MedCrave
Step into the World of Research

Journal of Cancer Prevention & Current Research

Research Article

Open Access

Genome-wide analysis of luhya and maasai genomes reveals signatures of selection for fertility, prostate cancer, prostatitis and vein disease

Abstract

Background & Aim: Selection pressure has left traceable molecular selection signatures within the human genomes. Identification of selection signatures is vital for understanding the evolutionary processes shaping human adaptation and diversity in disease susceptibility. In this study, genomes of Luhya and Maasai ethnic groups were compared to identify selection signatures that may explain their genetic divergence.

Materials and Methods: The population studied composed of 90 Luhyas' and 31 Maasais' from Webuye and Kinyawa, Kenya respectively. Data were sourced from 1000 genomes project phase 3. These data were processed through similar statistical quality control protocols.

Results: Datasets were processed individually then merged; resulting in 1,480,668 high-quality SNPs. Genomic scan using F-statistics (Fst) method revealed putative genomic selection signatures. Fifty-five genomic regions (high Fst, 5.6 to 9.8). These regions overlapped with 21 genes. Three genes (*C9orf24*, *BAZ2A*, and *SYMPK*) out of the 21 are associated with the missense variants that could explain partially the genetic divergence and variation in fertility and disease susceptibility between these populations. The genes are associated with spermatogenesis (*C9orf24 gene*), prostate cancer and prostatitis (*BAZ2A gene*) and vein disease (*SYMPK gene*).

Conclusion: The study suggests that the two populations have undergone diverged selection leading to variation in fertility and susceptibility to prostate cancer, prostatitis and vein disease. It highly recommended the ethnic groups should consider changing their culture to avoid challenges identified in this study.

Keywords: fertility, diseases, luhya, maasai, diversity

Volume 9 Issue 2 - 2018

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

Correspondence: Kiplangat Ngeno, Department of Animal Sciences, Egerton University, P.O. Box 536, 20115 Egerton, Kenya, Email aarapngeno@gmail.com

Received: February 20, 2017 | **Published:** April 02, 2018

Introduction

Kenyans are diverse populations that are clustered mainly based on culture, ethnicity/linguistic, phenotypic and geographically attributes. Luhya and Maasai are among the ethnic groups in Kenya. Luhya is a Bantu ethnic group composing of 18 sub-groups.[1] Luhya is crop-livestock farmers and their main foods include maize meal (ugali) combined with vegetables and meat (cattle, goat, fish or chicken meat). On the other hand, the Maasai, are Nilotic–ethnic group, and a member of Nilo-Saharan family. Maasai are nomadic pastoralists with the main food being milk, meat and occasionally blood. They mainly occupy areas with low rainfall and not suitable for crop farming.

Luhya and Maasai ethnic groups have thrived in diverse agro-ecological environments. The adaptation is attributed to changes in their genetic make-up which have been shaped by selection pressures imposed by the local environmental conditions, parasites, diets, and diseases. Different selection pressures result in the development of specific genetic variants that play a role in the local adaptation and microevolution.

Luhya and Maasai populations are diverse culturally, linguistically, and phenotypically; however, the genomic root of their notable diversity remains poorly understood. Therefore, genome-wide scans of these ethnic groups are required to dissect the genetic basis of their divergence. Analysis of selection signatures is vital in understanding the genomic basis of adaptation to survival under tropical conditions. It offers the opportunity of understanding the forces of historical selection, genes and mutations underlying phenotypic diversity and adaptation to local environment. Studying Signatures of selection allow unravelling of genetic differences contributing to rare as well as common diseases more precisely. Identification of the genetic underpinnings of the disease will aid in diagnostic tests and, treatments in some cases. In this study, genomic variation between Luhya and Maasai ethnic groups (Bantu and Nilotic group) were compared to identify selection signatures that may explain their divergence.

Materials and methods

Samples

The population studied composed of 90 Luhya and 31 Maasai from Webuye and Kinyawa, Kenya respectively. Data were sourced from Omni2.5 chip data of 1000 Genomes Project Phase 3. Data were obtained from the 1000 Genomes Project.[2]

Genotyping and SNP calling

The samples were genotyped using the Illumina HumanOmni 2.5MTquad BeadChip array. Sample collection, generation and processing of data is described in1000 Genomes Project.[3]

Statistical analysis

Individual in each population were subjected to the same quality control protocol; (i) Low quality variants (< 20), (ii) duplicates, (iii)

SNPs with less than 95% call rate, (iv) minor allele frequency (MAF) less than 5%, (v) Hardy Weinberg disequilibrium ($p<10^{-7}$) and (vi) samples with more than 10% missing genotypes were filtered out. Related individuals were identified by first estimating identity by descent (IBD). Individuals with estimated genome-wide IBD values above 0.05 were removed using. Datasets were processed individually then merged, resulting in 1,480,668 high-qualities SNPs. Sex chromosomes were filtered out. All statistical analysis was carried out using VCFtools v0.1.13,[4] SAMtools v0.1.31 and Plink v1.9 software.[5]

Genomic scans for selection signatures

Genomic scans for selection signatures were carried out by merging individual data into two datasets: the Luhya dataset comprising data from 91 individuals were considered as one single population, and the Maasai dataset comprising data from 31 individuals as a second population. Locus-specific fixation index (Fst) were calculated between Luhya and Maasai population using Weir and Cockerham formula.[6] The Fst values were estimated using VCFtools in bins of a 40kb window sliding 20kb each time over the entire genome. Bins with less than 10 variants were filtered out to reduce false positives. The skewed Fst values were normalized (ZFst) using the Z-transformation formula.[7] Bins with transformed Fst (ZFst) above six (top 0.005% of Fst values) were qualified as putative selection regions. The package ggplot2 in R environment[8] was used to display the graphs.

Functional annotation of genomic variants

The top 0.005% of Fst values (36 regions) were annotated for genes and their consequence using VEP.[9] The option of SIFT, Condel and PolyPhen 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 Fst regions were extracted and their associated genes were run against GeneCards human gene databases and MalaCards human database to search expression and diseases associated with the genes.

Results

Genomic signatures of selection

Luhya and Maasai genomes were scan using F-statistics (Fst) method to search for putative genomic selection signatures. Fifty-five genomic regions located on seven different chromosomes had high Fst values ranging from 5.6 to 9.8 (Figure 1). These regions overlapped with 21 genes (Table 1), three of which (*C9orf24, BAZ2A* and *SYMPK*) are associated with the missense variants.

Table 1 Different types of genes identified in the elevated Fst regions in the different chromosomes

Chromosome	Gene
3	*LOC101927518*
5	*CTB-114C7.4, LINC01366*
9	*FAM219A, C9orf24,Y_RNA, KIAA1161*
12	*BAZ2A, ATP5B, SNORD59A, SNORD59B,*
15	*DUT, LOC107984757, RP11-154J22.1,Y_RNA*
17	*ARHGAP23, AC124789.1, LOC101929494,*
19	*SYMPK, AC092301.3, FOXA3*

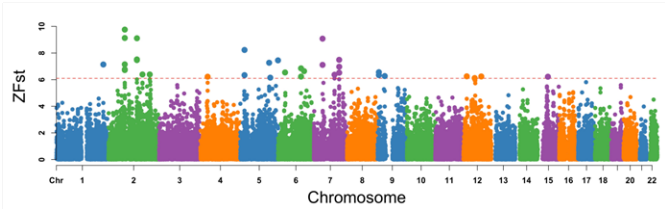

Figure 1 Distribution of the Fst values along the 22 chromosomes.

Functional annotation of genomic variants

Different types of genes identified in the elevated Fst regions in the different chromosomes and their respective biological process are depicted in Table 2. The prediction of the effect of substitution of amino acid on protein function revealed the consequence of the three genes (*C9orf24, BAZ2A* and *SYMPK*) to be deleterious or damaging.

The *C9orf24, BAZ2A* and *SYMPK* candidate genes were run against GeneCards and MalaCards human gene databases to search for expression and diseases associated with the genes. Results showed the genes are linked to spermatogenesis (*C9orf24* gene), prostate cancer and prostatitis (*BAZ2A* gene) and vein disease (*SYMPK* gene).

Table 2 Different types of genes identified in the elevated Fst regions and their respective biological process GO-terms

Chromosome	SNP	Position	Allele	SYMBOL	Amino acids	Codons	SIFT	PolyPhen	Condel
9	rs3808876	34382726-34382726	G	C9orf24	W/R	Tgg/Cgg	Tolerated low confidence (0.4)	Benign (0)	Neutral (0.016)
9	rs11790577	34397545-34397545	G	C9orf24	M/T	aTg/aCg	Deleterious (0.01)	Probably damaging (0.969)	Deleterious (0.817)
12	rs773663	56601739-56601739	T	BAZ2A	A/E	gCa/gAa	Deleterious (0.03)	Probably damaging (1)	Deleterious (0.865)
12	SNP12-54901530	56615263-56615263	G	BAZ2A	N/H	Aat/Cat	Tolerated (0.09)	Probably damaging (0.999)	Deleterious (0.787)
19	SNP19-50518206	45826366-45826366	A	SYMPK	S/F	tCc/tTc	Deleterious (0.03)	Benign (0.438)	Deleterious (0.507)

Discussion

Selection signatures

Information on signatures of selection is vital to understanding why people are different. In this study, a detailed genome-wide scan was performed to address the genetic basis of the differences between Luhya and Maasai populations. Results revealed 55 genomic regions, which are highly differentiated between the Luhya and Maasai ethnic groups. These regions under positive selection reflect unique historical population movements and evolution. The identified genomic regions overlapped with 21 genes, indicating the genes have been selected in the opposing direction.

Three genes (*C9orf24, BAZ2A,* and *SYMPK*) out of the 21 are associated with the missense variants. The genes have a deleterious effect and have changed the sequence of amino acid resulting in the differences between Luhya and Maasai population. Chromosome 9 open reading frame 24 (*C9orf24*) is a protein-coding gene that encodes a nuclear- or perinuclear-localized protein. This gene is overexpressed in reproductive system (testis and ovary) and it is expressed more in adults compared to young and embryos. Functionally, the gene plays a role in spermatogenesis and ciliogenesis (differentiation or function of ciliated cells) process. Variation tolerance analysis for *C9orf24* showed the residual variation intolerance score to be 79.8% and gene damage index score of 0.82 (17.26%), meaning the gene is more intolerant (likely to be disease-causing).

Bromodomain adjacent to zinc finger domain 2A (*BAZ2A*) is a protein-coding gene. This gene is overexpressed in nasal epithelium, testis, and kidney. The gene plays a vital role in the regulation of protein-coding genes and recruitment of *DNMT1, DNMT3B* and *HDAC1* chromatin silencing proteins. Together with *EZH2*, maintain gene's epigenetic silencing. Gene is associated with prostate cancer and Prostatitis.[10] It is overexpressed in individuals suffering from prostate cancer and plays a role in epigenetic alterations of prostate cancer and maintenance of growth of prostate cancer cells.[11] The high level of gene expression has been used as an independent predictor of biochemical recurrence of prostate cancer. It is well documented that frequent sexual activity lowers prostate which seems to be the case in Luhya group whom they are known to be fertile compared to the Maasai population. There is a positive association between intake of meat due to HCA and lethal prostate cancer.[12] Maasai's are more likely to be prone to prostate cancer because their meating eating culture exposes them to heterocyclic amines (HCA) compounds that are formed when meat is roasted or cooked at high temperatures.

Symplekin (*SYMPK*) is a protein-coding gene regulating the polyadenylation and promotion of gene expression.[13] This gene is found in a heat-sensitive complex and interacts with *HSF1* in heat-stressed cells.[14] The *SYMPK* gene is overexpressed in peripheral blood mononuclear cells, lymph node, and bone. Gene *SYMPK* is associated with vein disease.[10] Vein disease occurs more often in people who stand for long periods of time (Mosti, 2015), which is the case in the Maasai community when herding.

Information on selection signatures in this study offered information on the specific ecological adaptations of Luhya and Maasai populations. Luhya and Maasai live in different agro-ecological zones with varying altitudes and environmental conditions such as the frequency of droughts, dry spells duration, rainfall intensity, heat stress level and frequency of disease outbreaks. Luhya inhabits the western part of Kenya, which is a cool highland with a mean altitude of 1942m (1774 to 2084) above sea level, annual rainfall is 1422mm (1100-1500) and the temperature is 18^{C}(13-24). Maasai live in an arid and semiarid lowland with a mean altitude of 815m (739-914), annual rainfall of 243mm (120-430) and temperature of 30^{C} (24-35). In Maasai land, the frequency and length of droughts, level of heat stress is higher whereas rainfall intensity is low. Such varied climatic conditions are associated with varied climatic linked challenges such as the frequency of disease outbreaks. Their adaptation to the diverse environmental conditions, culture, food, disease, and parasites are at the root of their notable diversity in spermatogenesis and susceptibility to diseases; prostate cancer, prostatitis and vein disease.

Conclusion

Luhya and Maasai populations have suffered a strong selective pressure in an opposing manner. Findings in this study suggest that the two populations have undergone diverged selection leading to variation in fertility and susceptibility to different diseases; prostate cancer, prostatitis and vein disease. It highly recommended the ethnic groups should consider changing their culture to avoid challenges identified in this study.

Acknowledgements

Data used in this article is derived from 1000 Genomes Project.

Conflict of interest

There are no competing interests in this publication.

References

1. Bulimo SA. Luyia of Kenya: A cultural profile. Trafford Publishing, 2013.
2. Consortium GP. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56.
3. Consortium GP. A global reference for human genetic variation. *Nature.* 2015;526(7571):68.
4. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156–2158.
5. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The Am J Hum Genet.* 2007;81(3):559–575.
6. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution.* 1984;38(6):1358–1370.
7. Rubin C-J, Zody MC, Eriksson J, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature.* 2010;464(7288):587.
8. Team RC. R: A language and environment for statistical computing. 2013.
9. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome biol.* 2016;17(1):122.
10. Rappaport N, Twik M, Plaschkes I, et al. Mala Cards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic acids res.* 2016;45(D1):D877–D887.
11. Gu L, Frommel SC, Oakes CC, et al. BAZ2A (TIP5) is involved in epigenetic alterations in prostate cancer and its overexpression predicts disease recurrence. *Nat genet.* 2015;47(1):22.
12. Richman EL, Kenfield SA, Stampfer MJ, et al. Egg, red meat, and poultry intake and risk of lethal prostate cancer in the prostate-specific antigen-era: incidence and survival. *Cancer Prev Res (Phila).* 2011;4(12):2110–21.
13. Coordinators NR. Database resources of the national center for biotechnology information. *Nucleic acids rese.* 2017;45(D1):D7-19.
14. Fishilevich S, Nudel R, Rappaport N, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database(Oxford).* 2017;2017.