese L2, low-quality bases were removed with the Trimmomatric software [25], using the following parameter settings; leading = 10, trailing = 10, sliding windows = 4:15, and minlen = 40. The Samtools workflow [26–30] (http://www.htslib.org/workflow/) was used for mapping of the WGS data with 30X coverage.

To examine CNV in the *Fm* region, the reads from each of three pairs among Cemani, Silkie, and L2 WGS data were mapped onto nt positions 10,700,000–12,000,000

on chromosome 20 using the CNV-Seq software [31]. Default parameters (log2-threshold = 0.6, *p*-value = 0.01, and minimum windows = 4) were used to produce the CNV list.

### *2.2.8 Statistical and population genetics analyses*

The DNA sequences were aligned with the ATGC software (GENETIX). The number of nucleotide differences per site (*p*-distance) was calculated with Molecular Evolutionary Genetics Analysis (MEGA6) [32]. Neighbor-joining (NJ) trees [33] were constructed with 1000 bootstrap resampling with an option of complete deletion of gaps/missing nucleotides. The ratio of the extent of divergence to that of polymorphism between any of the nine 3-kb regions was tested using the HKA test [34] implemented in DnaSP [35]. Genetic components in Cemani, Silkie, other domesticated breeds, RJF, and GJF were examined using STRUCTURE (version 2.2) analysis [36] on the http://pritch.bsd.uchicago.edu/software_website. Heterozygosity at individual SNP sites was computed based on allele frequencies in the samples (S2.3 Figure).

Calculations of the allele age or the time span of artificial selection operation were based on the same idea used for adaptively introgressed tracts [37,38]. It was assumed that the probability of observing a tract length $\geq x$ follows the exponential distribution:

$$\mathrm{P}\{tract \geq x\} = \exp(-\frac{x}{L}) \qquad (1)$$

where $L$ is the mean tract length. This $L$ is given approximately by $L = \frac{1}{r^*t}$, in which $t$ is the number of generations elapsed during artificial selection, $r$ is the recombination rate between two adjacent sites per generation, and $r^* = r(1-f)$ is the recombination rate in the presence of inbreeding, with inbreeding coefficient $f$. If $L$ is equated to an observed mean tract length $\hat{x}$, I have an estimate of $t = \frac{1}{r\hat{x}(1-f)}$.

To estimate the selection coefficient, I used the following formula for the expected nucleotide diversity ($\pi$) at linked neutral sites under recent selective sweep. The ratio of $\pi$ to the diversity before the sweep ($\pi_0$) is given by

$$\frac{\pi}{\pi_0} = 1 - e^{-\frac{2r^*x}{s}\ln(2N_e s)} \quad (2)$$

where $s$ is the selection intensity for mutant homozygotes and $N_e$ is the effective population size [1,39, 40–43]. It is clear from the formula that the substantial reduction is expected only if $2r^*x/s$ in the exponent is as small as 0.01 [1] or roughly $s = 200r^*x$. I note that this estimate is almost independent of $2N_e s$, unless $N_e$ is unlikely large.

### *2.2.9 Data deposition*

The nucleotide sequence data were deposited in the DDBJ databank. Their accession numbers for NGS sequences were DRA0049442 (http://trace.ddbj.nig.ac.jp/dra/submission_e.html#Data_release), PRJDB5051 (http://trace.ddbj.nig.ac.jp/bioproject/submission_e.html#Release_of_projects), and SAMD00056706-SAMD00056837 (http://trace.ddbj.nig.ac.jp/biosample/submission_e.html#Data_release).

The accession numbers for 1-kb sequences are LC194635–LC194778 in the DDBJ databank.

## 2.3 Results

### *2.3.1 Single origin of the Fm phenotype*

As Cemani and Silkie exhibit the same Fm phenotype (S1 Figure), the chromosomal rearrangement including duplicated *EDN3* was suspected to be the common genetic cause of Fm in both breeds [12,14]. Therefore, I first confirmed that the genomic rearrangement is indeed of single origin and common to Cemani and Silkie.

First, I studied genetic variation at *EDN3* in nine chicken breeds - Ayam Cemani ($n = 5$), Silkie ($n = 3$), Ayam Arab ($n = 2$), Ayam Kedu ($n = 5$), White Kedu ($n = 2$), Ayam Kalosi ($n = 2$), Ayam Kate ($n = 1$), Ayam Sentul ($n = 1$), and Kampung Chicken from Lombok ($n = 1$), and two jungle fowl populations, RJF ($n = 6$) and GJF ($n = 6$) (Table 2.1). To obtain unambiguously phased genomic sequence data for possibly four different *EDN3* genes in certain individuals, I analyzed DNA sequences of about 1 kb in length spanning exons 4 to 5. I selected this rather short fragment to avoid any complication due to intragenic recombination in inferring ancestral relationships among the DNA sequences. Indeed when I used 3-kb sequences for determining the *EDN3* phylogeny, I found strong evidence for intragenic recombination in multiple haplotypes, though not in those of Cemani and Silkie for an obvious reason (see NJ tree of region 4 in S2.4 Figure). The 1-kb sequences obtained from the 36 individuals in our sample can be classified into 12 haplotypes. Of these, six (*Hap*2', 5, 6, 8, 10, and 11) are restricted to the jungle fowls, three (*Hap*1, 3, and 9) are restricted to domesticated chickens, and the remaining haplotypes (*Hap*2, 4, and 7) occur in both domesticated chickens and jungle fowls (Table 2). RJF and GJF individuals exhibit a relatively large number of distinct haplotypes and maintain higher haplotype diversity than domesticated chickens. Importantly, however, no individual possesses more than two distinct haplotypes, indicating that individuals with *EDN3* duplication are highly inbred and homozygous. All eight Cemani and Silkie individuals possess only the *Hap*2/*Hap*4 haplotype combination (Table 2.2). This is in sharp contrast to the presence of the *Hap4* homozygous BKL2 and *Ha*p2/*Hap*10 heterozygous RJF9527 in RJF. The absence of segregation of *Hap*2 and *Hap*4 in Cemani and Silkie indicates that they are homozygous with respect to the single *Hap*2-*Hap*4 haplotype. In other words, *Hap*2 and *Hap*4 are no longer allelic in these

breeds. This observation strongly suggests that *EDN3* was duplicated by unequal crossing over, and the two resulting loci produced permanent heterozygosity for these alleles.

Curiously, four Kedu (KD3, KD16, KDH3, and KDH8) and some other (STC13 and LOM39) individuals also show the same *Hap*2/*Hap*4 haplotype combination as Cemani and Silkie. As the phenotypes of KDH3 and KDH8 are quite similar to that of Cemani (S1 Figure), I speculate that these Kedu individuals are heterozygotes, each possessing one *Fm* chromosome with duplicated *EDN3* and one wild-type chromosome with a single *EDN3*. This suggests interbreeding between Cemani and Kedu, and is based on the likelihood of the allele on the wild-type chromosome being either *Hap*2 or *Hap*4, in light of their high frequencies in Indonesian chicken breeds. On the other hand, KD3 and KD16 show apparent wild-type phenotypes for comb and face color (S1 Figure), suggesting that they possess two wild-type chromosomes with distinct *Hap*2 and *Hap*4 alleles. In any case, as other individuals show different haplotypes (Table 2.2), the Kedu population appears to be much more heterogeneous than Cemani and Silkie with respect to haplotypes and copy number of *EDN3* genes.

I tested whether Cemani and the other chicken breeds have the same duplicated regions of DR1 and DR2 as Silkie does. I amplified the regions from DNA of 56 individuals using two sets of specific primers [14] (A2 and B2 in Fig 2.1) (Table 2.3). The A2 primer set is designed to amplify a region that may contain either the boundary between inverted DR1 and DR2 (1RD-DR2) or that between inverted DR2 and DR1 (2RD-DR1) in a head-to-head configuration, whereas the B2 primer set is designed to amplify a region that contains either the boundary between DR1 and inverted DR2 (DR1-2RD) or that between DR2 and inverted DR1 (DR2-1RD) in a tail-to-tail configuration. The control primer sets A1 and B1 successfully amplified target sequences in all the samples, as reported previously [14]. Amplification with A2 and B2 was consistently

successful for 12 Cemani and 12 Silkie individuals and for seven Kedu samples. These findings indicate that the nucleotide sequences of 1RD-DR2/2RD-DR1 and DR1-2RD/DR2-1RD amplification products from Cemani are identical to those from Silkie (S2.2 Figure).

I confirmed that the *Hap*2/*Hap*4 combination in STC13 and LOM39 does not result from duplicated DR1, but stems from the segregation of the *Hap*2 and *Hap*4 alleles. However, the results of the A2 and B2 amplifications for the 24 Kedu individuals were somewhat puzzling. Amplification was successful for KDH3, KDH8, and KD16, but not for KD3, despite the fact that all four carry the *Hap*2/*Hap*4 combination. These observations for KDH3, KDH8, and KD3 agree with the aforementioned speculation that KDH3 and KDH8 have at least one *Fm* chromosome, while KD3 has two wild-type chromosomes. However, the result for KD16 is unexpected and suggests that, despite its wild-type phenotype, KD16 might possess at least one *Fm* chromosome. This speculation is supported by the presence of noticeable black pigment on the comb of KD16 (S1 Figure). This may also be true for KD17, KD19, KD21, and KD22 samples which exhibited successful A2 and B2 amplification (Table 2.3 and S1 Figure). This observation corroborates the high heterogeneity of *Fm* in the Kedu population. Although it is conceivable that the heterogeneity is related to Cemani breeding in the same geographic area of Central Java, further investigation of the genotype-phenotype relationship is required to draw any definitive conclusion. The high heterogeneity in the Kedu Fm phenotype also suggests that the *Dermal Melanin* inhibitor (*ID*) locus, on chromosome Z [44], is worth further investigation.

Second, I studied CNV in *EDN3* using qPCR. I measured the absolute concentrations of amplified *EDN3* and *UCKL1* amplicons in reaction mixtures of each sample and normalized the copy number of *EDN3* based on the single-copy gene of

*UCKL1*. Cemani and Silkie have twice to four times larger copy numbers than non-Fm chickens (Fig 2.2). Although the exact number of *EDN3* copies in Cemani and Silkie genome is difficult to estimate by the qPCR alone, the Fm phenotype surely shows excessive *EDN3* copies [12]. In addition, I carried out WGS for a single Cemani individual (Cemani 41). Using CNV-Seq [31], I confirmed that approximately twice as many reads were mapped onto the DR1 and DR2 in Cemani as compared to Taiwanese native chicken L2 with non-Fm phenotype (Fig 2.3a) and a similar result was obtained in Silkie with respect to L2 (Fig 2.3b) [23]. However, when the Cemani genome was compared with the Silkie genome, neither DR1 nor DR2 showed any excess of reads (Fig 2.3c). Together, these results consistently indicate that the DR1 and DR2 arrangement in Cemani is identical to that in Silkie and strongly support a single common origin of the Fm phenotype in Cemani and Silkie.

### *2.3.2 Haplotype diversity and duplication of* EDN3

To study the origin of *Hap*2 and *Hap*4 at duplicated *EDN3* loci, I examined the sequence differences among the 12 haplotypes or alleles in more detail. The haplotype sequences of the 36 individuals (Table 2.2) contain 35 variable sites consisting of one 1-bp deletion, two 3-bp deletions, and 28 point mutations (S2.3 Table). Of these haplotypes, five are singletons in the sample, whereas *Hap*2 and *Hap*4 are represented in 17 and 23 individuals, respectively. *Hap*2' is one of the singletons and differs from *Hap*2 by a single point mutation at position 784. As it occurs in RJF, it has likely been derived from *Hap*2 in the RJF population. Likewise, *Hap*9 differs from *Hap*4 by a single 3-bp deletion and descends in indigenous Ayam Arab ARG19. More importantly, *Hap*10 was found only in RJF and differs from *Hap*2 and *Hap*4 by one and two point mutations, respectively. Therefore, *Hap*10 likely is a common ancestor of *Hap*2 and *Hap*4. Thus, the

allelic divergence among *Hap*2, *Hap*4, and *Hap*10 must have occurred in the RJF population, which still harbors all these alleles.

To examine the phylogenetic relationship among the 12 haplotypes, I constructed an NJ tree [33] rooted by the orthologous quail sequence and statistically evaluated with 1000 bootstraps based on all nucleotide substitutions in 34 1-kb fragments derived from 29 individuals (Fig 2.4).

Although the tree showed several intermingling patterns of ancestral allelic lineages leading to RJF and domesticated chickens owing to incomplete lineage sorting, it did support that *Hap*10 is a recent common ancestor of *Hap*2 and *Hap*4. Next, I estimated the divergence time between *Hap*2 and *Hap*4 based on two calibrated substitution rates. One is based on the published substitution rate in introns [8,45]. When the rate of $(1.7-1.8) \times 10^{-9}$ per site per year is applied to the average per-site nucleotide differences between *Hap*2 and *Hap*4 ($0.0020 \pm 0.0014$), the divergence time of $0.6 \pm 0.4$ million years is obtained. Alternatively, I can directly calibrate the substitution rate in the present *EDN3* sequences using the divergence time of RJF/domesticated chickens from GJF. This divergence time can be inferred from such a geological event as the emergence of Java island 3–4 million years ago [46]. Further evidence from fossil records regarding the 4–5 million year-old ancestor of *Gallus* (*Gallus bravardi*) consistently suggests that GJF originated around 4 million years ago [47,48]. In this calibration, however, it has to be noted that four distinct haplotypes (*Hap*5, 7, 8, and 11) exist in GJF, of which *Hap*11 clustered together with *Hap*6 in RJFs (Aceh1 and Aceh4), and *Hap*7 is the same haplotype as that in the domesticated KDP7. Provided that GJF is indigenous to Java and the Lesser Sunda Islands, these *Hap*7 and *Hap*11 in GJF raise the possibility of recent introgression from RJF and/or domesticated chickens [8]. Therefore, I excluded *Hap*7 and *Hap*11 when calculating the average nucleotide difference per site ($0.0107 \pm$

$0.0025$) between GJF and other chickens, which resulted in the substitution rate of $1.3 \times 10^{-9}$ per site per year. This rate is a little slower than the previous one and yields a somewhat earlier estimate of the divergence time ($0.8 \pm 0.5$ million years) between *Hap*2 and *Hap*4. In either case, a rough estimate of divergence time (0.6–0.8 myr) implies that these *EDN3* alleles in fact originated in the ancestral RJF population. At some point after this allelic divergence, the *EDN3* locus was duplicated, and the Fm phenotype appeared. I will discuss a lower limit of this allelic divergence that can be set by the divergence time between the Cemani and Silkie breeds, together with a re-evaluation of the above estimates with large standard errors.

### *2.3.3 High- homozygosity tracts (HHTs) in Cemani and Silkie*

To detect any genomic signature of artificial selection on the Fm phenotype, I investigated the pattern and degree of DNA polymorphism in DR1 and its surrounding genomic regions. Using 9 Cemani, 10 Silkie, 11 other domesticated chickens including a single RJF, and two GJFs, I first examined nine regions of about 3 kb long and separated by ~200-kb intervals. As a whole, they span a 1.4-Mb genomic region that includes the 254-kb duplicated DR1s, 342-kb duplicated DR2s, and 413-kb spacer (Fig 2.5 and S2.3 Figure). Table 2.4 shows summary statistics of the genetic variability in these nine regions (see S2.4 Figure for NJ trees). First, the number of haplotypes ($H_k$) in a sample of $k$ chromosomes is generally much smaller than the number of segregating sites ($S_k$) within the same region. Each region is thus in fairly strong linkage disequilibrium and is consistent with relatively large values of $|D'|$ (not shown) or the squared correlation coefficient ($r^2$): the absolute value of $r$ is greater than 0.56 in all regions of all four populations. Second, the pattern and level of DNA polymorphism in Cemani and Silkie greatly differ from those in "Others" and GJF. I note that region 3 is located upstream of DR1, region 4 is within DR1, and regions 5–7 are within the spacer, while regions 1, 2, 8,

and 9 are further away from the *Fm* region. Compared with regions 1, 2, and 6–9, regions 3–5 in Cemani and Silkie show a remarkable reduction in $H_k$, $S_k$, Watterson's $\theta_w$ and nucleotide diversity $\pi$ [49-51]. For instance, in regions 1, 2, and 6–9, the average number of segregating sites per kb is about 12 in Cemani and Silkie. The expected number in each 3-kb region is thus about 30; however the actual number observed in regions 3 and 5 is 0 in Cemani and at most 2 in Silkie. I regarded this extremely low level of genetic variability as evidence for selective sweep via artificial selection for *Fm*. To verify this, I carried out the HKA test for Cemani and Silkie [34] using the divergence data in comparison with GJFs. The test indicated a significantly lower level of polymorphism in regions 3–5 than in any other region (S2.5 Figure). Additionally, I applied the STRUCTURE analysis region by region [35]. Although Cemani and Silkie individuals are generally assigned to different multiple genetic components in regions 1, 2, and 6–9, they are assigned to a single common component in regions 3–5 (*DDX27*, *EDN3*, and *TH1L*) (S2.6 Figure).

Among regions 3–5, region 4 within DR1 shows an exceptionally high level of genetic variability; however, this is deceptive and results from the inevitable mixture of duplicated sequences. The homologous sequences within duplicated DR1 cannot be amplified separately by the present PCR method; genetic variability in region 4 within DR1 is simply owing to a mixture of the two paralogous sequences. Despite this caveat, the level of genetic variability in region 4 is somewhat higher in Silkie than in Cemani. This is largely due to the inclusion of three sequences of Indonesian Silkie (KPS16, 17, and 30) and is also visible in the STRUCTURE analysis (S2.6 Figure). The nucleotide sites at positions, 1828, 2015, 2049, 2230, 2279, 2630, 2637, 2664, 2837, and 3005 in region 4 (ranging from 11,157,992 to 11,158,169 in the reference genome) are variable with respect to two distinct haplotypes. One haplotype is identical to that in Silkie and the

other is identical to that in indigenous Indonesian chicken breed KAL28 (see NJ tree of region 4 in S2.4 Figure), suggesting frequent occurrence of interbreeding between Indonesian Silkie and other indigenous domesticated chickens.

I aimed to determine whether the regions with reduced genetic variability or the HHTs are identical for Cemani and Silkie. For practical purposes, I operationally defined an HHT as a consecutive genomic region over 10 kb with $\pi < 10^{-4}$ or $< 2\%$ of the normal level $\pi_0 = 0.005$ (Table 2.4). I determined the WGS of one Cemani individual and compared it with those of Silkie and Taiwanese L2. Cemani and Silkie exhibit a similar extent of reduction in DR1 and its surrounding regions, but the HHT length differed greatly between these two breeds (Fig 2.6). The Cemani HHT is long and extends toward regions 2 and 7, whereas the Silkie HHT is relatively short and limited from a little beyond regions 3 to 5. The left and right HHT lengths are, respectively, 118 and 387 kb in Cemani and 52 and 101 kb in Silkie, respectively. However, it is highly probable that such a tract can differ from individual to individual even within a breed. Therefore, using the same samples of 19 Cemani and Silkie in Table 2.4, I further examined the genetic variability surrounding the boundaries in nine windows, each about 1 kb each in length (Fig 2.6). The windows are localized into three clusters; the left HHT in Cemani still extends beyond the central cluster, while that of Silkie already ends there. Likewise, the right HHT in Cemani terminates just within the right cluster, whereas that in Silkie ends well within in this cluster. I measured the left and right HHT from the 5' and 3' ends, respectively, of *EDN3* (11,144,657–11,161,475) and estimated the tract lengths. The minimum left and right HHT lengths are 118 and 224 kb, respectively, in Cemani and 52 and 101 kb, respectively, in Silkie (Fig 2.6).

## 2.4 Discussion

### *2.4.1 Genomic configuration of the* Fm *region*

Dorshorst et al. [14] proposed three possible rearrangements (FM1–FM3) of duplicated DR1s and DR2s in the *Fm* region (Fig 2.1). Although all these rearrangements possess the same boundaries of 1RD-DR2/2RD-DR1 and DR1-2RD/DR2-1RD, one major difference exists with respect to the relative position of the 413-kb spacer. In models FM1 and FM3, duplicated DR1s sandwich the spacer. If either FM1 or FM3 is valid, the HHT is expected to cover the entire spacer as the two *EDN3* loci in DR1s are simultaneous targets of artificial selection. In contrast, in model FM2, the spacer is located outside the duplicated DR1s, and can therefore recombine with wild-type chromosomes without disrupting the Fm phenotype. In this case, the spacer region is expected to be polymorphic because of recombination. Our data (Figs 2.5, 2.6, and Table 2.4) clearly showed that the patterns and degrees of polymorphism exhibited by Cemani and Silkie are consistent with FM2, but inconsistent with FM1 and FM3.

### *2.4.2 DR1 duplication and emergence of the* **Fm** *phenotype*

I estimated an upper limit of *EDN3* duplication time of 0.6 ± 0.4 ~ 0.8 ± 0.5 million years based on the allelic divergence between *Hap*2 and *Hap*4. Although the standard error is large because of the usage of the short sequences, *Hap*2 and *Hap*4 seem to diverge from each other much earlier than is documented in any known archeological record of domesticated chickens. As mentioned earlier, *EDN3* duplication and the Fm phenotype emerged in the ancestral RJF population of chickens; therefore, this phenotypic variation was highly likely to be selected once domestication began in Asia. Xiang et al. [52] dated ancient mtDNA sequences from the earliest archaeological chicken bones in China back to 10,000 years ago.

The analysis of the Cemani and Silkie genome sequences revealed that the 71.4-kb region spanning nt 11,183,600 to 11,255,000 is located within the joint set of the right HHTs in both breeds (Fig 2.6 and S2.7 Figure). In this region, recombination has been

apparently prohibited by artificial selection on *EDN3* and therefore, the two breeds are most closely related to each other in terms of nucleotide substitutions (S2.7 Figure). Because of the paralogy between DR1s, 50 variable sites in Cemani and 51 in Silkie are observed within a stretch of an approximately 55 kb of 71.4-kb region. It is important to note that a great majority (49) of these variable sites are shared between the two breeds, implying that they accumulated in their common ancestor (S2.4 Table). As the per-site differences amount to approximately $(9.2 \pm 1.3) \times 10^{-4}$, I can estimate the sequence divergence time between the duplicated DR1s as $0.26 \pm 0.04 \sim 0.35 \pm 0.05$ million years ago. These are more recent, but more reliable than the previous estimates for the upper limit of DR1/*EDN3* duplication time. In either case, I conclude that the Fm phenotype caused by duplicated DR1/*EDN3* originated in RJF long before the domestication process began in China.

Additionally, I am interested in the divergence time between Cemani and Silkie, to use it as a lower limit for the DR1 duplication time. For this purpose, I used breed-specific substitutions. Provided that recombination is rare or absent within the 71.4-kb region, I treated such substitutions as derived variants and proportional to the divergence time between Cemani and Silkie breeds. In the entire region of $2 \times 55 + 16 = 126$ kb, there are one Cemani-specific and two Silkie-specific nucleotide substitutions (S2.4 Table). The mean per-site sequence differences are therefore given by $d = \frac{3}{126000} = 2.4 \times 10^{-5}$ for a pair of Cemani and Silkie genomes. Using the calibrated nucleotide substitution rate of $(1.3 \sim 1.8) \times 10^{-9}$, I obtain the divergence time of $6600 \sim 9100$ years and regard it as an upper limit of the divergence time between the Cemani and Silkie breeds. Because this also gives a lower limit of the DR1/*EDN3* duplication time, the estimate suggests that the Fm phenotype emerged by this time (Fig 2.7).

### *2.4.3 Ayam Cemani and Ayam Kedu in Indonesia*

The extent of nucleotide diversity in Cemani is almost the same as that in other domesticated chickens and jungle fowl, except near the *EDN3* locus (regions 3, 4, and 5), despite the fact that Cemani is a local breed of Kedu, in Indonesia. This is in sharp contrast to Silkie which I could be sampled in Indonesia, Japan, and the USA in the present study. There are two possible explanations for this high genetic variability in Cemani: a relatively large founding population, and frequent genetic exchanges with other domesticated chickens or jungle fowls. The presence of KDH3 and KDH8, which are heterozygous for the *Fm* and wild-type chromosomes, supports the latter hypothesis of interbreeding of Cemani with other domesticated chickens. In this case, there must have been intense selection to maintain the Fm phenotype. However, the effect of intense selection can be limited in genomic regions closely linked to a target site. Further study of Ayam Kedu with abundant *Fm* variation will provide useful information on their breeding schemes and the history of the Fm phenotype in Indonesia.

### *2.4.4 History and strength of artificial selection*

I used information on HHTs in two simple ways [53] without any sophistication for inference [54-56]. One is to infer the allele age or the history of artificial selection based on the idea underlying the inference of adaptively introgressed tracts [38,39], and the other is to infer the strength of artificial selection, as has been done for maize domestication [1]. Both estimates depend heavily on the recombination rate *r* per bp per generation. The recombination rate is known to vary considerably among as well as within chromosomes [7]. The rate is 3.7 cM/Mb if averaged over the entire chicken genome, whereas it is approximately 3.0–5.0 cM/Mb for chromosome 20. I assume here *r* = 3.0 cM/Mb = $3.0 \times 10^{-8}$ per bp per generation. When considering effect of inbreeding

in the domestication process, $r$ is replaced with the effective recombination rate $r^* = r(1-f)$ in which $f$ is the inbreeding coefficient [39].

As mentioned previously, I measured the minimum lengths of the left and right HHTs in sequences of Cemani or Silkie individuals. Based on both the population and NGS data, the left and right HHT lengths are 118 kb and 224 kb, respectively, in Cemani, and 52 kb and 101 kb, respectively, in Silkie. Using formula (1) [37, 38], with an observed mean tract length $\hat{x}$, I calculated the number of generations elapsed under artificial selection (see Materials and Methods). In the case of Cemani, $\hat{x} = \frac{118+224}{2} = 171$ kb so that $t \approx 200/(1-f)$, whereas in the case of Silkie, $\hat{x} = \frac{52+101}{2} = 77$ kb so that $t \approx 440/(1-f)$. It thus appears that the history of Cemani is approximately half of that of Silkie. In the absence of inbreeding, the tract erodes quickly, but even intense inbreeding such as full-sib mating, with $f = 1/4$, can increase the time only by 30%. Furthermore, if I define the generation time of chickens and fowls based on the mean age ($m$) of hens at a given time [39], I can convert the above $t$ generations into $m \times t$ years. If chickens can lay eggs for 7 years (with the age at first reproduction being 1 year and the mean longevity being approximately 15 years), it might be reasonable to assume $m$ = 3–4. Therefore, it appears that Cemani and Silkie have been bred for roughly $(600 \sim 800)/(1-f)$ years and $(1300 \sim 1700)/(1-f)$ years, respectively. Thus, the history of Indonesian Ayam Cemani appears to be rather short, whereas the relatively long history of Silkie is consistent with the relatively short HHT length in its *Fm* region.

Second, I used formula (2) for the expected nucleotide diversity $\pi$ at linked neutral sites under recent selective sweep with selection coefficient $s$ (see Materials and Methods). In the virtual absence of variation, I can have such a rough relationship as $s = 200r^*x$. With $r = 3 \times 10^{-8}$ [7,57] and $x \geq 100$ kb, I have $s \geq 0.6(1-f)$. This is

inevitably a crude estimate but it indeed suggests intense artificial selection in both the Cemani and Silkie breeds.

As a final caveat, it may be asked why the tract boundaries in the NGS data separate HHT very sharply from the neutral level (Fig 2.5). The two *Fm* chromosomes in an individual must be flanked by wild-type chromosomes and likely have different recombination breakpoints. However, the two *Fm* chromosomes have virtually identical DNA sequences in the focal site and nearby linked sites, but are different from wild-type chromosomes, which might also differ from each other. Therefore, I can identify only sharp HHT boundaries in a single diploid *Fm* individual. However, as these abrupt boundaries can differ among individuals, the tract boundaries might become more gradual for large samples or at the population level. This would explain intermediate values of $\pi$ observed in regions 1, 2 and 6 as well as in the right HHT in a sample of nine Cemani individuals (Table 2.4, Fig 2.6).

## 2.5 Conclusions

I showed that fibromelanosis (Fm) in Indonesian Ayam Cemani and Chinese Silkie resulted from the same genetic change involving *EDN3* duplication on chromosome 20. This genomic change of a single origin arose spontaneously in the ancestral population of RJF in Asia, probably well before the first domestication of chickens. Strong artificial selection for the Fm phenotype is evident in the genetic variability near the target site of duplicated *EDN3*, although the pattern and level of variability differ sensitively between these two breeds, which have undergone different domestication processes.

## 2.6 References


1. Wang RL, Stec A, Hey J, Lukens L, Doebley J. The limits of selection during maize domestication. Nature. 1999; 398: 236-239.
2. Cong B, Barrero LS, Tanksley SD. Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. Nat Genet. 2008; 40: 800-804.
3. Bovine HapMap Consortium. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 2009; 324: 528-532.
4. vonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, et al. Genome-wide SNP haplotype analyses reveal a rich history underlying dog domestication. Nature. 2010; 464: 898-902.
5. Amaral AJ, Ferretti L, Megens HJ, Crooijmans RPMA, Nie H, Ramos-Onsins SE, et al. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PLoS ONE. 2011; 4: e14782.
6. International Chicken Polymorphism Map Consortium. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature. 2004; 432: 717-722.
7. Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens HJ, Crooijmans RPMA, et al. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res. 2009; 19: 510-519.
8. Sawai H, Kim HL, Kuno K, Suzuki S, Gotoh H, Takada M, et al. The origin and genetic variation of domestic chickens with special reference to junglefowls *Gallus g. gallus* and *G. varius*. PLoS ONE. 2010; 5: e10639.
9. Rubin CJ, Zody MC, Erikson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010; 464: 587-591.
10. Johnsson M, Gering E, Willis P, Lopez S, Van Dorp L, Hellenthal G, Henriksen R, et al. Feralisation targets different genomic loci to domestication in the chicken. Nature Comm. 2016; 7: 12950.
11. Hutt FB. Genetics of the Fowl. New York: McGraw-Hill; 1949.
12. Shinomiya A, Kayashima Y, Kinoshita K, Mizutani M, Namikawa T, Matsuda Y, et al. Gene duplication of the endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (*Fibromelanosis*) in Silkie chickens. Genetics. 2011; 111.136705. E-pub 2011.Nov. 30.
13. Dorshorst B, Okimoto R, Ashwell C. Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the Silkie chicken. J Hered. 2010; 101: 339-350.
14. Dorshorst B, Molin AM, Rubin CJ, Johansson AM, Strömstedt L, Pham M-H, et al. A complex genomic rearrangement involving the *Endothelin 3* locus causes dermal hyperpigmentation in the chicken. PLoS Genet. 2011; 7: e1002412.
15. Saldana-Caboverde A, Kos L. Roles of endothelin signaling in melanocyte development and melanoma. Pigment Cell Melanoma Res. 2010; 23: 160-170.
16. Lahav R, Ziller C, Dupin E, Le Douarin NM. Endothelin 3 promotes neural crest cell proliferation and mediates a vast increase in melanocyte number in culture. Proc Natl Acad Sci U S A. 1996; 93: 3892-3897.
17. Lahav R, Lecoin L, Ziller C, Nataf V, Camahan JF, Martin FH, et al. Effect of the Steel gene product on melanogenesis in avian neural cell cultures. Differentiation. 1998; 58: 133-139.

18. Dupin E, Douarin NML. Development of melanocyte precursors from the vertebrate neural crest. Oncogene. 2003 22: 3016-3023.
19. Akiyama T, Shinomiya A. Overview on the melanocyte precursor migration from the neural crest. In: Smith J, Haworth M, editors. Skin Pigmentation. New York: Nova Science Publishers, Inc.; 2013, pp. 175-196.
20. Smyth JR Jr. Melanin pigmentation: Its Biological roles, Inheritance and Expression in the Chicken. University of Mass Amherst; 1994.
21. Kaelin CB, Xu X, Hong LZ, David VA, McGowan KA, Schmidt-Küntzel A, et al. Specifying and sustaining pigmentation patterns in domestic and wild cats. Science. 2012; 337: 1536-1541.
22. Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet. 2014; 10: e1004148.
23. Fan W-L, Ng CS, Chen C-F, Lu M-YJ, Chen Y-H, Liu C-J, et al. Genome-wide patterns of genetic variation in two domestic chickens. Genome Biol Evol. 2013; 5: 1376-1392.
24. Birnboim HC. A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol. 1983; 100: 243-255.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114-2120.
26. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM; 2013. Avilable: *arXiv:1303.3997v2*, [q-bio.GN]
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25: 2078-2079.
28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20: 1297-1303
29. DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43: 491-498
30. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinformatics 2013; 43: 11.10.1-11.10.33.
31. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics. 2009; 10: 80.
32. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30: 2725-2729.
33. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4: 406-425.
34. Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data. Genetics. 1987; 116: 153-159.
35. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25: 1451-1452.
36. Prichard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945-959.
37. Liang M, Nielsen R. The length of admixture tracts. Genetics. 2014; 197: 953-967.
38. Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 2015; 16: 359-371.
39. Charlesworth B, Charlesworth D. Elements of Evolutionary Genetics. Colorado: Roberts and Company Publishers; 2012.

40. Maynard Smith J, Haigh J. The hitch-hiking effect of a favorable gene. Genet Res. 1974; 23: 23-35.
41. Kaplan NL, Hudson RR, Langley CH. The “hitch-hiking” effect revisited. Genetics. 1989; 123: 887-899.
42. Barton NH. The effect of hitch-hiking on neutral genealogies. Genet Res. 1998; 72: 123-133.
43. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Res. 2005; 15: 1566-1575.
44. Bateson W, Punnett RC. The inheritance of the peculiar pigmentation of the silky fowl. J Genet. 1911; 1: 185-203.
45. Axelsson E, Smith NG, Sundstrom H, Berlin S, Ellengren H. Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. Mol Biol Evol. 2004; 21: 1538-1547.
46. Bellwood PS. Man’s conquest of the Pacific: The Prehistory of Southeast Asia and Oceania. Oxford University Press; 1979.
47. Lambrecht K. Handbuch der Paläornithologie. Berlin: Verlag Grebrüder, Borntaeger; 1933.
48. Brodkorb P. Catalogue of fossil birds. Bull Florida State Mus. 1964; 8: 195-335.
49. Ewens WJ. The sampling theory of selectively neutral alleles. Theor Popul Biol. 1972; 3: 87-112.
50. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975; 7: 256-276.
51. Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press; 1987.
52. Xiang H, Gao J, Yu B, Zhou H, Cai D, Zhang Y, et al. Early Holocene chicken domestication in northern China. Proc Natl Acad Sci U S A. 2014; 111: 17564-17569.
53. Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, et al. Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet. 1998; 62: 1507-1515.
54. Slatkin M. A Bayesian method for jointly estimating allele age and selection intensity. Genet Res. 2008; 90: 119-128.
55. Chen H, Slatkin M. Inferring selection intensity and allele age from multilocus haplotype structure. Genes Genomes Genet. 2013; 3: 1429-1442.
56. Chen H, Hey J, Slatkin M. A hidden Markov model for investigating recent positive selection through haplotype structure. Theor Popul Biol. 2015; 99: 18-30.
57. Mugal CF, Nabholz B, Ellegren H. Genome-wide analysis in chicken reveals that local levels of genetic diversity are mainly governed by the rate of recombination. BMC Genomics. 2013; 14: 86.

## 2.7 Figures and Table

**Fig 2.1**. **Three possible arrangements of duplicated DR1s and DR2s in the *Fm* region**. Duplication of DR1 and DR2 is absent in the wild-type chromosome (top bar). A2 and B2 primer sets are designed for detecting the duplication boundaries between DR1 and DR2; A1 and B1 are for amplification control. This figure is modified from Dorshorst et al. [14]

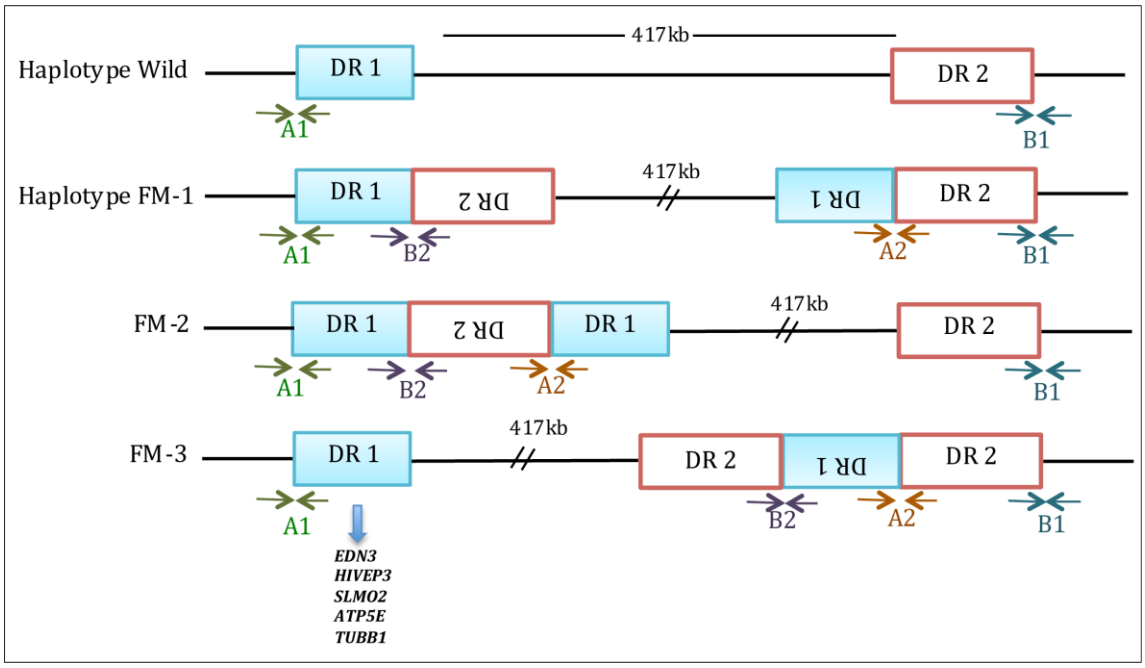


**Fig 2.2**. ***EDN3* CNV.** Copy numbers of *EDN3* were normalized to those of *UCKL1* in qPCR**.** Red bars represent the copy numbers of three Silkie and five Cemani individuals, and blue bars represent those of eight wild-type chickens. The average copy number in *Fm*-phenotype chickens is 3.39 ± 0.44 and that in wild types is 1.15 ± 0.09 ($P = 1.50 \times 10^{-6}$).

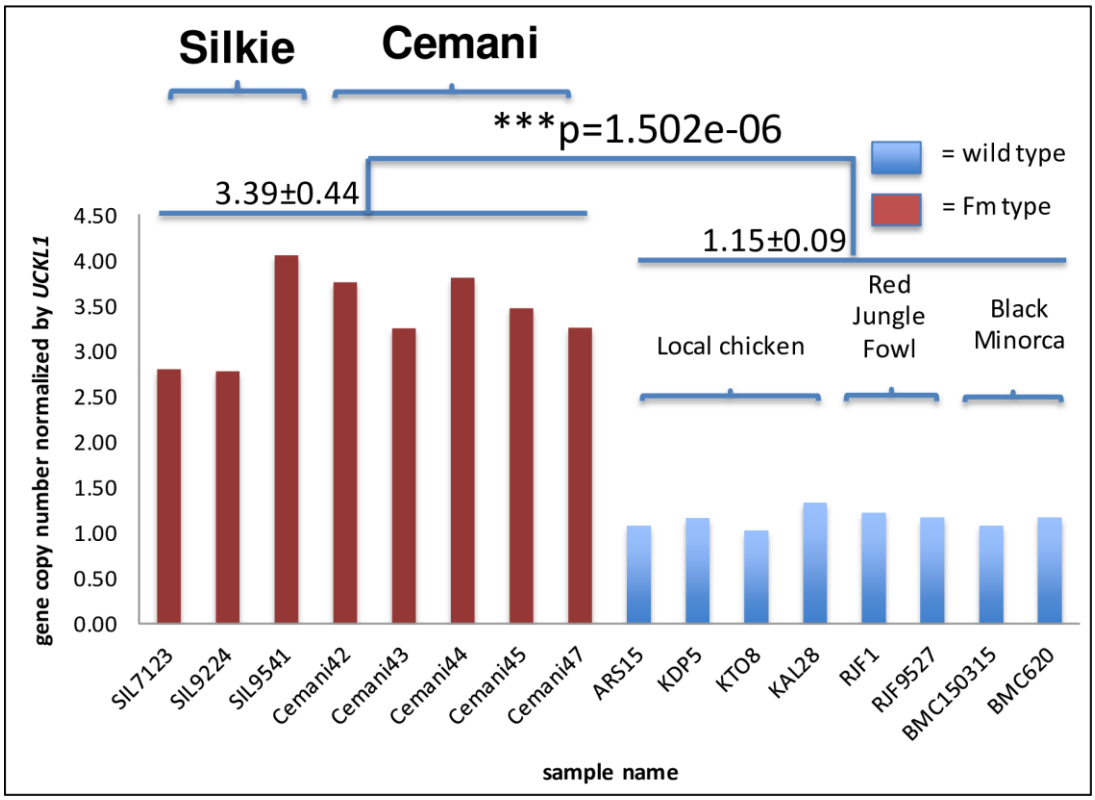


**Fig 2.3**. $Log_2$ **ratio for CNV in chicken chromosome 20 between nt positions 10,700,000 and 12,000,000.** (a) Comparison of read mapping between Cemani and Taiwanese L2. A blow-up of the DR1 region containing *EDN3* is shown below the map. Comparison of read mapping between combinations of (b) Silkie and Taiwanese L2 and (c) Cemani and Silkie.

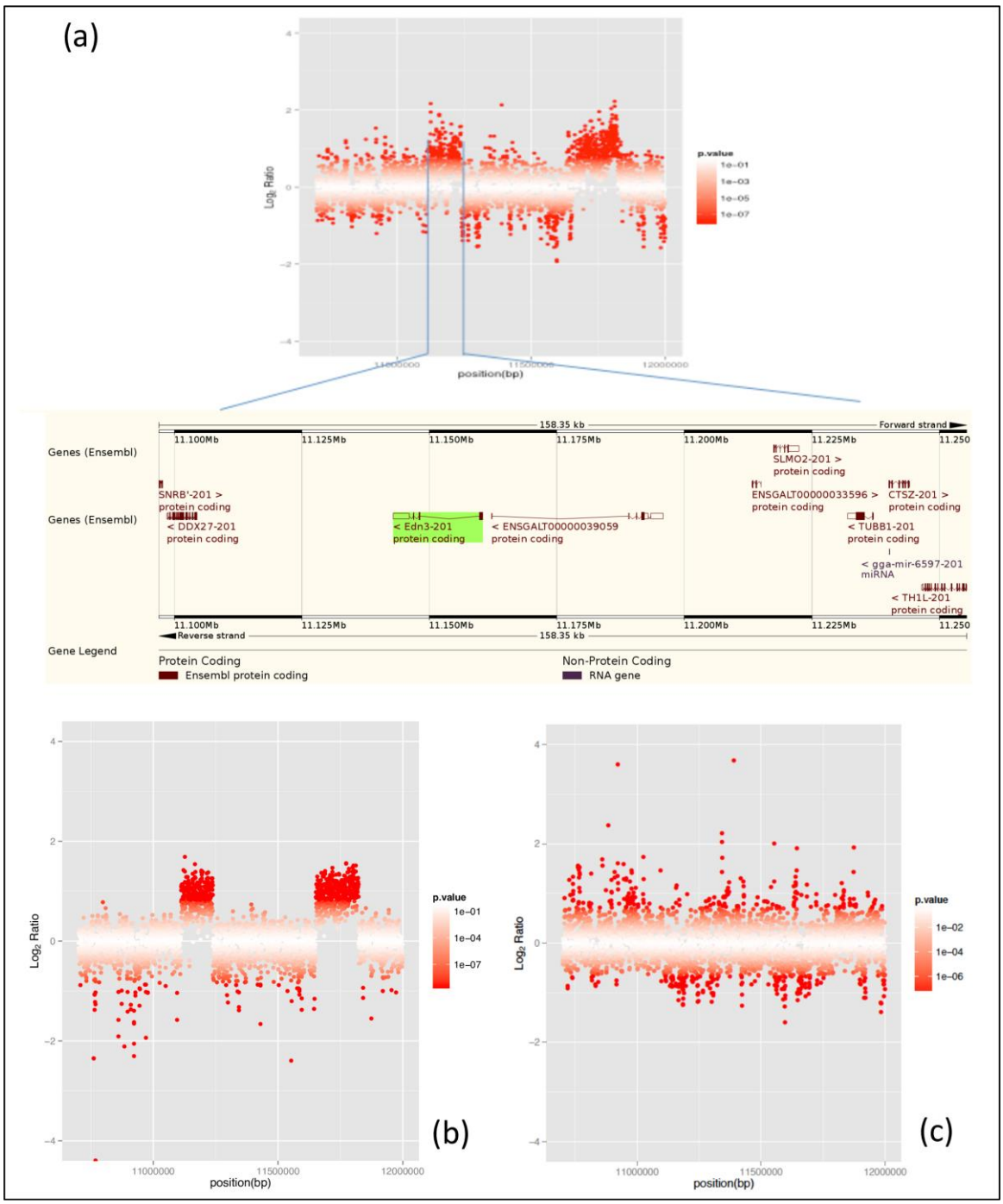


**Fig 2.4. NJ tree of *EDN3* haplotypes rooted by the quail sequence (accession number NC_029535)**. The tree was constructed with 1000 bootstrap resampling with an option of complete deletion of gaps/missing nucleotides [33]. The nucleotide divergence was measured by using the number of nucleotide differences per site, without multiple-hit correction.

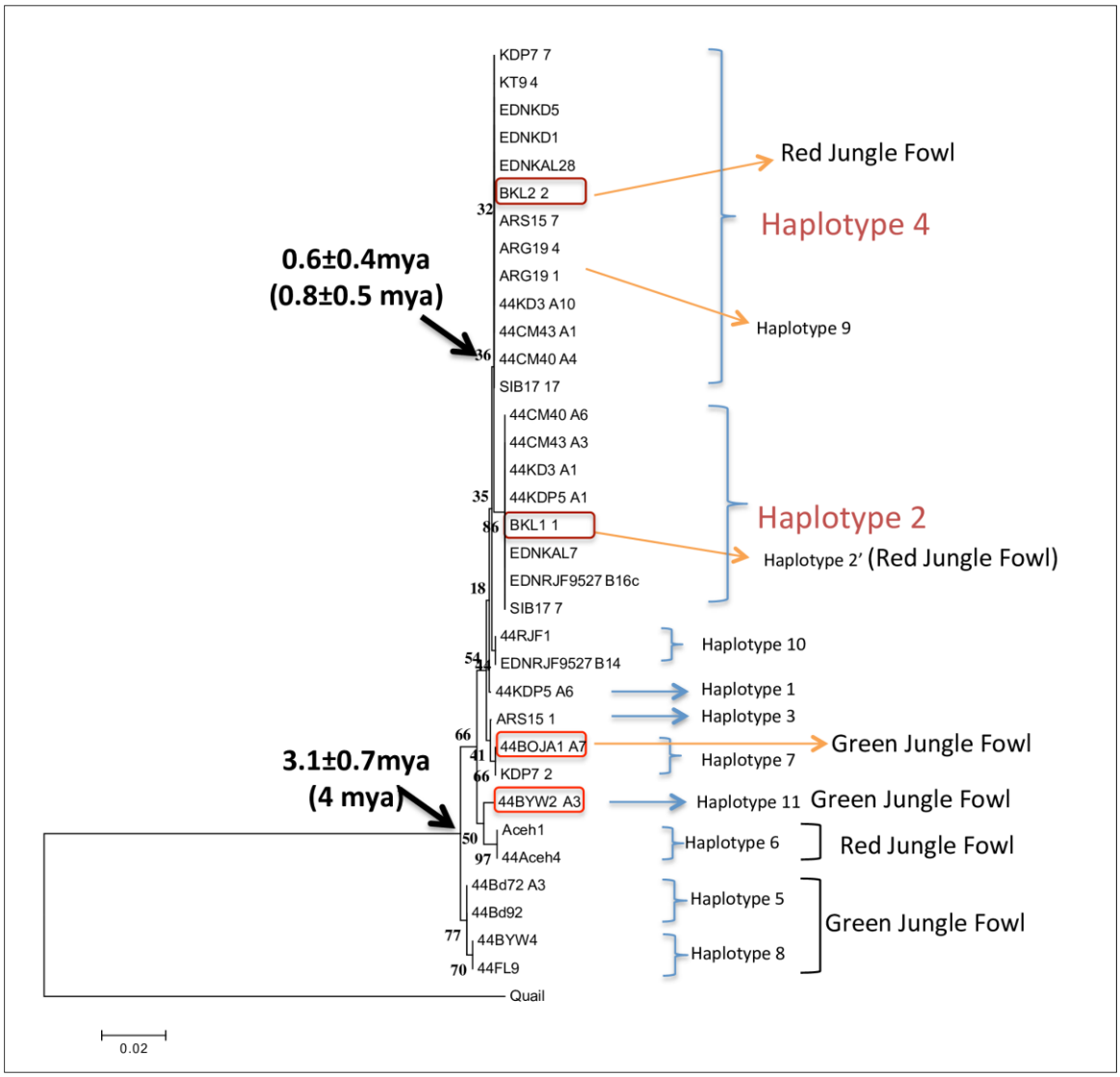


**Fig 2.5**. **The nucleotide diversity ($\pi$) in nine regions surrounding *END3* in chicken breeds**. The $\pi$ values in Cemani (green), Silkie (purple), other domesticated breeds (red), and GJF (blue) are shown under a schematic diagram of their locations together with duplicated DR1 and DR2 regions on chromosome 20. Each $\pi$ is measured in a 1 kb window with an overlapping sliding size of 100 bp. All regions except 7 and 8 are located within genes whose exon and intron structures are indicated below diversity plots.

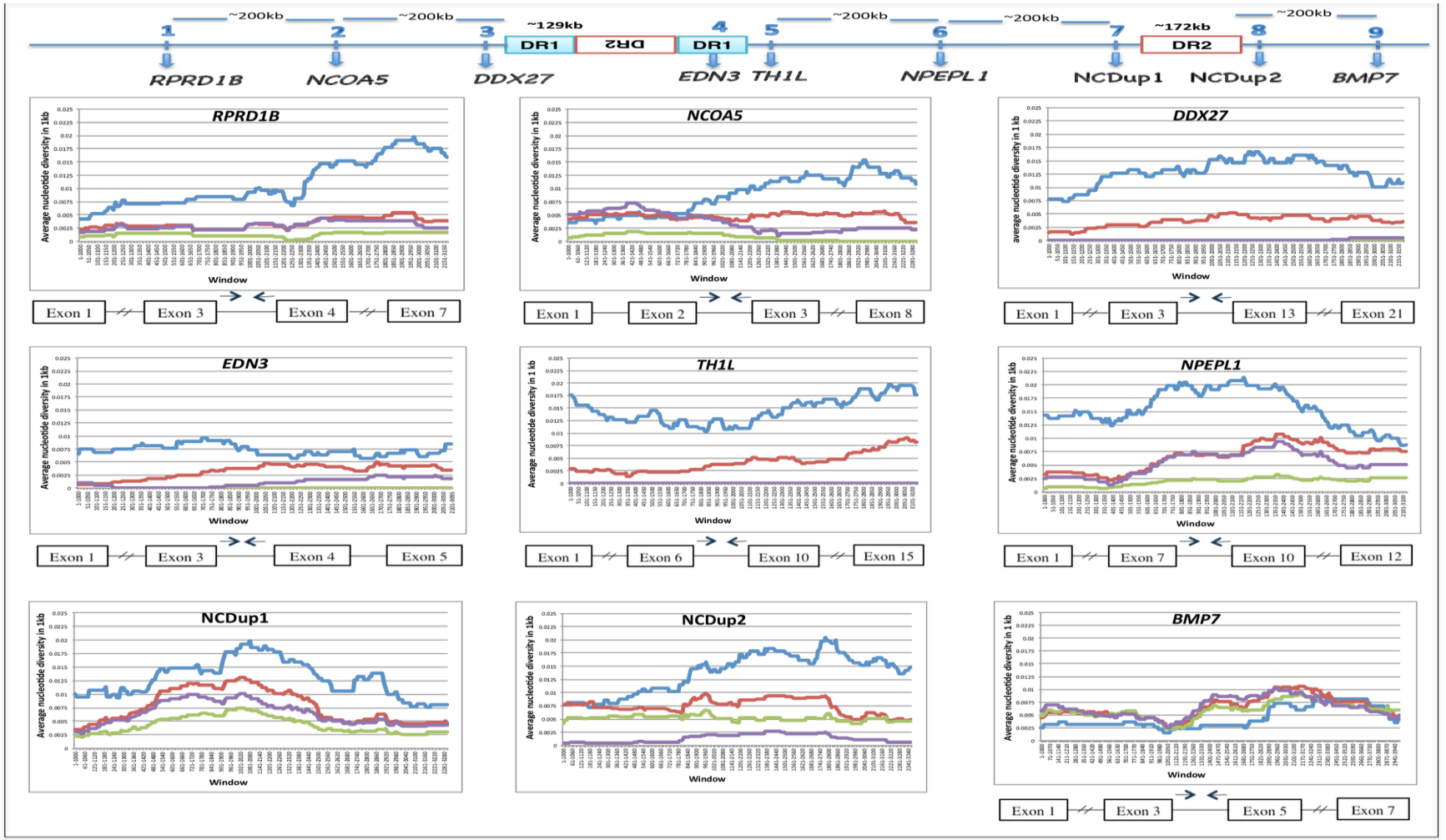


**Fig 2.6**. **Nucleotide diversity ($\pi$) based on NGS genotype data for one Cemani (red) and one Silkie (green) individual.** Diversity is measured in 10-kb non-overlapping windows. The left-shaded region represents DR1, in which nearly the same patterns and degrees of polymorphism are exhibited by Cemani and Silkie. This supports that DR1 was duplicated prior to their diversification and has since been frozen from recombination in both breeds. The same trend is observed in the right-shaded DR2; however, the ancestral haplotype is obscured by recombination. The three upper panels show the proportion of per-SNP heterozygotes at the population level for Cemani (red) and Silkie (green). Observations are made in 9 windows, each 1 kb long. The 9 windows are grouped into sets of 3, 2, and 4 windows. The three windows at the left are consecutive and contain 16 SNPs in total. The two windows in the middle and four at the right are separated by 16–40 kb and contain 22 and 19 SNPs, respectively

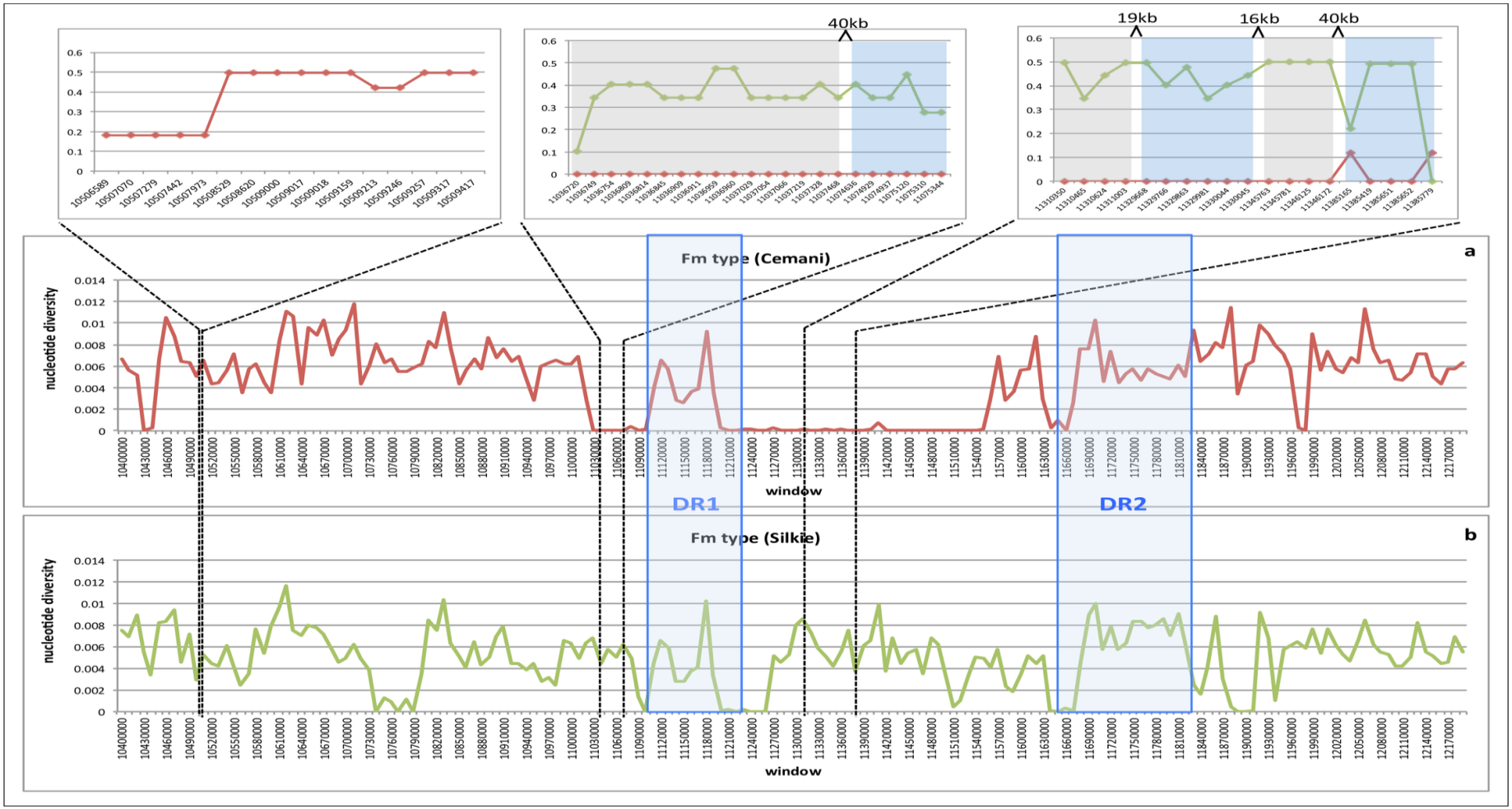



**Fig 2.6. Nucleotide diversity ($\pi$) based on NGS genotype data for one Cemani (red) and one Silkie (green) individual.**

**Fig 2.7**. **Evolutionary history of *EDN3* genes in Cemani and Silkie.** The green lines represent the allelic divergence of *Hap*2 and *Hap*4 at *EDN3* in the ancestral RJF population. The blue lines represent the ancestral lineages in the 71.4-kb region (shadow), from which the divergence time between Cemani and Silkie was estimated. The DR1 duplication is placed somewhere between 10,000 and 300,000 years ago. The horizontal red line corresponds to the beginning of the domestication process in China [52].

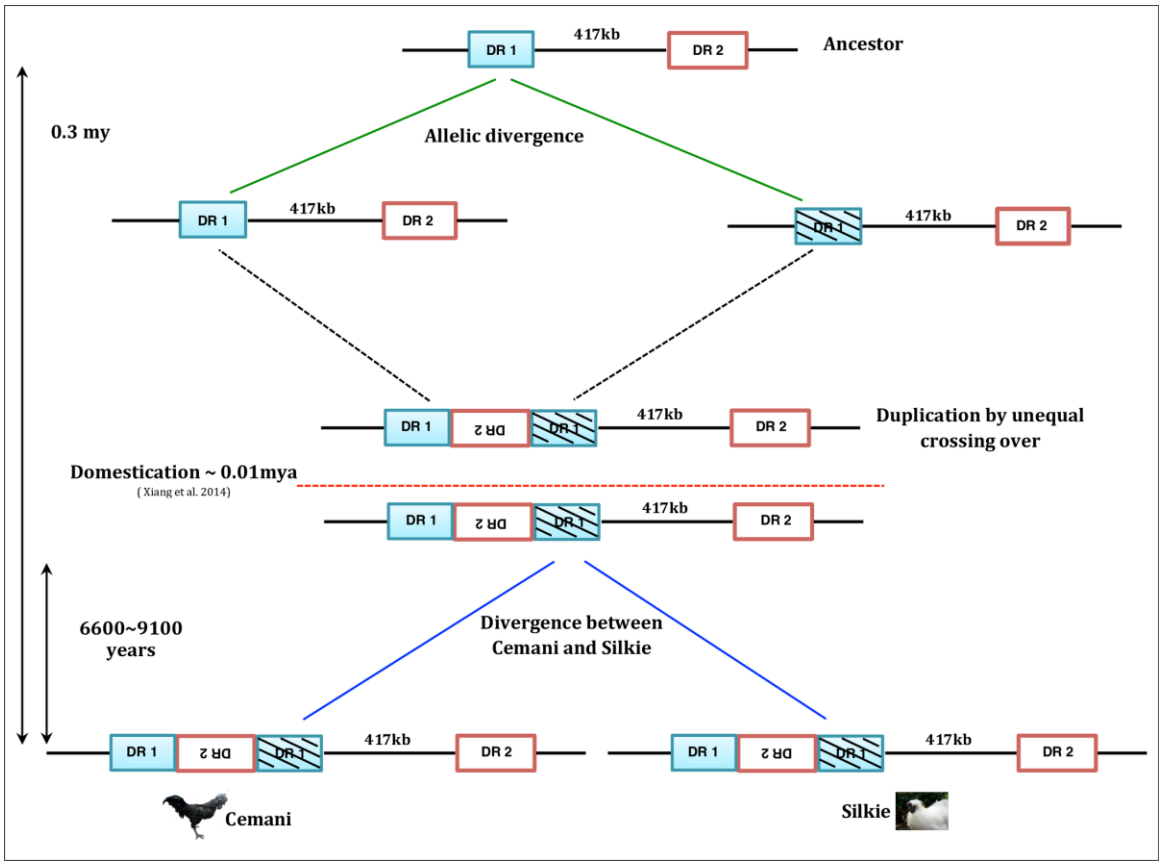


**Table 2.1.** IDs of 75 sampled domesticated chickens and jungle fowl together with their collection sites and sample sizes.

| No | Chicken breed | Sample ID | Collection site | Samples |
|---|---|---|---|---|
| 1 | Ayam Cemani | Cemani 40–47[a], CM (1, 6, 11, 23, 31)[a] | Kedu, Central Java, Indonesia | 13 |
| | | Cemani[b] | Nagoya, Japan | 1 |
| 2 | Silkie | SIB (2, 6, 7, 11, 14, 15, 16, 17)[c] | Murray McMurray Hatchery, Iowa, USA | 8 |
| | | WS (3741,BS3846)[b], SIL (7123, 7124, 9541)[b] | Japan | 5 |
| | | KPS (16, 17, 30)[a] | BPTU-Sembawa, Palembang-Sumatera | 3 |
| 3 | Ayam Arab (Silver) | ARS15[a] | BPTU Ayam, Sembawa, South | 1 |
| | (Golden) | ARG19[a] | Sumatera, Indonesia | 1 |
| 4 | Ayam Kedu (Hitam) | KD (1–5, 7–17,19, 21, 22)[a], KDH (3,8)[a] | Kedu, Temanggung, Central Java, Indonesia | 21 |
| 5 | White Kedu (Ayam Kedu Putih) | KDP (1, 5, 7)[a] | Kedu, Temanggung, Central Java, Indonesia | 3 |
| 6 | Ayam Kalosi | KAL (28, 7, 2) [a] | Gowa, South Sulawesi, Indonesia | 3 |
| 7 | Ayam Kate | KT9 [a] | Yogyakarta, DIY, Indonesia | 1 |
| 8 | Ayam Sentul | STC13 [a] | Sentul, West Java, Indonesia | 1 |
| 9 | Kampung Chicken | LOM39 [a] | Lombok, Indonesia | 1 |
| 10 | Black Minorca | BMC (610, 613)[b] | Japan | 2 |
| 11 | Red Jungle Fowl (RJF)* | BKL1[a], BKL2[a] | Bengkulu, Indonesia | 2 |
| | | Aceh (1, 4)[a] | Nangroe Aceh Darusalam, Indonesia | 2 |
| | | RJF (1, 9527)[b] | Nagoya, Japan | 2 |
| 12 | Green Jungle Fowl(GJF) | BYW (2, 4)[a] | Banyuwangi, East Java, Indonesia | 2 |
| | | BOJA1[a] | Boja, Kendal, Indonesia | 1 |
| | | Bd (72, 92)[a] | Sumbawa, West Nusa Tenggara | 2 |
| | | FL9[a] | Flores, East Nusa Tenggara, Indonesia | 1 |

* BKL1, BKL2, Aceh1 and Aceh4 are *Gallus gallus spadiceus*, while RJF1 and RJF9527 are *Gallus gallus* with unknown subspecies name: [a] Chicken breeds and genomic DNAs acquired from the MZB, LIPI in Indonesia: [b] Genomic DNA samples supplied from Keio University via ABRC of Nagoya University in Japan: [c] Genomic DNA samples provided through the JSAC.

**Table 2.2. *EDN3* haplotypes of 36 individuals**.

| | Populations[c] | | | | | |
|---|---|---|---|---|---|---|
| Haplotypes per individual[c] | GJF ($n^a = 6$) | RJF ($n = 6$) | Cemani ($n = 5$) | Silkie ($n = 3$) | Kedu ($n = 9$) | Others[b] ($n = 7$) |
| *Hap*2 | | | | | | KAL7 |
| *Hap*2' | | BKL1 | | | | |
| *Hap*4 | | BKL2 | | | KD1, 2, 5 | KT9 |
| | | | | | | KAL28 |
| *Hap*5 | Bd72, 92 | | | | | |
| *Hap*6 | | Aceh1, 4 | | | | |
| *Hap*7 | BOJA1 | | | | | |
| *Hap*8 | BYW5 | | | | | |
| | FL9 | | | | | |
| *Hap*10 | | RJF1 | | | | |
| *Hap*11 | BYW2 | | | | | |
| *Hap*1/*Hap*2 | | | | | KDP5 | |
| *Hap*2/*Hap*4 | | | CM1, 6, 31 | SIB2, 17 | KD3, 16 | STC13 |
| | | | Cemani40, 43 | WS3741 | KDH3, 8 | LOM39 |
| *Hap*2/*Hap*10 | | RJF9527 | | | | |
| *Hap*3/*Hap*4 | | | | | | ARS15 |
| *Hap*4/*Hap*7 | | | | | KDP7 | |
| *Hap*4/*Hap*9 | | | | | | ARG19 |

[a] *n* is the number of sampled individuals in each population : [b] "Others" consists of Ayam Sentul (*n* = 1), Ayam Lombok (*n* = 1), Ayam Arab (*n* = 2), Ayam Kate (*n* = 1), and Ayam Kalosi (*n* = 2) : [c] see Table 2.1 for sampled individuals in each population and S2.3 Table for the haplotype definition.

**Table 2.3. PCR amplification of the duplication boundaries between DR1 and DR2 in the *Fm* region**.

| Samples (no. of individuals) | Positive for duplication of DR1 and DR2 | Negative for duplication of DR1 and DR2 |
|---|---|---|
| Cemani (12) | Cemani40[*b], Cemani43[*b] Cemani41, Cemani42, Cemani44-47, CM6[*b], CM11, CM31[*b], Cemani | – |
| Silkie (12) | SIB2[*b], SIB6, SIB7, SIB11, SIB14, SIB15, SIB16, SIB17[*b], KPS16, KPS17, KPS18, KPS30 | – |
| Kedu (24) | KDH3[*b], KDH8[*b], KD16[*b], KD17, KD19, KD21, KD22 | KD1[*d], KD2[*d], KD3[*b], KD4, KD5[*d], KD7-15, KDP1, KDP5[*a], KDP7[*e], |
| Other chicken breeds (4) | - | STC13[*b], LOM39[*b], BMC610, BMC613, |
| Jungle fowls (4) | - | BKL1[*c], BKL2[*d], Bd72[*f], Bd92[*f] |

[*a-f] *EDN3* haplotypes of 21 individuals are the same as those indicated in Table 2.2; [*a] *Hap*1/*Hap*2, [*b] *Hap*2/*Hap*4, [*c] *Hap*2', [*d] *Hap*4, [*e] *Hap*4/*Hap*7, [*f] *Hap*5. In addition, 35 individuals in Table 2.1 with unknown *EDN3* haplotypes are examined for the duplication. The duplication boundary is identified in all Cemani and Silkie individuals, and also in some Kedu individuals.

**Table 2.4. DNA polymorphism in nine regions in three populations of chickens and GJF (see also Fig 2.5).**

| | | Populations ($k$ = nos. of chromosome examined) | | | |
|---|---|---|---|---|---|
| Region ($L$ bp) and gene in it | Statistics[a] | Cemani ($k$ = 18) | Silkie ($k$ = 20) | Others ($k$ = 22) | GJF ($k$ = 4) |
| R1 (3138) *RPR1D* | $H_k$ ($E(H_k)$) | 3 (7.0) | 8 (10.5) | 12 (12.1) | 4 (3.8) |
| | $S_k$ ($\theta_w$) | 21 (0.19) | 25 (0.23) | 40 (0.35) | 56 (0.97) |
| | $\pi = \theta$ (±SE) | 0.12 (0.03) | 0.26 (0.05) | 0.33 (0.06) | 1.06 (0.14) |
| | $D$ ($r^2$) | 0.064 (0.59) | 0.060 (0.44) | 0.044 (0.32) | 0.12 (0.49) |
| R2 (3299) *NCOA5* | $H_k$ ($E(H_k)$) | 2 (5.1) | 8 (11.9) | 12 (13.7) | 3 (3.8) |
| | $S_k$ ($\theta_w$) | 10 (0.09) | 33 (0.28) | 57 (0.47) | 44 (0.73) |
| | $\pi = \theta$ (±SE) | 0.06 (0.02) | 0.35 (0.07) | 0.44 (0.06) | 0.82 (0.12) |
| | $D$ ($r^2$) | 0.099 (1.0) | 0.054 (0.42) | 0.046 (0.35) | 0.14 (0.57) |
| R3 (3153) *DDX27* | $H_k$ ($E(H_k)$) | 1 (1) | 2 (2.8) | 12 (12.2) | 3 (3.8) |
| | $S_k$ ($\theta_w$) | 0 (0) | 2 (0.02) | 41 (0.36) | 54 (0.93) |
| | $\pi = \theta$ (±SE) | 0 | 0.02 (0.01) | 0.33 (0.07) | 1.14 (0.16) |
| | $D$ ($r^2$) | N.A. | 0.090 (1.0) | 0.043 (0.31) | 0.24 (0.96) |
| R4 (3093) *DDX27* | $H_k$ ($E(H_k)$) | 2 (3.4) | 3 (7.6) | 10 (11.3) | 3 (3.8) |
| | $S_k$ ($\theta_w$) | 2 (0.02) | 13 (0.12) | 40 (0.35) | 37 (0.65) |
| | $\pi = \theta$ (±SE) | 0.03 (0.02) | 0.13 (0.03) | 0.28 (0.05) | 0.72 (0.13) |
| | $D$ ($r^2$) | 0.099 (1.0) | 0.083 (0.77) | 0.051 (0.49) | 0.13 (0.52) |
| R5 (3110) *TH1L* | $H_k$ ($E(H_k)$) | 1 (1) | 1 (1) | 16 (14.2) | 4 (3.9) |
| | $S_k$ ($\theta_w$) | 0 (0) | 0 (0) | 57 (0.50) | 82 (1.44) |
| | $\pi = \theta$ (±SE) | 0 | 0 | 0.52 (0.09) | 1.55 (0.18) |
| | $D$ ($r^2$) | N.A. | N.A. | 0.048 (0.39) | 0.12 (0.48) |
| R6 (3086) *NPEPL1* | $H_k$ ($E(H_k)$) | 4 (8.4) | 9 (12.9) | 14 (14.6) | 4 (3.9) |
| | $S_k$ ($\theta_w$) | 34 (0.32) | 35 (0.32) | 58 (0.52) | 70 (1.24) |
| | $\pi = \theta$ (±SE) | 0.18 (0.03) | 0.47 (0.09) | 0.58 (0.08) | 1.42 (0.18) |
| | $D$ ($r^2$) | 0.078 (0.70) | 0.056 (0.45) | 0.045 (0.33) | 0.17 (0.66) |
| R7 (3286) *NCdup1* | $H_k$ ($E(H_k)$) | 8 (11.5) | 10 (13.7) | 13 (15.4) | 3 (3.9) |
| | $S_k$ ($\theta_w$) | 55 (0.49) | 55 (0.47) | 65 (0.54) | 64 (1.06) |
| | $\pi = \theta$ (±SE) | 0.39 (0.06) | 0.54 (0.08) | 0.66 (0.09) | 1.18 (0.15) |
| | $D$ ($r^2$) | 0.074 (0.64) | 0.057 (0.48) | 0.051 (0.44) | 0.13 (0.52) |
| R8 (3342) | $H_k$ ($E(H_k)$) | 9 (12.3) | 4 (7.3) | 20 (15.8) | 4 (3.9) |

| | | | | | |
|---|---|---|---|---|---|
| *NCdup2* | $S_k$ ($\theta_w$) | 38 (0.33) | 21 (0.18) | 76 (0.62) | 71 (1.16) |
| | $\pi = \theta$ (±*SE*) | 0.47 (0.09) | 0.11 (0.02) | 0.70 (0.09) | 1.28 (0.16) |
| | *D* ($r^2$) | 0.071 (0.59) | 0.073 (0.73) | 0.045 (0.33) | 0.092 (0.37) |
| | $H_k$ ($E(H_k)$) | 12 (13.4) | 14 (14.8) | 14 (15.7) | 3 (3.7) |
| R9 (3960) *BMP7* | $S_k$ ($\theta_w$) | 65 (0.48) | 76 (0.54) | 81 (0.56) | 27 (0.37) |
| | $\pi = \theta$ (±*SE*) | 0.56 (0.08) | 0.61 (0.08) | 0.58 (0.06) | 0.41 (0.07) |
| | *D* ($r^2$) | 0.060 (0.44) | 0.052 (0.38) | 0.044 (0.33) | 0.13 (0.53) |

[a] $H_k$ is the observed number of haplotypes in a sample of $k$ chromosomes. $E(H_k)$ is based on the formula for the expected number of neutral alleles with per-locus mutation rate $\theta L$, where $\theta$ is given by the observed $\pi$ value and $L$ is the number of nucleotides per region. $S_k$ is the observed number of segregating sites within a region. $\theta_w$ is Watterson's $\theta$ and $\pi = \theta$ is nucleotide diversity, both were multiplied by 100. $D$ is the mean value of linkage disequilibrium across all pairs of polymorphic sites within a region, and $r$ is the corresponding correlation coefficient given by $r = \frac{D}{\sqrt{p_A q_A p_B q_B}}$, where the denominator is proportional to heterozygosity at both sites *A* and *B*.

## 2.9 Supporting Information

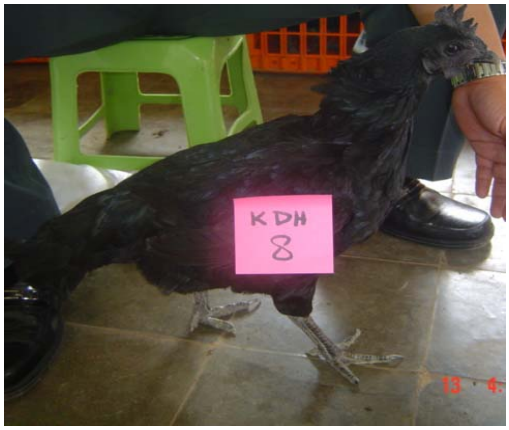
**S2.1 Figure. Characteristic morphological traits of several Indonesian chickens and Chinese Silkie**. (a) Female Cemani, (b) - (d) female Kedu, (e) - (i) male Kedu, (j) male white Silkie and (k) male black Silkie.

**a**

**b**


**KDH3**

**c**




**KDH 8**

**d**


**KD3**

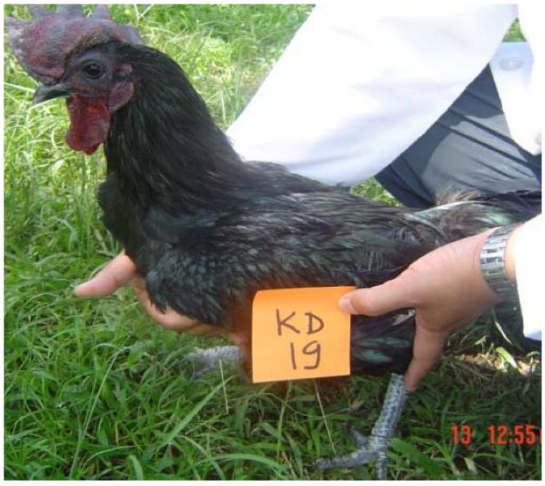
**e**


**KD16**

f


KD17

g




KD19

h


KD21

i


KD22

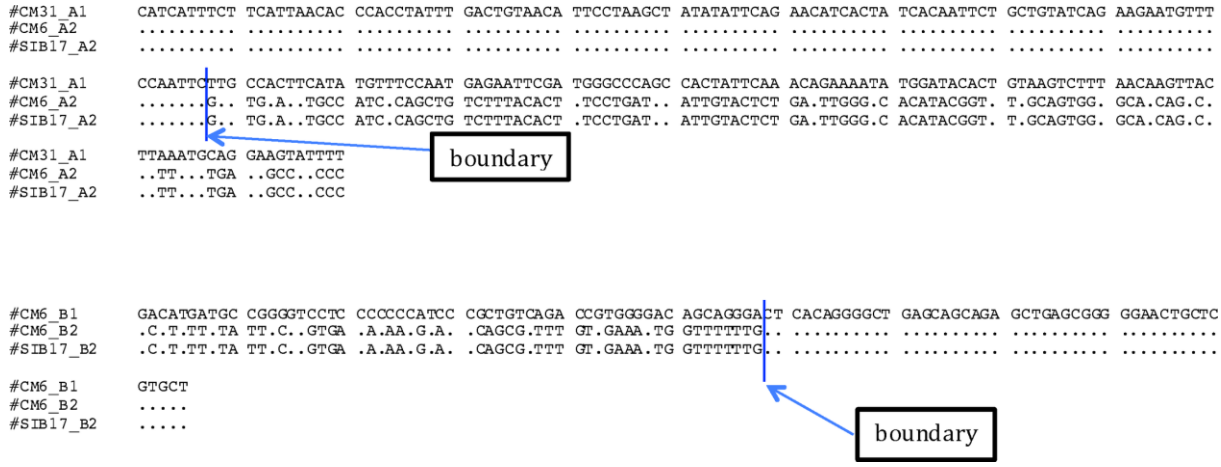
j

White Silkie

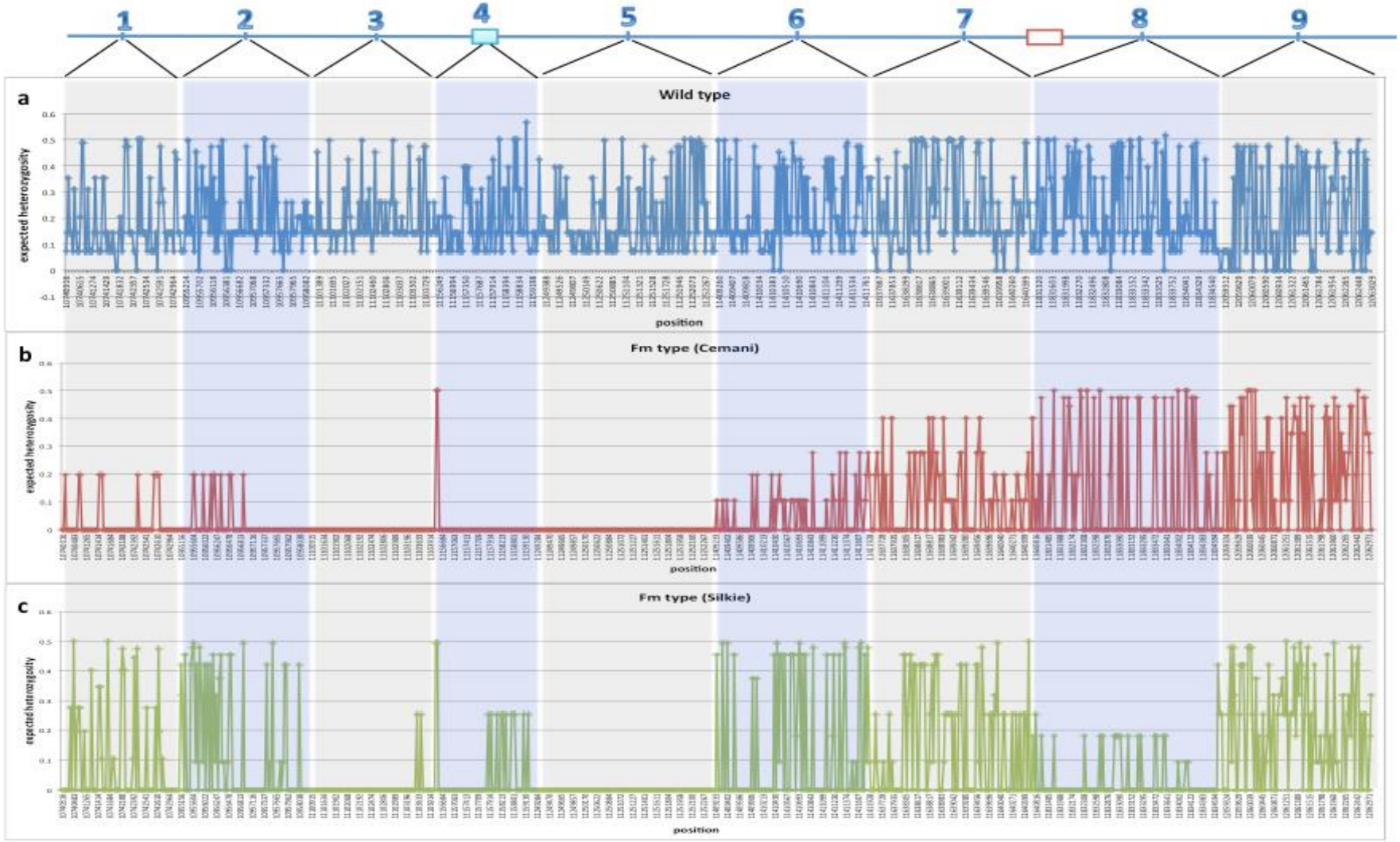
k

Black Silkie

**S2.2 Figure**. **Sequence information for duplication boundaries generated by the A2 and B2 primer sets**. The A2 and B2 sequences of Cemani (CM6_A2 and CM6_B2) are identical to those of Silkie (SIB17_A2 and SIB17_B2). The boundary was determined by comparison between A1 (CM31_A1) and A2 (upper panel), and between B1 (CM6_B1) and B2 (lower panel).

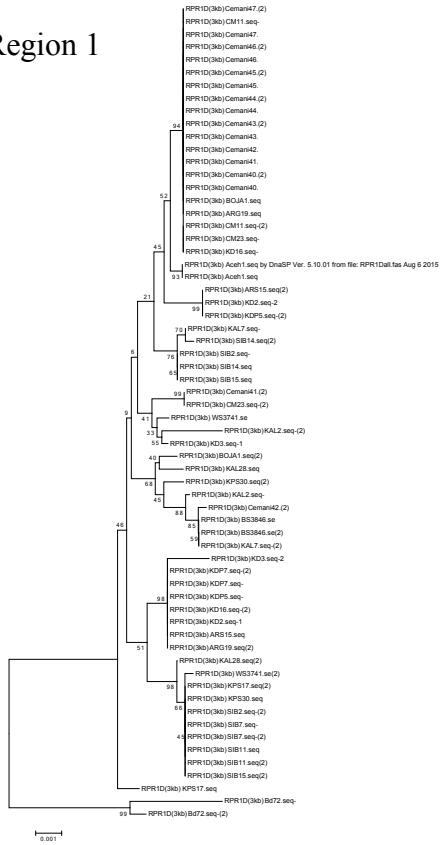


**S2.3 Figure**. **Expected heterozygosity at individual SNP sites in the nine regions in chicken breeds**. (a) Domesticated chickens, RJF, and GJF, (b) Ayam Cemani, (c) Silkie chicken.

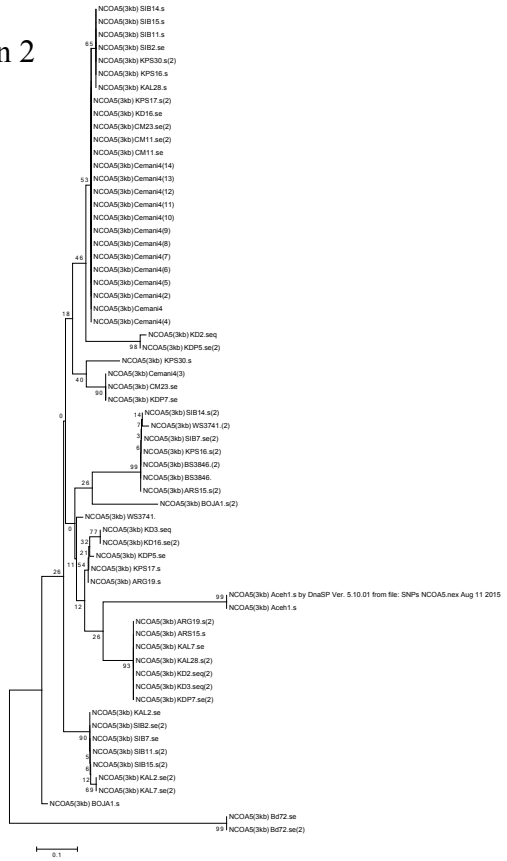


**S2.4 Figure**. **NJ trees for the nine regions in domesticated chicken breeds, and RJF**. The phylogenetic relationship differs greatly from region to region. Two GJF haplotype sequences were used as outgroups.

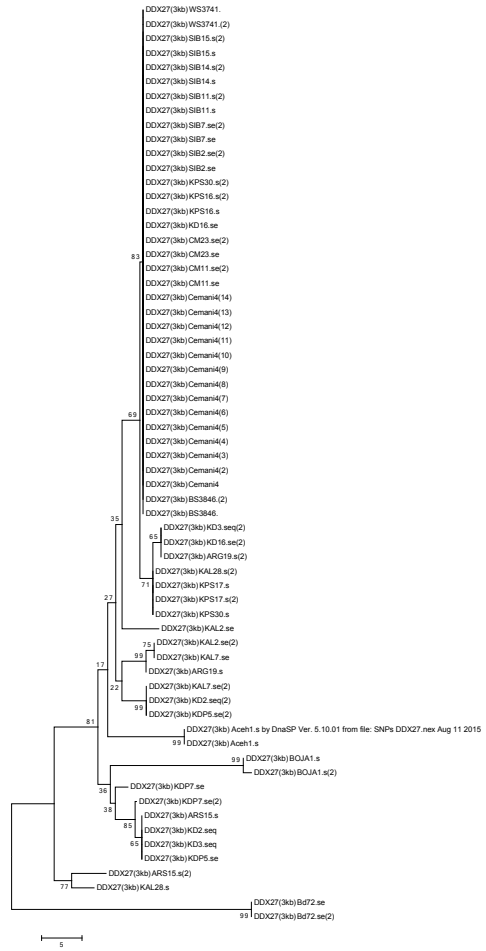
Region 1



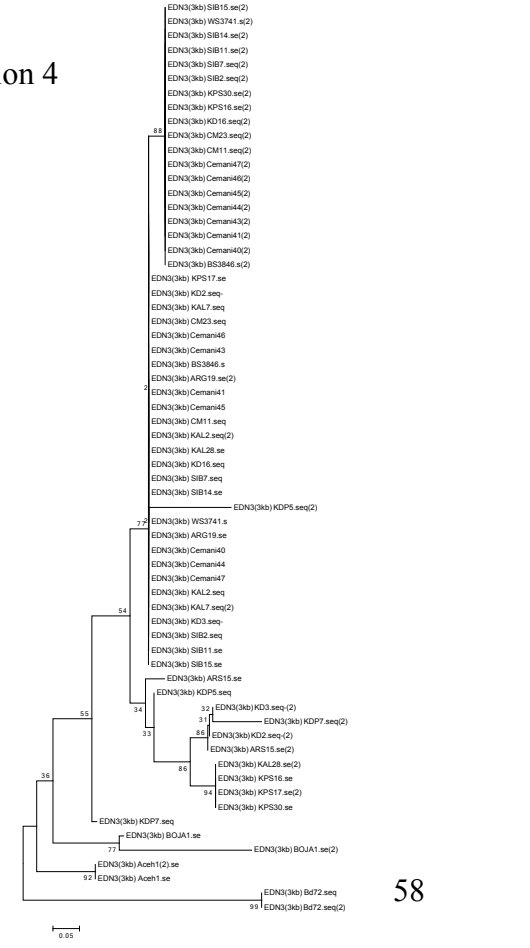

Region 2



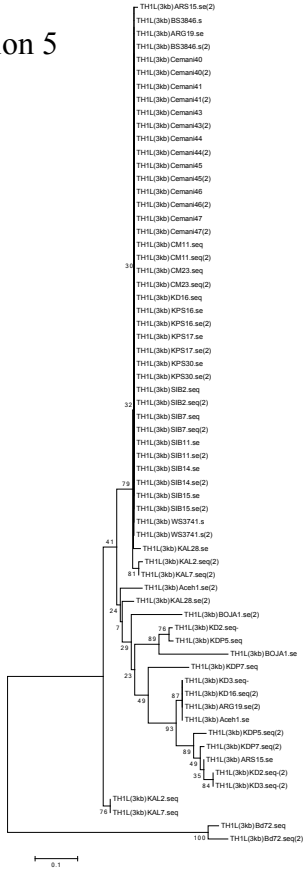

Region 3



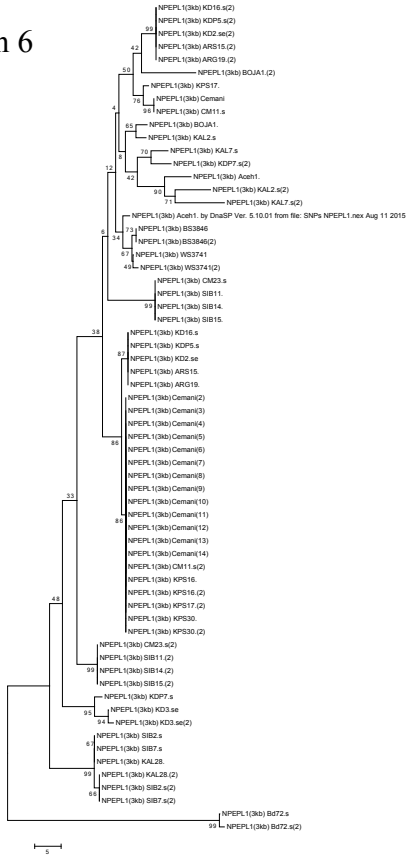

Region 4



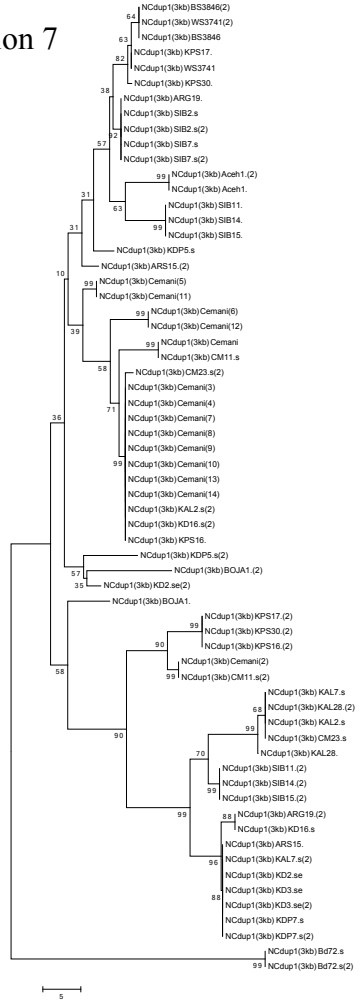
Region 5



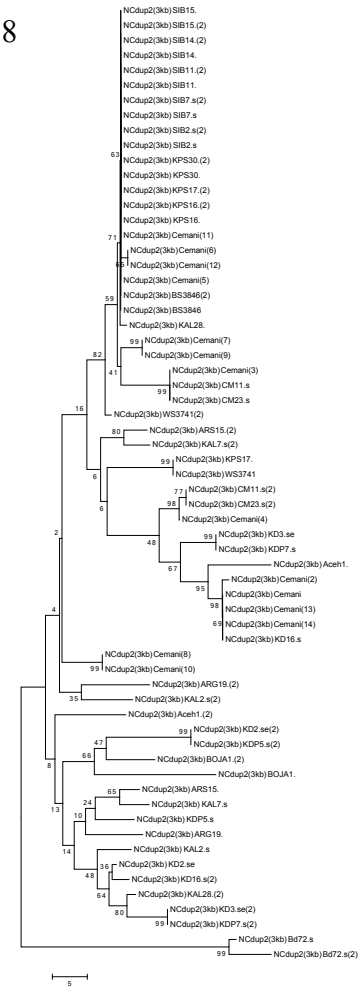

Region 6



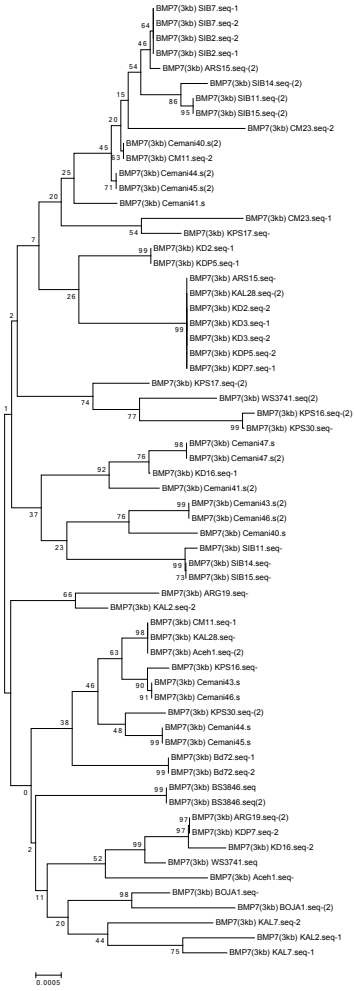

Region 7




Region 8

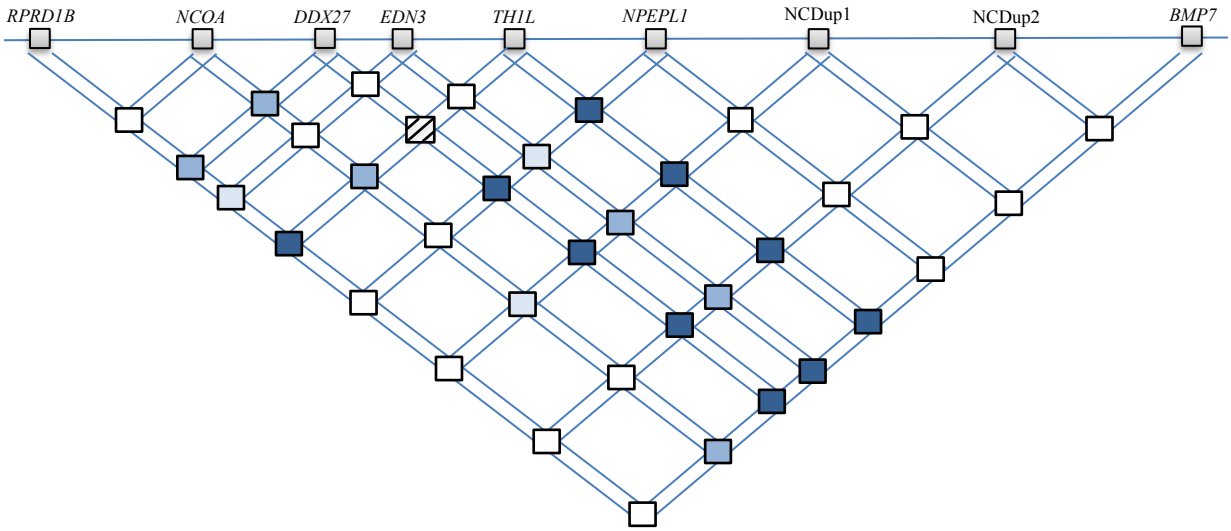


Region 9

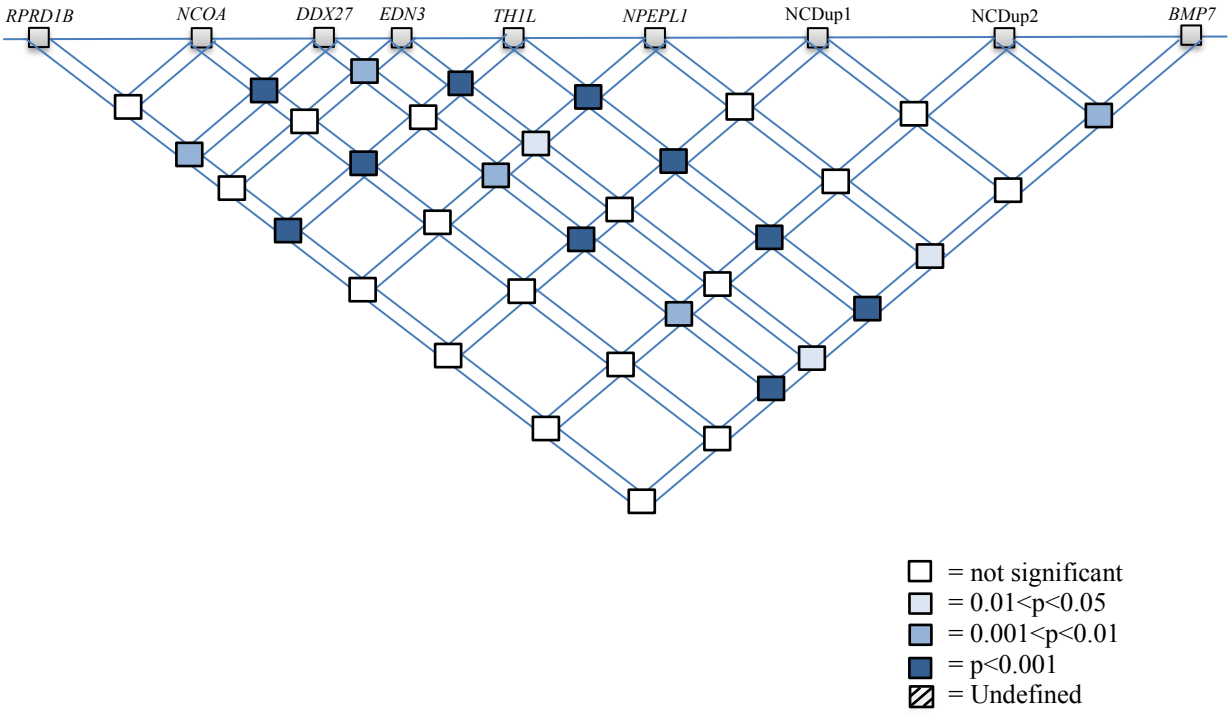

BMP7(3kb) SIB7.seq-1
BMP7(3kb) SIB7.seq-2
BMP7(3kb) SIB2.seq-2
BMP7(3kb) SIB2.seq-1
BMP7(3kb) ARS15.seq-(2)
BMP7(3kb) SIB14.seq-(2)
BMP7(3kb) SIB11.seq-(2)
BMP7(3kb) SIB15.seq-(2)
BMP7(3kb) CM23.seq-2
BMP7(3kb) Cemani40.s(2)
BMP7(3kb) CM11.seq-2
BMP7(3kb) Cemani44.s(2)
BMP7(3kb) Cemani45.s(2)
BMP7(3kb) Cemani41.s
BMP7(3kb) CM23.seq-1
BMP7(3kb) KPS17.seq-
BMP7(3kb) KD2.seq-1
BMP7(3kb) KDP5.seq-1
BMP7(3kb) ARS15.seq-
BMP7(3kb) KAL28.seq-(2)
BMP7(3kb) KD2.seq-2
BMP7(3kb) KD3.seq-1
BMP7(3kb) KD3.seq-2
BMP7(3kb) KDP5.seq-2
BMP7(3kb) KDP7.seq-1
BMP7(3kb) KPS17.seq-(2)
BMP7(3kb) WS3741.seq(2)
BMP7(3kb) KPS16.seq-(2)
BMP7(3kb) KPS30.seq-
BMP7(3kb) Cemani47.s
BMP7(3kb) Cemani47.s(2)
BMP7(3kb) KD16.seq-1
BMP7(3kb) Cemani41.s(2)
BMP7(3kb) Cemani43.s(2)
BMP7(3kb) Cemani46.s(2)
BMP7(3kb) Cemani40.s
BMP7(3kb) SIB11.seq-
BMP7(3kb) SIB14.seq-
BMP7(3kb) SIB15.seq-
BMP7(3kb) ARG19.seq-
BMP7(3kb) KAL2.seq-2
BMP7(3kb) CM11.seq-1
BMP7(3kb) KAL28.seq-
BMP7(3kb) Aceh1.seq-(2)
BMP7(3kb) KPS16.seq-
BMP7(3kb) Cemani43.s
BMP7(3kb) Cemani46.s
BMP7(3kb) KPS30.seq-(2)
BMP7(3kb) Cemani44.s
BMP7(3kb) Cemani45.s
BMP7(3kb) Bd72.seq-1
BMP7(3kb) Bd72.seq-2
BMP7(3kb) BS3846.seq
BMP7(3kb) BS3846.seq(2)
BMP7(3kb) ARG19.seq-(2)
BMP7(3kb) KDP7.seq-2
BMP7(3kb) KD16.seq-2
BMP7(3kb) WS3741.seq
BMP7(3kb) Aceh1.seq-
BMP7(3kb) BOJA1.seq-
BMP7(3kb) BOJA1.seq-(2)
BMP7(3kb) KAL7.seq-2
BMP7(3kb) KAL2.seq-1
BMP7(3kb) KAL7.seq-1
0.0005

**S2.5 Figure**. **Results of the HKA test in each of the nine regions of Cemani (a), Silkie (b), and other domesticated chickens (c).** The significant reduction in DNA polymorphism is found in Cemani and Silkie only for *DDX27* in region 3, *EDN3* in region 4, and *TH1L* in region 5 are compared.

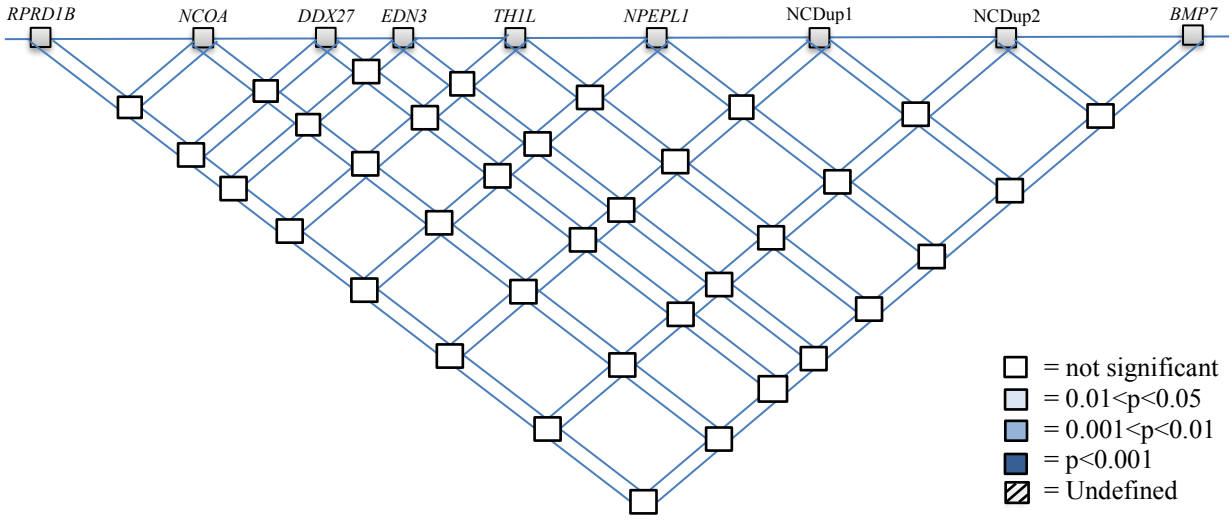
Ayam Cemani



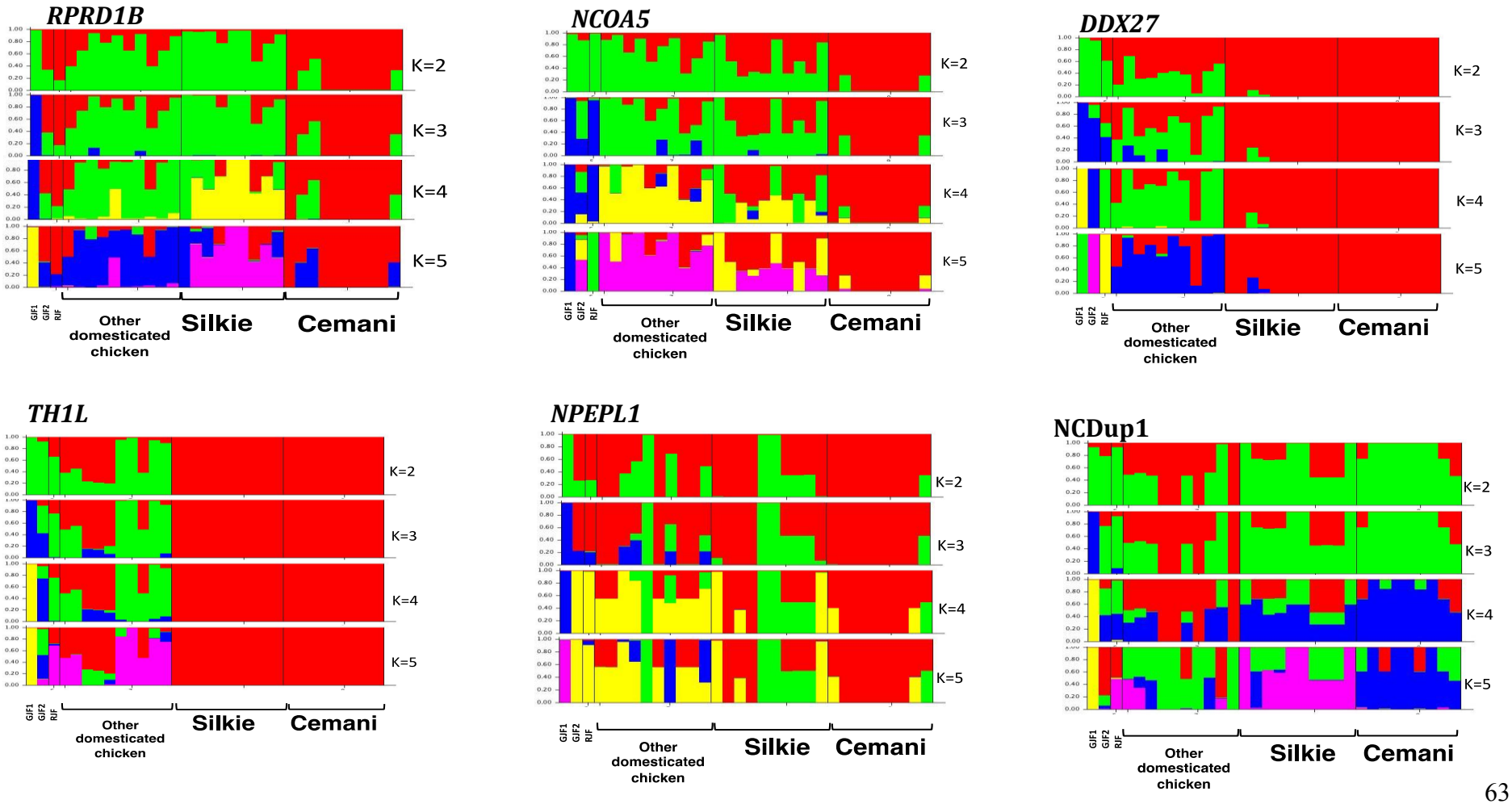

Silky



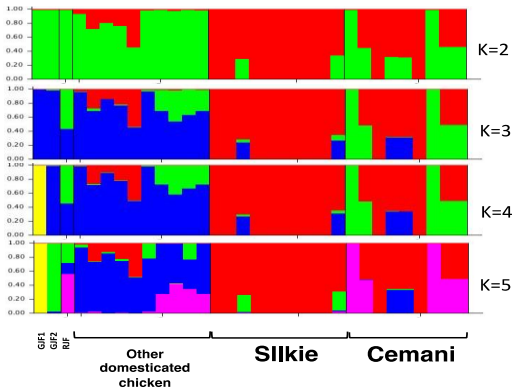
Other Domesticated Chickens



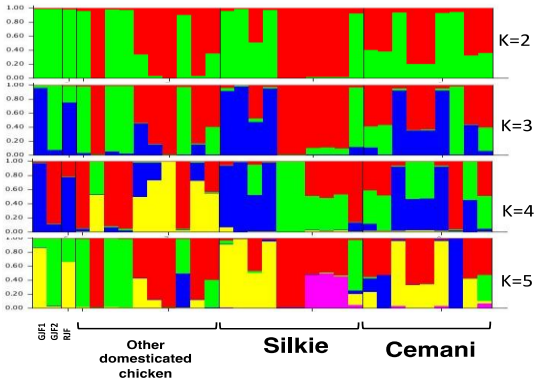
**S2.6 Figure**. **STRUCTURE analysis of each of the nine regions of GJF, RJF, Cemani, Silkie and other domesticated chickens.** For regions 3–5, Cemani and Silkie exhibit nearly identical genetic components, whereas in other regions, there are no noticeable structural differences among chicken breeds and RJF.



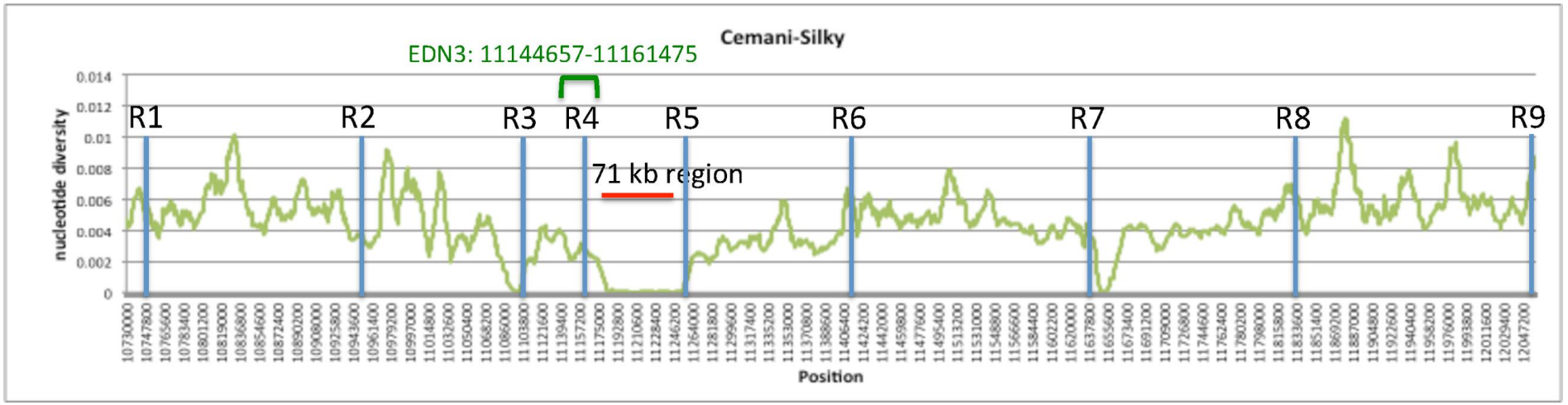
**NCDup2**


K=2
K=3
K=4
K=5
GIF1
GIF2
NP
Other domesticated chicken
SIlkie
Cemani

***BMP7***


K=2
K=3
K=4
K=5
GIF1
GIF2
NP
Other domesticated chicken
Silkie
Cemani

**S2.7 Figure**. **Nucleotide diversity between Cemani and Silkie based on NGS data.** Bars with R1–R9 indicate the positions of the nine regions. Green square parentheses indicate the position of *EDN3*, and a red bar indicates the 71.4-kb region with low divergence between the two breeds.



**S2.1 Table. Sequences of primers used in this study.**

| Primer name | Forward | Reverse | Position |
|---|---|---|---|
| RPRD1B (Region1) | 5'-GCACACACTCAAACTACGAGGTTACAGT-3' | 5'-GCTCAGTATCATTGGCTGTTAGATGGT-3' | 10740207 - 10743372 |
| NCO5 (Region 2) | 5'-TGAGTTATGCTAGGAAGGTCAGGCTTCT-3' | 5'-GGCATATGGCTCTCTGAGGAAAACAC-3' | 10955102 - 10958453 |
| DDX27 (Region 3) | 5'-CTACCACGCTGGATGAGAAGATTGAGA-3' | 5'-GTGCAGCCACATCTGTAGCTACCAAA-3' | 11100827 - 11104025 |
| EDN3 (Region 4) | 5'-AATCGTGTCTATGGGAACGGTAAACG-3' | 5'-TCGCTCTGTCCATCCAATTGTACTGAT-3' | 11156158 - 11159283 |
| TH1L (Region 5) | 5'-GAGTTCTGCGTACATCTCTGGCAACA-3' | 5'-TCATCCTTGTTTATGCTCACTCTCTTGTTC-3' | 11249249 - 11252418 |
| NPEPL1 (Region 6) | 5'-GTGCCTTCAAAGCCACTGTAAAGCA-3' | 5'-CCTGCCAGCTTTAACACAAGCCTTT-3 | 11409018 – 11412144 |
| NCdup (Region 7) | 5'-GACAATACGTCCAGCATCTCAGTTTACC-3' | 5'-CCTACATGCACCCATTAGCTTGAAGA-3' | 11637526 – 11640822 |
| NCdup2 (Region 8) | 5'-CTCTCGTCCACATTGAAATGAATCAAG-3' | 5'-TCCCATCTCAGTGTCTATGAGAGCAGTT-3' | 11831272 - 11834623 |
| BMP7 (Region 9) | 5'-GGTGTTTGACATCACTGCAACCAGTAAT-3' | 5'-GCTTTGCCCTTAACAACAACACAGCAT-3' | 12059056 - 12063056 |

| Primer list between region 2-3 and region 5-6 | | | |
|---|---|---|---|
| **Region 2 - 3** | | | |
| Primer name | Forward | Reverse | Position |
| D(2-3) | 5'-AGAGTTGCAGGTAAGAAAGCTATG-3' | 5'-AAGCTGACAAATTGCCTGAAG-3' | 11036545-11037716 |
| F(2-3) | 5'-TTGTGCTGTTCAGTCAGGTTCT-3' | 5'-TCAGCCTTCTGCATAAGTATCTGT-3' | 11074442-11075542 |
| **Region 5 - 6** | | | |
| Primer name | Forward | Reverse | |
| C(5-6) | 5'-ACCTTATTAACCCATGGCATGT-3' | 5'-ATAATGCTACGACAGCAGGAAAC-3' | 11310054-11311124 |
| D(5-6) | 5'-TTTCTTGCTCAGGTTTGCAGTA-3' | 5'-CAATTAGGTGAAGGCAGAATACAG-3' | 11329264-11330394 |
| E(5-6) | 5'-AGGCATTCCTCACTTTAAGCAT-3' | 5'-CACCTGTACTGATGTGCTTTGAG-3' | 11345368-11346407 |
| G(5-6) | 5'-AACAAAGCTGGTTTGGTTCTG-3' | 5'-TGCTTGTCTGACTCCTCTGTATG-3' | 11384999-11386123 |
| **Primer list for haplotype of *EDN3* variation and qPCR analysis** | | | |
| Primer name | Forward | Reverse | |
| AS044 (Shinomiya et al.2011) | 5’-CCCAGCCTTCATTTCGGTGC-3’ | 5’-CCCTCCAAGCTCTGCTACTG-3’ | 11146917-11148035 |
| qAS044 | 5’-CCTCATGTCTCGTAGCATAGGCTAACTC-3’ | 5’-TGACTTTATCACCATGTTTGAGCTTTCTC-3’ | 11147700-11147839 |
| q46 | 5’-ACACCTCCACCACCCAAGAAT-3’ | 5’-GCGAATGGAGAATGAACAACATCA-3’ | 9824248-9824394 |
| **Primer list for duplication boundary detection (Dorshort et al. 2010)** | | | |
| Primer set name | Forward | Reverse | |
| A1 | 232 (AGAAACAAGGGTCAAGGTGAGC) | 234 (TGGATCATTGGAGGAAGTGTTG) | |
| A2 | 200 (GGGATGGCTCTCACATAAAAGG) | 234 (TGGATCATTGGAGGAAGTGTTG) | |
| B1 | 201(CTTGGCTCAGA T A TTCGCCTCT) | 202 (AGGCACAGTCTGGCACATTAAA) | |
| B2 | 197 (GCAGCCTTT A TT A TTGCGTGTG) | 201(CTTGGCTCAGA T A TTCGCCTCT) | |

**S2.2 Table. Reaction mixtures and PCR conditions used in this study.**

| | PCR1 | PCR2 | PCR3 | PCR4 |
|---|---|---|---|---|
| Total reaction volume (µl) | 30 | 25 | 20 | 30 |
| Taq polymerase | (Takara LA TaqTM) 0.5 U/µl | (Takara ex Taq) 0.5 U/µl | SYBR® Premix Ex TaqTM II (Tli RNaseH Plus) | (Takara ex Taq) 0.5 U/µl |
| 2.5mdNTP | 3 µl | 2.5 µl | SYBR Premix 10µl | 3 µl |
| 10 X buffer | (plus Mg2+) 3µl | | (plus Mg2+) 3 µl | |
| 10pM of forward and reverse primer | 3µl | 2.5 µl | 1.6µl | 3 µl |
| template | 40 ng | 10 ~ 100 ng | 40 ng | 20-100 ng |
| Reaction condition | D (denaturation) at 94°C for 3min. | D at 95°C for 5min. | D at 95°C 1min. | D at 94°C for 5min. |
| | 30 cycles of D at 94°C for 3min., A(annealing) at 63~65°C for 1 min., and E(extenstion) at 72°C for 1min. Final E at 72°C for 5min. | 30 cycles of D at 95°C for 30 sec., A at 58°C for 30 sec., and E at 72°C for 1min. Final E at 72°C for 5min. | 40 cycles of D at 95°C for 5 sec. A/E at 60°C for 30 sec. | 35 cycles of D at 94°C for 30 sec., A at 58°C for 30 sec., and E at 72°C for 1min. Final E at 72°C for 5min. |

**S2.3 Table. Segregating sites in *EDN3* haplotypes.**

| Site position[*1] | 48 | 56 | 124 | 125 | 126 | 127 | 131 | 173 | 234 | 261 | 266 | 269 | 318 | 361 | 365 | 432 | 436 | 456 | 492 | 616 | 712 | 732 | 744 | 764 | 784 | 808 | 857 | 884 | 930 | 1013 | 1065 | 1066 | 1067 | 1094 | 1101 | No.[*2] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Haplotype 1 | A | C | C | T | G | A | C | T | A | G | C | A | G | A | - | G | C | G | A | G | C | A | A | T | A | G | C | A | T | A | A | C | A | G | C | 1 |
| Haplotype 2 | . | . | . | . | . | . | . | . | . | . | . | . | . | T | G | T | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 17 |
| Haplotype 2’ | . | . | . | . | . | . | . | . | . | . | . | . | . | T | G | T | T | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | 1 |
| Haplotype 3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | - | . | . | . | . | C | . | T | . | C | . | A | T | . | . | G | . | . | . | . | . | 1 |
| Haplotype 4 | . | . | . | . | . | . | . | . | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 23 |
| Haplotype 5 | G | T | T | - | - | - | T | . | . | . | T | C | A | . | - | . | . | . | G | . | . | . | G | C | . | . | . | . | . | . | . | . | . | A | T | 2 |
| Haplotype 6 | . | T | . | . | . | . | . | C | . | . | . | . | A | . | - | T | T | A | G | C | A | . | . | C | . | . | . | . | . | G | . | . | . | - | - | 2 |
| Haplotype 7 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | - | . | . | . | G | C | . | T | . | C | . | A | . | . | . | G | . | . | . | . | . | 2 |
| Haplotype 8 | . | T | T | . | . | . | T | . | G | . | T | C | A | . | - | . | . | . | G | . | . | . | G | C | . | . | . | G | G | . | . | . | . | . | . | 2 |
| Haplotype 9 | . | . | . | . | . | . | . | . | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | - | - | - | . | . | 1 |
| Haplotype 10 | . | . | . | . | . | . | . | . | . | . | . | . | . | T | G | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 2 |
| Haplotype 11 | . | T | . | . | . | . | . | C | . | A | . | . | A | . | - | . | . | A | G | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | 1 |

The top row shows relative positions of segregating sites in the cloned segment. A shadowed cell indicates 1-bp deletion.

*1 Site position 1 corresponds to 11005500 in the reference sequence (NC_006107).

*2: No. of individuals observed in the sample of size of 17 homozygotes and 19 heterozygotes.

**S2.4 Table. Variable sites in the 71.4-kb region in Cemani, Silkie and Taiwanese L2 as compared to the reference genome.** The region ranges from nt 11,183,600 to 11,255,000 and includes part of DR1. Insertions and deletions are excluded. The colored columns indicate the Silkie (green)- or Cemani (red)-specific mutations.

| position | reference | Cemani | genotype | Silkie | genotype | L2 Taiwanese | genotype |
|---|---|---|---|---|---|---|---|
| 11184107 | T | T/C | het | T/C | het | T | hom |
| 11184380 | C | C/T | het | C/T | het | C | hom |
| 11184505 | C | C/G | het | C/G | het | C | hom |
| 11184591 | C | C/A | het | C/A | het | C | hom |
| 11184672 | A | A | hom | A/G | het | A | hom |
| 11184694 | G | G/A | het | G/A | het | G | hom |
| 11184727 | T | T/G | het | T/G | het | T | hom |
| 11184734 | A | A/T | het | A/T | het | A | hom |
| 11185151 | C | C/A | het | C/A | het | C | hom |
| 11185935 | C | C/T | het | C/T | het | C | hom |
| 11186109 | C | C/T | het | C/T | het | C | hom |
| 11186125 | T | T/C | het | T/C | het | T | hom |
| 11186218 | G | G/A | het | G/A | het | G | hom |
| 11186220 | G | G/A | het | G/A | het | G | hom |
| 11186266 | T | T/G | het | T/G | het | T | hom |
| 11186267 | G | G/T | het | G/T | het | G | hom |
| 11186672 | C | C/T | het | C/T | het | C | hom |
| 11186808 | G | G/A | het | G/A | het | G | hom |
| 11187282 | T | T/C | het | T/C | het | T | hom |
| 11187334 | C | C/T | het | C/T | het | C | hom |
| 11187494 | G | A | hom | A | hom | G | hom |
| 11187816 | G | G/A | het | G/A | het | G | hom |
| 11187964 | T | T/C | het | T/C | het | T | hom |
| 11187996 | T | T/C | het | T/C | het | T | hom |
| 11188298 | C | C/T | het | C/T | het | C | hom |
| 11188585 | A | A/G | het | A/G | het | A | hom |
| 11188777 | G | G/T | het | G/T | het | G | hom |
| 11188845 | T | C | hom | C | hom | T | hom |
| 11189939 | A | A/G | het | A/G | het | A | hom |
| 11190074 | A | A/G | het | A/G | het | A | hom |
| 11190475 | T | T/C | het | T/C | het | T | hom |
| 11190652 | T | T/C | het | T/C | het | T | hom |
| 11190958 | T | T/A | het | T/A | het | G | hom |
| 11190958 | T | T/G | het | T/G | het | T | hom |

# of het_Cemani = 50
# of hom_Silkie = 51
# of het_Cemani = 106
# of hom_Silkie = 105

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 11191533 | A | A/G | het | A/G | het | A | hom |
| 11191638 | A | A/G | het | A/G | het | A | hom |
| 11191752 | C | C/T | het | C/T | het | C | hom |
| 11192379 | T | T/C | het | T/C | het | T | hom |
| 11192412 | C | C/T | het | C/T | het | C | hom |
| 11192860 | G | G/A | het | G/A | het | G | hom |
| 11193080 | G | G/C | het | G/C | het | G | hom |
| 11193105 | A | A/G | het | A/G | het | A | hom |
| 11193454 | G | G/A | het | G/A | het | G | hom |
| 11193610 | T | C | hom | C | hom | T | hom |
| 11193636 | C | A | hom | A | hom | C | hom |
| 11193723 | G | G/A | het | G/A | het | G | hom |
| 11193749 | T | T/G | het | T/G | het | T | hom |
| 11194503 | T | T/A | het | T/A | het | T | hom |
| 11194937 | A | A/G | het | A/G | het | A | hom |
| 11196053 | T | T/A | het | T | | T | hom |
| 11196105 | A | A/G | het | A/G | het | A | hom |
| 11197355 | A | A/C | het | A/C | het | A | hom |
| 11197426 | T | T/C | het | T/C | het | T | hom |
| 11201894 | C | A | hom | A | hom | C | hom |
| 11202063 | G | A | hom | A | hom | G | hom |
| 11202703 | T | G | hom | T | | T | hom |
| 11203349 | C | T | hom | T | hom | C/G | het |
| 11203758 | C | T | hom | T | hom | C | hom |
| 11204021 | C | A | hom | A | hom | C | hom |
| 11204052 | C | T | hom | T | hom | C | hom |
| 11204053 | A | G | hom | G | hom | A | hom |
| 11204131 | C | A | hom | A | hom | C | hom |
| 11204145 | T | C | hom | C | hom | T | hom |
| 11204268 | C | T | hom | T | hom | C | hom |
| 11204289 | G | A | hom | A | hom | G | hom |
| 11204553 | A | G | hom | G | hom | A | hom |
| 11205462 | G | A | hom | A | hom | G | hom |
| 11205485 | C | T | hom | T | hom | C | hom |
| 11205829 | C | A | hom | A | hom | C | hom |
| 11205845 | T | C | hom | C | hom | T | hom |
| 11205892 | T | G | hom | G | hom | T | hom |
| 11206338 | A | C | hom | C | hom | A | hom |
| 11206824 | A | G | hom | G | hom | A | hom |
| 11206998 | G | T | hom | T | hom | G | hom |
| 11207428 | G | T | hom | T | hom | G | hom |
| 11207467 | C | T | hom | T | hom | C | hom |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 11207724 | T | C | hom | C | hom | T | hom |
| 11207940 | G | C | hom | C | hom | G | hom |
| 11208255 | G | C | hom | C | hom | G | hom |
| 11208366 | G | C | hom | C | hom | G | hom |
| 11209096 | C | G | hom | G | hom | C | hom |
| 11209342 | G | A | hom | A | hom | G | hom |
| 11209633 | G | A | hom | A | hom | G | hom |
| 11209692 | G | G/A | het | G/A | het | G | hom |
| 11210453 | C | A | hom | A | hom | C | hom |
| 11213748 | C | T | hom | T | hom | C | hom |
| 11213786 | C | T | hom | T | hom | C | hom |
| 11214040 | T | A | hom | A | hom | T | hom |
| 11214913 | C | A | hom | A | hom | C | hom |
| 11214923 | A | G | hom | G | hom | A | hom |
| 11215207 | C | G | hom | G | hom | C | hom |
| 11215871 | G | A | hom | A | hom | G | hom |
| 11217777 | C | T | hom | T | hom | C | hom |
| 11217989 | C | T | hom | T | hom | C | hom |
| 11218209 | G | A | hom | A | hom | G | hom |
| 11218613 | T | A | hom | A | hom | T | hom |
| 11219048 | G | A | hom | A | hom | G | hom |
| 11219232 | G | T | hom | T | hom | G | hom |
| 11219370 | T | C | hom | C | hom | T | hom |
| 11220573 | G | C | hom | C | hom | G | hom |
| 11220623 | C | T | hom | T | hom | C | hom |
| 11221946 | G | A | hom | A | hom | G | hom |
| 11222140 | G | A | hom | A | hom | G | hom |
| 11222666 | C | A | hom | A | hom | C | hom |
| 11222766 | C | G | hom | G | hom | C | hom |
| 11223603 | A | G | hom | G | hom | A | hom |
| 11224007 | G | T | hom | T | hom | G | hom |
| 11224077 | A | G | hom | G | hom | A | hom |
| 11224374 | C | A | hom | A | hom | C | hom |
| 11224518 | C | A | hom | A | hom | C | hom |
| 11224519 | T | C | hom | C | hom | T | hom |
| 11224682 | A | G | hom | G | hom | A | hom |
| 11224974 | C | G | hom | G | hom | C | hom |
| 11225127 | C | T | hom | T | hom | C | hom |
| 11225316 | A | G | hom | G | hom | A | hom |
| 11226108 | T | C | hom | C | hom | T | hom |
| 11230298 | T | C | hom | C | hom | T | hom |
| 11231669 | A | C | hom | C | hom | A | hom |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 11232740 | A | G | hom | G | hom | A | hom |
| 11232802 | A | C | hom | C | hom | A | hom |
| 11233294 | A | G | hom | G | hom | A | hom |
| 11233373 | G | T | hom | T | hom | G | hom |
| 11233533 | C | G | hom | G | hom | C | hom |
| 11233849 | G | C | hom | C | hom | G | hom |
| 11234218 | C | T | hom | T | hom | C | hom |
| 11234374 | G | A | hom | A | hom | G | hom |
| 11234810 | T | T/G | het | T/G | het | T | hom |
| 11234814 | G | T | hom | T | hom | G | hom |
| 11234820 | A | T | hom | T | hom | A | hom |
| 11235522 | C | T | hom | T | hom | C | hom |
| 11235847 | T | G | hom | G | hom | T | hom |
| 11235959 | T | C | hom | C | hom | T | hom |
| 11236606 | C | C | | C/T | het | C | hom |
| 11237812 | T | C | hom | C | hom | T | hom |
| 11237813 | T | G | hom | G | hom | T | hom |
| 11238450 | G | A | hom | A | hom | G | hom |
| 11238851 | C | T | hom | T | hom | C | hom |
| 11239021 | C | T | hom | T | hom | C | hom |
| 11239313 | T | G | hom | G | hom | T | hom |
| 11239899 | C | G | hom | G | hom | C | hom |
| 11240721 | G | A | hom | A | hom | G | hom |
| 11241229 | G | C | hom | C | hom | G | hom |
| 11241403 | A | G | hom | G | hom | A | hom |
| 11241417 | C | T | hom | T | hom | C | hom |
| 11244911 | A | C | hom | C | hom | A | hom |
| 11244955 | C | T | hom | T | hom | C | hom |
| 11245067 | T | C | hom | C | hom | T | hom |
| 11245293 | T | C | hom | C | hom | T | hom |
| 11247151 | C | T | hom | T | hom | C | hom |
| 11248936 | G | A | hom | A | hom | G | hom |
| 11249067 | C | T | hom | T | hom | C | hom |
| 11249122 | T | C | hom | C | hom | T | hom |
| 11250412 | G | A | hom | A | hom | G | hom |
| 11251087 | A | G | hom | G | hom | A | hom |
| 11251132 | G | C | hom | C | hom | G | hom |
| 11251383 | C | A | hom | A | hom | C | hom |
| 11254430 | C | T | hom | T | hom | C | hom |
| 11254491 | G | A | hom | A | hom | G | hom |

1

# Chapter 3

## Analysis of Cemani chicken genome reveals candidate genes under selective sweeps

### 3.1 Introduction

Indonesian local chickens have been artificially selected over its long breeding history using criteria meeting human preferences, resulting in distinct phenotypes within each breed. Breeds have been selected as a source of food, pet, and hobby, or even for ceremonial purposes. Due to this phenotypic variation, Indonesian local breeds are valuable genetic resources that can be used as a model for genomic studies. One of the interesting domestic breeds from Indonesia is the Cemani chicken that possesses the fibromelanosis (Fm) phenotype, where hyperpigmentation occurs in skin, flesh and internal organs. In addition, Cemani is categorized as an Indonesian local chicken with high egg production and large body size [1], thus making them good candidates for identifying genes related to high performance for egg productivity.

The dramatic phenotypic differences between breeds in domestic chicken imply that there is alteration of genotypic trait due to selection for particular traits. Thus, by investigating the genomic regions that are under selective sweeps, the genetic variation underlying diversification of phenotypes can be identified. In addition, it can answer a question regarding how genetic mechanisms contributed markedly to the rapid evolution of domestic chickens.

Several studies that detected the genetic background of specific phenotypic traits in the domestic chicken have focused on commercial chicken breeds i.e. broiler (meat-type) and layer (egg-type) chicken, because these chickens have been intensively selected for growth and reproductive traits [2,3,4,5,6]. From these studies, several genes were detected as candidate genes associated with phenotypic changes in broiler and layer chickens. It is likely that growth and reproductive traits are complex traits that are

controlled by many genes. Another study that focused on a specific characteristic in local chicken was reported in Silkie chicken [7]. In Chapter 2, I discovered a selective sweep in the Fm region (associated with the Fibromelanosis phenotype), which contains a duplication of a region containing *EDN3* (Endothelin 3) on chromosome 20 of Cemani and Silkie. Since Cemani and Silkie showing different phenotypic characteristics, it is possible that Cemani possess specific genomic regions that may have been under selection for Cemani phenotypic traits.

The ability to assess evidence for selection at the genetic level represented a breakthrough for this pursuit. Most genome-wide studies using high throughput genotyping tools, like high-density SNP arrays and next generation sequencing, have been beneficial for detecting genomic footprints of artificial selection. Several approaches have been developed for identifying selected genomic regions associated with specific traits. (List of studies and statistical test are summarized in Table 1.1 in Chapter 1). In this Chapter, I examined the single nucleotide variants (SNVs) of Cemani chicken genome and applied homozygosity analyses in order to identify signs of selection in a particular region specific to Cemani chicken. I identified the genomic regions, which have been selected for and are associated with Cemani traits.

## 3.2 Materials and Methods

### *3.2.1 SNVs (Single Nucleotide Variations) data*

Single Nucleotides Variations (SNVs) data of Cemani, Silkie and L2 Taiwanese chicken used in this Chapter were obtained from the previous study in Chapter 2.

### *3.2.2 DNA samples*

DNA samples of Cemani chickens were used to analyze monomorphism in the Cemani-specific high homozygosity region (HHR). I sampled 12 individuals of Cemani (IDs Cemani 40-47, CM 11, 23, and 31), which were collected in Kedu, Central Java,

Indonesia, and DNA samples were extracted in the genetic laboratory, Museum Zoologicum Bogoriense (MZB), Indonesian Institute of Science (LIPI) using phenol-chloroform extraction method.

**Methods**

### *3.2.3 Comparison of high homozygosity region (HHR) in Cemani, Silkie and L2 Taiwanese chicken*

*High homozygosity region (HHR) in Cemani, Silkie and L2 Taiwanese*

SNVs data of Cemani, Silkie and L2 Taiwanese chicken were analyzed in order to obtain the HHRs from each breed. First, I extracted SNPs by removing insertion and deletions. Second, I identified the genotype (homozygoye or heterozygote) of each SNP site using ANNOVAR software [8]. After that, the ratio of homozygosity within 100 kb sliding windows of each breed was calculated using the following equation:

$$\text{Hom} = \frac{N_{hom}}{N_{het+hom}}$$

$N_{hom}$ = number of homozygous SNPs
$N_{het+hom}$ = total number of SNPs

*HHRs shared in Cemani, Silkie and L2 Taiwanese* and *HHR specific to Cemani chicken*

I compared the HHRs of Cemani, Silkie and L2 Taiwanese chicken in order to obtain HHRs shared in all three chicken breeds and HHRs specific to Cemani chicken. For obtaining such HHRs, a custom python script was written and executed. The shared regions between the three chicken breeds were extracted and then genes located within the shared region were identified. For detecting HHRs specific to Cemani, I excluded shared HHRs between Cemani, Silkie and L2 Taiwanese chicken. Then, HHRs shared between Cemani-Silkie chickens and Cemani-L2 Taiwanese chickens were also excluded. The remaining HHRs in Cemani were then referred to as HHRs specific to Cemani. This list of HHRs specific to Cemani was used for further analyses.

### *3.2.4 Identification and validation of genes in HHRs specific to Cemani.*

BLAST search (https://blast.ncbi.nlm.nih.gov) was used to identify genes located within the HHRs specific to Cemani. The identified genes were then validated using PCR and sequencing analysis in order to check the level of polymorphism in those genes. First, from the identified genes within the Cemani-specific HHRs, I selected several genes based on their functions. Second, using DNA samples of Cemani (n=12), I amplified 1kb length of the selected genes using PCR techniques and then directly sequenced. The gene names and primer sets are listed in S3.1 Table. PCR was performed in a 96-Well GeneAmp® PCR System 9700 from Applied Biosystems in initial denaturation step at 94$^{o}$C for 1 min followed by 35 cycles of 30 sec denaturation at 94$^{o}$C, 30 sec annealing at 56~58$^{o}$C, 1 min extension at 72$^{o}$C, and then a final extension step at 72$^{o}$C for 5 min. Amplified samples were then directly sequenced using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) on an ABI 3130xl sequencer (Applied Biosystems). The data sequences were assembled and analyzed using Codon Code Aligner software (https://www.codoncode.com/aligner/).

## 3.3. Results

### *3.3.1 High homozygosity regions (HHRs) in Cemani, Silkie and L2 Taiwanese chicken*

Including insertion deletions (INDELs)/substitutions, I identified a total of 7,543,624 SNVs in Cemani, 7,171,044 in Silkie, and 7,219,060 in L2 Taiwanese. After removal of INDELS, the total SNPs in Cemani, Silkie and L2 Taiwanese were 6,850,370, 6,506,777, and 6,503,381, respectively. These SNPs were then used for calculating homozygosity ratio in every 100 kb throughout the genome of a single individual each of Cemani, L2 Taiwanese and Silkie chickens. Only regions with homozygosity ratio above 0.95 were collated for downstream analyses. These regions were allocated as high

homozygosity regions (HHRs). Plot of homozygosity ratio on chromosome 1 of Cemani is shown in Fig 3.1 (for other chromosome see S3.1 Figure).

Analysis of high homozygosity regions (HHRs) from single nucleotide polymorphisms (SNPs) data of Cemani, Silkie and L2 Taiwanese revealed that in total there are 683, 2480, and 2505 HHRs in Cemani, Silkie and L2 Taiwanese, respectively. It is shown that Cemani has less HHRs compared with the two other chicken breeds, Silkie and L2 Taiwanese. The total number of HHRs in each chromosome after normalization with its chromosome length is shown in Fig 3.2.

### *3.3.2 Identification of High Homozygosity Regions (HHRs) shared in Cemani, Silkie and L2 Taiwanese*

Investigation of genes located within HHRs shared among the three breeds chicken revealed that there are 78 HHRs across eight chromosomes (chromosomes 1, 2, 3, 4, 6, 7, 16, 18, and 24). Blast search analyses showed that there were around 76 genes located in 78 HHRs shared among the three breeds. Among these genes, there are three genes (*SOX5, DAAM2, ARID4B*) that have already been reported with low heterozygosity [9]. Position of HHRs and genes located within HHRs shared in Cemani, Silkie and Taiwanese are listed in S3.2 Table.

### *3.3.3 High Homozygosity Region (HHRs) specific to Cemani*

Comparison of HHRs between Cemani, Silkie and L2 Taiwanese genomes obtained 249 HHRs specific to Cemani out of the 683 HHRs identified in Cemani. A single HHR is 100 kb in length, and several adjacent HHRs are called a consecutive region, of which varies in length from 200 kb to 1.2 Mb. In this study, I only focused on consecutive regions that are greater than 100 kb in length. The presence of consecutive regions in HHRs specific to Cemani was only detected in twelve chromosomes (see Table 3.1).

### *3.3.4 Identification and validation of genes in Cemani-specific HHRs*

Identification of genes located in the consecutive HHR regions (referred to as candidate of target region, CTR) as well its function determined using BLAST search (https://blast.ncbi.nlm.nih.gov) in twelve chromosomes showed that there are around 296 genes located in the CTR. Genes and its function are listed in S3.3 Table. Since I used only one individual of Cemani chicken to identify HHRs, it was then necessary to validate the homozygosity in the CTR using several individuals to ensure that the region is monomorphic among the Cemani population. For validation of homozygosity regions, I selected several genes based on their function. Firstly, I looked into genes belonging to the kinase family because these genes were distributed across ten (chromosomes 1, 2, 3, 4, 7, 10, 12, 13, 19 and 27) of twelve identified chromosomes. Secondly, I focused on several genes related to egg productivity traits on chromosome 23 (*BMP8A, PPIE, PABPC4, NT5C1A,* and *LOC419677*) reported from a previous study in Kauai chicken [10]. I used PCR technique and direct sequencing of 1 kb length of a total 28 selected genes in the Cemani population to clarify the homozygosity in CTR (PCR primers and conditions in S3.1 Table). This approach was also used for identification of surrounding gene regions for validation of monomorphism.

From the 28 selected genes, I found that only three genes were clarified as monomorphic in Cemani population, with the remaining genes classified as polymorphic (Table 3.2). For genes categorized in the protein kinase family, only one gene [*EGFR* (epidermal growth factor receptor)], was confirmed as monomorphic, which is located on chromosome 2. Meanwhile, from the five genes associated with fecundity on chromosome 23 [10] that I investigated, two genes (*NT5C1A* and *LOC419677*) were identified as monomorphic, while the other genes (*BMP8A, PPIE, PABPC4*) were polymorphic. I conducted further analysis by checking neighboring gene regions to

ensure that the signature of selective sweeps acted in *EGFR* on chromosome 2 and two genes, *NT5C1A* and *LOC419677,* on chromosome 23. The monomorphic analysis in surrounding region of candidate-selected gene (*EGFR* on chromosome 2 and *NT5C1A* and *LOC419677* on chromosome 23) revealed the signature of selective sweeps and consequently suggestive reduction of recombination (see Table 3.2, Fig 3.3 and 3.4).

## 3.4 Discussion

Through the process of natural (or artificial) selection, individuals with advantageous traits can survive and eventually produce offspring, and then from these offspring, populations with such traits evolve. Artificial selection in animals can also produce a population with specific traits according to human desires. This study intended to detect genes related to specific traits caused by artificial selection in domestic chicken using whole genome sequence data. Ideally, detection of selection signatures in the genome requires a sufficient sample size representing a population to be examined. However, since the invention of next generation sequencing technology, at least 4 - 6 samples is enough for representing a population when using a suitable statistical test [11]. Nevertheless, Fan (2013) only used whole genome sequence (WGS) data of two different individuals of two chicken breeds (Silkie and L2 Taiwanese) and could detect the signal of selection by scanning the entire genome and calculating heterozygosity values. Fan (2013) detected some candidate genes associated with domestication traits in chicken, with low heterozygosity values that have already been reported in other studies that used large sample sizes such as *TSHR, IGF1, PMCH, and TBC1D1* [9], *NELL1* [13] and *ESRP2* [14].

In this study, I only used whole genome sequence (WGS) of a single individual of Cemani chicken breed generated from NGS for studying the artificial selection in the Cemani breed. I also used WGS from single individuals of Silkie and L2 Taiwanese for

comparison. More than 6 million SNPs were extracted for each individual, reflecting that the genetic information from WGS of single individuals will give enough information for further analyses. Despite this, for detecting a signature of selection, the use of single individuals would still produce less reliable data. Therefore, in this study, to complement the single whole genome sequence data, and to make data more reliable, I validated the candidate regions of selective sweeps using several individuals by polymerase chain reaction (PCR) and direct sequencing approaches.

Homozygosity/heterozygosity approaches are widely used to discover putative regions under selection. Regions that may contain selected alleles are able to maintain homozygosity and these regions represent low nucleotide diversity or selective sweeps that also have an effect on surrounding regions. In this study, I used a simple equation (see Methods) to calculate homozygosity in order to find regions with low nucleotide diversity in the Cemani, Silkie and L2 Taiwanese chicken, and then compared these regions to extract Cemani-specific regions. Although only WGS data from one individual of each of the three breeds were used, detecting high homozygosity regions (HHRs) by scanning the entire genome of Cemani as well as Silkie and L2 Taiwanese chicken provided useful information. (i) Results of this study showed that Cemani has less HHRs compared with Silkie and L2 Taiwanese. Considering only a single individual was used in this study, it is likely that a large number of HHRs observed in Silkie and L2 Taiwanese could be attributed to random sampling. However, higher number of HHRs in L2 Taiwanese is also likely due to inbreeding. L2 Taiwanese is a type of commercial chicken breed known as a layer type chicken. It was reported that commercial chickens (broiler and layer) showed reduction of nucleotide diversity compared with non-commercial chickens due to a structured breeding system that only accommodated a limited number of chicken breeds that went into the formation of modern commercial

lines leading to inbreeding [15]. (ii) Homozygosity analyses identified that some HHRs of Cemani, Silkie and L2 Taiwanese were shared among the three breeds, while some of them were only specific to Cemani, Silkie and L2 Taiwanese. Three genes (*SOX5, DAAM2, ARID4B*) located in shared HHRs were detected in putative selective sweeps region in a previous study [9]. In addition, genes reported as a "domestication gene" in the chicken, such as *TSHR* and *BCDO2* [9,16], were also investigated in this analysis. These genes were suggested as "domestication genes" due to strong selective sweeps over the genes found in domesticated chickens, differentiating them from their ancestor. In this study, I found that *TSHR* was located in HHRs shared between Cemani and L2 Taiwanese chicken but not in Silkie. Meanwhile, *BCDO2* was located in HHRs shared between Silkie and L2 Taiwanese chicken but not in Cemani. Moreover, in HHRs specific to Cemani, I identified three genes on chromosome 1 (*CCDC53, MYBPC1, TBXAS1*) that also have been reported in the previous study [9].

Taken together, this study supports that the use of a single individual of WGS and a simple method for calculating homozygosity can provide reliable data for identifying regions of candidate selective sweeps. In addition, analysis of HHRs specific to individual breeds has detected the genomic regions containing gene(s), which may be associated with traits specific to Cemani, Silkie or L2 Taiwanese chickens, indicate that the breed-specific traits have possibly evolved independently. Therefore, I further examined the HHRs specific to Cemani for identifying candidate selective sweep regions related to Cemani's traits.

### *3.4.1 Genes located in High homozygosity regions (HHRs) specific to Cemani*

Investigation of genes located in high homozygosity regions (HHRs) specific to Cemani [candidate target region (CTR)] revealed that 10 of 12 chromosomes possess genes that were categorized into the protein kinase family. Protein kinase is an enzyme

that plays an important biological role in phosphorylating proteins that are essential for cellular activation processes like development and metabolism [17]. In addition, it was reported that Cemani have higher egg productivity compared with other Indonesian local chicken like Gaok and Merawang [1]. Hence, I extrapolated genes involved in the protein kinase family as they may contribute to cellular development in Cemani. In addition, I examined genes related to egg productivity which may be responsible for Cemani-specific traits.

The purpose of validating monomorphism in the Cemani population was to investigate whether CTRs had arisen by random chance in a single individual or actually due to the process of selection. In my study, only one gene (*EGFR*) was found to be monomorphic in the Cemani population from 24 genes identified as kinase family with high homozygosity. Similarly, on chromosome 23 from five egg production-related genes that were tested, only two genes (*NT5C1A* and *LOC419677*) were identified as monomorphic. Certainly, this study supports that validation of monomorphism in target regions is necessary when WGS from only a single individual is used. I further analyzed nucleotide diversity in regions surrounding *EGFR, NT5C1A* and *LOC419677* for confirming whether selective sweeps occurred on the target genes. I confirmed the presence of high homozygosity in neighboring target genes, indicating that *EGFR*, *NT5C1A* and *LOC419677* might be under selective sweeps (Fig 3.3 and 3.4).

### *3.4.2 Epidermal growth factor receptor (EGFR) as a candidate gene under selection*

*EGFR* is categorized into the tyrosine kinase family that is essential for cellular functions such as keratinocyte proliferation, differentiation and movement (Jost et al. 2000). Keratinocytes stimulate the proliferation of melanocytes, which are the pigment-producing cells derived from neural crest cells [19]. This gene was also suggested to be associated with human skin pigmentation in East Asia [20], since the gene showed

putative signatures of selective sweeps. In addition, another study revealed that polymorphisms in the genes *EGFR* and *OPRM1* were associated with skin pigmentation differences between Indigenous Americans and Europeans [21]. Therefore, high homozygosity in and surrounding *EGFR* in the Cemani could be correlated with its pigmentation in its entire body. This might be apart from the *EDN3* gene that has already been shown to be positively selected and correlated with skin and internal organ hyperpigmentation in the Cemani chicken [22. Chapter 2]. However, association between pigmentation and *EGFR* in Cemani chicken remains to be investigated.

Another important point to discuss is that *EGFR* has also been detected as a putative region under selection in commercial chicken breeds [9, 23]. Even though EGFR was reported to be under selection in commercial chicken breeds, there is a lack of information as to why it was under selection and uncertainty of the function of this gene. Nevertheless, it is known that *EGFR* is a transmembrane receptor that activates cell differentiation and proliferation by binding of one of its ligands; ligand binding receptor activates a tyrosine kinase and this tyrosine kinase phosphorylates a number of intracellular substrates that controls multiple aspects of cell and organism growth, differentiation, and function [24]. In mammals, it is reported that tyrosine kinase receptors play an important role as a controller for many biological processes such as development, differentiation, tissue repair and metabolic homeostatic mechanism [25]. In larval and pupal stages of Drosophila, *EGFR* controls growth, patterning and morphogenesis of multiple organs and structures, including wing veins and photoreceptor arrays [26,27]. Thus, it is possible that high homozygosity recorded in the *EGFR* locus of Cemani and commercial breeds may be related to the cellular growth and developmental process of these breeds.

Two different roles (cell pigmentation and cell growth controller) associated with *EGFR* indicate that this gene may have pleiotropic effects on phenotypic traits in Cemani and commercial chicken breeds. Cemani and commercial breeds may share a mutation in *EGFR* that has consequently undergone selective sweeps. However, this gene affects different characteristics in Cemani and commercial chicken because of differences in breeding histories. Therefore, future research about genetic mechanisms, such as pleiotropy in *EGFR* and how *EGFR* can affect several traits, is needed.

### *3.4.3 NT5C1A and LOC419677 as a candidate locus under selection*

I have identified high homozygosity in and surrounding the *NT5C1A* and *LOC419677* locus on chromosome 23 within the Cemani population, supporting the hypothesis of selective sweeps would occur in flanking regions with selected genes. It is reported that *NT5C1A* and *LOC419677* also showed low heterozygosity in layer chickens [9]. Interestingly, these genes were also candidate genes in and around putative sweeps in Kauai chicken and were reported having significant associations with fecundity traits in the wild and domestic intercross chickens [10], including total egg production (*LOC419677*), mean egg weight (*NT5C1A*) and egg number (*LOC419677*). The ability for high egg productivity in Cemani might be associated with *NT5C1A* and *LOC419677*, and this may also be the case for layer chicken. Selective sweeps in these genes within layer chickens seems conceivable because layer chickens were selected for the purpose of commercial egg production in the poultry farming industry, and as a result, the layer breeds have higher egg productivity compared with other chickens. However, in this study, the HHRs containing *NT5C1A* and *LOC419677* were not identified in L2 Taiwanese chicken, which is one type of layer chicken breeds. It might be that for L2 Taiwanese chicken, other genes that are related to productivity in layer chickens such as *PRLR*, *IGF1R, ITPR2* and *VIPR1* [6] could be responsible for traits in L2 Taiwanese

chicken. Furthermore, more research about this gene and its functional role in egg productivity in the Cemani population is necessary.

## 3.5 Conclusion

I have found that *EGFR, NT5C1A* and *LOC419677* showed high homozygosity in Cemani chickens and these genes share some similar selection patterns with other chicken breeds such as commercial chickens and Kauai chickens. Investigation of gene function indicates that these genes may be related to Cemani-specific traits such as black pigmentation and/or highly egg productivity. In addition, genes related to fecundity traits that might be selected in Cemani chickens. Altogether, the identification of these genes under selection supports that Cemani chicken is an Indonesian local chicken breed with desirable qualities related to productivity and genetic potential that could be applied to the commercial chicken industry.

## 3.6 Reference


1. Sulandari S, Zein MSA, Paryanti S, Sartika T, Astuti M, Widjastuti T, et al. Sumberdaya genetik ayam lokal Indonesia. In: Keanekaragaman Sumber Daya Hayati Ayam Lokal Indonesia: Manfaat dan Potensi. (Diwyanto, K and S. N. Prijono, Eds). Jakarta: 1st Ed. LIPI Press; 2007: 43-104.
2. Stainton JJ, Haley CS, Charlesworth B, Kranis A, Watson K, Wiener P. Detecting signatures of selection in nine distinct lines of broiler chickens. Anim Genet. 2015; 46: 37–49.
3. Gholami M, Erbe M, Gärke C, Preisinger R, Weigend A, Weigend S, et al. Population genomic analyses based on 1 million SNPs in commercial egg layers. PLoS ONE. 2014; 9:e94509.
4. Gholami M, Reimer C, Erbe M, Preisinger R, Weigend A, Weigend S, et al. Genome scan for selection in structured layer chicken populations exploiting linkage disequilibrium information. PLoS ONE. 2015; 10:e0130497.
5. Fu W, Lee WR, Abasht B. Detection of genomic signatures of recent selection in commercial broiler chickens. BMC Genetics. 2016; 17(1): 122.
6. Boschiero C, Moreira GCM, Gheyas AA, Godoy TF, Gasparin G, Mariani PDSC, et al. Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. BMC Genomics. 2018; 19: 83.
7. Dorshorst B, Molin AM, Rubin CJ, Johansson AM, Strömstedt L, Pham M-H, et al. A complex genomic rearrangement involving the Endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet. 2011; 7: e1002412.

8. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010; 38(16): e164.
9. Rubin CJ, Zody MC, Erikson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010; 464: 587-591.
10. Johnsson M, Gering E, Willis P, Lopez S, Van Dorp L, Hellenthal G, et al. Feralisation targets different genomic loci to domestication in the chicken. Nature Comm. 2016; 7: 12950.
11. Willing E-M, Dreyer C, van Oosterhout C. Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. PLoS ONE. 2012; 7(8): e42649. http://doi.org/10.1371/journal.pone.0042649
12. Fan W-L, Ng CS, Chen C-F, Lu M-YJ, Chen Y-H, Liu C-J, et al. Genome-wide patterns of genetic variation in two domestic chickens. Genome Biol Evol. 2013; 5: 1376-1392.
13. Elferink MG, Megens H-J, Vereijken A, Hu X, Crooijmans RPMA, Groenen MAM. Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken Breeds. Shioda T, ed. PLoS ONE. 2012; 7(2): e32720.
14. Zhang H, Wang S-Z, Wang Z-P, Da Y, Wang N, Hu X-X, et al. A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. BMC Genomics. 2012; 13: 704.
15. Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, et al. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. Proc Natl Acad Sci U S A. 2008; 105(45): 17312-17317.
16. Girdland Flink L, Allen R, Barnett R, Malmström H, Peters J, Eriksson J. et al. Establishing the validity of domestication genes using DNA from ancient chickens. Proc Natl Acad Sci U S A. 2014; 111(17): 6184-6189.
17. Hamilton, TA. Protein Kinase. Encyclopedia of Immunology 2nd edition; 1998; 2028-2033.
18. Jost M, Kari C, Rodeck U. The EGF receptor–an essential regulator of multiple epidermal functions. Eur J Dermatol. 2000; 10: 505–510.
19. Hirobe Tomohisha. Keratinocytes regulate the function of melanocytes. Dermatologica Sinica. 2014; 32(4): 200-204. https://doi.org/10.1016/j.dsi.2014.05.002.
20. Hider JL, Gittelman RM, Shah T, Edwards M, Rosenbloom A, Akey JM, et al. Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. BMC Evol Biol. 2013; 13:150.
21. Quillen EE, Bauchet M, Bigham AW, Delgado-Burbano ME, Faust FX, Klimentidis YC, et al. OPRM1 and EGFR contribute to skin pigmentation differences between Indigenous Americans and Europeans. Hum Genet. 2011; 13: 1073–1080.
22. Dharmayanthi AB, Terai Y, Sulandari S, Zein MSA, Akiyama T, Satta Y. The Origin and Evolution of Fibromelanosis in domesticated chicken: Genomic Comparison of Indonesian Cemani chicken and Chinese Silkie Breeds. Plos ONE. 2017; 12(4): e0173147. doi: 10.1371/jpurnal.pone.0173147.
23. Zhang X, Misztal I, Heidaritabar M, Bastiaansen JWM, Borg R, Okimoto R, et al. Prior genetic architecture impacting genomic regions under selection: An example using genomic selection in two poultry breeds. Livestock Science. 2015; 171: 1-11.

24. Voldborg BR, Damstrup L, Spang-Thomsen M, Poulsen HS. Epidermal growth factor receptor (EGFR) and EGFR mutations, function and possible role in clinical trials. Annals of Oncology. 1997; 8(12): 1197-1206.
25. Ullrich A, Schlessinger J. Signal transduction by receptors with tyrosine kinase activity. Cell. 1990; 61(2): 203-12.
26. de Celis JF. Pattern formation in the Drosophila wing: The development of the veins. BioEssays. 2003; 25: 443-451.
27. Roignant J-Y, Treisman JE. Pattern formation in the Drosophila eye disc. The International journal of developmental biology. 2009; 53(5-6): 795-804.

## 3.7 Figures and Table

**Fig 3.1. Homozygosity ratio plot of Cemani chromosome 1 using 100kb sliding window widths**. Blue dots represent homozygosity ratios and red bar line represents homozygosity ratio with value of 0.95.

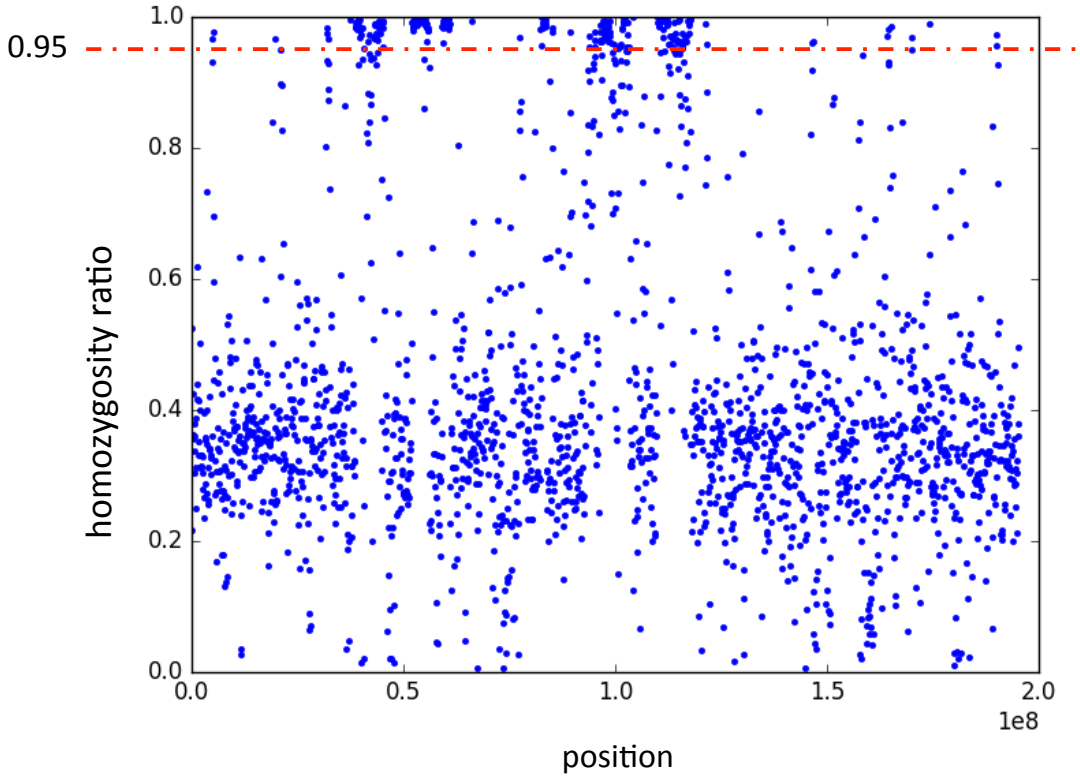


**Fig 3.2. Total number of high homozygosity regions (HHRs) in each chromosome.**

Cemani (red bar), Silkie (green bar), and L2 (purple bar).

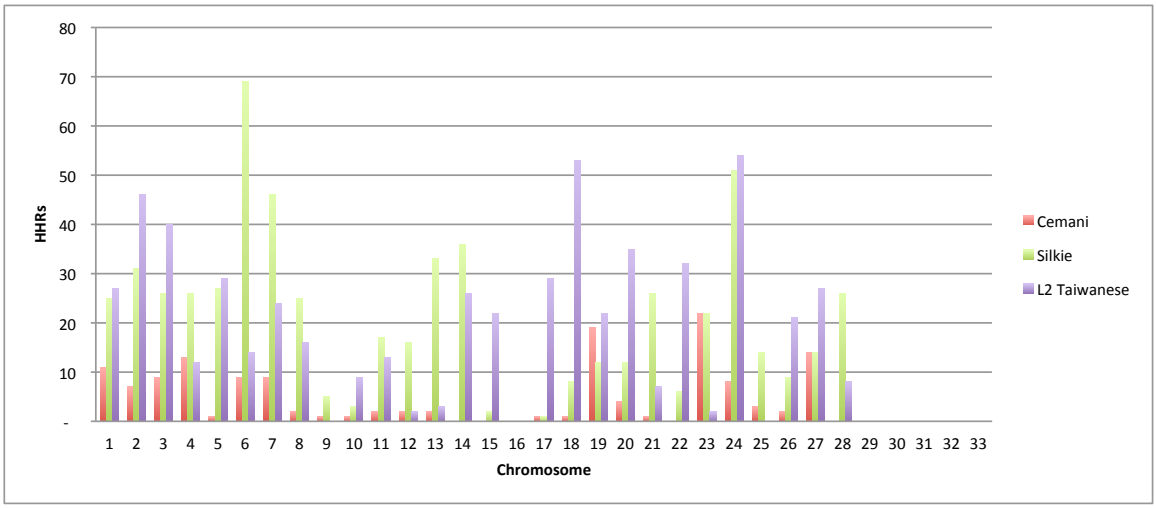


**Fig 3.3. Summary result of homozygosity analysis on position 52374486-52587311 in chromosome 2.** Monomorphism is displayed in and around *EGFR* and polymorphism displayed in the region far from *EGFR* as checked by using 1kb sequences of each gene.

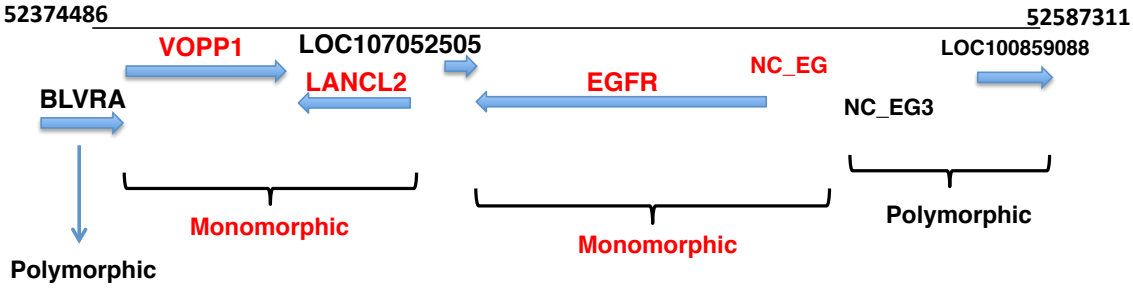



**Fig 3.4. Summary result of homozygosity analysis on position 5287786 - 5485334 in chromosome 23.** Monomorphism is found in and around *NT5C1A* and LOC419677, and polymorphism displayed in the region far from *EGFR*, as checked by using 1kb sequences of each gene.

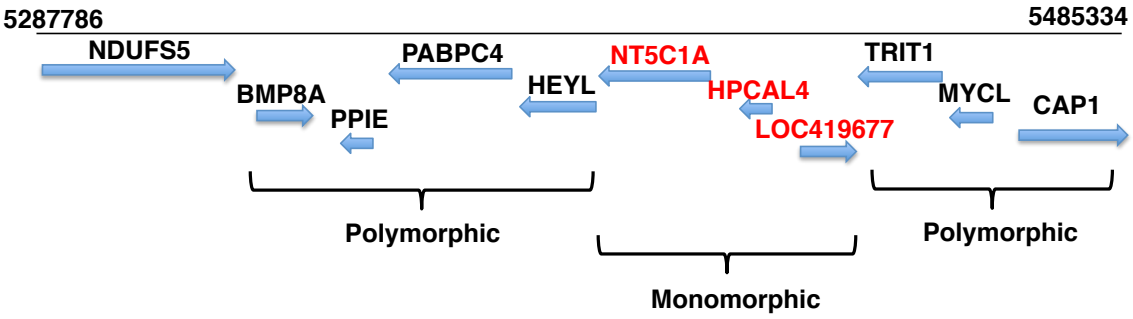


**Table 3.1.** Total number of HHRs specific to Cemani in each chromosome

| Chromosome | Total number of Cemani-specific HHRs (> 100kb) |
|---|---|
| 1 | 51 |
| 2 | 28 |
| 3 | 45 |
| 4 | 74 |
| 7 | 8 |
| 10 | 2 |
| 12 | 3 |
| 13 | 2 |
| 14 | 2 |
| 19 | 17 |
| 23 | 13 |
| 27 | 4 |
| Total | 249 |

**Table 3.2.** Genes identified inHHRs specific to Cemani and list of validated genes

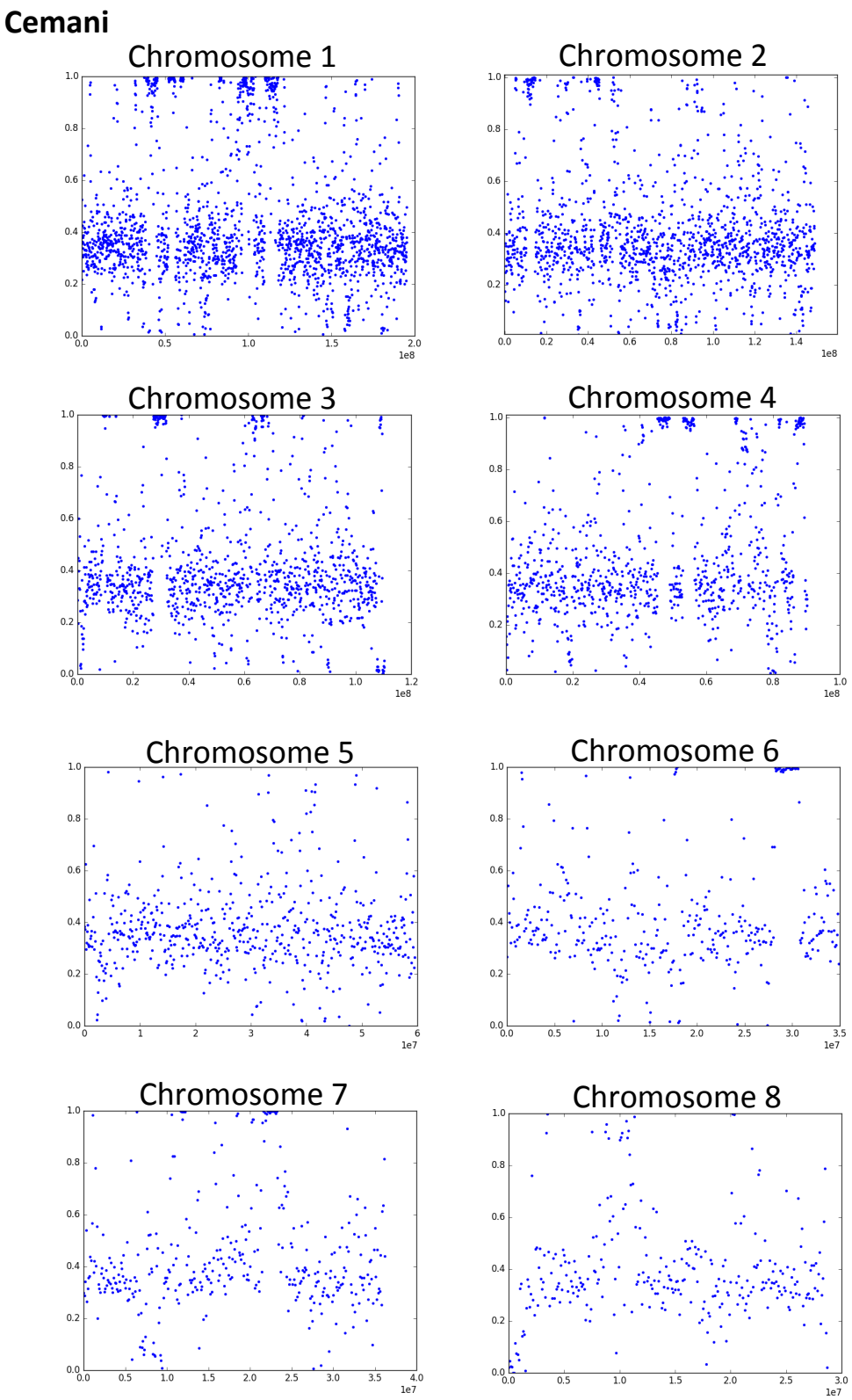
| Chromosome | Gene detected | | Gene name | Result |
|---|---|---|---|---|
| | Total gene | Validated | | |
| 1 | 91 | 3 | GNPTAB("N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits") | polymorphic |
| | | | HIPK2 ( "homeodomain interacting protein kinase 2") | polymorphic |
| | | | IGF1 | polymorphic |
| 2 | 23 | 8 | BLVRA | polymorphic |
| | | | EGFR (epidermal growth factor receptor) | **monomorphic** |
| | | | LANCL2 | **monomorphic** |
| | | | LOC100859088 | **monomorphic** |
| | | | NC_EG (non coding region) | **monomorphic** |
| | | | NC_EG2 (non coding region) | **monomorphic** |
| | | | NC_EG3 (non coding region) | polymorphic |
| | | | VOPP1 | **monomorphic** |
| | | | TRT1 | polymorphic |
| 3 | 41 | 3 | ROS1 (c-ros oncogene 1, receptor tyrosine kinase) | polymorphic |
| | | | DSE(dermatan sulfate epimerase) | polymorphic |
| | | | FIG4(polyphosphoinositide phosphatase precursor) | polymorphic |
| 4 | 68 | 3 | MAPK10 (mitogen-activated protein kinase 10) | polymorphic |
| | | | DUSP4 (dual specificity phosphatase 4) | polymorphic |
| | | | ANXA5(annexin A5); | polymorphic |
| 7 | 13 | 2 | SMARCAL1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a-like 1) | polymorphic |
| | | | MAP3K2 (mitogen-activated protein kinase kinase kinase 2) | polymorphic |
| 10 | 6 | 1 | MAP2K5(mitogen-activated protein kinase kinase 5) | polymorphic |
| 12 | 4 | 1 | PHF2 | polymorphic |
| 13 | 6 | 1 | PPP2R2B(serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B beta) | polymorphic |
| 14 | 2 | none | None | - |
| 19 | 42 | 4 | GEMIN4 (component of gems 4); | polymorphic |
| | | | OPNP(opsin 1 (cone pigments), long-wave-sensitive) | polymorphic |
| | | | RPS6KB1(ribosomal protein S6 kinase, 70kDa, polypeptide 1) | polymorphic |
| | | | TBX4 (T-box 4) | polymorphic |
| 23 | 75 | 14 | PPIE (peptidylprolyl isomerase E (cyclophilin E)) | polymorphic |
| | | | PABPC4 (poly(A) binding protein, cytoplasmic 4 (inducible form)) | polymorphic |
| | | | NT5C1A (5'-nucleotidase, cytosolic IA) | **monomorphic** |
| | | | MYCL1 (v-myc myelocytomatosis viral oncogene homolog 1,lung carcinoma derived) | polymorphic |
| | | | LOC419677 (protein FAM49A-like) | **monomorphic** |
| | | | CAP1 (CAP, adenylate cyclase-associated protein 1 (yeast)) | polymorphic |
| | | | BMP8A | polymorphic |
| | | | HPCAL4 | **monomorphic** |
| | | | TRIT1(2) | **monomorphic** |
| | | | MYCL1 | polymorphic |

| | | | | |
|---|---|---|---|---|
| | | | HEYL | polymorphic |
| | | | NC_HN | **monomorphic** |
| 27 | 8 | 2 | DCAF7 (DDB1 and CUL4 associated factor 7) | polymorphic |
| | | | IGFBPF | polymorphic |
| Total | 379 | 42 | | |

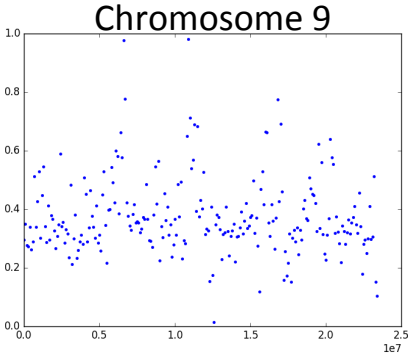
Note: gene in region surrounding *EGFR* are represented with light blue background and gene in region surrounding *NT5C1A* and *LOC419677* represented with light green background behind text.

## 3.8 Supporting Information

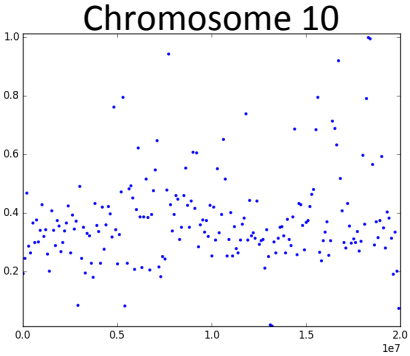
**S3.1 Figure**. **Plot homozygosity ratio of Cemani, Silkie and L2 Taiwanese on chromosome 1-28 using 100kb sliding window width**. Blue dot is the homozygosity ratio.



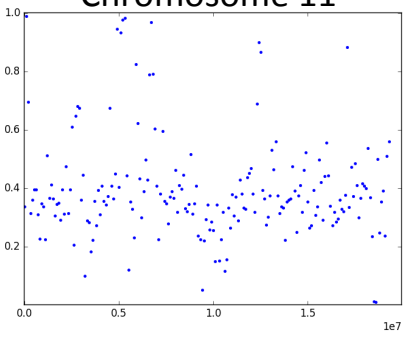
Chromosome 9



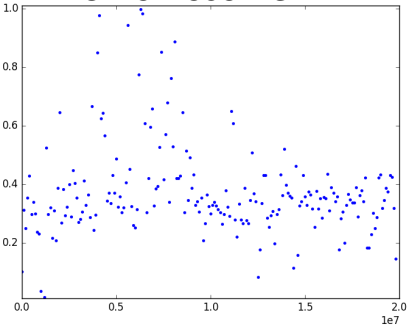

Chromosome 10



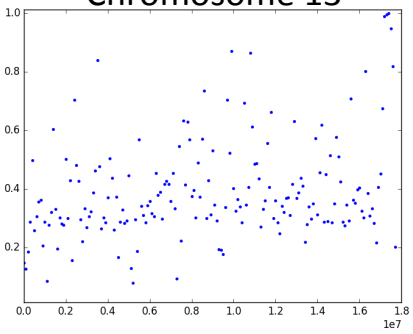

Chromosome 11



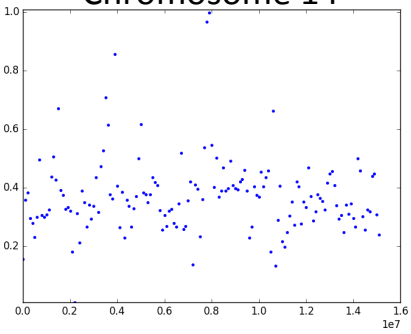

Chromosome 12



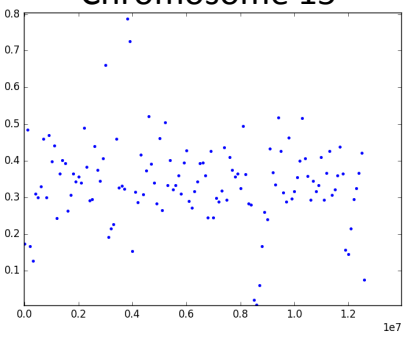

Chromosome 13



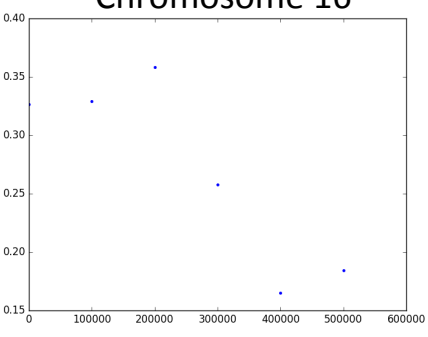

Chromosome 14

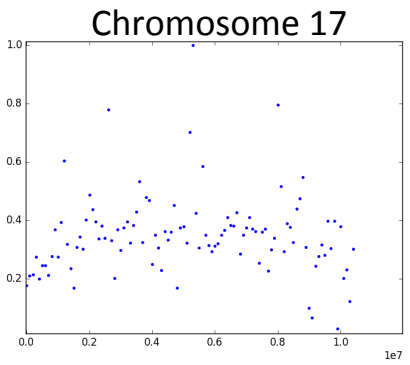


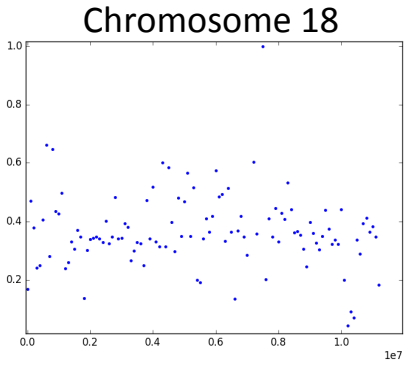

Chromosome 15

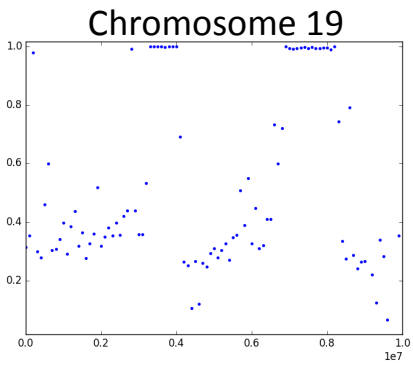


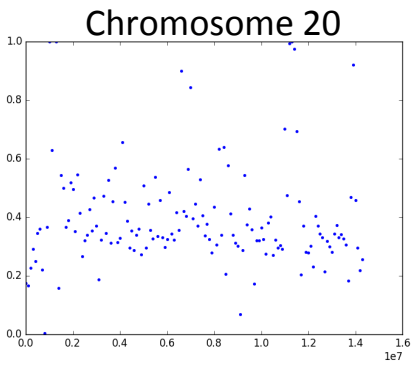

Chromosome 16

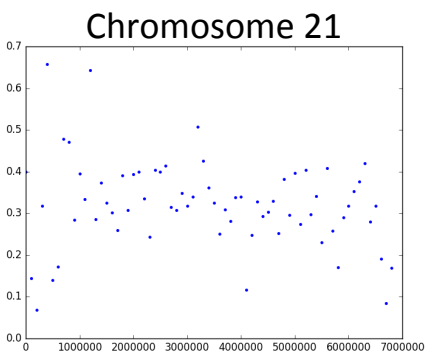


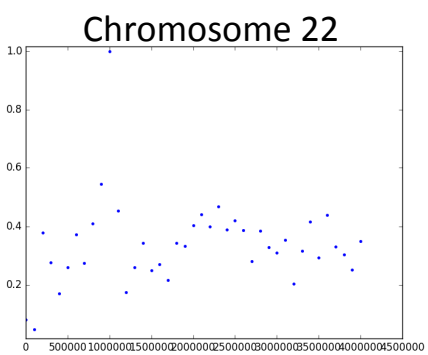

Chromosome 17
1e7

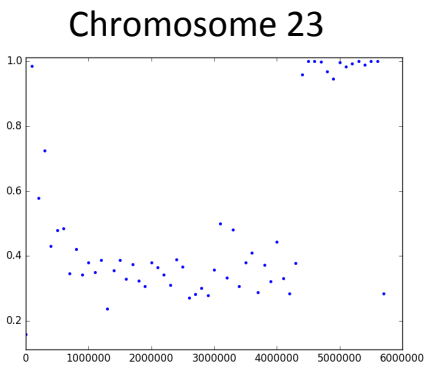

Chromosome 18
1e7

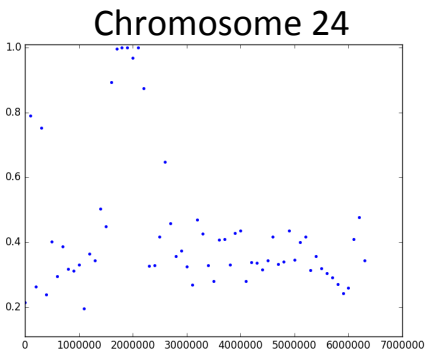

Chromosome 19
1e7

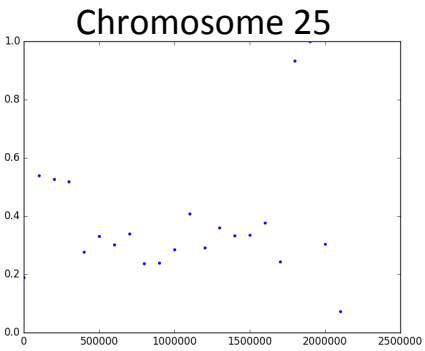

Chromosome 20
1e7

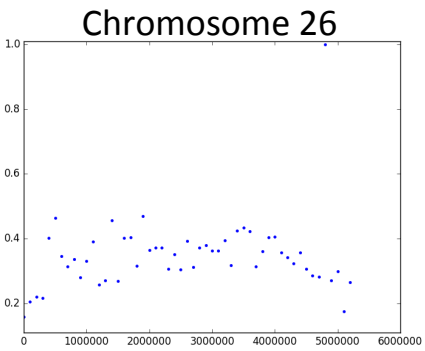

Chromosome 21

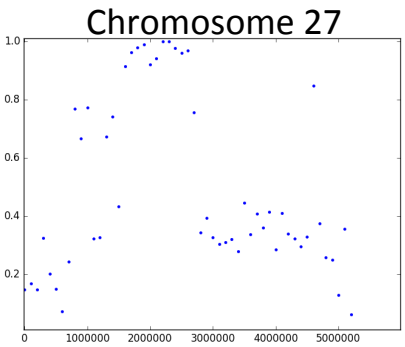

Chromosome 22

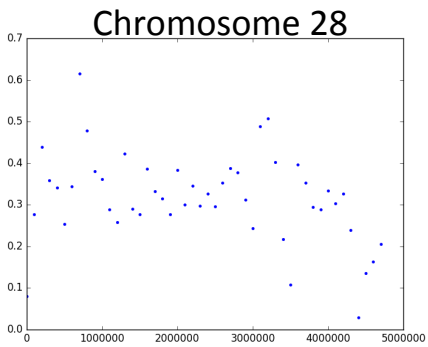

Chromosome 23


Chromosome 24

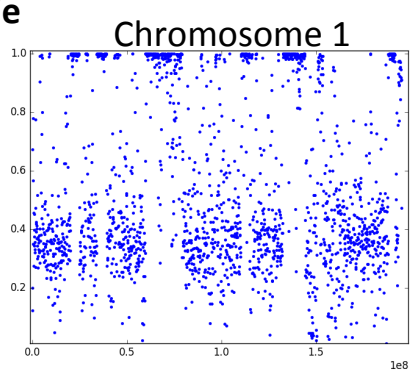

Chromosome 25
1.0
0.8
0.6
0.4
0.2
0.0
0
500000
1000000
1500000
2000000
2500000

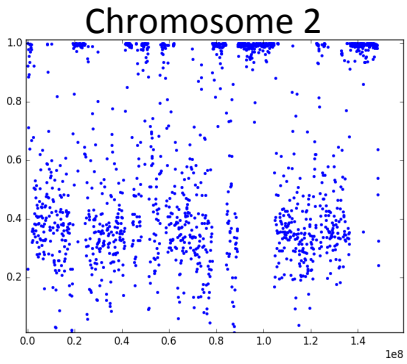

Chromosome 26
1.0
0.8
0.6
0.4
0.2
0
1000000
2000000
3000000
4000000
5000000
6000000

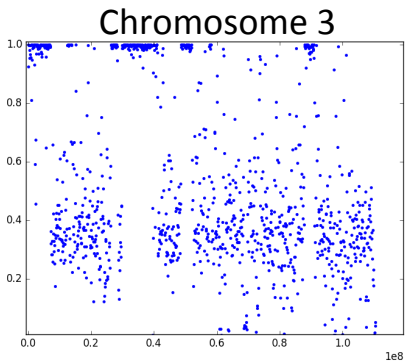

Chromosome 27
1.0
0.8
0.6
0.4
0.2
0
1000000
2000000
3000000
4000000
5000000

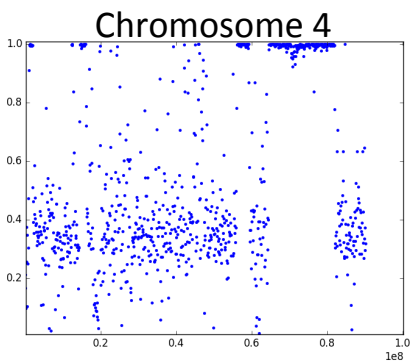

Chromosome 28
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1000000
2000000
3000000
4000000
5000000

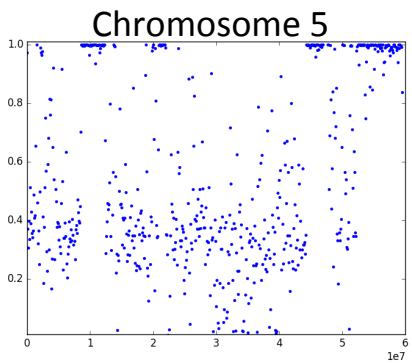
**Silkie**

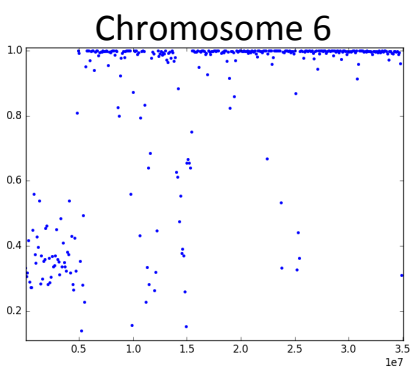

Chromosome 1
1e8

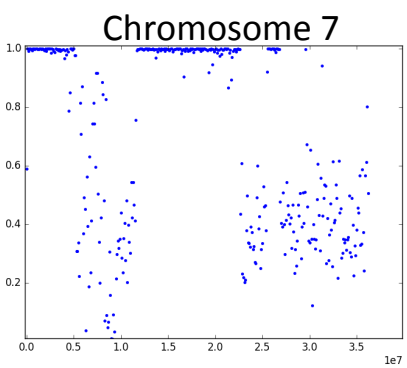

Chromosome 2
1e8

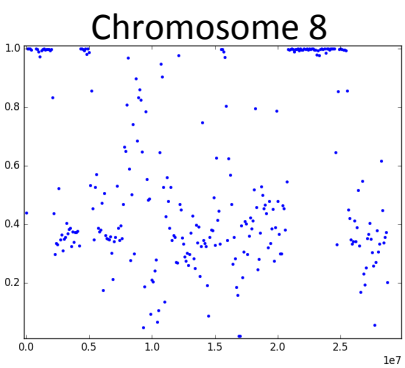

Chromosome 3
1e8


Chromosome 4
1e8

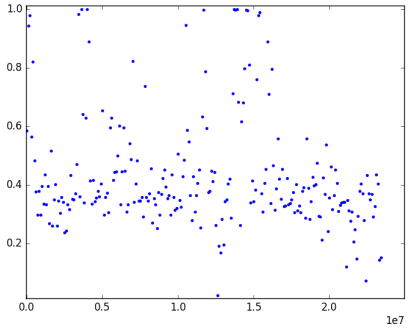

Chromosome 5
1e7


Chromosome 6
1e7

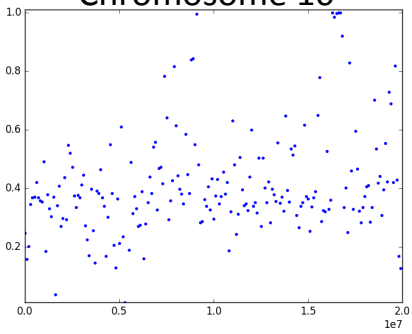

Chromosome 7
1e7


Chromosome 8
1e7

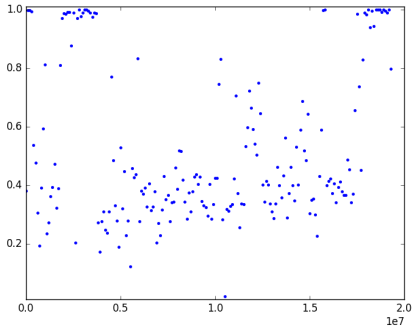
Chromosome 9



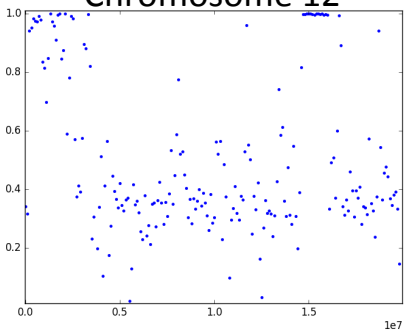

Chromosome 10



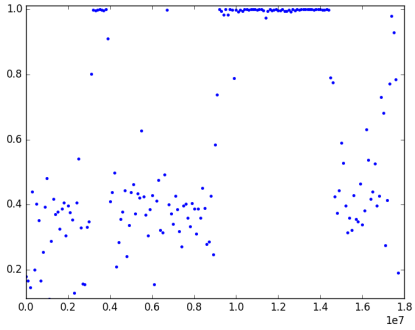

Chromosome 11



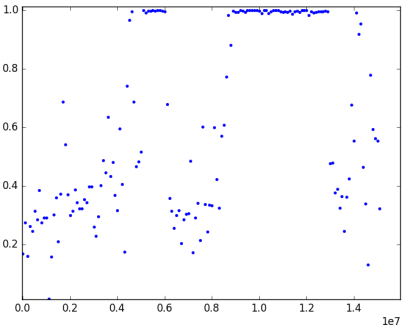

Chromosome 12



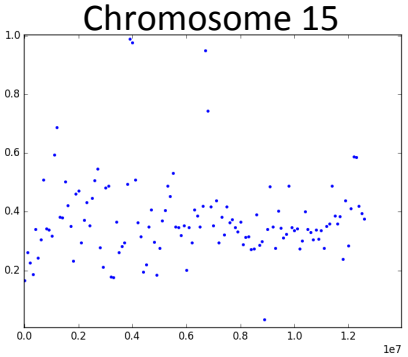

Chromosome 13



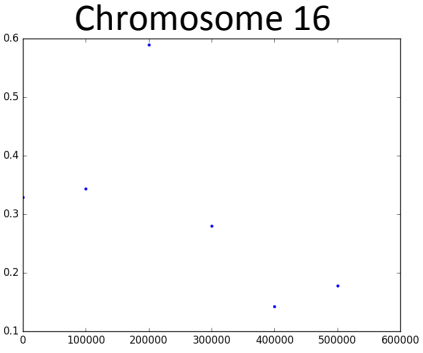

Chromosome 14



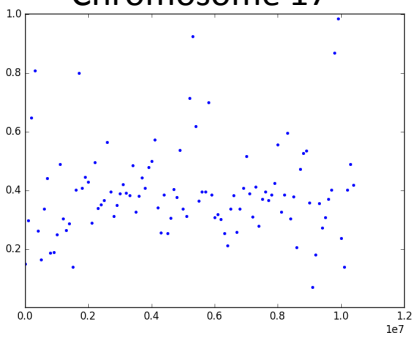

Chromosome 15



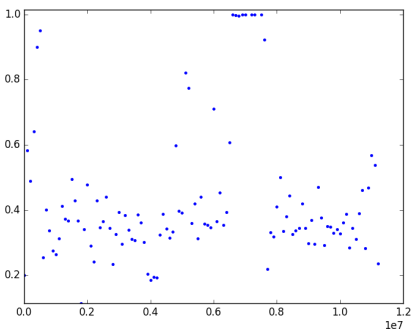

Chromosome 16



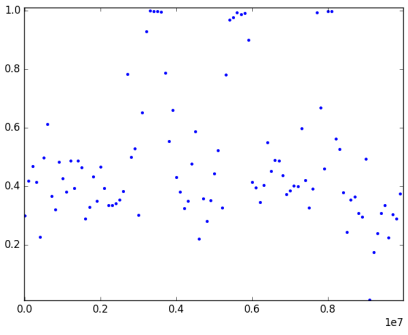
Chromosome 17



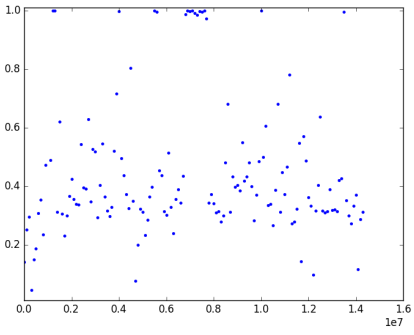

Chromosome 18



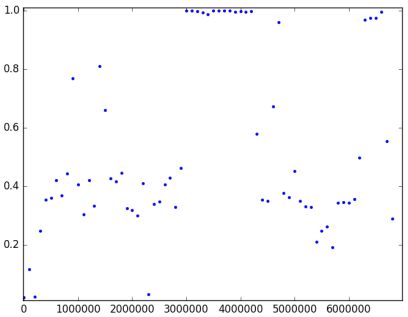

Chromosome 19



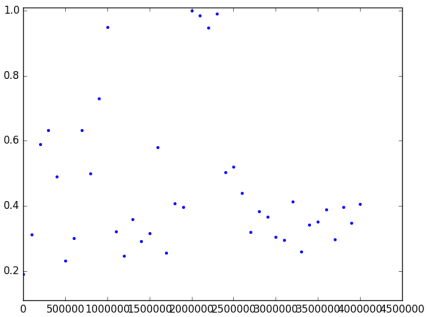

Chromosome 20



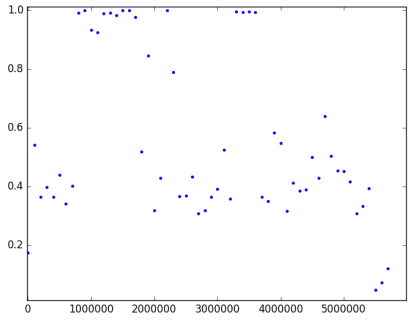

Chromosome 21



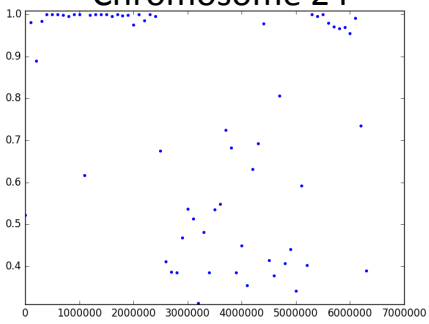

Chromosome 22



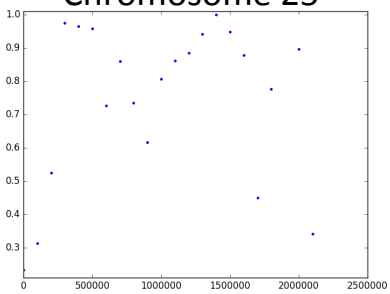

Chromosome 23



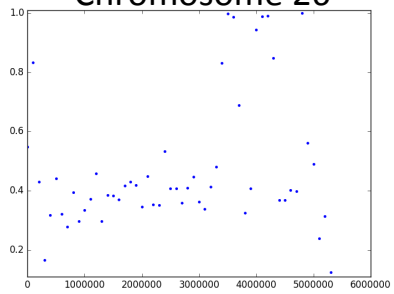

Chromosome 24



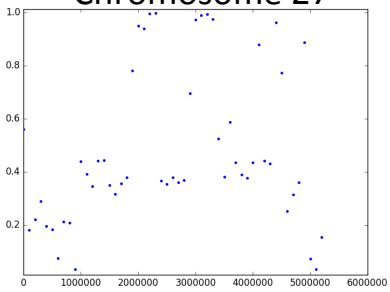
Chromosome 25



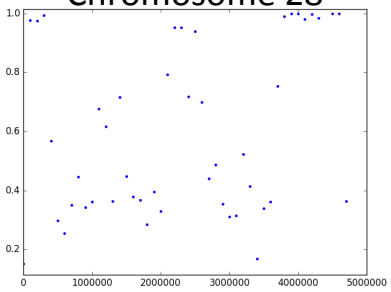

Chromosome 26



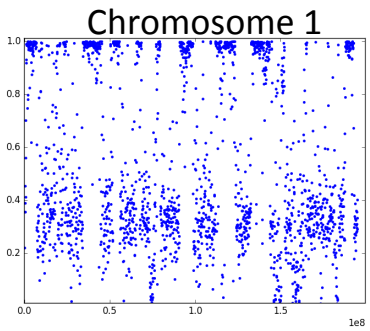

Chromosome 27

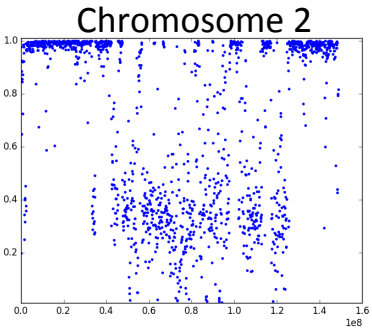


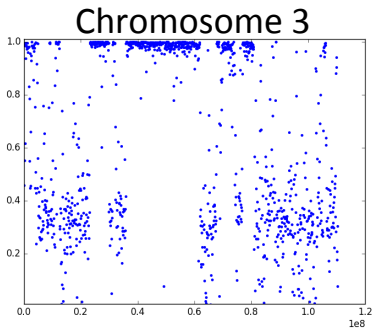

Chromosome 28

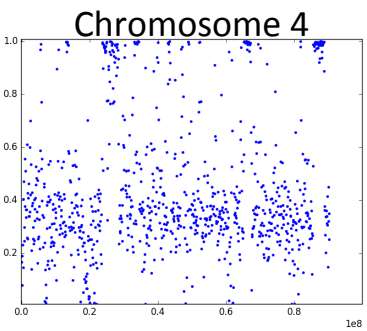


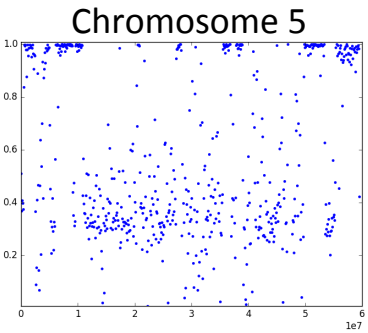
## L2 Taiwanese

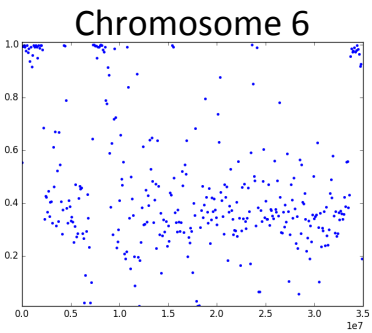


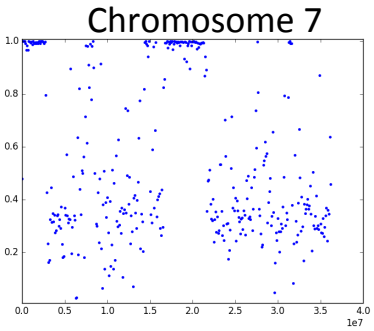



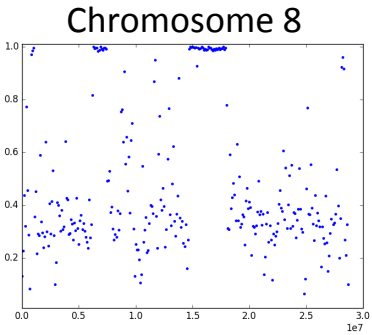



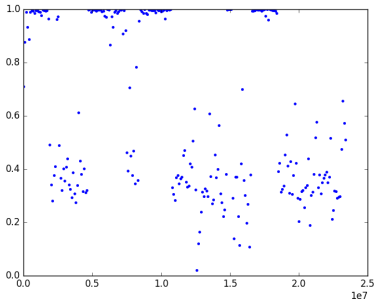



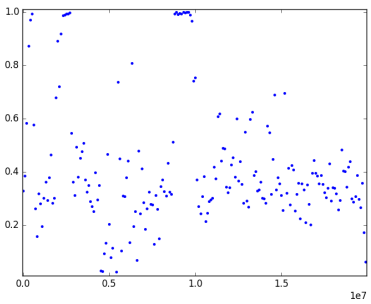



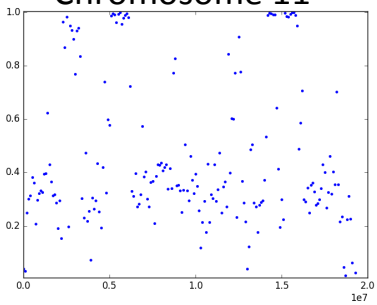
Chromosome 9



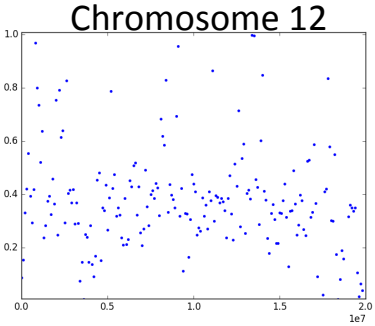

Chromosome 10



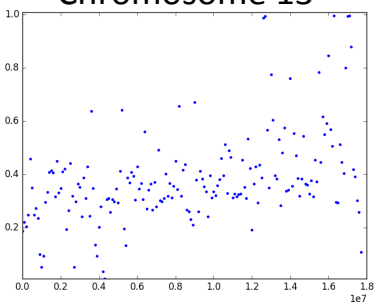

Chromosome 11



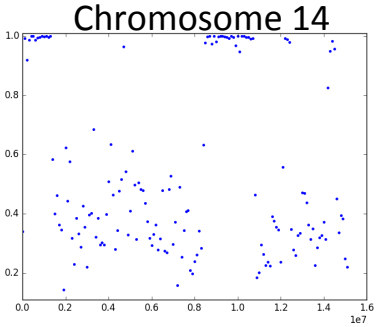

Chromosome 12



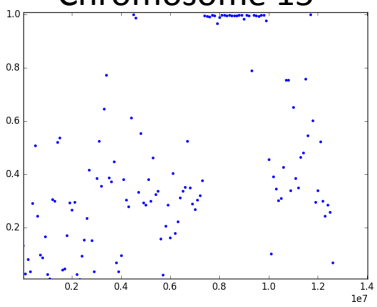

Chromosome 13



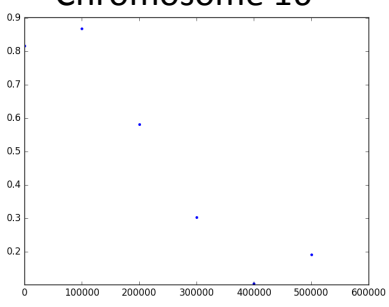

Chromosome 14



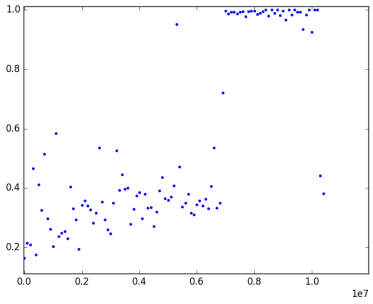

Chromosome 15



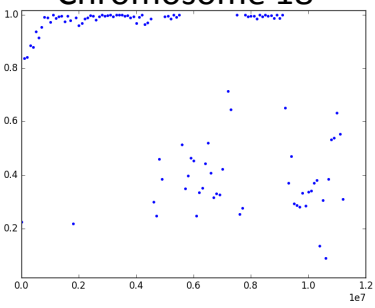

Chromosome 16



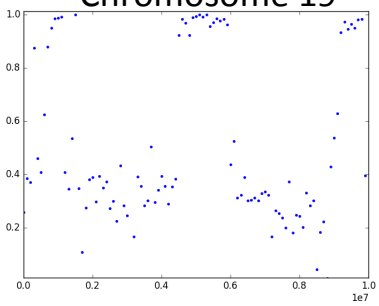
Chromosome 17



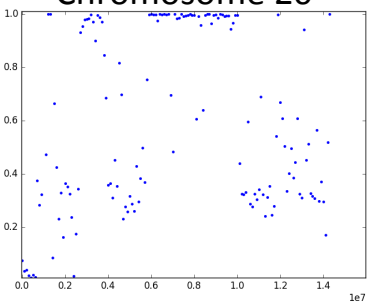

Chromosome 18



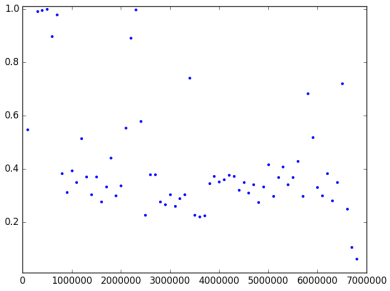

Chromosome 19



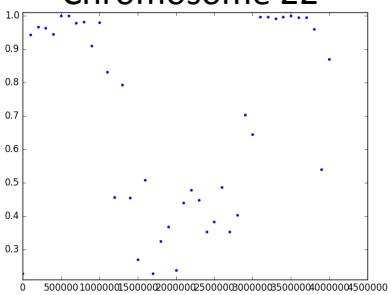

Chromosome 20



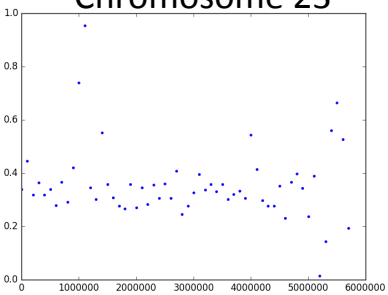

Chromosome 21



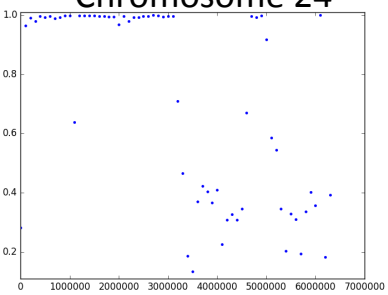

Chromosome 22

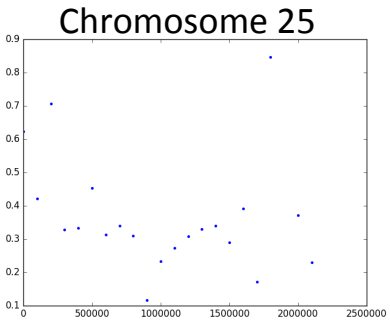


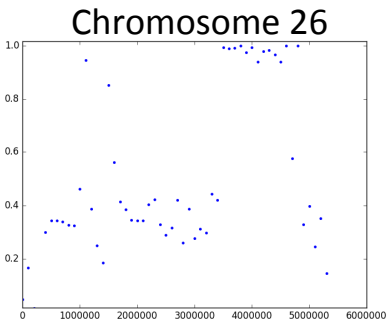

Chromosome 23

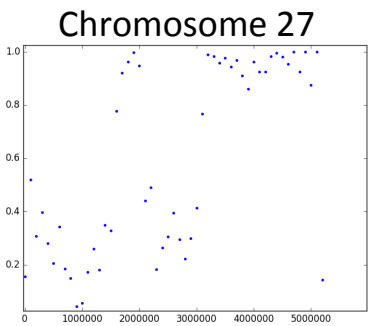


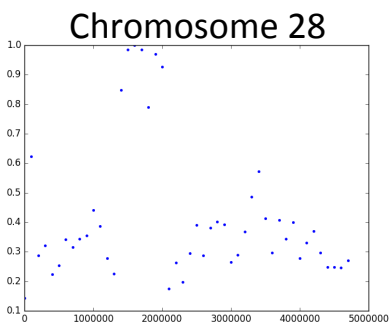

Chromosome 24




Chromosome 25
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
500000
1000000
1500000
2000000
2500000


Chromosome 26
1.0
0.8
0.6
0.4
0.2
0
1000000
2000000
3000000
4000000
5000000
6000000


Chromosome 27
1.0
0.8
0.6
0.4
0.2
0
1000000
2000000
3000000
4000000
5000000


Chromosome 28
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1000000
2000000
3000000
4000000
5000000

**S3.1 Table**. List of primer and primer sequence

| Chromosome | Primer name | Primer F sequence | Primer R sequence |
|---|---|---|---|
| Chromosome 1 | GNPTAB | 5'-TAGCTGGAGGAGTCAAAGCAT-3' | 5'-AAGCAGGCTGTACAACTTCAGA-3' |
| | HIPK2 | 5'-TTGCTGATGTGCATGTGAGA-3' | 5'-AATGCAAAGCAGAAGCAACA-3' |
| | IGF1 | 5'-TAGTGCCATTCCCTTGCATAC-3' | 5'-AGCCTATCACCTTAAGCATGGA-3' |
| Chromosome 2 | BLVRA | 5'-TTCTCGCTAAGCAGTTTCCA-3' | 5'-CTCATCTGGACTAACTGGATTAGG-3' |
| | EGFR2 | 5'-GGGATTTGCTGAACTTCAGTG-3' | 5'-CTTTCCCAGCTTGAATCCAT-3' |
| | LANCL2 | 5'-CTTCCTGCATTGGTTTGGTT-3' | 5'-CTGCACAGAAGTCTCCACCTTA-3' |
| | LOC100859088 | 5'-CAGCAGATGATGGAGAATGAG-3' | 5'-ATTGCTCTTGCTCCTTGACA-3' |
| | NC_EG | 5'-CTAATGAAATTGCTGGGAAGGT-3' | 5'-AATAGGGAGATCCCAGAAATCC-3' |
| | NC_EG2 | 5'-AATGGCAAACAGTGCACTCT-3' | 5'-GCTAACAATGAGGCTACAGCAT-3' |
| | NC_EG3 | 5'-TACTGGGAAACATGCTCCTTC-3' | 5'-CAGTCCGTCTCCTGTAACGTAT-3' |
| | VOPP1 | 5'-GATACCAGGATTTTCACATTGC-3' | 5'-GAACTTCAATAGCGCTCAGAAC-3' |
| Chromosome 3 | DSE | 5'-GAGCAGGGAAATGTAGTGCAT-3' | 5'-GCAGATACTGAATTCATCCTCTGA-3' |
| | FIG4 | 5'-TGGTGTTGAGCTGTTGGAAT-3' | 5'-AGCCAAAGGATACAAGCTCTCT-3' |
| | ROS1 | 5'-GCTACGGAGAATTCAGTCATCA-3' | 5'-AATGAGATAGGTCCCAGCAAAG-3' |
| Chromosome 4 | IGFBP7 | 5'-CTCGGAGACAAACTGACATCA-3' | 5'-GAACACGAGGCAATGTCAGT-3' |
| | ANXA5 | 5'-TTGTCCTTCTGCAAACTTCTG-3' | 5'-TCAGTTATCCTGGTCTCAGCTT-3' |
| | MAPK10 | 5'-ACTGCTGTATGACAGCAAACG-3' | 5'-TCCTGCTGACAAATTGAGACA-3' |
| Chromosome 7 | SMARCAL1 | 5'-AGGCTGAATCTCTAGCATCTGA-3' | 5'-CCACTATGGCTTTCAATCCA-3' |
| | MAP3K2 | 5'-TTTCACTCCTGAGCACTGAAG-3' | 5'-GTCATCTCCAAAGTGGTTTCAG-3' |
| Chromosome 10 | MAP2K5 | 5'-AGCACATATGCTCCACTCACA-3' | 5'-TGTGCAGGATTTAGCAATGC-3' |
| Chromosome 12 | PHF2 | 5'-GTCCACAGAAATGGTTCTGAC-3' | 5'-AAGACATCGACTGAGGAATCC-3' |
| Chromosome 13 | PPP2R2B | 5'-GCACAGAAACGTACGCTTGT-3' | 5'-CTCCAACCTTCATCTCATCCT-3' |
| Chromosome 19 | TBX4(3) | 5'-CCCTGACTTGCTGACTCAAA-3' | 5'-GTTTGCAGGCAGATCACAGT-3' |
| | RPS6KB | 5'-TTGAGTATCTCAGCGGTCAGTA-3' | 5'-GGCTACCTTAGATTCCTCCAA-3' |
| | OPNP | 5'-CCAACAACATCAATGGCTTCT-3' | 5'-GACAAACAGGGACAGGATGT-3' |

| | | | |
|---|---|---|---|
| Chromosome 23 | HPCAL4 | 5'-ATGCTGGAGATCATTGAGGTACT-3' | 5'-TCAGGGTAGTTCTGGAAGTCATC-3' |
| | TRIT1(2) | 5'-TCTGAGCACATCCAGATGATTC-3' | 5'-AATCGAAACACGGTGAGAAGA-3' |
| | MYCL1 | 5'-GTGGCTATGGAAGGAAATCTGT-3' | 5'-AACATGAGCCACACAGTGTGA-3' |
| | CAP1(2) | 5'-TGTAGGATTGGCTGCTCTGT-3' | 5'-TTCCTGCCATTAAGCACATC-3' |
| | HEYL | 5'-GTGATGCAGAGGTATCTGTGTTG-3' | 5'-TCCGTGATTCTCATCTCAGTGT-3' |
| | LOC419677 | 5'-ACATGCCCATTGCTATTAGTCA-3' | 5'-CAAATGGTGTGGATGTGGAA-3' |
| | BMP8A | 5'-TCTTTGCTTTAGGCTGTTTGC-3' | 5'-GACCACATCAAGCACACTGAA-3' |
| | PPIE(2) | 5'-TCTGCTGATGTTCTCCAGTGA-3' | 5'-CCACTCATCATCTGACCACACT-3' |
| | NT5C1A | 5'-GGAAACTGACAGCAGGATACAG-3' | 5'-GACAGCTCCAGGTAATGGAAA-3' |
| | PABPC4 | 5'-GGTTCAGAATGCCTTCTGTG-3' | 5'-AGCCCAGAACCAGTTCAGAT-3' |
| | NC_HN | 5'-TCTTTCCACCGATGAAACTGT-3' | 5'-TAGGAGGTGATGCTGACTGAAG-3' |
| Chromosome 27 | DCAF7 | 5'-TGCTCATCAAATCACTGGTACA-3' | 5'-TCTCTTGGCAACTGCTGAAG-3' |
| | IGFBPF | 5'-AACTTGGAGAATGAGCCAGTTC-3' | 5'-AGAAAGCTGTGACAGTCACAATG-3' |

**S3.2 Table.** Gene located in HHR shared between Cemani, Silkie and L2 Taiwanese

| Chromosome | Position HHR | Gene | Gene location |
|---|---|---|---|
| Chr 1 | 37500001 | KCNC2 | (37554139..37651977) |
| | 38600001-39100001 | NAV3 | (38593741..38853234) |
| | | LOC101749434 | (38966662..38986324) |
| | | LOC101747387 | (39034265..39087406) |
| | | SYT1 | (39159693..39381655) |
| | 43400001 | no gene | - |
| | 45000001-45100001 | SNAPC3 | (45113080..45114953) |
| | | CCDC41 | (45004373..45029064) |
| | | TMCC3 | (45035452..45158371) |
| | | NDUFA12 | (45179842..451873200) |
| | 53000001 | SYN3 | (53005381..53192098) |
| | 60200001 | IQSEC3 | (60139819..60140562) |
| | | LOC427923 | (60251705..60272380) |
| | 66200001 | **SOX5** | (66068226..66320260) |
| | 97000001 | no gene | - |
| | 97200001-97400001 | LOC418465 | (97271946..97296021) |
| | 97300001 | LIPI | ((97343775..97364332) |
| | | RBM11); | (97364416..97372233) |
| | | ABCC13 | (97373098..97406938) |
| | 104300001 | LOC100857701 | (104327992..104328622) |
| | | URB1 | (104330026..104350249) |
| | | C21ORF63 | (104359342..104395685) |
| | 111800001-112400001 | LOC101751408 | (111808731..111811586) |
| | | MID1IP1 | (112078681..>112107703) |
| | | LOC101751659 | (112107719..112110917) |
| | | OTC | (112238931..112264837) |
| | | RPGR | (112266480..>112305049) |
| | | LOC101751733 | (112268857..112272035) |
| | | SRPX | (112335114..112376061) |
| | | SYTL5 | (112379829..112458193) |
| | | DYNLT3 | (112473711..112480593) |
| | 112700001 | C1HXORF59 | (112721634..112984857) |
| | 112900001-113000001 | LOC100857426 | (113130026..113692265) |
| | 113200001 | TMEM47 | (113256360..113282643) |
| | 113400001-113500001 | LOC100857461 | (113450753..11346665) |
| | 113700001-113900001, 114100001, 114300001, 114500001-114700001 | DMD | (113936660..114932774) |
| | 115900001 | IL1RAPL1 | (115540697..116242849) |

| | | | |
|---|---|---|---|
| | 190000001-190100001 | no gene | |
| Chr 2 | 1391000001 | LOC101748709 | (139073189..139108712) |
| Chr 3 | 27400001-27800001 | LRFN2 | (27539623..27733444) |
| | | MOCS1 | (28046781..28077569) |
| | 28000001-28300001 | **DAAM2** | (28087883..28296003) |
| | | KIF6 | (28295247..28437368) |
| | 29800001 | LOC101750239 | (29867778..29872533) |
| | 31200001-31700001 | FAM82A1 | (31217050..31259985) |
| | | CYP1B1 | (31277077..>31281223) |
| | | NANP | (31342516..31346273) |
| | | EIF2AK2 | (31347176..31365169) |
| | | CCDC75 | (31365762..3137214) |
| | | HEATR5B | (31372192..31425246) |
| | | STRN | (31428645..31493583) |
| | | CRIM1 | (31632337..>31793506) |
| | 37300001-37400001 | LYST | (37276366..37338630) |
| | | GNG4 | (37340874..3735442) |
| | | GGPS1 | (37413899..37414450) |
| | | **ARID4B** | (37437365..37523556) |
| | | LOC101747771 | (37462118..37465542) |
| Chr4 | 81700001-81800001 | HTT | (81695894..81773478) |
| | | GRK4 | (81780286..81810730) |
| | | NOP14 | (81810957..81832280) |
| | | MFSD10 | (81835253..81860398) |
| | | ADD1 | (81861183..81918087) |
| Chr 6 | 8300001 | NON-CODING | |
| Chr7 | 1100001 | CALCRL | (1099279..1125855) |
| | | LOC101751439 | (1128312..1244919) |
| | 18400001-18500001 | G6PC2 | (18415592..18426666) |
| | | SPC25 | (18426610..18430129) |
| | | CERS6 | (18441491..18552521) |
| | | STK39 | (18583965..18668725) |
| | | MIR1733 | (18452934..18453017 |
| | 20200001-20400001 | LOC101749581 | (20178615..20243928) |
| | | KCNH7 | (20331328..20536768) |
| Chr18 | 7500001 | CACNG1 | (7507047..7513432) |
| Chr 24 | 1700001-2100001 | SNX19 | (1748837..1792934) |
| | | NTM | (1860917..2105601) |
| | | MIR1601 | (2010029..2010105) |
| Chr 26 | 4800001 | LOC100858652 | (4781990..4822012) |

**S3.3 Table.** List of gene and gene function in consecutive HHR specific to Cemani chicken.
Chromosome 1

| Position | Gene name |
|---|---|
| 31900001 - 32000001 | LOC101749889 |
| | SLC16A7 (solute carrier family 16, member 7 (monocarboxylic acid transporter 2) |
| 44300001 - 44600001 | PLEKHG7 (pleckstrin homology domain containing, family G (with RhoGef domain) member 7) |
| | EEA1 (early endosome antigen 1, transcript variant X3) |
| | LOC101750717; LOC101750670, |
| | NUDT4 (nudix (nucleoside diphosphate linked moiety X)-type motif 4, transcript variant 2) |
| 55600001 - 55900001 | CCDC53 (coiled-coil domain containing 53, transcript variant X3) |
| | DRAM1 (DNA-damage regulated autophagy modulator 1) |
| | GNPTAB (N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits); |
| | SYCP3 (synaptonemal complex protein 3) |
| 59100001 - 59600001 | CHPT1 (choline phosphotransferase 1) |
| | MYBPC1 (myosin binding protein C, slow type) |
| | TRNAV-CAC (tRNA-Val)"transfer RNA valine (anticodon CAC) |
| | SPIC (Spi-C transcription factor (Spi-1/PU.1 related) |
| | PARP12 (poly (ADP-ribose) polymerase family) |
| | TBXAS1 (thromboxane A synthase 1 (platelet) |
| | HIPK2 (homeodomain interacting protein kinase 2) |
| | DNM1L (dynamin 1-like) |
| | BICD1 (bicaudal D homolog 1 (Drosophila) |
| | C1H12ORF35 (chromosome 1 open reading frame, human C12orf35) |
| | LOC101751557 |
| | AMN1 (antagonist of mitotic exit network 1 homolog (S. cerevisiae) |
| | METTL20 (methyltransferase like 20) |
| | DENND5B (DENN/MADD domain containing 5B) |
| | FAM60A (family with sequence similarity 60) |

Chromosome 2

| Position | Gene name |
|---|---|
| 44200001 - 45200001 | ULK4 (unc-51-like kinase 4 (C. elegans) |
| | TRAK1 (trafficking protein, kinesin binding 1) |
| | CCK (cholecystokinin) |
| | LOC101750378; LOC420716; LOC101750697; |
| | EIF1B (eukaryotic translation initiation factor 1B) |
| | MYRIP (myosin VIIA and Rab interacting protein) |
| | TMC2 (transmembrane channel-like 2) |
| | RPSA (ribosomal protein SA) |
| | SLC25A38 (solute carrier family 25, member 38) |
| | CX3CR1 (CX3C chemokine receptor 1) |
| | CCR4 (chemokine (C-C motif) receptor 4) |
| | GLB1 (galactosidase, beta 1) |
| | TMPPE (transmembrane protein with metallophosphoesterase domain) |
| | CRTAP (cartilage associated protein) |
| | LOC420721 (sushi domain-containing protein 5-like) |
| | FBXL2 (F-box and leucine-rich repeat protein 2) |
| | UBP1 (upstream binding protein 1 (LBP-1a); |
| | CLASP2 (cytoplasmic linker associated protein 2) |
| 51900001 - 52000001 | HECW1 (HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1) |
| | STK17A (serine/threonine kinase 17a) |
| 52400001 -52500001 | EGFR (epidermal growth factor receptor) |

Chromosome 3

| Position | Gene name |
|---|---|
| 62800001 - 62900001 | GOPC (golgi-associated PDZ and coiled-coil motif containing) |
| | DCBLD1 (discoidin, CUB and LCCL domain containing 1) |
| 63100001 - 63400001 | LOC101749082; LOC100858979 |
| | ROS1 (c-ros oncogene 1, receptor tyrosine kinase) |
| | VGLL2 (vestigial like 2 (Drosophila) |
| | RFX6 (regulatory factor X, 6) |
| | GPRC6A (G protein-coupled receptor, family C, group 6, member A) |
| | FAM162B (family with sequence similarity 162, member B) |
| | KPNA5 (karyopherin alpha 5 (importin alpha 6)) |
| | LOC421740 (sulfotransferase family 3A, member 1-like) |
| | RWDD1 (RWD domain containing 1) |
| | BET3L (BET3 like (S. cerevisiae)) |
| | FAM26D (family with sequence similarity 26, member D) |
| | BET3L (trafficking protein particle complex subunit 3-like protein) |
| | FAM26E (family with sequence similarity 26, member F) |
| | FAM26F (family with sequence similarity 26) |
| | DSE (dermatan sulfate epimerase) |
| | NT5DC1 (5'-nucleotidase domain-containing protein 1) |
| | FRK (fyn-related kinase) |
| 64100001 - 64200001 | HS3ST5 (heparan sulfate (glucosamine) 3-O-sulfotransferase 5); |
| | HDAC2 (histone deacetylase 2); |
| | LOC101750983; LOC101748060; LOC101749291; LOC101749451 |
| | C3H6ORF186 (chromosome 3 open reading frame, human C6orf186) |
| | CDC40 (cell division cycle 40 homolog (S. cerevisiae)) |
| | WASF1 (WAS protein family, member 1) |

| | |
|---|---|
| 66000001 - 67000001 | GPR6 (G protein-coupled receptor 6) |
| | FIG4(polyphosphoinositide phosphatase precursor) |
| | AKD1 (chromosome 6 open reading frame 199) |
| | CD164 (CD164 molecule, sialomucin) |
| | CEP57L1 (centrosomal protein 57kDa-like 1) |
| | SESN1 (sestrin 1) |
| | ARMC2 (armadillo repeat containing 2) |
| | ARMC2 (armadillo repeat containing 2) |
| | FOXO3 (forkhead box O3) |
| | LACE1 (lactation elevated 1) |
| | SNX3 (sorting nexin 3) |
| | NR2E1 (nuclear receptor subfamily 2 group E member 1) |
| | OSTM1 (osteopetrosis associated transmembrane protein 1) |
| | SEC63 (SEC63 homolog (S. cerevisiae)); |
| | SCML4 (sex comb on midleg-like 4 (Drosophila)) |
| 81100001 - 81200001 | KCNQ5 (potassium voltage-gated channel, KQT-like subfamily, member 5) |
| 108500001 - 108600001 | CDC5L (CDC5 cell division cycle 5-like (S. pombe)) |
| | SUPT3H (suppressor of Ty 3 homolog (S. cerevisiae)) |
| 108800001 - 109200001 | RUNX2 (runt-related transcription factor 2) |
| | CLIC5 (chloride intracellular channel 5) |

Chromosome 4

| Position | Gene name |
|---|---|
| 11400001 - 11500001 | KLF8 (Kruppel-like factor 8) |
| | LOC101749773; LOC101750275 |
| | RRAGB (Ras-related GTP binding B) |
| | MIR1790 (gga-mir-1790) |
| | MTMR8 (myotubularin related protein 8) |
| 40700001 - 40800001 | LOC100859119 (translation initiation factor IF-2-like) |
| | MAPK10 (mitogen-activated protein kinase 10) |
| 45400001 - 45800001 | ARHGAP24 (Rho GTPase activating protein 24) |
| | COPS4 (COP9 constitutive photomorphogenic homolog subunit 4 (arabidopsis) |
| | LIN54 (lin-54 homolog (C. elegans)) |
| | THAP9 (DNA transposase THAP9) |
| | SEC31A (SEC31 homolog A (S. cerevisiae)) |
| 46200001 - 46700001 | SCD5 (stearoyl-CoA desaturase 5) |
| | TMEM150C (transmembrane protein 150C) |
| | ENOPH1 (enolase-phosphatase 1) |
| | HNRPDL (heterogeneous nuclear ribonucleoprotein D-like) |
| | MIR1804 (microRNA mir-1804) |
| | PLAC8 (placenta-specific 8) |
| | COQ2 (coenzyme Q2 homolog, prenyltransferase (yeast)) |
| | HPSE (heparanase) |
| | HELQ (helicase, POLQ-like) |
| | MRPS18C (mitochondrial ribosomal protein S18C) |
| | FAM175A (family with sequence similarity 175, member A) |
| | LOC422609 (1-acylglycerol-3-phosphate O-acyltransferase 9-like) |
| | NKX6-1 (NK6 homeobox 1) |

| | |
|---|---|
| | CDS1 (CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 1) |
| | WDFY3 (WD repeat and FYVE domain containing 3) |
| | PPM1K (protein phosphatase, Mg2+/Mn2+ dependent, 1K) |
| | LOC101748637; LOC100859459; LOC101748838; LOC101748937 |
| 46900001 - 47000001 | Non coding region |
| 47200001 - 48700001 | LOC101749070; LOC101749033; LOC101748088 |
| | LPHN3 (latrophilin 3) |
| | MIR1730 (gga-mir-1730) |
| | TECRL (trans-2,3-enoyl-CoA reductase-like) |
| | IGFBP7 (insulin-like growth factor binding protein 7) |
| | POLR2B (polymerase (RNA) II (DNA directed) polypeptide B, 140kDa) |
| | NOA1 (nitric oxide associated 1) |
| | REST (RE1-silencing transcription factor) |
| | TMEM66 (transmembrane protein 66 precursor) |
| | LEPROTL1 (leptin receptor overlapping transcript-like 1) |
| | SRP72 (signal recognition particle 72kDa) |
| | ARL9 (ADP-ribosylation factor-like 9) |
| | HOPX (HOP homeobox); |
| | SPINK2 (serine peptidase inhibitor, Kazal type 2 (acrosin-trypsin inhibitor) |
| | DUSP4 (dual specificity phosphatase 4); |
| | TNKS (tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase) |
| 53000001 - 54600001 | FGF2 (fibroblast growth factor 2 (basic)) |
| | BBS12 (Bardet-Biedl syndrome 12) |
| | CETN1 (centrin, EF-hand protein, 1) |
| | IL21 (interleukin 21); IL2 (interleukin 2) |
| | ADAD1 (adenosine deaminase domain containing 1(testis-specific)) |
| | KIAA1109; LOC101751215; LOC101751990; LOC101752030; LOC101747787 |

| | |
|---|---|
| | TRPC3 (transient receptor potential cation channel,subfamily C, member 3) |
| | BBS7 (Bardet-Biedl syndrome 7) |
| | CCNA2 (cyclin A2) |
| | EXOSC9 (exosome component 9) |
| | ANXA5 (annexin A5) |
| | QRFPR (pyroglutamylated RFamide peptide receptor) |
| | TNIP3 (TNFAIP3 interacting protein 3) |
| | NDNF (neuron-derived neurotrophic factor) |
| | PRDM5 (PR domain containing 5) |
| | MAD2L1 (MAD2 mitotic arrest deficient-like 1 (yeast)) |
| | RPL7L1 (60S ribosomal protein L7-like 1) |
| | PDE5A (phosphodiesterase 5A, cGMP-specific) |
| | FABP2 (fatty acid binding protein 2, intestinal) |
| | USP53 (ubiquitin specific peptidase 53) |
| | SYNPO2 (synaptopodin 2) |
| | SEC24D (SEC24 family, member D (S. cerevisiae)) |
| | PRSS12 (protease, serine, 12 (neurotrypsin, motopsin)) |
| | NDST3 (N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 3) |
| 55200001<br>-<br>56000001 | LOC10175120; LOC101747253<br>NDST4 (N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 4) |
| | UGT8 (UDP glycosyltransferase 8) |
| | ARSJ (arylsulfatase family, member J) |
| 82000001<br>-<br>82200001 | TNIP2 (TNFAIP3 interacting protein 2) |
| | LOC101747497; LOC422887 (trichohyalin-like) |
| | FAM193A (family with sequence similarity 193, member A) |
| | RNF4 (ring finger protein 4) |
| | ZFYVE28 (zinc finger, FYVE domain containing 28) |

| | |
|---|---|
| 88700001<br>-<br>89100001 | MIR1654-1(microRNA mir-1654-1) |
| | ATRN (attractin) |
| | GFRA4 (GDNF family receptor alpha-4 precursor) |
| | LOC101750886 |
| | ADAM33 (ADAM metallopeptidase domain 33) |
| | ALMS1 (alstrom syndrome 1) |
| | EGR4 (early growth response protein 4) |
| | FBXO41 (F-box protein 41) |

Chromosome 7

| Position | Gene name |
|---|---|
| 22700001<br>-<br>23200001 | IGFBP5 (insulin-like growth factor binding protein 5) |
| | RPL37A (ribosomal protein L37a) |
| | SMARCAL1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a-like 1) |
| | MAR-04 (membrane-associated ring finger (C3HC4) 4) |
| | XRCC5 (X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining) |
| | TMEM169 (transmembrane protein 169) |
| | PECR (peroxisomal trans-2-enoyl-CoA reductase) |
| | MAP3K2 (mitogen-activated protein kinase kinase kinase 2) |
| | ERCC3 (excision repair cross-complementing rodent repair deficiency, complementation group 3) |
| | CYP27C1 (cytochrome P450, family 27, subfamily C, polypeptide 1) |
| | BIN1 (bridging integrator 1) |
| | MIR1582 (microRNA mir-1582) |
| | MIR1553 (gga-mir-1553) |

Chromosome 10

| Position | Gene name |
|---|---|
| 18300001 - 18400001 | SMAD3 (SMAD family member 3) |
| | AAGAB (alpha- and gamma-adaptin binding protein) |
| | LOC415554 (IQ motif containing H-like) |
| | C10H15orf61 (chromosome 10 open reading frame, human C15orf61) |
| | MAP2K5 (mitogen-activated protein kinase kinase 5) |
| | C10H15orf61 (uncharacterized protein C15orf61 homolog) |

Chromosome 12

| Position | Gene name |
|---|---|
| 6300001 - 6400001 | PHF2 (PHD finger protein 2) |
| | LOC101751708 |
| | FAM120A (family with sequence similarity 120A) |
| | WNK2 (WNK lysine deficient protein kinase 2) |

Chromosome 13

| Position | Gene name |
|---|---|
| 17200001 - 17300001 | GRXCR2 (glutaredoxin, cysteine rich 2) |
| | SH3RF2 (SH3 domain containing ring finger 2); LARS (leucyl-tRNA synthetase) |
| | RBM27 (RNA binding motif protein 27) |
| | TCERG1 (transcription elongation regulator 1) |
| | PPP2R2B (serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B beta) |
| | MIR1576 (gga-mir-1576) |

Chromosome 14

| Position | Gene name |
|---|---|
| 7800001 - 7900001 | LOC101747913 |
| | XYLT1 (xylosyltransferase I) |

Chromosome 19

| Position | Gene name |
|---|---|
| 3700001 - 4000001 | MYL10 (myosin, light chain 10, regulatory) |
| | LOC101749458; LOC417507; LOC101749861 (homeobox protein cut-like 1-like) |
| 6900001 - 7900001 | NXN (nucleoredoxin) |
| | RNMTL1 (RNA methyltransferase like 1) |
| | GEMIN4 (gem (nuclear organelle) associated protein 4) |
| | GLOD4 (glyoxalase domain containing 4) |
| | GEMIN4 (component of gems 4) |
| | FAM57A (family with sequence similarity 57, member A) |
| | VPS53 (vacuolar protein sorting 53 homolog (S. cerevisiae)) |
| | VPS53 (vacuolar protein sorting-associated protein 53 homolog) |
| | FAM101B (family with sequence similarity 101, member B) |
| | RPH3AL (rabphilin 3A-like (without C2 domains)) |
| | DOC2B (double C2-like domains, beta) |
| | OPNP (opsin 1 (cone pigments), long-wave-sensitive) |
| | TEX14 (testis expressed 14) |
| | RAD51C (RAD51 homolog C (S. cerevisiae)) |
| | LOC101750845 |
| | PPM1E (protein phosphatase, Mg2+/Mn2+ dependent, 1E) |
| | TRIM37 (tripartite motif containing 37) |
| | SKA2 (spindle and kinetochore associated complex subunit 2) |

| | |
|---|---|
| | MIR301B (microRNA mir-301b); MIR130C (gga-mir-130c) |
| | PRR11 (proline rich 11) |
| | SMG8 (protein SMG8) |
| | GDPD1 (glycerophosphodiester phosphodiesterase domain containing 1) |
| | YPEL2 (yippee-like 2 (Drosophila)) |
| | MIR1697 (microRNA mir-1697) |
| | DHX40 (DEAH (Asp-Glu-Ala-His) box polypeptide 40) |
| | CLTC (clathrin, heavy chain (Hc)) |
| | PTRH2 (peptidyl-tRNA hydrolase 2, mitochondrial) |
| | VMP1 (vacuole membrane protein 1 ); MIR21 (microRNA mir-21) |
| | TUBD1 (tubulin, delta 1); |
| | RPS6KB1 (ribosomal protein S6 kinase, 70kDa, polypeptide 1) |
| | RNFT1 (ring finger protein, transmembrane 1); |
| | MED13 (mediator complex subunit 13) |
| | LOC101747313; LOC772381; LOC101747835 (translation initiation factor IF-2-like); LOC101747794; |
| | TBX4 (T-box 4); BCAS3 (breast carcinoma-amplified sequence 3) |
| | PPM1D (protein phosphatase, Mg2+/Mn2+ dependent, 1D) |
| | APPBP2 (amyloid beta precursor protein (cytoplasmic tail) binding protein 2)) |

Chromosome 23

| Position | Gene name |
|---|---|
| 4400001 - 4800001 | GJA4 (gap junction protein, alpha 4, 37kDa) |
| | GJB3 (gap junction protein, beta 3, 31kDa) |
| | GJB5 (gap junction protein, beta 5, 31.1kDa) |
| | C1ORF94 (chromosome 23 open reading frame, human C1orf94) |
| | CSMD2 (CUB and Sushi multiple domains 2) |
| | COL9A2 (collagen, type IX, alpha 2) |
| | SMAP2 (small ArfGAP2) |
| | SMAP2 (stromal membrane-associated protein 2) |
| | RIMS3 (regulating synaptic membrane exocytosis 3) |
| | NFYC (nuclear transcription factor Y, gamma) |
| | MIR30E (microRNA mir-30e) |
| | MIR30C-1 (microRNA mir-30c-1) |
| 5000001 - 5600001 | PTP4A2 (protein tyrosine phosphatase type IVA, member 2) |
| | MIR1780 (microRNA mir-1780) |
| | KHDRBS1 (KH domain containing, RNA binding, signal transduction associated 1) |
| | TMEM39B (transmembrane protein 39B) |
| | MARCKSL1 (MARCKS-related protein) |
| | HDAC1 (histone deacetylase 1) |
| | LCK (lymphocyte-specific protein tyrosine kinase) |
| | FAM167B (family with sequence similarity 167, member B) |
| | MTMR9LP (myotubularin related protein 9-like, pseudogene) |
| | EIF3I (eukaryotic translation initiation factor 3 subunit I) |
| | TMEM234 (transmembrane protein 234) |
| | DCDC2B (doublecortin domain containing 2B) |
| | CCDC28B (coiled-coil domain containing 28B) |
| | TXLNA (taxilin alpha) |

| |
|---|
| KPNA6 (karyopherin alpha 6 (importin alpha 7)) |
| FAM229A (family with sequence similarity 229, member A) |
| TSSK3 (testis-specific serine kinase 3) |
| BSDC1 (BSD domain containing 1) |
| ZBTB8B (zinc finger and BTB domain containing 8B) |
| LOC101747327; LOC419662 (zinc finger and BTB domain containing 8A-like) |
| ZBTB8A (zinc finger and BTB domain-containing protein 8A) |
| ZBTB8OS (zinc finger and BTB domain containing 8 opposite strand) |
| RBBP4 (retinoblastoma binding protein 4) |
| SYNC (syncoilin, intermediate filament protein) |
| KIAA1522 |
| YARS (tyrosyl-tRNA synthetase) |
| HPCA (hippocalcin) |
| RNF19B (ring finger protein 19B) |
| AK2 (adenylate kinase 2) |
| ADC (arginine decarboxylase) |
| NDUFS5 (NADH dehydrogenase (ubiquinone) Fe-S protein 5,15kDa (NADH-coenzyme Q reductase)) |
| LOC101748731 (microtubule-actin cross-linking factor 1, isoforms 1/2/3/5-like) |
| LOC101750034 (microtubule-actin cross-linking factor 1-like); LOC101749968 |
| BMP8B (bone morphogenetic protein 8b) |
| PPIE (peptidylprolyl isomerase E (cyclophilin E)) |
| PABPC4 (poly(A) binding protein, cytoplasmic 4 (inducible form)) |
| HEYL (hairy/enhancer-of-split related with YRPWmotif-like) |
| NT5C1A (5'-nucleotidase, cytosolic IA) |
| HPCAL4 (hippocalcin like 4) |
| LOC419677 (protein FAM49A-like) |
| TRIT1 (tRNA isopentenyltransferase 1) |

| | |
|---|---|
| | MYCL1 (v-myc myelocytomatosis viral oncogene homolog 1,lung carcinoma derived) |
| | MFSD2A (major facilitator superfamily domain-containing protein 2A) |
| | CAP1 (CAP, adenylate cyclase-associated protein 1 (yeast)) |
| | PPT1 (palmitoyl-protein thioesterase 1) |
| | RPL11 (60S ribosomal protein L11) |
| | TCEB3 (transcription elongation factor B polypeptide 3) |
| | PITHD1 (PITH (C-terminal proteasome-interacting domain of thioredoxin-like) domain containing 1) |
| | LYPLA2 (lysophospholipase II); GALE (UDP-galactose-4-epimerase) |
| | HMGCL (hydroxymethylglutaryl-CoA lyase, mitochondrial isoform 2) |
| | HMGCL (3-hydroxymethyl-3-methylglutaryl-CoA lyase) |
| | FUCA1 (fucosidase, alpha-L- 1, tissue) |
| | CNR2 (cannabinoid receptor 2 (macrophage)) |
| | PNRC2 (proline-rich nuclear receptor coactivator 2) |
| | SRSF10 (serine/arginine-rich splicing factor 10) |
| | GRHL3 (grainyhead-like 3 (Drosophila)) |
| | LBFABP (liver basic fatty acid binding protein) |
| | MYOM3 (myomesin family, member 3) |
| | IL22RA1 (interleukin 22 receptor, alpha 1) |
| | IL28RA (interleukin 28 receptor, alpha (interferon, lambda receptor)) |
| | GRHL3 (grainyhead-like 3 (Drosophila)) |
| | NIPAL3 (NIPA-like domain containing 3) |

Chromosome 27

| Position | Gene name |
|---|---|
| 2400001 - 2600001 | MRC2 (mannose receptor, C type 2) |
| | LOC101750687 |
| | TRNAS-GCU |
| | TANC2 (tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2) |
| | CYB561 (cytochrome b-561) |
| | ACE (angiotensin I converting enzyme (peptidyl-dipeptidase A) 1) |
| | DCAF7 (DDB1 and CUL4 associated factor 7) |
| | KCNH6 (potassium voltage-gated channel, subfamily H (eag-related)) |

# Chapter 4

## The origin of *Ayam* Cemani and its relationship with other black plumage chicken breeds from Indonesia, America and China.

### 4.1 Introduction

Humans have manipulated the genotypes of chicken to produce new breeds for their benefit for a long time, resulting on the phenotypic variation of chickens. Nowadays, most of domestic chickens exist around the world as a result of artificial selection by cross breeding. Cross breeding between the chickens has produced a hundred of chicken breeds and hybrids of that have different phenotypes compared to its ancestor. For example, cross breeding between green jungle fowl (*Gallus varius*) with Indonesian local chicken (especially *Kampung* chicken) produced *Ayam* Bekisar [1,2]. Even though very limited, such cross breeding among chicken breeds has not only taken place within local geographic regions but also between continents through migration of humans. Because of the recurrence of cross breeding, it is difficult to elucidate the track record of the origin of a breed unless such breed has a good historical record.

Since the invention of DNA technology, researchers had been attempting to identify the starting point of the domestication process and the history of chicken distribution combined with the historical record of chicken breeds. Studies using mitochondrial DNA revealed that Southeast Asia is considered as the origin of the domestication process of chicken [3,4] and Indonesia is suggested as one of the locations where domestication events took place due to a special phylogenetic clade forming for Indonesian chicken [5], along with the existence of red jungle and green jungle fowl in Indonesia.

From all of the local chicken in Indonesia, Kedu *Hitam* [Black Kedu (BK)] chicken has the longest breeding history. Chicken breeds such as Black Java (BJ) in

America and Cemani in Indonesia are considered as descendants of BK chicken. BJ was developed by crossbreeding between BK chicken and unknown chicken breeds [6], and were brought to the United States and Europe around the 1800's from China, India and other Asian countries; BJ chicken was admitted to the American Standard in 1910 [7,8] Meanwhile, Cemani chicken was traditionally developed by selection and crossbreeding of BK chicken by local breeders, based on a well-known story [9]. This intentional selection resulted in a specific characteristic of Cemani chicken, the Fibromelanosis (Fm) phenotype; due to this Fm phenotype, Cemani has become a popular breed with high economic value. In addition, there is another story mentioning that Cemani is an independent breed brought to Kedu village from Kalikuto, Magelang, Central Java in 1920 [10].

Morphological characteristics of Indonesian Cemani, BK and American BJ chicken are almost similar, denoted by black plumage in their body. Another black plumage chicken from America with similar morphology with Cemani, BK and American BJ is the Black Sumatra (BS) chicken. American BS chicken is derived from Sumatran chicken which originated from Sumatra, Indonesia [6]. In addition, there are also several chickens in China which share a similar black plumage phenotype with Indonesian and American black plumage chicken: Muchuan, Tianfu, Jiuyuan, and Emei black plumage chicken [11]. Interestingly, black plumage chicken breeds have varying comb colors. Cemani, Indonesian Sumatra and Chinese chicken like Muchuan and Tianfu share similar black color of comb as well as the Fibromelanosis (Fm) phenotype (Table 4.1). However, Indonesian BK and American black plumage chicken (BJ and BS) display red comb color. Therefore, to study the relationship between black plumage chickens across the world is intriguing to elucidate the evolutionary history of black plumage chicken, especially their relationship with Cemani chicken in Indonesia.

In Chapter 2, I revealed that Cemani and Silkie are closely related based on phylogenetic tree analysis of the *Fm* region. Based on this, I presumed that since the fibromelanosis region involving *EDN3* under selective sweeps is located on chromosome 20, then analyzing genomic patterns of black plumage chicken population structure on chromosome 20 could clearly resolve the relationship between black plumage chicken breeds. Therefore, in this study, I examined the variability in the duplication boundary in the *Fm* region and analyzed genetic relationships between Cemani and other black plumage chicken on chromosome 20. I revealed the genetic differentiation among black plumage chicken in Indonesia, America and China.

## 4.2 Materials and Methods

***Materials***

### *4.2.1 Genomic DNA samples*

Genomic DNA samples used in Chapter 4 for detecting variation in the duplication boundaries are Fm type chicken: Cemani (Cemani 40), Black Kedu (KD 2, 3, 5, and 16) and non-Fm type chicken: White Kedu (KDP5, 7 and 13).

### *4.2.2 Whole Genome Sequence of chicken breeds*

WGS of chicken breeds were retrieved from Genbank database with accession number of project SRP067615 (http://www.ncbi.nlm.nih.gov/) for Chinese black plumage chicken and one Chinese local chicken non Fm-phenotype [12]; DRA003951 (http://www.ddbj.nig.ac.jp) for Indonesian black plumage chickens, American black plumage chickens, White Leghorn, Red and Green Jungle Fowl; SRS420686 (http://www.ncbi.nlm.nih.gov/) for Chinese Silkie chicken [13]; and SRS426963 (http://www.ncbi.nlm.nih.gov/) for Taiwanese L2 [13]. WGS data of Cemani chicken was obtained from the previous study in Chapter 2. This study only used some WGS data of

Chinese breeds provided in the databases as well as Indonesian black plumage chickens. Data sample of WGS used in this study are listed on Table 4.1.

***Methods***

### *4.2.3 Analysis of variation in duplication boundaries*

The multiplex-PCR primers for identifying the variation in duplication boundaries of Cemani, Black Kedu and White Kedu chicken are listed in Table 4.2 (Kinoshita's method). DNA amplification of each individual bird was performed according to the following conditions: the multiplex-PCR was performed in a total volume of 25 µL mix solution, containing 3µL buffer 10X (Takara), 3µL 25mM MgCl2, 2µL of deoxynucleotide triphosphate (dNTP) mixture, 40ng of genomic DNA, 10 pmol of each oligonucleotide primer, and 1U of Taq DNA polymerase; PCR reaction cycle parameters were denaturation for 4 min at 95°C then 35 cycles of 94°C for 30 sec, 60°C annealing for 30 sec, and 72°C for 1 min, with a final extension step for 5 min at 72°C. The multiplex-PCR products with length 558 and/or 664 bp (see Fig 4.1) were digested at 37°C overnight with 10 U of *MIuI* restriction enzyme (Kinoshita's method). Restriction digests were electrophoresed for 50 minute at 50 V on a 2% agarose gel.

### *4.2.4 Mapping and SNP calling*

Before mapping sequence data to the reference, sequences of each bird were trimmed using trim function in CLC genomic workbench 9.0 (https://www.qiagenbioinformatics.com/). Trimmed sequences were then mapped separately to the *Gallus gallus* reference sequence (*Gallus gallus* 4.0) with the following parameter: length and similarity length = 0.9. After that, the coverage of mapped sequences was calculated using Qualimap software [16]. After mapping, I used SNVs detection function in CLC Genomic workbench for calling SNVs of each individual. I

applied different parameters for minimum coverage and minimum read calling based on the coverage of each individual mapped reads (Table 4.3). Furthermore, I applied similar parameter to each individual mapped reads for minimum central base=20, average base quality=15 and minimum allele frequency=35%. To extract SNPs from each breed, I excluded insertion and deletions (INDELs) from the SNVs data that I had obtained using vcftools. The total SNVs and SNPs were also calculated using vcftools [17].

### *4.2.5 Population structure and evolutionary history analysis*

The relationship between Cemani and other black plumage chickens was examined using principle component analysis. Only SNPs on chromosome 20 were used for PCA analysis. Using vcftools software [17], I extracted SNPs only in chromosome 20 of each individual and then merged the SNPs data from all individuals. Principal components of the variance-standardized relationship matrix of merged data were extracted using PLINK software [18] with default parameters. Then, the diagram plot was constructed using Genesis software [19].

## 4.3 Results

### *4.3.1 Variation in duplication boundary*

Restriction fragment length polymorphism, or RFLP is a technique for identifying variation in DNA sequences involves fragmenting a sample of DNA by a restriction enzyme. For detecting the variation in duplication boundary of Black Kedu chicken, I used Cemani with Fm phenotype as a positive control (Cem40) for Fm type chicken and Kedu *Putih* [White Kedu (KDP)] with non-Fm phenotype as a positive control for non-Fm type. Primers set for multiplex-PCR were used to amplify the boundary region. Two bands with 664 and 588 bp length would be amplified in positive control (i.e. chicken with Fm/Fm genotype). Fragment with 588 bp length is specific for Fm type chickens.

Therefore, the Fm specific boundary would not be detected in positive control for non-Fm type (i.e. wild type chicken), resulting only in amplification of a single 664 bp band (Fig 4.1). PCR result showed that two bands (588 and 664 bp) were amplified in Cemani (Cem40) as Fm type and BK (KD16) sample [Fig 4.2(a)]. However, for other chickens BK (KD2, KD3, KD5) and White Kedu (KDP 7 and KDP 13), only one band (664 bp) was amplified [Fig 4.2(a)]. Restriction enzyme *MluI* was used to identify the wild type allele- two bands of 400 and 264 bp lengths would be produced after digestion of the plasmid [Fig 4.2 (b)].

From restriction enzyme analysis in the Cemani sample (Cem 40), as positive control for Fm type, two bands (588 and 664 bp) were detected, indicating that Cemani possess heterozygous Fm allele. In addition, the cleavage site located within 664 bp fragment involve duplication boundary. Restriction enzyme *MluI* will digest and produce two bands (440 bp and 224 bp) in White Kedu sample, positive control for non-Fm type. However, in White Kedu (KDP) sample (KDP5, 7 and 13), restriction enzyme analysis showed that there are two different restriction patterns, either samples with two bands (440 bp and 224 bp) or samples with three bands (664 bp, 440 bp and 224 bp); this indicates that there are two categories of wild type, heterozygous wild type and homozygous wild type. For homozygous wild type, KDP5, there are two bands formed (440 bp and 224 bp), indicating that both alleles have cut position [fm(+)/fm(+)]. However, for heterozygous wild type seen in KDP7 and KDP13, there are three bands formed, which means that one of the alleles was not digested by the restriction enzyme *MluI* thus producing a 664 bp band, and other allele was digested to produce 440 bp and 224 bp bands [fm(-)/fm(+)]. Interestingly, for BK chicken, there are individuals that are heterozygous Fm-type and heterozygous wild type. KD16 with blackish comb showed that one allele possesses the duplication boundary and other alleles is wild type allele

[Fm/fm(+)]. Meanwhile, other BK chicken (KD2, KD3, KD5) produced three bands after restriction digestion, indicating they are all heterozygous wild type [fm(-)/fm(+)].

### *4.3.2 Mapped reads, SNP calling and SNP identification*

The reference genome sequence was covered with differences in average depth in each individual, ranging from the lowest 3.1-fold (BS85) to the highest 57.8-fold (Tianfu25) (Table 4.1). The alignments between the uniquely mapped reads and the reference genome were used to identify SNVs. Mapping of the sequence reads of each individual based on its coverage identified a range of 1,282,018 (in Black Java) to 7,501,295 (in GJF) SNVs including Insertions/deletions (INDELs) and MNVs (Table 4.3). After removing INDELs and MNVs, remaining average SNPs identified a range of 873,425 (Black Kedu) to 6,708,442 (in GJF) (Table 4.3).

### *4.3.3 PCA analysis of chromosome 20*

PCA analysis was used to detect population stratification between black plumage chicken and other local chicken. Only SNPs in chromosome 20 were used for PCA analysis. SNPs on chromosome 20 from each individual breed are listed in Table 4.3.

PCA analysis revealed that there are five clusters formed: Green Jungle Fowl, Red Jungle Fowl, American, Chinese and Indonesian clusters. Cemani chicken clustered together with Indonesian black plumage chickens including BK, while Silkie clustered together with Chinese local chicken. American black plumage chickens were clustered together with White Leghorn and L2 Taiwanese. RJF were in two separate clusters: two samples were in the Indonesian cluster and three samples in the RJF cluster (Fig 4.3).

## 4.4 Discussion

In Chapter 2 (Table 2.2), haplotype analysis of *EDN3* revealed that four BK individuals (KD3, KD16, KDH3, and KDH8) were heterozygous with hap2/hap4, which

is similar to Cemani and Silkie chicken with Fm phenotype. Furthermore, detection of duplication boundaries by PCR showed that three BK chickens (KD16, KDH3, and KDH8) had duplication boundaries which could be amplified (Table 2.3). These results indicate heterogeneity in the *Fm* region within BK chicken. However, haplotype analysis and duplication boundaries detection could not determine the genotypic allele of BK chicken, that is, whether BK chicken has two *Fm* chromosomes like Cemani or not. Therefore, in this study, for detecting the haplotype allele in BK samples, restriction enzyme analysis was used. Restriction enzyme analysis revealed that Cemani and BK chicken (KD16) showed different restriction digest patterns. Cemani as a positive control for Fm/Fm has two bands (588 and 664 bp). However, apart from the two bands representing the *Fm* allele (588 and 664 bp), KD16 also has another two bands (400 and 264bp) representing the wild type allele. In addition, KD2 and KD5 individuals have the wild type allele. This confirms the previous results from Chapter 2 about the variation of *Fm* allele in the BK chicken in which some BK individuals (KD2, 3 and 5) have two wild allele phenotypes and others (KD16) have one *Fm* locus and one wild type locus. Even though restriction enzyme analysis proved that BK chicken (KD16) also possesses a Fm chromosome and high diversity in the *Fm* region, it still could not reveal whether Cemani originated from BK chicken or not. This is because the variation in duplication boundary of BK chicken could be due to recent interbreeding between Cemani and BK chicken.

I further examined genetic relationships using whole genome sequence of three lines of Indonesian black plumage chicken (Cemani, Sumatra and BK), six lines of Chinese local chicken (Emei, Jiuyuan, Muchuan, Tianfu, Silkie and Pengxian), three lines of American local chicken [white leghorn, Black Java (BJ), Black Sumatra (BS)], one layer line (L2 Taiwanese), and Green and Red Jungle fowl as ancestor. Since the resources of WGS provided by the Genbank database was generated from different

studies [6,11], the quality of WGS is variable. Therefore, I processed the analysis of sequence reads by using different parameters of SNP calling based on its quality in order to obtain reliable SNP data (Table 4.3). Then, I used SNPs data for PCA analysis. PCA analysis is a common method used for identifying ancestry differences among sampled individuals, inferring population structure and investigating demographic history [20,21,22]. A previous study in chickens used PCA analysis revealed the presence of at least three distinct clusters among the six geographically representative populations of Tibetan fowls [11].

In Chapter 2, based on phylogenetic tree analysis of the *Fm* region on chromosome 20, Cemani and Silkie are closely related. Therefore, I expected that Black plumage chicken might cluster together due to selection acting on the *Fm* region of chromosome 20. However, in contrast, my study showed a distinct cluster and distribution pattern of Indonesian, Chinese, and American black plumage chickens based on principle component analysis of SNPs from chromosome 20. I identified five clusters among the three geographically representative populations: (i) Green Jungle fowl cluster, (ii) Red Jungle fowl cluster, (iii) the chickens inhabiting Indonesia were genetically closer to some RJF individuals (Indonesia cluster), (iv) the chickens inhabiting America (America cluster), and (v) chickens inhabiting China (China cluster).

American BJ and BS chicken, which have a recorded history with Indonesian BK and Sumatra chicken, respectively, were in a distinct cluster with Indonesian chickens. In addition, Chinese black plumage chickens that possess similar morphology with Cemani chicken showed different population structure with Cemani chicken. In contrast with my previous study in Chapter 2 which showed close relatedness between Cemani and Silkie, in this Chapter, Silkie and Cemani were found in distinct clusters. Moreover, BJ which is known to be originated from crossing over between BK and unknown local Indonesian

chicken was also in a distinct cluster from the BK chicken. The distinct clustering between Indonesia, America and Chinese chickens may be due to the background of breeding and distribution of the chicken breeds. This distinct distribution pattern between black plumage chicken in America, Indonesia and China suggests that the divergence of black plumage chicken might have originated independently across different geographical regions which influenced the genetic differentiation of chicken breeds. The results from this Chapter supports the previous study which revealed that American (BJ and BS) and Indonesia black plumage chickens (Sumatra and BK) were in a different clade based on phylogenetic tree constructed from SNPs of chicken whole genome sequences [6].

Red Jungle Fowl which was sampled in Sumatra (Indonesia) is in a cluster together with Indonesian local chicken, indicating that introgression has occurred between red jungle fowl and local chicken in Indonesia. This result corroborates with previous studies that revealed introgression between red jungle fowl and local chicken based on nuclear DNA markers [23] and mitochondrial DNA [24].

*Cemani as an independent breed*

In Chapter 2, Cemani and BK (KD16) was located in the same clade based on phylogenetic tree analysis in and around *EDN3*. In this Chapter, PCA analysis of chromosome 20 indicate that they are closely related to each other by sharing similar ancestral polymorphisms. The possibility that Cemani and BK clustered together in the *Fm* gene region is because of recent crossbreeding between Cemani and BK. The Fm phenotype in Cemani is a homozygous dominant trait, such that cross breeding with BK that have homozygous recessive traits will produce heterozygous traits in BK chicken which is confirmed by the presence of a heterozygous allele in individual KD16. The genetic alteration may have partly occurred in *Fm* region including *EDN3* on chromosome 20.

One possible reason that Cemani and BK were in a clade together with other Indonesia chicken on chromosome 20 is because Indonesian chickens including BK and Cemani have a long breeding history. Similarity in mating systems and environment as well as the occasional transportation of chickens between villages [25] may affect the genetic diversity of Indonesian chicken. Therefore, it is likely that Cemani, BK and other Indonesian chicken reciprocally transferred genetic variation due to interbreeding between these breeds.

However, in contrast with the close relationship of Cemani and BK chickens based on the *Fm* region and chromosome 20, Ismoyowati (2012) calculated the genetic distance of BK chicken (Red, White, Black Kedu and Cemani) using microsatellite DNA and revealed that BK and Cemani had the farthest genetic distance (0.236) compared with genetic distance between Black and Red Kedu (0.126) or Black and White Kedu (0.072). In that study, Black Kedu was characterized with blackish comb and white skin color. Such characterization of comb of BK chicken is similar with the KD16 individual in my study. In addition, the phylogenetic tree analysis of 14 breeds of Indonesian local chicken, including Cemani, Black and White Kedu chicken, based on 20 microsatellite markers revealed that Cemani and BK chicken were in separate clades [27]. Ashari [27] included the same BK samples (KD16, KDH3 and KDH8) that were used in Chapter 2. Moreover, Sulandari (2008) calculated genetic distance of Indonesian chickens including Kedu and Cemani based on D-loop sequence. The BK population from the study by Sulandari (2008) also include samples KD16, KDH3 and KDH8 used in my study and BK with red comb and black plumage (wild type phenotype) and revealed genetic distance between Cemani and BK was 0.043, the value is higher compared with Black and White Kedu was 0.003.

Combining the two results above, I conclude that Cemani might be an independent breed that was brought to Kedu village and experienced interbreeding and selection with BK chickens resulting in genetic introgression in the *Fm* region between two breeds. The separate clustering of Indonesian chicken (which include Cemani and BK), American chicken and Chinese chicken can be explained by "isolation effect". Different geographical distribution of Indonesian, American and Chinese chicken causes limited contact or crossbreeding between the breeds, influencing genetic variation by limiting gene flow between the chickens. The distant relationship between Cemani and BK based on microsatellite and mitochondrial DNA indicates that Cemani and BK chicken possess genetic differentiation between populations (Fig 4.4).

Nevertheless, this study still could not reveal the relationships between BK and Cemani chicken in other chromosomes aside from chromosome 20 within Indonesian local chicken and between other chickens across the world. In addition, this study also only provided information of genetic structure of black plumage chicken on chromosome 20. Therefore, more analysis in other chromosomes is needed to clarify the history of domestication process in Indonesia, especially in Cemani chicken and its relationship with other Black plumage chickens.

## 4.5 Conclusion

In conclusion, I assessed the variation within Black Kedu chicken in duplication boundaries. This variation in Kedu chicken may be due to recent interbreeding with Cemani chicken. In addition, I acquired five clusters (Green and Red Jungle Fowl, America, Indonesia and China) based on principle component analysis on chromosome 20, indicating that chicken breeds in America, Indonesia and China have different ancestral polymorphisms in chromosome 20 due to geographical isolation. In addition, Cemani and Silkie shared similar genetic information in the region that under selective

sweeps associated with the Fm phenotype on chromosome 20, yet were in separate clusters based on PCA analysis of chromosome 20, supporting that the region selected for the Fm phenotype has occurred recently.

## 4.6 Reference


1. Sartika T, Iskandar S. Mengenal Plasma Nutfah Ayam Indonesia dan Pemanfaatannya. Bogor (ID): Balai Penelitian Ternak. Pusat Penelitian dan Pengembangan Peternakan. Badan Penelitian dan Pengembangan Pertanian; 2007.
2. Nataamijaya AG. The native chickens of Indonesia. Buletin Plasma Nutfah. 2000; 6(1): 1-6.
3. Fumihito A, Miyake T, Sumi S, Takada M, Ohno S, Kondo N. One subspecies of the red junglefowl (Gallus gallus gallus) suffices as the matriarchic ancestor of all domestic breeds. Proc Natl Acad Sci U S A. 1994; 91(26): 12505-12509.
4. Fumihito A, Miyake T, Takada M, et al. Monophyletic origin and unique dispersal patterns of domestic fowls. Proc Natl Acad Sci U S A. 1996; 93(13): 6792-6795.
5. Sulandari S, Zein, MSA, Sartika T. Molecular Characterization of Indonesian Indigenous Chickens based on Mitochondrial DNA Displacement (D)-loop Sequences. HAYATI Journal of Biosciences. 2008; 15(4): 145-154.
6. Ulfah M, Kawahara-Miki R, Farajalllah A, Muladno M, Dorshorst B, Martin A, et al. Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam. BMC Genomics. 2016; 17: 320.
7. American Poultry Association. The American standard of perfection, illustrated. A complete description of all recognized varieties of fowls. Boston, Mass: American Poultry Association; 1910.
8. American Poultry Association. The American standard of perfection, illustrated. A complete description of all recognized varieties of fowls. Guthrie, Okla: American Poultry Association; 1956.
9. Muryanto. Hasil-hasil penelitian dan sumbangan pemikiran pengembangan Ayam Kedu. Prosiding Lokakarya Nasional Inovasi Teknologi Pengembangan Ayam Lokal. 2005; 114-118.
10. Sulandari, S., M.S.A. Zein, S. Paryanti, T. Sartika, M. Astuti, T. Widjastuti, E. Sujana, S. Darana, I. Setiawan, and D. Garnida. Sumberdaya genetik ayam lokal Indonesia. In: Keanekaragaman Sumber Daya Hayati Ayam Lokal Indonesia: Manfaat dan Potensi. (Diwyanto, K and S. N. Prijono, Eds). Jakarta: 1st Ed. LIPI Press; 2007; 49-50.
11. Li D, Che T, Chen B, Tian S, Zhou X, Zhang G, et al. Genomic data for 78 chickens from 14 populations. GigaScience. 2017; 6(6): 1-5.
12. Li D, Che T, Chen B et al. Supporting data for "Genomic data for 78 chickens from 14 populations." Giga- Science Database. 2017
13. Fan W-L, Ng CS, Chen C-F, Lu, M-Y J, Chen Y-H, Liu C-J, et al. Genome-Wide Patterns of Genetic Variation in Two Domestic Chickens. Genome Biol Evol. 2013; 5(7): 1376-1392.
14. Dharmayanthi AB, Terai Y, Sulandari S, Zein MSA, Akiyama T, Satta Y. The Origin and Evolution of Fibromelanosis in domesticated chicken : Genomic Comparison of Indonesian Cemani chicken and Chinese Silkie Breeds. Plos ONE. 2017; 12(4): e0173147.

15. Pham M-H, Chang W-H, Berthouly-Salazar C, Lin DY, Yungrahang S, Wang C-C, et al. Genetic Characterization of Taiwan Commercial Native Chickens Ascertained by Microsatellite Markers. Breeding and Genetics. 2013; 50(4): 290-299.
16. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics. 2012; 28(20): 2678-2679.
17. Petr D, Adam A, Goncalo A, Cornelis A A, Eric B, Mark A D, et al. 1000 Genomes Project Analysis Group; The variant call format and VCFtools. Bioinformatics. 2011; 27(15): 2156–2158
18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3): 559–575.
19. Buchmann R and Hazelhurst S. Genesis Manual. Johannesburg: University of the Witwatersrand; 2014. http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf
20. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. Allison DB, ed. PLoS Genetics. 2006; 2(12): e190.
21. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genet. 2006; 38(8): 904-909.
22. Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen et al. Quantifying Population Genetic Differentiation from Next-Generation Sequencing Data. Genetics. 2013; 195(3): 979-992.
23. Sawai H, Kim HL, Kuno K, Suzuki S, Gotoh H, Takada M, et al. The Origin and Genetic Variation of Domestic Chickens with Special Reference to Junglefowls Gallus g. gallus and G. varius. Orban L, ed. PLoS ONE. 2010; 5(5): e10639.
24. Miao Y-W, Peng M-S, Wu G-S, Ouyang Y-N, Yang Z-Y, Yu N, et al. Chicken domestication: an updated perspective based on mitochondrial genomes. Heredity. 2013; 110(3): 277-282.
25. Kinoshita K, Okamoto S, Shimogiri T, Kawabe K, Nishida T, Kakizawa R, et al. Gene constitution of egg white proteins of native chicken in Asian countries. Asian-Australas J Anim Sci. 2002; 15: 157-165.
26. Ismoyowati, Sukardi and Agus Susanto. Genetic Diversity of Kedu Chicken Based on Phenotypic Characteristics and Microsatellite Loci. International Journal of Poultry Science. 2012; 11(9): 605-610.
27. Ashari H, Sulandari S, Zein MSA, Han JL. Genetic Diversity and Relationship of Indonesian Indigenous Chickens Inferred from Microsatellite DNA Markers. Tropical Agricultural Research. 2015; 26(4): 642–654.

## 4.7 Figure and Table

**Fig 4.1. Location of multiplex-PCR primers.**

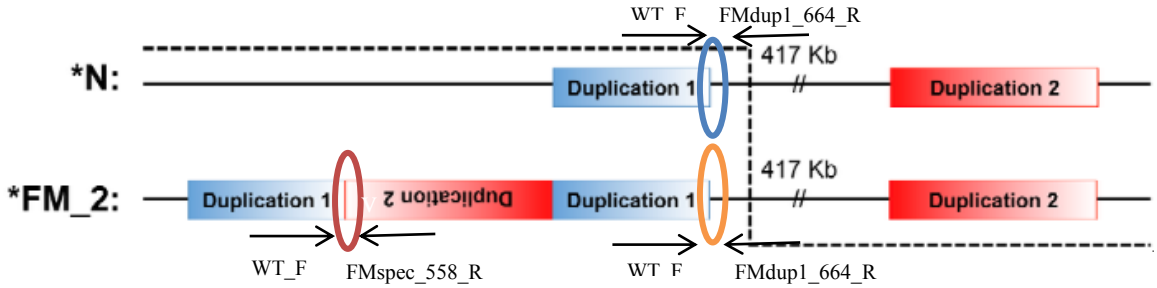



**Fig 4.2. The result of multiplex-PCR and RFLP** (a) PCR amplified by multiplex-PCR primers. (b) multiplex-PCR primer-amplified products digested with *MluI*.

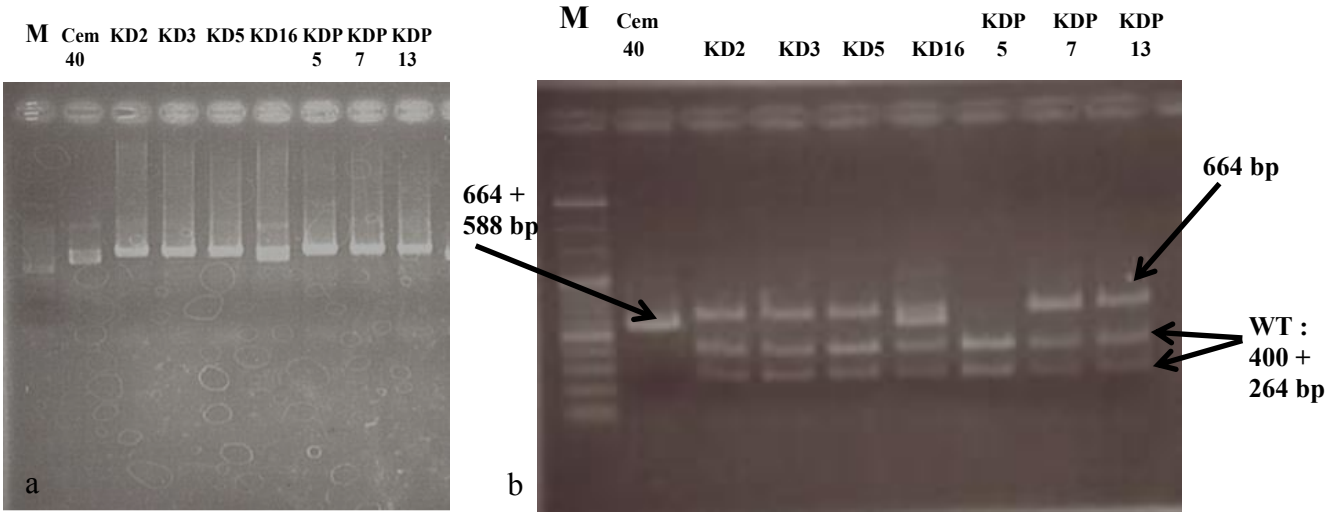


**Fig 4.3**. **Principle component analysis of chromosome 20 of domestic chicken breeds.** Each circle represents one of five clades: GJF (red), RJF (blue), America (orange), China (black), and Indonesia (green).

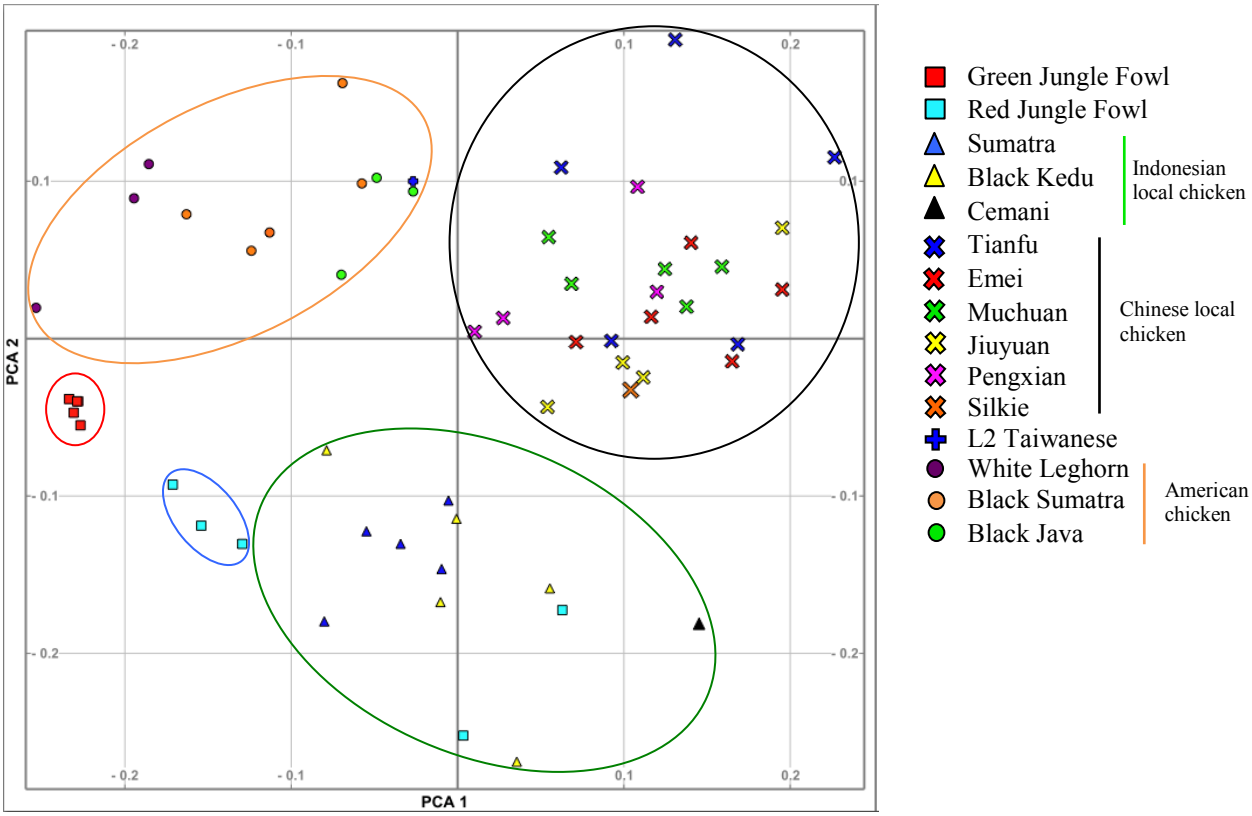


**Fig 4.4. Interpretation of relationship between Cemani and Kedu *Hitam*** based on (a) PCA analysis on chromosome 20, (b) phylogenetic tree of Fibromelanosis region, and (c) microsatellite and mitochondrial DNA analysis.

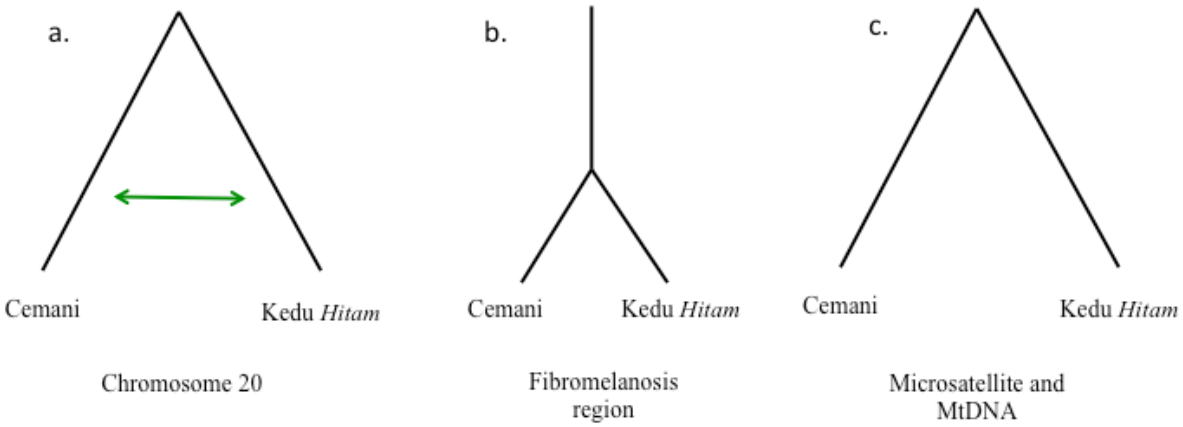
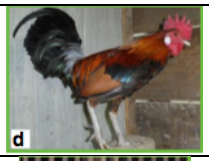
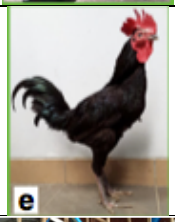


**Table 4.1.** List of chicken breeds, location and accession number.

| Location | Sample | Accession number | Figure | Reference (Accession; figure) |
|---|---|---|---|---|
| **INDONESIA** | **GJF (Java)** | DRR089960<br>DRR089961<br>DRR089962<br>DRR089964<br>DRR089965 |  | Ulfah et al. 2016; Ulfah et al. 2016 |
| | **RJF (Java)** | DRR089968<br>DRR089969<br>DRR089970 |  | Ulfah et al. 2016; Ulfah et al. 2016 |
| | **RJF (Sumatra)** | DRR089971<br>DRR089972 |  | Ulfah et al. 2016; Ulfah et al. 2016 |
| | **SUMATRA** | DRR089973<br>DRR089974<br>DRR089975<br>DRR089976<br>DRR089977 |  | Ulfah et al. 2016; Ulfah et al. 2016 |
| | **KEDU *HITAM* (BLACK KEDU(BK))** | DRR089989<br>DRR089990<br>DRR089992<br>DRR089996<br>DRR089997 |  | Ulfah et al. 2016; Ulfah et al. 2016 |
| | **CEMANI** | This study |  | This study; Dharmayanthi et al. 2017 |
| **AMERICA** | **BLACK SUMATRA (BS)** | DRR089983<br>DRR089984<br>DRR089985<br>DRR089986<br>DRR089987 |  | Ulfah et al. 2016; Ulfah et al. 2016 |
| | **BLACK JAVA (BJ)** | DRR089998<br>DRR089999<br>DRR090002 |  | Ulfah et al. 2016; Ulfah et al. 2016 |
| **JAPAN** | **WHITE LEGHORN** | DRR090008<br>DRR090009<br>DRR090010 |  | Ulfah et al.2016; Pham et al. 2013 |

| | | | | |
|---|---|---|---|---|
| **CHINA** | **EMEI** | SRR3036337<br>SRR3041115<br>SRR3041116<br>SRR3041121<br>SRR3041122 |  | Li et al. 2017; Li et al. 2017 |
| | **JIUYUAN** | SRR3041124<br>SRR3041125<br>SRR3041127<br>SRR3041128 |  | Li et al.2017; Li et al. 2017 |
| | **MUCHUAN** | SRR3041135<br>SRR3041136<br>SRR3041137<br>SRR3041138<br>SRR3041364 |  | Li et al. 2017; Li et al. 2017 |
| | **TIANFU** | SRR3041423<br>SRR3041425<br>SRR3041426<br>SRR3041427<br>SRR3041428 |  | Li et al. 2017; Li et al. 2017 |
| | **PENGXIAN** | SRR3041414<br>SRR3041415<br>SRR3041416<br>SRR3041417 |  | Li et al.2017; Li et al. 2017 |
| | **SILKIE** | SRX286765<br>SRX286766<br>SRX286773<br>SRX286776<br>SRX286777 |  | Fan et al. 2013; Dharmayanthi et al. 2017 |
| **TAIWAN** | **L2 TAIWANESE** | SRX286779<br>SRX286780<br>SRX286781<br>SRX286798<br>SRX286799 |  | Fan et al. 2013; Pham et al. 2013 |

**Table 4.2.** List of primer, primer sequence and restriction enzyme (Kinoshita's Method).

| Primer name | Sequence | Position | *Mlu*I RFLP |
|---|---|---|---|
| WT_F | TTCAGCAGCATTCACTGAAGGC | 11238619 - 11238640 | WT allele: 400 +264bp |
| FMdup1_664_R | ACCAACCCAGTAACCACAAGTG | 11239282 - 11239261 | Fm allele: uncut 664bp |
| FMspec_558_R | TGTCCATCTCACATTCTGGTGC | 11822233 - 11822254 | Uncut 588bp |

**Table 4.3.** List of chicken samples used in this study, parameters for SNP calling and result of total SNPs in all chromosomes, without insertions-deletions and total SNPs in chromosome 20 of chicken samples.

| Sample | Parameter of SNP calling (MC;MB;MF) | Result | | | |
|---|---|---|---|---|---|
| | | Coverage | # of SNVs | # of SNPs (no indels) | # of SNPs (chr20) |
| **Green Jungle Fowl** | | | | | |
| GJF60 | 5;2;35 | 5.7 X | 7024405 | 6258689 | 97708 |
| GJF61 | 6;2;35 | 6.2 X | 4983019 | 4473455 | 68848 |
| GJF62 | 6;2;35 | 7.3 X | 7501295 | 6708442 | 107771 |
| GJF64 | 6;2;35 | 7.0 X | 6542511 | 5867607 | 98614 |
| GJF65 | 6;2;35 | 4.8 X | 5972714 | 5369878 | 90108 |
| **Red Jungle Fowl** | | | | | |
| RJF68 | 6;2;35 | 7.4 X | 4526334 | 4040714 | 57797 |
| RJF69 | 6;2;35 | 7.1 X | 4480269 | 3999117 | 56949 |
| RJF70 | 6;2;35 | 6.8 X | 4161790 | 3716857 | 54184 |
| RJF71 | 10;4;35 | 10.9 X | 2535922 | 2274957 | 33980 |
| RJF72 | 10;4;35 | 10.9 X | 1577532 | 1409076 | 17863 |
| **Sumatra (Indonesian chicken)** | | | | | |
| Sumatra73 | 4;2;35 | 3.9 X | 2555834 | 2283547 | 37291 |
| Sumatra74 | 4;2;35 | 4.5 X | 3405198 | 3037201 | 46164 |
| Sumatra75 | 4;2;35 | 4.5 X | 3229795 | 2870750 | 46830 |
| Sumatra76 | 4;2;35 | 4.5 X | 3255559 | 2897449 | 46152 |
| Sumatra77 | 4;2;35 | 4.7 X | 3320934 | 2988595 | 49502 |
| **Black Sumatra (American chicken)** | | | | | |
| BS83 | 4;2;35 | 3.2 X | 1645966 | 1467772 | 22173 |
| BS84 | 4;2;35 | 3.4 X | 1896827 | 1689493 | 25295 |
| BS85 | 4;2;35 | 3.1 X | 1521712 | 1359945 | 25202 |
| BS86 | 4;2;35 | 3.4 X | 1969015 | 1753446 | 27350 |
| BS87 | 4;2;35 | 3.7 X | 2260281 | 2013435 | 33319 |
| **Black Kedu (Indonesian chicken)** | | | | | |
| BK89 | 8;2;35 | 7.1 X | 1831999 | 1649628 | 28878 |
| BK90 | 8;2;35 | 7.3 X | 1609332 | 1450860 | 22639 |
| BK92 | 6:2:35 | 5.9 X | 967984 | 873425 | 13382 |
| BK96 | 6;2;35 | 5.8 X | 2471851 | 2218158 | 31561 |
| BK97 | 6;2;35 | 6.0 X | 2653695 | 2381349 | 36793 |
| **Black Java (American chicken)** | | | | | |
| BJ98 | 8;2;35 | 7.7 X | 1939634 | 1747145 | 48379 |
| BJ99 | 6;2;35 | 6.3 X | 2111328 | 1899731 | 49484 |
| BJ02 | 8;2;35 | 6.8 X | 1503230 | 1353923 | 32695 |
| **White Leghorn (American chicken)** | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| WL08 | 15;6;35 | 34.6 X | 3545977 | 3238460 | 52753 |
| WL09 | 15;6;35 | 30.2 X | 3025831 | 2770793 | 41388 |
| WL10 | 15;6;35 | 36.9 X | 3742320 | 3406306 | 52163 |
| **Emei black (Chinese black plumage chicken)** | | | | | |
| Emei37 | 8;2;35 | 7.8 X | 2476377 | 2256887 | 52939 |
| Emei15 | 15;6;35 | 25.5 X | 4034067 | 3647662 | 72059 |
| Emei16 | 10;4;35 | 10.5 X | 2580477 | 2342169 | 59488 |
| Emei21 | 15;4;35 | 47.2 X | 6765086 | 5993202 | 94343 |
| Emei22 | 10;4;35 | 15.2 X | 2864339 | 2585836 | 63153 |
| **Jiuyuan Black-bone (Chinese black plumage chicken)** | | | | | |
| Jiuyuan24 | 8;2;35 | 8.7 X | 3414634 | 3097152 | 59732 |
| Jiuyuan25 | 15;6;35 | 31.1 X | 4823062 | 4345850 | 78207 |
| Jiuyuan27 | 15;4;35 | 15 X | 2450243 | 2209815 | 34556 |
| Jiuyuan28 | 15;6;35 | 16.4 X | 3463318 | 3126039 | 49755 |
| **Muchuan Black-bone (Chinese black plumage chicken)** | | | | | |
| Muchuan35 | 10;4;35 | 9.0 X | 2649486 | 2413382 | 55066 |
| Muchuan36 | 14;5;35 | 14.5 X | 2961841 | 2696306 | 62413 |
| Muchuan37 | 15;6;35 | 34.1 X | 6217926 | 5557196 | 83778 |
| Muchuan38 | 10;4;35 | 11.2 X | 3245833 | 2952902 | 62328 |
| Muchuan64 | 15;6;35 | 19.9 X | 4329104 | 3874469 | 59283 |
| **Tianfu Black-bone (Chinese black plumage chicken)** | | | | | |
| Tianfu23 | 15;6;35 | 19.3 X | 4163382 | 3770420 | 73237 |
| Tianfu25 | 30;10;35 | 57.8 X | 6524756 | 5826434 | 89583 |
| Tianfu26 | 8;2;35 | 7.9 X | 2680418 | 2435850 | 57219 |
| Tianfu27 | 10;4;35 | 11.2 X | 3552385 | 3224068 | 66617 |
| Tianfu28 | 15;4;35 | 15.4 X | 2685389 | 2416346 | 41149 |
| **Pengxian Yellow (Chinese local chicken)** | | | | | |
| Pengxian14 | 15;6;35 | 17.5 X | 3948442 | 3562619 | 59041 |
| Pengxian15 | 15;6;35 | 17.0 X | 3628814 | 3268119 | 49695 |
| Pengxian16 | 8;2;35 | 7.3 X | 3821299 | 3401932 | 51452 |
| Pengxian17 | 15;6;35 | 18.1 X | 4157369 | 3722002 | 52566 |
| **Cemani (Indonesian chicken)** | | | | | |
| CM41 | 15;4;35 | 31 X | 5392351 | 4790288 | 88340 |
| **Silkie (Chinese chicken)** | | | | | |
| Silkie | 15;4;35 | 28 X | 6338929 | 5598210 | 91358 |
| **L2 Taiwanese (Taiwanese chicken)** | | | | | |
| L2 Taiwanese | 15;4;35 | 33 X | 6400459 | 5591572 | 82486 |

(MC: minimum coverage; MB: minimum base; MF: minimum frequency; SNVs: Single Nucleotide Variants; SNPs: Single Nucleotide Polymorphism)

# Chapter 5.

## Conclusion and Future Perspective

The approach for detecting selection signatures using a single individual of whole genome sequence complemented with monomorphic validation on selected region was conducted in **Chapters 2** and **3** of my thesis. In **Chapter 2**, I detected selective sweeps in a duplicated region containing the target gene, *EDN3,* on chromosome 20; and this positive selection is associated with the Fm phenotype in Cemani and Silkie. Furthermore, I calculated the timing of selection and selection intensity in the *Fm* region containing *EDN3* of Cemani and Silkie chicken. Statistical analyses of sequences in flanking region of *EDN3* were also performed. In **Chapter 3**, I identified additional target genes under selection, despite using limited statistical analyses but rather approaches to detect monomorphism of 1 kb sequences in flanking regions of the target genes. Signatures of artificial selection was detected in *EGFR, NT5C1A* and LOC419677 by assessing monomorphism within Cemani population in the region surrounding the target genes. This suggested that such genes might be correlated with Cemani's specific traits, although more reliable evaluation for regions detected under selection is needed in the future to support my conclusion that *EGFR, NT5C1A* and LOC419677 were under selective sweeps. In addition, another approach for detecting the length of selection and calculating selection intensity in the genes is necessary.

In the future, I would like to determine and analyze the region surrounding *EGFR* on chromosome 2, *NT5C1A* and LOC419677 on chromosome 23 in order to provide more thorough data to support my conclusion in **Chapter 3**. To achieve this, I will use BAC clones that covered the surrounding region of target genes and use the BAC clone to sequence the regions by capturing method. Apart from Cemani, I will also use other

domesticated chickens including Silkie, Red and Green Jungle Fowl for comparison. The statistical tests for detecting selective sweeps in previous studies (Table 1.1 **Chapter 1**) such as Fst, EHH (extended haplotype homozygosity) and/or LD (Linkage Disequilibrium) can also be applied in future.

In **Chapter 2**, I identified that some Black Kedu chicken individuals had amplified duplication boundaries indicating these BK might have one *Fm* chromosome and one wild type chromosome, and this result was confirmed in **Chapter 4** by restriction enzyme digestion analysis. I also analyzed the genetic relationship of Cemani and Black Kedu chicken based on phylogenetic tree of 3 kb sequence in *Fm* region and principle component analysis (PCA) of SNPs on chromosome 20. The results revealed that these chickens shared ancestral polymorphisms. However, this analysis still did not support that Cemani was originated from Black Kedu chicken which contrasts with the previous studies based on microsatellite and mitochondrial DNA. Instead, my study in **Chapter 4** supports that Cemani is an independent breed that experienced introgression with Black Kedu chicken.

PCA analysis on chromosome 20 showed distinct clustering of chicken breeds based on their origin/location, suggesting that the main effect of geographical isolation on genetic differentiation of chicken breeds. Nevertheless, there is limited data to fully elucidate the relationship between Cemani and other black plumage chickens around the world. In the future, I want to analyze whole chromosome of genomes from various black plumage chicken breeds as well as the *Fm* region to reveal the history of black plumage chickens in the world. Moreover, for the origin of Cemani chicken, analysis of genetic differentiation within Indonesian chickens will be conducted to corroborate the hypothesis of independent breeds.

In general, despite only utilizing a single individual of whole genome sequence of each chicken breed, my thesis provides a better understanding of the phenotype-genotype relationship in chicken breeds and the effect of different factors on selection signature detection, but yet, many challenges remain. Further advances in selection signature detection within chicken genomes, especially in Indonesian Cemani chickens, require more accurate statistical approaches for determination of putative regions under selection. In addition, the results discussed in **Chapter 4** may give a better insight about the history of domestication in chickens, particularly in Cemani chicken, based on molecular approaches. In addition, further insight can be gained regarding the effect of geographical isolation to elucidate the genetic differentiation among chicken breeds with different geographical origins.