THE POPULATION GENOMIC SIGNATURE OF ENVIRONMENTAL SELECTION IN CHICKENS FROM MALAWI, SOUTH AFRICA AND ZIMBABWE

K. Hadebe[1], E.F. Dzomba[2] and F.C. Muchadeyi[1]

[1]Agricultural Research Council, Biotechnology Platform, Private Bag X5, Onderstepoort, 0110 South Africa
[2] Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville, 3209 South Africa

SUMMARY

Indigenous chickens in Africa are found across heterogeneous landscapes, and heritable adaptive variations across environmental gradients suggest local adaptation. The direction of adaptive differentiation is still underestimated and may have negative impact on the conservation programs. This study examined 60K genotyping data from 311 village chickens from Zimbabwe, Malawi and South Africa, and conserved flocks (Venda, Naked Neck, Potchefstroom Koekoek, and Ovambo) to identify runs of homozygosity (ROH) and selection signatures using ROH islands and association of SNPs with bioclimatic and geographic variables. Overall, 5537 ROH were detected, with short segments more prevalent across all populations. Larger windows (>40 Mb) were found in the South Africa and Zimbabwe flocks only, suggesting less genetic diversity. Thirty-three ROH islands (50% of population) were only found in Naked Neck, Potchefstroom Koekoek and Venda and were located in 4352 genes. Two SNPs Gga_rs14045047 (chromosome 12) and Gga_rs13560712 (chromosome 6) were associated with 7 variables and longitude, altitude, BIO8, BIO17 was common for both. This suggests their importance and complexity of genetic adaptation. This study identifies regions potentially under selection pressure of production system and environmental adaptation and provides baseline for identifying populations adapted to local environment.

INTRODUCTION

Indigenous chickens in Africa have an extended geographic distribution across agro-ecological zones and production systems, thriving in environments with limited resources due to their unique adaptive traits. After an initial description and characterisation of the extensively raised village chickens populations from Zimbabwe and Malawi (Muchadeyi *et al*. 2007) and South Africa (Mtileni *et al*. 2011), A detailed population genetic studies using the Illumina 60K SNP BeadChip were completed (Khanyile *et al*. 2015a; 2015b). These studies observed genetic divergence with sufficiently strong geographic barrier of village genetic groups among African countries (Muchadeyi *et al*. 2007), differentiation of South African conservation populations to founder village populations due to reproductive isolation (Mtileni *et al*. 2011; Khanyile *et al*. 2015a; 2015b), regions with high linkage disequilibrium suggestive of selection of signatures (Khanyile *et al*. 2015a). Understanding the role of natural and artificial selection in the shaping diversity may provide new insights into the genetic mechanisms underlying their adaptation to their production environment. ROH have been used in livestock genomic studies, confirming the correlation between shared ROH and genomic regions putatively under selection (ROH island) (Mastrangelo *et al*. 2018). In addition, signatures of past climatic trends have played large roles in shaping genetic structure of livestock species. Therefore, the objectives of our study were to detect runs of homozygosity (ROH) and detect key bioclimatic and geographic factors that drive adaptive differentiation and assess their association with SNPs using landscape genomics approach.

MATERIALS AND METHODS

Genomic data and quality control. Illumina chicken iSelect SNP60 Beadchip genotype data (SNP= 57636) of 311 chickens from different regions of Malawi, South Africa and Zimbabwe was used and has been previously described (Khanyile *et al*. (2015a; 2015b). Briefly, 135 village chickens were from three Zimbabwean agro-ecological zones (AEZ, AEZ1 = 92, AEZ3 = 34, and AEZ5 = 10) and 30 chickens were sampled from Malawi. South African village chickens (SAFIELD = 76) were ecotypes from Limpopo (n = 15), Eastern Cape (n = 26) and Northern Cape (n = 35) provinces. In addition, four conserved flocks (n = 70, Venda (VD = 20), Naked Neck (NN = 20), Potchefstroom Koekoek (PK = 20) and Ovambo (OV = 10) at the Agricultural Research Council Poultry Breeding Resource, Irene Pretoria, South Africa. Genotypes with a failed call rate of > 0.95, minor allele frequency of > 0.05 and Hardy-Weinberg equilibrium > 1^{-5} were used in this study. Accordingly, 46160 SNPs from 290 individuals were used for further analyses.

Runs of homozygosity (ROH) and ROH islands. Runs of homozygosity (ROH) was defined per animal as 1) 50 or more consecutive homozygous SNPs, 2) a minimum physical length of 1 Mb to exclude short ROH deriving from LD, 3) a density of 50 Kb/SNP and 4) maximum of 3 heterozygous calls within ROH using *detectRuns* (Biscari *et al*. 2018). ROH islands were defined by ROHs that occurred in 50% of the individuals.

Environmental contribution to genetic structure and selection. Bioclimatic variables over the 30-year period (1970 to 2000) were available from the WorldClim version 2 (Fick and Hijmas 2017) using the GPS coordinates (latitude and longitude) at each district level for each individual. Districts level was used as individual farm data were not available for all countries. In Zimbabwe, Temperature (°C) and precipitation (mm) variables included annual mean temperature (BIO1), Mean Diurnal Range (BIO2), Isothermality (BIO3), Temperature Seasonality (BIO4), Maximum Temperature of Warmest Month (BIO5), Minimum Temperature of Coldest Month (BIO6), Temperature Annual Range (BIO7), Mean Temperature of Wettest Quarter (BIO8), Mean Temperature of Driest Quarter (BIO9), Mean Temperature of Warmest Quarter (BIO10), Mean Temperature of Coldest Quarter (BIO11), Annual Precipitation (BIO12), Precipitation of Wettest Month (BIO13), Precipitation of Driest Month (BIO14), Precipitation Seasonality (BIO15), Precipitation of Wettest Quarter (BIO16), Precipitation of Driest Quarter (BIO17), Precipitation of Warmest Quarter (BIO18), Precipitation of Coldest Quarter (BIO19).. To prevent overestimation on the contribution to the genetic structure, correlation analysis was performed on the bioclimatic (BIO1-BIO19), geographic (longitude, latitude and latitude) variables using ggcorr in *GGally* package (Schloerke *et al*. 2013). Redundancy analysis (RDA) detected the contribution of the variables on the spatial genetic structure, using *vegan* package (Oksanen *et al*. 2015). Association analysis using latent factor mixed model (lfmm) was then performed using the *LEA* package (Frichot *et al*. 2013). Parameters included 10,000 sweeps, 5,000 burn-in sweeps, 10 repetitions and 6 latent factors (Khanyile *et al*. 2015). SNPs with a false discovery rate of P< 0.001 were considered as significantly associated.

RESULTS AND DISCUSSION

Runs of homozygosity (ROH) and ROH islands. A total of 5537 ROHs were detected across the 7 chicken populations. The frequency of ROHs and their length-distribution differed across populations (Table 1). In all populations, shorter segments of between 1 to 10Mb predominated the homozygosity present and accounted for approximately 82% of all ROH detected suggestive of more ancient relatedness, inbreeding and long-term selection within these populations. OV had the least number of ROH for all categories. Zimbabwe had the highest number of segments larger than 10Mb suggestive of the more likely that recent inbreeding occurred within a pedigree (Khanyile *et al*., 2015b) which remains unaccounted for in village populations, due to lack or recording system. The

increase of homozygous regions in NN, PK, and VD could be a consequence of inbreeeding, bottleneck effect and the decline in effective population size because they have been a closed populations since 25 years ago (Mtileni *et al*. 2011). Across all populations, the mean ROH length was 2.34 Mb and the longest segment was 50.09Mb in length (1994 SNPs) which was found on chromosome 3 in Zimbabwe population. The number of ROH per chromosome decreased with chromosome length and was greater for chromosome 1 (796 ROH) and lower for micro-chromosomes including chromosome 23 (23 ROH). High level of homozygosity in chromosome 1 was consistent with the presence of high number haploblocks (Khanyile *et al*. 2015b), which could be due to differences in recombination rates, genetic drift and selection across the different geographical distribution.

Table 1. Number of runs of homozygosity (*n*ROH) and length (in Mb) categorised by ROH length class (ROH_{1-5Mb}, ROH_{5-10Mb}, $ROH_{10-20\ Mb}$, $ROH_{20-40\ Mb}$, and $ROH_{>40\ Mb}$)

Class	NN	OV	PK	VD	SAFIELD	MALAWI	ZIMBABWE
1-5Mb	765	178	750	767	649	309	1102
5-10Mb	125	10	77	142	104	38	230
10-20Mb	31	1	12	46	35	21	88
20-40Mb	4	0	2	7	5	6	31
>40Mb	0	0	0	0	1	0	1
Total	925	189	841	962	794	374	1452

Thirty-three ROH islands, which indicate regions of strong selection were evident across the genome of NN (n = 5), PK (n = 7) and VD (n = 21) only. ROH islands were not found on other populations observed to have highly admixed individuals (Khanyile *et al*. 2015a). The longest ROH island was observed in VD on chromosome 7 (46.65Mb), while the shortest one was observed on chromosome 4 (1.73Mb). Within all of the ROH islands reported, we identified from 4352 genes (827 NN, 324 PK, 3202 VD). Functions of the genes varied and included metabolic, cardiac muscle and vascular smooth muscle contraction and signaling pathways,

Landscape genomics. Landscape genomics studies in indigenous livestock have gained momentum in past years. Highly correlated (r> 0.90) bioclimatic and geographic variables and those that did not explain genomic variation using RDA were removed, whilst annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), temperature annual range (BIO7), mean temperature of wettest quarter (BIO8), mean temperature of driest quarter (BIO9), mean temperature of warmest quarter (BIO10), precipitation of wettest month (BIO13), precipitation of driest quarter (BIO17), altitude, longitude, latitude were retained for LFMM analysis. RDA 1 and RDA 2 explained 3.82% and 1.23% of the variance, respectively. BIO3, BIO9, BIO1 and altitude explained most of the genetic variation (P <0.001) and BIO8 explained the least. RDA showed evidence of population structuring, consistent of the population structure described in Khanyile *et al*. (2015). Village populations clustered together despite the geographic distances between the countries potentially due to similar production environments. Reproductive isolation and sharing of production environment of the conservation flocks resulted in the populations clustering together despite different genetic backgrounds. Overall, a total of 3090 SNPs (6.69%) were associated with one or more variables, whilst 1888 SNPs were associated with a specific variable. BIO2 has been associated with thermoregulation and was the most significant variable in 283 SNPs suggestive of their role in adaptation to diurnal environmental conditions and BIO4 had the least (n= 184). SNP Gga_rs14045047 on chromosome 12 was associated with BIO1, BIO2, BIO8, BIO10, BIO17, alti-

tude and longitude. Association of BIO3, BIO7, BIO8, BIO13, BIO17, altitude and longitude with Gga_rs13560712 on chromosome 6 indicates their complexity in shaping the genetic diversity. Five SNPs (Gga_rs13705188 on chromosome 5, Gga_rs14836389 on chromosome 1, Gga_rs14091452 on chromosome 15, Gga_rs15243798 on chromosome 27, Gga_rs14142376 on chromosome 2) were only associated with longitude and latitude suggestive of role in the local adaptation. Genes were related to pathogen and disease defence and adaptive phenotypic traits including weight and fast growth.

CONCLUSIONS

The existence of ROH and islands demonstrated the role of the production systems in increasing homozygosity in specific regions of the genome. In low input village production systems in the sampled regions, climate ranges from heat and drought, pose selection pressure. Although different regions were identified between the two analysis, gene functions overlapped showing the complexity of response to production and environmental pressure.

ACKNOWLDEGMENTS

Agricultural Research Council Biotechnology Platform funded the study.

REFERENCES

Biscarini, F., Cozzi, P., Gaspa, G., Marras, G. (2018) https://github.com/bioinformatics-ptp/detectRUNS/tree/master/detectRUNS.

Frichot E., Schoville S.D., Bouchard G. and François O. (2013) *Mol. Biol. Evol* **30**: 1687.

Fick, S.E., and Hijmans, R.J. (2017) *International Journal of Climatology*.

Khanyile, K., Dzomba, E.F., and Muchadeyi, F.C. (2015a) *Frontiers in Genetics*. **6**:1.

Khanyile, K., Dzomba, E.F., and Muchadeyi, F.C. (2015b) *Genetics of Molecular Research* **14**:12276:.

Mastrangelo, S., Sardina, M.T., Tolone, M, Di Gerlando, R., Sutera, A.M., Fontanesi, L., and Portolano, B. (2018) *Animal.* **12**: 2480

Mtileni, B.J., Muchadeyi, F.C., Maiwashe, A., Groeneveld, E., Groeneveld ,L.F., Dzama K. and Weigend, S. (2011) *J. Anim. Breed. Genet.* **128**: 209.

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.

Oksanen. J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara R.B, Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. (2015). http://cran.r-project.org, https://github.com/vegandevs/vegan.

Schloerke, B., Crowley, J., Cook, D., Hofmann, H., Wickham, H., Briatte, F., Marbach, M. (2013). http://CRAN.R-project. org/package=GGally.