Cell Genomics

CellPress
OPEN ACCESS

Preview

Leveraging our common African origins to understand human evolution and health

Ambroise Wonkam[1,*] and Adebowale Adeyemo[2]
[1]McKusick-Nathans Institute & Department of Genetic Medicine. Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[2]Center for Research on Genomics and Global Health, National Human Genome Research Institute, The National Institutes of Health, Bethesda, MD, USA
*Correspondence: awonkam1@jhmi.edu
https://doi.org/10.1016/j.xgen.2023.100278

In the March 2023 issue of *Cell*, Fan et al.[1] report whole-genome sequencing across 12 indigenous African populations and analyze local adaptation and evolutionary history. Here, Wonkam and Adeyemo highlight their findings and how this contributes to African and global genomic research.

Less than 2% of human genomes analyzed so far have been those of people with African ancestry, even though Africa, where humans originated, contains more genetic diversity than any other continent and millions of uncaptured variants accumulated over 300,000 years of modern humans' evolutionary history. It is now clear that the detailed study of African genomic variation is a scientific imperative.[2] Therefore, the research reported from Fan et al. adds much needed data to our knowledge of African genomes and the information they contain about humanity's history within Africa.[1] The authors used whole-genome sequencing (WGS) of 180 individuals from 12 indigenous African populations encompassing all four African language phyla to provide evidence of multiple introgression events from "ghost" archaic populations and revealed additional insights on within-Africa admixture, migration, and signatures of local adaptations.

The authors data suggest that all modern humans derive approximately 5%–15% of their ancestry from a lineage that may have diverged as long as 1–3 mya, with multiple introgression events (Figure 1A), which is in line with previous reports.[3] Generation of high-coverage reference genomes for archaic hominid species such as Neanderthals has allowed the identification of approximately 2% introgression portion of that genome in present-day Europeans, ostensibly enriched for variation in genes involved in dermatological phenotypes, neuropsychiatric disorders, and immunological functions, including host susceptibility to coronavirus disease 2019 (COVID-19).[4] In a similar vein, Fan et al. showed evidence of 2% content from Neanderthal genomes in African populations from Ethiopia, most likely as a result of admixture with non-Africans. However, these findings should be further interrogated directly with comparative analysis of more genomes from Ethiopians and available Neanderthal genomes. The largest challenge in expanding this area of research in Africa has been the inability to obtain high-quality ancient DNA from regions with a tropical climate where the heat and humidity rapidly degrade DNA. The recent report of the successful isolation and sequencing of DNA from four children buried at Shum Laka (Cameroon) 3,000–8,000 years ago[5] indicates that studies of ancient DNA are now possible in western Central Africa. Of note, DNA has been successfully obtained from individuals in eastern and south-central Africa up to 18,000 years ago.[6] Similar successes in future studies have the potential to further inform our knowledge of the distribution of ancestry from archaic ghost populations (that never moved out of Africa) and possibly lead to discovery of novel associations between variants from archaic humans and human traits and diseases.

Genetic studies in African populations have reconstructed multiple migrations and movements of people across the continent (Figure 1B). While confirming some of these population movements, particularly the so-called Bantu expansion as previously reported,[7] Fan et al. further emphasized that ancestors of southern African San and central African rainforest hunter-gatherers (RHGs) diverged from other populations >200 kya and uniquely revealed gene flow between eastern and southern Khoesan-speaking hunter-gatherer populations lasting until ~12 kya, even though these groups are currently isolated from each other. As expected, the San and RHG individuals have the greatest number of SNPs and the highest levels of genetic diversity, whereas individuals from populations that experienced strong non-African admixture (e.g., Amhara from Ethiopia) carry the fewest SNPs and have the lowest genetic diversity.[1]

Fan et al. reported more than 5 million novel variants, with 78% that are population specific. Similarly, previous H3Africa consortium WGS analyses of 426 individuals (comprising 50 ethnolinguistic groups) uncovered more than 3 million undescribed variants.[6] Because many of these novel variants were predicted to be functionally important, these novel findings could have implications for refining variant deleteriousness interpretation. For example, "unusually" high frequencies of targeted "pathogenic" variants in *DNH* among Fulani can be interpreted as most suggestive of variant misclassifications in ClinVar and other databases. While the small numbers of individuals studied per population (15) have the limitation of wide confidence intervals around the reported allele frequencies, these observations emphasize a strong need to include ethnically diverse and particularly African populations in human genetic studies, especially because rarity

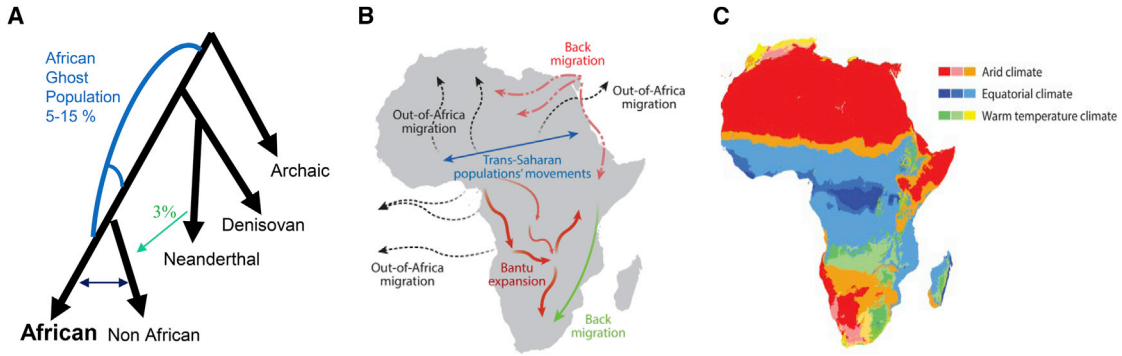

Figure 1. Major driving factors contributing to African genome complexity

(A) Complex demographic history of African populations consists of ancient population evolutionary divergence, including multiple introgressions of archaeid DNA within African genomes, as illustrated by a simplified demographic model indicating at least two introgression events from archaic human that never move out of Africa (blue lines).

(B) Migration and admixture events over 300 years (e.g., the Bantu expansion [solid red arrows], back migration [dotted-dashed red and solid green arrows], out-of-Africa migrations [dotted black arrows], and trans-Saharan populations' movements [solid blue arrow]).

(C) Combined variability of environment desert, tropical rainforest, savanna, swamps, and high-altitude mountains and related diverse selection pressures from factors such as climate, diet, and pathogen exposure have drove local adaptation, leading to far-reaching effects on African genomes.

(B) and (C) are adapted from Wonkam et al.[2]

is used as a criterion for determining a variant's pathogenicity in clinical studies. Another important implication of these uncaptured variants is their potential utility in mapping complex traits. The deep evolutionary tree within Africa, with more time for recombination to separate linked alleles, combined with complex patterns of admixture across the continent creates highly heterogeneous haplotypes with limited linkage disequilibrium.[2] As further illustrated by the present paper, on average, a genome of African ancestry carries three to five times more variants than those from other ancestries.[7] In concert, there is a similar imperative to develop faster and better methods of accounting for admixture in the context of natural selection by teasing apart local ancestry, i.e., the likely ancestral parental haplotypes at a locus. Therefore, to promote discovery and produce reliable clinical tools, genotyping and analysis must be re-optimized using genomes from more African populations. For example, a recently designed genome-wide association study (GWAS) SNP array developed by the H3Africa consortium is already showing some promising results with the discovery of new quantitative trait loci, i.e., a novel LDL-C association in the *GATB* region while replicating several well-known lipid-trait loci including *LDLR*, *PMFBP1*, and *LPA*. The transferability of signals detected in two large global studies consistently improves with an increase in the size of the African replication cohort.[8] A major focus of the present study was to bridge the gaps in African population-scale WGS data. However, we note the very low inclusion of samples from understudied geographical regions such as far western and northern regions of Africa. Owing to the extent of variants that are population specific (78%), with only 15% shared by population in the same country and 7% shared by populations in different countries, it is critical that more populations within more countries should be investigated to better refine African genomic variation.

The north-south axis orientation of the African continent is associated with a range of climates, cultures, languages, and biodiversity across the continent (Figure 1C), both in the present and in the past. It has shaped differential genomic variant frequencies among populations via natural selection as an adaptation to environmental pressures. Quintessential examples include (1) variants in the β-globin gene (*HBB*) that cause sickle cell disease but confer resistance to malaria, (2) variants in *APOL1* that are protective against trypanosomes but increase susceptibility to chronic kidney disease, (3) variants of *OSBPL10* and *RXRA* that protect against dengue fever, and (4) variants in genes that play a role in lactose tolerance.[2] Another example is the Duffy allele *ACKR1*_rs2814778-C, whose global map reflects adaptation to malaria, with a high frequency of the Duffy-null genotype in sub-Saharan Africa (>90%) and up to 65% among African Americans (who are primarily African-European admixed). A recent study showed that decisions on safety and discontinuation of azathioprine (a cancer medication) made upon deviations of neutrophil count values defined as "normal" in Europeans (absolute neutrophil count $< 1.5 \times 10^9$ cells/L) can be inappropriate for Africans. Such decisions can result in unnecessary and potentially harmful azathioprine discontinuations in African American patients if made in the absence of genetic information that could provide an alternate explanation for neutropenia (such as the Duffy-null phenotype).[9] Still more variants of relevance to persons of African ancestry are being added to the above examples. An H3Africa survey of seven African populations identified 60 novel coding loci and 34 novel noncoding loci that exhibited evidence of natural selection, and most variants were in genes involved in viral pathogen infections and basic cellular metabolism.[7] Therefore, it is not surprising that Fan et al. revealed additional signatures of local adaptation—and natural selection—for several traits related to skin color (e.g., *OCA2*, *TYRP1*, *SLC24A5*, *MITF*) immune response, height, and metabolic processes. The functional validation provided for a positively selected variant, rs77665059, in the lightly

pigmented San is noteworthy, suggesting *in vitro* a regulation of the enhancer activity and gene expression of *PDPK1*. Similarly, in the Mursi, Amhara, and Dizi in East Africa, they observed that enrichment for genes involved in pathways related to kidney development and morphology could reflect an adaptation to environments that are often arid, in line with findings from previous multi-ancestry GWASs.

In conclusion, this study is a clear demonstration that increased availability of African genomes will improve our understanding of evolutionary history of humans, refine our studies of genomic variation associated with complex traits, and improve how we can leverage knowledge of population-specific variants in health and disease to improve genetic medicine for Africans and potentially in all populations. The extent of novel variants found in this study within a relatively limited number of African genomes suggests that hundreds of thousands, if not millions, of African genomes still need to be sequenced. Sequencing of African genomes need to be done not only at scale (large numbers) but also to capture the wide diversity of populations across Africa.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Fan, S., Spence, J.P., Feng, Y., Hansen, M.E.B., Terhorst, J., Beltrame, M.H., Ranciaro, A., Hirbo, J., Beggs, W., Thomas, N., et al. (2023). Whole-genome sequencing reveals a complex African population demographic history and signatures of local adaptation. Cell *186*, 923–939. https://doi.org/10.1016/j.cell.2023.01.042.
2. Wonkam, A., Munung, N.S., Dandara, C., Esoh, K.K., Hanchard, N.A., and Landoure, G. (2022). Five priorities of African genomics research: The next frontier. Annu. Rev. Genomics Hum. Genet. *23*, 499–521.
3. Durvasula, A., and Sankararaman, S. (2020). Recovering signals of ghost archaic introgression in African populations. Sci. Adv. *6*, eaax5097.
4. Zeberg, H., and Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature *587*, 610–612.
5. Lipson, M., Ribot, I., Mallick, S., Rohland, N., Olalde, I., Adamski, N., Broomandkhoshbacht, N., Lawson, A.M., López, S., Oppenheimer, J., et al. (2020). Ancient West African foragers in the context of African population history. Nature *577*, 665–670. https://doi.org/10.1038/s41586-020-1929-1.
6. Lipson, M., Sawchuk, E.A., Thompson, J.C., Oppenheimer, J., Tryon, C.A., Ranhorn, K.L., de Luna, K.M., Sirak, K.A., Olalde, I., Ambrose, S.H., et al. (2022). Ancient DNA and deep population structure in sub-Saharan African foragers. Nature *603*, 290–296. https://doi.org/10.1038/s41586-022-04430-9.
7. Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y.J., et al.; TrypanoGEN Research Group; H3Africa Consortium (2020). High-depth African genomes inform human migration and health. Nature *586*, 741–748.
8. 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
9. Dickson, A.L., Daniel, L.L., Jackson, E., Zanussi, J., Yang, W., Plummer, W.D., Dupont, W.D., Wei, W.Q., Nepal, P., Hung, A.M., et al. (2022). Race, genotype, and azathioprine discontinuation: A cohort study. Ann. Intern. Med. *175*, 1092–1099.