University of Natural Resources and Life Sciences,
Vienna Department of Sustainable Agricultural Systems
Division of Livestock Sciences
Animal Breeding Working Group

ISLANDS AND DESERTS:
Patterns of runs of homozygosity in chicken breeds

Masters Thesis

Juliet Orazietti

Supervisor:
Univ.-Prof. Dipl.-Ing. Dr. rer. nat. Johann Sölkner
Co-Supervisors:
Dr. ing. agr., Dipl. Ing. agr. Steffen Weigend
Dr. Gabor Meszaros
Prof. Dr. Erling Strandberg
Solomon Antwi Boison, MSc

Vienna
June 2015

Student number 1341480

Statutory Declaration

I certify, that the master thesis was written by me, not using sources and tools other than those quoted and without use of any other illegitimate support.

Furthermore, I confirm that I have not submitted this master thesis either nationally or internationally in any form.

Vienna, June 5, 2015

Signature (Juliet Orazietti, BSc)

Abstract

Runs of homozygosity are long, contiguously homozygous portions of the genome. They give us insight into inbreeding in populations. Regions have emerged where a very large or very small proportion of the population share the same ROH; these regions are knows as ROH islands and deserts. The overarching importance of these patterns of ROH is largely in terms of inbreeding. By understanding how much diversity exists in a population and where that diversity, or lack thereof, is distributed along the genome, animal breeders will be better able to preserve genetic diversity, target production-related alleles, and understand how individuals and populations are related. Studies on ROH have been done in humans, cattle, pigs, and chickens. ROH islands, have, until now, only been identified in humans and cattle. In this paper we will answer the following questions:

Do we find ROH islands and deserts that occur across a wide spectrum of chicken breeds with different breed history and origin? How are ROH islands and deserts in the chicken genome distributed? How gene dense are islands and deserts and how do they compare with the rest of the genome? What genes do we find in islands and deserts? What level of haplotype diversity is found in islands and how are the various haplotypes related?

Eighty-two diverse populations of chicken breeds originating from Europe and Asia were analysed. The breeds include many fancy breeds, commercial lines of layers and broilers, and two populations of Red Jungle Fowl. All populations were genotyped using the Affymetrix Axion ® Genome-Wide Chicken Genotyping Array 600k chip.

The analysis has found that ROH islands and deserts occur frequently in the chicken genome in both macro- and microchromosomes, and in all regions of chromosomes. The distribution of deserts in the subtelomeric chromosome regions is poorly understood and requires further analysis. Islands are found across all breed groups and are likely ancestral in origin. A wide range of gene densities is found in ROH islands. While nearly half of all islands contain no genes, others are quite gene dense. A wide range of genes with varying functions and pathways are found in islands. The haplotypes discovered in ROH islands are mainly shared across breed groups and indicate admixture. As with other characteristic of the chicken genome, the patterns of islands and deserts found in chickens is clearly different from those found in other species.
This necessitates a unique approach to the search for, and interpretation of, these fascinating regions.

Acknowledgements

For their guidance and support during the completion of my thesis, I would like to thank the following people. First and foremost, Univ. Prof. Johann Sölkner for his enthusiastic and inspirational support throughout my time at BOKU. Secondly, to Dr. Steffen Weigend for introducing me to the Synbreed project and providing the data that made this thesis possible. Also, Annett Weigend for her assistance during my stay at the Friedrich Loeffler Institute. I would like to thank Dr. Gábor Mészáros and Solomon Antwi Boison for all their help and advice in the research and writing process, and Prof. Dr. Erling Strandberg for his support.

I must express my huge debt of gratitude to my parents, Ann Sperling and Phil Orazietti, and family in always supporting me in my ambitions. Without their confidence in me I would not be where I am today. Special thank to Martin Weber for always pushing me. Finally to my friends all over the world, for always being there in spirit when I need them.

Table of Contents

1. Introduction

1.1 Importance & topicality

The development of high-density single nucleotide polymorphism (HD SNP) chips has opened up the field of genetics and the genomes of many species to much greater, more efficient analysis. With a greater number of SNPs analysed, the need for whole genome sequencing is largely eliminated since patterns and details can be discerned from the SNP data alone. For instance, with the 777k chip in cattle, the 64k chip in pigs, and the 600k chip in chickens it is possible to state with a high degree of certainty that if 100 contiguous SNPs are homozygous, it is almost certain that all the base pairs in between are as well. These long, homozygous regions of the genome known as runs of homozygosity (ROH), a term coined by Lencz et al. (2007), were first identified in humans by Broman and Weber (1999) who realised their potential for understanding inbreeding in populations. Since then, similar studies have looked into the homozygosity of various livestock species including cattle (Curik et al., 2014; Sölkner et al., 2014), pigs (Bosse et al., 2012), and chickens (Weigend et al., 2014).

Once ROH were identified in humans and cattle, patterns began to emerge (McQuillan et al., 2008; Nothnagel et al., 2010; Pemberton et al., 2012; Sölkner et al, 2014; Curik et al., 2014). Specifically, regions where a very large or very small proportion of the population share the same ROH. McQuillan et al. (2008) were the first to describe these regions as ROH islands and deserts. Pemberton et al. (2012) call them ROH hotspots and coldspots. We will use the original terminology in this study. For chickens, limited research into ROH in chickens has been done (e.g. Weigend et al., 20xx), and no one has investigated the patterns of ROH in chickens yet.

The overarching importance of these patterns of ROH is largely in terms of inbreeding; mapping of recessive disease alleles can be done, and more accurate estimations of inbreeding are possible. By understanding how much diversity exists in a population and where that diversity, or lack thereof, is distributed along the genome, animal breeders will be better able to preserve genetic diversity, target production-related alleles, and understand how individuals and populations are related.

To date, ROH have been used to estimate levels of inbreeding in individuals and populations. The way the calculations are done, however, does not allow for any diversity within ROH or ROH islands. A single ROH has no diversity, it is completely homozygous; On the other hand, two ROH may be very different from one another, even if they are found in the exact same position in the chromosome. To our knowledge, this study is the first to investigate the haplotypes found in ROH islands of any species. Knowing the genetic code of ROH islands, and how haplotypes relate to one another will give valuable insight into the relatedness of diverse chicken populations, and give a more colourful picture of the diversity within these fascinating, homozygous regions.

1.2 Aims

The main goals of this paper will be to investigate the following questions:

1. Do we find ROH islands and deserts that occur across a wide spectrum of chicken breeds with different breed history and origin?
2. How are ROH islands and deserts in the chicken genome distributed?
3. How gene dense are islands and deserts and how do they compare with the rest of the genome? What genes do we find in islands and deserts?
4. What level of haplotype diversity is found in islands and how are the various haplotypes related?

2. Literature review

2.1 Chicken domestication and migration

Chicken domestication began in Asia between 6000 and 8000 years ago, but intensive selection of breeds only started several decades ago (~1940s). Although the details are still controversial, it is thought that the wild ancestors of today's modern chicken originated in India and South-East Asia (Hillel et al., 2003; Tixier-Boichard et al., 2011). Of the ancestral species existing today, the family *Gallus gallus* most closely resembles modern chickens (*Gallus gallus gallus*) in its morphology. They are also still capable of cross breeding with domesticated chickens. Even the plumage of modern chickens still closely resembles that of *G. gallus*. Evidence has shown that domestication occurred independently in several locations of Asia and India (Kanginakudru et al., 2008). Archaeological evidence has shown that humans have been raising chickens for thousands of years; In Neolithic sites in northern China, Western Asia, and Europe chicken remains have been found from as far back at 6000 BC and in regions such as the Indus Valley of India from 2500 BC (Zeuner 1963; Crawford, 1990).

It is thought that chickens came to Europe via two main routes; the first though Persia and Greece (the Southern route), the second through China and Russia (the Northern route), however others claim that both routes began in Iran and travelled across the Mediterranean and the Black Sea. Historical records show that chickens had arrived in Greece by the year 700 BC. The modern Mediterranean chicken breeds are considered to be the most ancestral of the European chicken breeds based on morphology and molecular genetic data (Crawford, 1990). Chickens were bred by the Romans for food and leisure, and for physical appearance; the muff, beard, and rose comb are examples of unique phenotypes seen in Roman-age chickens (Tixier-Boichard et al., 2011). European chicken breeding began in the second half of the 19th century (Weigend et al., 2014). Of the modern chicken breeds, the Mediterranean types (the first chickens brought to Europe) are the most closely related to the RJF (Moiseyeval et al., 1996; Hillel et al., 2003). Pritchard et al. (2000) have found that when modern chickens are genotyped, Asian and European types can be clearly differentiated with brown layers deriving from the Asian type, white layers from the European type, and broilers from both.

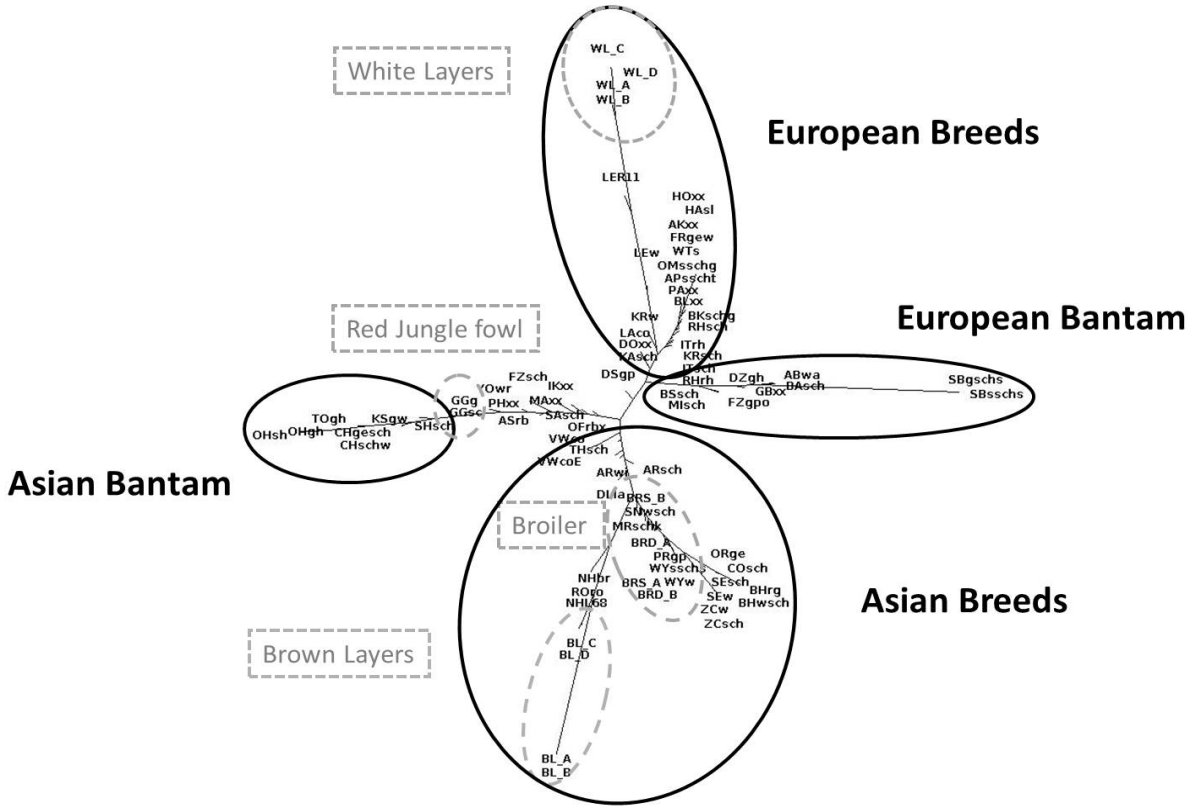


Figure 1: Neighbour Joining tree of 82 chicken populations based on PCA.

2.2 The chicken genome

The chicken genome was first published by the consortium headed by Hillier et al. in 2004. This has allowed for much greater understanding of the structure and function of the chicken genome. They discovered that the chicken genome, like most other avian karyotypes, is made up of chromosomes of significantly variable length. They describe them as macro- and microchromosomes. This is one of the significant differences between avian and mammalian genomes. The chicken genome is made up of 38 autosomes and a pair of sex chromosomes, Z (male), and W (female) with the female being heterogametic unlike in mammals (2n = 78) (Hillier et al., 2004). To this date, only the first 28 autosomes have successfully been mapped. The first 11 chromosomes are categorized as macrochromosomes with the rest being microchromosomes. One of the important characteristics described by Hillier et al. (2004) is the meiotic recombination rate of chromosomes. They report that macrochromosomes have a significantly lower rate of recombination than microchromosomes. This is the principle factor that leads to a lower rate of ROH in microchromosomes as will be discussed later. Since the publication of the genome, there has been constantly increasing knowledge of the genes determining phenotype, and disease and production characteristics (Weigend et al., 2014).

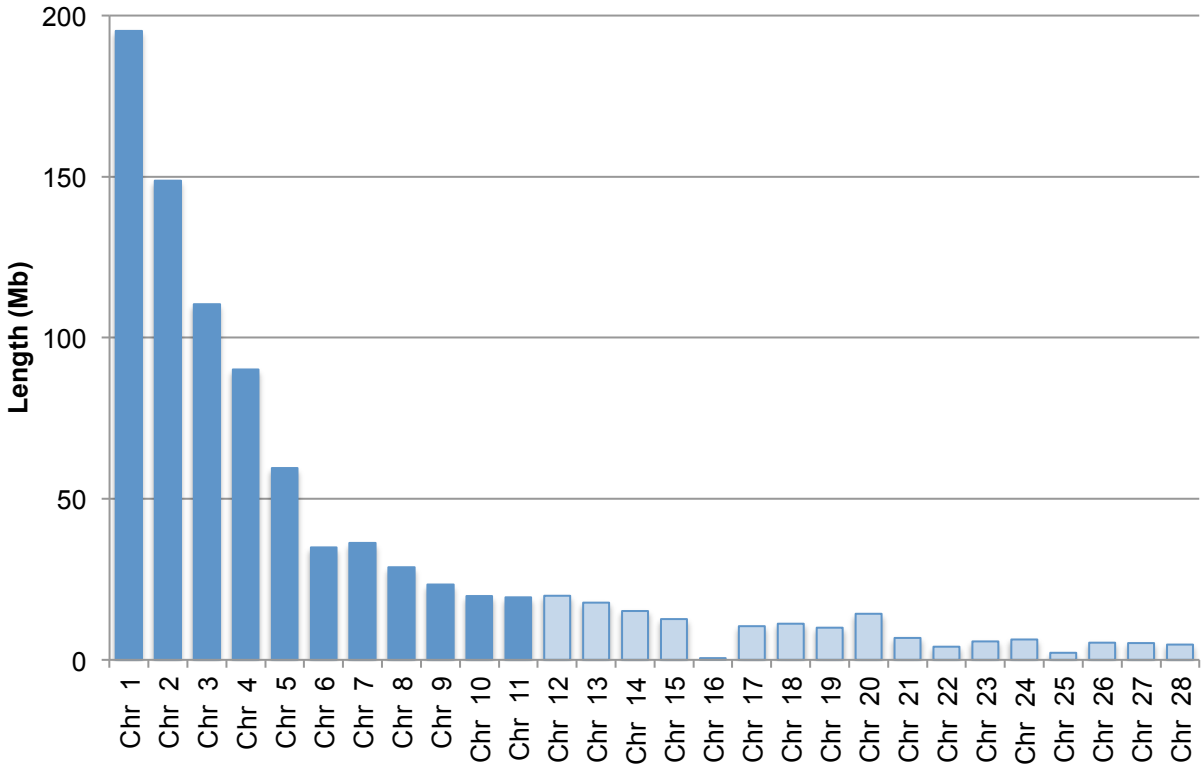


Figure 2: The chicken genome is broken down by length into macro- (dark blue) and microchromosomes (light blue). An additional 10 chromosomes have yet to be sequenced.

2.3 Runs of Homozygosity

Broman and Weber (1999) were the first to recognize that long stretches of homozygous segments in human populations likely reflect autozygosity. They proposed that these regions could have widespread effects on health and disease. They attributed this discovery to the availability of high-density panels of genetic polymorphisms. The term “run of homozygosity” was first used by Lencz et al. (2007). They defined a ROH as a window of at least 100 consecutive SNPs occurring on a single chromosome, that does not receive a heterozygous call when uncalled SNPs are permitted. Said plainly, that is 100 contiguous homozygous SNPs. These ROH are formed because the parents of the carrying individual have passed on identical haplotypes to their offspring (Lencz et al., 2007; Sölkner et al., 2014). The length and formation of these haplotypes are determined by the frequency of recombination events that take place during gamete production, specifically in meiosis (Sölkner et al., 2014).

Thanks to the development of HD SNP panels, it is now possible to determine with a high degree of accuracy whether or not specific SNPs occur in a ROH (Sölkner et al., 2014). As the concentration of SNPs in the genome increases, we can make better conclusions about the molecular traits of an individual, and with greater certainty than in the past. In the case of ROH, when SNPs are sufficiently close together we can say that if several adjacent SNPs are in homozygous state that the base pairs in between them are also homozygous. If a great enough number of homozygous SNPs are found contiguously, we can say that they form a run of homozygosity. For this reason it is important that SNPs are distributed equally along all chromosomes.

Since their discovery, ROH have been explored for a variety of things including the development of a new genomic inbreeding coefficient, F_{ROH} (McQuillan et al., 2008; Ferenčaković et al., 2013, Silio et al., 2013), genome-wide autozygosity mapping (Curtis et al., 2008; McQuillan et al., 2008), investigation of the effects of genome-wide autozygosity on important traits (Keller et al., 2012). Mapping ROH allows deleterious recessive variants to be identified in affected individuals so long as the inheritance is Mendelian in nature (Broman and Weber, 1999; Pemberton et al., 2012).

Pemberton et al. (2012) and many more (e.g. Frazer et al., 2007; McQuillan et al., 2008; Nothnagel et al., 2010) have found that ROH occur at relatively high frequencies in humans. The ROH seen in humans range in length from hundreds of kb to several Mb (Pemberton et al., 2012) and occur in either a small or large percentage of populations (as little as 2% in outbred, or as much as 90% in more consanguineous populations) (Pemberton et al, 2012). The length of a ROH is influenced by three main factors: how recent an inbreeding event occurred, natural selection, and a lack of recombination. An unusually low rate of recombination would allow long ROH to persist for longer than they would under normal conditions. In their investigation of 63 human populations from around the world, Pemberton et al. (2012) divided ROH into size-based groups and found that both the length of ROH and the cumulative ROH length depend on the geographical origin of the individual. They also found that ROH are more frequent in regions that have been associated with autosomal-dominant diseases (Pemberton et al., 2012). ROH also vary in animal populations depending on breed (Groenen et al., 2009).

2.4 Islands and Deserts

As ROH were studied more in depth, certain non-uniform patterns were found in the distribution of ROH (McQuillan et al., 2008; Nothnagel et al., 2010; Pemberton et al., 2012). McQuillan et al. (2008) described regions with unusually high ROH frequency as ROH islands and those with unusually low occurrences of ROH as ROH deserts. Pemberton et al. (2012) describe the same regions and hotspots and coldspots. ROH islands are regions with decreased genetic diversity (homozygous) while deserts are regions with unusually high levels of diversity (heterozygous). Pemberton et al. (2012) defined their ROH hotspots as the top 0.5% of SNPs that occur in ROH in a population; in their case SNPs that occurred in ROH in at least 30.34% of the population. Coldspots were defined as the 0.5% least common SNPs in ROH (2.72% of the population).

The reasons why these islands and deserts exist are not yet fully understood, and there are likely a variety of reasons for their development (linkage disequilibrium, inbreeding, selection, population history, mating system) (Nothnagel et al., 2010; Pemberton et al., 2012). Pemberton et al. (2012) attribute islands and deserts partially to stochasticity in recombination events or variation across the genome. In humans, demographic processes influence the development of hotspots for genetic diversity (deserts) and positive selection may increase the level of homozygosity around the target locus (islands). They also suggest that deserts may be enriched with loci for critical functions or coding lethal or negative recessive variants (Pemberton et al., 2012).

In their, study Pemberton et al. (2012) identified autosomal recessive diseases in ROH islands. Additionally, they searched for islands that occurred within populations. They were able to find patterns in the lengths and frequencies of islands dependent on geography and migration patterns. Specifically they found that ROH islands were shorter the further the population was from Africa. They attribute this to persistent migration of humans away from Africa, similar to the migration patterns of chickens out of Asia (Pemberton et al., 2012). Islands can also be used as strong signals of selection, indicating regions that have been targeted in breeding programs (Nothnagel et al., 2010, Pemberton et al., 2012, Kim et al., 2013).

2.5 Inbreeding

Previous studies have defined inbreeding in various terms. At the most general level it refers to the mating of two parents who share a common ancestor (Charlesworth and Charlesworth, 1987). All animals in a species share a common ancestor at some level of their pedigree though, and so this concept of inbreeding is subjective. A more useful way to think about inbreeding is as the consequence of mating two individuals who are more related to each other than the average relatedness of the overall population (Curik et al., 2014). The fact that every animal has some level of inbreeding (relatedness) is useful when we think about the principles of evolutionary genetics and how various chicken breeds are connected.

The crucial fact about inbreeding is that it increases homozygosity in the genome. When two parents pass on the same allele it is identical by descent, or autozygous. Changes in autozygosity can result in genetic variations within and between populations or breeds of animals, which allows us to differentiate between groups at the molecular level (Fernandez et al., 1995). If enough homozygous alleles line up we are left with a ROH. Inbreeding is not always self-evident. Although the main cause of a loss of heterozygosity may be poor breeding practices, bottlenecks, geographical barriers, and other natural selection phenomena can also cause a loss of diversity (Pemberton et al., 2012). When heterozygosity is lost, the chances of a deleterious or detrimental recessive allele being passed on in a homozygous state increases, which reduces the overall fitness of the population, a phenomenon known as inbreeding depression (Pemberton et al., 2012). If an animal is homozygous for a recessive disease allele, they are autozygous for the disease state and are bound to pass it on to their offspring. When homozygous portions of the genome are passed on, the longer the homozygous haplotype, the greater the chance that it is identical by descent (IBD), that is they have been passed on fro the same ancestral haplotype (parent) without recombination or mutation (Curik et al., 2014). This relationship is a tangible way in which relatedness and inbreeding are connected.

If a region is homozygous and IBD it can inform about the history of the population; the longer the haplotype, the more recent the inheritance is presumed to have been. This is because we expect a certain rate of recombination in meiosis. If the parent is homozygous, however, recombination will not result in a change in haplotype since there is only one possible allele to be inherited (Curik et al., 2014). When this is the situation we often find an increase in recessive-disease inheritance in a population due to autozygosity for the disease state (Pemberton et al., 2012). This is one of the reasons that it is important to maintain genetic diversity in animal populations. In their study of 52 chicken breeds, Hillel et al. (2003) found that chicken breeds range in heterozygosity, but overall have a higher level of homogeneity than other important species. Specifically, chickens were found to have a heterozygosity level of 0.47 on average while humans range from 0.7-0.8, cattle are 0.6, and pigs are 0.68 heterozygous. White layers were found to have the least diversity, consistent with their breeding background of selection from within a single breed (White Leghorns), brown layers are more diverse from their cross-bred origins (Rhode Island Red, New Hampshire, Plymouth Rock, and Australorp), and broilers are the most diverse as they originate from crossings of both Asian and European breeds (White Cornish, White Plymouth Rock) (Hillel et al., 2003). This is consistent with the findings of Granevitze et al. (2009) who investigated 65 chicken breeds from diverse origins.

Outside of mutation, we cannot regain lost genetic diversity. Chickens are an important species to humans across the globe as a source of nutrition and income. It is important to maintain what diversity still exists so that populations are able to adapt to evolving conditions in the future. As a result of the success of commercial chicken breeds, many dual-purpose breeds are now at risk of being lost. The high rate of homozygosity in commercial breeds is a risk to the future

success of commercial chicken production if production conditions change. More diverse, fancy breeds could carry the answer to future genetic challenges in commercial chicken production. For this reason, and for the preservation of our cultural heritage, it is important to understand and maintain the genetic diversity present in our extent breeds. In order to understand how best to preserve a population, we must first have a detailed, molecular understanding of the unique traits of that population (Hillel et al., 2003).

3. Materials and methods

3.1 The Data

3.1.1 Synbreed Chicken Diversity Panel

The Synbreed Chicken Diversity Panel (SCDP) has been established as part of the Synergistic Plant and Animal Breeding (Synbreed) project which aims to advance the knowledge and technologies needed to promote sustainable agricultural production. It was funded by the German Federal Ministry of Education and Research (BMBF). Overall, the SCDP contains samples from approximately 3000 individuals of 160 breeds from more than 22 countries around the world collected between 2010 and 2012. These breeds were selected to have a high degree of genetic and phenotypic diversity (Weigend et al., 2014). As an initial insight into the data, 2000 samples from 82 breeds have been genotyped. It is this 2000 animal data set that is discussed in this study.

Genetic and phenotypic data was collected with the consent of breeders under German Animal Welfare regulations and the authorities of Lower Saxony were notified in accordance with §8 of German Animal Welfare Act. Blood samples were collected from a wing vein using EDTA (ethylenediamine tetraacetic acid) as an anticoagulant (Weigend et al., 2014).

For the purposes of this study, only populations (breeds) with at least 15 (maximum of 20) sampled animals were considered. These 1677 individuals belong to 82 breeds (Table 1). While animals from 72 of the breeds were collected in Germany, they are presumed to come from diverse genetic backgrounds and to be purebred according to the German Association of Poultry Breeders (Bund Deutscher Rassgeflügelzüchter e.V., BDRG) and Society for the Conservation of Old and Endangered Livestock Breeds (GEH). Despite having been raised in Germany for about 150 years, the Asian breeds listed in Table 1 are purebred, and genetically distinct from the European breeds (Figure 1). Additionally, samples from two Red Jungle Fowl (RJF) and twelve commercial sire and dam lines (white layer, brown layers, and broilers) were used sourced from AVIANDIV (w3.tzv.fal.de/aviandiv/) and SYNBREED respectively (Weigend et al., 2014).

Table 1: Breeds of chickens analysed

Origin	Type	Breeds*	# breeds
Asia	Long tailed breeds	*PHxx, SAsch, YOwr*	3
	Game type and related breeds	*ASrb, IKxx, Maxx, OFrbx, SHsch*	5
	Asian type breeds	*BHrg, BHwsch, COsch, DLla, MRschk, NHbr, NHL68, ORge, PRgp, ROro, SNwsch, TOgh, WYsschs, WYw*	14
	Crested breeds	*SEsch, Sew*	2
Europe	Intermediate type breeds	*ARsch, ARwi, DOxx, VWco, VWcoE*	5
	Mediterranean type breeds	*ITrh, ITsch, KAsch, LER11, LEw, Misch*	6
	Northwest-European breeds	*AKxx, BKschg, BLxx, BSsch, DSgp, FRgew, HAsl, KRsch, KRw, LAco, OMsschg, RHrh, RHsch, THsch, WTs*	15
	Crested breeds	*APsscht, HOxx, PAxx*	3
Bantam	Asian type breeds	*CHgesch, CHschw, KSgw, OHgh, OHsh, ZCsch, ZCw*	7
	European type breeds	*ABwa, BAsch, DZgh, FZgpo, FZsch, GBxx, SBgschs, SBsschs*	8

Wild/RJF	Gallus gallus gallus	*GGg*	1
	Gallus gallus spadiceus	*GGsc*	1
Commercial	Broilers	*BRD_A, BRD_B, BRS_A, BRS_B*	4
lines	White Layer	*WL_A, WL_B, WL_C, WL_D*	4
	Brown Layer	*BL_A, BL_B, BL_C, BL_D*	4

** Chicken breed names*

ABwa - Barbue d'Anvers quail	KAsch - Castilians black	SHsch - Shamo black
AKxx - Carlise Old English Game any colour	KRsch - Creeper black	SNwsch - Sundheimer light
APsscht - Appenzeller Pointed Hood silver spangled	KRw - Creeper white	THsch - Thuringian Bearded Chicken black
ARsch - Rumpless Araucana black	KSgw - Ko Shamo black-red	TOgh - Toutenkou black breasted red
ARwi - Rumpless Araucana black breasted red	LAco - Lakenvelder black and white	VWco - Vorwerk buff columbian
ASrb - Aseel red mottled	LER11- White Leghorn line R11	VWcoE - Vorwerk conservation program
BAsch - Rosecomb Bantam black	LEw - White Leghorn	WTs - Westphalian Chicken silver
BHrg - Brahma gold	Maxx - Malay black red	WYsschs - Wyandotte silver laced
BHwsch - Brahma light	Misch - Minorca black	WYw - Wyandotte white
BKschg - Bergische Crower	MRschk - Marans copper black	YOwr - Yokohama red saddled white
BLxx - Brakel silver	NHbr - New Hampshire red	ZCsch - Pekin Bantam black
BSsch - Berg-Schlotter black	NHL68 - New Hampshire line 68	ZCw - Pekin Bantam white
CHgesch - Japanese Bantam black tailed buff	OFrbx - Orloff red spangled	
CHschw - Japanese Bantam black mottled	OHgh - Ohiki red duckwing	GGg - Gallus Gallus Gallus
COsch - Cochin black	OHsh - Ohiki silver duckwing	GGsc - Gallus Gallus Spadiceus
DLIa - German Faverolles salmon	OMsschg - East Friesian Gulls silver pencilled	
DOxx - Dorking any colour	ORge - Orpington buff	BRD_A - Broilers_A
DSgp - German Grey Chickens cuckoo	PAxx - Poland any colour	BRD_B - Broilers_B
DZgh - German Bantam gold partridge	PHxx - Phoenix golden or golden duckwing	BRS_A - Broilers_C
FRgew - Frisian Fowl chamois penciled	PRgp - Plymouth Rocks barred	BRS_B - Broilers_D
FZgpo - Booted Bantam millefleur	RHrh - Rhinelander Chicken brown	WL_A - White Leghorn line A
FZsch - Booted Bantam black	RHsch - Rhinelander Chicken black	WL_B - White Leghorn line B
GBxx - Barbue du Grubbe any colour	ROro - Rhode Island Red red	WL_C - White Leghorn line_C
HAsl - Hamburgh silver spangled	SAsch - Sumatra black	WL_D - White Leghorn line_D
HOxx - Poland White Crested black	SBgschs - Sebright Bantam golden	BL_A - Rhode Island Red line A
IKxx - Indian Game dark	SBsschs - Sebright Bantam silver	BL_B - Rhode Island Red line B
ITrh - Leghorn brown	SEsch - Silkies black	BL_C - White Rock line C
ITsch - Leghorn black	SEw - Silkies white	BL_D - White Rock line D

A principal component analysis (PCA) was carried out on the genetic and phenotypic data. This information was used to map out the relationships between the breeds and groups (Figure 1). The PCA showed that the breeds do segregate according to their supposed evolutionary origin (Weigend et al., 2014). Unsurprisingly, the PCA showed that white layers are a part of the European breed group, while brown layers are part of the Asian breed group. It is however, interesting that these two breeds of chicken are so disparate despite the fact that they have been selected for the same traits over many generations. In fact, these two breed groups are

found on opposite ends of the PCA plot. Asian and European bantam breeds were found to segregate independent of the Asian or European breed groups and from each other although the Asian bantams were found to be most similar to the RJF. Broiler lines segregated with Asian breeds but were close to the European group indicating a crossing of the two groups. These results are concurrent with the migration pattern and breeding records of the breeds (Tixier-Boichard et al., 2011; Weigend et al., 2014).

3.1.2 Genotyping

Genetic information was derived from blood samples using the Affymetrix Axion ® Genome-Wide Chicken Genotyping Array (referred to in this paper as the HDSNP array) which contains 580 961 SNPs in 28 autosomal chromosomes, two sex chromosomes (Z and W), and two linkage groups (LGE64 and LGE 22C19W28) (Kranis et al., 2013).

DNA isolation from the collected blood samples was done using the standard phenol-chloroform method **(CITE)**. SNP genotyping was carried out at the Chair of Animal Breeding, Technische Universität München. Annotation of the SNP data used Affymetrix "Axiom_GW_GT_Chicken.na33.annot.csv" (Weigend et al., 2014).

3.1.3 Quality control

Quality control measures were applied to the whole data set (2000 individuals, Table 1). A 99% SNP call rate and 95% animal call rate were set. This left the 1677 individuals and 82 populations with 15 to 20 individuals per population included in this study, and 445 264 SNPs.

3.1.4 Selection of SNP markers

The SNPs in the HD SNP array were selected based largely on polymorphic sites in sequences of commercial chicken lines (Kranis et al., 2013). This situation leads to ascertainment bias and means that the diversity of other breeds such as the fancy breeds in this study may be underestimated (Albrechtsen et al., 2010). In order to capture a more accurate picture, only SNPs found to be polymorphic in RJF were used in further analyses. This criteria eliminated about 45% of the HD SNP array and left 311 006 SNPs (Weigend et al., 2014). By using only RJF polymorphic SNPs we change the nature of the data presented. Most importantly, this may influence the number of ROH found in RJF populations since ROH are, by their nature, monomorphic.

The remaining SNPs were found to be equally distributed across the genome with the exception of the major histocompatibility complex (MHC) in chromosome 16 which was eliminated. It should be noted that the chicken genome shows a much higher gene density in the microchromosomes. The SNP loci on the HD SNP array have been uniformly distributed according to genetic, and not physical distance (Hillier et al., 2004; Kranis et al., 2013).

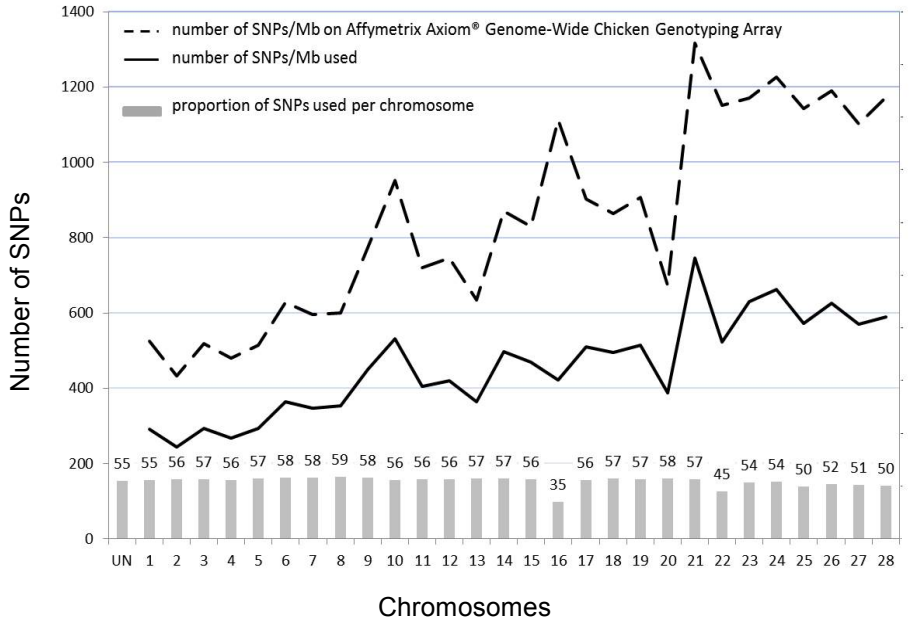


Figure 3: Chromosome-wise density of SNPs of the chicken HD SNP array (dashed line) and of the reduced SNP panel (solid line) used in this study. Gray bars indicate the proportion of loci of the reduced SNP panel per chromosome compared to the full set of loci of HDSNP array (Weigend et al., 2014)

3.2 Data analysis

3.2.1 Identification of Runs of Homozygosity

ROH were identified using SNP & Variation Suite v8.8 (SVS) from Golden Helix (www.goldenhelix.com). The ROH found were compared with those identified using PLINK software (Purcell et al., 2007) and the results were found to have a very high correlation of 0.99 (Curik et al., 2014; **personal communication Weigend, 2015)**. ROH were searched for with a window size of 300 kb (correlation of 0.98 if window size set to 100 kb) (personal communication Weigend, 2015). For this study a ROH must have been a minimum of 100 consecutive SNPs long, with a maximum gap size of 1000 kb between SNPs. One heterozygous and two missing SNPs were allowed per window to allow for genotyping errors (Weigend et al., 2014). These criteria were set based on those of Lencz et al. (2007) and ad hoc scripts in the R software package (R Core Team, 2013). After the ROH data were isolated, missing values were imputed based on the most common allele in that position in the rest of the population with a ROH in that location. Heterozygous values were left as is since it is impossible to know if they were coding errors.

3.2.2 Identification of Islands and Deserts

Once ROH were identified, each SNP was recoded in binary as 1 (in a ROH) or 0 (not in a ROH). The definition of a ROH island is made up of arbitrary limits, for example Nothnagel et al. (2010) defined islands as SNPs that occur in ROH in greater than 50% of the population. Pemberton et al. (2012) defined “hotspots and coldspots” as the top and bottom 0.5% of SNPs occurring in ROHs. Here, two definitions of island were used for investigation of the chicken genome.

Fixed percentile islands are defined as the top 1% of SNPs that occur in ROH across all breeds. Fixed percentile deserts are defined as the 1% of SNPs that occur least frequently in ROH across all breeds (Curik et al., 2014; Sölkner et al., 2014). This means that islands and deserts, as defined, are regions of homozygosity that occur in a significant portion of the population across breeds. Percentiles were calculated in SVS using sums of binary ROH data for each SNP. SVS tallied the number of times a SNP occurred in a ROH across all individuals. SNPs that occurred more than 906 times or in at least 56.6% of individuals (99^{th} percentile = islands), or less than 467 times or in no more than 29.2% (1^{st} percentile = deserts) were then isolated for further analysis. The width of an island or desert region was determined by the positions of the SNPs at the left and right limit that were over the 56.6% or 29.2% threshold. Fixed width islands and deserts use the same binary sum data but give each island or desert the same width (2 Mb). Fixed width islands and deserts are centred on the peak or trough of the fixed percentile island or desert and extend 1 Mb upstream and downstream from this most or least frequent SNP respectively.

Chromosome number, start and end position on the chromosome, SNP names, and the number of SNPs in each region were recorded. Islands and deserts containing three or fewer SNPs were discarded.

3.2.3 Island and Desert distribution

Island and desert distribution trends were calculated using data from SVS and Ensembl (Cunningham et al., 2015) in Excel (Microsoft, 2010). The number, and length of these regions as well as the proportion of each chromosome taken up by islands was analysed. Relationships between chromosome number and type (micro-, or macrochromosome), and chromosome or island length were also found via linear correlations.

3.2.4 Gene Density

Gene density in island and desert regions was calculated (number of annotated genes per Mb) and compared to overall chromosome gene density (number of genes per Mb). Genes (annotated regions) in islands and deserts were recorded from Ensembl.

3.2.5 Haplotype Analysis

The relationships between the haplotypes found in each fixed percentile island is shown via Integer Neighbour Joining Networks (INJN) made using the software PopART (http://popart.otago.ac.nz). The algorithm used to calculate these networks is fully described in Bandelt et al., 1999. INJNs show the number of mutations between haplotypes between the nodes. Additional nodes (shown in black) are added to infer a missing ancestor or haplotype (http://popart.otago.ac.nz).

The algorithm used in calculating haplotype frequencies in a given ROH island was based on the following: for a given ROH island, both the maternal and paternal haplotypes are identical, thus for a given haplotype, frequencies were computed by counting the number of times a consecutive homozygous genotype appear within a population divided by two times the total number animals within a population (only individuals in the island were considered). Haplotype frequencies were calculated within each population and overall populations. Custom R scripts were used in building haplotypes and computing summary statistics. The frequencies of haplotypes within groups, and the proportion of each haplotype across all groups was calculated and plotted in R (R Core Team, 2013).

4. Results and discussion

4.1 Identification of islands and deserts and their distribution

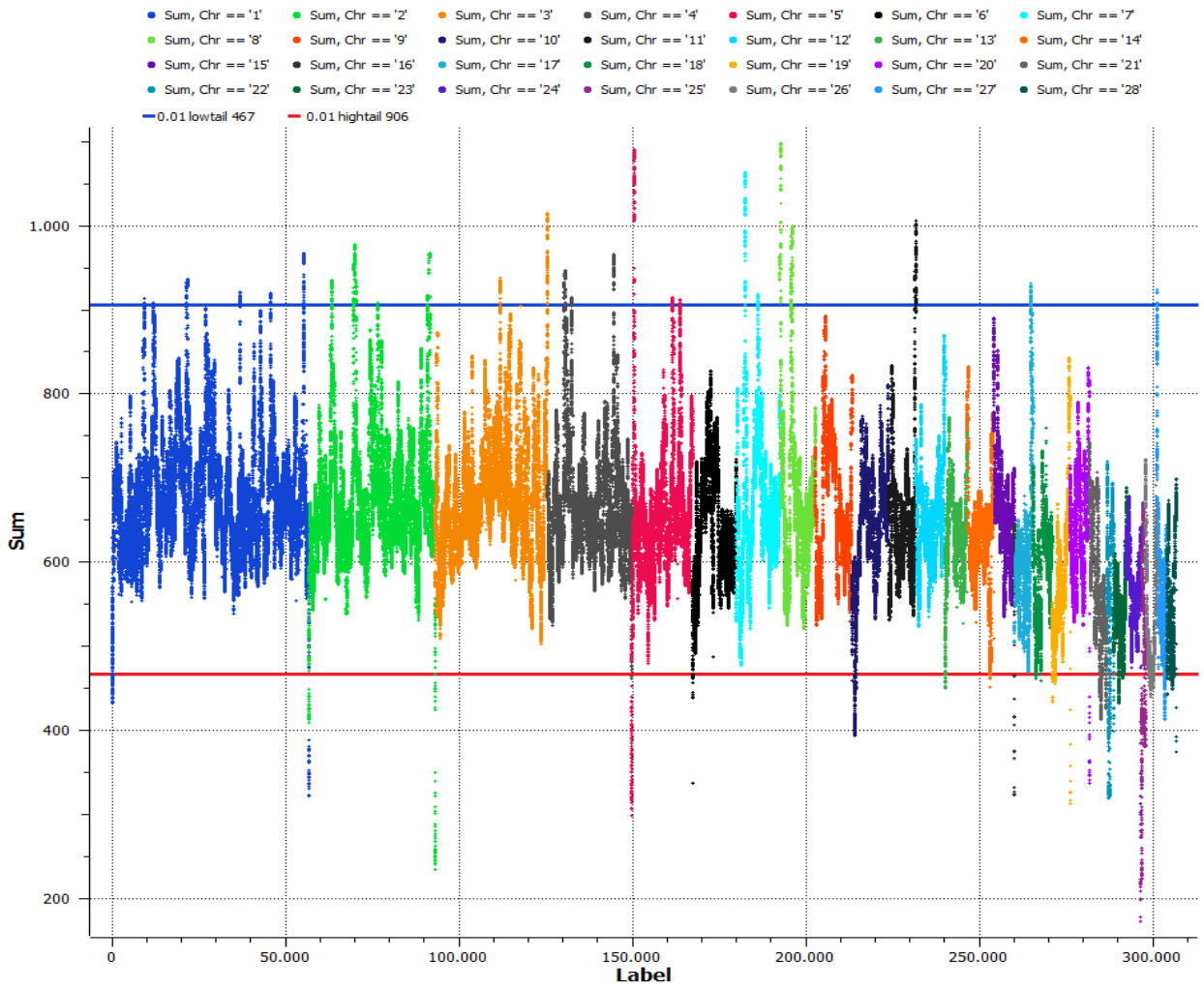


Figure 4. All peaks above the blue line (99th percentile) indicate islands. Troughs below the red line (1st percentile) indicate deserts.

Figure 4 shows the frequencies at which SNPs occurred in ROH. Each data point represents a SNP that was found in a ROH at least once. Twenty-seven islands were identified with greater than three SNPs. In several instances islands were found that were separated by only several SNPs which were just below the 99th percentile threshold. In cases where five or fewer SNPs separated islands, the two regions were combined into one. Many deserts were identified but these were found to be mainly in microchromosomes, and exclusively in subtelomeric regions and so were eliminated (Figures 5 and 6). Subtelomeric regions are the first and last 10 Mb of a chromosome in the chicken genome and are poorly understood. They tend to have different properties from the rest of the chromosome. For instance they show a much higher rate of synonymous substitution and polymorphism which would inherently reduce the possibility of ROHs occurring in these regions (Hillier et al., 2004) Islands are most frequently found in macrochromosomes (25 of 27). There are only two microchromosomes with islands (17, 27). These microchromosome islands may be especially significant given the high rate of

recombination seen in chromosomes of this size (Hillier et al., 2004). If recombination is not occurring at the expected rate in these parts of the microchromosomes it could indicate that these genes must remain together to retain their functionality, and must be important for the success of the species.

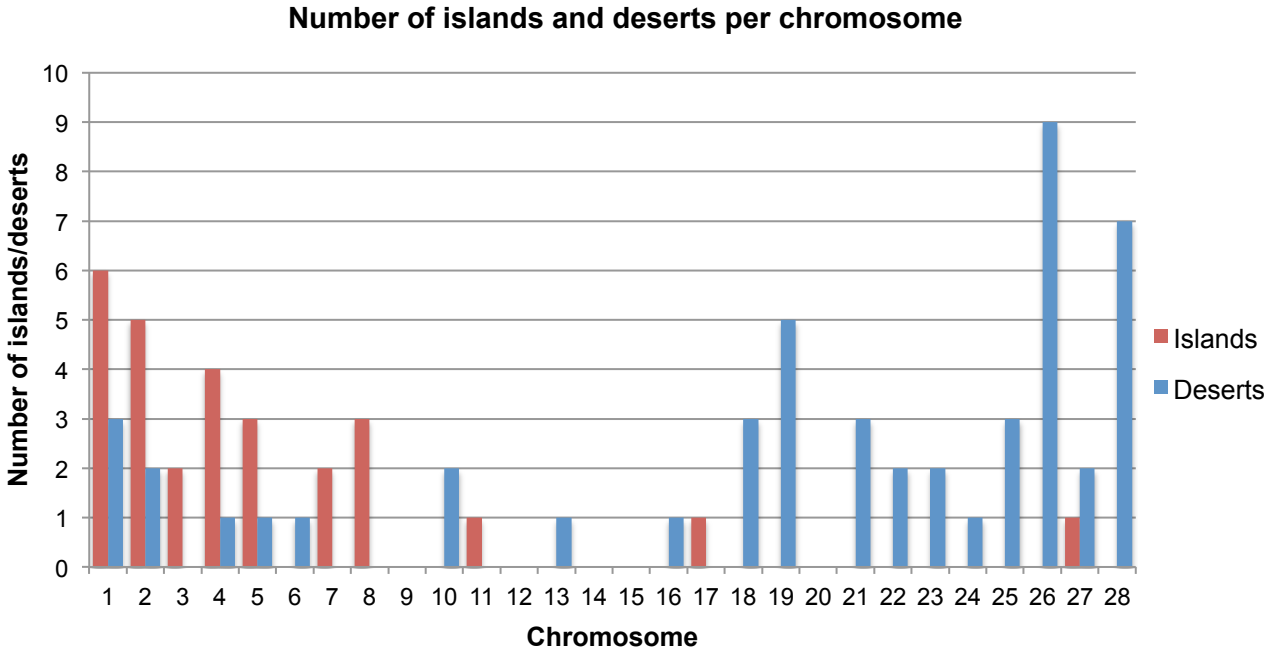


Figure 5: The distribution of islands and deserts in the chicken genome

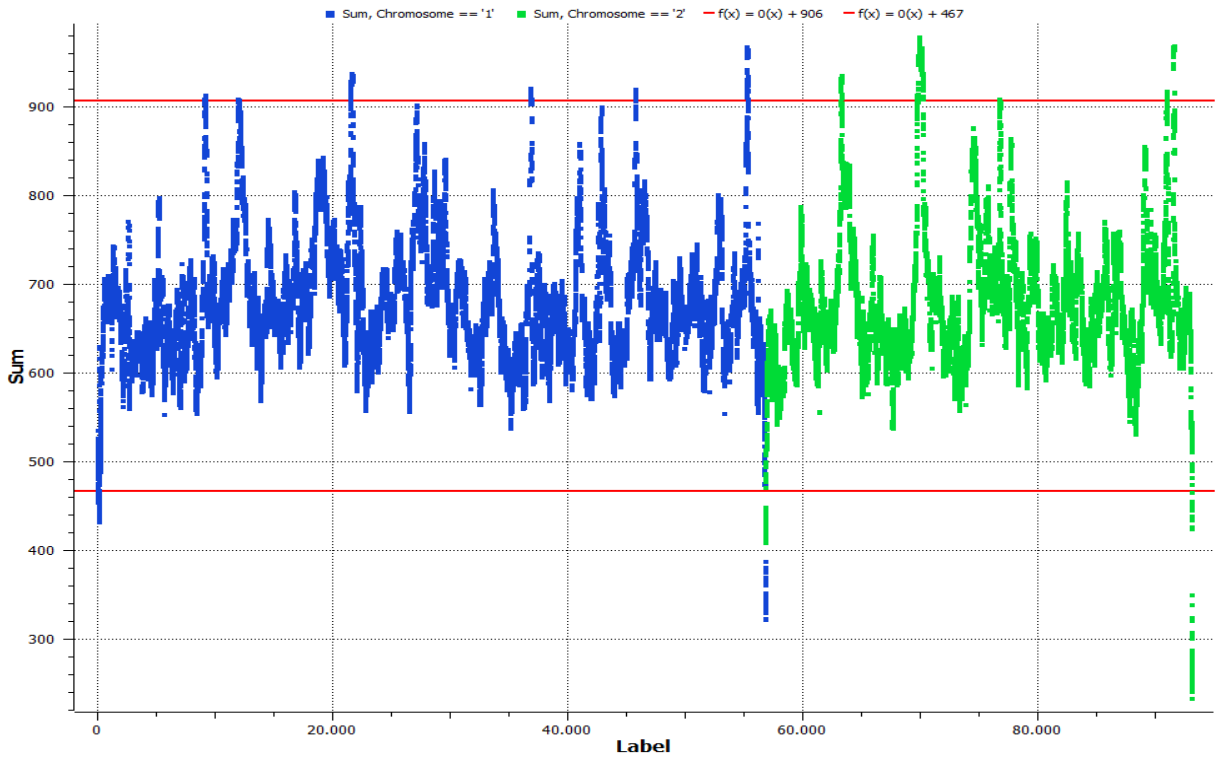


Figure 6. The first four deserts occur at the beginnings and ends of chromosomes 1 and 2.

The lack of clear deserts in the chicken genome is surprising. Karimi (2013), Curik et al. (2014), and Sölkner et al. (2014) found clear patterns of islands and deserts when similar protocols were used to investigate cattle breeds, and deserts were found in the human genome by Nothnagel et al. (2010), and Pemberton et al. (2012). Perhaps a different definition of a desert is needed for the chicken genome. Due to the lack of a clear pattern in deserts and limited knowledge of the end of chromosomes (Pemberton et al. (2012) also discarded the first and last 60 SNPs in their human SNP chip data due to a lack of knowledge of the beginnings and ends of chromosomes), we will no longer consider deserts in the analysis.

The islands found ranged in length from 8 to 481 SNPs or 0.0243 to 2.492 Mb with an average length of 115.070 SNPs or 0.608 Mb. The islands also contained between 0 and 27 genes (average 6.44) and had a gene density ranging from 0 to 50.765 genes/Mb (average 8.122)

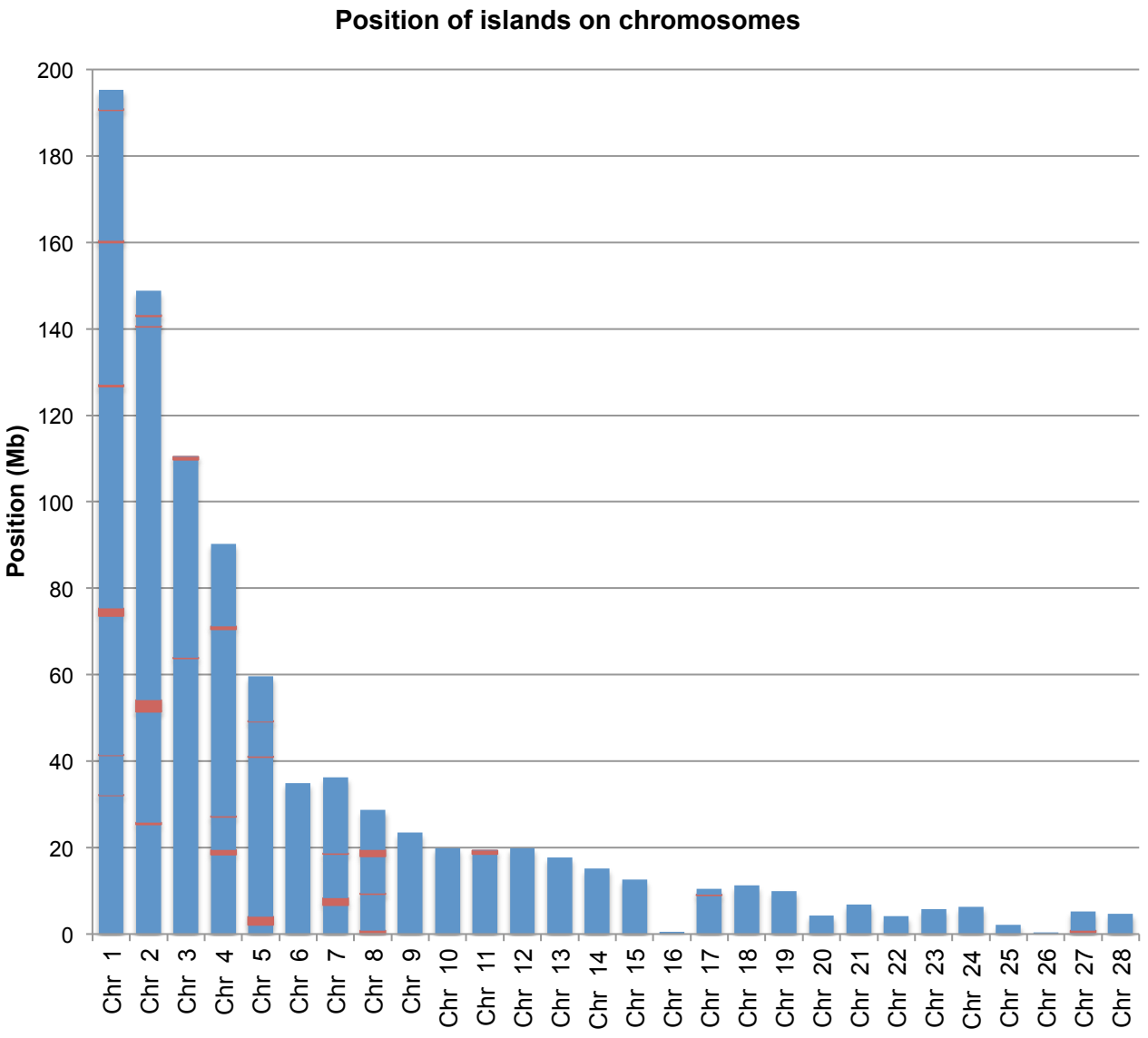


Figure 7: The positions of islands on chicken chromosomes

ROH are known to be less common in microchromosomes due to their high rate of polymorphism, recombination, and synonymous substitution (Mefford and Trask, 2002; Hillier et al., 2004). In fact, microchromosomes have a recombination rate eight times greater (2.5 to 21 cM/Mb) than that seen in macrochromosomes. Even the lower rate seen in chicken macrochromosomes is much higher than in mammals (0.5–2cM/Mb) (Stehelin et al., 1976; Hillier et al., 2004; Groenen et al., 2009). Pigs, like chickens, display varied levels of recombination depending on the location of the genome (Bosse et al., 2012). Furthermore, recombination rates have been shown to vary between breeds, possibly as a result of intensive breeding (Groenen et al., 2009). Figure 4 shows a decreasing trend of SNP occurrence in ROH in the microchromosomes as is expected. The variation in recombination rate along the chicken genome suggests that perhaps a varying definition of ROH, and islands and deserts, is needed depending on the region of the genome. This is something that could be tested in future but which has no precedent in other species. Secondly, we may find different or additional islands in different breeds or groups of breeds if analysed separately. This should be done in future and could provide valuable insight into how breeds or groups of breeds differ from one another.

This analysis has found that the number and length of islands tends to be greater in larger chromosomes (Figure 5). Conversely, the proportion of each chromosome taken up by fixed percentile islands tends to increase with decreasing chromosome length except in the microchromosomes where the data is insufficient to find any trend (Figure 8**)**. We also see that chromosome 8 has the greatest proportion of its length in islands with nearly 8.4%.

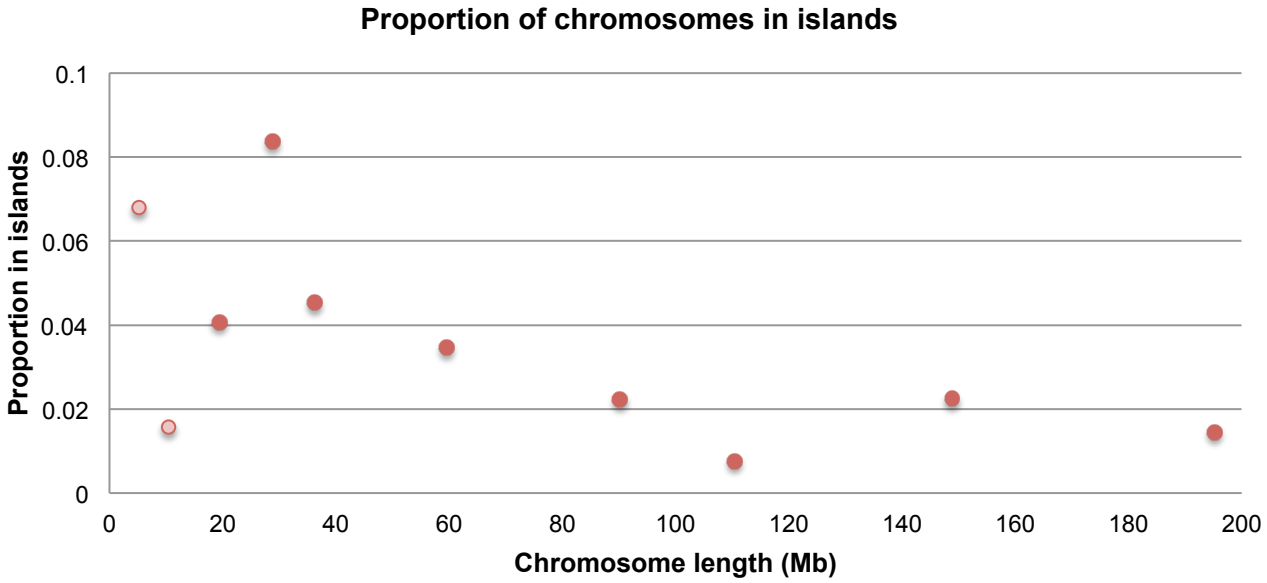


Figure 8: The proportion of microchromosomes (pink) in islands does not increase in the same way as in macrochromosomes.

Figure 9 shows that the length of a fixed percentile island is proportional to the number of SNPs in the island. Weigend et al. (2014), found that the SNPs used in this study are genetically equally distributed (Figure 1). This could account for the imperfect linear relationship (r = 0.925).

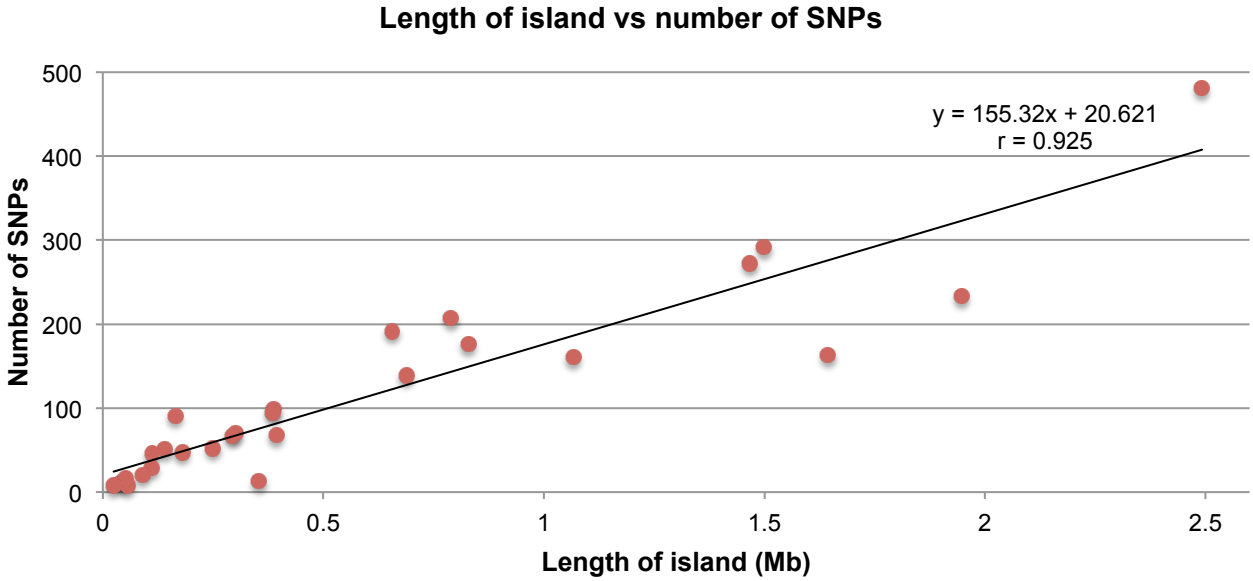


Figure 9: SNPs occur at about the same density in all islands.

4.2 Gene Density

Thus far, the distribution and characteristics of islands across the genome seems to follow the expected trends; islands are more common in regions of lower recombination (macrochromosomes), and SNPs are distributed equally within islands. When the contents of these regions are further analysed, however, we find, unexpectedly, that a large number of the islands contain no annotated regions (Cunningham et al., 2015). This is surprising since we would expect highly conserved regions of the genome to contain important genetic material. Others (Karimi 2013; Curik et al., 2014**)** have found ROH islands to be especially gene-rich in cattle. These non-coding islands must, then, be areas influenced by inbreeding. Given that the islands found in this study are presumed to be ancestral, since they are shared across all populations, the common ancestor must have experienced a bottleneck or other intense selection pressure which caused this high rate of homozygosity (especially) in chromosome 1 and which it passed on to its modern progeny**.** Specifically, 44% of islands contained no coding regions at all. These non-coding islands range from 8 to 139 SNPs in length (average of 51.91) and 0.024 to 0.393 Mb (average 0.221 Mb). About 42% of the non-coding islands are found on chromosome 1.

Table 2: Island location and gene data

Island	Location	# SNPs	# genes	Genes
1.1	32105873-32150135	12	0	
1.3	41305981-41361805*	8	0	
1.5	73659173-75302834	163	20	KCNA1, ENSGALG00000025798, KCNA5, NTF3, ANO2, VWF, CD9, NOBOX, ARHGEF5, PRMT8, PARP11, TSPAN9, TEAD4, TULP3, FOXM1, ITFG2, NECAP1, FOXJ2, FKBP4, C12ORF32
1.6	126607488-126994702 **	99	0	
1.7	159978139-	68	0	

	160371348			
1.8	190571514-190871495	70	0	
2.1	25348292-25734199	94	1	NXPH1
2.2	51437340-51547872****	29	1	GLI3-201
2.3	51580308-54072586	481	14	ENSGALG00000012332, C7orf25, HECW1, MRPL32, STK17A, ENSGALG00000012353, VOPP1, BLVRA, LANCL2, SEC61G, TPK1, 7SK, VOPP1, BLVRA, LANCL2, SEC61G, TPK1, 7SK, ENSGALG00000026036, ENSGALG00000012389,
2.5	140506904-140558276	16	1	ADCY8
2.6	142925250-143218775	67	0	
3.1	63776707-63957074	47	0	
3.2	109772365-110428731	191	5	C6orf138, EVA1A, ENSGALG00000016731, ENSGALG00000016732, ENSGALG00000029012
4.1	18391954-19459250	161	8	AFF2, FMR1MB, FMR1, SLITRK2, TLR2-2, TRIM2, MND1, KIAA0922
4.2	27055953-27306201**	52	0	
4.4	70532800-71221544	139	1	PCDH7-201
5.1	2091956-4038938	163	22	PRMT3, SLC6A5, NELL1, gga-mir-1775, ANO5, SLC17A6, ENSGALG00000026649, FANCF, GAS2, SVIP, ANO3, SLC5A12, FIBIN, BBOX1, CCDC34, LGR4, BDNF, ENSGAL00000012162, METTL15, 7SK, gga-mir-1760, KIF18A
5.2	40954167-41043854*	21	0	
5.4	49119327-49143659	8	0	
7.1	6744362-8243239	292	23	COL6A1, COL6A2, FTCD, ENSGALG00000006133, MCM3AP, FUT13, S100B, ENSGALG00000002213, LLS, TUBA1C, PCNT, C21orf58, KMO, FAM207A, ITGB2, GLS, STAT1, STAT4, MYO1B, ENSGALG00000007724, NABP1, PRKCDBP, TMEFF2
7.3	18530576-18671971	51	2	CERS6, STK39-201
8.1	9100-838224	176	3	AMY2A-201, ENSGALG00000021244, NTNG1-201
8.2	9229140-9340962	46	0	
8.3	9428942-10896059	272	25	ENSGALG00000028389, PLA2G4A, PTGS2, PDC, C8H1orf27, TPR, PRG4, RNPC3, COL11A1, gga-mir-6561, OLFM3, S1PR1, ENSGALG00000020884, DPHS, SLC30A7, CDC14A, VCAM1, RTCA, DBT, LRRC39, TRMT13, SASS6, gga-mir-1610, EXTL2, GPR88
11.1	18521596-19330411	207	19	ZFHX3, ENSGALG00000019806, ENSGALG00000019804, SNF821, AP1G1, PHLPP2-201, TAT-201, CHST4-201, KARS-201, ADATI-210, PSMD7-201, gga-mir-1699, ENSGALG00000019805, FHX38-201, IST1, ATXN1L-201, ENSGALG00000000882, TERF2IP-201, SNORD71
17.1	8930249-9094716	91	3	STRBP, CRB2, DENDD1A
27.1	446512-801088	13	18	ENSGALG00000026257, ENSGALG00000028366, ENSGALG00000028287, ENSGALG00000027679, FK21 (X7), ENSGALG00000027859, ENSGALG00000025813, FK27 (X2), ENSGALG00000025863, ENSGALG00000026660, ENSGALG00000026219

* Indicates the number of SNPs below the 99th percentile threshold

Table 3: Island gene density statistics.

	Minimum gene density (genes/Mb)	Maximum gene density (genes/Mb)	Average gene density (genes/Mb)
Fixed percentile	0.000	50.765	8.122
Fixed width	6.000	32.500	11.222

When the number of genes per island was counted, the gene density in islands (number of annotated regions per Mb in an island) was found to increase with decreasing chromosome size (r = 0.815), which is consistent with the general trend of gene density in chromosomes (Cunningham et al., 2015). When only islands containing genes are considered the correlation remains high at r = 0.857 (see appendix).

The way islands are defined means that they are flanked by SNPs less frequently in ROHs; the islands capture the SNPs most commonly in ROH, but there are many more SNPs upstream and downstream that would have been cut off the ends of the islands. For this reason, a second island description was defined. The largest island set with the percentile threshold was over 2 Mb long. A fixed width of 2 Mb was thus given to all the islands in an attempt to capture ROH SNPs lost due to the strict percentile threshold. The 2 Mb islands are centred on the central peak (SNP most frequently in a ROH) of the fixed percentile islands. The fixed width islands and the fixed percentile island show a moderate correlation of r = 0.789 in their gene densities (Figure 12). This can be explained by the fact that none of the fixed width islands are gene-free, while a large number of the fixed percentile islands are. The concept of ROH islands is still relatively new and the definitions and limits set upon them are arbitrary. Therefore, at this point we cannot say whether the fixed percentile or fixed width islands are more valid, just that they show similar trends. The effect of lengthening the islands can be seen in Figure 11 where there are no longer any regions that do not contain any coding elements (minimum of one).

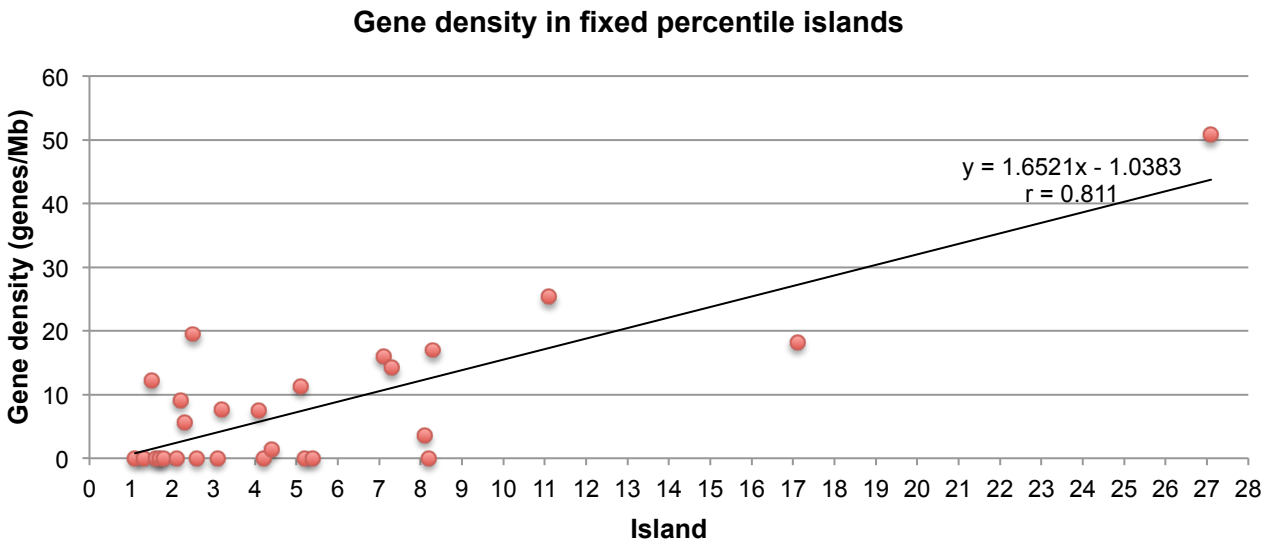


Figure 10: Many macrochromosome islands do not contain any genes, decreasing the overall gene density.

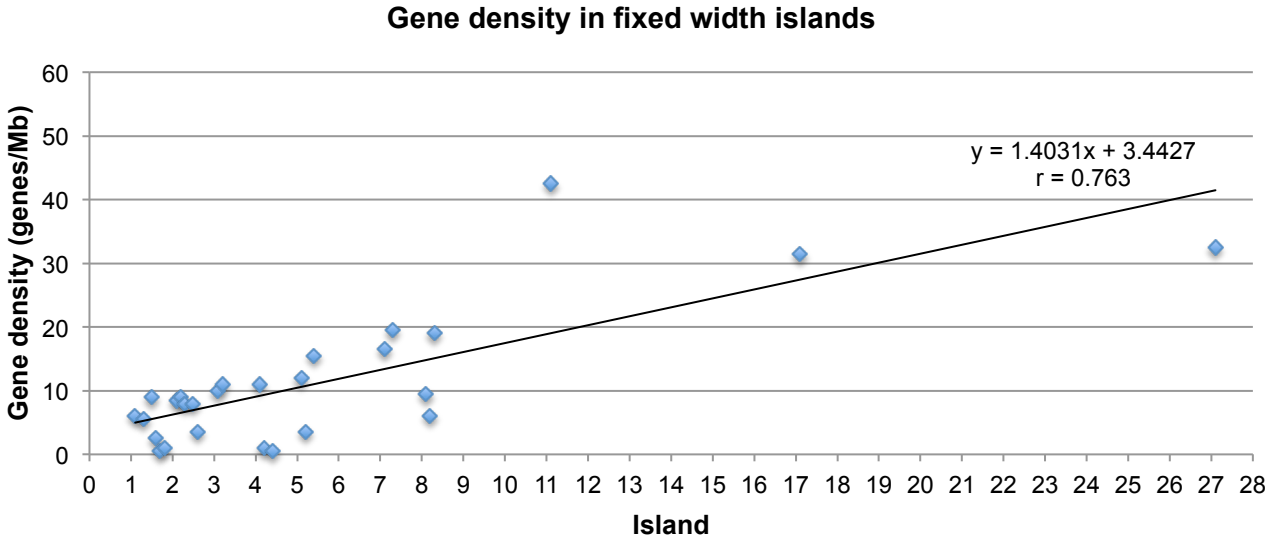


Figure 11: All islands contain at least one gene, increasing the overall gene density.

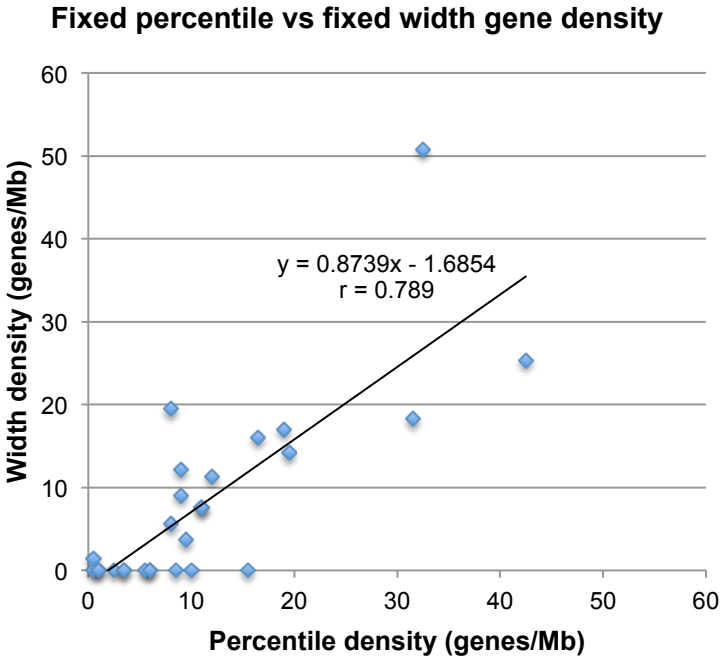


Figure 12: Gene density in fixed width islands increases at a slightly greater rate than in fixed percentile islands.

Knowing that chromosome length decreases from chromosome 1 to 28 we can see from Figures 10 and 11 that gene density in islands increases as chromosome length decreases. This is also true of chromosomes as a whole (Hillier et al., 2004; Cunningham et al., 2015). If the gene density of islands is the same as the overall chromosome gene density we should expect that island gene density will increase at the same rate as chromosome gene density. In

fact we see only a moderate correlation between gene density and chromosome length whether we look at the fixed percentile islands (r = 0.653) or the fixed width islands (r = 0.663) (Figure 13). The low correlation indicates that there is an increase in gene density in islands as compared with the rest of the chromosome but whether or not this is significant is unclear. Figure 13 clearly shows the final two shortest islands (those on chromosomes 17, and 27) are outliers from this trend which weakens the relationship. This is another example of how microchromosomes should perhaps be treated differently in the search for ROH and ROH islands. The trend is further obscured when only macrochromosomes are considered (r = 0.216-0.336), and is only slightly improved when only island containing genes are included (r = 0.482) (Figure 14).

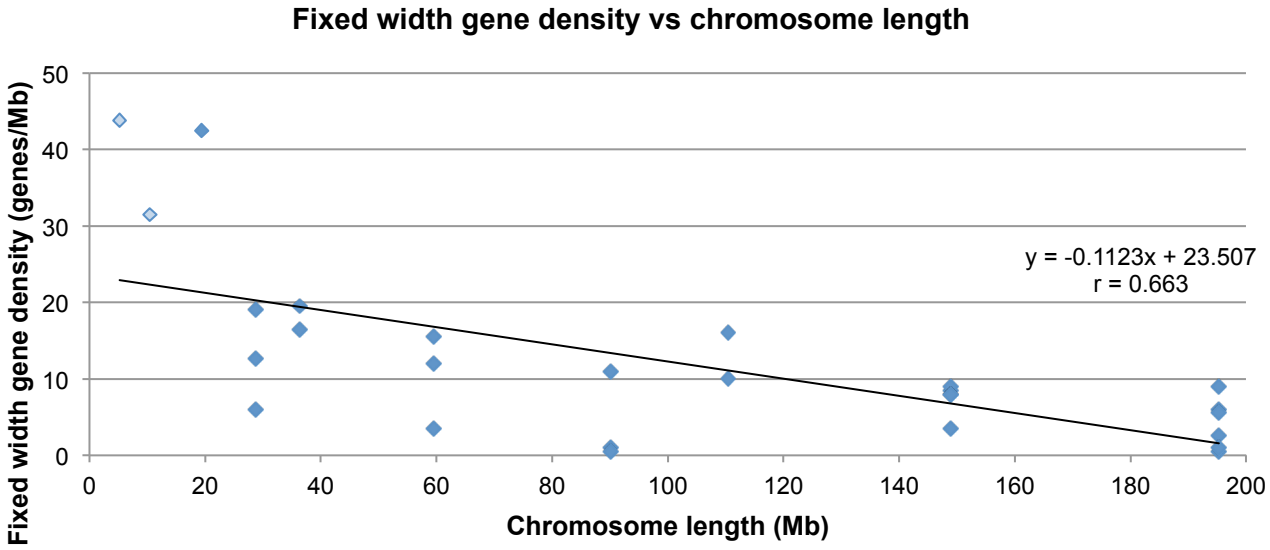


Figure 13: Microchromosomes have a higher gene density than macrochromosomes

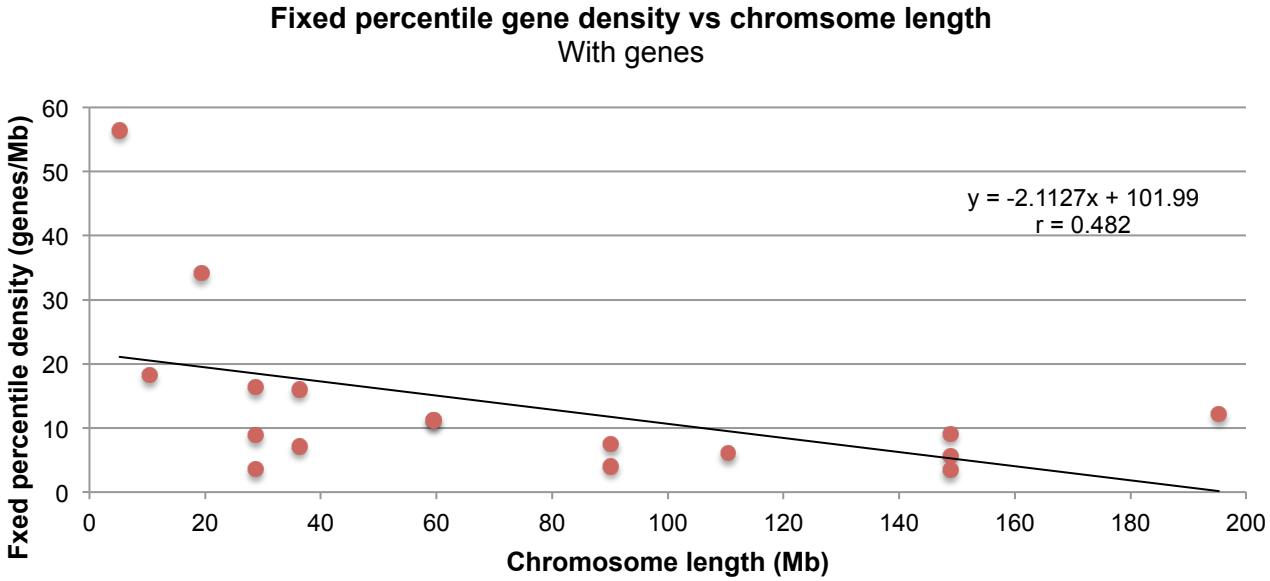


Figure 14: Islands containing genes do not have proportionately linearly increasing gene densities when compared with chromosome length

Figure 15 shows the relationship between island density and chromosome density. The increase in gene density of chromosomes increases linearly with the increase in gene density in islands (r = 0.778 and 0.672) showing that while there is not a large increase in gene density in islands as compared with the rest of the genome as has been seen in cattle (Karimi, 2013). Similar to the comparison with chromosome length, when we look only at the macrochromosomes we see that there is a moderate relationship (r = 0.463) between island and chromosome gene density (Figure 16).

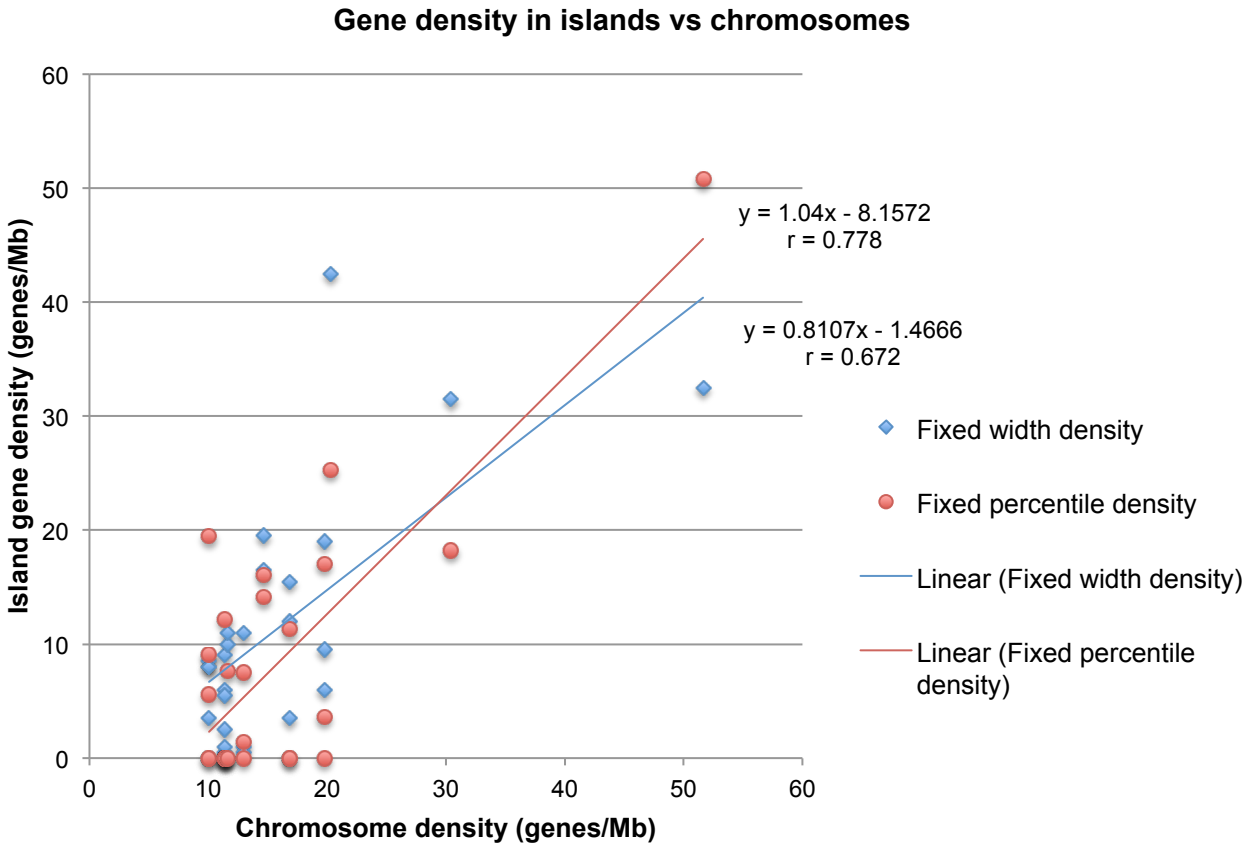


Figure 15: Gene density in fixed percentile islands is closer to chromosome gene density than is fixed width gene density

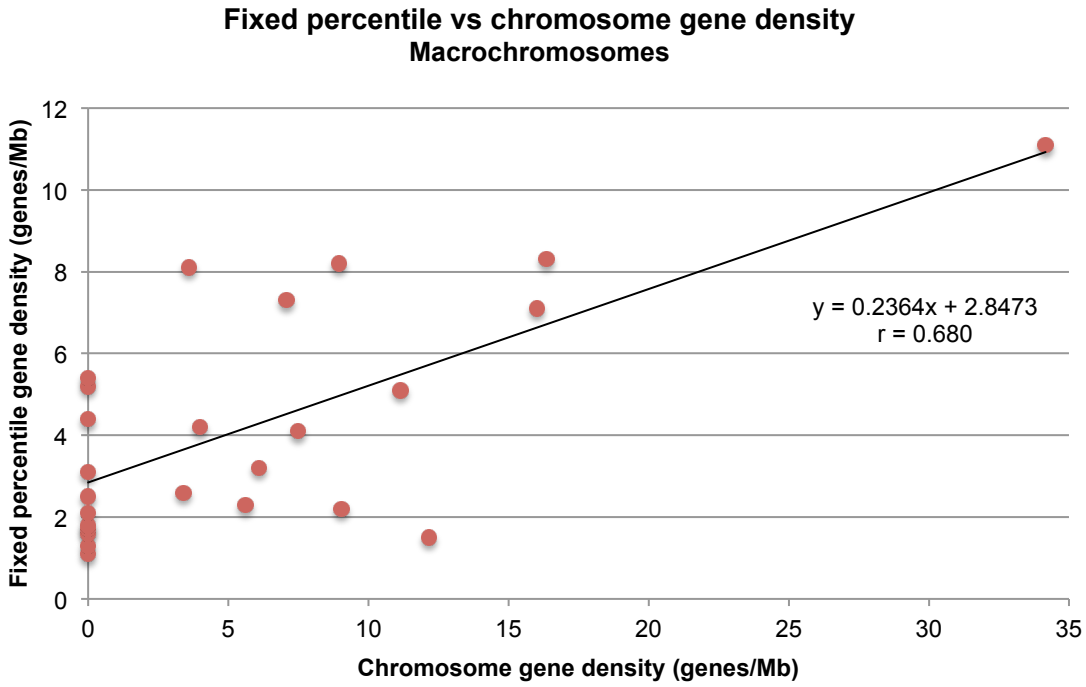


Figure 16:

There is a no trend related to the length of an island and its gene density (r = 0.161) although longer islands do seem to have disproportionately more genes than do shorter ones (Figure 17).

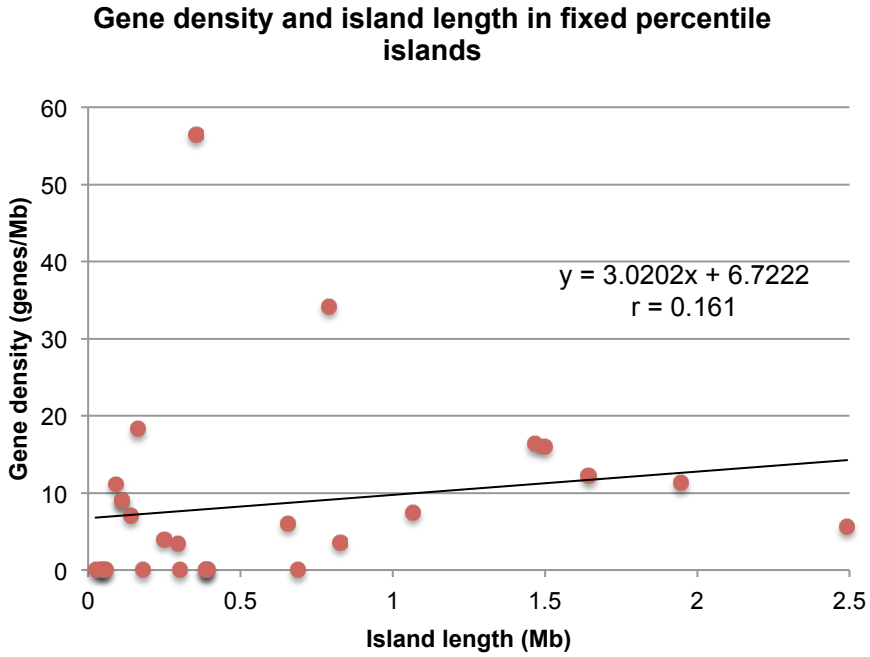


Figure 17: Longer islands are not more gene dense than shorter ones.

QTLs were found in all island regions (see appendix), however, any potential significance was not investigated. In future, the relationships between islands, their genes, and the known QTLs should be investigated to better understand the potential functions and uses of islands.

4.3 Haplotype analysis

ROH are significant in that they are long stretches of DNA that are in a homozygous state. ROH islands are significant in that they are regions that are in ROHs across a large proportion of individuals in a population (at least 55.6% in this case). This homozygosity has been used as a measure of inbreeding in populations (McQuillan et al., 2008; Ferenčaković et al., 2013, Silio et al., 2013). What has not been known up until this point is how much diversity exists within ROH and more specifically, in ROH islands.

This study has analysed the haplotypes found in each fixed percentile island to determine the number of haplotypes, the frequency of each haplotype within and across groups, and finally, how each haplotype is related to the others within an island. As can be seen in Table 3, there is a wide range of haplotype diversity across the islands with diversity ranging from as many as 4.375 to as few as 0.276 haplotypes per SNP (average 1.181 ± 0.921). The islands not containing coding elements have been highlighted in blue (Table 3). The non-coding regions range in diversity from 0.587 up to 4.375 haplotypes per SNP (Table 4). The average number of haplotypes per SNP (1.479 ± 1.017) is higher than the average in coding islands (0.942 ± 0.792), however the difference is not significant. There does seem to be a trend to fewer haplotypes in the coding islands, however. This might become clearer with a larger sample size. If substantiated, this difference could indicate that the genetic material in the coding islands has been better conserved over the generations.

Table 4: Non-coding islands in blue.

Island	# Haplos	# SNPs	H/SNP
1.1	25	12	2.083
1.3	15	8	1.875
1.5	84	163	0.515
1.6	70	99	0.707
1.7	73	68	1.073
1.8	87	70	1.243
2.1	86	94	0.915
2.2	43	29	1.483
2.3	**133**	**481**	**0.276**
2.5	27	16	1.687
2.6	60	67	0.895
3.1	49	47	1.042
3.2	124	191	0.649
4.1	84	161	0.522
4.2	74	52	1.423
4.4	93	139	0.669
5.1	84	163	0.515
5.2	32	21	1.524
5.4	**35**	**8**	**4.375**
7.1	133	292	0.455
7.3	67	51	1.314
8.1	108	176	0.614
8.2	27	46	0.587
8.3	106	272	0.390
11.1	159	207	0.768
17.1	81	91	0.890
27.1	44	13	3.385

Table 5

	Avg # Haplo	Avg # SNPs	Avg H/SNP
All islands	74.185 ± 37.325	112.481 ± 109.233	1.181 ± 0.921
With genes	91.333 ± 37.023	163.000 ± 123.320	0.942 ± 0.792
Without genes	52.750 ± 25.421	49.333 ± 31.890	1.479 ± 1.017

Integer Neighbour Joining Networks are an effective way of representing the relationships between haplotypes. An INJN begins with a Median Joining Network, but the number of SNP differences (mutations) is represented in the length of the branches (and in bracketed numbers) between the nodes (Bandelt et al., 1999; Woolley et al., 2008; http://popart.otago.ac.nz). The more ancestral haplotypes are located centrally with those thought to be derivative branching out from the ancestral haplotype(s). The distance (number of mutations) between haplotypes indicates how closely or distantly two haplotypes might be related (http://popart.otago.ac.nz). In order to effectively represent the haplotype diversity and relationships found across populations, only those haplotypes occurring in at least 2.5% of the animals with ROHs in each island are shown in the analysis and graphs below.

Due to the nature of SNP selection in this analysis, variations between the groups and ancestral relationships to the RJF populations may not be apparent, however, they may not exist at all. Mitochondrial DNA analysis of chickens has found that there are no breed-specific haplotypes, only regional haplotypes, which is similar to what we've seen in our analysis of island haplotypes (Tixier-Boichard et al., 2011). In future, all haplotypes may be included in order to search for clearer relationships or more steps connecting the haplotypes between populations.

As the RJF is known to be the major ancestor of the modern domesticated chicken one might presume that the RJF island haplotypes would be found to be ancestral to those found in more modern breeds. Additionally, breeds coming from similar ancestral, geographical, or trait backgrounds (ex. white and brown layers have had the same selection pressures, or European and Asian bantams have similar physical traits) might be expected to share clusters of closely related haplotypes at least in some islands. The results shown here (Island graphs below), however, indicate that there is no significant clustering of breed type haplotypes and that the RJF haplotypes, when present, are often not ancestral. There are several possible explanations for this finding.

Islands were searched for across all populations; therefore we found regions that frequently shared ROH in all populations. This would highlight similarities between breeds, not differences, at least in terms of homozygosity. It then, should not be surprising that the haplotypes we find in these shared ROH islands show sameness in their relationships rather than distinct differences between populations. Extensive crossbreeding early in the development of European breeds may also explain why the ancient haplotypes found here are shared widely between breeds of diverse origin (Hillel et al., 2003). Nothnagel et al. (2010) also looked for and found shared islands across European human populations. This is quite the opposite of what has been done previously looking at breeds of similar origin to find similarities within and differences between groups (Pemberton et al., 2012; Karimi, 2013; Curik et al., 2014; Sölkner et al., 2014). Looking across all populations may have uncovered the most ancestrally preserved portions of the genome; those which are descended from an ancient population of ancestors and which have been preserved in all breeds throughout domestication and breed development. Hillel et al. (2003) also found RJF and modern breeds to have alleles not shared between the two groups. Based on the varying lengths of the islands we may be able to infer that longer islands have come from a more recent common ancestor while shorter islands are more ancient. It is also possible that the certain islands have been more preserved than others which could make them appear to be newer than they really are. In future, effects of linkage disequilibrium could be investigated to determine why certain regions are being maintained in a contiguous homozygous state.

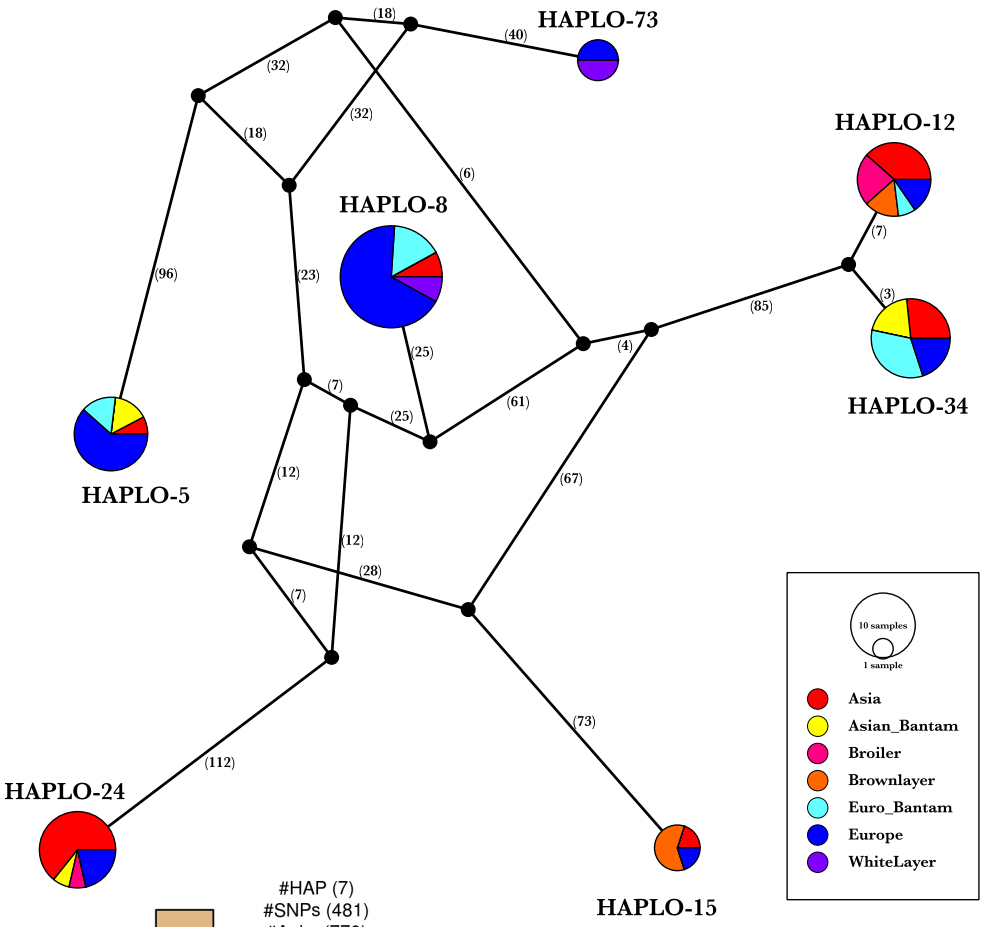
In our findings, ROH islands are quite absent from the RJF populations. This is likely explained by the way in which SNPs were chosen for the initial analysis. Since only SNPs found to be polymorphic in RJF were used to investigate the entire population, if there were any SNPs that

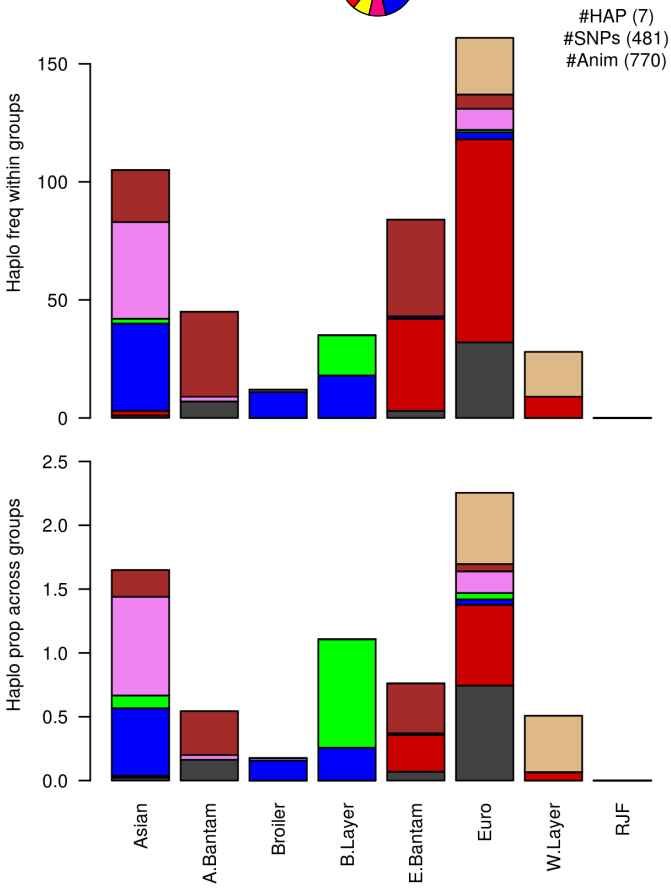
were homozygous in all RJF individuals, they may have been in ROH and those ROH may have been in islands. Since the analysed SNPs are equally distributed, though, the homozygous SNPs found to be in ROH should also be representative of those SNPs between them that may have been excluded during SNP selection. This could be confirmed by using a wider range of SNPs (not eliminating those monomorphic in RJF) or by sequencing the entire genome to ensure that the base pairs between homozygous SNPs in a ROH are, in fact, also homozygous. As is, the data show a small number of RJF individuals contain a small number of island haplotypes, and when present, these haplotypes are not found centrally in the INJNs and so do not seem to be ancestral to the rest of the population. Some RJF haplotypes are very similar while others are very diverse. The small number of RJF sampled (40) may also limit the number of haplotypes in the sampled population. This surprising finding may have another explanation. While the RJF sampled come from captured populations, they are now being raised in semi-domesticated conditions and are exposed to domestic populations allowing for crossbreeding (Berthouly et a., 2009). Also, importantly, these RJF birds are separated from their pre-domestication ancestors by 3000 generations or more. Selection pressures of the last millennia such as environmental change, inbreeding and genetic drift, are likely to have shifted the RJF in a different direction than its prehistoric ancestors. We cannot go back in time to sample these ancient chicken ancestors, however, it may be possible to sample the oldest available RJF specimens (for example in museums) to detect if there has been a genetic shift over the last centuries. Two species of *Gallus gallus* were sampled in this study but there are five extant *Gallus gallus* species which contribute to the modern chicken, *G. g. spadiceus* and *G. g. gallus* (mainly) but also *G. g. jabouillei*, *G. g.murghi* and *G. g. bankiva*. Finally, while the RJF is the primary ancestral species of the modern chicken, some believe that *Gallus sonneratii,* the Grey Jungle Fowl (GJF), has also contributed and though samples have been collected, they have not yet been analysed (Crawford, 1984; Hillel et al., 2003; Tixier-Boichard et al., 2011). These additional GJF samples could be added to the analysis to search for missing haplotypes. The figures below demonstrate the trends seen in the islands.

The following graphs show the relationships between the haplotypes of some of the islands (additional graphs can be found in the appendix). The first INJNs show how each haplotype is related to the others and which breed groups occur in which haplotypes. The first bar graph shows the number of animals that occur in each haplotype in this island with haplotypes shown in different coloured bars. The second bar graph shows how each haplotype is dispersed across breed groups. The proportion of each haplotype is shown and adds up to 1.0 (100%) for each haplotype.

Island 2.3

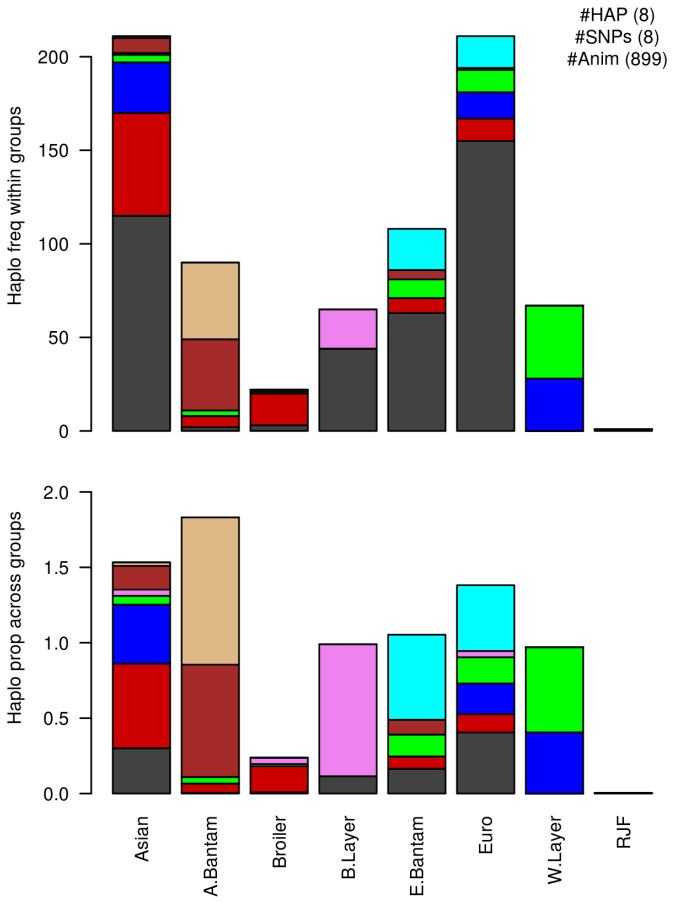
The length, and the small number of haplotypes per SNP are what make this island most interesting. Island 2.3 is a coding region and has lowest H:SNP ratio (0.0145) of all islands indicating that these haplotypes are highly conserved and that this island has less diversity than the others. This is significant considering that island 2.3 is by far the largest island found in the chicken genome with 481 consecutive homozygous SNPs (~2.5 Mb). Long ROH are interpreted as being more modern than short ROH, which are thought to be more ancient in ancestry (Curik et al., 2014; Sölkner et al., 2014). This effect is weaker if there have been strong selection pressures acting on the population. This is in strong contrast to island 5.4 (below) which has a very high H:SNP ratio and does not code for any genes. It could be possible that island 2.3 is being held together by some functional constraints while island 5.4 has become a ROH by chance. We see some separation of haplotypes by breed with haplotypes 5, 8, and 73 mainly in European type animals, but the trend is not clear, and there are no RJF in this island. Bar graph one shows that the majority of animals in this island are from European type breeds.

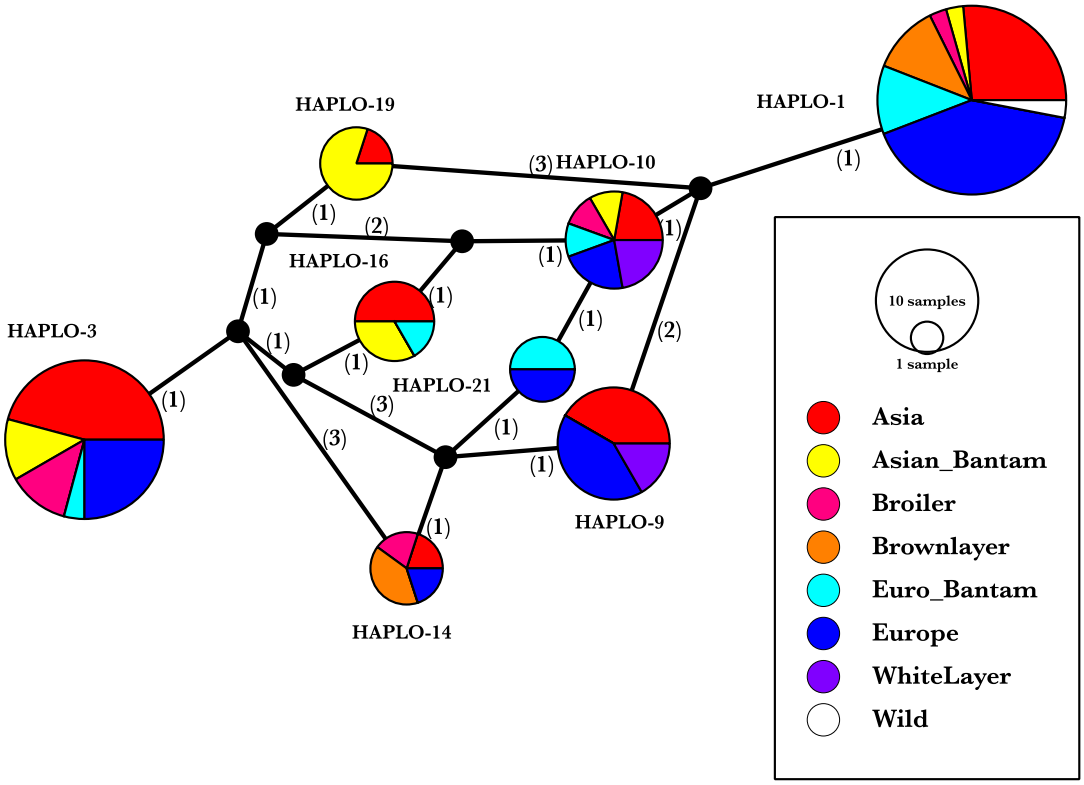




Island 5.4

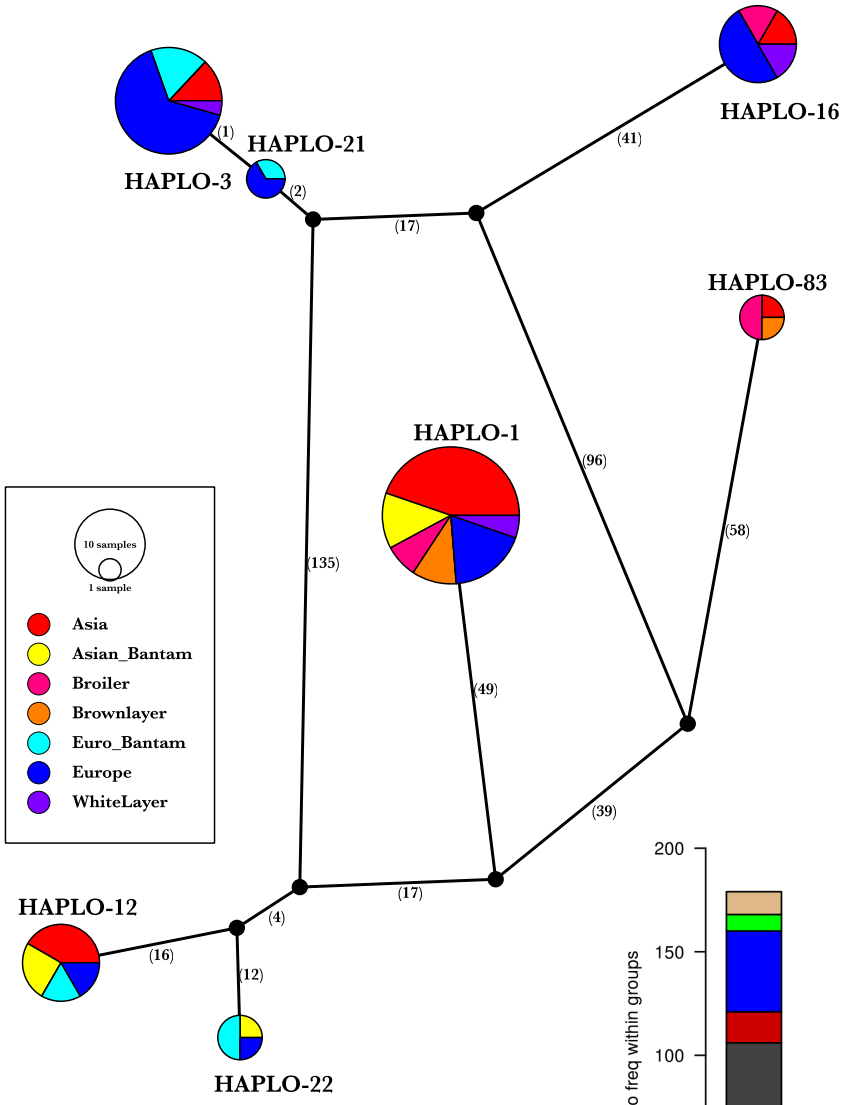
Island 5.4 contains no coding segments and has the highest haplotype to SNP (H:SNP) ratio of all islands (1.000) indicative of a less conserved region of the genome. It also means that there is more diversity in this island. This is more representative of the non-coding islands in that they tend to have more haplotypes per SNP (diversity) than the coding islands. All the haplotypes of island 5.4 vary from one another by greater than 10% each with two main haplotypes being separated by seven mutations (out of 8 SNPs). In this island we see how each haplotype occurs in the majority of populations. We do see some evidence for haplotype separation by breed groups with haplotypes 1, 3, 14, and 19 being mainly Asian type breeds but there is also evidence of admixture as is seen in haplotypes 1, 10, and, 3.

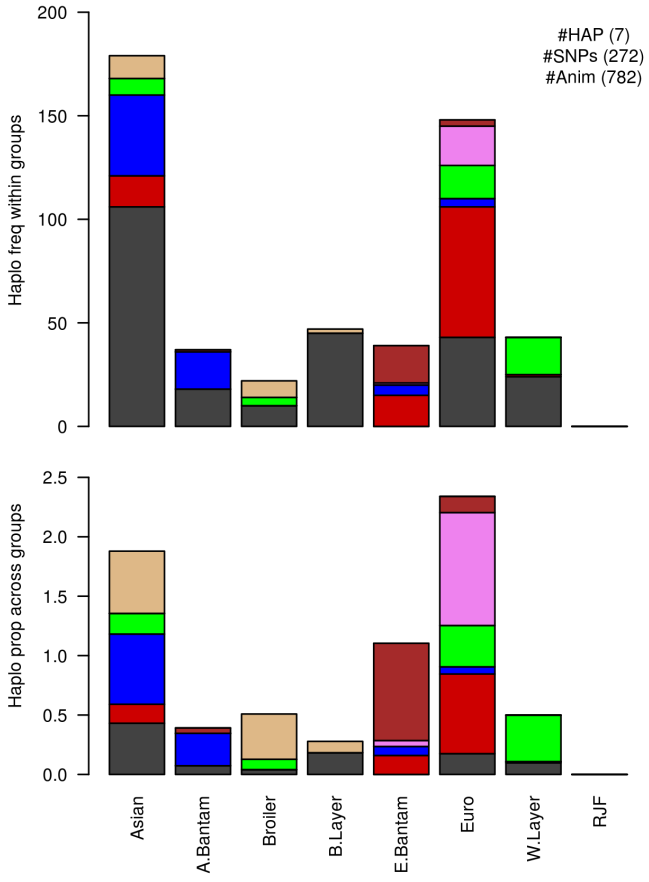




Island 8.3

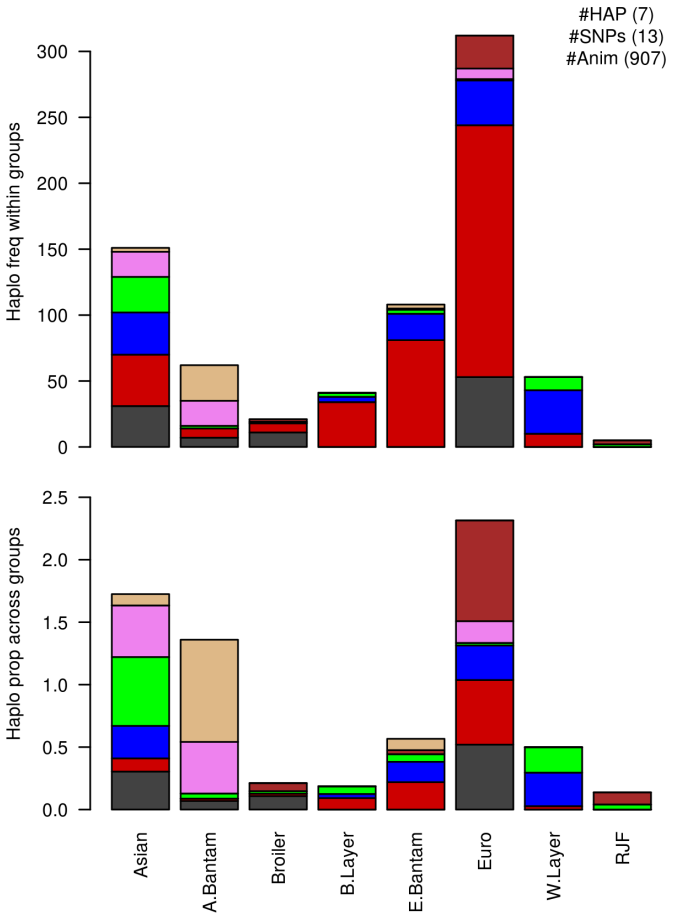
Island 8.3 is a coding region with a moderately high gene density (17 genes/Mb). There are no RJF with ROH in this island which is interesting because it seems from the INJN that all of the ancestral haplotypes are missing. It is possible that all the present haplotypes are unrelated and became homozygous due to functional constraints in the region, or that the ancestral haplotypes have been lost or were not present in the sampled populations of RJF.

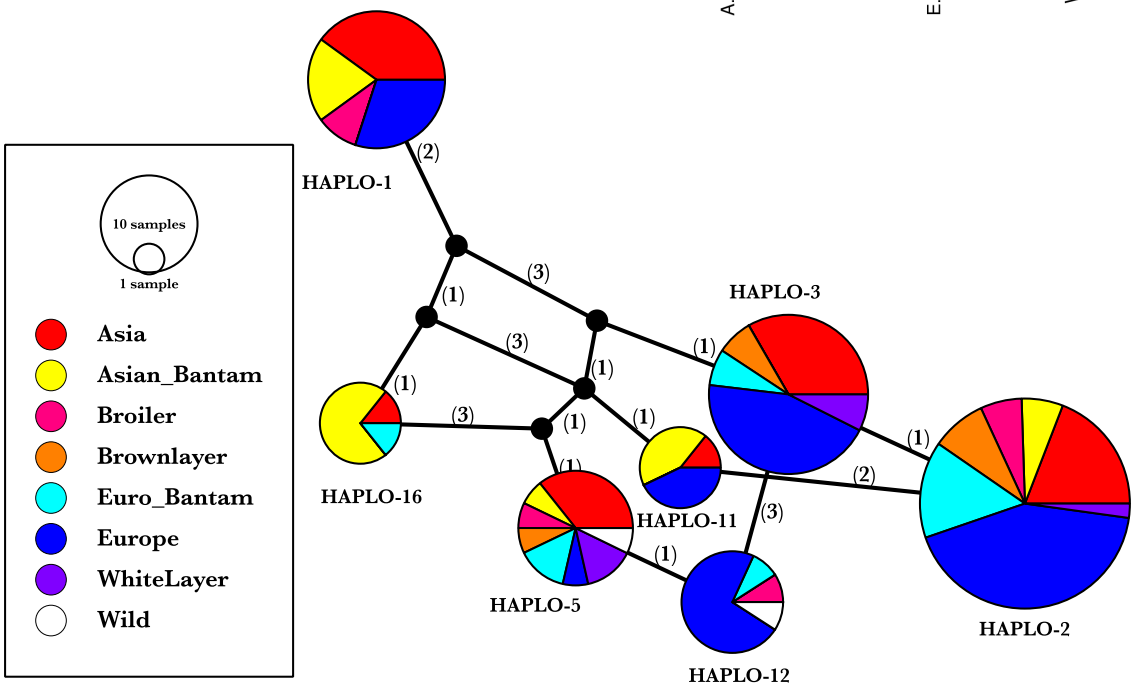




Island 27.1

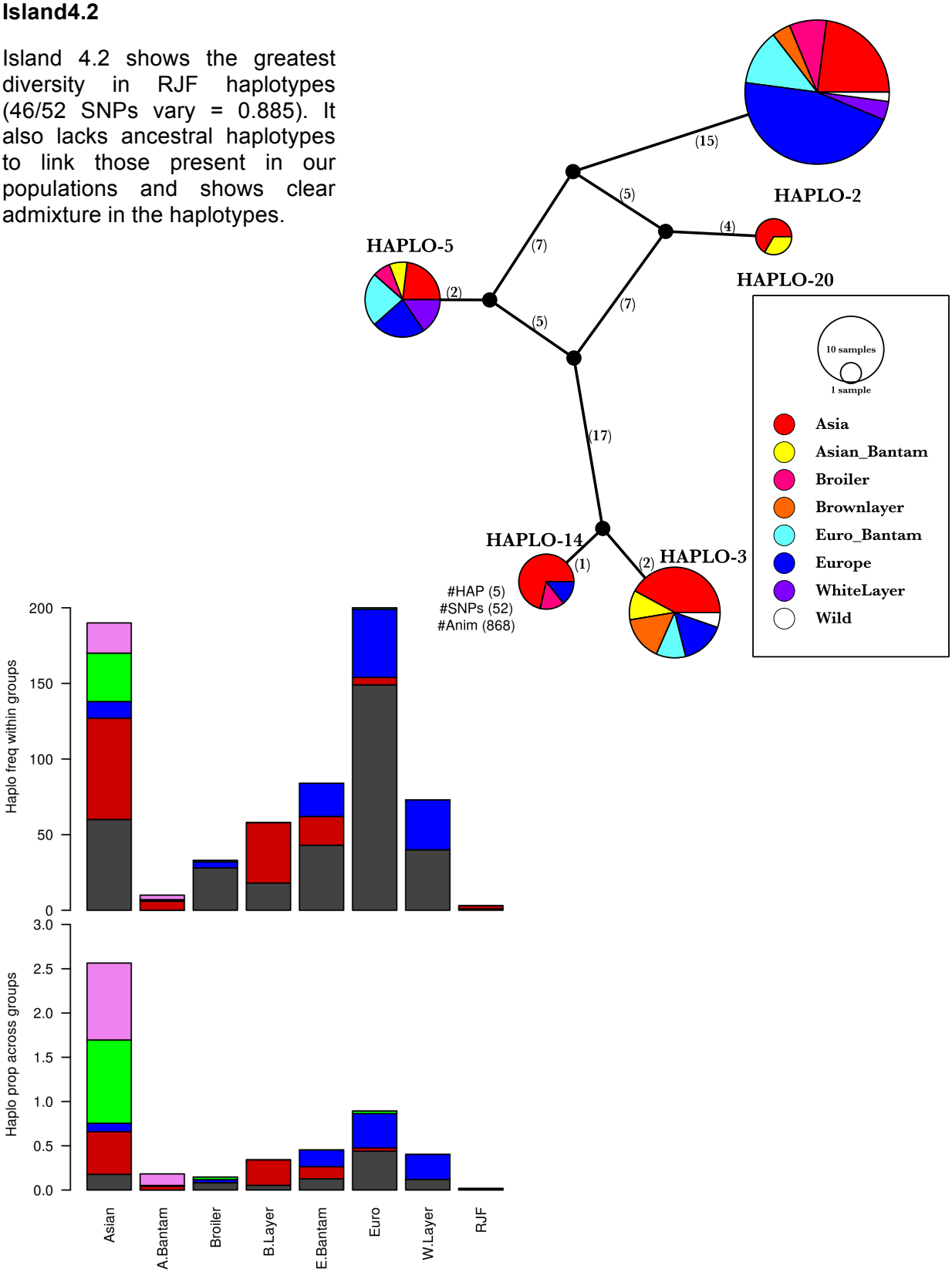
This is one of the microchromosome islands, and the island with both the highest gene density and the highest H:SNP ratio of islands containing genes. The RJF haplotypes seen in this island have the lowest amount of diversity (1/13 = 0.0769 SNPs vary). There is clear evidence of admixture from the composition of each haplotype, but from graph a we see that most of the animals in this islands are of European descent. We also see that haplotypes 2 and 3 are closely related (1 SNP different), and make up the majority of the population.





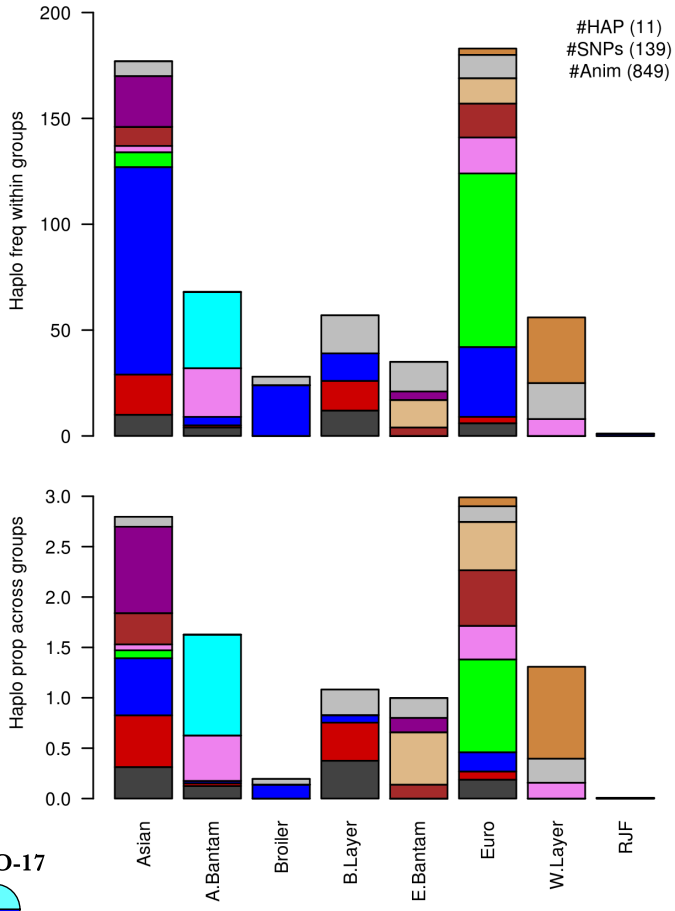
Island4.2

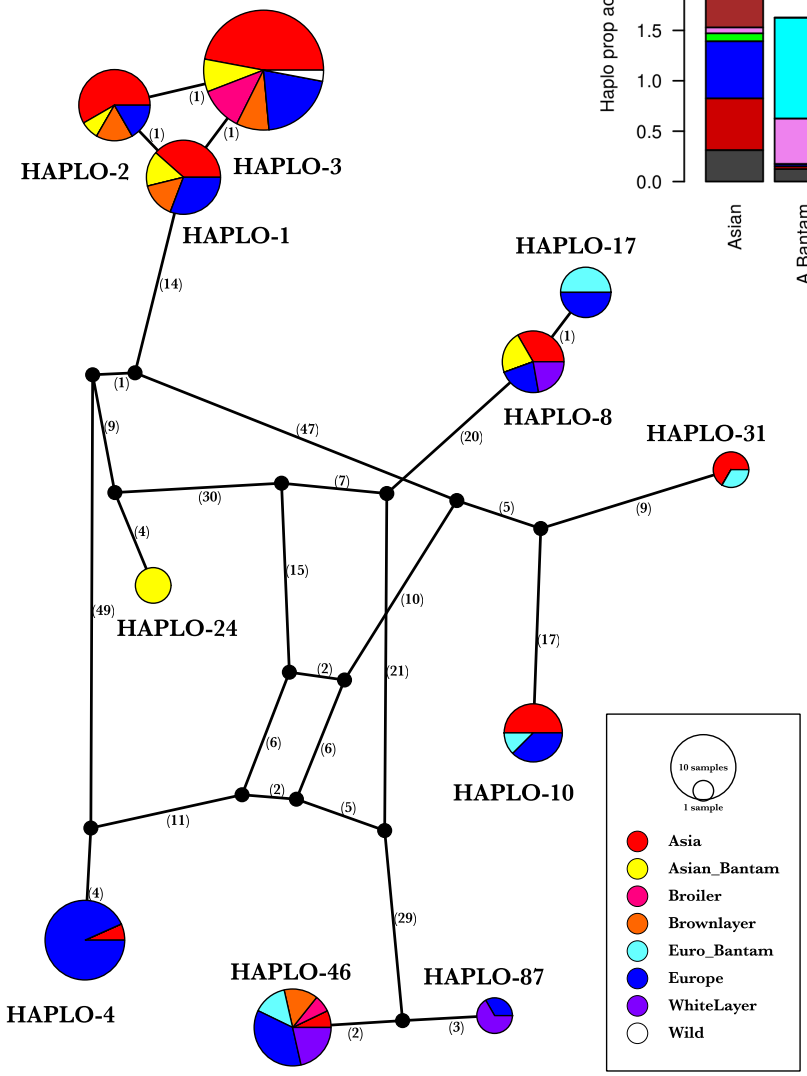
Island 4.2 shows the greatest diversity in RJF haplotypes (46/52 SNPs vary = 0.885). It also lacks ancestral haplotypes to link those present in our populations and shows clear admixture in the haplotypes.



Island 4.4

This island contains many small, highly diverse haplotypes. Interestingly, we see a cluster of closely related, mainly Asian type haplotypes in the top left corner and a cluster of mainly European haplotypes at the bottom. The Asian type haplotypes are the most common haplotypes in the population although there are a similar number of Asian and European type birds in this island. The RJF haplotype present is farthest from the centre of the graph indicating more lost ancestral haplotypes to look for in future.





5. Conclusions

The analysis has found that ROH islands and deserts occur frequently in the chicken genome. Islands appear clearly in both macro- and microchromosomes, and in all regions of chromosomes. While islands are found less frequently in microchromosomes, this is expected due to the high rate of recombination in microchromosomes. The distribution of deserts in the subtelomeric chromosome regions is poorly understood and requires further analysis. Islands are found across all breed groups and are likely ancestral in origin. A wide range of gene densities is found in ROH islands. While nearly half of all islands contain no genes, others are quite gene dense, although not significantly more so than the average gene density of the chromosomes. A wide range of genes with varying functions and pathways are found in islands. Future attention should be focused on how the genes in islands function within the chicken, and if they are somehow linked through the conservation of functionality. The haplotypes discovered in ROH islands are mainly shared across breed groups indicating admixture. There is no clear separation of island haplotypes by breed or breed group. By looking at subgroups of each breed type, and by including all the haplotypes found in the islands, clearer subdivisions may be seen. As with other characteristic of the chicken genome, the patterns of islands and deserts found in chickens is clearly different from those found in other species. This necessitates a unique approach to the search for, and interpretation of, these fascinating regions.

References

AVIANDIV http://aviandiv.tzv.fal.de/

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, *16*(1), 37-48.

BDRG (Bund Deutscher Rassegeflügelzüchter). Rassegeflügelstandard für Europa. 2010, Howa Druck & Satz GmbH (ISBN 3-9806597-1-4).

Berthouly, C., Leroy, G., Van, T. N., Thanh, H. H., Bed'Hom, B., Nguyen, B. T., ... & Rognon, X. (2009). Genetic analysis of local Vietnamese chickens provides evidence of gene flow from wild to domestic populations. *BMC genetics*, *10*(1), 1.

Bosse, M., Megens, H. J., Madsen, O., Paudel, Y., Frantz, L. A., Schook, L. B., ... & Groenen, M. A. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS genetics*,*8*(11), e1003100.

Broman, K. W., & Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *The American Journal of Human Genetics*, *65*(6), 1493-1500.

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual review of ecology and systematics*, 237-268.

Curtis, D., Vine, A. E., & Knight, J. (2008). Study of regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations. *Annals of human genetics*, *72*(2), 261-278.

Crawford R.D., Domestic fowl, in: Mason I.L. (Ed.), Mason, I. L. (1984). *Evolution of domesticated animals*. Longman, London, 1984, pp. 298–311.

Crawford, R.D., Origin and history of poultry species, in: Crawford. R.D. (Ed*.), Poultry Breeding and Genetics*, Elsevier, 1990, pp. 1–42.

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science*, *166*, 26-34.

Fernandez, A., Toro, M. A., & Lopez-Fanjul, C. (1995). The effect of inbreeding on the redistribution of genetic variance of fecundity and viability in Tribolium castaneum. *Heredity*, *75*(4), 376-381.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., ... & Shen, Y. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*(7164), 851-861.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I., & Sölkner, J. (2013). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics*, *130*(4), 286-293.

Granevitze, Z., Hillel, J., Feldman, M., Six, A., Eding, H., & Weigend, S. (2009). Genetic structure of a wide-spectrum chicken gene pool. *Animal genetics*,*40*(5), 686-693.

Groenen, M. A., Wahlberg, P., Foglio, M., Cheng, H. H., Megens, H. J., Crooijmans, R. P., ... & Andersson, L. (2009). A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research*, *19*(3), 510-519.

Hillel, J., Groenen, M. A., Tixier-Boichard, M., Korol, A. B., David, L., Kirzhner, V. M., ... & Weigend, S. (2003). Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. *Genetics Selection Evolution*, *35*(5), 533-558.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., ... & Aerts, J. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*,*432*(7018), 695-716.

Karimi, S., 2013. Runs of Homozygosity Patterns in Taurine and Indicine Cattle Breeds (Master thesis). BOKU-University of Natural Resources and Life Sciences, Vienna.

Kanginakudru, S., Metta, M., Jakati, R. D., & Nagaraju, J. (2008). Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. *BMC evolutionary biology*, *8*(1), 174.

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., ... & Burt, D. W. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC genomics*, *14*(1), 59.

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., ... & Malhotra, A. K. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences*, *104*(50), 19942-19947.

Moiseeva, I. G., Chzhan, I., Zakharov, I. A., & Nikiforov, A. A. (1996). [Comparative analysis of morphologic traits in Mediterranean and Chinese chicken breeds. The problem of the origin of the domestic chicken]. *Genetika*,*32*(11), 1553-1561.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., ... & Wilson, J. F. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, *83*(3), 359-372.

Mefford, H. C., & Trask, B. J. (2002). The complex structure and dynamic evolution of human subtelomeres. *Nature Reviews Genetics*, *3*(2), 91-102.

Microsoft. (2010). Microsoft Excel [computer software]. Redmond, Washington: Microsoft.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* 81. PLINK (v.1.07) URL http://pngu.mgh.harvard.edu/purcell/plink/

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/.

Silió, L., Rodríguez, M. C., Fernández, A., Barragán, C., Benítez, R., Óvilo, C., & Fernández, A. I. (2013). Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *Journal of Animal Breeding and Genetics*, *130*(5), 349-360.

SNP & Variation Suite (Version 8.x) [Software]. Bozeman, MT: Golden Helix, Inc. Available from http://www.goldenhelix.com.

Sölkner, J. (2014, August). Extremely Non-uniform: Patterns of Runs of Homozygosity in Bovine Populations. In *10th World Congress on Genetics Applied to Livestock Production*. Asas.

Stehelin, D., Varmus, H. E., Bishop, J. M., & Vogt, P. K. (1976). DNA related to the transforming gene (s) of avian sarcoma viruses is present in normal avian DNA.

Weigend, S., Janßen-Tapken, U., Erbe, M., Baulain, U., Weigend, A., Sölkner, J., & Simianer, H. In: Conference Information and Proceedings of XIV European Poultry Conference (ed. Birger Svihus), Stavanger, Norway, 23. - 27. June 2014, pp. 164 – 176.

Weigend, S. (2015): Personal communication.

PopART [Software]. *Available from http://popart.otago.ac.nz*

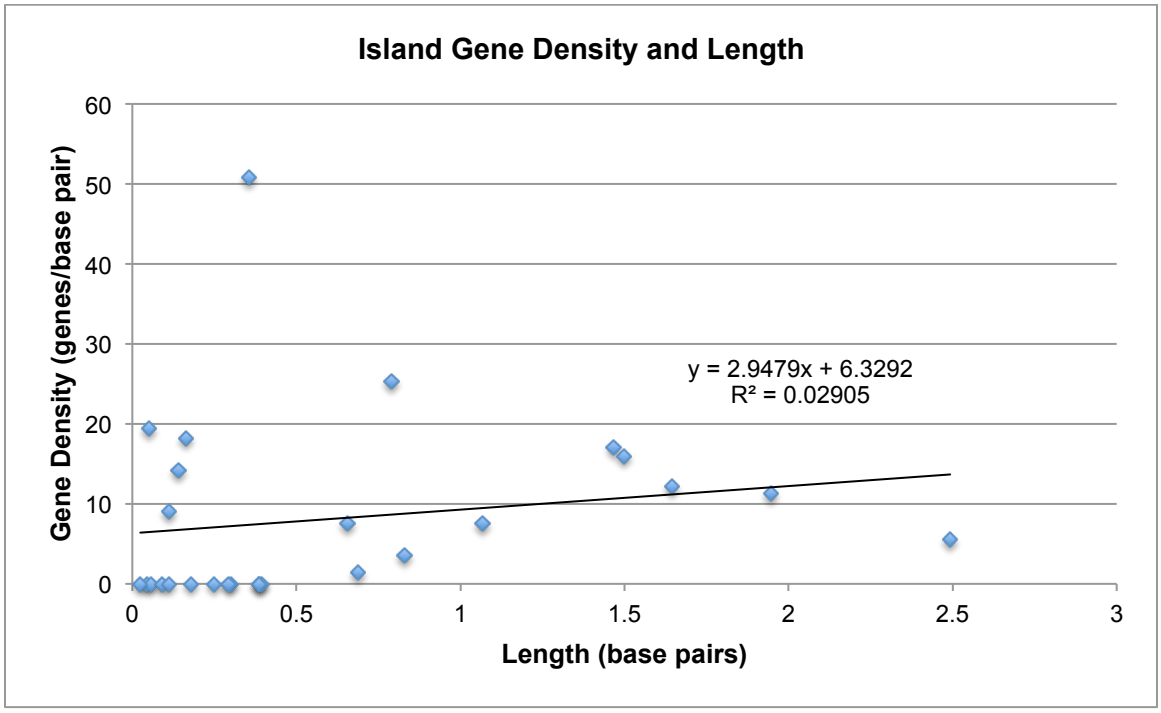
Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., ... & Flicek, P. (2015). Ensembl 2015. *Nucleic acids research*, *43*(D1), D662-D669. Version 79

Zeuner, F. E. (1963). A history of domesticated animals. *A history of domesticated animals.*

Abbreviations

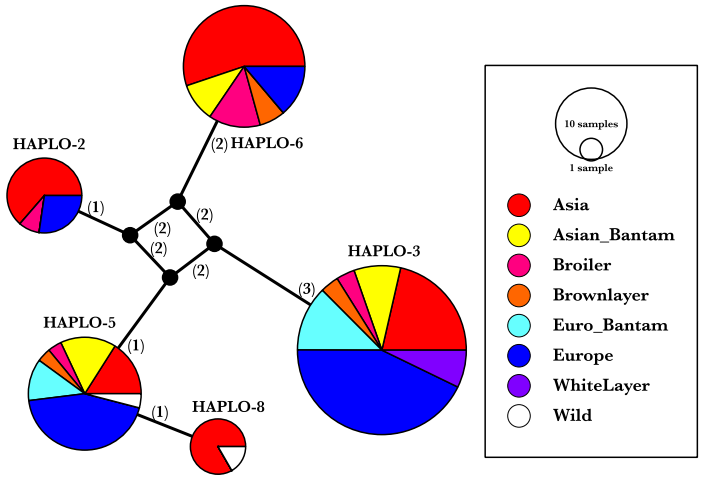
Abbreviation	Term
BDRG	Bund Deutscher Rassgeflügelzüchter
EDTA	ethylenediamine tetraacetic acid
GEH	Society for the Conservation of Old and Endangered Livestock Breeds
GJF	Grey Jungle Fowl
HD SNP	high density single nucleotide polymorphism
IBD	identical by descent
INJN	integer neighbour joining network
kb	kilobase
Mb	megabase
MHC	major histocompatibility complex
PCA	principal component analysis
RJF	Red Jungle Fowl
ROH	run of homozygosity
SCDP	Synbreed Chicken Diversity Panel
SNP	single nucleotide polymorphism
SVS	SNP and Variation Suite

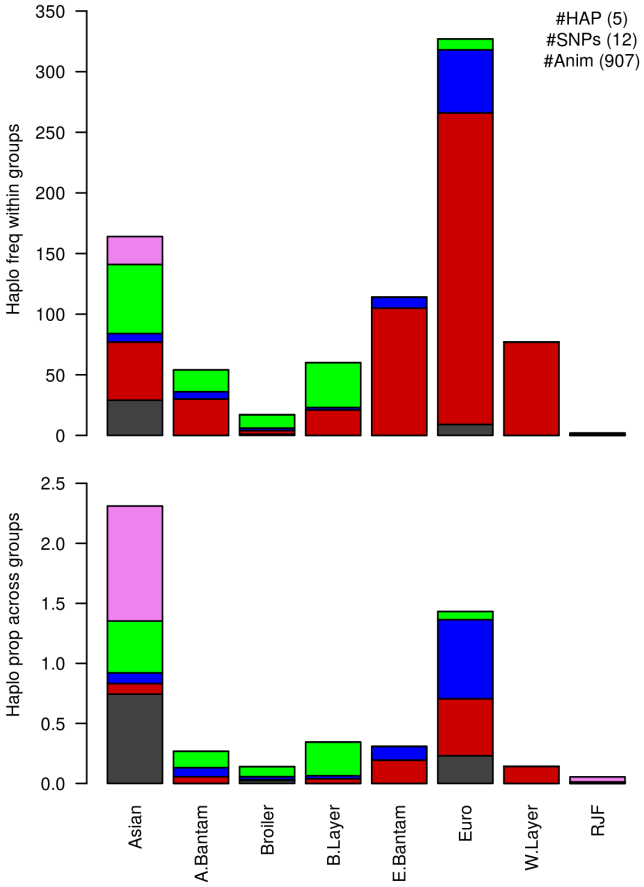
Appendix


Island Gene Density and Length
Gene Density (genes/base pair)
60
50
40
30
20
10
0
0
0.5
1
1.5
2
2.5
3
Length (base pairs)
y = 2.9479x + 6.3292
R² = 0.02905

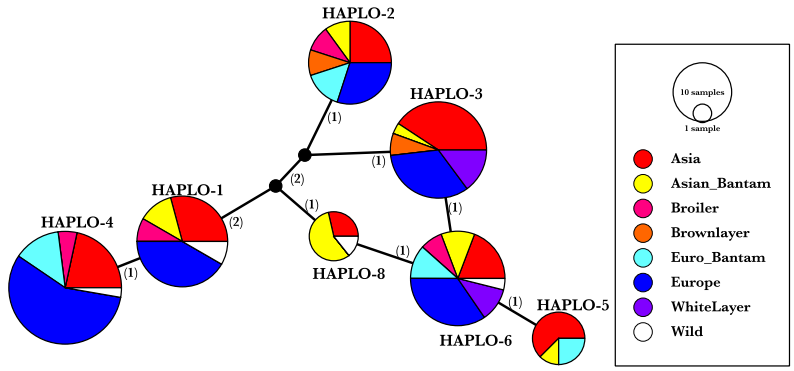
Haplotype graphs

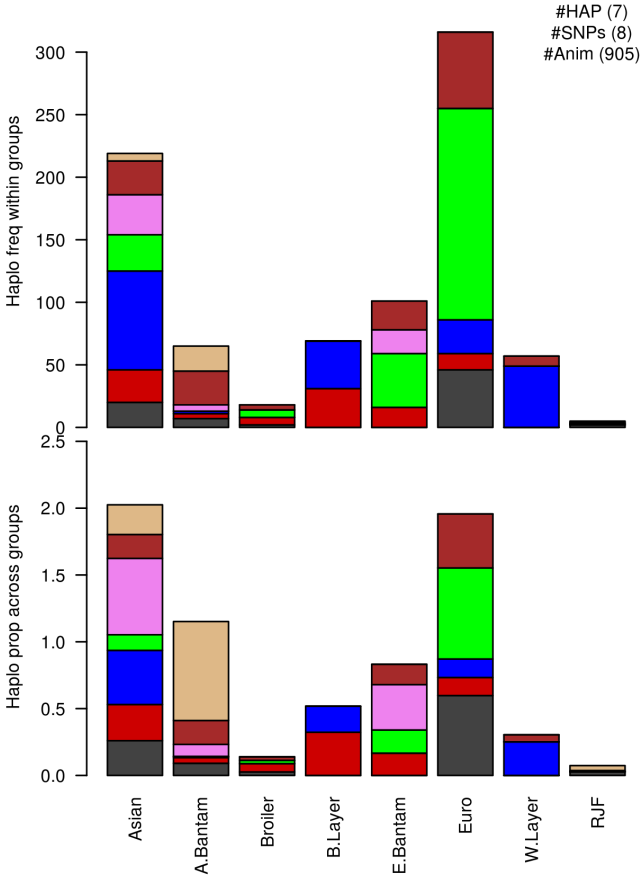
Island 1.1



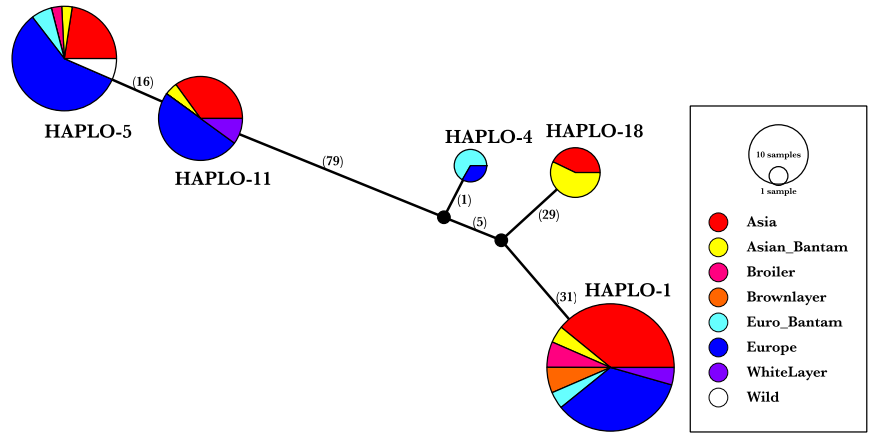


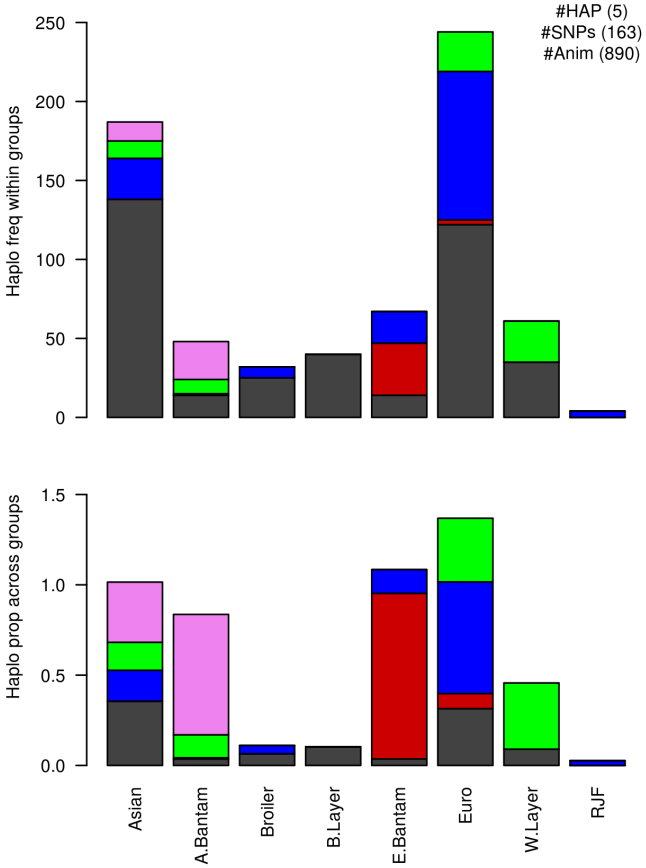
Island 1.3


HAPLO-2
HAPLO-3
HAPLO-1
HAPLO-4
HAPLO-8
HAPLO-5
HAPLO-6
(1)
(2)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild

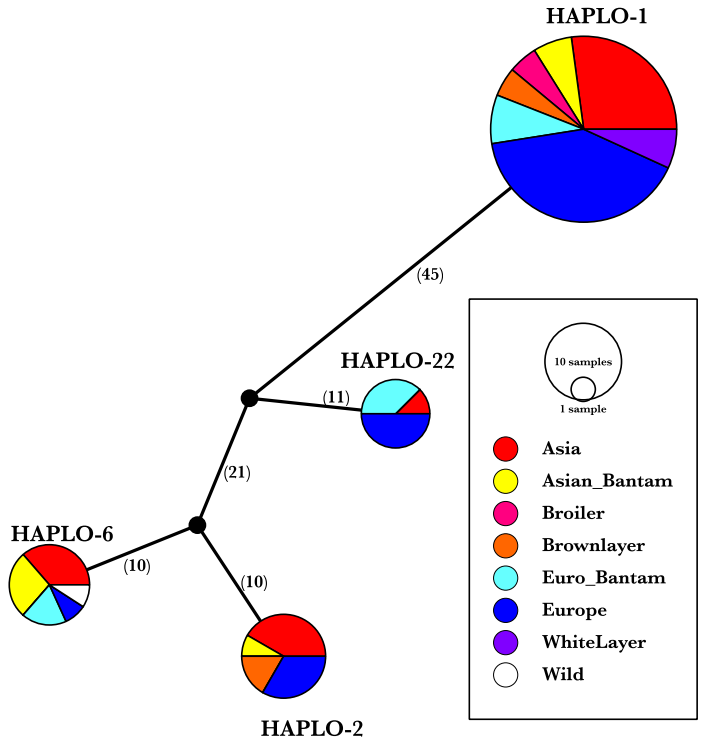

#HAP (7)
#SNPs (8)
#Anim (905)
Haplo freq within groups
Haplo prop across groups
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

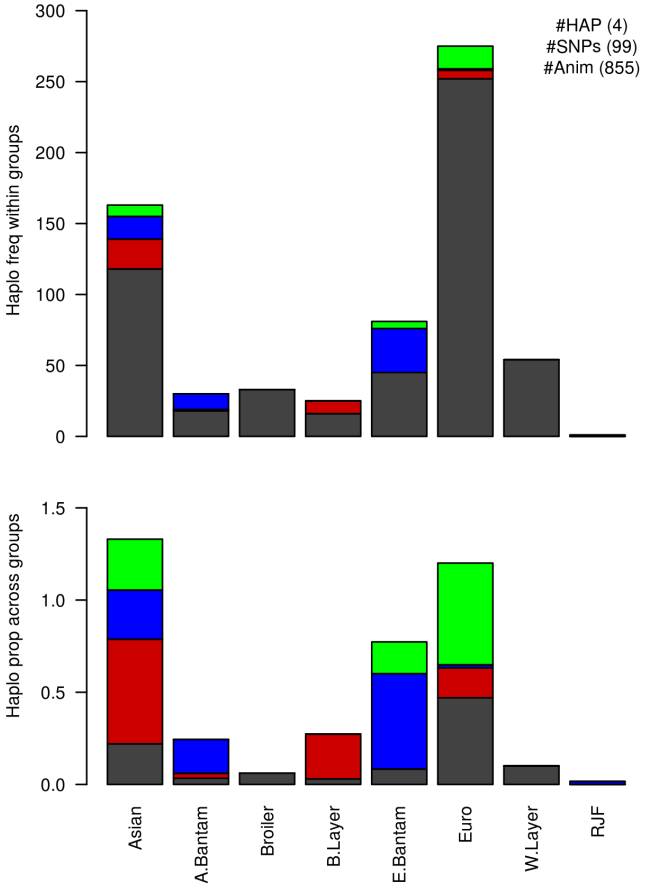
Island 1.5


HAPLO-5
(16)
HAPLO-11
(79)
HAPLO-4
HAPLO-18
(1)
(5)
(29)
(31)
HAPLO-1
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild

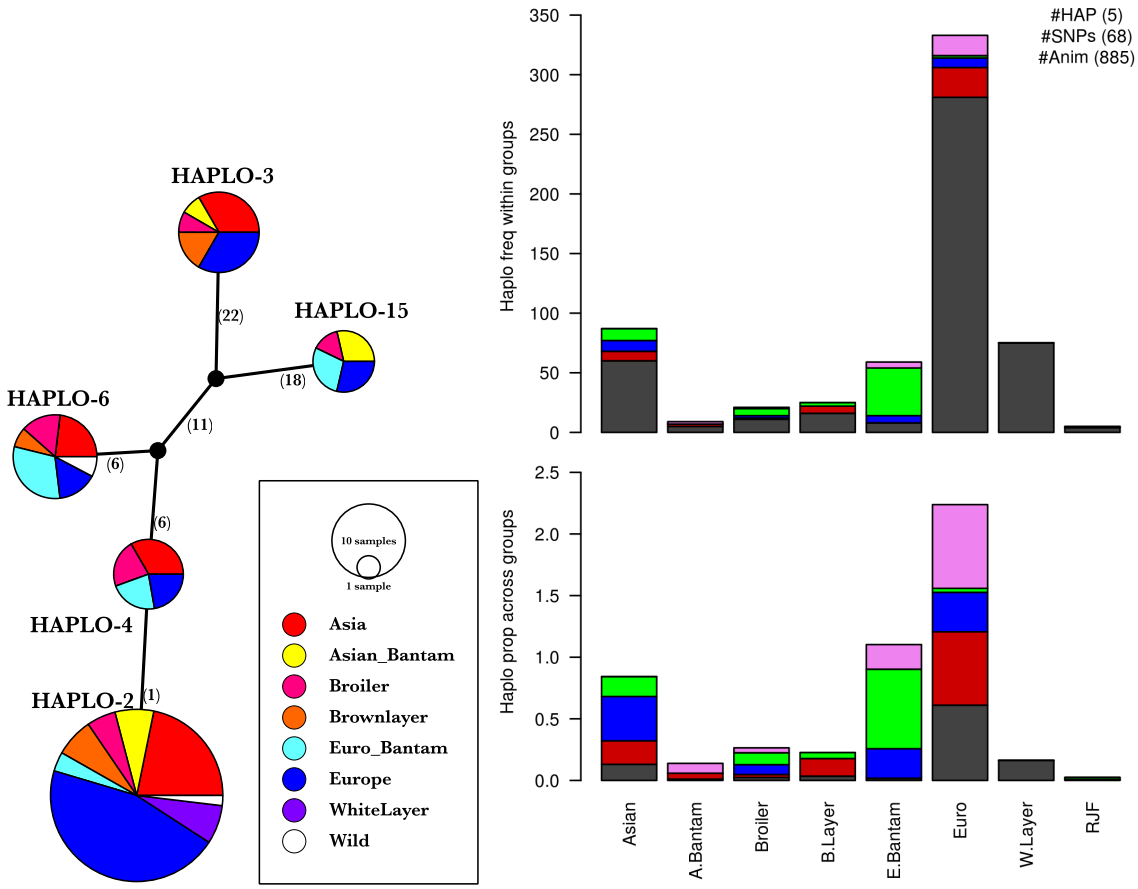

#HAP (5)
#SNPs (163)
#Anim (890)
Haplo freq within groups
250
200
150
100
50
0
Haplo prop across groups
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 1.6

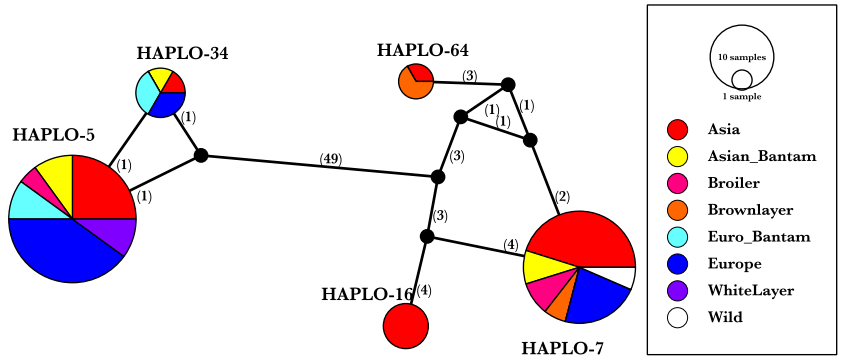

HAPLO-1
(45)
HAPLO-22
(11)
(21)
HAPLO-6
(10)
(10)
HAPLO-2
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild

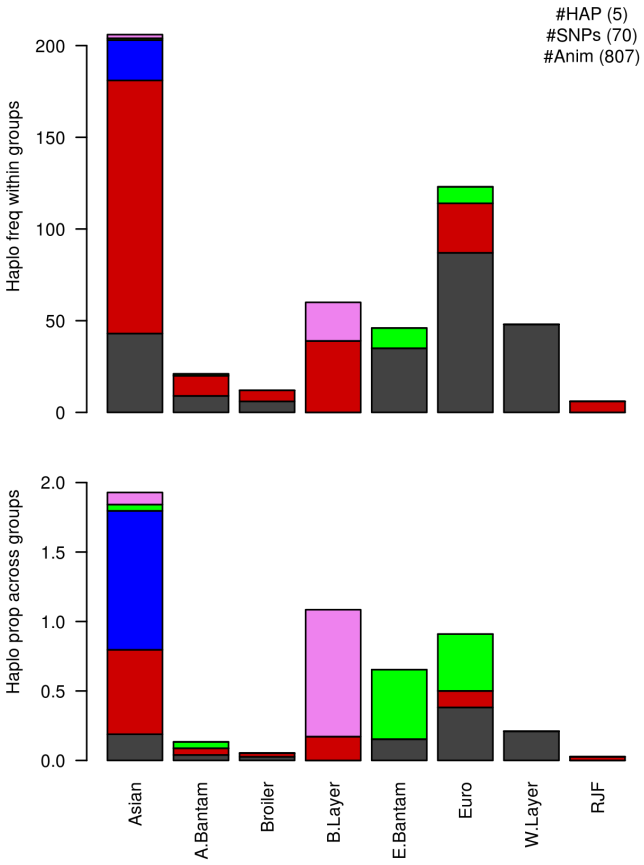

#HAP (4)
#SNPs (99)
#Anim (855)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 1.7

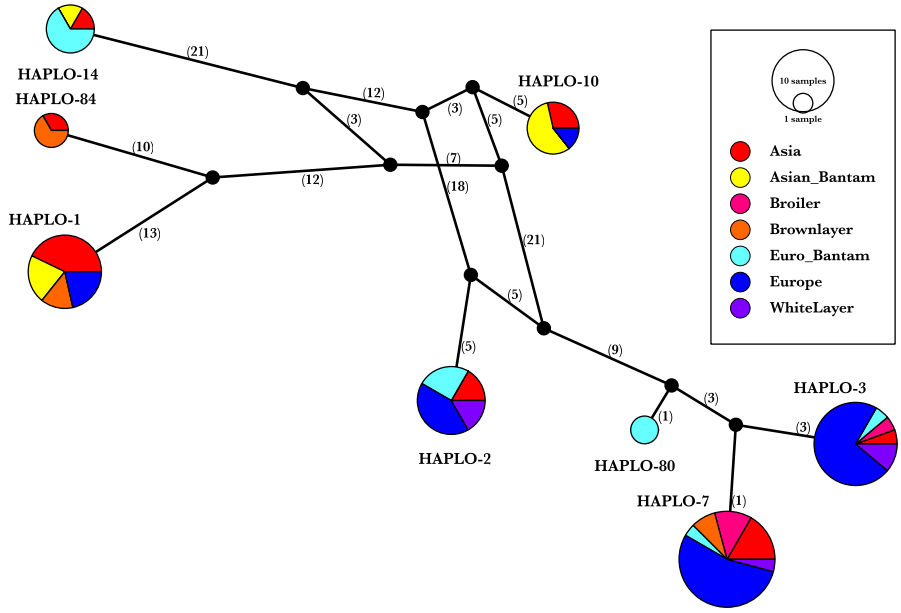

HAPLO-3
HAPLO-15
HAPLO-6
HAPLO-4
HAPLO-2
(22)
(18)
(11)
(6)
(6)
(1)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild
#HAP (5)
#SNPs (68)
#Anim (885)
Haplo freq within groups
Haplo prop across groups
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

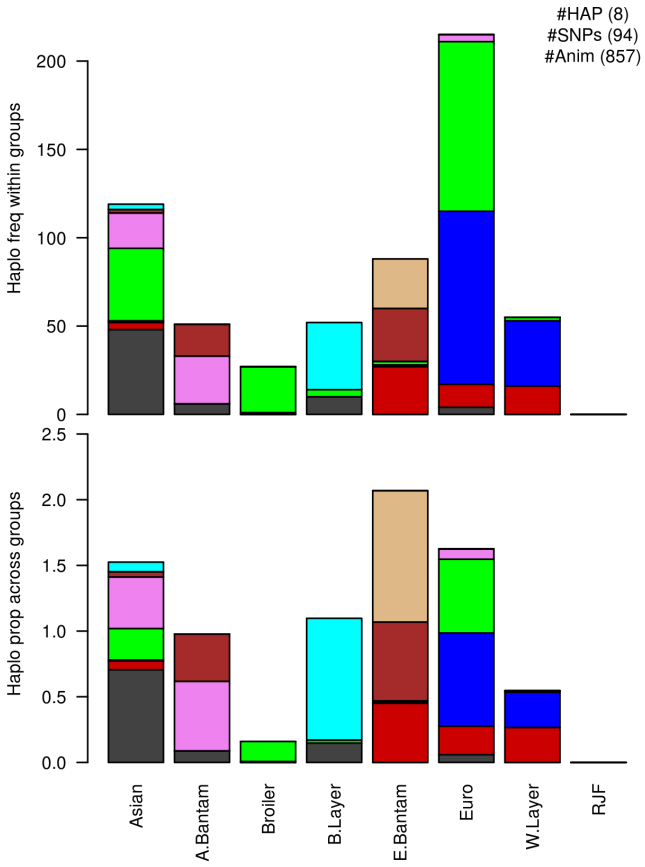
Island 1.8


HAPLO-34
HAPLO-64
HAPLO-5
HAPLO-16
HAPLO-7
(1)
(49)
(3)
(2)
(4)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild

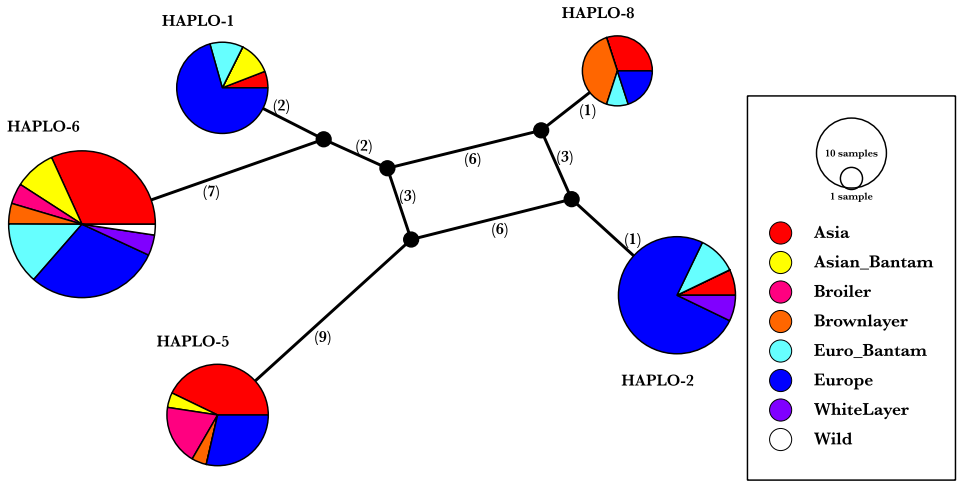

#HAP (5)
#SNPs (70)
#Anim (807)
Haplo freq within groups
200
150
100
50
0
Haplo prop across groups
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

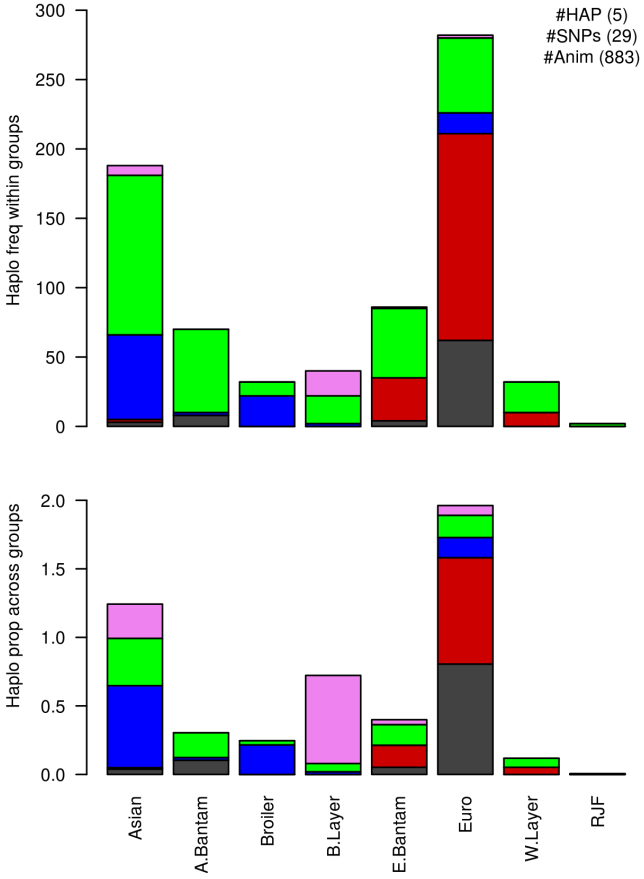
Island 2.1


HAPLO-14
HAPLO-84
HAPLO-10
HAPLO-1
HAPLO-2
HAPLO-80
HAPLO-7
HAPLO-3
(21)
(12)
(3)
(3)
(5)
(5)
(10)
(12)
(7)
(18)
(13)
(21)
(5)
(5)
(9)
(3)
(1)
(3)
(1)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer

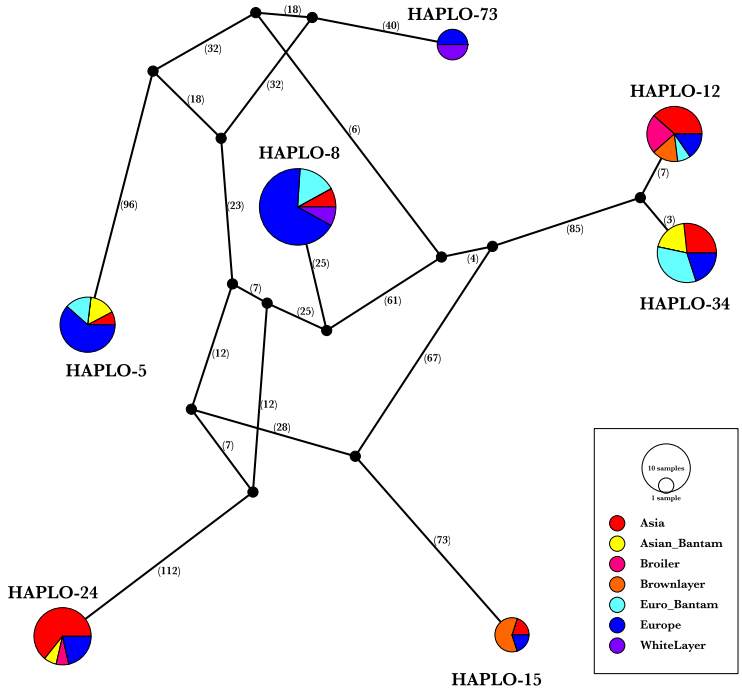

#HAP (8)
#SNPs (94)
#Anim (857)
Haplo freq within groups
200
150
100
50
0
Haplo prop across groups
2.5
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

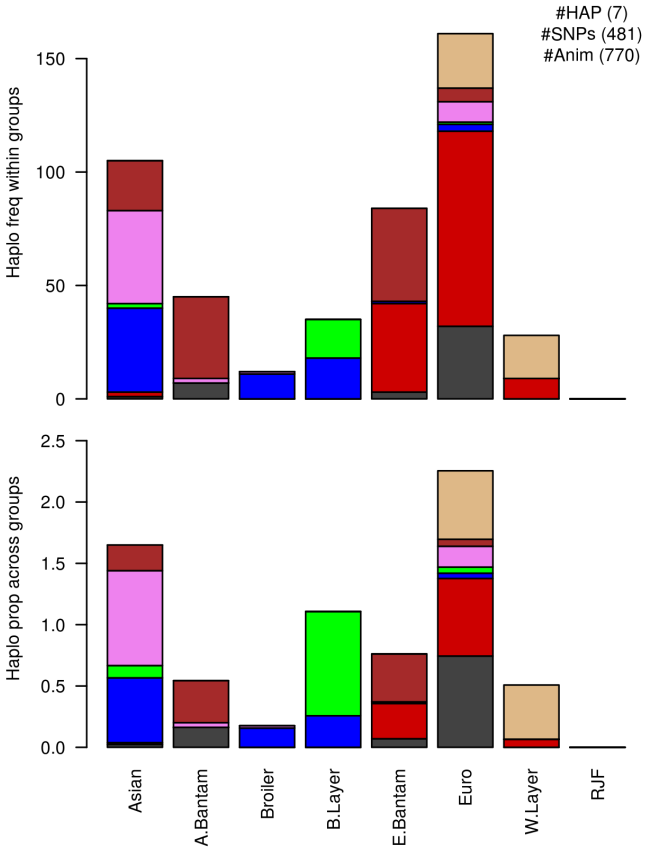
Island 2.2


HAPLO-1
HAPLO-8
HAPLO-6
HAPLO-5
HAPLO-2
(2)
(2)
(6)
(1)
(3)
(7)
(3)
(6)
(1)
(9)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild

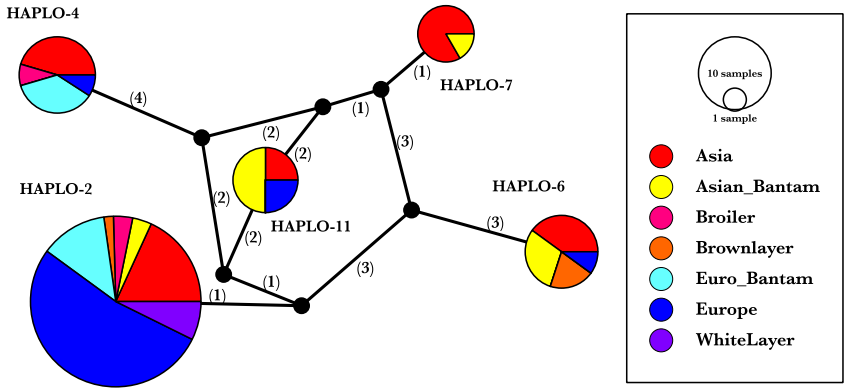

#HAP (5)
#SNPs (29)
#Anim (883)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

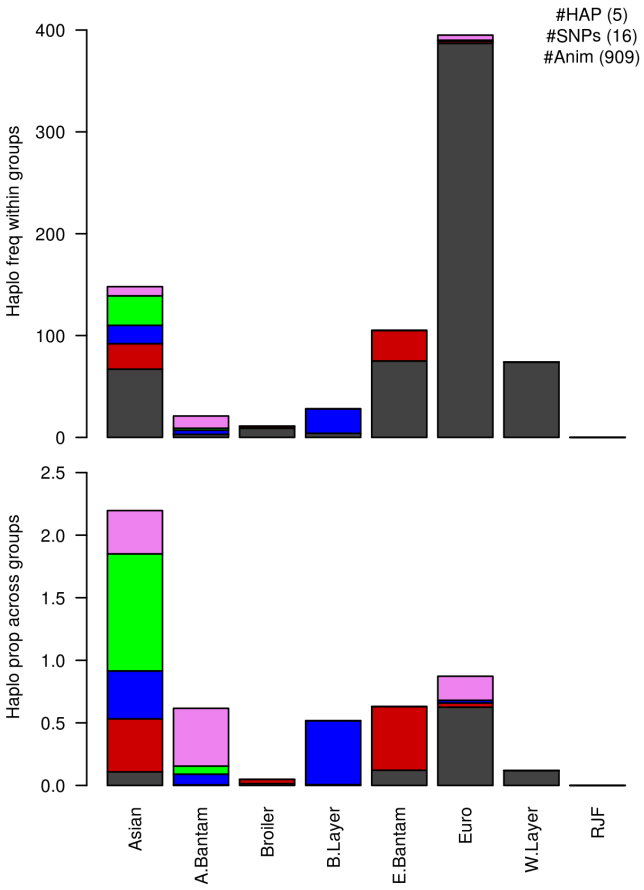
Island 2.3


HAPLO-73
HAPLO-12
HAPLO-8
HAPLO-34
HAPLO-5
HAPLO-24
HAPLO-15
(18)
(40)
(32)
(32)
(18)
(6)
(7)
(96)
(23)
(3)
(85)
(4)
(25)
(7)
(25)
(61)
(12)
(67)
(12)
(28)
(7)
(73)
(112)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer

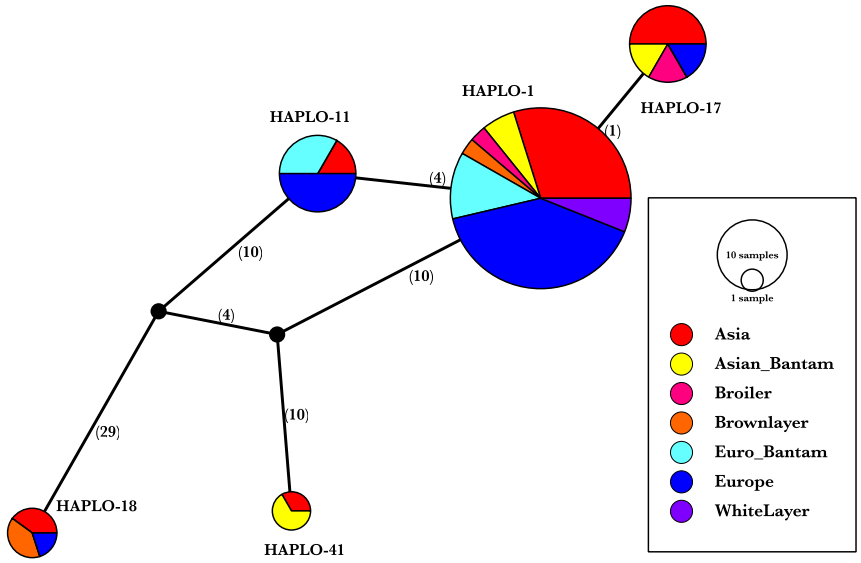

#HAP (7)
#SNPs (481)
#Anim (770)
Haplo freq within groups
150
100
50
0
Haplo prop across groups
2.5
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

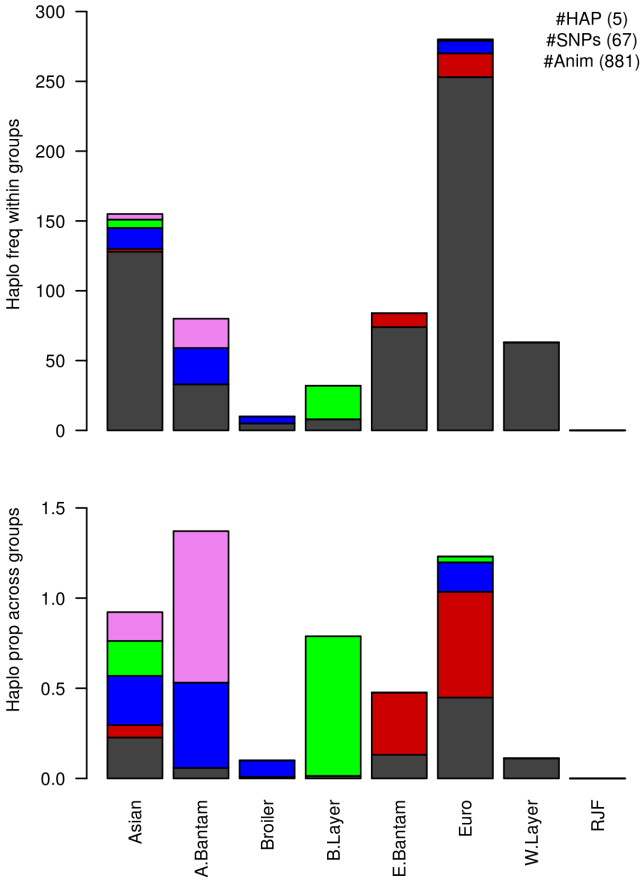
Island 2.5


HAPLO-4
HAPLO-7
HAPLO-2
HAPLO-11
HAPLO-6
(4)
(1)
(1)
(2)
(2)
(3)
(2)
(2)
(3)
(1)
(1)
(3)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer

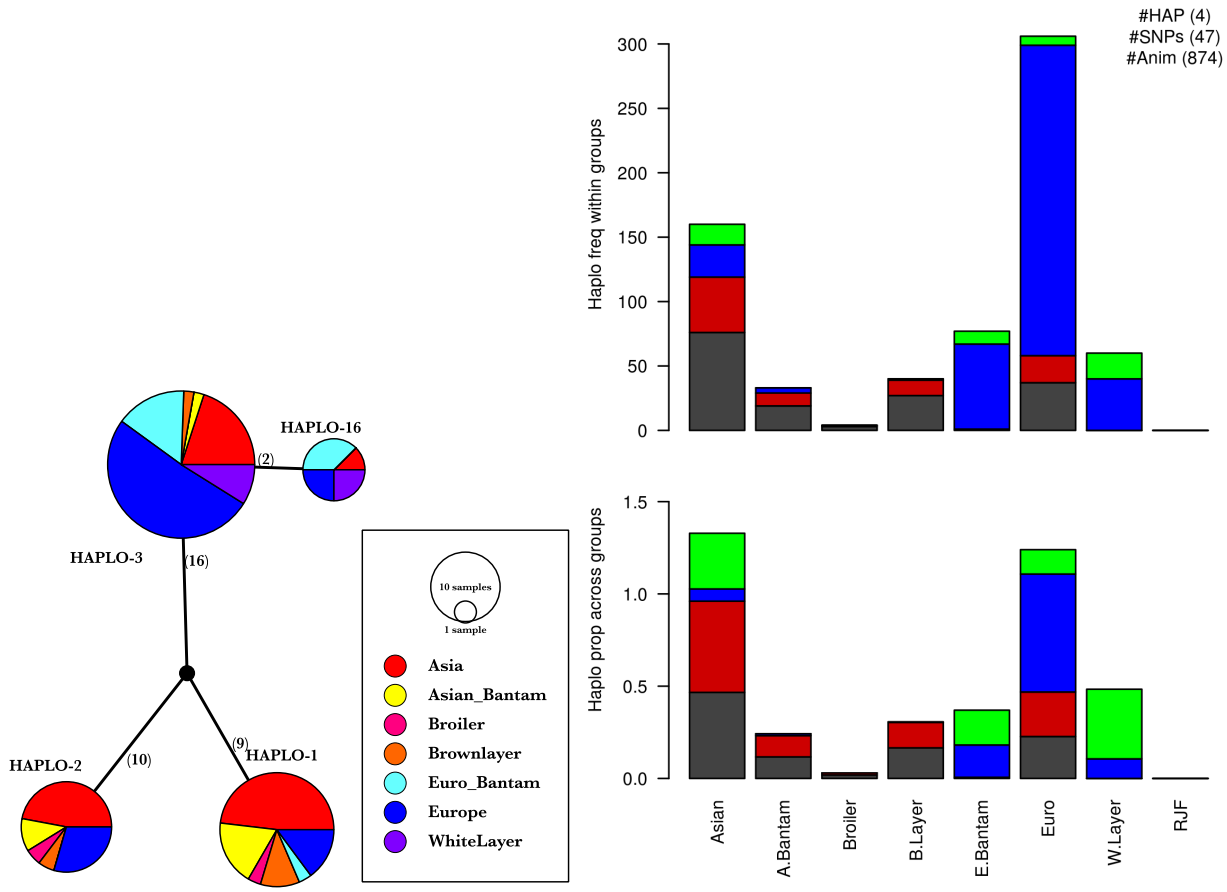

#HAP (5)
#SNPs (16)
#Anim (909)
Haplo freq within groups
400
300
200
100
0
Haplo prop across groups
2.5
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 2.6

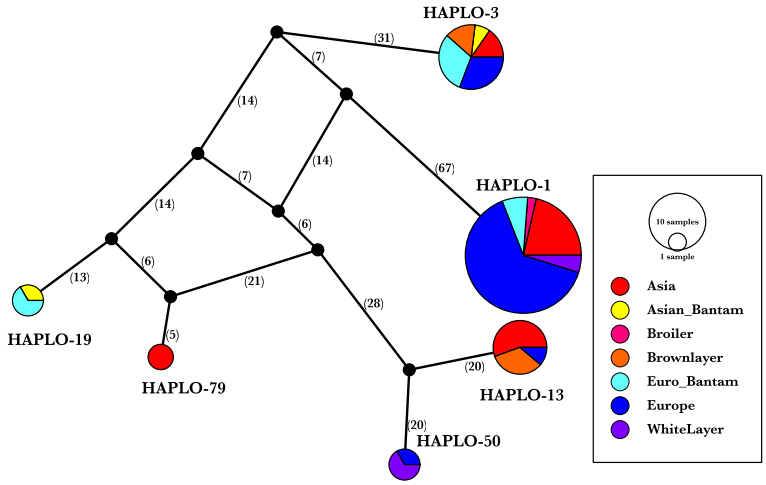

HAPLO-1
HAPLO-17
HAPLO-11
(1)
(4)
(10)
(10)
(4)
(10)
(29)
HAPLO-18
HAPLO-41
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer

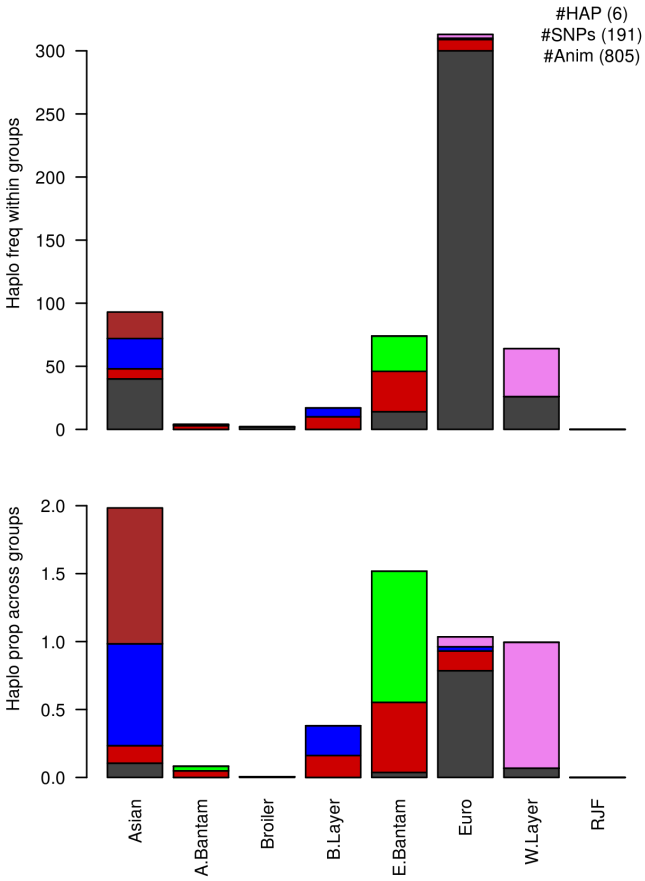

#HAP (5)
#SNPs (67)
#Anim (881)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 3.1

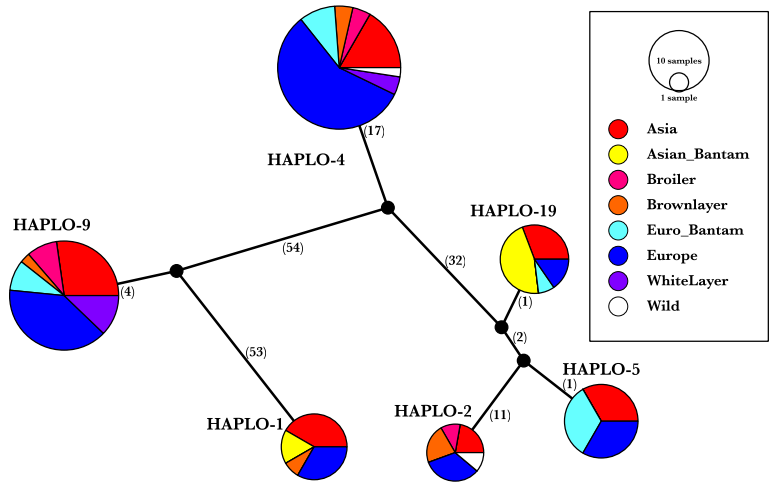

#HAP (4)
#SNPs (47)
#Anim (874)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF
HAPLO-16
(2)
HAPLO-3
(16)
HAPLO-2
(10)
(9)
HAPLO-1
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer

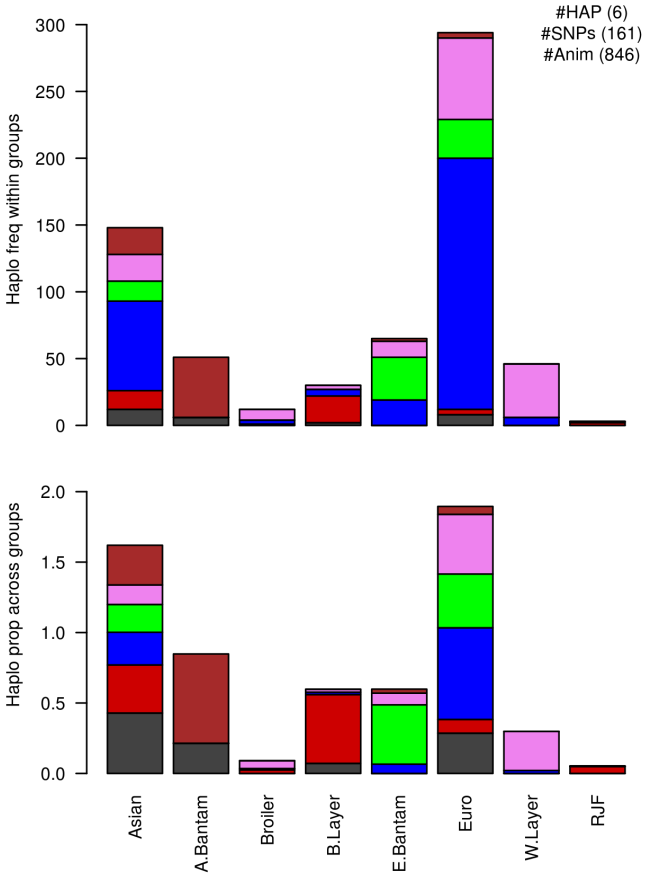
Island 3.2


HAPLO-3
HAPLO-1
HAPLO-19
HAPLO-79
HAPLO-13
HAPLO-50
(31)
(7)
(14)
(14)
(7)
(67)
(14)
(6)
(6)
(13)
(21)
(28)
(5)
(20)
(20)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer

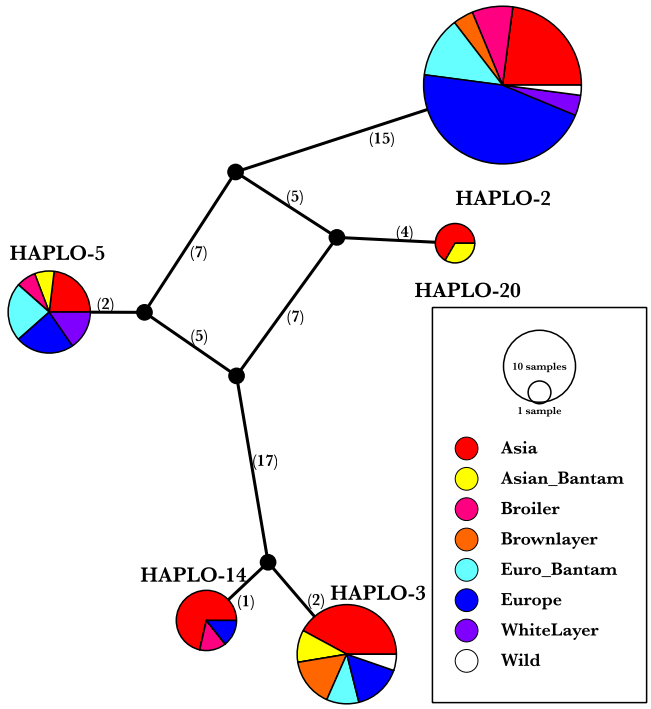

#HAP (6)
#SNPs (191)
#Anim (805)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

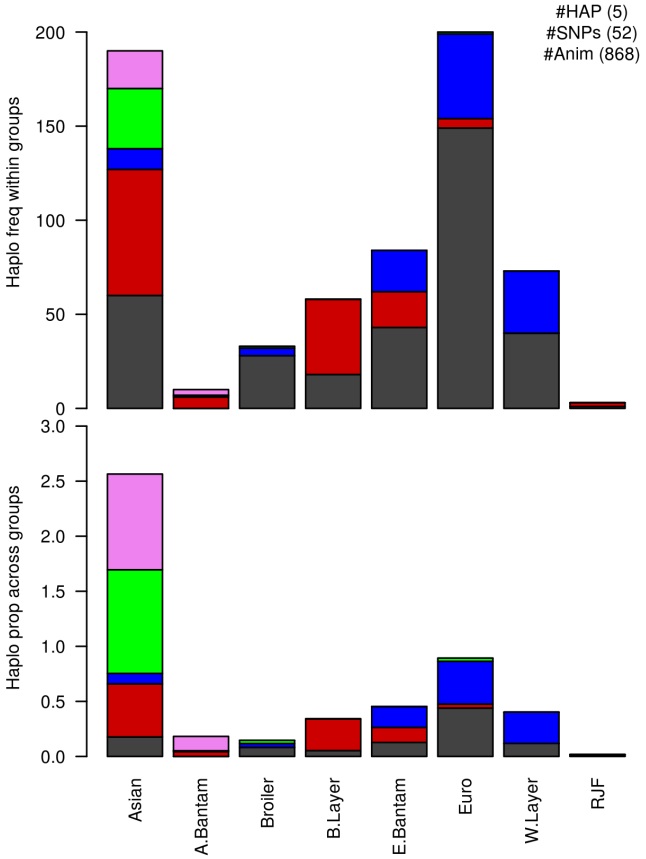
Island 4.1


10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild
HAPLO-4
(17)
HAPLO-9
(4)
(54)
HAPLO-19
(32)
(1)
(2)
(53)
HAPLO-5
(1)
HAPLO-2
(11)
HAPLO-1

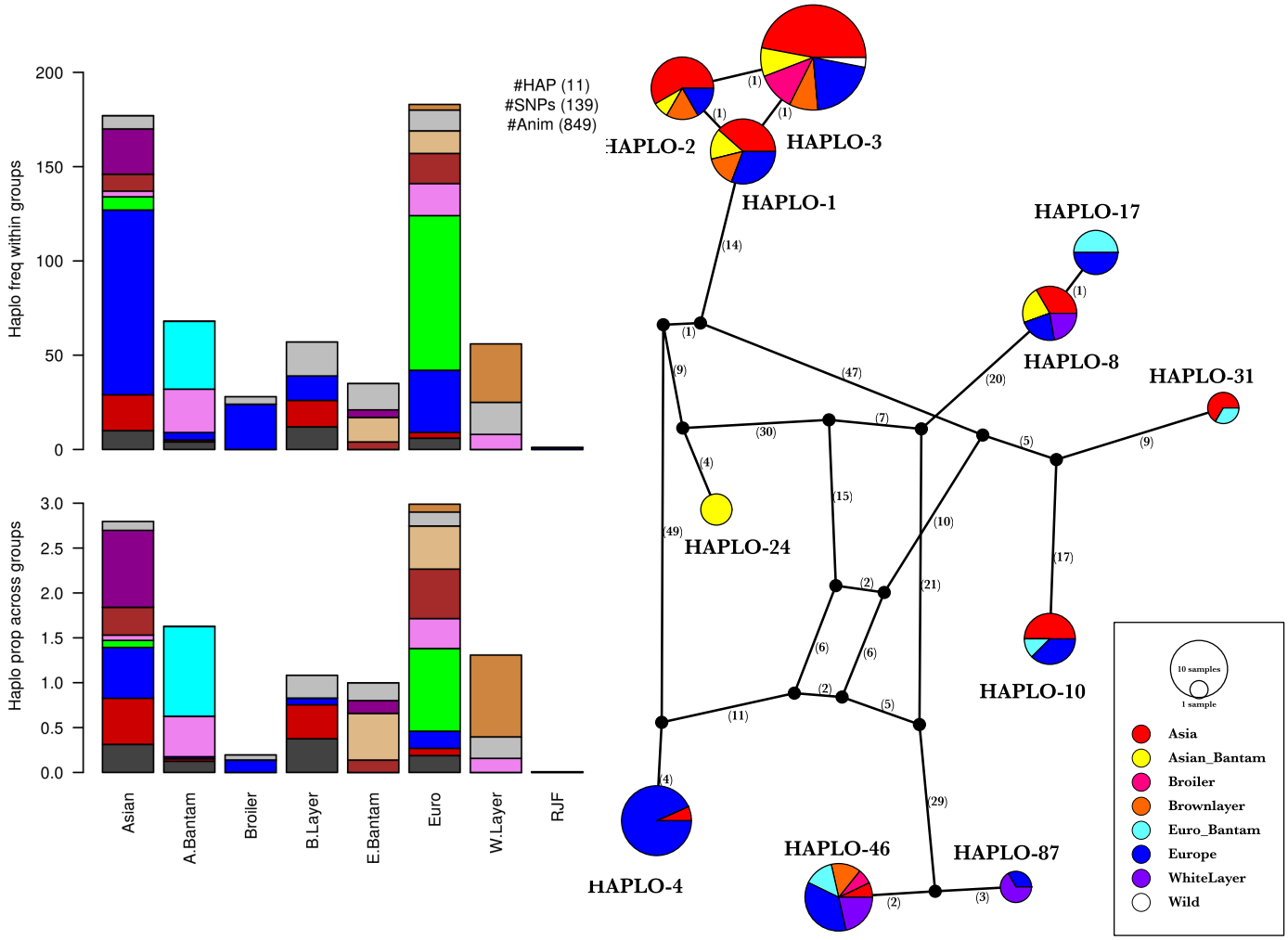

#HAP (6)
#SNPs (161)
#Anim (846)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 4.2

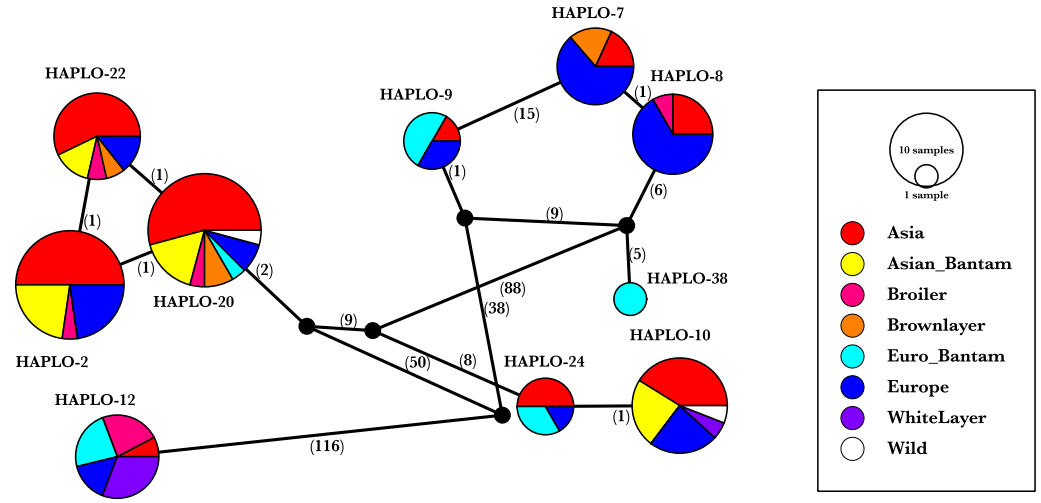

(15)
(5)
HAPLO-2
(4)
HAPLO-5
(7)
(2)
HAPLO-20
(5)
(7)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild
(17)
HAPLO-14
(1)
(2)
HAPLO-3

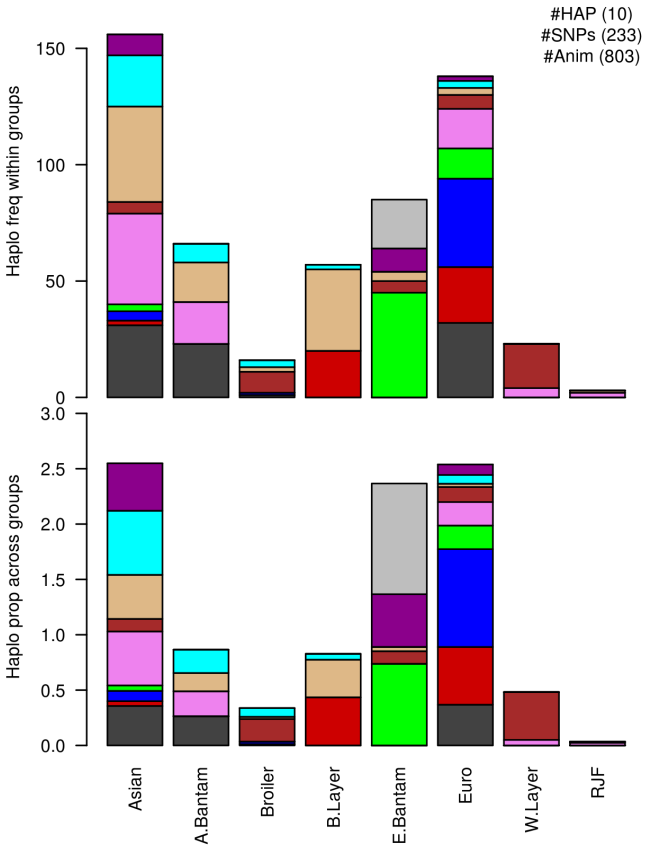

#HAP (5)
#SNPs (52)
#Anim (868)
Haplo freq within groups
200
150
100
50
0
Haplo prop across groups
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 4.4

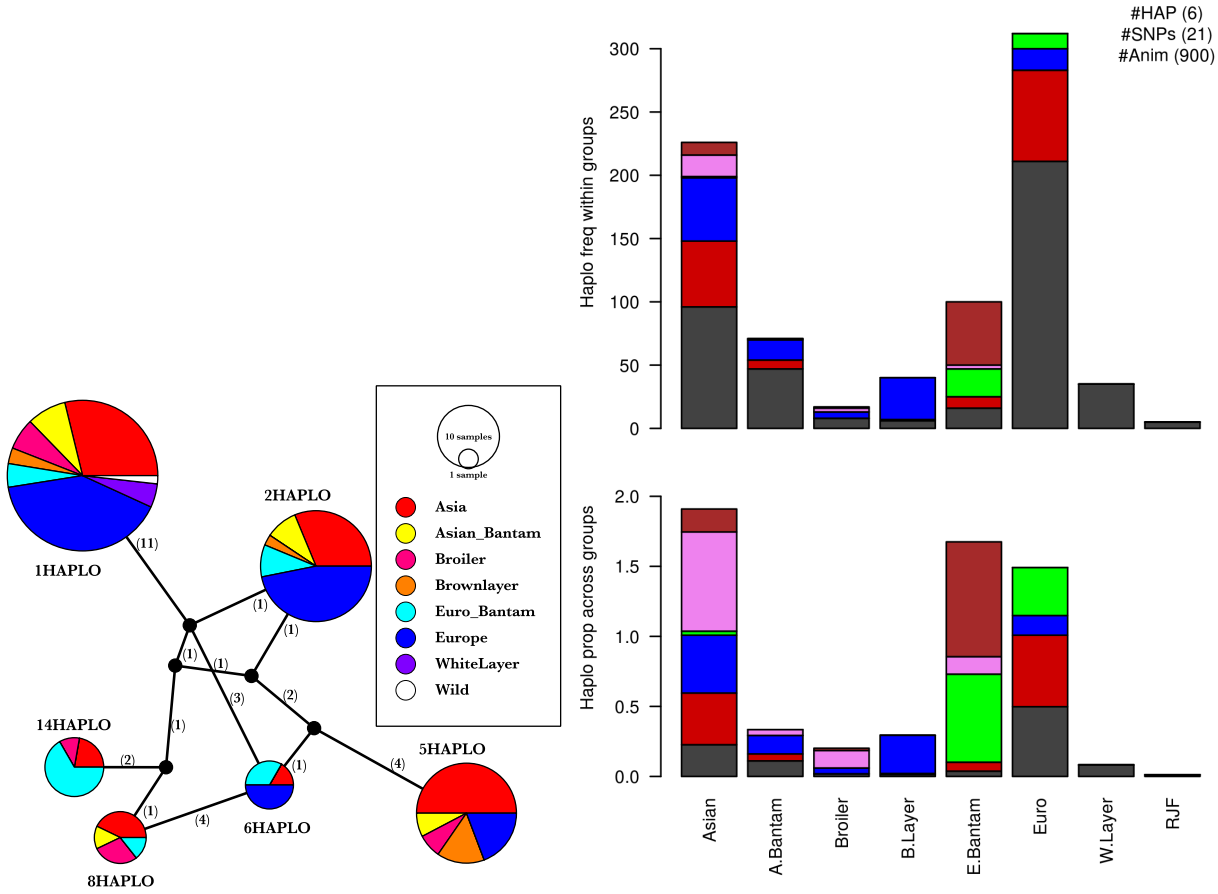

#HAP (11)
#SNPs (139)
#Anim (849)
Haplo freq within groups
Haplo prop across groups
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF
HAPLO-1
HAPLO-2
HAPLO-3
HAPLO-4
HAPLO-8
HAPLO-10
HAPLO-17
HAPLO-24
HAPLO-31
HAPLO-46
HAPLO-87
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild

Island 5.1

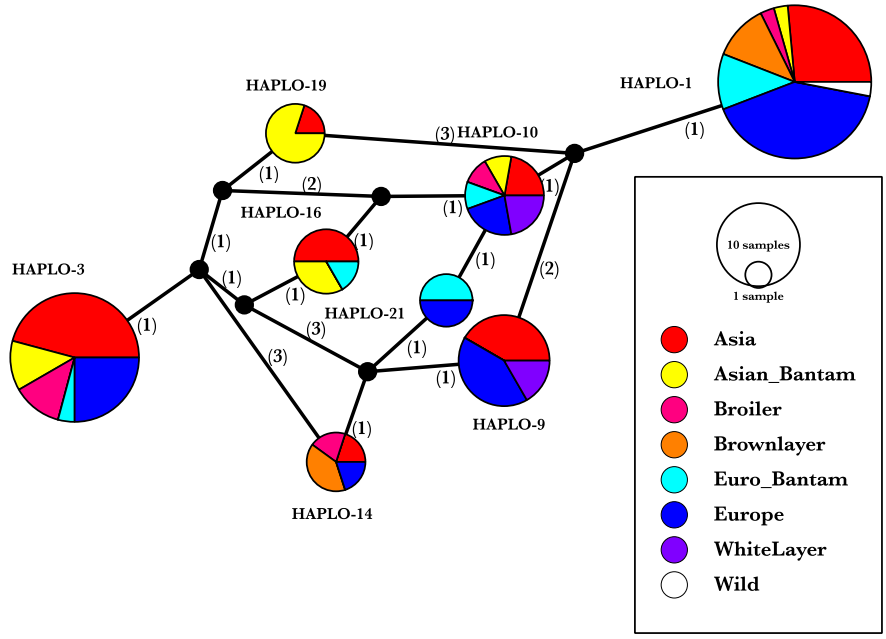

HAPLO-7
HAPLO-22
HAPLO-8
HAPLO-9
(15)
(1)
(1)
(6)
(1)
(9)
(1)
(2)
(5)
HAPLO-20
HAPLO-38
(88)
(38)
(9)
HAPLO-10
HAPLO-2
(50)
(8)
HAPLO-24
HAPLO-12
(1)
(116)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild

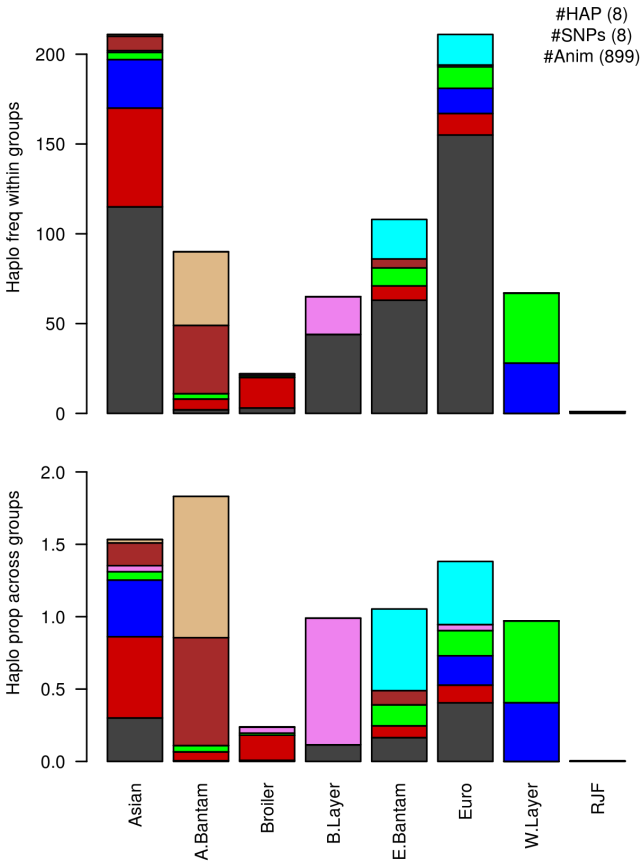

#HAP (10)
#SNPs (233)
#Anim (803)
Haplo freq within groups
150
100
50
0
Haplo prop across groups
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 5.2

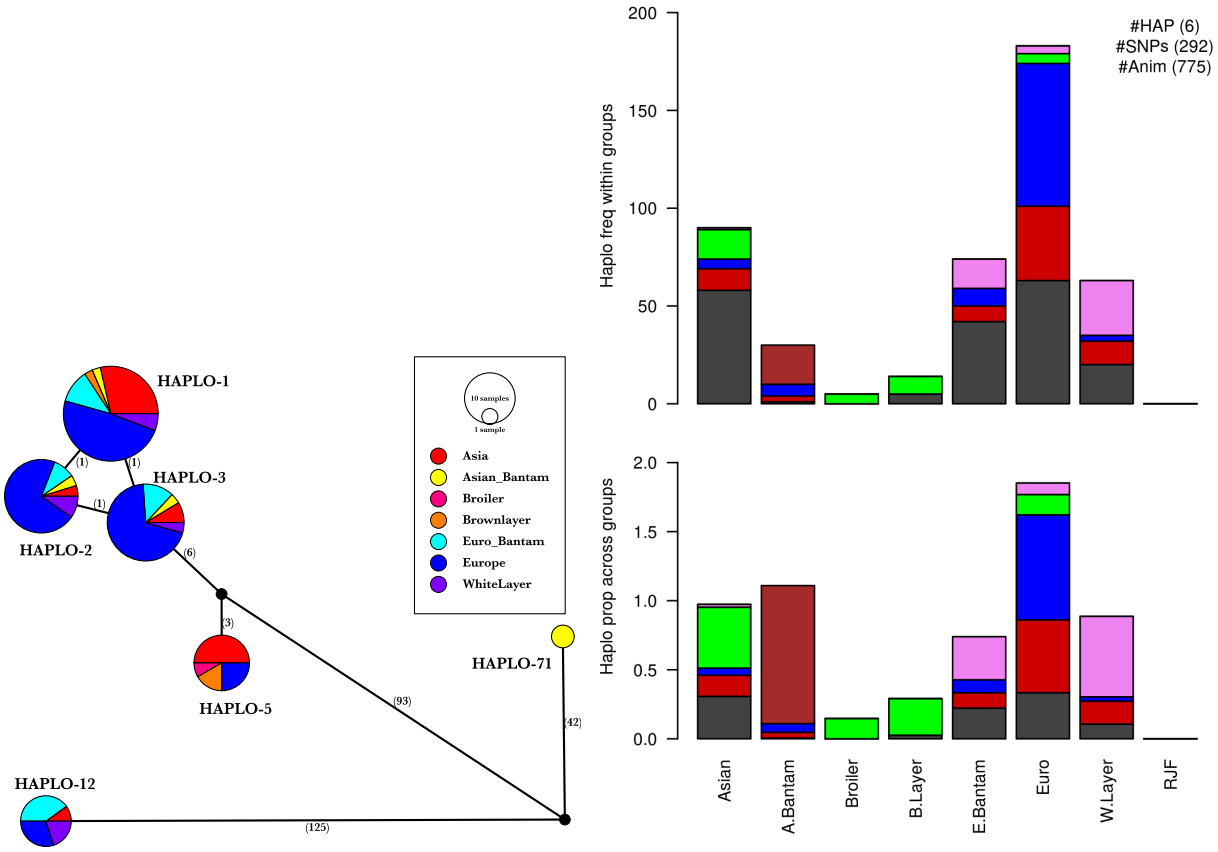

#HAP (6)
#SNPs (21)
#Anim (900)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild
1HAPLO
2HAPLO
5HAPLO
6HAPLO
8HAPLO
14HAPLO
(11)
(1)
(1)
(1)
(1)
(3)
(2)
(1)
(2)
(1)
(4)
(4)

Island 5.4

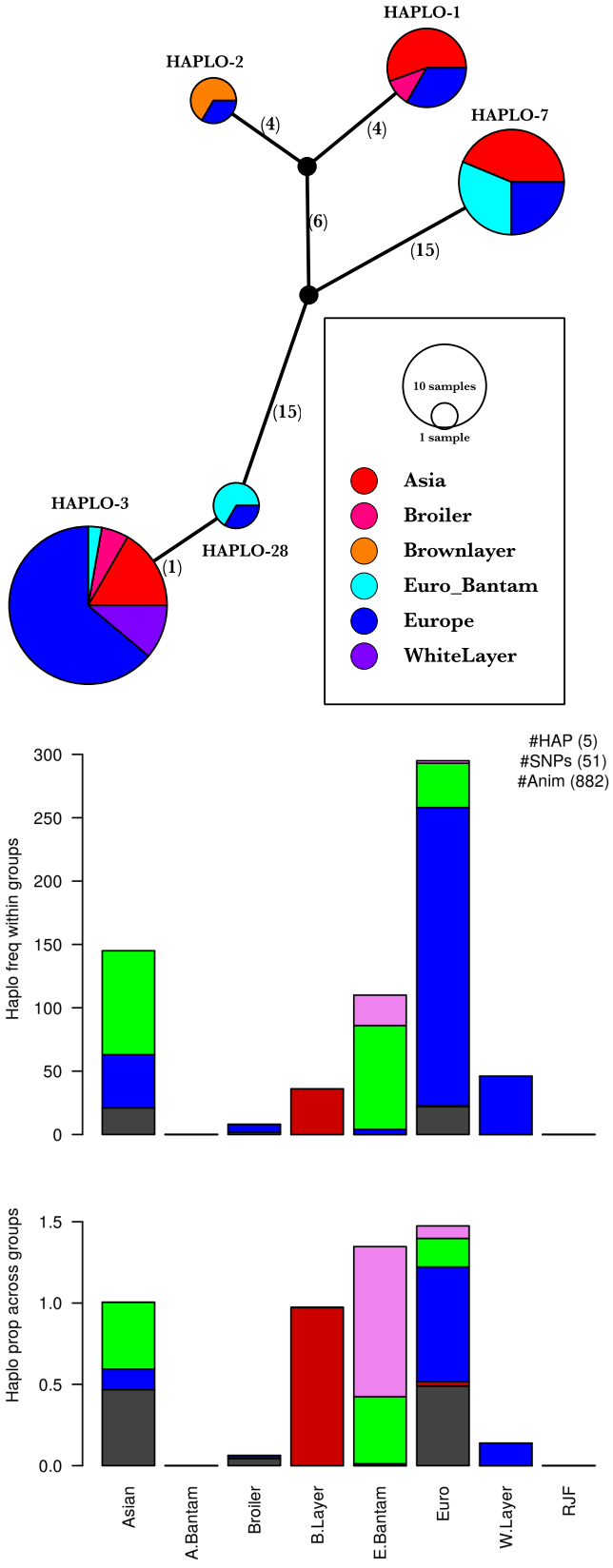

HAPLO-19
HAPLO-1
(3) HAPLO-10
(1)
(1)
(2)
(1)
HAPLO-16
(1)
(1)
HAPLO-3
(1)
(1)
(1)
HAPLO-21
(1)
(2)
(3)
(3)
(1)
(1)
HAPLO-9
(1)
HAPLO-14
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild


#HAP (8)
#SNPs (8)
#Anim (899)
Haplo freq within groups
200
150
100
50
0
Haplo prop across groups
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

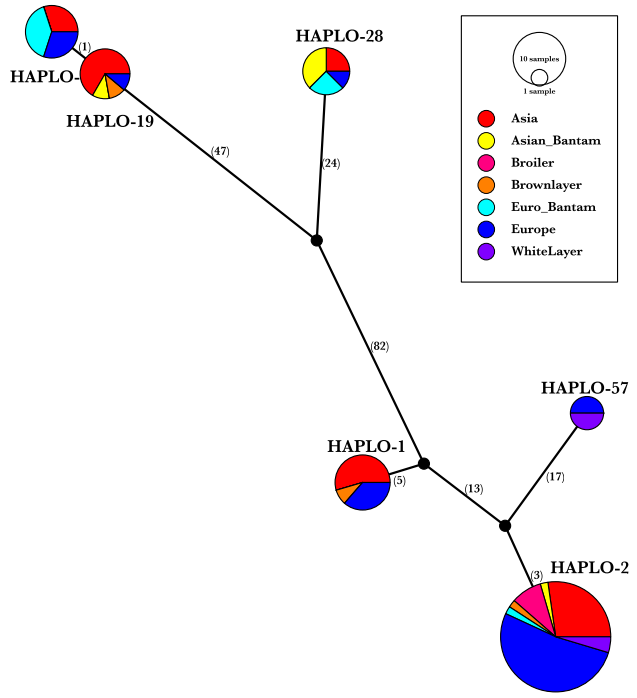
Island 7.1

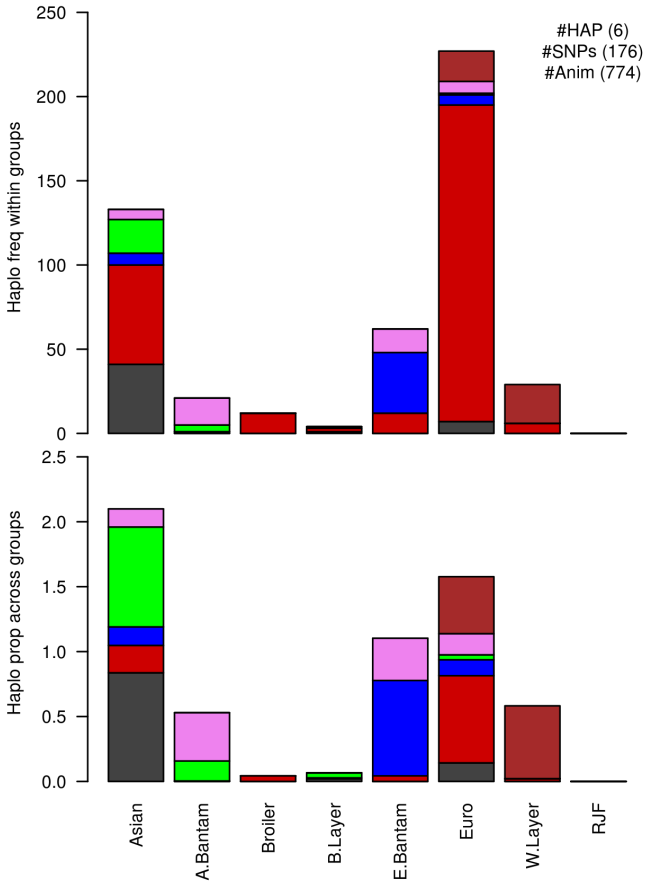

HAPLO-1
HAPLO-2
HAPLO-3
HAPLO-5
HAPLO-12
HAPLO-71
(1)
(1)
(1)
(6)
(3)
(93)
(42)
(125)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
#HAP (6)
#SNPs (292)
#Anim (775)
Haplo freq within groups
Haplo prop across groups
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 7.3

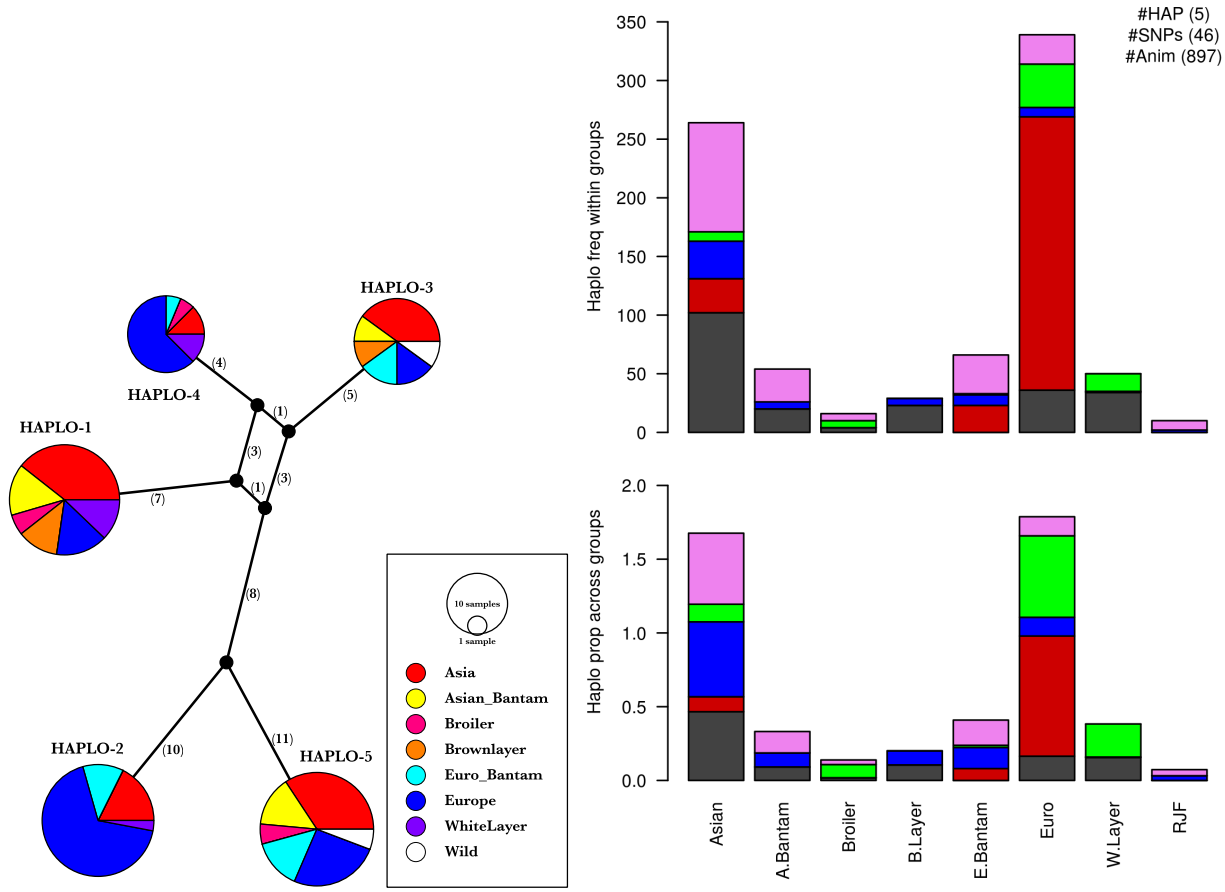

HAPLO-1
HAPLO-2
HAPLO-7
HAPLO-3
HAPLO-28
(4)
(4)
(6)
(15)
(15)
(1)
10 samples
1 sample
Asia
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
#HAP (5)
#SNPs (51)
#Anim (882)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 8.1

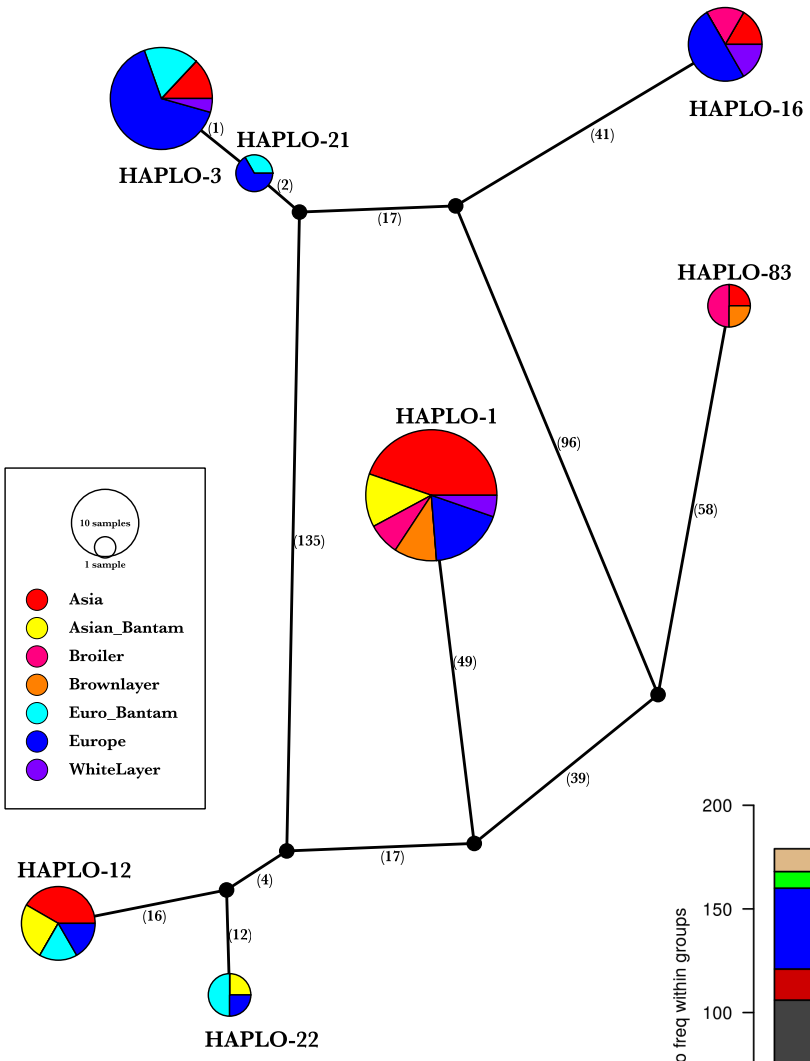

HAPLO-28
(1)
HAPLO-
HAPLO-19
(47)
(24)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
(82)
HAPLO-57
HAPLO-1
(5)
(13)
(17)
(3)
HAPLO-2

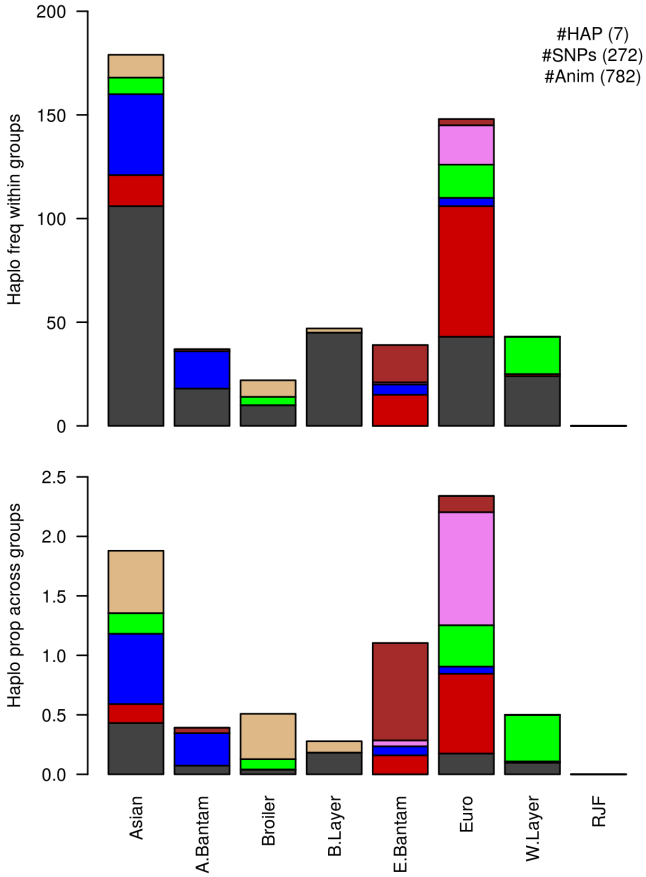

#HAP (6)
#SNPs (176)
#Anim (774)
Haplo freq within groups
250
200
150
100
50
0
Haplo prop across groups
2.5
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 8.2

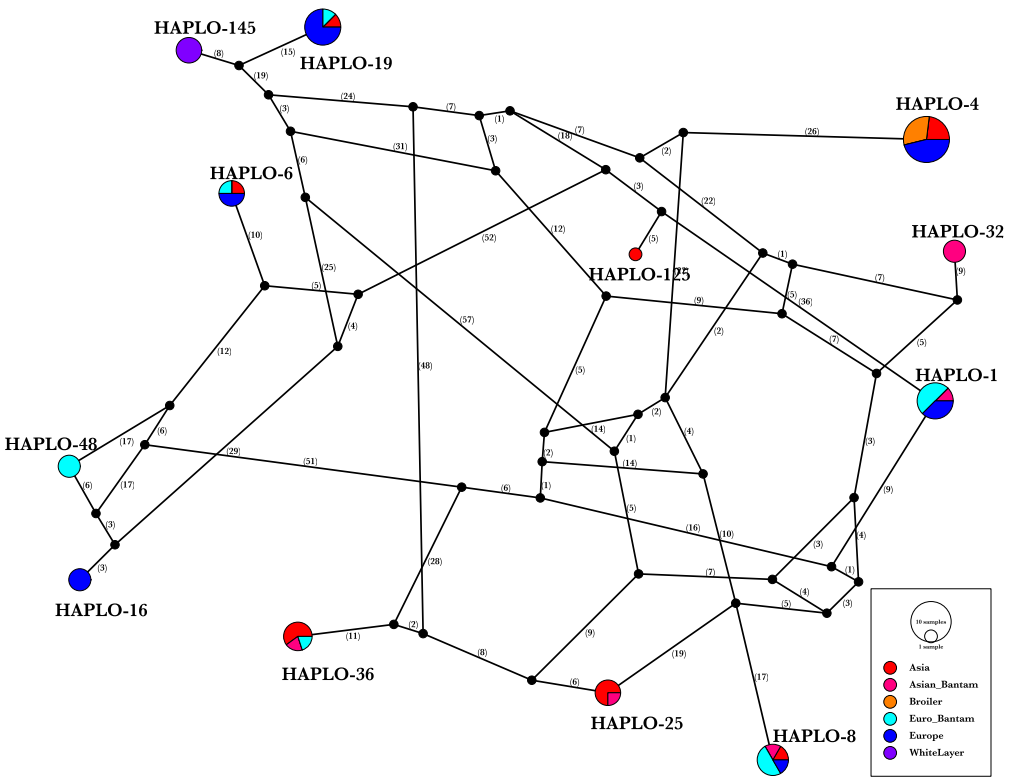

HAPLO-1
HAPLO-2
HAPLO-3
HAPLO-4
HAPLO-5
(1)
(3)
(4)
(5)
(7)
(8)
(10)
(11)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild
#HAP (5)
#SNPs (46)
#Anim (897)
Haplo freq within groups
Haplo prop across groups
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

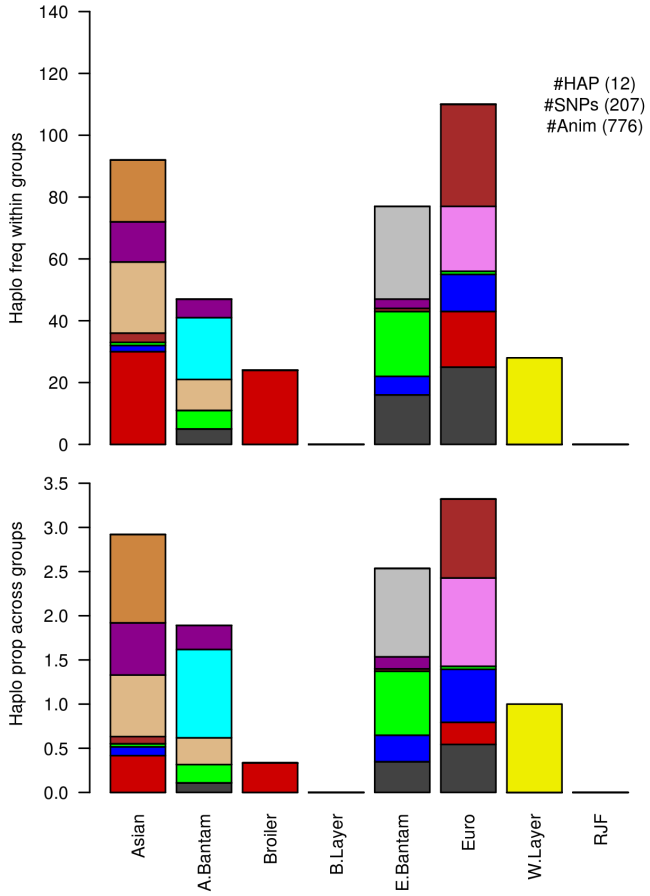
Island 8.3


HAPLO-16
HAPLO-21
HAPLO-3
HAPLO-83
HAPLO-1
HAPLO-12
HAPLO-22
(1)
(2)
(41)
(17)
(96)
(58)
(135)
(49)
(39)
(17)
(4)
(16)
(12)
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer

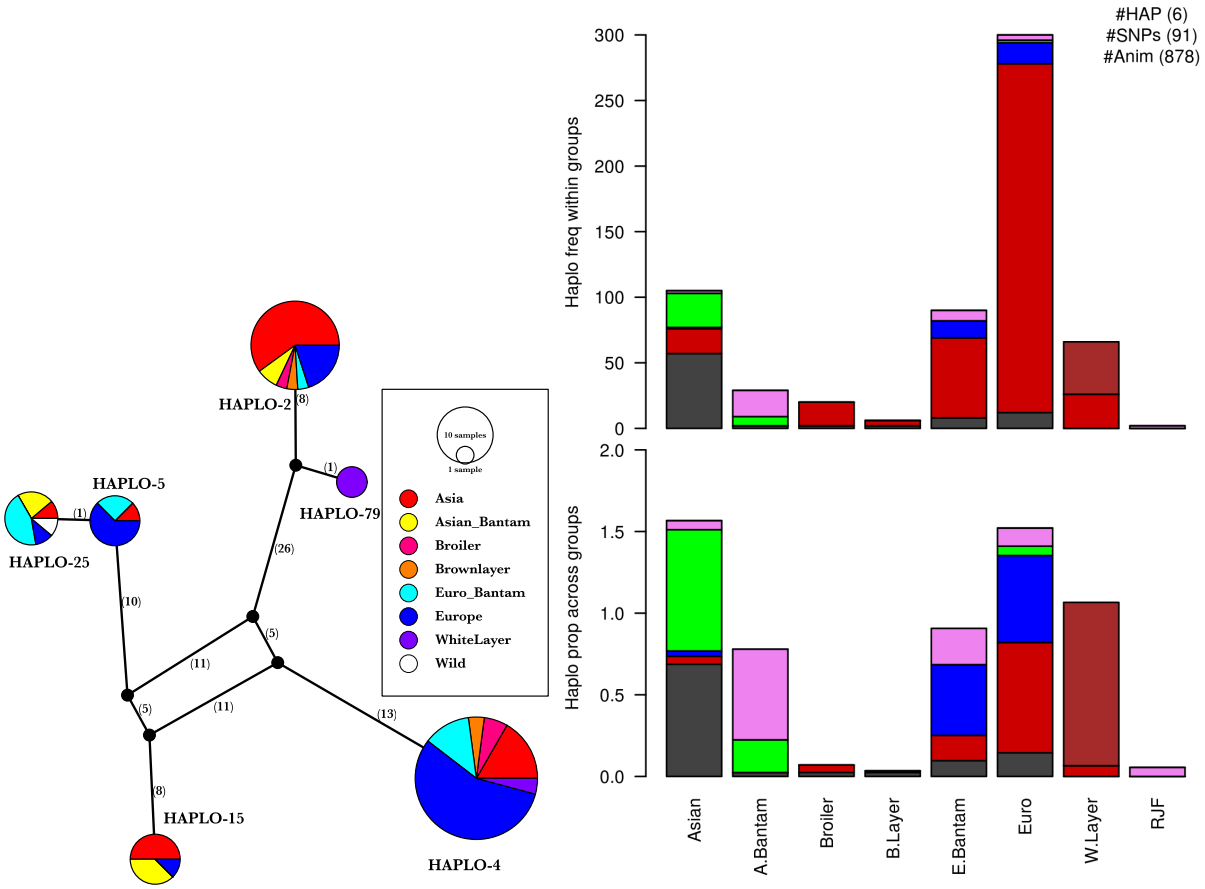

#HAP (7)
#SNPs (272)
#Anim (782)
Haplo freq within groups
Haplo prop across groups
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 11.1

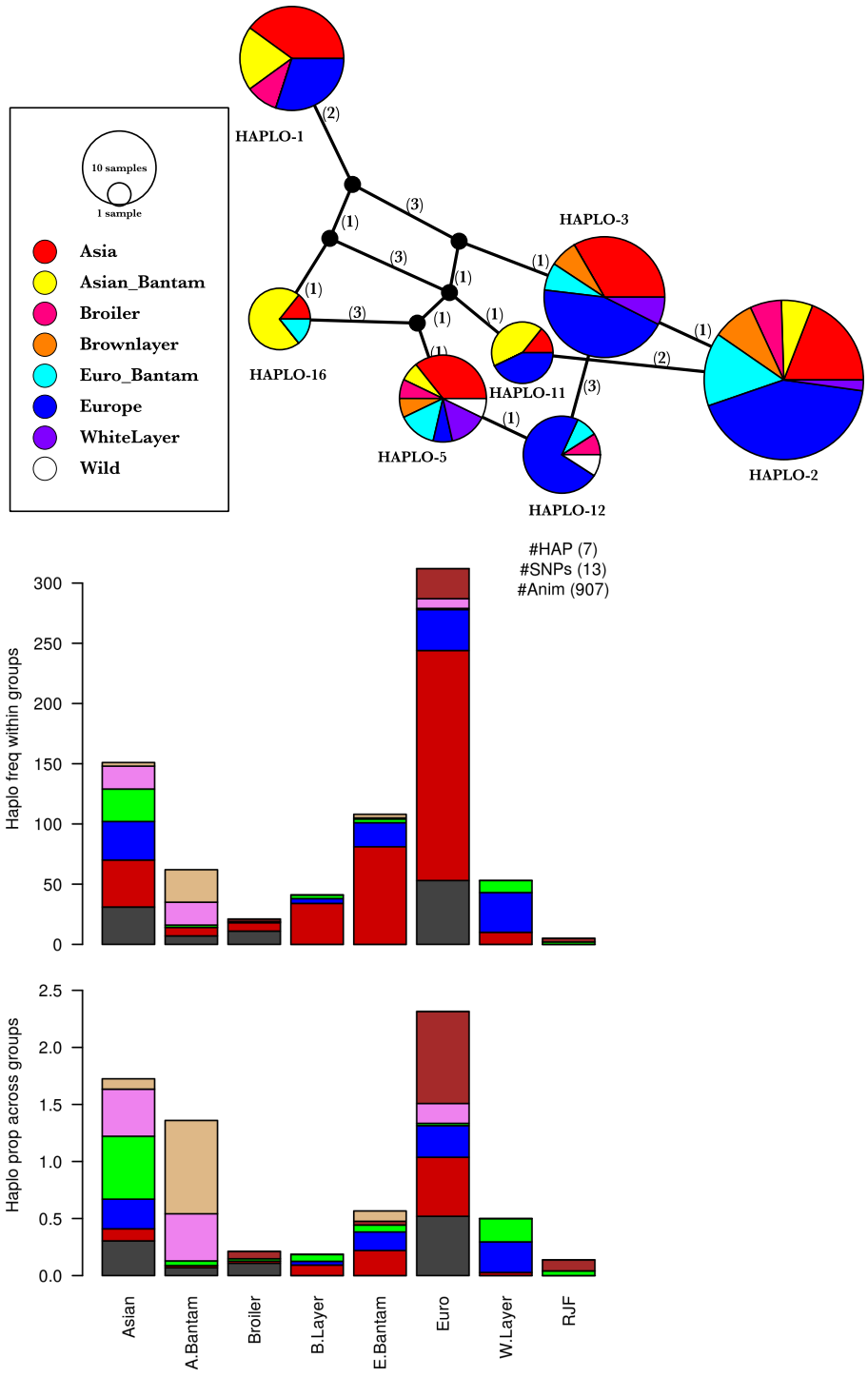

HAPLO-145
HAPLO-19
HAPLO-4
HAPLO-6
HAPLO-32
HAPLO-125
HAPLO-1
HAPLO-48
HAPLO-16
HAPLO-36
HAPLO-25
HAPLO-8
Asia
Asian_Bantam
Broiler
Euro_Bantam
Europe
WhiteLayer


#HAP (12)
#SNPs (207)
#Anim (776)
Haplo freq within groups
Haplo prop across groups
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 17.1


HAPLO-2
(8)
(1)
HAPLO-79
HAPLO-5
(1)
HAPLO-25
(26)
(10)
(5)
(11)
(5)
(11)
(13)
(8)
HAPLO-15
HAPLO-4
10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild
#HAP (6)
#SNPs (91)
#Anim (878)
Haplo freq within groups
300
250
200
150
100
50
0
Haplo prop across groups
2.0
1.5
1.0
0.5
0.0
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island 27.1


10 samples
1 sample
Asia
Asian_Bantam
Broiler
Brownlayer
Euro_Bantam
Europe
WhiteLayer
Wild
HAPLO-1
HAPLO-2
HAPLO-3
HAPLO-5
HAPLO-11
HAPLO-12
HAPLO-16
#HAP (7)
#SNPs (13)
#Anim (907)
Haplo freq within groups
Haplo prop across groups
Asian
A.Bantam
Broiler
B.Layer
E.Bantam
Euro
W.Layer
RJF

Island	Position	QTL
1.1	32105873 - 32150135	- Abdominal fat percentage - Antibody response to SRBC antigen - Body weight (8, 21, 42, 46, 63x2, 112, 200 days) - Body weight (x3) - Breast muscle weight - Carcass weight - Chest width - Cingular fat width - Cloacal bacterial burden after challenge with Salmonella E - Drumstick and thigh muscle weight (x2) - Egg weight (x5) - Feat factor - Femur length - Growth (1-8, 8-46, 46-112, 112-200 days) - Head width - Heart weight - Insulin level - Leg twisting - Lung weight - Pectoralis major weight - Shank length (x3) - Subcutaneous fat thickness - Subcutaneous neck fat weight - Thigh weight - Tibia marrow diameter - Total white fat weight - Visceral fat weight - Wing weight
1.3	41305981 - 41361805	- Age at first egg - Age at sexual maturity - Antibody response to SRBC antigen - Body weight (8, 21, 42, 48, 63x2, 112, 140, 200 days) - Body weight (x3) - Breast muscle weight - Carcass weight - Cingular fat width - Comb size - Conformation score - Drumstick and thigh muscle weight - Drumstick and thigh percentage - Drumstick and thigh weight (x2) - Drumstick muscle weight - Egg weight (x4) - Fear-walk latency - Feed conversion ratio - Femur bending strength - Growth (1-8, 8-46, 46-112, 112-200 days) - Leg twisting - Lung weight (x2) - Marek's disease-related traits - Residual feed intake - Shank length (x4) - Spleen weight - Subcutaneous fat thickness - Subcutaneous neck fat weight - Thigh weight - Tibia marrow diameter - Total white fat weight - Visceral fat weight - Wing weight
1.5	73659173 - 75302834	- Abdominal fat percentage (x2) - Abdominal fat weight (x2) - Age at first egg - Antibody response to SRBC antigen - Body weight (14, 42, 49, 140 days) - Body weight (x2) - Breast muscle percentage - Breast muscle weight - Carcass weight (x2) - Chest width (x2) - Cingular fat width - Cloacal bacterial burden after challenge with Salmonella T - Comb size
		- Creatine kinase level - Drumstick and thigh weight - Egg weight (x3) - Head percentage - Humerus area - Insulin-like growth factor level - Intramuscular fat - Shank length (x2) - Spleen weight - Subcutaneous fat thickness - Subcutaneous neck fat weight - Total white fat weight - Visceral fat weight - Wing percentage
1.6	126607488 - 126994702	- Abdominal fat weight (x2) - Age at first egg - Body weight - Body weight (21, 35, 42x3, 49, 70, 84, 140 days) - Breast pH - Carcass weight - Chest width - Corticosterone level - Corticosterone response - Drip loss - Egg weight (x3) - Femur weight - Receiving feather pecking - Shank length (x2) - Shank weight - Spleen weight - Subcutaneous fat thickness - Total white fat weight - Visceral fat weight
1.7	159978139 - 160371348	- Abdominal fat weight (x2) - Age at first egg - Average daily gain - Blood and meat spots - Body depth - Body weight (8, 21, 35, 42x2, 46, 49, 112, 168, 200 days) - Body weight (x3) - Breast muscle weight - Carcass weight - Drumstick and thigh muscle weight - Egg weight (x2) - Feet weight - Gizzard weight (x2) - Growth (1-8, 8-46, 46-112, 112-200 days) - Head width - Heart weight - Intestine length (x2) - Shank circumference - Skin fat weight - Spleen weight - Total white fat weight - Wing weight
1.8	190571514 - 190871495	- Abdominal fat weight (x2) - Average daily gain - Body weight (8, 21, 42, 46, 112, 168, 200 days) - Body weight (x2) - Carcass ash content, dry matter basis - Carcass fat content, dry matter basis (x2) - Carcass protein content, dry matter basis - Drip loss - Drumstick and thigh muscle weight - Growth (0-14, 1-8, 8-46, 46-112, 112-200 days) - Insulin-like growth factor level - Intestine length - Pectoralis major weight - Shank weight - Shank weight percentage - Skin fat weight
2.1	25348292 - 25734199	- Antibody response to KLH antigen - Body weight (42 days) - Body weight (x3) - Drumstick and thigh muscle weight

		- Haugh units - Insulin level - Insulin-like growth factor level - Marek's disease-related traits - Subcutaneous neck fat weight - Total white fat weight
2.2	51437340 - 51547872	- Body weight (42, 46, 140, 200 days) - Breast muscle weight (x2) - Breast percentage - Carcass weight - Cloacal bacterial burden after challenge with Salmonella E - Drumstick and thigh percentage - Factor fear - Fear-stand latency - Femur bone mineral density - Head percentage - Shank weight percentage - Subcutaneous neck fat weight - Total white fat weight - Visceral fat weight
2.3	51580308 - 54072586	- Body weight (42, 46, 140, 200 days) - Breast muscle weight (x2) - Breast percentage - Carcass weight - Cloacal bacterial burden after challenge with Salmonella E - Drumstick and thigh percentage - Factor fear - Fear-stand latency - Femur bone mineral density - Head percentage - Shank weight percentage - Subcutaneous neck fat weight - Total white fat weight - Visceral fat weight
2.5	140506904- 140558276	- Average daily gain - Body weight (63 days) - Body weight (test end) - Breast muscle weight - Cloacal bacterial burden after challenge with Salmonella E - Fat distribution - Liver weight - Shank diameter - Tibia bone mineral density
2.6	142925250- 143218775	- Average daily gain - Body weight (63 days) - Body weight (test end) - Breast muscle weight - Cloacal bacterial burden after challenge with Salmonella E - Fat distribution - Liver weight - Shank diameter - Tibia bone mineral density
3.1	63776707 - 63957074	- Antibody response to KLH antigen - Antibody response to SRBC antigen - Body weight (8, 35, 49, 63, 112, 200 days) - Body weight (x4) - Fear-stand latency - Glucose level - Growth (1-8 days) - Leg twisting - Lung weight - Proventriculus weight - Skin fat weight - Thigh muscle weight - Tibia area - Tibia breaking force - Tibia length - Tibia stress
3.2	109772365- 110428731	- Abdominal fat percentage - Abdominal fat weight - Average daily gain - Body weight (49 days) - Carcass weight - Chest width
4.1	18391954 - 19459250	- Age at first egg - Average daily gain (x2) - Body weight - Body weight (21, 42, 84, 168, 336, 504 days) - Body weight (day of first egg) - Conformation score - Creatine kinase level - Drumstick and thigh muscle weight - Egg weight - Fear-related behaviour - Feed conversion ratio - Head percentage - Marek's disease-related traits - Residual feed intake - Shank length - Thigh muscle weight - Thigh weight - Tibia strength - Visceral fat weight
4.2	27055953 - 27306201	- Age at first egg - Average daily gain (x2) - Body weight (21, 42, 84, 168, 336, 504 days) - Body weight (day of first egg) - Conformation score - Creatine kinase level - Drumstick and thigh muscle weight - Egg number - Egg weight (x3) - Fear-related behaviour - Feed conversion ratio - Head percentage - Marek's disease-related traits - Residual feed intake - Shank length - Thigh muscle weight - Tibia strength - Visceral fat weight
4.4	70532800 - 71221544	- Abdominal fat percentage - Body temperature - Fear-related behaviour
5.1	2091956-4038938	- Body weight (7,14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84 days) - Carcass weight - Head width
5.2	40954167 - 41043854	- Abdominal fat percentage - Abdominal fat weight (x2) - Antibody titer to LTA antigen - Body weight - Body weight (140 days) - Breast muscle weight - Carcass weight - Cloacal bacterial burden after challenge with Salmonella E - Egg number - Egg weight (x2) - Gizzard weight - Glucose level - Growth (42-63 days) - Head width - Non-esterified fatty acid level - Receiving feather pecking - Shank diameter - Shank length - Troponin T concentration - Wing weight
5.4	49119327 - 49143659	- Abdominal fat weight - Body weight (140 days) - Body weight (21, 42 days) - Breast muscle weight - Carcass weight - Egg number - Egg weight (x2) - Gizzard weight - Growth (1-8, 42-63 days) - Head width - Non-esterified fatty acid level - Receiving feather pecking

		- Shank diameter - Troponin T concentration - Wing weight
7.1	6744362-8243239	- Abdominal fat weight (x2) - Average daily gain - Body weight (112, 200 days) - Body weight (x4) - Breast muscle weight - Carcass weight - Drumstick and thigh muscle weight - Drumstick muscle weight (x2) - Drumstick weight - Fat distribution - Heart weight - Shank weight - Skin fat weight - Thigh muscle weight (x2) - Thigh weight - Wing weight
7.3	18530576 - 18671971	- Abdominal fat weight - Average daily gain - Body weight (14, 35, 77, 112, 200 days) - Body weight (x6) - Breast muscle weight - Carcass weight - Drumstick and thigh muscle weight - Drumstick muscle weight - Drumstick weight - Fat distribution - Intestine length - Skin fat weight - Thigh muscle weight - Thigh weight - Wing weight
8.1	9100-838224	
8.2	9229140-9340962	- Tibia width - Body weight (63, 504 days) - Tibia bone mineral density (x2) - Drumstick and thigh weight - Abdominal fat percentage - Body weight (x2) - Thigh meat-to-bone ratio
		- Body weight (day of first egg) - Breast muscle weight - Chest width - Drumstick muscle weight - Growth (21-42 days) - Tibia marrow diameter - Tibia plateau angle - Tibia strength - Tibia weight - Tibia width - Wing angle - Marek's disease-related traits - Crooked digits
8.3	9428942-10896059	- Abdominal fat percentage - Body weight (63, 504 days) - Body weight (day of first egg) - Body weight (x2) - Breast muscle weight - Chest width - Crooked digits - Drumstick and thigh weight - Drumstick muscle weight - Growth (21-42 days) - Marek's disease-related traits - Thigh meat-to-bone ratio - Tibia bone mineral density (x2) - Tibia marrow diameter - Tibia plateau angle - Tibia strength - Tibia weight - Tibia width - Tibia width - Wing weight
11.1	18521596 - 19311908	- Body weight (40, 46, 140 days) - Breast color (x2) - Carcass weight (x2) - Spleen weight - Thigh meat-to-bone ratio
17.1	8930249-9094716	- Egg production rate - Troponin T concentration
27.1	446512-801088	- Antibody response to KLH antigen - Antibody response to SRBC antigen - Wing weight