University of KwaZulu-Natal

Estimation of genetic and demographic parameters of extensively raised chicken populations using genome-wide single nucleotides polymorphism (SNP) data

By

Khulekani Sedwell Khanyile

207505346

A dissertation submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN GENETICS

School of Life Sciences

Supervisor: Mr. Edgar Farai Dzomba

Co-Supervisor: Dr. Farai Catherine Muchadeyi

March 2015

DECLARATION

I declare that the entirety of the work contained herein is my own original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by University of KwaZulu-Natal will not infringe any third party rights and that I have not previously, in its entirety or in part, submitted it for obtaining any qualification. The experimental work described in this dissertation was conducted under the supervision of Mr. E.F. Dzomba and Dr. F.C. Muchadeyi. All assistance towards the production of this work and all the references contained herein have been duly acknowledged.

_____ _____

K.S. Khanyile **Date:**

Approved in style and content by:

Mr E.F. Dzomba _____ **Date:** _____

Dr F.C. Muchadeyi _____ **Date:** _____

RESEARCH OUTPUT AND AUTHOR CONTRIBUTIONS

The following conference proceedings and conference poster were generated during the study

JOURNAL ARTICLES

K.S. Khanyile, E.F. Dzomba and F.C. Muchadeyi. 2015. Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa. Frontiers in Genetics, 6: 13. This proceeding is part of chapter 3 of the thesis and is original work from Mr Khanyile.

CONFERENCE PROCEEDINGS

K.S. Khanyile, E.F. Dzomba and F.C. Muchadeyi. 2013. The extent and distribution of linkage disequilibrium in extensively raised chicken population of Southern Africa. The 20th Association in Advance Animal Breeding and Genetics, held on the 20-23 October 2013, Napier, South Africa. Proceeding paper. Proc. Assoc. Advmt. Anim. Breed. Genet. 20: 364-367. Available at http://www.aaabg.org/aaabghome/AAABG20papers/khanyile20364.pdf. This proceeding is part of chapter 3 of the thesis and is original work from Mr Khanyile.

ORAL PRESENTATION

K.S. **Khanyile**, E.F. Dzomba and F.C. Muchadeyi. 2014. Haplo-block structure of Southern African village chicken population. The work presented at the School of Life Sciences Postgraduate Research Day 2014, held 26 May 2014, University of KwaZulu-Natal Westville campus, South Africa. This presentation was chapter 4 and is entirely the work of Mr Khanyile.

POSTERS

K.S. Khanyile, E.F. Dzomba and F.C. Muchadeyi. 2014. Haplo-block structure of Southern African village chicken populations. Poster presented at the 10th World Congress on Genetics Applied to Livestock Production (WCGALP) held 17–22 August 2014 in Vancouver, Canada. This poster forms part of chapter 4 and is entirely the work of Mr Khanyile.

K.S. Khanyile, E.F. Dzomba and F.C. Muchadeyi. 2013. Population genetics of Southern African village chickens using genome-wide SNP data. A poster presented for the 46th Congress of the South African Society for Animal Science, held 23-26 June 2013 at the University of the Free State, Bloemfontein, South Africa. This poster contribution form part of chapter 3 and is entirely the work of Mr Khanyile.

K.S. Khanyile, E.F. Dzomba and F.C. Muchadeyi. 2013. The Extent and Distribution of Linkage Disequilibrium in Extensively Raised Chicken Populations of Southern Africa. A poster presented at the 2013 Annual College of Agriculture, Engineering and Science Post-graduate Research Day – University of KwaZulu-Natal Howard College Campus, 1 November 2013. This poster contribution forms part of Chapter 3 and is entirely the work of Mr Khanyile.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to the following:

My supervisors Mr E.F. Dzomba and Dr F.C. Muchadeyi for their professional guidance, data collection and analysis, advice and dedication to the last moment of this work.

Mrs S.O. Makina, Ms K.T. Ncube for assisting with DNA sample preparation, processing and genotyping at ARC-Biotechnology Platform and for the peer-to-peer advice.

The Agricultural Research Council-Biotechnology Platform (ARC-BTP) for funding genotyping of samples. I held fellowships from the ARC-Professional Development Program, National Research Foundation, and the University of KwaZulu-Natal postgraduate program for which I am very thankful.

ABSTRACT

Village chicken populations are raised under a farming system that faces a number of challenges such as small flock size, lack of animal performance and pedigree records, lack of proper husbandry and poorly defined and structured mating systems all of which can negatively influence the genetic structure of the populations. Understanding of the evolutionary history, demographic structure, inbreeding levels and risk of a population to extinction is important in facilitating genetic improvement programs while maintaining biodiversity of extensively raised chicken populations. Linkage disequilibrium (LD) is an important source of information about historical events of recombination in a population and together with an understanding of the haplotype structure can provide valuable guidelines for breed improvement. This study was undertaken to investigate the existing LD level, inbreeding levels, effective population size and haplotype structure of extensively raised chicken populations from Southern Africa. A total of 312 village chickens from Malawi (n = 30, from one ecotype), South Africa, (n = 146, from three different ecotypes) and Zimbabwe (n = 135, from three different ecotypes) were genotyped using the Illumina chicken iSelect SNP60K bead chip. LD was calculated for each population from a total of 43,175 SNP after pruning for minor allele frequency (MAF) <0.05, genotyping call rate of <0.95, and deviation from Hardy Weinberg Equilibrium (HWE) $p < 0.001$ and missing genotypes of more than 5%. Linkage disequilibrium averaged 0.41±0.006 and was observed to extend up to a marker distance of 100 kb. From the LD, effective population size was estimated that indicated reduced size of the breeding population over the past 40 generations to less than 20 individuals. Haplo-block structure analysis resulted in a total of 649, 2104 and 2442 blocks from Malawi, South Africa and Zimbabwe, respectively. Most of the observed blocks were less than 20 kb with a few that were more than 500 kb. Haplo-block genome coverage was 39 Mbp, 64.4 Mbp and 54.5 Mbp for Malawi, South Africa, and Zimbabwe, respectively. Large haplo-blocks on chromosome 8 spanned QTL regions associated mostly with body composition traits. The LD pattern was consistent with low effective population sizes and loss of heterozygosity in the village chicken populations. Potentially useful haplo-blocks spanning regions of known QTLs should be targeted for

further analysis and identification of genes conferring optimal production performance of village chickens under harsh and marginalized production systems. Overall, the study provides baseline information on the utility of genome wide SNP data in studying extensively raised village chicken populations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1

1.1 INTRODUCTION

Smallholder communal farmers in Southern Africa and other developing countries raise indigenous chickens under low input production systems. This extensive system of production is characterized by minimal management and veterinary interventions for maximizing production (Gondwe & Wollny 2007; Muchadeyi et al. 2007a; Mtileni et al. 2010) These chickens however, have a major contribution to many village communities as a source of protein (from meat and eggs), source of income and other cultural contributions. Furthermore, unlike other domestic livestock that are raised under intensive systems of production with management and protection from being extinct, village chickens receive little or no attention for conservation and genetic improvement in Southern Africa and most developing countries (Abdelqader et al. 2007; Granevitze et al. 2007; Kaya & Yildiz 2008). Village chickens are generally described as non-descript birds that have not been developed as a breed and with uncharacterized genetic attributes. Studies have revealed that chickens raised by small communal farmers are characterized by their morphological traits, plumage pattern, locally preferred phenotypic merits, and are also categorized by their farm specific production traits (Abdelqader et al. 2007). Local chickens are found in the most marginalized farming systems, to which they have adapted and survived the harsh environmental conditions, exposure to parasitic and bacterial pathogens and the associated diseases, and to poor or no sheltering and low quality nutrition (FAO 1995; Ekue F.N. 2002; Illango J. 2002). Such survival and adaptive attributes need to be characterized, conserved, and utilized (Muchadeyi et al. 2006; van Marle-Köster et al. 2008; Mtileni et al. 2010).

Several studies suggested that village chicken populations hold valuable genetic diversity. Diversity studies using autosomal microsatellite (Nidup et al. 2005; Muchadeyi et al. 2007a; Mtileni et al. 2011a) and mitochondria-DNA sequences (Muchadeyi et al. 2009; Mtileni et al. 2011b) on Southern African local chickens have unveiled high levels of genetic variation within and among the village chicken

populations and identified maternal origins of Southern African chicken populations. The above-mentioned studies have also indicated that village chickens contribute genetic variation that is unique and different from diversity exhibited by commercial and specialized chicken populations.

There is a need to improve the genetics of these village chickens in such a way that productivity is improved and genetic diversity retained in the population. To achieve this requires a good understanding of the evolutionary history and demographic structure, inbreeding levels and risk of the populations to extinction. Knowledge of the level of divergence between populations gives insight into the genetic uniqueness of the populations and together with an identification of the genetic merits of each population for key production traits will inform conservation and breed management and utilization policies. The investigation of demographic structure and other population parameters is not feasible without complete pedigree and performance data. Most village chicken populations are raised under scavenging systems of production in which farmers do not keep records on parentage and performance of their flocks (Muchadeyi et al. 2009).

The availability of whole genome SNP data facilitates the use of powerful statistical methods to investigate the evolutionary history and population demographics of village chickens. The availability of the whole genome sequence of chickens (Hillier et al. 2004) provided opportunities for using high-density SNP arrays.

Linkage disequilibrium (LD) is non-random association between two or more allele at different loci (Wang 2005). Linkage disequilibrium as a function of effective population size (N_e) has been used to estimate other demographic parameters such as rate and level of inbreeding as well as selection pressures in the absence of pedigree and performance information (Li & Merilä 2009; Qanbari et al. 2010a). Linkage disequilibrium has been suggested as a robust tool to use especially in the presence of high density SNP data. The extent and distribution of LD in the genome has been exploited to shed light on the origin of domestication of domestic animals, demographic history and facilitate an understanding of

breeds' relatedness (Du et al. 2007; Amaral et al. 2008; De Roos et al. 2008; Corbin et al. 2010). Evolutionary biology, population, and quantitative genetics have benefited from LD measures to study evolutionary mechanisms in natural populations. Such examples include the "out of Africa human theory" that was investigated using LD analyzed for humans sampled from different points of ancestral origin using an array 22 million SNPs (Tenesa et al. (2007). The ancestral breeding population size of commercial pigs that was analyzed using 371 SNPs in three genomic regions (Amaral et al. 2008) and the origin of Hanwoo Korean cattle domestication (Lee et al. 2011) investigated using LD of 4.525 SNPs.

Haplotype blocks are defined as two or more alleles at a locus that are bound to be linked together at close proximity and be transferred together (Zhao et al. 2003; Crawford & Nickerson 2005). Haplo-blocks are more informative than single alleles in a genome because markers spanning a haplo-block can be useful in gene mapping for breeding programs and disease variants detection (Tishkoff & Verrelli 2003). Haplotype sharing between populations can indicate transferability of genetic parameters between populations, such as QTLs (Megens et al. 2009). Haplotypes can be useful in association mapping of disease variants and in association studies for phenotypes of economic interest. The investigation of haplo-blocks in chicken populations has mainly been limited to selected genomic regions on selected chromosomes. In studies using dense SNP marker sets, haplotype analysis has shown that different breeds have different haplo-block structures, which was considered to be due to differences in breed/population demographic history (Andreescu et al. 2007; Megens et al. 2009).

There is no information on the extent of LD and haplo-block structure of Southern African village chicken populations. Such information is considered important in facilitating the understanding of the evolutionary history and population parameters of these chickens. The Illumina iSelect chicken 60K SNP chip was developed by using SNPs from four commercial lines of the Brown and White broilers and egg layers. The chip has over 53 000 SNPs that have found application in population and quantitative genetics studies (Qanbari et al. 2010a; Wragg et al. 2012). The investigation of level of LD using this SNP panel

has revealed demographic history, effective population size and level of genetic erosion, facilitating the knowledge of the level of genetic diversity in chicken populations studied (Qanbari et al. 2010a).

In this study, the Illumina iSelect chicken 60K SNP chip was used to estimate linkage disequilibrium, effective population size, haplo-block structure and haplo-block sharing of extensively raised chickens from Zimbabwe, Malawi and South Africa. LD based estimates of N_e, were computed which provided an understanding of whether the extensively raised chicken population are at risk of extinction, or went through population bottlenecks that might result in increases of inbreeding and loss of genetic diversity. The study aimed to provide additional information on genetic diversity and population structure of village chickens to that given by previous studies and provides an understanding of these important animal genetic resources.

1.2 HYPOTHESIS

The study hypothesized that extensively raised chicken populations have evolved over time to adapt to specific and harsh environmental conditions and genetic diversity and genetic structure can be purely defined by agro-ecological zone where these animals are raised.

1.3 GENERAL OBJECTIVES

The overall objective of this study was to investigate the population genetic parameters of extensively raised chicken populations of Southern Africa using genome-wide SNP data.

The specific objectives of the study were to:

a) To investigate the existing Linkage disequilibrium in extensively raised chicken populations using genome-wide SNP data.
b) To estimate effective population size in extensively raised village chicken populations using an LD-based method.

c) To investigate LD bound haplo-blocks structure, haplo-block sharing, and indications of genetic changes supporting adaptation between Zimbabwean, South African, and Malawian extensively raised chicken populations.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Smallholder communal farmers in Africa mainly raise indigenous village chickens. These chickens play a major role in the socio-economic aspects of smallholder communities as well as being an integral part of their nutrition. Generally, households within rural communities raise these chickens with minimal management interventions (FAO 1995). The management varies between households depending on the demographic and socio-economic status (Ekue et al. 2002). In the smallholder farming system, majority of women are responsible for managing chickens with less farming skills under poorly resourced production systems. Village chickens scavenge in their surrounding environment with little supplements from the farmers. Available feed material for village chickens includes insects, exposed garbage, vegetables, and weeds from the gardens, which they convert to meet their growth, reproduction and maintenance needs (Ekue et al. 2002).

In addition, these chickens face daily challenges, such as exposure to parasitic and bacterial pathogens as well as harsh environmental conditions. Pathogenic diseases are the major constrain in these production systems and almost ~75% of the chicken population can be wiped-out by Newcastle and others diseases such as fowl pox, Gumboro and respiratory diseases. The poor sheltering also results in vulnerability of these animals to predators and to extreme environmental conditions. These factors contribute towards low productivity and survivability of village chickens under the extensive system (Ekue et al. 2006; Abdelqader et al. 2007). Studies have been conducted, characterizing and assessing existing breeds and the current production practices in Southern Africa (Muchadeyi et al. 2007b; Mtileni et al. 2009; Muchadeyi et al. 2009; Mtileni et al. 2010). However, information is still scarce particularly on the underlying mechanism that shaped the phenotypic and genetic diversity observed is these village chicken populations.

The understanding of the genetic architecture of these populations is a vital step for future conservation and genetic improvement programs. An understanding of the evolutionary mechanism that shaped the genetic structures of these populations is also vital. Effective population size (N_e) is an ideal population size that reflects evolutionary processes that shaped a particular population and determines its risk to extinction (Wang 2005). Studies on population genetic structures were previously limited by the sparse genome coverage of the microsatellite and other autosomal markers available at that time and to the use of mtDNA D-loop sequences which were mostly limited to the use few base pairs of biallelic markers (Cuc et al. 2006; Muchadeyi et al. 2006; Granevitze et al. 2007; Muchadeyi et al. 2008; Mtileni et al. 2011a; Mtileni et al. 2011b; Mwacharo et al. 2011).

The first draft of the chicken genome paved way to the use of genome wide high-density SNP markers and powerful statistical tools (Hillier et al. 2004). The whole genome SNP arrays for chickens have been used in determining the demographic history of layers (Qanbari et al. 2010a) and in understanding the genetic structure and genetics of Mendelian traits in traditional and commercial chicken breeds (Wragg et al. 2012). These whole genome SNP chips have the potential to unravel the needed genetic information in extensively raised chicken populations. By applying linkage disequilibrium analysis, one will be able to infer on the demographic and evolutionary processes at play in these populations. Genome wide LD profile can be analyzed to reflect the population history, the breeding system and the pattern of geographic subdivision in populations (Slatkin 2008). The availability of whole genome SNP and sequence data has played a crucial role towards the use of LD in the understanding of the evolutionary history of populations and in estimating N_e in the absence of records and pedigree information.

This review looks at the challenges associated with communal village chicken production systems, measures of effective population size and factors affecting N_e. The review also looks at linkage disequilibrium and the factors that affect it as well as how it can be used to estimate population demographic parameters.

2.2 VILLAGE CHICKEN PRODUCTION SYSTEMS

Chicken production in the village communities is mainly under the extensive or scavenging system whereby the chickens are free to move around and search for feed in the surrounding environment (Ekue et al. 2006; Abubakar et al. 2007). In this system, the environmental conditions play a crucial role on what feed is available for these animals as well as what disease pathogens and climatic stress they are exposed to. The genetic composition of these animals determines how the populations cope with the production challenges. Genotypes that are adapted to the environment and are able to utilize the available feed and those that can sustain, tolerate or resist diseases will be able to survive in such environments. The environmental conditions play a selective role (natural selection) in the village chicken population by removing animals that cannot utilize poor quality feeds and those that are susceptible to diseases (Ekue et al. 2006; Mtileni et al. 2009). Factors that affect production in this system vary across the seasons of the year and between villages and farming systems.

2.2.1 Sheltering

Poor housing contributes to the low productivity in chicken production system (Abdelqader et al. 2007; Abubakar et al. 2007). The extensively raised chickens are constantly exposed to predators (Moges et al. 2010). In most village chicken production systems, chickens find shelter in bushes, homestead verandah, trees, and sometimes they will be housed in housing structures meant for humans (e.g. kitchens and spare living rooms). In cases where farmers built the barns, locally available but sometimes substandard materials such as tree logs or wood, plastic, iron sheets, and cardboard boxes are used. The building of chicken housing is mainly dependent on men with assistance from children and is usually done in summer during heavy thunderstorms and during extreme winter condition (Ekue et al. 2006; Moges et al. 2010; Olwande et al. 2010). Chicks at the ages of one to four months are housed in boxes or baskets in household kitchens (FAO 1995; Ekue et al. 2006; Okeno et al. 2012). Substandard shelter poses a selection pressure on these chicken populations and animals that cannot survive or escape predators and those that cannot tolerate the harsh unprotected environment get to be eliminated in the system. Only

animals with the ability to withstand these challenges get to survive and contribute to the next generations.

2.2.2 Feeding

The feed in most village communities is mainly controlled by season and other farming practices (Okeno et al. 2012). During the wet seasons, the environment has plenty to offer to these animals from insects, weeds, field crops and vegetable gardens. Supplements from farmers include grains such as maize, rice, and barley (FAO 1995; Mapiye et al. 2008). Other supplements may come from the household leftovers such as porridge, old bread, and vegetable pills (FAO 1995; Ekue et al. 2002; Muchadeyi et al. 2005; Mtileni et al. 2009). In such environments, the genetic profiles of the animals tend to be biased towards those genotypes that can survive and utilize poor quality and limited feed supplies. Chicken body sizes and structure in these systems tend to be determined by the ability of animals to convert the available low quality feed resources into high quality protein and energy for growth, maintenance and reproduction. Genotypes associated with slow growth rates and lower mature body sizes seem to be supported by this low-input system (Muchadeyi et al. 2009).

2.2.3 Chicken Health

The prevalence of diseases in village chicken communities is one of the major constrains that contribute towards the low production (Olwande et al. 2010). The pathogenic and parasitic environment these chickens are exposed to leaves them vulnerable to infections (Mwale et al. 2005). Studies have been conducted on village chicken populations to investigate mortality caused by pathogens across Africa. Diseases like Newcastle, fowl pox, Infectious basal disease (Gumboro), intestinal and ectoparasitic infections have been classified as the most prevalent causes of chicken mortality (Ekue et al. 2002; Illango et al. 2002). Newcastle disease is the most pathogenic, eradicating about 80% of chickens during outbreaks (FAO 1995). Effects of Newcastle and other diseases such as those caused by parasite vary depending with seasons (Mapiye et al. 2008; Mwale et al. 2005).

Diseases effects and management is not well documented in village chicken production systems (FAO 1995). Farmers have different ways of managing chicken diseases depending on their socio-economic status (Ekue et al. 2002; Illango et al. 2002). The uses of ethno-veterinary medicines such as aloe species (Mwale et al. 2005) and the culling of the affected animal are common ways by which farmers manage infections under such systems (Muchadeyi et al. 2009). Farmers do not keep records of animal performance, mortality and its cause as well as the disease symptoms in their flocks. This information gets to be investigated by relevant public officers such as extension officials and non-governmental organizations during outbreaks. The village chicken production system has no bio-security systems in place to manage outbreaks (Muchadeyi et al. 2009). Chickens are constantly exposed to the environment and other wild animals (Ekue et al. 2002), and within flocks different age groups are raised together. Such management practices result in compromised bio-security and complicates disease management and prevention programs (Muchadeyi et al. 2009). The animal health care service varies depending on a number of factors such as costs of veterinary interventions and the farmer's ability to pay for services (Gondwe & Willony 2007). Inaccurate diagnosis usually due to inadequate information from the farmers who do not keep records also makes the management of diseases difficult (Muchadeyi et al. 2009). Those farmers who afford veterinary services normally use broad-spectrum antibiotic drugs to treat a wide range of diseases (Ekue et al. 2002; Illango et al. 2002).

Regardless of the high exposure to diseases and disease causing pathogens, it has been found that generally village chicken flocks tend to survive and in some instances are even able to produce optimally to meet farmer's subsistence needs (Abdelqader et al. 2009). The observation that some animals survive during disease outbreaks, without having received any form of treatment raise some interest on the disease resistance profiles of these animals and in understanding the physiological and genetic differences between animals that will be associated with variation in response to infections. Similar to other production challenges, diseases in these village chicken farming systems acts as natural selection forces that would structure the genetics of the population over time. (Ekue et al. 2002; Illango et al. 2002). It has

been hypothesized by other researchers that such selection pressures would result in unique and advantageous genotypes being selected and promoted in such populations (Hall 2004; Muchadeyi et al. 2009). Genetic differences associated with such advantageous traits can be investigated and be manipulated to improve resistance of chicken populations to disease infections.

2.2.4 Flock sizes, composition and mating systems

Flock sizes of smallholder farming system range from 5 to 20 animals with the mixture of different phenotypic traits and production purposes (Illango et al. 2002; Mtileni et al. 2009; Moges et al. 2010). The number of breeding animals is biased towards hens meaning more than 3 hens can be kept on a farm, where one or two cocks are kept based on performance, behavior and phenotypic traits (Olwande et al. 2010). A hen: cock ratio of 3:1 has been reported in most village chicken production systems (Mtileni et al. 2009; Moges et al. 2010), whilst Abdelqader et al. (2007) have reported a different hen: cock sex ratio of 6:1 in Jordan chicken populations. Farmers preferred phenotypic features are used to select breeding animals (Muchadeyi et al. 2009; Moges et al. 2010). It has however been observed that farmers keep breeding animals on the farm for long periods of time in order to maintain their flocks (Muchadeyi et al. 2009). Breeding animals are generally selected for block body size, mothering, and fertility (Muchadeyi et al. 2009). Keeping breeding animals for extended periods of time usually results in overlapping of generations where parent mate with offspring's, and vice versa (Ekue et al. 2002; Illango et al. 2002). Indiscriminate mating of animals from different household within villages also takes place when chickens are out scavenging for feed (Muchadeyi et al. 2007b).

2.3 EFFECTIVE POPULATION SIZE, INBREEDING AND SUSTAINABILITY OF A POPULATION

Effective population size is defined as the ideal population size that experience the same amount of inbreeding, allele frequency variation, or loss of heterozygosity as the natural population under question (Hedrick 2004; Tenesa et al. 2007). Effective population size is used to investigate factors that have an

impact on changing the genetic composition of a population over a period of time. Effective population size determines the rate of loss of genetic variation, inbreeding and rate of fixation of deleterious allele (Hedrick 2004). One can make use of N_e to investigate the genetic properties of any given population (Schwartz et al. 1998). Populations with low genetic variation are associated with high levels of homozygosity particularly at deleterious loci that often lead to reduce performance in reproduction growth and in the survival rate of populations (Hedrick 2004). Inbreeding is defined as mating of related individuals resulting in increased homozygosity (Wang 2005). Reduced number of breeding individuals results in mating of closely related animals and to the loss of genetic variation in any given population. Individuals that originate from small populations turn to have high levels of inbreeding. Effective population size has been used as an indicator of the risk of populations to extinction (Lee et al. 2011). Reduced N_e is usually observed in livestock that have been exposed to intense artificial or natural selection as it has been reported for Hanwoo Korean cattle (Lee et al. 2011), Holstein Friesian (Kim et al. 2009) beef cattle (Lu et al. 2012), and pig breeds (Uimari et al. 2011). The above studies have reported the importance of knowing and managing the levels of effective population size for the sustainability of genetic diversity of a breed or population over time.

2.4 FACTORS INFLUENCING EFFECTIVE POPULATION SIZE OF A POPULATION

Effective population is affected by various demographic factors. These are variation of sex ratio of breeding individuals, variance in reproduction rate and number of breeding individuals over generations (Hedrick 2004; Wang 2005).

2.4.1 Sex Ratio

The uneven number of male and female breeding animals has an impact on the genetic contribution of breeding individuals into the next generation (Hedrick 2004). In a population where there are only few males or female that contributes in reproduction system, there are high chances of reduced N_e. This is so because it is not guaranteed that all males or female will get equal chance of reproduction in a population

with disrupted sex ratio. In undisrupted populations, an increased N_e is observed as a result of equal male and female ratios and individuals having equal chances of reproducing (Wang 2005).

In a natural population unequal sex ratio may be as results of catastrophic climatic evens. Village chicken populations are exposed to harsh environmental condition where mostly the survival rate of chicks is reduced. A smallholder farmer keeps one cock and 3 to 6 hens and mostly the flock sizes range from 5 to 20 animals (Mtileni et al. 2009). The effective population size of village chicken population size is not well studied because of lack of pedigree records including number of breeding males and females.

2.4.2 Progeny variation

The number of progeny per parent varies in a population, which can be as a result of random, genetic, fertility, environment and accidental factors contributing to change on effective population size over time (Falconer & Mackay 1996). The fluctuation of number of reproductive offspring due inbred, harsh environmental conditions and disease infestation in a scavenging system may play a role in genetic diversity and genetic structure of village chicken populations (Abubakar et al. 2007).

2.4.3 Variation in natural factors over time

Populations can experience different evolutionary events that might shape population structure over time. Disasters, climatic conditions, diseases and other factors varies in occurrence with time, that mean population size after and before those forces won't be the same over time. The effective population size will differ in different generations depending on what evolutionary events the population went through that shaped the breeding population size. In most case for instance village chickens are raised under poor infrastructure which exposes these animals to all disastrous climatic condition in which case only those that can escape survive (Ekue et al. 2002). Diseases out break such as Newcastle diseases as one of the eradicative diseases in avian species it eliminates certain genotypes and animals resulting to reduced gene pool for that population. The hot-wet season and hot-dry seasons have been reported as the times where

there are major losses of village chickens due various factors (Abdelqader et al. 2007; Mapiye et al. 2008).

2.4.4 Inbreeding

Inbreeding decreases the gene pool and results in an increase in homozygosity that can be lethal when there is an increase in deleterious alleles and reduction of fertility (Keller & Waller 2002; Hedrick 2004). Effective population size can be affected typically by inbreeding depression reducing fertility in the population, where there will be few individuals that can reproduce to contribute to the next generation. Populations that have experienced population size reduction from environmental disasters and outbreak of diseases can foster inbreeding. Inbreeding in a population with effective population size of less than 100 is most likely to increases number of individuals sharing genotypes from common ancestry resulting to loss of population fitness (Keller & Waller 2002). The flock sizes of village chicken range from 5 to 30 animals and with limited gene flow can contribute to mating of relatives in the same flock. The mating system is not monitored and managed properly in such a way that individuals can mate with offspring's or close relatives within a flock (Moges et al. 2010; Uimari et al. 2011).

2.5 METHODS OF MEASURING EFFECTIVE POPULATION

Different methods to estimate N_e including use of evolutionary history, demographic, genetic data, and population census and pedigree data have been proposed and evolved in the past (Wang 2005). Demographic, pedigree and population census methods are however, challenging in extensively raised populations because of absence of pedigree data and the intermixing of flocks. For extensively raised and wild populations, molecular methods are the most reliable as they do not rely on availability of pedigree data (Wang 2005). Molecular methods include temporal and LD-based approaches (Wang 2005; Lee et al. 2011). The two later approaches can assist in detecting the changes in effective population size that might be associated with allele frequency variance and loss of heterozygosity in populations (Wang 2005). There are few studies that have been conducted to compare the reliability of these concepts. Cervantes et al. (2011) proposed that N_e estimates varies according to the approach used and that research should have an understanding of the strengths and limitations of each method.

Temporal methods measure changes in allele frequency (F) over time (Hedrick 2004). These changes would be as a result of genetic drift, assuming that all other force such as mutation, migration and selection are kept constant (Wang 2005). Changes in allele frequency (F) over time (t) will therefore be a function of N_e, which will be estimated by taking two samples from the same population at random and at different time intervals. The sampling assumes that the population has discrete non-overlapping generations (Waples 1989), an assumption that's often violated in free-range chicken populations. The calculated standardized variance in the temporal change of allele frequency will therefore reflects on genetic drift, which will be inversely proportional to effective population size (Waples 1989; Wang 2005).

In natural populations the estimation of N_e using a temporal method can be challenging requiring the sampling structure to be designed and account for age/sex structure and to assume discreet generation interval. In reality, breeding animals are kept for longer extended periods of time resulting in overlapping generations, which is a one of the challenges of using temporal methods of estimating N_e in village

chicken populations (Wang 2005). In natural population gene flow or admixture of population cannot be easily eliminated so accountability of migration in temporal method is also necessary (Wang 2005). Gene flow introduces new alleles thereby changing the allele frequency of the population (Hedrick 2004; Wang 2005). In village chickens the level of gene flow is not understood and lack of pedigree record makes it difficult to understand the age and relatedness of individuals. This could make it difficult to use temporal methods for the estimation of N_e.

Linkage disequilibrium is the nonrandom association of alleles at two or more loci (Reich et al. 2001; Hedrick 2004). LD is used to determine the evolutionary forces that contributes to the genetic variation of a population under consideration and is now well implemented in many population genetics studies (Hedrick 2004; Andreescu et al. 2007; Slatkin 2008). LD is also used for gene mapping (Slatkin 2008). LD provides information on evolutionary history and population demography. The distribution and extent of LD across a genome have been used to assess the current and ancestral genetic forces that structure current populations (Tenesa et al. 2007). In humans, LD measures have indicated the origin and expansion of humans out of Africa into the rest of the world. In domesticated animals the selection measures for breed development were found to cause an elevated LD value that suggested reduced effective population of the founder population (Hill & Robertson 1968; Backström et al. 2006; Du et al. 2007; Thevenon et al. 2007; Khatkar et al. 2008; Lu et al. 2012). Previous work has also indicated that LD varies between populations and chromosomes (Andreescu et al. 2007; Prasad et al. 2008; Megens et al. 2009).

Linkage disequilibrium is one of the most common methods used to estimate N_e in the absent of population pedigree (Uimari et al. 2010). It also the most convenient as it requires only a single sampling comparing to temporal methods that requires that a population is sampled more than once to compare changes in allele frequencies (Hedrick 2004). Linkage disequilibrium as a squared correlation coefficient of marker pair between two loci is a function of distance between makers where recombination rate decline with a decrease of marker distance. Historic effective population size of N_e and recombination rate

have been reviewed and modified to take into account of the species chromosome size and mutation rates within populations (Hayes et al. 2003; Tenesa et al. 2007; Corbin et al. 2012). This has increased the precision of estimation of N_e using LD data from natural population. The LD of marker pairs at a close distance reflects a past N_e and LD of marker pair at two different loci far apart reflect recent N_e of the population of interest (Hayes et al. 2003). Theoretically recombination events will be high between markers that are far apart than those markers that are close to each other.

There are two commonly used measures of LD (Slatkin 2008). The measure D´ is a standardized LD value which is dependent on allele frequencies (Slatkin 2008). D´ is calculated from D which is the difference between the frequency of gametes carrying allele A and B (p_{AB}) at two Loci and the product of the frequencies of those alleles (p_A and p_B), $D_{AB} = p_{AB}\text{-}p_A p_B$ (Hedrick 2004).

The alternative (r^2) LD measure is defined as squared correlation co-efficient between two loci (Hedrick 2004; Slatkin 2008). It can be used for multiple pairs of loci, as it is less dependent on allele frequencies (Lee et al. 2011). The r^2 has been used over D´ because it takes into account recombination rate, occurrence of mutation, and the effect of population sample size and is considered to be more precise.

Linkage disequilibrium as a statistical measure is computed under certain assumptions that are applicable to different populations (Slatkin 2008). Under the assumption of infinite size with random mating, no genetic drift and no selection LD should approach zero (Corbin et al. 2010). The consideration of these factors when calculating LD is vital as the estimation can be bias (Corbin et al. 2010). These assumptions however, do not hold in the natural populations that are of finite size (Tenesa et al. 2007; Slatkin 2008). Such finite populations are bound to experience genetic drift as well as non-random mating (Hayes et al. 2003; Slatkin 2008; Corbin et al. 2010). The populations are confined to the boundaries defined by farmers and are therefore sub-structured with some controlled gene flow between subpopulations.

New alleles introduced by mutation are generally considered to have a very low impact on LD since the rate of mutation is very low (Ardlie et al. 2002). In practice, the product of mutation is driven by other factors such as genetic drift and selection, which will determine its influence on LD. These factors can keep or eliminate new mutations (alleles) in a population that would results in its frequencies being either elevated or reduced in a population. A reduction of LD in a population can be observed if the new gametes are maintained and their frequency increases (Hedrick 2004). Avian species have a different karyotype from most species in that they have macro-, intermediate and micro-chromosome that have different genetic properties. Micro-chromosomes have high recombination rate compared to macro-chromosome (Megens et al. 2010) resulting in lower LD levels.

The decay of LD is affected by a degree of inbreeding in a finite population (Rao et al. 2008). Mating closely related individuals assumed to have linked ancestral loci may results in elevated LD. Linkage disequilibrium values will provide information about the mating system of village chicken population since the level of inbreeding is proportional to reduced effective population size (Corbin et al. 2010). Genetic drift in village chicken population can have an impact on changing the level of LD as it can eliminate or increase the frequency of an allele (Wang 2005; Corbin et al. 2010). Genetic drift can be in favour of increased of homozygosity of which its effects will be also similar to inbreeding (Hayes et al. 2003). It has been observed that village chicken population have high genetic diversity and are not sub-structured to promote inbreeding (Cuc et al. 2006; Hassen et al. 2009).

New alleles can be introduced into a population through gene flow (Falconer & Mackay 1996). Gene flow or mixing of gametes from different population also has an effect on LD (Li & Merilä 2011). The mixing up of two populations carrying different allele at different frequencies will usually result in a new gene pool characterized by higher heterozygosity levels over time. This increased heterozygosity level will result in reduced levels of LD. The extent of reduction in LD will depend on the origin of the chicken subpopulation in such a way that greatly diverse subpopulations will result in high genetic variation and lower LD values than when genes are exchanged between closely related populations (Li & Merilä 2009).

Linkage disequilibrium has been well investigated and reviewed as a method of choice for the understanding of evolutionary history and genetic processes that contributes to the existence of natural populations. Slatkin (2008) observed that LD reflects on the population background in terms of its genomic architecture, the mating system, and its ecological distribution. Furthermore, the estimate of LD for each locus reflects on the effects of natural selection, genetic drift, and inbreeding in a population (Slatkin 2008). The above factors are important to consider when estimating LD and corrective measures have to be made to avoid biases. Normally assumptions are made on the influence of such factors on LD (Wang 2005; Corbin et al. 2010). The use of LD and the assumptions made will depend on the research questions to be addressed.

Linkage disequilibrium has been used to reveal the demographic history including trends in effective population sizes of egg laying commercial chicken populations (Qanbari et al. 2010a). In other studies in a number of species, LD have been used to construct LD maps (Nordborg & Tavare 2002; Meadows et al. 2008; Lee et al. 2011), fine mapping quantitative traits locus (QTL) and population parameters (Tenesa et al. 2007; De Roos et al. 2008), analysis of haplotype diversity (Amaral et al. 2008) and mating systems (Uimari & Tapio 2010; Lee et al. 2011). Such findings are of great interest as they provide a new platform of understanding the dynamics of the populations and in managing population biodiversity. LD based analysis have found use in population studies where the pedigree data is not readily available. Wild populations and some extensively raised livestock are such good examples. The availability of whole genome high density SNP data has even increased the use of LD based estimates of Ne and other population parameters particularly in humans and domestic livestock without pedigree data.

2.6 HAPLOTYPE ANALYSIS

The availability of genome wide sequences for many livestock species has made it possible to intensify SNPs across the genome to investigate population parameters and associate disease variant traits (Zhao et al. 2003). The possibility of analyzing markers independently has been reviewed to be less informative than a combination of markers particularly in determining the evolutionary history and possible disease

causing variants of a population (Ardlie et al. 2002). It has been observed that linked markers on the same chromosome can be non-randomly associated with each other and measured by a squared correlation coefficient denoted by r^2 (Hedrick 2004; Tenesa et al. 2007). The patterns of LD are considered as confined borders of a haplotypes structure on a chromosomal region, where we can infer demographic history of a population (Zhao et al. 2003; Tishkoff et al. 2003). Haplotype blocks have been defined by patterns of LD created by recombination hot and cold spot (Ardlie et al. 2002; Tishkoff et al. 2003; Megens et al. 2009). In human studies haplo-blocks have been used to investigate association of mutant variants to human disorders (Slatkin 2008). These are done using case and control designs of individual with/without a disorder and associate them to the haplo-blocks observed (Zhao et al. 2003; Slatkin 2008). In livestock studies haplo-blocks has been characterized to understand their diversity within and between breeds and to infer population demography (Qanbari et al. 2010b). Different haplo-blocks of different lengths have been observed in many population genetic studies where they focused on different chromosomal regions with different marker panels. The focus on different chromosomal regions makes it difficult to compare haplo-block structure in livestock species from different studies. There has been great variation in the average lengths of haplo-blocks observed. Previous studies suggested that, the theory that longer haplo-blocks have low LD and vice versa do not always hold in different populations (Qanbari et al. 2010b). Longer haplo-blocks have been observed in populations that went through bottleneck event and sub-structuring suggestive of genetic drift and inbreeding effects (Qanbari et al. 2010b).

2.7 LD ESTIMATION AND HAPLOTYPE ANALYSIS IN THE GENOMICS ERA

The earlier detection level of LD on chicken genome was revealed by the use of microsatellite markers that were sparsely distributed across the genome (Heifetz et al. 2005). Microsatellite markers were used for the estimation of LD for specific chromosomal region of interest. Different statistical methods including X^2 test statistics have been used in calculating LD between the pairs of microsatellite markers (Heifetz et al. 2005; Meadows et al. 2008). After the completion of the first chicken genome in 2004

(Hillier 2004), studies have been conducted to reveal polymorphisms in the chicken genome, which consequently led to the development of genome-wide high-density SNPs arrays.

The 60k SNP bead chip was made available for the public and number of studies has been conducted to understand the population parameters underlying the evolutionary traits of chickens. The extent and distribution of LD have revealed useful LD of 0.3 in cattle (Qanbari et al. 2010b; Sargolzaei et al. 2012), pigs, (Uimari & Tapio 2011) and wild dogs (Gray et al. 2009) for the association of phenotypic trait with quantitative traits locus (Rao et al. 2008). The extent of LD varied between certain cattle (Sargolzaei et al. 2012), sheep (Meadows et al. 2008), dogs (Gray et al. 2009) and pig breeds (Uimari & Tapio 2011) due to evolutionary and population demographic difference between and within breeds. In commercial chicken lines, LD extended over a long distances compared to that of wild fowl (Wragg et al. 2012). The investigation of LD became a success after the development of high-density genome wide SNP chips in most livestock species. The estimation of LD has made it possible to infer the population demography in chickens (Andreescu et al. 2007; Qanbari et al. 2010a), dairy cattle (Sargolzaei et al. 2008) and pigs (Uimari & Tapio 2011). Linkage disequilibrium has also been used to estimate trends in effective population size in some livestock and domesticated species (Tishkoff & Verrelli 2003; Zhao et al. 2003; Kim et al. 2006; Amaral et al. 2008; Megens et al. 2009).

2.8 CONCLUSION

The extensive smallholder production systems present some challenges on the survivability and productivity of village chicken populations in Southern Africa and most developing countries. Such challenges that range from poor and low quality nutrition, exposure to diseases pathogens, reduced flock sizes and unorganized mating systems have a negative impact on biodiversity and sustainability of these important animal genetic resources. The small flock sizes, reduced gene flow between geographically isolated populations and lack of performance records could lead to isolated genetic change/evolution resulting to localized inbred sub-populations. The estimation of trends in effective population size in these animals will provide insight on the number of breeding individuals, loss of genetic variation and facilitate the knowledge whether these animals are at risk of extinction or not. Such information is important for future effective breeding and conservation programs. The knowledge of changes in effective population size over time will assist on understanding the demographic history and evolutionary mechanism that have played a role in shaping up the genetic structure of extensively raised Southern African chicken populations.

Genome wide SNP data and LD analysis present opportunities of understanding effective population sizes and other demographic parameters in populations such as village chickens where pedigree records are missing. The Illumina chicken iSelect 60K SNP chip have proven to be useful in analyzing background LD, haplo-block structure and estimate trends in effective population sizes as well as identify regions under selection and conferring to Mendelian traits.

2.9 POTENTIAL OUTPUT

The use of genome wide high density SNPs will be a new application in Southern African village chickens and is expected to provide a new knowledge on the genetic composition and evolutionary factors at play in these populations. These populations have adapted and survived low input communal farming characterized by extreme environmental conditions, exposure to disease pathogens, and low and

fluctuating nutritional supplies. This study will look at the LD and haplo-block structure and provide information that can be used to better understand the village chicken population structure; its population diversity and insight into the evolutionary factors that contribute significantly in these populations. This study will also provide baseline information to further characterize genes that might be under selection and facilitate inferences on the local production pressures acting on the studied populations.

2.10 REFERENCES

Abdelqader A., Wollny C. & Gauly M. (2007) Characterization of local chicken production systems and their potential under different levels of management practice in Jordan. *Tropical Animal Health and Production* **39**, 155-64.

Amaral A.J., Megens H.-J., Crooijmans R.P., Heuven H.C. & Groenen M.A. (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* **179**, 569-79.

Andreescu C., Avendano S., Brown S.R., Hassen A., Lamont S.J. & Dekkers J.C. (2007) Linkage disequilibrium in related breeding lines of chickens. *Genetics* **177**, 2161-9.

Ardlie K.G., Kruglyak L. & Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**, 299-309.

Backström N., Qvarnström A., Gustafsson L. & Ellegren H. (2006) Levels of linkage disequilibrium in a wild bird population. *Biology Letters* **2**, 435-8.

Cervantes I., Goyache F., Molina A., Valera M. & Gutiérrez J.P. (2011) Estimation of effective population size from the rate of coancestry in pedigreed populations. *Journal of Animal Breeding and Genetics* **128**, 56-63.

Corbin L., Blott S., Swinburne J., Vaudin M., Bishop S. & Woolliams J. (2010) Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics* **41**, 8-15.

Corbin L., Liu A., Bishop S. & Woolliams J. (2012) Estimation of historical effective population size using linkage disequilibria with marker data. *Journal of Animal Breeding and Genetics* **129**, 257-70.

Crawford D.C. & Nickerson D.A. (2005) Definition and clinical importance of haplotypes. *Annual Review Medicine* **56**, 303-20.

Cuc N.T.K., Muchadeyi F., Baulain U., Eding H., Weigend S. and Wollny C. (2006) An assessment of genetic diversity of Vietnamese H‚Äômong chickens. *International Journal of Poultry Science* **5**, 912-20.

De Roos A., Hayes B.J., Spelman R. & Goddard M.E. (2008) Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics* **179**, 1503-12.

Du F.-X., Clutter A.C. & Lohuis M.M. (2007) Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences* **3**, 166.

Ekue F.N. P.K.D., Mafeni M.J., Nfi A.N. & Njoya J. (2002) Survey of the traditional poultry production system in the Bamenda area, Cameroon. In: *Characteristics and parameters of family poultry production in Africa*. FAO/IAEA, Vienna, Austria, p. 15

Ekue F., Pone D., Nfi J.M.A. & Njoya J. (2006) The effect of vaccination against Newcastle disease and feed supplementation on production in village chicken in Bamenda area of Cameroon. In: *Improving farmyard poultry production in Africa: interventions and their economic assessment: proceedings of a final research coordination meeting.* International Atomic Energy Agency. p. 51.

Falconer D.S. & Mackay T.F.C. (1996) Qauntitative Genetics. Pearson Education Limited, Harlow, UK.

FAO (1995) Global project for the maintenance of domestic animal genetic diversity (MoDAD). Draft project formulation report. Food and Agriculture Organization of United Nations, Rome.

Flury C., Tapio M., Sonstegard T., Drögemüller C., Leeb T., Simianer H., Hanotte O. & Rieder S. (2010) Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *Journal of Animal Breeding and Genetics* **127**, 339-47.

Gondwe T. & Wollny C. (2007) Local chicken production system in Malawi: Household flock structure, dynamics, management and health. *Tropical Animal Health and Production* **39**, 103-13.

Granevitze Z., Hillel J., Chen G., Cuc N., Feldman M., Eding H. & Weigend S. (2007) Genetic diversity within chicken populations from different continents and management histories. *Animal Genetics* **38**, 576-83.

Gray M.M., Granka J.M., Bustamante C.D., Sutter N.B., Boyko A.R., Zhu L., Ostrander E.A. & Wayne R.K. (2009) Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics* **181**, 1493-505.

Hall S.J. (2004) Livestock biodiversity: genetic resources for the farming of the future. Cambridge University Press.

Hassen H., Neser F., De Kock A. & van Marle-Köster E. (2009) Study on the genetic diversity of native chickens in northwest Ethiopia using microsatellite markers. *African Journal of Biotechnology* **8(7),** 1347-1353.

Hayes B.J., Visscher P.M., McPartlan H.C. & Goddard M.E. (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* **13**, 635-43.

Hedrick P.W. (2004) Genetics of Populations. Jones and Bartlett, Canada.

Heifetz E.M., Fulton J.E., O'Sullivan N., Zhao H., Dekkers J.C. & Soller M. (2005) Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* **171**, 1173-81.

Hill W. & Robertson A. (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226-31.

Hillier L.W., Miller W., Birney E., Warren W., Hardison R.C., Ponting C.P., Bork P., Burt D.W., Groenen M.A. & Delany M.E. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716.

Illango J. E.A., Olupot H., Mabonga J. (2002) Rural poultry production in two agro-ecological zone of Uganda. In: *Characteristics and parameters of family poultry production in Africa* (p. 117. FAO/IAEA, Vianna, Austria.

Kaya M. & Yildiz M.A. (2008) Genetic diversity among Turkish native chickens, Denizli and Gerze, estimated by microsatellite markers. *Biochemical Genetics* **46**, 480-91.

Keller L.F. & Waller D.M. (2002) Inbreeding effects in wild populations. *Trends in Ecology and Evolution* **17**, 230-41.

Khatkar M.S., Nicholas F.W., Collins A.R., Zenger K.R., Cavanagh J.A., Barris W., Schnabel R.D., Taylor J.F. & Raadsma H.W. (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics* **9**, 187.

Kim K.J., Lee H.-J., Park M.-H., Cha S.-H., Kim K.-S., Kim H.-T., Kimm K., Oh B. & Lee J.-Y. (2006) SNP identification, linkage disequilibrium, and haplotype analysis for a 200-kb genomic region in a Korean population. *Genomics* **88**, 535-40.

Kim E.S. & Kirkpatrick B. (2009) Linkage disequilibrium in the North American Holstein population. *Animal Genetics* **40**, 279-88.

Koenig S. & Simianer H. (2006) Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. *Livestock Science* **103**, 40-53.

Lee S., Cho Y., Lim D., Kim H., Choi B., Park H., Kim O., Kim S., Kim T. & Yoon D. (2011) Linkage disequilibrium and effective population size in hanwoo korean cattle. *Asian-Australasian Journal of Animal Sciences* **24**, 1660-5. DOUBLE CHECK JOURNAL NAME?????????

Li M. & Merilä J. (2009) Extensive linkage disequilibrium in a wild bird population. *Heredity* **104**, 600-10.

Li M.H. & Merilä J. (2011) Population differences in levels of linkage disequilibrium in the wild. *Molecular Ecology* **20**, 2916-28.

Lu D., Sargolzaei M., Kelly M., Changxi L., Vander Voort G., Wang Z., Plastow G., Moore S. & Miller S.P. (2012) Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Frontiers in Genetics* **3**, 10.

Mapiye C., Mwale M., Mupangwa J., Chimonyo M., Foti R. & Mutenje M. (2008) A Research Review of Village chicken production constraints and Opportunities in Zimbabwe. *Asian-Australasian Journal of Animal Science* **21**, 1680-8.

Meadows J.R., Chan E.K. & Kijas J.W. (2008) Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics* **9**, 61.

Megens H.-J., Crooijmans R.P., Bastiaansen J.W., Kerstens H.H., Coster A., Jalving R., Vereijken A., Silva P., Muir W.M. & Cheng H.H. (2009) Comparison of linkage disequilibrium and haplotype diversity on macro-and microchromosomes in chicken. *BMC Genetics* **10**, 86.

Moges F., Mellesse A. & Dessie T. (2010) Assessment of village chicken production system and evaluation of the productive and reproductive performance of local chicken ecotype in Bure district, North West Ethiopia.

Mtileni B., Muchadeyi F., Maiwashe A., Phitsane P., Halimani T., Chimonyo M. & Dzama K. (2009) Characterisation of production systems for indigenous chicken genetic resources of South Africa. *Applied Animal Husbandry and Rural Development* **2**, 18-22.

Mtileni B., Muchadeyi F., Maiwashe A., Groeneveld E., Groeneveld L., Dzama K. & Weigend S. (2010) Domestication of South African chicken genetic resources. In: *Proceedings International*.

Mtileni B., Muchadeyi F., Maiwashe A., Chimonyo M., Groeneveld E., Weigend S. & Dzama K. (2011a) Diversity and origin of South African chickens. *Poultry Science* **90**, 2189-94.

Mtileni B., Muchadeyi F., Maiwashe A., Groeneveld E., Groeneveld L., Dzama K. & Weigend S. (2011b) Genetic diversity and conservation of South African indigenous chicken populations. *Journal of Animal Breeding and Genetics* **128**, 209-18.

Muchadeyi F., Sibanda S., Kusina N., Kusina J. & Makuza S. (2005) Village chicken flock dynamics and the contribution of chickens to household livelihoods in a smallholder farming area in Zimbabwe. *Tropical Animal Health and Production* **37**, 333-44.

Muchadeyi F.C., Weigend S., Wollny C. & Makuza S. (2006) Use of Microsatellites and MtDNA to Assess Genetic Diversity Within and Between Zimbabwe Chicken Eco-types. *Bonn*.

Muchadeyi F., Eding H., Wollny C., Groeneveld E., Makuza S., Shamseldin R., Simianer H. & Weigend S. (2007a) Absence of population substructuring in Zimbabwe chicken ecotypes inferred using microsatellite analysis. *Animal Genetics* **38**, 332-9.

Muchadeyi F., Wollny C., Eding H., Weigend S., Makuza S. & Simianer H. (2007b) Variation in village chicken production systems among agro-ecological zones of Zimbabwe. *Tropical Animal Health and Production* **39**, 453-61.

Muchadeyi F., Eding H., Simianer H., Wollny C., Groeneveld E. & Weigend S. (2008) Mitochondrial DNA D-loop sequences suggest a Southeast Asian and Indian origin of Zimbabwean village chickens. *Animal Genetics* **39**, 615-22.

Muchadeyi F., Wollny C., Eding H., Weigend S. & Simianer H. (2009) Choice of breeding stock, preference of production traits and culling criteria of village chickens among Zimbabwe agro-ecological zones. *Tropical Animal Health and Production* **41**, 403-12.

Mwacharo J., Bjørnstad G., Mobegi V., Nomura K., Hanada H., Amano T., Jianlin H. & Hanotte O. (2011) Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Molecular phylogenetics and evolution* **58**, 374-82.

Mwale M., Bhebhe E., Chimonyo M. & Halimani T.E. (2005) Use of herbal plants in poultry health management in the Mushagashe small-scale commercial farming area in Zimbabwe. *International Journal of Applied Research in Veterinary Medicine* **3**, 163-70.

Nidup K., Penjor P.D., Gurung R.B., Arasta P. & Moran C. (2005) Read Origin And Genetic Diversity Of Indigenous Chickens of Bhutan As Determined By Mitochondrial DNA. *SAAR Journal of Agriculture* **3**, 69-89.

Okeno T.O., Kahi A.K. & Peters K.J. (2012) Characterization of indigenous chicken production systems in Kenya. *Tropical Animal Health and Production* **44**, 601-8.

Olwande P.O., Ogara W.O., Okuthe S.O., Muchemi G., Okoth E., Odindo M.O. & Adhiambo R.F. (2010) Assessing the productivity of indigenous chickens in an extensive management system in southern Nyanza, Kenya. *Tropical Animal Health and Production* **42**, 283-8.

Prasad A., Schnabel R., McKay S., Murdoch B., Stothard P., Kolbehdari D., Wang Z., Taylor J. & Moore S. (2008) Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Animal Genetics* **39**, 597-605.

Qanbari S., Hansen M., Weigend S., Preisinger R. & Simianer H. (2010a) Linkage disequilibrium reveals different demographic history in egg laying chickens. *BMC Genetics* **11**, 103.

Qanbari S., Pimentel E., Tetens J., Thaller G., Lichtner P., Sharifi A. & Simianer H. (2010b) The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics* **41**, 346-56.

Rao Y.S., Liang Y., Na Xia M., Shen X., Jun Du Y., Glong Luo C., Hua Nie Q., Zeng H. & Quan Zhang X. (2008) Extent of linkage disequilibrium in wild and domestic chicken populations. *Hereditas* **145**, 251-7.

Reich D.E., Cargill M., Bolk S., Ireland J., Sabeti P.C., Richter D.J., Lavery T., Kouyoumjian R., Farhadian S.F. & Ward R. (2001) Linkage disequilibrium in the human genome. *Nature* **411**, 199-204.

Sargolzaei M., Schenkel F., Jansen G. & Schaeffer L. (2008) Extent of linkage disequilibrium in Holstein cattle in North America. *Journal of Dairy Science* **91**, 2106-17.

Schwartz M.K., Tallmon D.A. & Luikart G. (1998) Review of DNA□based census and effective population size estimators. *Animal Conservation* **1**, 293-9.

Slatkin M. (2008) Linkage disequilibrium- understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**, 477-85.

Tenesa A., Navarro P., Hayes B.J., Duffy D.L., Clarke G.M., Goddard M.E. & Visscher P.M. (2007) Recent human effective population size estimated from linkage disequilibrium. *Genome Research* **17**, 520-6.

Thevenon S., Dayo G.-K., Sylla S., Sidibe I., Berthier D., Legros H., Boichard D., Eggen A. & Gautier M. (2007) The extent of linkage disequilibrium in a large cattle population of western Africa and its consequences for association studies. *Animal Genetics* **38**, 277-86.

Tishkoff S.A. & Verrelli B.C. (2003) Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. *Current opinion in Genetics and Development* **13**, 569-75.

Uimari P. & Tapio M. (2011) Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *Journal of Animal Science* **89**, 609-14.

van Marle-Köster E., Hefer C.A., Nel L.H. & Groenen M. (2009) Genetic diversity and population structure of locally adapted South African chicken lines: Implications for conservation. *South African Journal of Animal Science* **38**, 271-81.

Wang J. (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1395-409.

Waples R.S. (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379-91.

Wragg D., Mwacharo J., Alcalde J., Hocking P. & Hanotte O. (2012) Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. *Heredity* **109**, 6-18.

Zhao H., Pfeiffer R. & Gail M.H. (2003) Haplotype analysis in population genetics and association studies. *Pharmacogenomics* **4**, 171-8.

CHAPTER 3

THE EXTENT AND DISTRIBUTION OF LINKAGE DISEQUILIBRIUM IN EXTENSIVELY RAISED CHICKEN POPULATIONS OF SOUTHERN AFRICA

ABSTRACT

Linkage disequilibrium (LD) is an important source of information about historical events of recombination and allows inferences about genetic diversity and on genomic regions that have undergone selection. Linkage disequilibrium is equally important in studying effective population size and rate of inbreeding particularly in extensively raised and wild populations, where pedigree records are scarce. The objective of this study was to investigate LD and estimate effective population size of Southern African village chicken populations. These chickens are raised under scavenging systems of production characterized by uncontrolled breeding and frequent population bottlenecks due to disease outbreaks and fluctuations in feed supplies. DNA from 312 village chickens of South Africa (n = 146), Malawi (n = 30) and Zimbabwe (n = 136) were genotyped using the Illumina iSelect chicken SNP60K Bead Chip. A panel of 43,157 out of the total 57,636 (74.8%) SNPs was used in the final analysis after screening for those that had a minor allele frequency of less than 5%, deviated from Hardy-Weinberg equilibrium ($P<0.001$) and had a call rate of less that 95%. Results indicated that LD averaged between 0.48 ± 0.08 and 0.50 ± 0.08 for SNPs that had a pairwise distance of less than 20 kb. LD dropped to 0.41 ± 0.006 for SNPs between 20 and 100 kb after which it remained constant. Significant differences in LD ($P<0.05$) were observed between chromosomes and LD decayed significantly over marker distance. There were no differences in LD between geographically isolated populations. LD based effective population size analysis indicates that effective population size for the three populations was less than 20 individuals forty generations ago having decreased from 120 animals 400 generations prior. The findings indicated reduced genetic variation in village chickens. The utility of the iSelect chicken SNP60K bead chip in investigating free-range chicken population genetics was demonstrated.

3.1 INTRODUCTION

Extensively raised chickens are kept under a low input management system. These chickens scavenge for their feed in open fields and feed supplementation and other management interventions by humans are minimal (Ekue et al. 2002; Muchadeyi et al. 2009; Mtileni et al. 2010). The chickens are considered non-descript and have not been bred or selected for commercial traits or for breed development (Ekue et al. 2006). Extensively raised village chickens are considered a valuable source of biodiversity that is needed for future advances and improvement in response to climate change and consumer demands (Delany 2004). The genetic variation found in these chickens enables both adaptive evolutionary changes and artificial selection. Village chickens are regarded as diverse and to be holding valuable genotypes that have been developed over thousands of years, and survived successfully in extreme and unusual environmental conditions with limited veterinary and management input (Hall & Bradley 1995). The diversity found in these chickens ought to be characterized and conserved and could be manipulated in similar production systems such as free-range organic farming systems. Diversity studies from previous research (Muchadeyi et al. 2007a; Mtileni et al. 2010; Mtileni et al. 2011a) used microsatellite markers that were of sparse density and could not be used to extensively estimate the population demographic estimates. Surveys that can reveal the effective populations sizes of these populations, their inbreeding levels, the effects of natural and artificial selection, along with population bottleneck events that shaped the current genetic structures will provide valuable information that can be used to manage and conserve these valuable indigenous animal genetic resources. In the presence of dense marker sets, molecular based methods can now be used to calculate linkage disequilibrium, which can then be analyzed and used to determine these population demographic parameters in the absence of pedigree, and performance record as it is the case in these extensively raised chicken populations.

Linkage disequilibrium (LD) is defined as a non-random association of alleles at two or more loci (Hedrick 2004; Qanbari et al. 2010a). It is a useful tool in population and quantitative genetics and also in ecology and evolutionary biology. Linkage disequilibrium can be useful in understanding the level of inbreeding and the genetic background of animal populations and assist in the fine mapping of genes and

quantitative traits loci (QTL) of economically important traits (Wragg et al. 2012). LD analysis also help in understanding the biological and demographic processes such as recombination, mutation, selection and founder effects that could have shaped the population structures (Rao et al. 2008). The decay and extent of LD at a pair-wise distance can be used to determine the evolutionary history of populations (Andreescu et al. 2007; Lu et al. 2012; Wragg et al. 2012). Linkage disequilibrium will therefore be of use particularly in extensively raised chicken populations raised in smallholder farming systems where it can be used to calculate population parameters that could otherwise not be estimated due to lack of pedigree and performance data.

The advent of whole genome sequencing and high density SNPs genotyping technologies have resulted in increased marker density and facilitated estimation of LD in a number of domesticated animals including chicken. The completion of the first draft of the chicken genome (Hillier et al. 2004) made it possible for the development of high density markers (Groenen et al. 2011; Kranis et al. 2013). The Illumina iSelect chicken SNP 60K Bead Chip consist of a panel of 57,636 SNPs (Groenen et al. 2011) that have found utility in LD analysis in various commercial (Qanbari et al. 2010a) and traditional chicken populations as well as in other analysis such as mapping of Mendelian traits (Wragg et al. 2012) and in copy number variations screening (Jia et al. 2012).

This study sought to investigate the extent and decay of LD in extensively raised village chicken populations of South Africa (SA), Malawi (Mal) and Zimbabwe (Zim). These are chicken populations that are raised by smallholder communal farmers under village chicken farming systems characterized by low input management, uncontrolled mating systems and intermixing of flocks within and between villages (Muchadeyi et al. 2007a). Population genetic structures of these village chicken populations could be a function of small flock sizes, inbreeding as farmers retain breeding animal from within flocks over a couple years as well as natural selection from disease outbreaks, extreme weather conditions and poor quality feeds. The objectives of the study were therefore to (i) investigate the extent and decay of LD in these village chicken populations and (ii) to estimate LD-based effective population sizes and provide

baseline information that could be used in further analysis of the population for management and conservation purposes.

3.2 MATERIALS AND METHODS

3.2.1 Chicken populations, blood collection and SNP genotyping

A total of 312 village chickens were randomly sampled from South Africa, Malawi and Zimbabwe. South African village chickens were represented by chickens from Limpopo (n = 15), Eastern Cape (n = 26) and Northern Cape (n = 35) provinces, and four conserved flocks of Venda (VD, n = 20), Naked Neck (n = 20), Potchefstroom Koekoek (n = 20) and, Ovambo (n = 10) that are kept at the Agriculture Research Council, Poultry Breeding Resource Unit at Irene in Pretoria. Detailed sampling of these populations are described by (Mtileni et al. 2011b). A total of 136 village chickens were sampled from three agro-ecological zones of Zimbabwe i.e.: AEZ1 (n = 92), AEZ3 (n = 34) and AEZ5 (n = 10). The detailed sampling of Zimbabwe chicken populations are described by (Muchadeyi et al. 2007a). Thirty chickens sampled from one region of Malawi were also used in the study. Basically the study selected individuals, households, villages, and regions to get genetically unrelated individuals representing a wide geographical location. The distances between villages within a district ranged from 20 to 40 km, and 100 to 500 km between districts within a province and over 1000 km between provinces. The number of individuals varied from 2 to 10 per village depending on household chicken density in each village. All chickens used in this study were not selected for any commercial production traits and were raised by communal farmers under a scavenging system of production.

Blood samples had been collected on FTA Micro Cards (Whatman Bio Science, UK) described in the previous studies (Muchadeyi et al. 2007a; Mtileni et al. 2011b). DNA was extracted from these FTA cards using a modified Qiagen® DNeasy Blood and Tissue protocol. DNA quality was checked on a 1% agarose gel where bright sharp bands where observed indicating an intact DNA (no degradation) and DNA concentration of 50ng/µl for each sample was used for genotyping.

3.2.2 SNP genotyping and data preparation

SNP genotyping was done using the Illumina chicken iSelect SNP 60K Bead chips using the Infinium assay compatible with the Illumina HiScan SQ genotyping platform. This Infinium whole genome genotyping assay is designed to interrogate a large number of SNPs at unlimited levels of loci multiplexing (Illumina 2013). SNP calling was done using Illumina Genome Studio v2.0. The genotype input file was converted into a PLINK v1.07 (Purcell et al. 2007) input file using a plug-in compatible with the Genome Studio program.

SNP quality control was done to remove SNPs that deviated from Hardy-Weinberg equilibrium (HWE) ($p < 0.001$) and SNPs showing a minor allele frequency (MAF) of at least 5%, missing genotypes (>95%) and individuals with missing genotype (>95%) using PLINK v1.07 (Purcell et al. 2007). SNPs that were located on unknown chromosomes, linkage groups and on sex chromosomes were excluded from further analysis. After filtering, 43157, 45676, 46905, 44667 SNPs on 28 autosomal chromosomes were used for overall population, Malawi, Zimbabwe and South African populations, respectively.

3.2.3 Minor allele frequency analysis, heterozygosity and deviation from HWE

Plink v1.07 program (Purcell et al. 2007) was used to measure minor allele frequency for all the 53 476 SNPs in each population. Bins were set for minor allele frequencies of 0.0001-0.05; 0.05-0.1, 0.1-0.2; 0.2-0.3; 0.3-0.4 and 0.4-0.5. The proportion of SNPs per bin was calculated by dividing the number of markers per bin by total number of markers included in MAF estimation.

3.2.4 Linkage Disequilibrium

A pair-wise r^2 estimation was used to measure LD between pairs of loci using PLINK v1.07 program (Purcell et al. 2007) for SNPs on autosomal chromosomes 1 to 28 that had passed the quality control as described earlier and for SNP pairs that had a minimum LD of 0.05. The r^2 measure was chosen because it is independent of allele frequency and it correlates multiples of SNPs at two independent loci (Lu et al. 2012). Briefly, its calculation, considers 2 loci, A and B, each locus having 2 alleles (denoted A_1, A_2; B_1,

B_2, respectively). The frequencies of the haplotypes will then be denoted as f_{11}, f_{12}, f_{21}, and f_{22} for haplotypes A_1B_1, A_1B_2, A_2B_1, and A_2B_2, respectively and as f_{A1}, f_{A2} f_{B1}, and f_{B2} for A_1, A_2, B_1, and B_2, respectively. From this r^2 was then being calculated as:

$$r^2 = \frac{(f_{11}f_{22} - f_{12}f_{21})^2}{fA_1fA_2fB_1fB_2}$$

A Generalized Linear Model procedure (Proc. GLM) in the Statistical Analysis System (SAS 2011) was used to determine the effects of chromosome, population, SNP marker interval (bp), and the interaction of chromosome-by-population on LD, using the following model:

$r^2_{ij} = \mu + Pop_i + Gga_j + (Pop \times Gga)_{ij} + bSNP_{int} + e_{ik}$,

where: Pop_i was the effect of *i*th chicken population from Malawi, Zimbabwe or South Africa; Gga_j was the effect of the *j*th chromosome 1-28; and SNP_{int} the effect of SNP interval fitted as a covariate with regression coefficient *b*.

Linkage disequilibrium decay was estimated for all autosomes and specifically for those chromosomes that had significantly different LD based on the Proc. GLM analysis. Sliding window bins for LD decay were set at 10; 20; 40; 60; 100; 200; 500; 1000 and 2000 kb.

3.2.5 Trends in effective population size

The relationship between *Ne*, recombination frequency and expected LD (r^2) was determined using the following equation (Hayes et al. 2003);

$$E\left(r^2_{adj} = \frac{1}{(\alpha + 4N_e c)} + \frac{1}{n}\right)$$

where α = 1 when assuming no mutations and 2 if mutation was considered, $r^2_{adj} = r^2 - \frac{1}{2n}$, *c* was the recombination rate, and *n* was the chromosomal sample size. The effective population size *Ne*, in $\frac{1}{2c}$

generations ago, was estimated from the observed r^2 values related to a given genetic distance *d*, assuming, $c = d$.

For each pair of SNP on each chromosome, recombination rate was estimated by converting physical marker interval length x_i (Mb) to the corresponding genetic length c_i using the formula: $c_i = \bar{o}_i x_i$, where $\bar{o}_i$ is the average ratio of Morgan per kilo base pair on chromosome *i*, which will be taken from the physical lengths of the chicken genome v74 (Ensembl 2013). The genetic length of chromosomes was adopted from Hillier et al. (2004). The r^2 value ranges between 0 and 1, whereby a zero value indicates uncorrelated SNPs while one reflect SNPs perfectly correlated (Qanbari et al. 2010a).

Setting bins then estimated the trends in effective population sizes at 10; 20; 60; 100 and 1000 kb. The bins were designed to cover the genome in tens, hundreds, thousands and hundred thousand base pairs.

3.3 RESULTS

3.3.1 SNP marker characteristics

Over 8.5% of the SNPs on the Illumina iSelect chicken SNP50 panel had a minor allele frequency of less than 0.05 (Table 3.3.1). Chickens from Malawi had a higher percent of alleles occurring at a frequency below 0.05 followed by Zimbabwe and South African chickens. Malawi chickens had a higher proportion of monomorphic loci (3526) compared to 626 and 1879 for South African and Zimbabwean chickens respectively. The observed level of heterozygosity ranged from 0.25±0.12 in Malawi chickens to 0.31±0.08 in South African chickens. Overall H_o in all populations was lower than was expected. For all populations, over 80% of the SNPs were used in the final analysis with majority of SNPs pruned out due to MAF, monomorphic and missing genotypes (Table 3.3.1).

Table 3.3.1: Distribution of markers after SNP quality control and the minor allele frequency, observed and expected heterozygosities of the Malawi, South African and Zimbabwean chicken populations.

	Malawi	South Africa	Zimbabwe
Total SNPs	57636	57636	57636
SNPs MAF < 0.05	6579	4532	4773
Unknown chromosome	135	162	30
Monomorph	3526	631	1879
HWE (p<0.001)	209	3264	1278
Missing genotypes > 0.05	1490	1949	1353
SNPs in use	45676	44667	46905
Mean MAF ±SD	0.26 ±0.14	0.28 ±0.14	0.28 ±0.14
Mean Obs. Heterozygosity	0.25 ±0.15	0.31 ±0.08	0.30 ±0.09
Mean Exp. Heterozygosity	0.40 ±0.09	0.40 ±0.10	0.38 ±0.11

Quality control pruning resulted in a large fraction of SNPs being eliminated from the analysis based on MAF, monomorphic and missing genotypes. SNPs that were sitting on unknown chromosome, linkage groups, and sex chromosomes were also excluded from further analysis. The proportion of SNPs used for further analysis was 85% for Malawi, 83% for South Africa and 87% for Zimbabwe (Table 3.3.1).

3.3.2 Minor allele frequency distribution

Minor allele frequency averaged 0.26±0.14 for chickens from Malawi and was 0.28±0.14 for Zimbabwe and South African chickens (Figure 3.3.1). An analysis of the distribution of MAF across all populations showed that over 80% markers had a minor allele frequency greater than ten percent

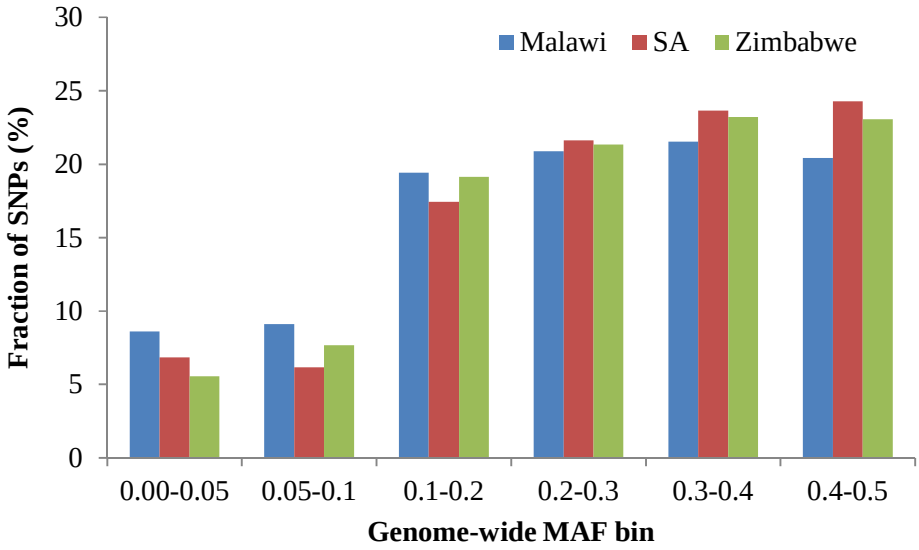

Figure 3.3.1: Genome wide distribution of minor allele frequency of SNPs for each population.

3.3.3 LD estimates and the effects of chromosome, SNP intervals and breed

Table 3.3.2 summarizes the total length, number of SNPs, and average SNP interval and r^2 values for the 28 autosomal chromosomes in the three village chicken populations. The SNP interval was not consistent across the genome ranging from a distance of 0.01 to 0.1 Mb. Macro-chromosome showed to have highest marker distance followed by intermediate chromosome and SNP intervals were shorter for micro-chromosome. Number of SNPs per chromosome varied with chromosome size between the macro (chromosomes 1-5) that had highest number of SNPs ranging from 900 to 3000 and micro (chromosome 16-28) with number of SNPs ranging from 70 to 700 per chromosome.

Overall population LD ranged from 0.34±0.2 to 0.52±0.26 and averaged 0.40±0.04 (Figure 3.3.2). Per population LD ranged from 0.35±0.14 to 0.46±0.25 for Malawi, 0.35±0.2 to 0.50±0.25 for Zimbabwe and 0.34±0.12 to 0.42±0.2 for South African chickens. Pairwise LD varied significantly (*P<0.001*) among chromosomes with high LD observed on chromosomes 8, 22, and least on chromosome 13 and 18 (Table 3.3.3). There were no significant differences in LD among populations (P>0.05). A population by chromosome interaction on LD was also observed (Table 3.3.3) whereby chromosome 8 of Zimbabwe had a highest LD of 0.50±0.25 whilst chromosome 22 had the highest LD in the SA chicken population

(Figure 3.3.3). Chromosome 18 had the lowest LD with an average value of 0.35±0.14 Malawi and Zimbabwe 0.35±0.2, where South Africa had lower values of 0.34±0.14.

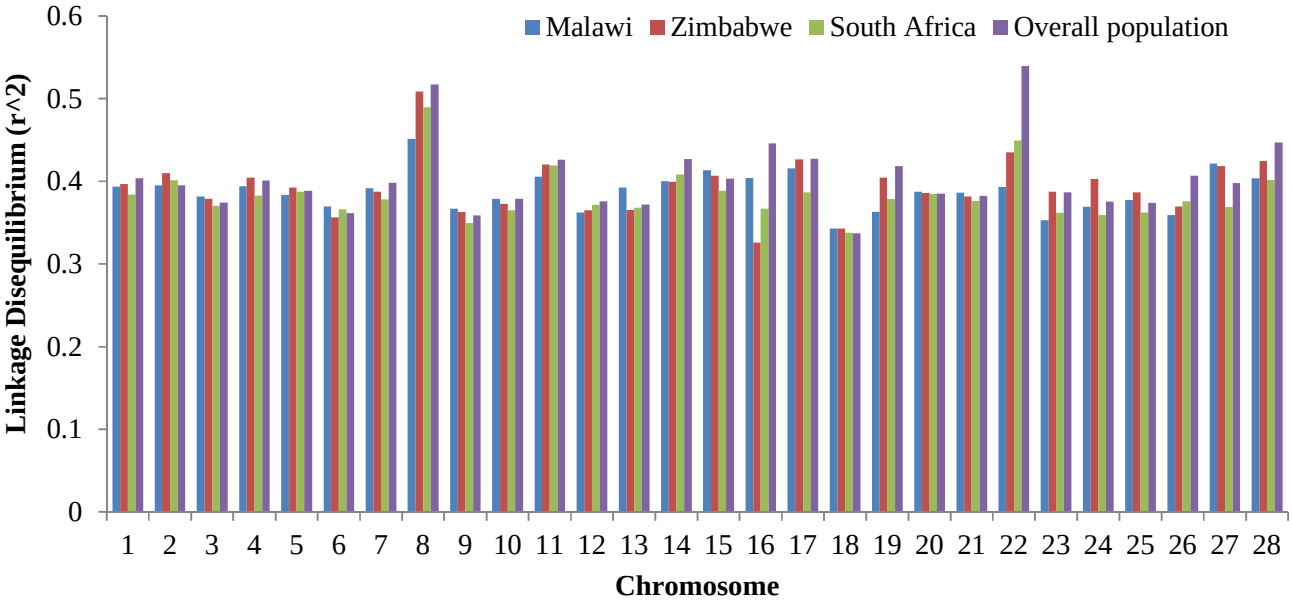

Figure 3.3.2: LD distribution across the 28 autosomal chromosomes for each population.

Linkage disequilibrium was also affected by SNP interval. Figures 3.3.3 (a-e) illustrates the rate of LD decay over marker distance for all 28 chromosomes (Figure 3.3.3a) and for chromosomes 22, 8, 18, and 13 (Figures 3.3.3b-e respectively). Overall, a sharp decline in LD from 0.58±0.28 to ~0.45±0.23 at 10 kb pairwise distance was observed followed by a gradual decline to 0.4 at a pairwise distance of less than 200 kb after which it was maintained well over 1000 kb (Figure 3.3.3a). The LD on chromosome 22 of South Africa and Zimbabwe chickens was high (0.85±0.0 for South Africa and 0.7±0.09 for Zimbabwe) at a pairwise distance of 1 kb after which it sharply decreased to 0.49±0.23 at 10 kb and remained constant beyond 200 kb (Figure 3.3.3b). LD was maintained at 0.40 over all sliding windows on chromosome 22 for Malawi chicken population. The LD decay at chromosome 8 for Malawi chickens continued to decline after 40 kb.

Unexpected trend was observed on chromosome 18 where a very low LD of below 0.25 were observed at a pairwise distance of 1 kb in Malawi chickens (Figure 3.3.3d). LD in these chicken populations then increased to 0.35 at 10 kb after which it maintain the same genetic patterns as that of South Africa and Zimbabwe.

Table 3.3.2: Summary of analyzed SNPs.

GGA	Length (Mb)	Average SNPs (interval in Mb)	Number of SNP	Max distance between SNPs (Mb)	Min distance between SNPs (base)	Average r^2			
						Malawi	Zimbabwe	South Africa	Overall population
1	195.3	0.1±0.08	3443	0.4	18	0.40±0.2	0.40±0.2	0.40±0.2	0.40±0.2
2	148.8	0.1±0.9	2399	0.6	1	0.41±0.2	0.42±0.2	0.41±0.2	0.40±0.2
3	110.4	0.09±0.08	1996	0.5	46	0.39±0.2	0.40±0.2	0.40±0.2	0.37±0.2
4	90.2	0.09±0.08	1602	0.4	5	0.40±0.2	0.41±0.2	0.40±0.2	0.40±0.2
5	59.6	0.1±0.09	976	0.5	46	0.40±0.2	0.40±0.2	0.40±0.2	0.40±0.2
6	34.9	0.06±0.5	832	0.3	16	0.38±0.2	0.37±0.2	0.40±0.2	0.36±0.2
7	36.2	0.08±0.07	844	0.4	145	0.40±0.2	0.39±0.2	0.40±0.2	0.40±0.2
8	28.8	0.09±0.07	650	0.4	124	0.46±0.2	0.50±0.2	0.40±0.3	0.52±0.3
9	23.4	0.07±0.05	474	0.3	772	0.38±0.2	0.37±0.2	0.36±0.2	0.36±0.2
10	19.9	0.05±0.03	621	0.2	1	0.40±0.2	0.38±0.2	0.40±0.2	0.40±0.2
11	19.4	0.08±0.1	648	0.7	3	0.41±0.2	0.42±0.2	0.42±0.2	0.42±0.2
12	19.9	0.05±0.4	583	0.2	145	0.37±0.2	0.37±0.2	0.40±0.2	0.40±0.2
13	17.8	0.05±0.4	572	0.2	56	0.40±0.2	0.37±0.2	0.40±0.2	0.37±0.2
14	15.2	0.06±0.04	456	0.2	94	0.40±0.2	0.40±0.2	0.41±0.2	0.43±0.2
15	12.7	0.05±0.04	551	0.2	15142	0.42±0.2	0.41±0.2	0.40±0.2	0.40±0.2
16	0.535	0.05±0.03	5	0.08	1	0.40±0.2	0.33±0.1	0.40±0.2	0.45±0.2
17	10.4	0.05±0.03	379	0.2	51	0.42±0.2	0.43±0.2	0.40±0.2	0.43±0.1
18	11.2	0.05±0.4	358	0.2	90	0.35±0.1	0.35±0.2	0.34±0.1	0.34±0.2
19	9.9	0.04±0.05	340	0.3	6	0.37±0.2	0.40±0.2	0.37±0.2	0.42±0.2
20	14.3	0.04±0.03	755	0.2	73	0.39±0.2	0.40±0.2	0.38±0.2	0.38±0.2
21	6.8	0.03±0.03	359	0.1	90	0.39±0.2	0.38±0.2	0.38±0.2	0.38±0.2
22	4.1	0.05±0.46	135	0.2	1	0.40±0.2	0.44±0.25	0.45±0.3	0.54±0.3
23	5.7	0.04±0.04	174	0.2	13	0.36±0.2	0.38±0.2	0.37±0.2	0.40±0.2
24	6.3	0.03±0.02	254	0.1	20	0.37±0.2	0.41±0.2	0.36±0.2	0.37±0.2
25	2.2	0.03±0.02	71	0.1	20	0.40±0.2	0.40±0.2	0.37±0.0.2	0.37±0.2
26	5.3	0.03±0.03	138	0.2	6	0.36±0.2	0.36±0.2	0.37±0.2	0.41±0.2
27	5.2	0.06±0.09	99	0.5	160	0.42±0.2	0.42±0.2	0.37±0.2	0.40±0.2
28	4.7	0.03±0.03	252	0.2	12	0.41±0.2	0.42±0.2	0.40±0.2	0.45±0.2

Table 3.3.3: The effects of country of origin, chromosome and SNP marker interval on LD.

Factor	Df	Sum sq.	Mean sq.	F-value	Sign.
Country	2	0.03	0.01	0.41	0.67
Chromosome	27	88.43	3.27	86.32	***
SNPs interval.	1	49.95	49.95	1316.40	***
Country x chromosome	54	12.23	0.23	5.97	***
Residuals	136804	5191.10	0.04	42.78	***

****p < 0.0001*

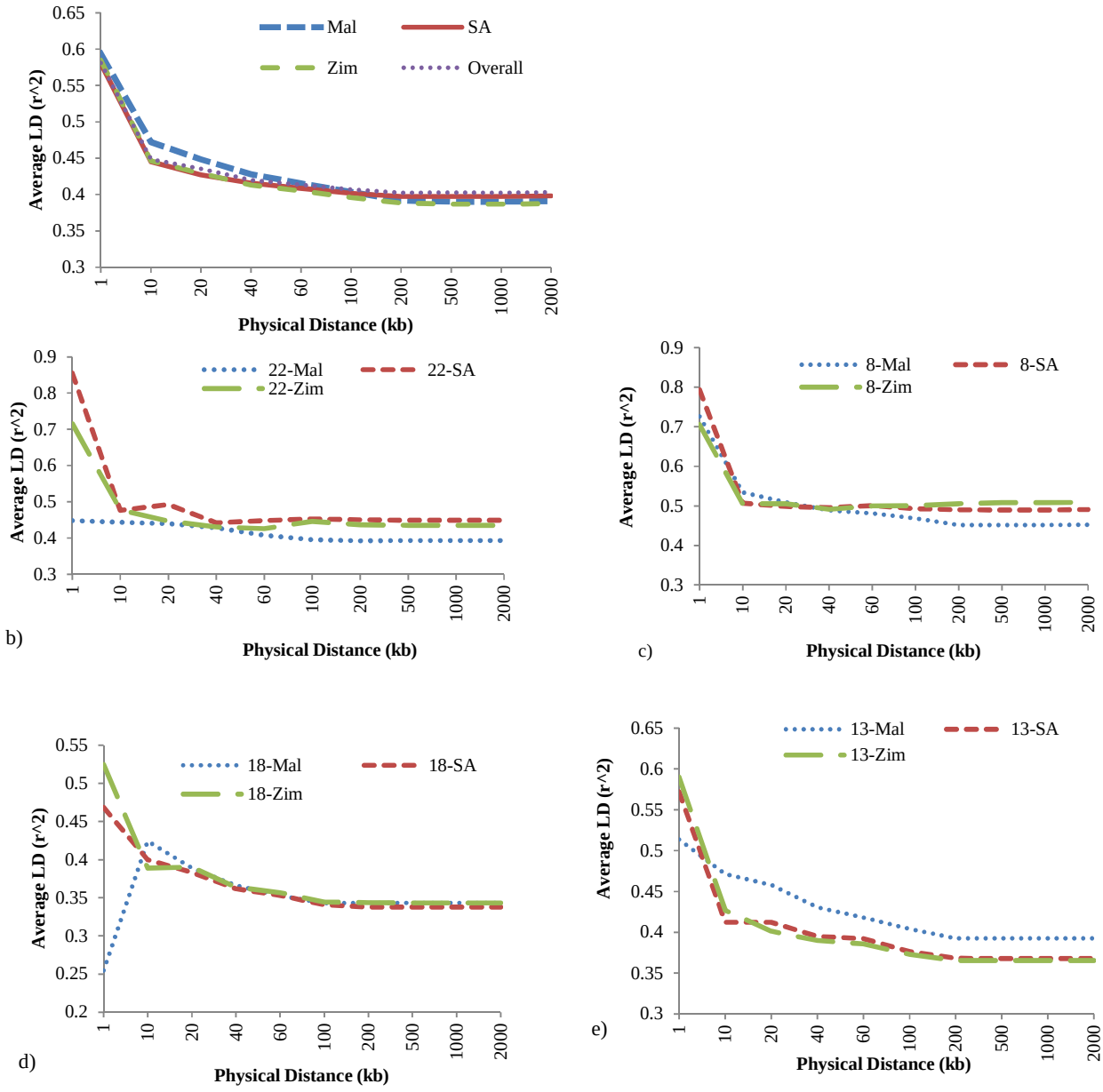

Figure 3.3.3: Average LD decay with n increase in physical distance between SNPs for a) chromosomes 1-28, b) chromosome 22, c) chromosome 8, d) chromosome 18 and e) chromosome 13 for Malawi (Mal), South Africa (SA) and Zimbabwe (Zim) chicken populations.

3.3.4 Effective population size over the past generations

Figure 4 plots the estimated effective population size N_e (t) at t generations ago. The adjusted LD based estimates of N_e indicated an effective population size of less than 20 individuals for all the populations' 40 generations prior. The major reduction of effective population size was observed 80 generations prior.

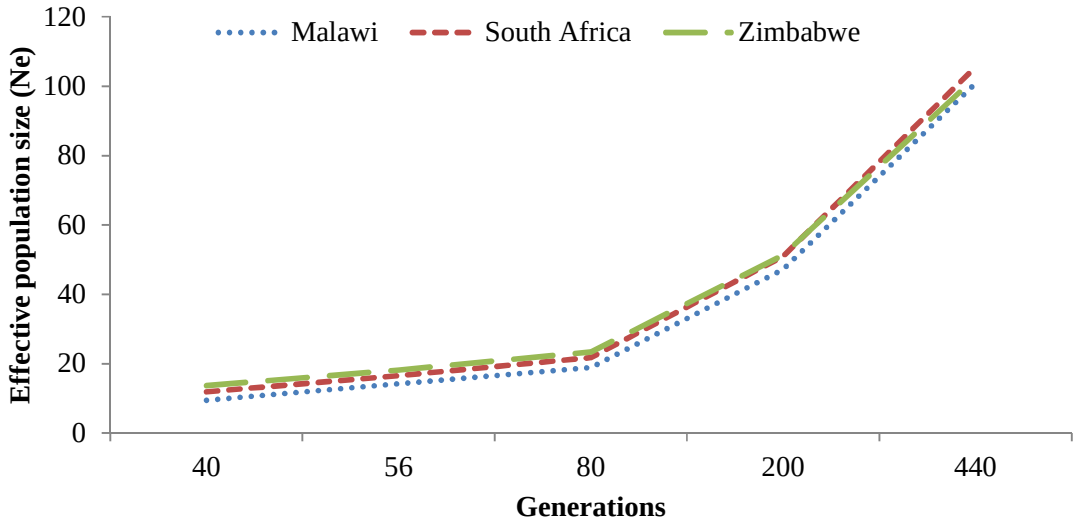

Figure 3.3.4: Trends in effective population size over 440 generations ago.

3.4 DISCUSSION

Village chicken populations of sub-Sahara Africa has not been well studied to understand the genetic and demographic parameters that are in play in shaping their genetic structures. Previous studies have suggested that village chickens hold valuable genetic diversity particularly for smallholder resource-limited farmers due to their existence over a wide geographical distribution that is characterized by extreme climatic and environmental conditions. There is however limited information on population demographics of village chickens populations kept by smallholder farmers under extensive systems of production. The random mating systems and absence of pedigree and performance data makes it difficult to estimate the effective population sizes and key population genetic parameters such as inbreeding and selection in these populations. Absence of these demographic estimates hinders genetic improvement

programs that are needed to increase productivity. Efforts have been made to understand genetic diversity using microsatellite markers (Muchadeyi et al. 2006; Muchadeyi et al. 2007a; Kaya & Yildiz 2008; Hassen et al. 2009). These markers were however, not informative enough to provide accurate estimates of such population parameters. High density SNP chip has been successfully used in previous studies to characterize LD (Qanbari et al. 2010a; Megens et al. 2010), map Mendelian traits and screen for other genetic variants in both commercial and traditional chicken populations raised under similar production systems such as village chicken populations (Wragg et al. 2012). There are however few case studies on the utility of this panel of markers for village chickens particularly from Southern Africa. This study therefore seeks to use genome-wide SNP data to estimate linkage disequilibrium and population demographic history of extensively raised chicken populations of Southern Africa.

Over 80 % of the SNPs on the panel could be used for further analysis after pruning (Table 3.3.3). The proportion of usable SNPs was higher than what was used in commercial egg laying chickens (Qanbari et al. 2010a) but similar to that of study conducted by Wragg et al. (2012) where they also included traditional breeds and village chicken populations from Ethiopia, Kenya and Chile. Although not used in the development of 60K SNP chip, Wragg et al. (2012) has proven the utility of the 60K chip to non-descript village chicken populations.

Linkage disequilibrium was calculated using 28 of the 38 chicken autosomal chromosomes that were represented on the Illumina iSelect SNP 60K bead chip. SNPs on linkage groups and sex chromosomes as well as those of unknown marker positions were excluded from the analysis. Most SNPs were pruned due to monomorphism and minor allele frequency. The number of monomorphic markers and SNPs of lower MAF is not comparable to that from previous studies. Qanbari et al. (2009) had about 5 fold more monomorphic markers excluded and 3 fold fewer markers excluded based on MAF < 0.05 compared to our study. Wragg et al. (2012) had 2-3 fold less monomorphic markers excluded before further analysis in comparison to our study. Variations in the number of alleles pruned for MAF and monomorphism could be explained by the different genetic backgrounds that would result in different alleles segregating in the

commercial lines used by Qanbari et al. (2009) and the traditional and village chicken population used by Wragg et al. (2012). A threshold of MAF < 0.05 was used in this and other studies which according to Corbin et al. (2010) can increase accuracy on LD measures when sample size is large. It was observed (Corbin et al. 2010; Corbin et al. 2012) that pruning MAF of more than 0.1 can lead to ascertainment bias on the measures of effective population size particularly in small to moderate sample sizes. The distribution of marker in the different MAF categories can be explained by optimization of the iSelect chicken SNP 60K panel in accordance with uniform marker spacing and MAF distribution (Hillier et al. 2004).

There were no significant differences on the LD between populations. This can be an indication of a homogenous mixture of village chicken populations, and also indicating that these populations might be experiencing similar evolutionary mechanism such as bottleneck effect, genetic drift, selection, recombination rate and therefore be of similar effective population sizes. The Level of LD was unevenly distributed across the chromosomes where they were four chromosome with extreme levels of LD. Amount of LD was high for chromosome 8 and 22 for almost all populations and was significantly low for chromosome 13 and 18 . The effect of chromosomal difference is in support with observations by Andreescu et al. (2007); Megens et al. (2009) and Qanbari et al. (2010a) even though Andreescu et al. (2007) and Megens et al. (2009) studies were of selected genomic regions on selected chromosomes. Such findings indicate that different evolutionary forces affecting LD are acting at the chromosomal level in these populations. The variation in LD among chromosomes could be a function of the different genetic factors that might be in play and affecting different chromosomes differently. The significantly variable LD on chromosomes 8, 13, 18 and 22 could be an indication of selection (Hedrick 2004) particularly natural selection as these chicken populations are raised under extensive low input production systems where human selection pressures are minimal (Mtileni et al. 2010). An analysis of these chromosomes showed existence of known QTLs on chromosome 8, 13, 18 and 22 that play a role in body composition, feed conversion ratio, fat percentage and some antibody related traits. A QTL for antibody

response to Newcastle Disease Virus (NDV) was found on chromosome 22 (Chicken QTL Database 2013).

Results in this study also indicated a significant LD decay with increased marker intervals, which generally according to Megens et al. (2009) is a function of increased recombination events with increased genetic distance. The GC content and high density of genes on micro-chromosome compared to macro-chromosome can influence extend and distribution of LD on the chicken genome (Megens et al. 2009) and results from this study agreed with the expected trends.

However, over and above the expected trends in LD decay with increased marker distances, LD in this population was generally high and remained well above 0.4 at marker distances over 1000 kb. This high average LD that is persistent over long distances could be a reflection of the population sub-structuring that is common among village chicken populations. Village farmers are known to keep small flocks of chickens ranging from 1-20 chickens per household. Although cock sharing is expected within villages, there is no evidence from previous studies of farmers sharing hens. It could be that there are very small flock sizes and some level of population sub-structuring observed at village level. Cock/breeding animal sharing between villages is very limited. Although this study assumed animals in a country as one big population, there could be multiple sub-populations within each country, which will be chickens from different villages. Such population sub-structuring if present would explain the high and persistent LD values. It was however challenging to define these population boundaries in the absence of pedigree records during sampling. Moreover results from microsatellite and mtDNA analysis (Muchadeyi et al. 2007; Mtileni et al. 2011) had suggested low level of population sub structuring between farming regions within countries.

Analysis of trends in effective population size from LD values suggested low effective population size that are below the recommended safe thresholds as defined by the FAO–DAD guidelines (FAO 2013). Results showed a decrease in genetic variation which could be due to poor management, inbreeding due

to population sub-structuring within villages or population bottlenecks that could have been experienced 80 generations prior (Figure 3.3.4). The reduced levels of effective population sizes could explain the reduced heterozygosity observed in these populations (Table 3.3.1). It shows that these populations are generally inbred within subpopulations as suggested in previous study (Muchadeyi et al. 2007a).

Factors that could lead to inbreeding in these populations could be the small flock sizes and genetic isolation between flocks among villages, unmonitored mating system as well as population bottlenecking due to disease outbreaks and predation. The overlapping generations promotes animals on the same flock or closely related animals to mate, subsequently increasing levels of inbreeding. These populations are recognized as outbred and diversity studies have indicated that these populations are highly diverse. These studies were using less dense microsatellite markers that had a lower precision power than SNPs (Muchadeyi et al. 2006; van Marle-Köster et al. 2008; Mtileni et al. 2011a). Population bottlenecking could have happened during the introduction of chickens into South Africa, which according to archeological findings is thought to have happened year 1600 (van Marle-Köster et al. 2008, Mtileni et al. 2010).

3.5 CONCLUSION

A relatively high LD that persisted over short SNP intervals was observed in the South African, Zimbabwean and Malawian chicken populations. This LD pattern seems to be consistent with low effective population sizes and loss of heterozygosity in the village chicken populations. The study recommends for a further investigation on the role of selection and population bottlenecks on chromosomes 8 and 22 that had significantly high LD value. Haplo-blocks partitioning across the genome will also provide useful information on recombination effects on LD.

3.6 REFERENCES

Andreescu C., Avendano S., Brown S.R., Hassen A., Lamont S.J. & Dekkers J.C. (2007) Linkage disequilibrium in related breeding lines of chickens. *Genetics* **177**, 2161-9.

Chicken QTL Database (2013) Browse the Chicken QTLdb. Available: URL http://www.animalgenome.org/cgi-bin/QTLdb/GG/browse. [Accessed August 2013].

Corbin L., Blott S., Swinburne J., Vaudin M., Bishop S. & Woolliams J. (2010) Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics* **41**, 8-15.

Corbin L., Liu A., Bishop S. & Woolliams J. (2012) Estimation of historical effective population size using linkage disequilibria with marker data. *Journal of Animal Breeding and Genetics* **129**, 257-70.

Delany M.E. (2004) Genetic variants for chick biology research: from breeds to mutants. *Mechanisms of Development* **121**, 1169-77.

Ekue F., Pone D., Nfi J.M.A. & Njoya J. (2006) The effect of vaccination against Newcastle disease and feed supplementation on production in village chicken in Bamenda area of Cameroon. In: *Improving farmyard poultry production in Africa: interventions and their economic assessment: proceedings of a final research coordination meeting*, International Atomic Energy Agency, p. 51.

Ensembl (2013) Chicken genome. Available: URL http://www.ensembl.org/Gallus_gallus/Location/Genome. [Accessed August 2013].

FAO (2013) *In vivo conservation of animal genetic resources*, Rome.

Groenen M., Megens H.-J., Zare Y., Warren W., Hillier L., Crooijmans R., Vereijken A., Okimoto R., Muir W. & Cheng H. (2011) The development and characterization of a 60K SNP chip for chicken. *BMC Genomics* **12**, 274.

Hayes B.J., Visscher P.M., McPartlan H.C. & Goddard M.E. (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* **13**, 635-43.

Hall S.J. & Bradley D.G. (1995) Conserving livestock breed biodiversity. *Trends in Ecology and Evolution* **10**, 267-70.

Hedrick P.W. (2004) *Genetics of Populations.* Jones and Bartlett, Canada.

Hillier L.W., Miller W., Birney E., Warren W., Hardison R.C., Ponting C.P., Bork P., Burt D.W., Groenen M.A. & Delany M.E. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716.

Illumina I. (2013) Infinium HD Assay Ultra Protocol Guide. (ed. by Illumina), p. 240, San Diego.

Jia X., Chen S., Zhou H., Li W. & Yang N. (2012) Copy number variations identified in the chicken using a 60K SNP BeadChip. *Animal Genetics* **44**, 276-84.

Kaya M. & Yildiz M.A. (2008) Genetic diversity among Turkish native chickens, Denizli and Gerze, estimated by microsatellite markers. *Biochemical Genetics* **46**, 480-91.

Kranis A., Gheyas A.A., Boschiero C., Turner F., Yu L., Smith S., Talbot R., Pirani A., Brew F. & Kaiser P. (2013) Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* **14**, 59.

Lu D., Sargolzaei M., Kelly M., Changxi L., Vander Voort G., Wang Z., Plastow G., Moore S. & Miller S.P. (2012) Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Frontiers in Genetics* **3**, 10.

Megens H.-J., Crooijmans R.P., Bastiaansen J.W., Kerstens H.H., Coster A., Jalving R., Vereijken A., Silva P., Muir W.M. & Cheng H.H. (2009) Comparison of linkage disequilibrium and haplotype diversity on macro-and microchromosomes in chicken. *BMC genetics* **10**, 86.

Mtileni B., Muchadeyi F., Maiwashe A., Chimonyo M., Groeneveld E., Weigend S. & Dzama K. (2011a) Diversity and origin of South African chickens. *Poultry Science* **90**, 2189-94.

Mtileni B., Muchadeyi F., Maiwashe A., Groeneveld E., Groeneveld L., Dzama K. & Weigend S. (2010) Domestication of South African chicken genetic resources. In: *Proceedings International*.

Mtileni B., Muchadeyi F., Maiwashe A., Groeneveld E., Groeneveld L., Dzama K. & Weigend S. (2011b) Genetic diversity and conservation of South African indigenous chicken populations. *Journal of Animal Breeding and Genetics* **128**, 209-18.

Muchadeyi F.C., Weigend S., Wollny C. & Makuza S. (2006) Use of Microsatellites and MtDNA to Assess Genetic Diversity Within and Between Zimbabwe Chicken Eco-types. *Bonn*.

Muchadeyi F., Eding H., Wollny C., Groeneveld E., Makuza S., Shamseldin R., Simianer H. & Weigend S. (2007a) Absence of population substructuring in Zimbabwe chicken ecotypes inferred using microsatellite analysis. *Animal Genetics* **38**, 332-9.

Muchadeyi F., Wollny C., Eding H., Weigend S. & Simianer H. (2009) Choice of breeding stock, preference of production traits and culling criteria of village chickens among Zimbabwe agro-ecological zones. *Tropical Animal Health and Production* **41**, 403-12.

Purcell S.N., B, Todd-Brown K., Thomas L., Ferreira M., Bender D., Maller J., Sklar P., de Bakker P. & PC D.M.S. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. URL http://pngu.mgh.harvard.edu/purcell/plink/.

Qanbari S., Hansen M., Weigend S., Preisinger R. & Simianer H. (2010) Linkage disequilibrium reveals different demographic history in egg laying chickens. *BMC Genetics* **11**, 103.

Rao Y.S., Liang Y., Na Xia M., Shen X., Jun Du Y., Glong Luo C., Hua Nie Q., Zeng H. & Quan Zhang X. (2008) Extent of linkage disequilibrium in wild and domestic chicken populations. *Hereditas* **145**, 251-7.

SAS (2011) SAS User's Guide: Statistics. (ed. by Inc. SI). Cary, North Carolina, USA.

van Marle-Köster E., Hefer C.A., Nel L.H. & Groenen M. (2008) Genetic diversity and population structure of locally adapted South African chicken lines: Implications for conservation. *South African Journal of Animal Science* **38**, 271-81.

Wragg D., Mwacharo J., Alcalde J., Hocking P. & Hanotte O. (2012) Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. *Heredity* **109**, 6-18.

CHAPTER 4

HAPLO-BLOCK STRUCTURE OF SOUTHERN AFRICAN VILLAGE CHICKEN POPULATIONS

ABSTRACT

The observed block like linkage disequilibrium in the human genome has increased interest in the investigation of possible genetic variants associated with haplo-blocks in livestock and model species. This study investigated the haplotype structure, sharing and diversity in extensively raised chicken populations of Southern Africa. A total of 290 animals from Malawi (n = 30), South Africa, (n = 132) and Zimbabwe (n = 128) were included in the study from which 649, 2104, and 2442 haplo-blocks were observed respectively. Large numbers of haplo-blocks were observed in macro-chromosomes (ranging from 118 to 402) compared to micro-chromosome (ranging from 13 to 125). The median block lengths were 13.6 kb, 10.8 kb and 9.7 kb whilst the genome coverage was 39 Mbp, 64.4 Mbp followed by 54.5 Mbp for Malawi, South Africa, and Zimbabwe respectively. Haplo-block sharing was observed between populations with 2325 haplo-blocks common between Zimbabwe and Malawi and 2689 between South Africa and Zimbabwe. Majority of haplo-blocks were < 25kb and only five blocks were more than 2 Mbp. A large number of haplotype had a frequency ranging from 0.25 to 0.5 across all chromosomes. The observed haplo-blocks structure is in agreement with results from other chicken studies analyzed on the 60K SNP panel. Slight variation could be due to the different genetic background of the chicken populations. The low chromosomal coverage of haplo-blocks across the genome suggests high recombination events that brake the ancestral haplo-blocks into blocks sizes less than 10 kb and suggesting the incorporation of useful SNP markers on bead chip for Southern African population. The haplotype sharing indicates overlapping haplo-blocks between extensively raised chicken populations implying transferability of population genetic tools between populations. The observed unique haplo-

blocks suggest isolated evolution accumulating haplo-blocks that are specific to the agro-ecological zone of the sampled countries. Quantitative trait loci (QTLs) analysis revealed genes related to body composition where spanned by haplo-blocks investigated. These traits (Feed conversion ratio, abdominal fat percentage, and body weight) are of importance to these animals to enable them to convert the minimal feed they get on their environment into body requirement, such as body development and maintenance of feather density.

4.1 INTRODUCTION

The evolution of village chicken populations of Southern Africa and other developing countries is not clearly understood. It is assumed that communal farmers who raise village chickens do not impose selection pressures for breed's development as is done in the commercial sector. However, it is also hypothesized that the agro-ecological regions in which these chickens exists impose natural selection pressures that have shaped the gene pool of these extensively raised livestock species (Muchadeyi et al. 2007a) and generated sub-populations that might have accumulated long ancestral haplotypes supporting isolated genetic change. Genetic variation is expected to be high in these village chickens due to the fact that the animals have not experienced artificial selection for specific traits and are thought to be structured more or less like natural populations (Muchadeyi et al. 2006; Muchadeyi et al. 2007a; Muchadeyi et al. 2007b; Mtileni et al. 2011b). Genetic variation is also expected to be influenced by geographical isolation of these populations, and pockets of inbred animals are expected due to low effective population size in each sub-population due to poor husbandry and mating system as well as natural selection pressures from uncontrolled disease outbreaks coupled by poor and fluctuating feed supplies (Li & Merilä 2011).

Studies of linkage disequilibrium in these populations as described in Chapter 3 indicated high LD that extend over long genetic distance which is consistent with reduced effective population size and high levels of population sub-structuring. The shift from sparsely distributed markers to high density marker has made it possible to investigate haplo-block structures in a number of livestock species (Amaral et al. 2008; Megens et al. 2009; Qanbari et al. 2010a) and humans (Zhao et al. 2003). Several studies on

haplotype diversity in chicken have been conducted after the whole genome sequence was made available and the development of high-density markers panel became feasible (Tishkoff & Verrelli 2003; Zhao et al. 2003; Lindblad-Toh et al. 2005; Amaral et al. 2008; Megens et al. 2009). Haplotype blocks are defined as the number (two or more) of alleles in loci that are at bound to be linked together at a close proximity and transferred together (Zhao et al. 2003; Crawford & Nickerson 2005). Haplo-blocks are more informative than single allele in a genome because markers spanning are more stable. Haplo-blocks can be useful in gene mapping for breeding programs and disease variants detection. Haplotype and haplo-blocks have been found to span across quantitative trait regions for traits of commercial importance and across diseases variants traits. In population diversity studies, haplotype sharing amongst individuals can indicate conserved genomic regions that could have been a result of intensive and directional natural or artificial selection thereby finding use in fine mapping of QTLs of disease variants and in association studies for traits of economic importance (Cuc et al. 2006).

Chickens belong to the avian species and have a unique genetic architecture that comprise of macro-, intermediate and micro-chromosome (Megens et al. 2009; Qanbari et al. 2010a; Ensembl 2013). Each chromosome set has different genetic structure and genetic composition and is characterized by different LD patterns. Studies on linkage disequilibrium have suggested block-like structures of markers distributed across the chicken genome. It has been observed that genomic regions that have low recombination events form a block-like structure that can be shared among individuals within and between populations (Cuc et al. 2006). Haplotype boundaries are assumed to be structured by recombination hot spots while recombination cold-spots introduce variation within haplo-blocks (Gabriel et al. 2002; Zhao et al. 2003). The extent and distribution of haplo-blocks across the genome are considered a function of the genetic structure, genetic variation as well as the demography of populations. Megens et al. (2009) have indicated that micro-chromosomes have high GC content as well as lower levels of intragenic and intergenic regions.

Genome-wide analysis of haplotypes and the distribution of haplo-blocks in chickens can assist in identifying regions that can be associated with phenotypic traits or adaptive features. Information on haplotypes has been investigated on selected genomic regions and chromosomes prior the genomics era mainly because of absence of tools to scan whole genomes. Analysis of selected regions made it difficult to compare populations analyzed in different studies on different genomic regions. Regardless of such limitations such analysis added valuable information on the genetic architecture of populations. In the next generation sequencing genomics era, genome-wide haplo-blocks partitioning reveals the degree of haplotype sharing and diversity within and among breeds (Amaral et al. 2010). The next generation sequencing and genome-wide SNP genotyping technologies have made available large SNP data sets that can be used to study most domestic livestock species. The first draft of the chicken genome was made available in 2004 (Hillier et al. 2004) and from it over 2.8 million SNPs have been discovered. The Illumina chicken iSelect 60K SNP chip has been found useful in studying LD as well as haplo-blocks partitioning in commercial (Qanbari et al. 2010a) as well as traditional free ranging chickens (Wragg et al. 2012).

There is no information on haplo-blocks structure of extensively raised chicken population of Southern Africa. These are chicken populations raised under scavenging village production systems and where the genetic structure of the populations seems to be affected by the natural selection pressures from the extreme environments as well as inbreeding due to small sub-structured populations. An analysis of LD in these populations indicated chromosomal differences in LD with high LD associated with chromosomes 8 and 22 and least LD on chromosomes 13 and 18 (Chapter 3). The influence of recombination and existence of haplo-blocks associated with the observed LD profile has not been investigated. An analysis of haplo-blocks structure in these extensively raised chickens will shed more light into utility of the available SNP panel in studying village chicken populations.

The objective of the study was therefore to screen for haplo-blocks and investigate the haplotype structure, sharing and diversity within and between village chicken populations. It was hypothesized in

this study that there were differences in the haplo-blocks in chicken populations from different geographical origins due to isolated evolutionary processes.

4.2 MATERIALS AND METHODS

4.2.1 Animal populations

A total of 312 village chickens were randomly sampled from three different countries and in different ecotypes within a country, South Africa (three ecotypes), Malawi (one ecotype), and Zimbabwe (three ecotypes). From each country a set of animals were sampled by following sampling method describedin Chapter 3.

4.2.2 SNPs quality control and pruning

SNPs pruning was performed using PLINK v1.07 for SNPs that had a minor allele frequency of 0.05, had over 5% missing genotypes, SNPs that deviated from Hardy-Weinberg equilibrium at P = 0.001 and individuals with over 5% missing genotypes. Using these quality control criteria, 45676; 46905 and 44667 markers were used for further analysis for overall populationof Malawi, Zimbabwe and South Africa, respectively.

4.2.3 Haplo-blocks partitioning

Haplo-blocks were estimated in PLINK v1.07 (Purcell et al. 2007) which uses default procedure from Haploview http://www.broad.mit.edu/mpg/haploview/. Individuals within a population were considered similar and therefore treated as cases. Pairwise LD was calculated on SNP distance of 10000 Kb for autosomal chromosome except chromosome 16 since it had less than 20 markers. For blocks partitioning --*blocks* function was used in PLINK as default algorithm by Gabriel et al. (2002) as implemented in Haploview. Blocks were created if 95% confidence bounds on r^2. Haplo-blocks frequency was estimated in PLINK using the --*hap-freq* function. Blocks of different frequencies was generated per chromosome from those blocks occurring at frequency of < 0.1, 0.1-0.25, 0.25-0.5, 0.5-0.75 and 0.75-1.0.

The percentage of SNPs making up a haplo-blocks was calculated by dividing number of SNPs within a block by total number of SNPs used in the haplo-block partitioning multiply by hundred. Unique number of haplo-blocks was defined as those blocks that occurred in one population and not found in the other.

4.2.4 Haplotype Diversity and QTLs detection

Haploview v4.2 was used for LD plots and haplotype frequency within blocks per population. Chromosome 8 and 22 were selected for further analysis based on the LD results showing higher LD than all the other chromosomes. We hypothesized that haplotype diversity is high on these chromosomes that could be harboring QTLs under selection for the adaptability of these animal populations. On chromosome 8 haplo-blocks with the size of more than 50 kb and on chromosome 22 haplo-blocks with size of more than 10 kb were selected and for each haplo-block minimum number of SNPs was set at two. The first and last position of SNP markers was used to search for possible QTLs spanned by the block for each chromosome on the Chicken QTL Database: http://www.animalgenome.org/cgi-bin/QTLdb/GG. Haplotype diversity was considered as the number of haplotypes found within a haplo-block.

4.3 RESULTS

4.3.1 Haplo-blocks characteristics per chromosome

Mean number of haplo-blocks per chromosome was 100.9 and ranged from 13 at chromosome 25 to 402 at chromosome 1 (Table 4.4.1). Haplo-blocks length averaged 18.8 kb and ranged from 6.2 kb to 53.1 kb. Longest haplotypes were observed on chromosome 8 whilst shortest haplotypes were on chromosome 25. Chromosome coverage averaged 0.08 Mbp and high chromosome coverage was observed on chromosomes 8 and 11 and least on chromosomes 3, 6, 9, 24, and 25 (Table 4.4.1). Figure 4.4.1 illustrates the distribution of haplo-blocks of varying lengths across the different chromosomes. Most haplo-blocks were observed between the frequency of 0.25 and 0.5 followed by those that ranged from 0.1 to 0.25

across all chromosomes (Figure 4.4.1) and most haplo-blocks were on macro-chromosomes (chromosomes 1 to 5) and fewer on micro-chromosomes (chromosomes 16 to 28) (Table 4.4.1).

Table 4.3.1: Haplotype characteristics per chromosome.

Chromosome	Blocks	Chromosome Coverage (Mbp)	Mean Block Length (kb)±SD	Median Block Length (kb)±SD	SNPs (%)	Mean nSNPs	Mode SNPs per	Max nSNPs
1	402	0.05	25.8±157.8	11.5±2.7	13.7	2.4	2	54
2	314	0.05	23.5±56.3	12.2±1.2	14.1	2.4	2	12
3	222	0.04	20.2±47.8	12.1±1.1	13	2.3	2	13
4	186	0.065	31.4±210.0	10.9±2.8	13.6	2.4	2	39
5	118	0.06	30±142.3	10.2±3.5	14	2.6	2	39
6	104	0.04	14.4±22.3	10.4±0.6	13.5	2.2	2	5
7	121	0.125	37.5±220.1	12.2±2.7	18.1	2.5	2	31
8	92	0.17	53.1±249.1	12.5±4.1	21.2	3.1	2	31
9	72	0.04	13.5±17.4	12±1.0	13.8	2.2	2	10
10	94	0.06	12.6±17.09	9.5±0.7	16.6	2.2	2	6
11	95	0.14	28.4±118.5	8.2±2.5	20.6	2.6	2	21
12	99	0.07	13.2±25.7	9.8±1.4	18.5	2.3	2	15
13	93	0.08	14.8±23.0	10.6±1.1	19.2	2.3	2	8
14	69	0.09	19.34±41.4	8.9±2.1	18.5	2.7	2	16
15	89	0.11	16.1±21.0	10.3±1.3	22.6	2.5	2	9
17	66	0.11	16.9±28.5	7.5±1.9	21.5	2.7	2	12
18	58	0.06	11.5±22.6	8.6±0.9	15.2	2.2	2	9
19	53	0.11	20±49.4	8.9±1.8	16.5	2.6	2	13
20	125	0.11	12.3±36.4	6.5±2.2	20.7	2.4	2	25
21	63	0.08	8.9±14.8	6.1±1.2	19.5	2.3	2	10
22	16	0.08	21.6±40.7	6.9±2.8	16.9	3.2	2	12
23	26	0.06	12.4±32.3	6.2±1.4	11.4	2.3	2	9
24	34	0.04	8.02±20.0	3.7±1.7	11.4	2.3	2	12
25	13	0.04	6.2±5.1	5.1±0.4	16.6	2.2	2	3
26	32	0.07	11.1±28.5	4.9±1.4	12.5	2.5	2	9
27	27	0.07	13.6±27.6	7.2±1.7	16.4	2.7	2	10
28	41	0.09	10.5±29.22	5.2±1.9	18.7	2.6	2	14
Average ± SD	100.9±89.2	0.08±0.03	18.8±10.4	8.8±2.6	16.6±3.3	2.5±0.26	2±0.0	16.5±12.2

SD – standard deviation

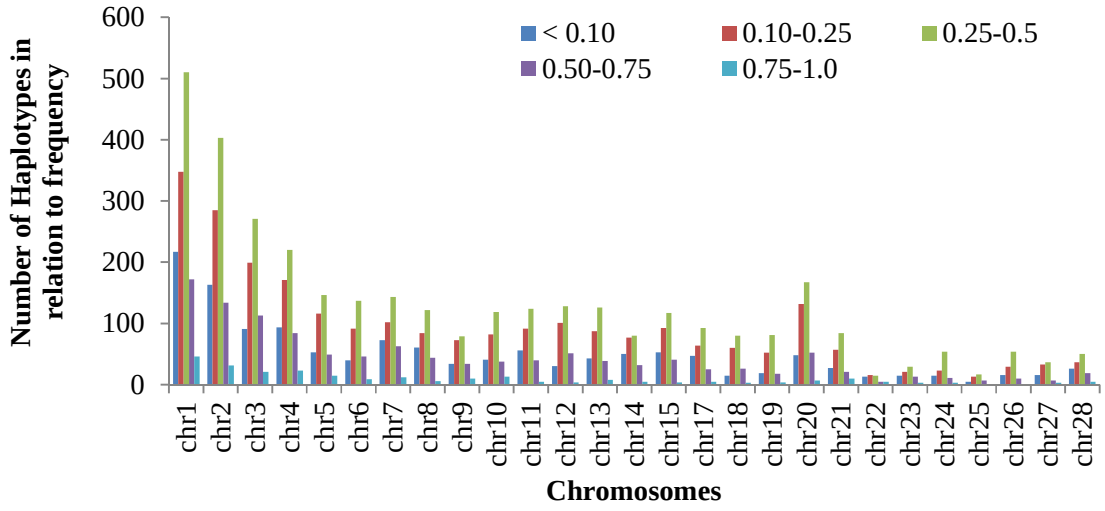

Figure 4.3.1: The distribution of haplotype frequency across the village chicken genome.

4.3.2 Haplotype distribution per population

Number of haplo-blocks was less in Malawi (649) followed by Zimbabwe (2104) and South Africa had the highest of 2442 (Table 4.3.2). Genome coverage was low in Malawi (39 Mbp) followed by Zimbabwe (54.5 Mbp) and higher in South Africa (64.4 Mbp). Average block length was high in Malawi (60.7 kb) and more less the same between Zimbabwe (25.9 kb) and South Africa (26.4 kb) (Table 4.3.2). The percentage of SNPs making blocks was low in all populations with Zimbabwe having SNP percentage of 11.0 and South Africa with 14.1 percent, whilst Malawi had lowest (5.0 % SNP).

Table 4.3.2: Characteristic of haplo-blocks structure for Malawi, South Africa, and Zimbabwe.

	Malawi	South Africa	Zimbabwe
Blocks	649	2442	2104
Genome Coverage (Mbp)	39	64.4	54.5
Mean Block Length (kb)±SD	60.7±239.6	26.4±132.9	25.9±140.5
Median Block Length (kb)±SD	13.6±239.6	10.8±132.9	9.7±140.5
SNPs (%)	5.0	14.1	11.5
Mean nSNPs	3.5±3.9	2.6±2.7	2.6±2.7
Mode SNPs per Block	2	2	2
Max nSNPs	49	55	60

Table 4.4.3 shows the variation of haplotype size per population in which most haplotypes ranged from 10 kb to 25 kb. Haplotype block size of more than 1 MB was observed in all populations as well a few blocks of more than 2 MB across populations.

Table 4.3.3: Number of haplo-blocks in relation to their sizes for each population.

	Malawi	South Africa	Zimbabwe	Overall
< 10 kb	250	1132	1078	1374
10-25 kb	214	1023	792	1115
25-50 kb	44	90	78	64
50-100 kb	56	81	63	83
100-250 kb	63	87	70	71
250-500 kb	13	14	9	10
500-2000 kb	3	4	4	4
>2000 kb	5	5	5	4

4.3.3 Haplotype sharing between populations

Zimbabwean and South African populations shared a relatively large number of haplo-blocks (2689 haplo-blocks) whilst Zimbabwean and Malawian populations shared the least number of 2325 haplo-blocks (Table 4.3.4). The proportion of haplo-blocks shared among populations was less compared to the

total number of 2792 haplo-blocks observed in the overall population. Malawi had 103 unique haplotypes, while Zimbabwe had 256 and South Africa had 467 unique haplotypes.

Table 4.3.4: Haplo-block sharing between populations (Unique haplotypes on the diagonal).

	Malawi	**South Africa**	**Zimbabwe**
Malawi	103	2536	2325
South Africa	2536	467	2689
Zimbabwe	2325	2689	256

4.3.4 Haplotype diversity and characteristics

The chromosomal region in this study was defined as a section or part of chromosome where blocks were found which is in contrast to other cases where a block itself is labeled as a region. Haplotype blocks of less than 20 kb were observed on region 1 followed by region 3. Region 2 had large haplo-blocks (Figure 4.3.2) across populations. Malawi had fewer blocks of less than 10 kb compared to those carried by Zimbabwe and South African chickens. Region 3 exhibited similar patterns of haplo-block structure across all three populations where Malawi had few haplo-blocks with gaps in between them and South Africa and Zimbabwe had a number of large blocks that were close to each other. Haplotype blocks of less than 20 kb were observed more frequently in Zimbabwe and South Africa than in Malawi chickens (Figure 4.3.2). Haplotype blocks sizes of larger than 50 kb were observed across all the genomic regions considered. Variation was observed among genomic region within populations. Overall, Zimbabwean and South African chickens shared similar patterns of haplo-blocks across the genomic regions under investigation. The large blocks in Zimbabwe and South Africa corresponded to regions that had high LD value as observed in Chapter 3.

Selected regions had variations of haplotype within populations. On region 1, at least five haplo-blocks per population were observed where South Africa and Zimbabwe seemed to share similar blocks pattern

having high haplotype diversity within each block as compared to Malawi. Zimbabwe had a different block pattern from that of Malawi and South Africa on region 2 (Figure 4.3.3). Further more, all populations had haplo-block sizes greater than 2000 kb (Table 4.3.3). A number of haplotypes within each blocks had low haplotype frequency and bigger block had high haplotype diversity within and in between populations (Figure 4.3.3). There were few haplotype combinations that had high frequency and majority of the haplotype combinations occurred at low frequencies.

Malawi population had less number of blocks (10 blocks) as compare to Zimbabwe and South Africa, which had 13 blocks on chromosome 8. Haplotype diversity varies between blocks where Zimbabwe had high haplotype diversity on block 5 but very low on block 9, where Malawi had the highest haplotype diversity (Figure 4.3.4).

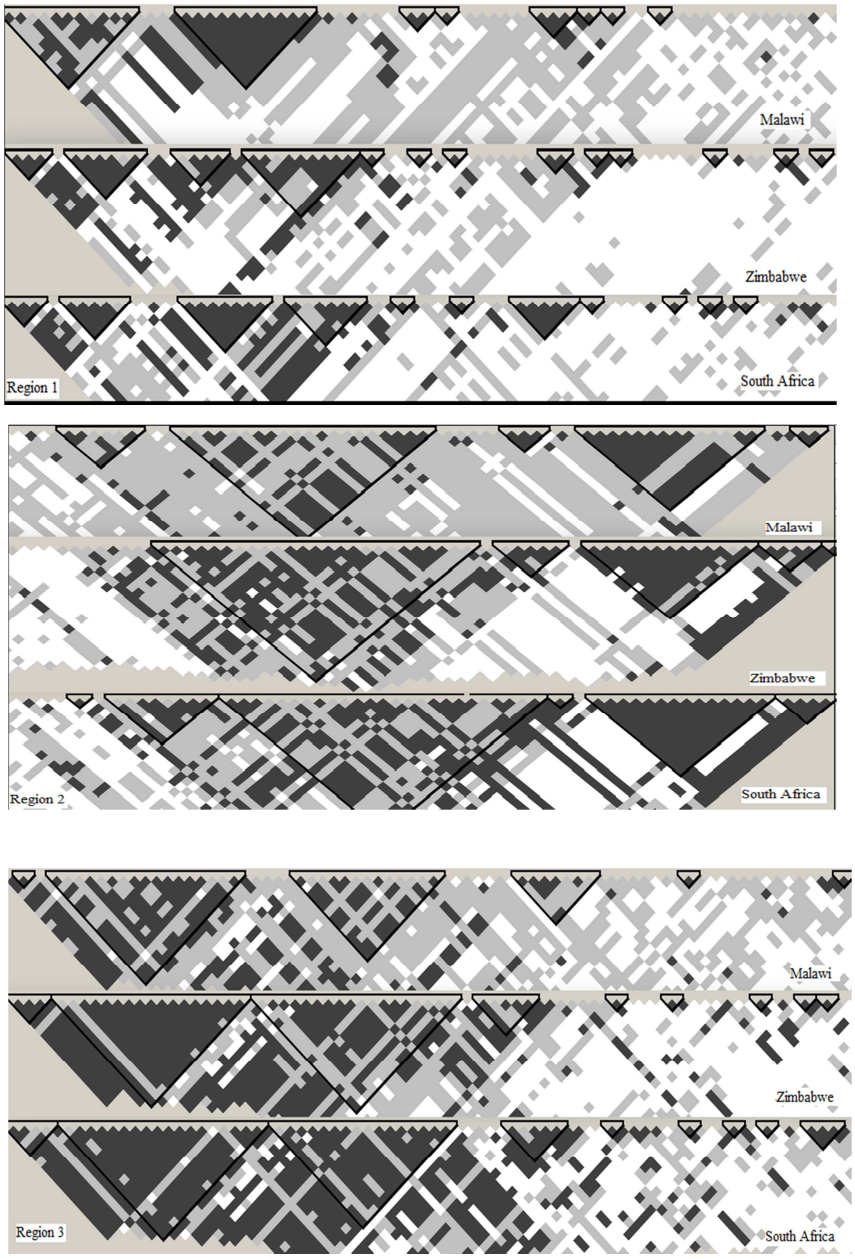

Figure 4.3.2: LD bound haplo-block plots of three chromosomal regions on chromosome 8. Region 1 was found at the edge of *p*-arm, region 2 was found close to centromere of the *p*-arm and region 3 was found on *q*-arm close to the centromere.

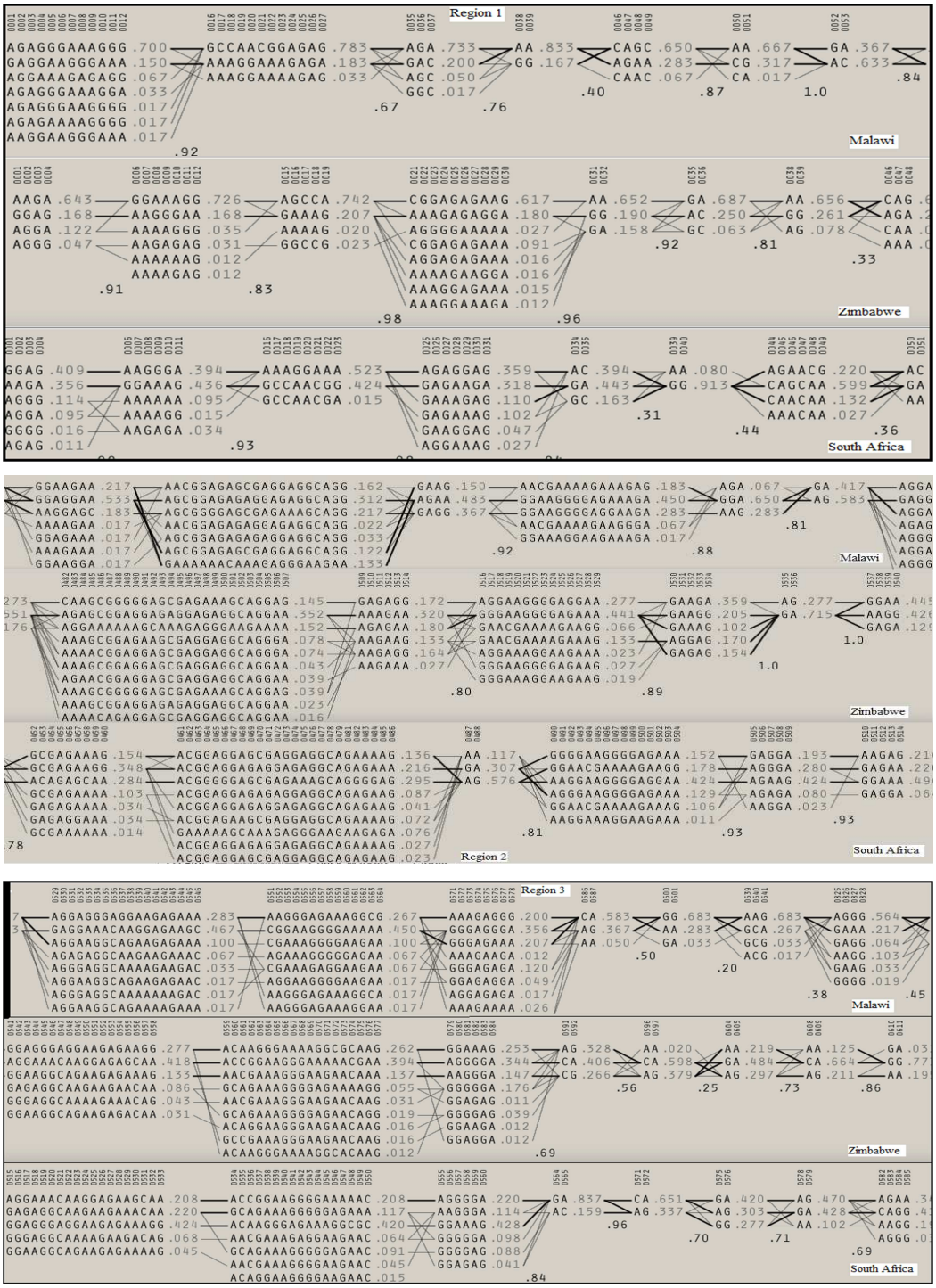

Figure 4.3.3: Haplotype plots of chromosome 8 from three chromosomal portions.

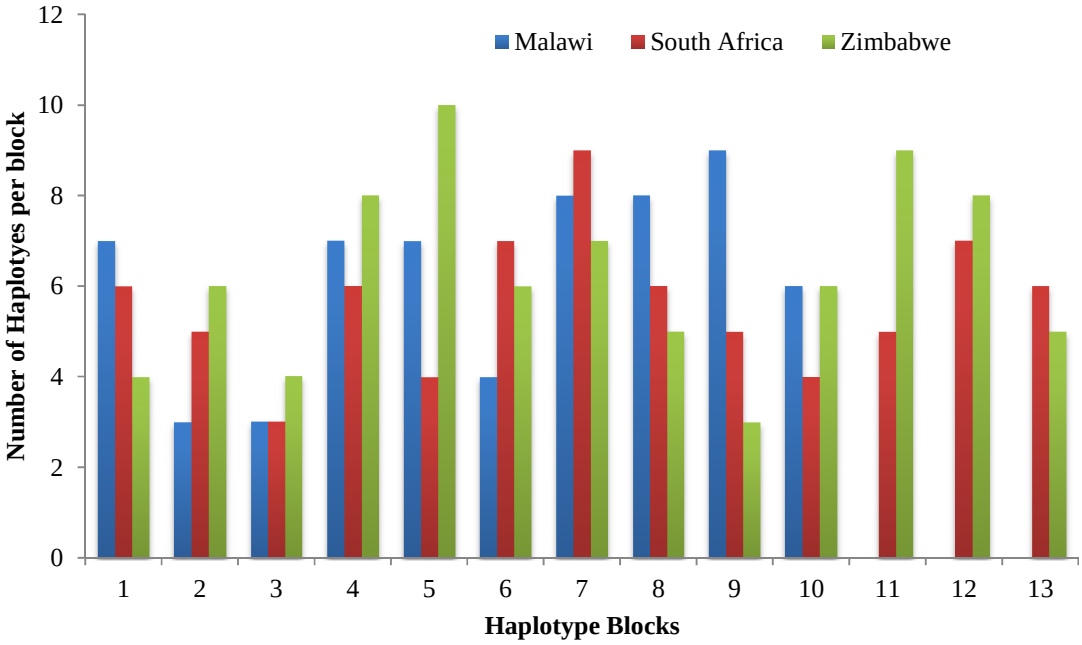

Figure 4.3.4: Haplotype diversity within blocks observed on chromosome 8 between Malawi, South Africa and Zimbabwe extensively raised chicken population.

The haplotype diversity was high on haplo-block 31 (11 haplotypes), 33 (7 haplotypes), 36 (7 haplotypes) and 37 (7 haplotypes) which also corresponded to the size of blocks, 477 kb, 197kb, 486 kb and 269 kb respectively. QTLs on the haplo-blocks around the telomeric region did not have match of QTLs except spleen percentage QTL that was found on Zimbabwean and South African chicken population. The QTLs found were related to body composition traits such as Drumstick muscle weight, Feed conversion ratio, Thigh meat-to-bone ratio, Femur bone mineral density, Breast muscle weight, Body weight (day of first egg), Shank weight to length ratio, Abdominal fat percentage, Head percentage (Table 4.3.5).

Table 4.3.5: Overall haplotype characteristics and *QTLs associated with regions covered by haplo-block Chromosome 8 overall population.

Blocks No.	No. Haplotypes	Location (bp)	Size (kb)	No. SNPs	QTLs
1	5	5597-63486	57.89	4	No match
2	4	129472-221139	91.67	4	SPLP
3	6	286940-492518	196	9	No match
5	5	584764-676818	81	5	No match
30	6	8999746-9090157	90.05	6	DSMWT, FBMD, SHKWLR, TW
31	11	9135643-9613172	477.5	25	DSMWT, TMTBR, FBMD, BMWT, BW, SHKWLR, ABFP, HEADP
32	6	9639311-9770642	131	6	DSMWT, TMTBR, FBMD, BMWT, BW, SHKWLR, ABFP, HEADP
33	7	9818354-10011078	192.7	15	DSMWT, TMTBR, FBMD, BMWT, BW, SHKWLR, ABFP, HEADP
34	5	10016422-10222972	113	5	DSMWT, TMTBR, FBMD, BMWT, BW, SHKWLR, ABFP, HEADP
35	4	11776856-11841126	100	5	DSMWT, TMTBR, FBMD, BMWT, BW, SHKWLR, ABFP, HEADP
36	7	11858695-12359447	486	17	DSMWT, TMTBR, FBMD, BMWT, BW, SHKWLR, ABFP, HEADP
37	7	12429560-12628528	269	10	DSMWT, THBWT, FCR, CHWID, CRDIG, TIBWT, TW, TIBMD, TIBPA, TIBSTR, BMWT, BW, GROW, BMWT, DSMWT, SHKWLR, BMWT, MD, HEADP, WINGWT
38	6	12757762-12838840	81	6	DSMWT, THBWT, FCR, CHWID, CRDIG, TIBWT, TW, TIBMD, TIBPA, TIBSTR, BMWT, BW, GROW, SHKWLR, MD, HEADP, WINGWT
39	5	12923972-12987908	63.9	5	DSMWT, THBWT, FCR, CHWID, CRDIG, TIBWT, TW, TIBMD, TIBPA, TIBSTR, BMWT, BW, GROW, BMWT, DSMWT, SHKWLR, BMWT, MD, HEADP, WINGWT

***List of full names of QTLs on the next page:**

DSMWT	Drumstick muscle weight
THBWT	Thigh bone weight
FCR	Feed conversion ratio
CHWID	Chest width
CRDIG	Crooked digits
TIBWT	Tibia weight
TW	Tibia width
TIBMD	Tibia marrow diameter
TIBPA	Tibia plateau angle
TIBSTR	Tibia strength
BMWT	Breast muscle weight
BW	Body weight (63 days)
GROW	Growth (21-42 days)
BMWT	Breast muscle weight
DSMWT	Drumstick muscle weight
SHKWLR	Shank weight to length ratio
BMWT	Breast muscle weight
MD	Marek's disease-related traits
HEADP	Head percentage
BW	Body weight
WINGWT	Wing weight

4.4 DISCUSSION

The aim of this study was to investigate the haplo-blocks structure in extensively raised village chicken populations using high-density SNP data in order to determine the relationship between the high LD value observed on chromosome 8 and 22 and haplotype diversity along with possible QTLs spanned by haplo-blocks found on these chromosomes. This was done in order to understand whether these chromosomes had region that can constitute of LD associated haplo-blocks that could explain any adaptation qualities observed in extensively raised chicken populations genome. Genomic regions have been investigated in few previous studies working with different commercial chicken lines to analyze the haplo-blocks structure and haplotype diversity using microsatellite markers. As a result few studies e.g. (Megens et al. 2009) are available to make comparisons with. The insight on haplo-blocks on village chicken populations of Southern Africa will shed light into the genomic structure and provide baseline

information for future investigations on haplotype variation in association with phenotypic traits of interest.

Although the number of haplo-blocks was high on the macro-chromosomes, the chromosomal coverage was low indicating presence of many short haplo-blocks in all populations. Average length of haplo-blocks in these regions was 26 kb. Multiple short haplo-blocks imply high recombination events that break the historical ancestral blocks into small segments. The observed results are expected from outbred populations with high genetic diversity that has not been under artificial selection pressures for specific production traits (Wragg et al. 2012). The average median block length of 8.8 kb (Table 4.3.1) observed in this study is comparable with those reported in chickens (Megens et al. 2009), wild dogs (Zhao et al. 2003; Lindblad-Toh et al. 2005), pigs (Amaral et al. 2008) and humans (Gabriel et al. 2002; Wall & Pritchard 2003). The number of SNPs forming blocks was above 25 SNPs in some chromosomes such as chromosomes 1, 4, 5, 7, and 8. Such genomic regions that have long stretch of haplo-blocks should be further investigated for association with morphological or quantitative traits that could be spanned by these blocks. Majority of haplo-blocks were less than 10 kb with some falling between 10 and 25 kb across all populations. Very few haplo-blocks were more than 500 kb. Megens et al. (2009) reported haplo-blocks lengths of less than 10 kb on targeted genomic regions on four chromosomes (chromosome 1, 2, 26 and 27), from both macro- and micro-chromosomes. Haplo-blocks of less than 10 kb were observed in regions of low LD (~0.2) in the presumptively outbred populations (Megens et al. 2009).

A large proportion of haplo-blocks occurred at a frequency greater than 20% in the overall population (Figure 4.3.1). The moderately prevalent haplo-blocks can be used to assess haplotype diversity and genetic variation within and between populations.

The number of observed haplo-blocks varied between populations. Malawi had lower number of haplo-blocks compared to South Africa and Zimbabwe. Sampling structure where by in Malawi animals were sampled from one ecotype whilst three ecotypes per country were used for South Africa and Zimbabwe

might cause this. The number of haplo-blocks in Malawi is similar to those observed by Wragg et al. (2012) in traditional and village chicken populations. Number of haplo-blocks of South Africa and Zimbabwe are similar to those observed by Qanbari et al. (2010a) in commercial lines (Broilers and layers). Overall, high number of haplo-blocks was expected from the Southern African chicken populations that are considered as diverse outbred chickens and have not been selected for any specific traits. Findings from this study shows that South African and Zimbabwean chicken populations have accumulated more blocks compared to Malawi populations. These blocks accumulated by South African and Zimbabwean chicken population are small in size (less than 25 kb). Results suggest higher effective population sizes (see Figure 3.3.4, Chapter 3) and low inbreeding levels for South African and Zimbabwean chickens compared to Malawi chicken populations (as observed in Table 3.3.1, Chapter 3).

The level of haplo-blocks sharing determines the transferability of genetic parameters between populations (Megens et al. 2009). Haplotype sharing in this study varied with a considerable number of haplo-blocks shared between populations. Malawi showed to have more haplo-blocks shared with other populations and only a few haplo-blocks unique to the country. Further analysis into the sizes of haplotypes shared, their genomic content and the frequency between populations can give more insight into genomic regions spanning economic important traits. The variation in number of unique haplo-blocks within population indicates independent genomic sub-structuring and evolution of populations. There is a need to understand what genomic regions are shared between populations with the aim of associating haplotypes with adaptive traits in these extensively raised populations.

Haplo-blocks observed in this study had 10 to 16 folds lower genome coverage in comparison to Qanbari et al. (2010a), where blocks covered more than 300 Mbp of the 1.05 Gbp chicken genome in both broilers and layers, respectively. The median block length and mode SNPs per block were similar to those observed in traditional chicken populations by Wragg et al. (2012). These results indicates that extensively raised chickens still resemble ancestral block length as in other species such as dogs even though they had different domestication history and demographic evolutionary history (Lindblad-Toh et

al. 2005). The number of SNPs forming a block was low in comparison to studies conducted on commercial lines (Qanbari et al. 2010a). Denser SNP panels such as the 600K SNPs chips would probably improve on the number and proportion of SNPs forming haplo-blocks. This will also improve on the LD fine mapping of QTLs and association studies.

On the regions investigated, QTLs have been found to be associated with growth and body composition related traits such as body weight, muscle weight, tibia, wings and thigh size. Results indicate the importance of such traits in village chicken populations. Growth and body composition traits could be key to the adjustment of village chicken populations to fluctuating and scarcity of feed under the extensive systems of management most of these birds are kept under. Village chickens are known to be slow growers (Muchadeyi et al. 2007b), a characteristic that possibly helps them to reduce the risk of failing to cope and survive during periods of feed shortage. The rate of growth and the body weight might therefore associate with the observed haplo-block regions in these chicken populations.

There could be other traits of interest that could be covered by lager haplo-blocks in the southern African chicken genome but attention was drawn to those chromosomes that had high LD values (Chapter 3) which were chromosomes 8 and 22. Chromosome 22 had fewer haplo-blocks that were greater than 10 kb and none that were greater than 50 kb, no QTLs were found on the observed blocks on this chromosome. Most blocks were found around the centromere and telomere region of chromosome 8.

The presence of known QTLs related to body composition on the observed haplo-blocks indicates the possibility of LD fine mapping of QTLs in extensively raised chicken populations using genome-wide high density SNP panels such the Illumina iSelect chickens SNP 50K array. Further investigation of chromosomal regions flacked by large haplo-blocks and QTL association in other regions can be useful for genomics assisted selection to improve fertility, growth and feed conversion efficiency.

4.5 CONCLUSION

The observed optimal number of haplo-blocks had a low genome coverage of which most blocks were of small size. This might limit the use of haplo-blocks in QTL mapping in certain genomic region, however, genomic regions covered by long haplo-block were observed to span QTL regions associated with economic important traits. The haplotype diversity on chromosome 8 and difference in haplo-block structure with and in between populations might indicate biodiversity and evolution imposed by agro-ecological zone and farming system where these animals exist.

4.6 REFERENCES

Abdelqader A., Wollny C.B.A. & Gauly M. (2007) Characterization of local chicken production systems and their potential under different levels of management practice in Jordan. *Tropical Animal Health and Production* **39**, 155-64.

Abubakar M., Ambali A. & Tamjdo T. (2007) Rural chicken production: constraints limiting rural chicken production in some areas of Nigeria and Cameroon. *FAMILY POULTRY*, 57.

Amaral A.J., Megens H.-J., Crooijmans R.P., Heuven H.C. & Groenen M.A. (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* **179**, 569-79.

Andreescu C., Avendano S., Brown S.R., Hassen A., Lamont S.J. & Dekkers J.C. (2007) Linkage disequilibrium in related breeding lines of chickens. *Genetics* **177**, 2161-9.

Ardlie K.G., Kruglyak L. & Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**, 299-309.

Backström N., Qvarnström A., Gustafsson L. & Ellegren H. (2006) Levels of linkage disequilibrium in a wild bird population. *Biology Letters* **2**, 435-8.

Besbes B., Tixier-Boichard M., Hoffmann I. & Jain G. (2008) Future trends for poultry genetic resources. In: *Proc. Int. Conf. Poultry Twenty first Cent.: avian influenza and beyond, Bangkok, Thailand. Avaialble from* ftp://ftp*. fao. org/docrep/fao/011/i0323e/i0323e. pd.*

Cervantes I., Goyache F., Molina A., Valera M. & Gutiérrez J.P. (2011) Estimation of effective population size from the rate of coancestry in pedigreed populations. *Journal of Animal Breeding and Genetics* **128**, 56-63.

Chicken QTL Database (2013) Browse the Chicken QTLdb. URL http://www.animalgenome.org/cgi-bin/QTLdb/GG/browse.

Corbin L., Blott S., Swinburne J., Vaudin M., Bishop S. & Woolliams J. (2010) Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal genetics* **41**, 8-15.

Corbin L., Liu A., Bishop S. & Woolliams J. (2012) Estimation of historical effective population size using linkage disequilibria with marker data. *Journal of Animal Breeding and Genetics* **129**, 257-70.

Crawford D.C. & Nickerson D.A. (2005) Definition and clinical importance of haplotypes. *Annual Review of Medicine* **56**, 303-20.

Cuc N.T.K., Muchadeyi F., Baulain U., Eding H., Weigend S. & Wollny C. (2006) An assessment of genetic diversity of Vietnamese H,Äômong chickens. *International journal of poultry science* **5**, 912-20.

De Roos A., Hayes B.J., Spelman R. & Goddard M.E. (2008) Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics* **179**, 1503-12.

Delany M.E. (2004) Genetic variants for chick biology research: from breeds to mutants. *Mechanisms of development* **121**, 1169-77.

Du F.-X., Clutter A.C. & Lohuis M.M. (2007) Characterizing linkage disequilibrium in pig populations. *International journal of biological sciences* **3**, 166.

Ekue F., Pone D., NFI J.M.A. & NJOYA J. (2006) The effect of vaccination against Newcastle disease and feed supplementation on production in village chicken in Bamenda area of Cameroon. In: *Improving farmyard poultry production in Africa: interventions and their economic assessment: proceedings of a final research coordination meeting*, p. 51. Intl Atomic Energy Agency.

Ekue F.N. P.K.D., Mafeni M.J., Nfi A.N., Njoya J. (2002) Survey of the traditional poultry production system in the Bamenda area, Cameroon. In: *Characteristics and parameters of family poultry production in Africa* (p. 15. FAO/IAEA, Vienna, Austria.

Ensembl (2013) Chicken genome URL http://www.ensembl.org/Gallus_gallus/Location/Genome.

Falconer D.S. & Mackay T.F.C. (1996) *Qauntitative Genetics*. Pearson Education Limited, Harlow, UK.

FAO (1995) Global project for the maintenance of domestic animal genetic diversity (MoDAD). Draft project formulation report. Food and Agriculture Organization of United Nations, Rome.

FAO (2013) *In vivo conservation of animal genetic resources*, Rome.

Flury C., Tapio M., Sonstegard T., Drögemüller C., Leeb T., Simianer H., Hanotte O. & Rieder S. (2010) Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *Journal of Animal Breeding and Genetics* **127**, 339-47.

Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A. & Faggart M. (2002) The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9.

Gondwe T. & Wollny C. (2007) Local chicken production system in Malawi: Household flock structure, dynamics, management and health. *Tropical Animal Health and Production* **39**, 103-13.

Granevitze Z., Hillel J., Chen G., Cuc N., Feldman M., Eding H. & Weigend S. (2007) Genetic diversity within chicken populations from different continents and management histories. *Animal genetics* **38**, 576-83.

Groenen M., Megens H.-J., Zare Y., Warren W., Hillier L., Crooijmans R., Vereijken A., Okimoto R., Muir W. & Cheng H. (2011) The development and characterization of a 60K SNP chip for chicken. *BMC genomics* **12**, 274.

Hall S.J. (2004) *Livestock biodiversity: genetic resources for the farming of the future*. Cambridge University Press.

Hall S.J. & Bradley D.G. (1995) Conserving livestock breed biodiversity. *Trends in Ecology and Evolution* **10**, 267-70.

Hassen H., Neser F., De Kock A. & van Marle-Köster E. (2009) Study on the genetic diversity of native chickens in northwest Ethiopia using microsatellite markers. *African Journal of Biotechnology* **8**.

Hayes B.J., Visscher P.M., McPartlan H.C. & Goddard M.E. (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome research* **13**, 635-43.

Hedrick P.W. (2004) *Genetics of Populations*. Jones and Bartlett, Canada.

Heifetz E.M., Fulton J.E., O'Sullivan N., Zhao H., Dekkers J.C. & Soller M. (2005) Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* **171**, 1173-81.

Hill W. & Robertson A. (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226-31.

Hillier L.W., Miller W., Birney E., Warren W., Hardison R.C., Ponting C.P., Bork P., Burt D.W., Groenen M.A. & Delany M.E. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716.

Illango J. E.A., Olupot H., Mabonga J. (2002) Rural poultry production in two agro-ecological zone of Uganda. In: *Characteristics and parameters of family poultry production in Africa* (p. 117. FAO/IAEA, Vianna, Austria.

Illumina I. (2013) Infinium HD Assay Ultra Protocol Guide. (ed. by Illumina), p. 240, San Diego.

Jia X., Chen S., Zhou H., Li W. & Yang N. (2012) Copy number variations identified in the chicken using a 60K SNP BeadChip. *Animal genetics* **44**, 276-84.

Kaya M. & Yildiz M.A. (2008) Genetic diversity among Turkish native chickens, Denizli and Gerze, estimated by microsatellite markers. *Biochemical genetics* **46**, 480-91.

Keller L.F. & Waller D.M. (2002) Inbreeding effects in wild populations. *Trends in Ecology and Evolution* **17**, 230-41.

Khatkar M.S., Nicholas F.W., Collins A.R., Zenger K.R., Cavanagh J.A., Barris W., Schnabel R.D., Taylor J.F. & Raadsma H.W. (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC genomics* **9**, 187.

Kim E.S. & Kirkpatrick B. (2009) Linkage disequilibrium in the North American Holstein population. *Animal genetics* **40**, 279-88.

Kim K.J., Lee H.-J., Park M.-H., Cha S.-H., Kim K.-S., Kim H.-T., Kimm K., Oh B. & Lee J.-Y. (2006) SNP identification, linkage disequilibrium, and haplotype analysis for a 200-kb genomic region in a Korean population. *Genomics* **88**, 535-40.

Koenig S. & Simianer H. (2006) Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. *Livestock Science* **103**, 40-53.

Kranis A., Gheyas A.A., Boschiero C., Turner F., Yu L., Smith S., Talbot R., Pirani A., Brew F. & Kaiser P. (2013) Development of a high density 600K SNP genotyping array for chicken. *BMC genomics* **14**, 59.

Lee S., Cho Y., Lim D., Kim H., Choi B., Park H., Kim O., Kim S., Kim T. & Yoon D. (2011) Linkage disequilibrium and effective population size in hanwoo korean cattle. *Asian-Australasian Journal of Animal Sciences* **24**, 1660-5.

Li M.H. & Merilä J. (2009) Extensive linkage disequilibrium in a wild bird population. *Heredity* **104**, 600-10.

LI M.H. & Merilä J. (2011) Population differences in levels of linkage disequilibrium in the wild. *Molecular ecology* **20**, 2916-28.

Lindblad-Toh K., Wade C.M., Mikkelsen T.S., Karlsson E.K., Jaffe D.B., Kamal M., Clamp M., Chang J.L., Kulbokas E.J. & Zody M.C. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803-19.

Lu D., Sargolzaei M., Kelly M., Changxi L., Vander Voort G., Wang Z., Plastow G., Moore S. & Miller S.P. (2012) Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Frontiers in Genetics* **3**, 10.

Mapiye C., Mwale M., Mupangwa J., Chimonyo M., Foti R. & Mutenje M. (2008) A Research Review of Village chicken production constraints and Opportunities in Zimbabwe. *Asian-Australasian Journal of Animal Science* **21**, 1680-8.

Megens H.-J., Crooijmans R.P., Bastiaansen J.W., Kerstens H.H., Coster A., Jalving R., Vereijken A., Silva P., Muir W.M. & Cheng H.H. (2009) Comparison of linkage disequilibrium and haplotype diversity on macro-and microchromosomes in chicken. *BMC genetics* **10**, 86.

Moges F., Mellesse A. & Dessie T. (2010) Assessment of village chicken production system and evaluation of the productive and reproductive performance of local chicken ecotype in Bure district, North West Ethiopia. *African Journal of Agricultural Research* **5**, 1739-48.

Mtileni B., Muchadeyi F., Maiwashe A., Chimonyo M., Groeneveld E., Weigend S. & Dzama K. (2011a) Diversity and origin of South African chickens. *Poultry science* **90**, 2189-94.

Mtileni B., Muchadeyi F., Maiwashe A., Groeneveld E., Groeneveld L., Dzama K. & Weigend S. (2010) Domestication of South African chicken genetic resources. In: *Proceedings International*, pp. 03-7, Germany.

Mtileni B., Muchadeyi F., Maiwashe A., Groeneveld E., Groeneveld L., Dzama K. & Weigend S. (2011b) Genetic diversity and conservation of South African indigenous chicken populations. *Journal of Animal Breeding and Genetics* **128**, 209-18.

Mtileni B., Muchadeyi F., Maiwashe A., Phitsane P., Halimani T., Chimonyo M. & Dzama K. (2009) Characterisation of production systems for indigenous chicken genetic resources of South Africa. *Applied Animal Husbandry and Rural Development* **2**, 18-22.

Muchadeyi F., Eding H., Simianer H., Wollny C., Groeneveld E. & Weigend S. (2008) Mitochondrial DNA D-loop sequences suggest a Southeast Asian and Indian origin of Zimbabwean village chickens. *Animal genetics* **39**, 615-22.

Muchadeyi F., Eding H., Wollny C., Groeneveld E., Makuza S., Shamseldin R., Simianer H. & Weigend S. (2007a) Absence of population substructuring in Zimbabwe chicken ecotypes inferred using microsatellite analysis. *Animal genetics* **38**, 332-9.

Muchadeyi F., Sibanda S., Kusina N., Kusina J. & Makuza S. (2005) Village chicken flock dynamics and the contribution of chickens to household livelihoods in a smallholder farming area in Zimbabwe. *Tropical Animal Health and Production* **37**, 333-44.

Muchadeyi F., Wollny C., Eding H., Weigend S., Makuza S. & Simianer H. (2007b) Variation in village chicken production systems among agro-ecological zones of Zimbabwe. *Tropical Animal Health and Production* **39**, 453-61.

Muchadeyi F., Wollny C., Eding H., Weigend S. & Simianer H. (2009) Choice of breeding stock, preference of production traits and culling criteria of village chickens among Zimbabwe agro-ecological zones. *Tropical Animal Health and Production* **41**, 403-12.

Muchadeyi F.C., Weigend S., Wollny C. & Makuza S. (2006) Use of Microsatellites and MtDNA to Assess Genetic Diversity Within and Between Zimbabwe Chicken Eco-types. In: *Bonn*, Tropentag.

Mwacharo J., Bjørnstad G., Mobegi V., Nomura K., Hanada H., Amano T., Jianlin H. & Hanotte O. (2011) Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Molecular phylogenetics and evolution* **58**, 374-82.

Mwale M., Bhebhe E., Chimonyo M. & Halimani T.E. (2005) Use of herbal plants in poultry health management in the Mushagashe small-scale commercial farming area in Zimbabwe. *International Journal of Applied Research in Veterinary Medicine* **3**, 163-70.

Nidup K., Penjor P.D., Gurung R.B., Arasta P. & Moran C. (2005) Read Origin And Genetic Diversity Of Indigenous Chickens of Bhutan As Determined By Mitochondrial DNA. *SAAR Journal of Agriculture* **3**, 69-89.

Nordborg M. & Tavaré S. (2002) Linkage disequilibrium: what history has to tell us. *TRENDS in Genetics* **18**, 83-90.

Okeno T.O., Kahi A.K. & Peters K.J. (2012) Characterization of indigenous chicken production systems in Kenya. *Tropical Animal Health and Production* **44**, 601-8.

Olwande P.O., Ogara W.O., Okuthe S.O., Muchemi G., Okoth E., Odindo M.O. & Adhiambo R.F. (2010) Assessing the productivity of indigenous chickens in an extensive management system in southern Nyanza, Kenya. *Tropical Animal Health and Production* **42**, 283-8.

Prasad A., Schnabel R., McKay S., Murdoch B., Stothard P., Kolbehdari D., Wang Z., Taylor J. & Moore S. (2008) Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Animal genetics* **39**, 597-605.

Purcell S.N., B, Todd-Brown K., Thomas L., Ferreira M., Bender D., Maller J., Sklar P., de Bakker P. & PC D.M.S. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. URL http://pngu.mgh.harvard.edu/purcell/plink/.

Qanbari S., Hansen M., Weigend S., Preisinger R. & Simianer H. (2010a) Linkage disequilibrium reveals different demographic history in egg laying chickens. *BMC genetics* **11**, 103.

Qanbari S., Pimentel E., Tetens J., Thaller G., Lichtner P., Sharifi A. & Simianer H. (2010b) The pattern of linkage disequilibrium in German Holstein cattle. *Animal genetics* **41**, 346-56.

Rao Y.S., Liang Y., Na Xia M., Shen X., Jun Du Y., Glong Luo C., Hua Nie Q., Zeng H. & Quan Zhang X. (2008) Extent of linkage disequilibrium in wild and domestic chicken populations. *Hereditas* **145**, 251-7.

SAS (2011) SAS User's Guide: Statistics. (ed. by Inc. SI). Cary, North Carolina, USA.

Schwartz M.K., Tallmon D.A. & Luikart G. (1998) Review of DNA-based census and effective population size estimators. *Animal Conservation* **1**, 293-9.

Slatkin M. (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**, 477-85.

Tenesa A., Navarro P., Hayes B.J., Duffy D.L., Clarke G.M., Goddard M.E. & Visscher P.M. (2007) Recent human effective population size estimated from linkage disequilibrium. *Genome research* **17**, 520-6.

Thevenon S., Dayo G.-K., Sylla S., Sidibe I., Berthier D., Legros H., Boichard D., Eggen A. & Gautier M. (2007) The extent of linkage disequilibrium in a large cattle population of western Africa and its consequences for association studies. *Animal genetics* **38**, 277-86.

Tishkoff S.A. & Verrelli B.C. (2003) Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. *Current opinion in Genetics and Development* **13**, 569-75.

van Marle-Köster E., Hefer C.A., Nel L.H. & Groenen M. (2008) Genetic diversity and population structure of locally adapted South African chicken lines: Implications for conservation. *South African Journal of Animal Science* **38**, 271-81.

Wall J.D. & Pritchard J.K. (2003) Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics* **4**, 587-97.

Wang J. (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1395-409.

Waples R.S. (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379-91.

Wragg D., Mwacharo J., Alcalde J., Hocking P. & Hanotte O. (2012) Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. *Heredity* **109**, 6-18.

Zhao H., Pfeiffer R. & Gail M.H. (2003) Haplotype analysis in population genetics and association studies. *Pharmacogenomics* **4**, 171-8.

CHAPTER 5

GENERAL DISCUSSION & CONCLUSION

The availability of modern genomic tools and powerful statistical methods has made it possible to investigate genetic and demographic parameters that are at play in shaping up the existing genetic resource of extensively raised chicken populations in the absence of pedigree and performance data.. The possible impact of the management systems under which the extensively raised chicken populations are kept on the genomic architecture, diversity and risk to extinction have not been fully evaluated. An understanding of the inbreeding levels, effective population size and possible genetic structure supporting adaptation features could facilitate effective conservation and proper breeding programs. The small flock sizes, random mating systems and lack of supporting genetic parameters are thought to influence genetic variation and the evolution of village chickens. It is important to quantify the levels of genetic parameters that contribute to genetic variation and evolution of the village chicken populations. The investigation of linkage disequilibrium (LD), effective population size and haplo-block structure has not been done on Southern African village chicken populations. This study sought to provide a platform for the use of SNP data in extensively raised chicken populations to understand their demographic and genetic parameters that exists and shaping the genetic structure of these populations. The estimation of LD and the use of LD-based methods to measure effective population size and detection of haplotypes could unravel the genetic architecture of the village chickens.

In Chapter 3, variation in LD levels between populations and chromosomes was observed. In general, all populations showed similar trends of LD decay. Chromosome 8 and 22 were observed to have high LD levels of 0.52±0.26 and 0.54±0.3, respectively whilst the chromosome 13 and 18 showed lower levels of LD of 0.37±0.2 and 0.34±0.2, respectively compared to the chromosomal average of 0.42. A chromosome by population interaction was however observed whereby the South African and Zimbabwean chickens had similar patterns of LD decay that was different from Malawi. Results indicated that even though

population did not influence LD, genomic selection and adaptation that shape genetic structure of these animals differed between chromosomes at a country level. LD was further used to estimate effective population size over time and all populations showed a constant decline of effective population size over the past generations showing that less than 20 individuals contributed to the current genetic pool. Even though it has been assumed that farmers who keep village chickens do not impose selection pressure for breed development, some decisions taken in these extensive production systems could contribute towards genetic variability (gain and loss of genetic diversity) such as uncontrolled mating systems, sharing of chickens and culling .

In chapter 4, we sought to evaluate whether haplo-blocks on chromosomes that had high LD (chromosome 8 and 22) could be spanning QTLs for traits of economic importance, and supporting isolated evolution of the different populations from different geographic and production systems. Results showed that populations shared a majority of haplo-blocks and that most haplo-blocks had a frequency ranging between 0.2-0.5. The number of haplo-blocks varied per population with the majority of haplo-blocks observed being small in size of > 15 Kb with fewer larger haplo-blocks of < 1Mb. However, the median block length of 10 kb was comparable to that observed in other species such as in humans and pigs. There is a need to use a denser SNP panel for a better haplotype resolution. However, the transferability of genetic tools is important for the investigation of population genetic parameters and the haplotype sharing between these populations indicate the potential use of haplotype for investigation of population parameters in village chickens. Haplotype blocks observed on chromosome 8 were highly diverse having few haplotypes combination that had high frequency and a number of haplotypes combinations occurring at low frequency. The average size of the haplo-block was 126 kb. Some of these haplo-blocks were spanning QTL regions that are associated with body composition traits such as body weight, feed conversion ratio, thigh muscle, wing weight, shrunk size, breast weight and abdominal fat percentage. These traits are found on regions with low or no recombination events indicating the importance of conservation of these traits for adaptation to survive under the smallholder farming system.

The extent and distribution of LD across the genome was in support on the observed haplo-block with high confidence bound to be in LD. The observed haplo-blocks spanning over coding regions and none coding regions can be a useful tool in identifying conserved regions of economic importance.

In conclusion, this study has shown the utility of 60K SNP chip panel to investigate demographic parameters of village chicken populations raised by communal farmers in the different agro-ecological zones of Southern Africa. The geographic origin of animals did not have significant difference on LD measure, and the reduced level of heterozygosity was in support of reduced effective population size. This indicated that different agro-ecological zones and farming system have similar impact on the gene pool of the studied population. Farming systems in which communal farmers keep these animals have an influence on reduced genetic diversity and proper management needs to be practiced to conserve them. The characterized haplo-block showed some similarities in between populations and high genetic variation. The observed QTL spanned by haplo-blocks can be used in association with performance records and extract haplotypes that can be associated with production performance for genetic improvement programs.

APPENDIX

Proc. Assoc. Advmt. Anim. Breed. Genet. 20:364-367

THE EXTENT AND DISTRIBUTION OF LINKAGE DISEQUILIBRIUM IN EXTENSIVELY RAISED CHICKEN POPULATIONS OF SOUTHERN AFRICA

K.S. Khanyile[1,2], E. F. Dzomba[2] and F. C. Muchadeyi[1*]

[1]Biotechnology Platform, Agricultural Research Council, P Bag X5, Onderstepoort, 0110, South Africa
[2]University of KwaZulu-Natal, Discipline Genetics, School of Life Sciences, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa
Correspondence: MuchadeyiF@arc.agric.za

SUMMARY

The amount of linkage disequilibrium (LD) is an important source of information about historical events of recombination and allows inferences about genetic diversity and genomic regions that have undergone selection. Linkage disequilibrium is equally important in studying effective population size and rate of inbreeding particularly in extensively raised and wild animal populations where pedigree records are scarce. The objective of this study was to investigate LD in village chicken populations of Southern Africa. These chickens are raised under scavenging systems of production characterized by uncontrolled breeding and frequent population bottlenecks due to disease outbreaks and fluctuations in feed supplies. DNA samples from 312 extensively raised chickens from South Africa, Malawi and Zimbabwe were genotyped using the Illumina iSelect chicken SNP60K BeadChip. A panel of 43,157 out of the total 57,636 (74.8%) SNPs was used in the final analysis after screening for those that had a minor allele frequency of less than 5%, were out of Hardy-Weinberg equilibrium ($P<0.01$) and had a call rate of less that 95%. Results indicated that LD averaged between 0.45 and 0.58 for SNPs that had a pairwise distance of less than 20 kb. LD dropped to 0.34 for SNPs between 20 and 100 kb after which it remained constant. LD was further analyzed for its decay over marker distance and differences between populations from different geographic locations. Results are discussed in terms of historical changes in effective population size and resultant recombination rates. The utility of the iSelect chicken SNP60K beadchip in investigating free-range chicken population genetics is demonstrated.

INTRODUCTION

Linkage disequilibrium (LD) is defined as a non-random association of alleles at two or more loci (Hendrik 2005; Qanbari *et al*. 2010). The importance of LD is in providing information about historical events of recombination thereby explaining genetic diversity in genomic regions undergoing selection. LD also allows estimation of effective population size and rate of inbreeding in extensively raised and wild animal populations without pedigree records (Wragg *et al*. 2012).

The village chicken production system in Africa is mainly based on scavenging village chickens (Kitalyi 1998), that are used to meet the multiple household social, economic and cultural needs and are crucial to biodiversity (Delany 2003). However, very little is known about the genetic composition of village chickens in developing regions like Southern Africa. Diversity studies using autosomal microsatellite (Muchadeyi *et al*. 2007) and mtDNA sequences (Mtileni *et al*. 2011) have not defined the genetic stability of these populations. Demographic population parameters such as effective population size and inbreeding levels, that influence the risk to extinction of these populations, remain uncharacterized due to the absence of pedigree and other population census records in these village chicken production systems. The availability of large-scale sequence data in chickens has resulted in an increase in the marker density and achieved a

comprehensive SNP coverage of the chicken genome. The chicken 60K SNP genotyping chip has the potential to unravel the genetic information in extensively raised chicken populations. Applying LD analysis will permit estimation of demographic and evolutionary parameters of these populations. The aim of this study was to investigate the extent and distribution of LD in extensively raised chicken populations of South Africa, Zimbabwe and Malawi using the Illumina iSelect chicken SNP60K BeadChip.

MATERIALS AND METHODS

Chicken populations, blood collection and DNA isolation. A total of 312 village chicken samples were collected from South Africa (n =147), Malawi (n = 30) and Zimbabwe (n = 135). In South Africa, village chickens representing Limpopo (n = 15), Eastern Cape (n = 26) and Northern Cape (n = 35) populations, and four conservation flocks of the Naked Neck (n = 20); Potchefstroom Koekoe (n = 20); Ovambo (n = 10) and Venda (n = 20) chickens kept at Agriculture Research Council Poultry Breeding Resource, were sampled as described in Mtileni et al. (2011). The sampling of the village chickens from Zimbabwe (n = 135) and Malawi (n = 30) populations is described in Muchadeyi et al. (2007). Blood was collected from the selected chickens onto FTA Micro Cards (Whatman Bio Science, UK) and DNA was isolated using a modified protocol of the Qiagen® DNA blood and tissue kit.

SNP genotypes and quality control. The chicken DNA samples were genotyped using the iSelect chicken SNP60K bead chip produced by Illumina Inc. SNP quality control was done using Plink (1.07) software to remove SNPs that were either out of Hardy-Weinberg equilibrium (HWE) *(P< 0.01)*, showing a minor allele frequency (MAF) of at least 5%, had low call rate (< 95%) and with missing genotypes (> 5%). SNPs that were on unknown chromosomes, mtDNA, linkage groups and/or sex chromosomes were excluded from further analyses. After filtering, 45676, 44667,46905 and 43157 SNPs on 28 autosomal chromosomes were used for each of the Malawi, South Africa, Zimbabwe and combined populations, respectively.

Linkage Disequilibrium analysis. A pair-wise LD (r^2) was estimated using PLINK (1.07) software for SNPs on chromosome 1 to 28 for the individuals belonging to the three populations using the following formula:

$$r^2 = \frac{(f11f22 - f12f21)^2}{fA1fA2fB1fB2}$$

A Generalized Linear Model procedure (Proc GLM) in the Statistical Analysis System (SAS) was used to determine the effects of SNP marker interval (bp), chromosome, and population group and interaction of chromosome-by- population on the decay of LD using the following model:

$$r^2_{ij} = \mu + Pop_i + Gga_j + (Pop \times Gga)_{ij} + b\,SNPint + e_{ik},$$

Where: Pop_i was the effect of *i*th chicken population of either, Malawi, Zimbabwe or South Africa; Gga_j was the effect of the *j*th chromosome 1-28; and SNPint the effect of SNP interval fit as a covariate with *b* the regression coefficient.

RESULTS AND DISCUSSION

Effects of chicken population, chromosome and distance between SNPs on LD. LD was calculated on 28 of the 38 chicken autosomes. The chromosome size, SNP interval distance and number of SNPs per chromosome support the differences between macrochromosome 1-5 that had high number of SNPs and large intervals between SNPs and micro-chromosomes 16-28, which are smaller and had less SNPs that were relatively close together (Megens *et al.* 2009). Linkage disequilibrium (r^2±SD) averaged 0.38 ± 0.20 and ranged from 0.34 ± 0.14 -0.45 ± 0.24 in Malawi,

0.34 ± 0.15 - 0.52 ± 0.27 in Zimbabwe and 0.34 ± 0.14 - 0.50 ± 0.27 in South African chicken populations. Overall, there was no significant difference in r^2 values (*P<0.05*) between populations indicating similarities between the Malawian, Zimbabwean and South African village chicken populations. However, LD varied significantly between chromosomes (*P<0.001*) with chromosome 8 having the highest LD of 0.52 ± 0.26 followed by chromosome 22 with an r^2 ± SD value of 0.49 ± 0.28. The high LD might be an indication of selection at genes on these chromosomes (Hendrick 2005) particularly natural selection pressures as these chicken populations are raised under extensive systems of production where human selection pressures are minimal (Mtileni *et al*. 2010). Although population did not influence genome-wide LD, a population by chromosome interaction was observed whereby the Zimbabwean chicken population had the highest LD on chromosome 8 (0.52 ± 0.267) and the South African chicken population was highest on chromosome 22 (0.49 ± 0.29). Such interactions need to be further investigated as they might indicate different selection pressures in different populations (Wragg *et al*. 2012).

Another factor that influenced LD was the SNP interval. To further understand this, LD was computed at different distance interval of 0-1 kb, 1-10 kb, 10-20 kb, 20-40 kb, 40-60 kb, 60-100 and 100kb plus using SNP data from chromosomes 1-28 (Fig 1a) and from chromosomes 8; 22 and 13 as indicated in Figures 2b, c and d respectively.

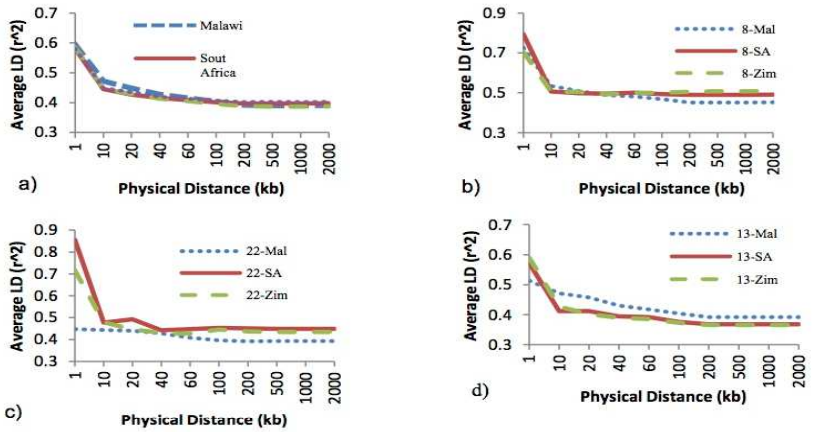

Figure 1. Average LD decay with an increase in physical distance between SNPs for a) chromosomes 1-28, b) chromosome 8; c) chromosome 22; and d) chromosome 18.

The LD averaged 0.58 for SNPs within a 10 kb interval and decayed to 0.45 -0.47 for SNPs between 10-30 kb after which they remained constant. The LD decay at chromosome 8 of the Malawi chickens continued to decline after 40kb. In Zimbabwe and South African chickens, LD at chromosome 22 made a sharp decay from 0.7 (Zimbabwe) and 0.85 (South Africa) to an r^2 below 0.5 at 10kb after which it stayed constant. On the same chromosome LD was maintained around 0.45 over all sliding windows in the Malawi chicken population.

Overall, a higher LD was observed in the Southern African chicken populations compared to

other chicken populations observed in other studies (Qanbari *et al*. 2010; Wragg *et al*. 2012). For example, in a commercial egg laying flock, r^2 averaged 0.32 ± 0.33 with a minimum 0.21 ± 0.26 (Qanbari *et al*. 2010) whereas it was maintained around 0.38 in this study.

CONCLUSION

A relatively high LD that persisted over long SNP intervals was observed in the South African, Zimbabwean and Malawian chicken populations. This LD pattern seems to be consistent with low and steady effective population sizes. The study recommends for a further investigation on the role of selection and population bottlenecks on chromosomes 8 and 22 that had significantly high LD.

ACKNOWLEDGEMENTS

Genotyping of samples was funded by the Agricultural Research Council-Biotechnology Platform (ARC-BTP) while Mr KS Kanyile held fellowships from the ARC-Professional Development Program and the University of KwaZulu-Natal postgraduate program.

REFERENCES

Delany M.E., (2004) *Mec. of Dev*. **121**: 1169.

Hendrick P. W. (2005) Genetics of populations 3rd. Jones and Bartlett publishers, Canada

Illumina Technical notes: DNA analysis, Analyzing GoldenGate Genotyping Data

Kitalyi A. J. (1998). Village chicken production systems in rural Africa: Household food security and gender issues. FAO Animal Production and Health Paper 142, Rome, Italy.

Megens H., Crooijmans R.P.M.A., Bastiaansen J. V. M., Kersten H. H. D., Coster A., Jalving R., Vereijken A., Silva P., Muir W. M., Chen H. H., Hanotte O. and Groene M. A. M. (2009) *BMC Genetics*. **10**: 86.

Mtileni B. J., Muchadeyi F. C., Maiwashe A., Groeneveld E., Groeneveld L. F., Dzama K. and Weigend S. (2011) *Journal of Animal Breeding and Genetics*. **128**: 209.

Mtileni B. J., Muchadeyi F. C., Maiwashe A., Phitsane P. M., Halimani T. E., Chimonyo M. and Dzama, K. (2010) *Applied Animal Husbandry and Rural Develop*. **2**:18.

Muchadeyi F. C., Eding H., Wollny C. B. A., Groeneveld E., Makuza S. M., Shamseldin R., Simianer H. and Weigend S. (2007) *Animal Genetics*. **38**: 332.

Qanbari S., Hansen M., Weigend S., Rudolf P. and Simianer H. (2010) *BMC Genetics*. **11**: 471.

Wragg D., Mwacharo J., Alcalde J., Hocking P. and Hannotte O. (2012) *Heredity*. **109**: 6.