

Identifying molecular signatures from sex chromosomes: a case for indigenous chicken based on the analyses of the z chromosome

Abstract

Domestications and breeding of chicken have left traceable molecular selection signatures within their genomes. In this study, a detailed Z-chromosome-wide genomic scan using whole genome re-sequenced data was performed to address the genetic basis of differences between indigenous and exotic chicken. The data were processed individually through the similar statistical quality protocol and then merged, resulting in 1,826,332 high-quality SNPs. Results showed SNP density between 66.11 to 142.33 per 50kb. Comparison analysis detected five genes (*FOXEL1*, *FOXEL3*, *LOC107052388*, *LOC101749568*, and *ENSGALG00000046172*) within the regions affected by selection, out of which two genes were associated with missense variants. The genes within the selection hotspots are associated with metabolism, muscle development, response to starvation and reproduction. The study suggests that the Z chromosome has undergone selection and have a contribution in the divergence between indigenous chicken and commercial layer populations.

Keywords: signatures of selection, sex chromosome, indigenous chicken

Volume 4 Issue 3 - 2017

Ngeno K, Khobondo JO

Department of Animal Sciences, Animal Breeding and Genomics Group, Egerton University, Kenya

Correspondence: Kiplangat Ngeno, Animal Breeding and Genomics Group, Department of Animal Sciences, Egerton University, PO Box 536, 20115 Egerton, Kenya, Tel 254708246317, Email aarapngeno@gmail.com

Received: May 26, 2017 | **Published:** December 11, 2017

Abbreviations: Fst, f-statistics; ZFst, transformed fst; BWA, burrows-wheeler aligner; SNP, single nucleotide polymorphism; MAF, minor allele frequency; HW, Hardy-Weinberg equilibrium; Ts, transition; Tv, transversion ratio; kb, kilobases; VCF, variant call format

Introduction

Domestications and selection have led to phenotypic variation in chicken and has left traceable signatures within their genome. Indigenous chicken has been shaped more by natural selection¹ imposed by the environmental conditions, parasites, diets and diseases. Several candidate genes that contribute to adaptation to harsh scavenging tropical conditions, for instance, those linked to response to endoplasmic reticulum stress and cell redox homeostasis (e.g. *PDIA4* gene), suppression of viral infection and mediation of inflammatory responses (*MAPK13*), have been discovered.² Nevertheless, most of the detection of selection signatures is based on the genomic scans of autosomes, though chicken chromosome hosts both autosomes and allosomes. Allosome (sex chromosome) is more specialized than an autosome and plays a significant role in evolution.³ Particularly, allosome has been suggested to be a good target for detection of signatures of selection because of its high gene density and undergoes more drift than autosomes.⁴ In pigs, a selective scan of the X chromosome revealed genes contributing to immunity, reproduction and meat quality, in the potential regions affected by selections. In sheep, the X chromosome hosts several genes associated with tail fat deposition.⁵ However, only a few genomic studies have scanned the allosome of the chicken. In this study, a Z-chromosome-wide genomic scan using F-statistics (Fst) method was performed to search for genomic selection signatures in chicken kept under low input production system compared to commercial layers. Identification of the regions affected by selection offers the opportunity of understanding the forces of historical selection, genes,

and mutations underlying a phenotype difference between indigenous and commercial chicken.

Materials and methods

Birds

A total of eight indigenous chickens from eight different regions in Kenya; Kakamega and Siaya in the Western region, West Pokot, and Turkana in North Rift Valley, Bomet and Narok in South Rift Valley, Lamu and Taita-Taveta in the coastal region were sampled. Sequences of two unrelated individuals of a commercial exotic layer chicken line were included in the study.

DNA isolation, reads mapping and SNPs calling

Genomic DNA was obtained by standard phenol-chloroform extraction. Library preparation and 100bp paired-end sequencing were performed according to manufacturer's protocol (Illumina) on the DNA samples. Short reads were trimmed using Sickle Version 1.33.⁶ The reads were aligned to the reference chicken genome (release: galGal4) using Burrows-Wheeler Aligner (BWA) v0.7.13 (Li and Durbin, 2009). SAM tools v0.1.31⁷ was used to obtain the variant call format (VCF) files.⁷ Variations were filtered for a minimum genotype SNP quality of 20. Single nucleotide polymorphism (SNP) with a call rate less than 90%, minor allele frequency (MAF) less than 0.05 and Hardy-Weinberg equilibrium (HWE) below p-value of 10^{-6} were filtered out. INDELs were removed using VCF tools v0.1.13.⁸

Selection scans on the Z chromosome

The eight indigenous chickens were grouped into a single population and the two commercial layers in another group. Fixation index (Fst) values between indigenous chicken and EL were estimated using VCF tools in bins of a 40kb window sliding 20kb over the entire chromosome. Out of 39,498 windows, nine windows which

had less than 10 variants were filtered out, to minimize the number of false positives. The F_{st} values were normalized (ZFst) using Z-transformation method.⁹ One percent top level F_{st} was qualified as putative selection regions. The package ggplot2 in R environment (<http://www.r-project.org/>) was used to display the distribution of F_{st} values.

Functional annotation of genomic variants and gene ontology

Functional annotations of genomic variants in the regions affected by selection were determined using Variant Effect Predictor.¹⁰ The option of SIFT predictions within VEP was used to predict if the substitution of amino acid has an effect on protein function using the annotations found in Ensembl 87. Missense variants within the elevated F_{st} regions were extracted and their associated genes analyzed for the Gene Ontology (GO) enrichment. BinGO v3.0.3 within Cytoscape v3.4.0¹¹ was used to check overrepresentation of GO terms enrichment for the biological process using UniProt-GOA for chicken.¹² Benjamini and Hochberg correction method designed for multiple comparisons was used to test significance at 0.05.

Table 1 Number of SNPs, SNP density and transition/transversion (Ts/Tv) ratios

Population	SNPs	SNP density per 50kb	SNP density per kb	Ts/Tv ratio
Bomet	222,448	134.9	2.7	2.01
Kakamega	234,708	142.33	2.85	1.97
Lamu	217,631	131.98	2.64	1.92
Narok	205,231	124.46	2.49	1.94
Siaya	197,900	120.01	2.4	1.88
Turkana	231,724	140.52	2.81	1.97
Taita-taveta	171,377	103.93	2.08	1.9
West-Pokot	132,508	80.36	1.61	1.83
Exotic layer 1	109,023	66.11	1.32	1.82
Exotic layer 2	103,782	62.94	1.26	1.79

SNP-single nucleotide polymorphism; Ts-transition; Tv-transversion ratio; kb-kilobases

Selective footprints of the Z chromosome

Figure 1 shows the distribution of the F_{st} values on the Z chromosome. Putative selection signatures were identified by searching genome sites with high F_{st} values. Outliers were detected based on a total of 1,826,332 SNPs from the two populations using a threshold of 1% top-level F_{st} values ($F_{st} > 2.5$ standard deviation). A total of 753 SNPs were detected as outliers. The outlier genomic regions overlapped with five genes (*FOXEL*, *FOXEL3*, *LOC107052388*, *LOC101749568*, and *ENSGALG00000046172*), out of which two genes were associated with missense variants (Table 2). Identification of the regions affected by selection offers the opportunity of understanding the forces of historical selection; farmer's breeding objectives, genes and mutations underlying bird's phenotypic differences. Most genomic studies have regularly omitted the allosome in the detection of molecular signatures of selection since it represents a different effective population size. In this study, the population genetic variation at each SNP in the Z chromosome was detected using F_{st} test. The F_{st} values were calculated between local populations from Kenya and commercial layers for the detection of selection signatures. Indigenous chicken and commercial exotic

Results and discussion

Summary of mapping and SNPs calling

After filtering out low-quality SNPs, SNPs with a call rate less than 90%, MAF less than 0.05 and HWE below p-value of 10^{-6} , an average between 103,782 and 234,708 high-quality SNPs per individual remained (Table 1). The SNP density in the populations ranged from 66.11 to 142.33 per 50 kb and 1.32 to 2.85 per kb (Table 1). This study presents vital insights for understanding the role of sex chromosomes in the genetic diversity. Compared to results from autosomes by Ngeno,² the genetic diversity assessments established that the Z chromosome had lower genetic diversity across all the populations. The SNP density per kilobase indicated abundant genetic variation within indigenous chicken genome compared to commercial layer (Table 1). Transition/Transversion ratio (Ts/Tv) were 1.82 to 2.01 (Table 1). The observed differences in Tv/Ts revealed the different frequency with which substitutions occur in the populations. Such in-depth information on the Z-chromosome genetic diversity is vital for genetic improvement.

layers represent chicken kept in two different environments (tropical extensive vs temperate intensive production systems) and with different breeding goals. The impact of such diverse production conditions and selection pressure for egg production in commercial layers has acted on the commercial layer and indigenous chicken differently resulting in the diverse genetic patterns.

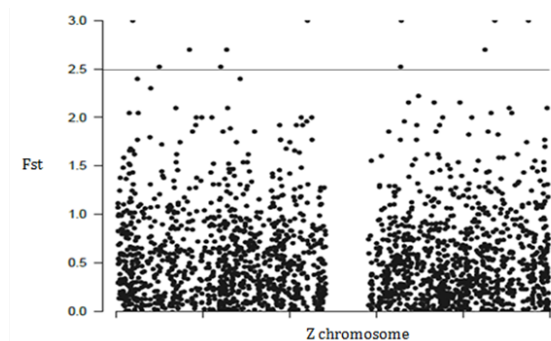


Figure 1 Distribution of the F_{st} values along the Z chromosome.

Table 2 Regions affected by selection on the Z chromosome and the genes associated with missense variants

Position	Allele	Gene	Biotype	cDNA position	CDS position	Protein position	Amino acids	Codons
70689749	G	ENSGALG00000046172	Protein coding	463	463	155	T/A	Acc/Gcc
70689950	A	ENSGALG00000046172	Protein coding	664	664	222	G/S	Ggt/Agc
70689954	A	ENSGALG00000046172	Protein coding	668	668	223	V/E	gTa/gAa
70689987	C	ENSGALG00000046172	Protein coding	701	701	234	R/P	cGa/cCa
70689749	G	LOC101749568	Protein coding	3563	463	155	T/A	Acc/Gcc
70689950	A	LOC101749568	Protein coding	3764	664	222	G/S	Ggt/Agc
70689954	A	LOC101749568	Protein coding	3768	668	223	V/E	gTa/gAa
70689987	C	LOC101749568	Protein coding	3801	701	234	R/P	cGa/cCa

Functional annotation of genomic variants and gene ontology

Based on the annotations, the consequences for most of the variants in the regions affected by selection were introns (46%) (Figure 2). Missense variants contributed about 1%. Gene Ontology (GO) enrichment analyses of the *LOC101749568* and *ENSGALG00000046172* candidate genes linked with missense variants associated them with metabolism, muscle development, and reproduction. Missense variants in the high *Fst* regions were studied since they change the sequence of the amino acid result in phenotypic differences.¹³ Five genes overlapped with the high *Fst* peaks, indicating they have been selected in opposing direction. These genes included the avian *FOXE1* that is expressed in growing feathers and play role in the regulation of the evagination and morphogenesis of thyroid.¹⁴ *FOXE3* gene play role in cellular response to starvation, glucose homeostasis and spermatogenesis processes. Comparison analysis detected two genes (*ENSGALG00000046172* and *LOC101749568*) associated with missense variants within the selection hotspots. The two genes have to change the sequence of amino acid, resulting in phenotypic differences. Indigenous chicken is known to be poor in growth and egg production (20 to 100 eggs per year), unlike commercial layers. Such large variation in performance might be explained by these candidate genes, which are associated with metabolism, muscle development, and reproduction between indigenous chicken and commercial layer.

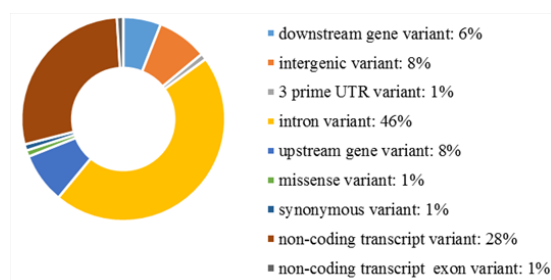


Figure 2 Types of genomic variants in the regions affected by selection.

Conclusion

The study suggests that the Z chromosome has undergone selection and have a contribution in the divergence between indigenous chicken and commercial layer populations. Conversely, identification of candidate genes provided a further understanding of the source of their phenotypic variation.

Acknowledgements

None.

Conflict of interest

Author declares that there is no conflict of interest.

References

- Wragg D, Mwacharo J, Alcalde J, et al. Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. *Heredity*. 2012;109(1):6–18.
- Ngeno K. Breeding program for indigenous chicken in Kenya. *Analysis of diversity in indigenous chicken populations*. Netherlands: PhD Thesis, Wageningen University; 2015. 154p.
- Zhu C, Fan H, Yuan Z, et al. Detection of Selection Signatures on the X Chromosome in Three Sheep Breeds. *Int J Mol Sci*. 2015;16(9):20360–20374.
- Heyer E, Segurel L. Looking for signatures of sex-specific demography and local adaptation on the X chromosome. *Genome Biol*. 2010;11(1):203.
- Zhang W, Shen M, Li H, et al. Detection and analysis of polymorphisms of 59571364 and 59912586 loci on X chromosome in fat-tail and thin-tail sheep flocks. *Heredity*. 2013;35(12):1384–1390.
- Joshi N, Fass J. *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files*. 2009.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
- Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–2158.
- Rubin C-J, Zody MC, Eriksson J, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*. 2010;464:587–591.
- McLaren W, Pritchard B, Rios D, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010;26(16):2069–2070.
- Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005;21(16):3448–3449.
- Dimmer EC, Huntley RP, Alam-Faruque Y, et al. The UniProt-GO annotation database in 2011. *Nucleic Acids Res*. 2012;40(Database issue):D565–D570.
- Strachan T, Read AP. *Instability of the human genome: mutation and DNA repair*. USA: Oxford University Press; 1999.
- Sergey, Y, Diana D, Maricela P, et al. Accelerated evolution of 3' avian *FOXE1* genes, and thyroid and feather specific expression of chicken *FoxE1*. *BMC Evol Biol*. 2011;11:302.