IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Exploring the Role of Ethnicity and Genetic Ancestry in Disease Prevalence and Risk

Fatima Z. Al-Sakkaf

The Sheffield Private School, Al-Nahda 2, Dubai, United Arab Emirates; fatima.z.alsakkaf15@gmail.com
Mentor: Dr Kif Liakath-Ali

ABSTRACT: This article aims to encompass propositions, varying from environmental selection pressures to genetic predisposition that summarizes why certain populations are at risk of specific diseases or risk factors. A focus on pathophysiological pathways of inherited susceptibility and current modern analysis methods is drawn. Both multitudes of epidemiological and genetic ideas will be explored, as well as their closely linked research studies will be analyzed. Various approaches, including genetic strategies such as admixture linkage disequilibrium mapping (MALD), are discussed and related to their applications in recessive and Mendelian disorders, both common and rare. The biological foundation of genetic diseases, and their origin and evolution must, to an extent, rely on the past. Therefore, encouraging a more sophisticated analysis of the importance of ancestry and ethnicity can stimulate scientific development to fill gaps in medical therapies' current understanding and progression.

KEYWORDS: Cellular and Molecular Biology; Genetics; Admixture Linkage Disequilibrium Mapping (MALD); Consanguinity; Genetic Drift.

Introduction

Race and ethnicity are social constructs incorporating biological and sociocultural components. Genetic ancestry is a concept that describes the construct of genome variation between populations. All genetic variation begins locally as a new mutation in an individual, and thus, all new mutations are initially geographically localised.[1] As a result of this population history, a pattern may be established between genetic variants found in one geographic region and their differentiated expression in the respective pan-ethnic group.

Diversity has led to population-specific variants – genetic variations more prevalent within certain ethnic or geographic groups. These variants play a significant role in shaping individual susceptibility to diseases, and their study offers insights into the intricate dynamic between genetics, environment, and health outcomes. Population-specific variants are rooted in historical migration patterns and local adaptations. Their impact on disease susceptibility and resistance is well-documented. Human populations migrated to different regions and encountered unique environments, pathogens, and dietary sources. Natural selection favored genetic variants that applied survival advantages in these specific contexts. For instance, the higher prevalence of sickle cell anemia in populations originating from malaria-endemic regions highlights the protective role of specific variants against the effects of this infectious disease. This shows how variants that once provided an evolutionary advantage can inadvertently increase the risk of certain diseases in the present day.

Another, lesser-known example is that variants associated with skin pigmentation may have evolved as adaptations to varying levels of ultraviolet (UV) radiation.[2] However, these same variants can also influence susceptibility to skin cancer in modern populations exposed to different UV environments. Conversely, population-specific variants can confer resistance to diseases prevalent in certain geographic areas. The CCR5-Δ32 variant, which protects against HIV infection, is more common in populations with a history of exposure to the virus.[3] Such examples underscore the complex relationship between genetic diversity, historical context, and disease risk.

Discussion

Reproductive Isolation Mechanisms:

Isolation and genetic drift are both important concepts in evolutionary biology. Isolation separates a population into smaller groups, usually due to geographic barriers or other factors. This isolation can lead to genetic differentiation as each group faces its unique environment and selection pressures. Ethnic groups isolated geographically or culturally can develop distinct genetic traits due to limited gene flow with other populations.[4] This isolation can lead to a higher prevalence of certain genetic disorders within these groups if those disorders were present in the founding population. Some genetic diseases are more common in specific ethnic groups due to these historical factors.

Genetic drift, conversely, is a random process that can cause changes in the frequency of alleles in a population over time. It's particularly significant in small populations where chance events can have a big impact. Genetic drift can lead to the fixation of certain alleles, meaning they become the only alleles present in the population, or the loss of alleles, reducing genetic diversity.[5] When a small ethnic group experiences genetic drift, it can lead to changes in the frequency of specific alleles, including those related to disease susceptibility.[6] If a disease-associated allele becomes more common due to genetic

DOI: 10.36838/v6i6.4

drift in a particular ethnic group, the prevalence of that disease could increase among members of that group.[7] Isolation often plays a role in the occurrence of genetic drift. When populations are isolated, genetic drift can have a stronger effect on each isolated group, leading to divergence between them. This can ultimately result in the formation of new species if the genetic differences become significant enough. Isolation by distance best describes the pattern of variation. Differences in populations' phenotypic and genotypic expression are roughly proportional to the geographic distance between them.[8] Together, isolation and genetic drift are crucial components of the evolutionary process, shaping populations' genetic diversity and divergence over time.

The Founder Effect is a specific type of genetic drift that occurs when a small number of individuals from a larger population establish a new population, often in a new geographic area. If the founding individuals carry alleles that increase the risk of certain diseases, those diseases might be more prevalent in the new population.[9]

Inherited risk details how Isolation and genetic drift can concentrate certain genetic variations within ethnic groups, including protective and risk-associated alleles. This means some ethnic groups may have a higher genetic predisposition to specific diseases, while others may have protective genetic factors. For example, certain genetic variations might confer resistance to certain infectious diseases or increase the risk of developing certain metabolic disorders.

Consanguinity:

Consanguinity refers to the marriage or reproduction between close blood relatives (typically first or second cousins), which can affect disease prevalence and risk in several ways, primarily due to the increased likelihood of inheriting autosomal recessive genetic disorders.[10,11] When closely related individuals have children, there is a higher chance that both parents carry the same recessive disease-causing allele. If both parents are carriers (heterozygous) for the same autosomal recessive disorder, their offspring have a 25% chance of inheriting two copies of the mutated gene (being affected by the disorder). This increased risk can lead to a higher prevalence of specific genetic diseases in populations with a history of consanguinity.[12]

Another implication is the amplification of recessive alleles. Specific autosomal recessive alleles may be more common in consanguineous populations due to shared ancestry. This can amplify the prevalence of certain genetic disorders, making them more frequent in these populations compared to populations with less consanguinity.[13]

While consanguinity significantly impacts recessive genetic disorders, it may also contribute to the prevalence of certain complex disorders or traits influenced by multiple genes and environmental factors. Consanguinity can increase the likelihood of sharing genetic susceptibility factors, which could contribute to the risk of multifactorial conditions. This encompasses type 1 and 2 diabetes, high blood pressure and clotting factor inheritance.[14]

In some cases, consanguinity can lead to genetic isolates, where a population is relatively isolated due to geographic, cultural, or historical factors. This isolation can enhance the effects of consanguinity, leading to the accumulation of specific genetic variants, both beneficial and detrimental, within the isolate.

Risk VS Resistance: Sickle Cell Disease:

Red blood cells are the basic foundational component of blood, giving it the ability to carry oxygen and its distinct red color. They differ from typical somatic cells by their characteristic biconcave shape, allowing maximum surface area for carrying oxygen and physiological ease of navigating blood vessels.

Sickle Cell Disease (SCD) is a genetic disorder that originated due to a mutation in the HBB gene, which encodes the beta-globin subunit of haemoglobin, the protein responsible for carrying oxygen in red blood cells.[15,16] This mutation has a fascinating evolutionary history linked to protection against malaria.

The mutation responsible for SCD is a point mutation that changes one amino acid in the beta-globin protein, causing the red blood cells to take on a "sickle" shape under certain conditions. The sickle-shaped cells can cause blockages in small blood vessels, leading to pain, anemia, and other complications associated with the disease.[17] The interesting aspect of this mutation is that being a carrier of the sickle cell trait (heterozygous for the mutation) provides some protection against malaria, a deadly infectious disease transmitted by mosquitoes. This encompasses people with one normal HBB gene (A) and one mutated HBB gene (S). While the trait can cause some health issues, it also confers a degree of resistance to malaria (Figure 1).

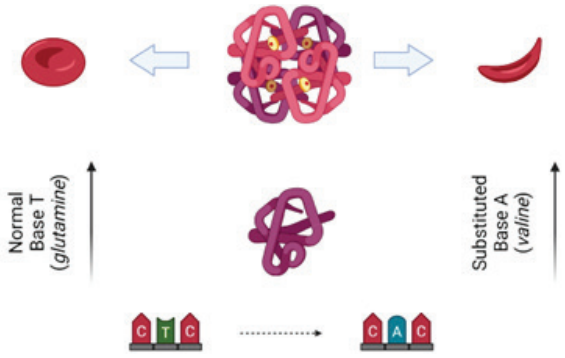

Figure 1: Normal vs. SCD Point Mutation in HBB Gene Comparison.

When a point mutation occurs on the CTC triplet codon in the beta-globin protein, the amino acid valine is produced, and the red blood cell takes on a curved 'sickle' shape, as depicted.

Created with BioRender.com

In regions where malaria is prevalent, individuals with the sickle cell trait are more likely to survive and reproduce, passing on the protective trait to their offspring. This selective advantage has contributed to the persistence of the sickle cell mutation in certain populations, particularly those from sub-Saharan Africa, the Mediterranean, and other regions where malaria historically posed a significant threat.[18] However, when an individual inherits two copies of the mutated gene (HbS) from both parents (homozygous for the mutation),

the resulting condition is sickle cell disease, which has severe health consequences.

This evolutionary perspective helps explain the geographical distribution of the sickle cell mutation and why it's more prevalent in certain ethnic groups with a history of exposure to malaria. While the mutation protects against malaria, it also leads to the development of a serious genetic disorder in homozygous individuals, leading to heterozygous advantage.

Screening:

The ability to predict future risk hinges on the capacity to identify carriers before they have affected offspring. Until recently, achieving this has been challenging due to the wide variety of gene variants underlying most single-gene disorders, rendering DNA-based carrier screening impractical. As a result, current carrier screening is primarily viable for prevalent disorders that can be detected by assessing the protein end-products, such as Tay-Sachs disease and haemoglobin disorders. Carrier screening helps identify couples at risk of producing AR-affected babies and provides a basis for genetic and reproductive counseling. In application, this can drastically decrease recessive disease prevalence. Currently, mutation-specific DNA-based screening is accessible only for certain disorders like cystic fibrosis and for disorders that exhibit heightened prevalence within specific population groups—like Tay-Sachs among French Canadians and Ashkenazi Jews.[26] It's important to differentiate between race as a statistical risk indicator and the actual genetic factors that contribute to causation.

Analysis Using Statistical Methods:

Aside from comparing individual DNA sequences, genetic variations among different populations can be assessed using the FST statistic.[19,20] This statistic quantifies the portion of overall genetic variance, determined by variations in allele frequencies at specific genetic sites within a collection of sampled populations, that exists between these populations as summarized in the example graph below (Figure 2). The range of this statistic spans from 0, indicating no genetic differences, to 1, which signifies complete genetic differentiation between populations. The common benchmark for identifying biological subspecies has been an FST value exceeding 0.25.[8]

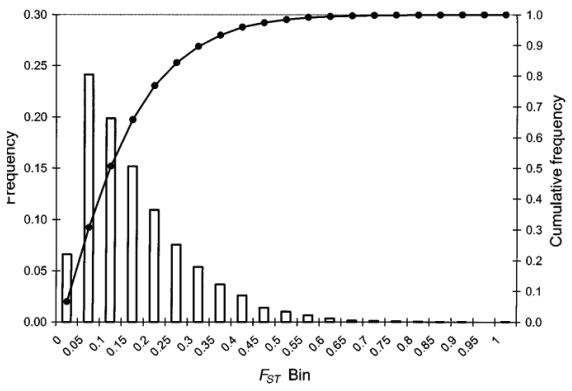

Figure 2: Created by Kittles, R. A. & Weiss, K. M. Figure 4 Distribution of FST values for sites sampled in European Americans, Asians, and African Americans—the major human Old World continental regions. The line is cumulative distribution. Courtesy of Mark Shriver and (1).[8]

To apply this method to a case, the level of heterogeneity within African populations varies across different genetic markers, which holds significance for a genetic mapping technique known as admixture mapping. This approach harnesses the distinct historical backgrounds of admixed populations like African Americans to identify genes linked to disease susceptibility. The concept revolves around the notion that if there are substantial genetic-based disparities in disease prevalence among groups, the risk within each group might stem from shared ancestral alleles. This method involves selecting genome-spanning markers that exhibit varying frequencies within the same groups due to differential ancestry (reference to Figure 2). These markers are then utilized in comparative analyses like case-control studies to pinpoint associations with susceptibility alleles in individuals of mixed ancestry. The occurrence of admixture creates linkage disequilibrium between the disease-causing variants and nearby marker alleles, enhancing the effectiveness of this mapping strategy.[8, 21, 22]

Admixture Mapping and Linkage Disequilibrium (MALD):

Admixture mapping is a statistical technique aimed at identifying associations. It doesn't rely on the assumption of distinct and isolated ancestral populations or imply that such populations were completely isolated. The gradual shift in human genetic diversity globally indicates the absence of strict isolation. This method employs a statistical approach to gene mapping that doesn't make assumptions about the studied samples' overarching distinctions, similarities, evolutionary paths, or biological characteristics. It operates on a scale-dependent basis, proving effective if there has been a separation between the assumed ancestral populations of the admixed individuals and if the admixture has occurred relatively recently (since recombination eventually diminishes linkage disequilibrium). Admixture mapping can be applied to any pair of populations. Still, its success is more probable in cases of recent admixture between individuals from geographically distant regions, especially when genetic differences in disease causation exist. For instance, the phenomenon of linkage disequilibrium stemming from the intermingling of previously isolated west and central African populations with northern Europeans in post-Columbian America can be harnessed, potentially surpassing the effectiveness of conventional mapping methods.[23,24]

Ancestry informative markers (AIMs) are specific genetic locations within the genome that exhibit distinct alleles with significant frequency disparities between populations. These markers are employed in admixture mapping applications to estimate ancestral origins at various levels—ranging from entire populations to smaller subgroups like disease cases and controls, as well as individuals.[8, 25] Despite being a developing approach further investigation is required to establish the optimal criteria for selecting informative AIMs. Key attributes sought in these markers include a substantial degree of genetic differentiation (measured by FST) between parental populations and a correlation between marker frequency and the alleles responsible for the trait being studied.[8] Determining this correlation is challenging. For instance, when conducting admixture mapping among African Americans, it becomes essential to gauge the extent of genetic diversity within West-

DOI: 10.36838/v6i6.4

ern and Central African populations. Disparities in allele frequencies among potential African source populations can significantly impact the accuracy of estimating admixture proportions.[8] This consideration has been explored in ancestral populations from these regions.

Nonetheless, a cautious approach is advisable, particularly when dealing with high FST markers. It's important to ensure that these markers are appropriately derived from identified parental populations that align with the specific admixed study population. Furthermore, these markers must accurately reflect the genomic variability in the parental populations, serving as a reliable proxy.

Ethical Considerations:

While population-specific variants offer valuable insights, ethical considerations must guide their exploration and interpretation. The potential for misusing genetic information to reinforce stereotypes or discrimination requires responsible research practices and communication. Emphasizing that genetic diversity is a natural outcome of human evolution helps counter any misconceptions that may arise from the study of population-specific variants and eliminate bias in subsequent research studies.

Conclusion

Accurately predicting disease susceptibility based on genetic ancestry analysis is still in infancy. Due to the multitude of ethnic groups and debatable forms of categorization, it is impossible to be as specific as one would prefer. In addition, because of other factors such as socioeconomic environment and demographics, genetic predisposition based on unique alleles in certain groups of people may result from correlation, not causation.

This topic may benefit from further research into various ethnic groups, and stronger statistical analysis for elevated credibility. The resulting data can be a powerful resource for clinical geneticists to hone population-adjusted genetic screening programs, especially in otherwise understudied or ignored ethnic groups. Efforts to diversify genetic research participants are essential to gain a more comprehensive understanding of disease risk across different populations. It is important to note that because genetic differences can impact drug metabolism, efficacy, and potential side effects, pharmacogenomics will greatly benefit from integrating such considerations. This will contribute to the current development of personalized medicine and targeted therapies, and make it better applied in everyday clinical healthcare.

Acknowledgments

I would like to sincerely thank my mentor, Dr Kif Liakath-Ali, whose invaluable teaching and expert insight helped guide this paper, and the rest of the incredible team at Indigo Research for their support, resources, and advice. I also want to thank my family for their encouragement and motivation.

References

1. Dries, D. L. Genetic Ancestry, Population Admixture, and the Genetic Epidemiology of Complex Disease. *Circ. Cardiovasc. Genet.* **2**, 540–543 (2009).
2. Jablonski, N. G. The evolution of human skin pigmentation involved the interactions of genetic, environmental, and cultural variables. *Pigment Cell Melanoma Res*. **34**, 707–729 (2021).
3. McLaren, P. J. & Fellay, J. HIV-1 and human genetic variation. *Nat. Rev. Genet*. **22**, 645–657 (2021).
4. Andrews, C. A. (2010) Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. Nature Education Knowledge 3.
5. Lenart, P., Bienertová-Vašků, J. & Berec, L. Learning mitigates genetic drift. *Sci. Rep*. **12**, 20403 (2022).
6. Gurdasani, D., Barroso, I., Zeggini, E. & Sandhu, M. S. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet*. **20**, 520–535 (2019).
7. Di Rienzo, A. Population genetics models of common diseases. *Curr. Opin. Genet. Dev*. **16**, 630–636 (2006).
8. Kittles, R. A. & Weiss, K. M. Race, Ancestry, and Genes: Implications for Defining Disease Risk. *Annu. Rev. Genomics Hum. Genet*. **4**, 33–67 (2003).
9. Barton, N. H. & Charlesworth, B. Genetic Revolutions, Founder Effects, and Speciation. *Annu. Rev. Ecol. Syst*. **15**, 133–164 (1984).
10. Blencowe, H. et al. Rare single gene disorders: estimating baseline prevalence and outcomes worldwide. *J. Community Genet*. **9**, 397–406 (2018).
11. Corry, P. C. Consanguinity and Prevalence Patterns of Inherited Disease in the UK Pakistani Community. *Hum. Hered*. **77**, 207–216 (2014).
12. Ahmed, A. E. *JPHDC* **2**, (2016).
13. Fareed, M. & Afzal, M. Genetics of consanguinity and inbreeding in health and disease. *Ann. Hum. Biol*. **44**, 99–107 (2017).
14. Oniya, O., Neves, K., Ahmed, B. & Konje, J. C. A review of the reproductive consequences of consanguinity. *Eur. J. Obstet. Gynecol. Reprod. Biol*. **232**, 87–96 (2019).
15. Piel, F. B., Steinberg, M. H. & Rees, D. C. Sickle Cell Disease. *N. Engl. J. Med*. **376**, 1561–1573 (2017).
16. Kato, G. J. et al. Sickle cell disease. *Nat. Rev. Dis. Primer* **4**, 18010 (2018).
17. Rees, D. C., Williams, T. N. & Gladwin, M. T. Sickle-cell disease. *The Lancet* **376**, 2018–2031 (2010).
18. Williams, T. N. Sickle Cell Disease in Sub-Saharan Africa. *Glob. Hematol*. **30**, 343–358 (2016).
19. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
20. Michalakis, Y. & Excoffier, L. Michalakis, Y, Excoffier, L. A generic estimation of population subdivision using distances between alleles with special interest to microsatellite loci. *Genetics* **142**, 1061–4 (1996).
21. Chakraborty, R. & Weiss, K. M. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl. Acad. Sci.* **85**, 9119–9123 (1988).
22. Collins-Schramm, H. E. et al. Ethnic-Difference Markers for Use in Mapping by Admixture Linkage Disequilibrium. *Am. J. Hum. Genet*. **70**, 737–750 (2002).
23. Anderson, E. C. & Thompson, E. A. A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics* **160**, 1217–1229 (2002).

24.McKeigue, P. M., Carpenter, J. R., Parra, E. J. & Shriver, M. D. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann. Hum. Genet*. **64**, 171–186 (2000).
25.Pfaff, C. L. et al. Population Structure in Admixed Populations: Effect of Admixture Dynamics on the Pattern of Linkage Disequilibrium. *Am. J. Hum. Genet*. **68**, 198–207 (2001).
26.Palomaki, G. E., Williams, J., Haddow, J. E. & Natowicz, M. R. Tay-sachs disease in persons of French-Canadian heritage in northern New England. *Am. J. Med. Genet*. **56**, 409–412 (1995).

Author

Fatima Zahra Al-Sakkaf has been enrolled onto the Neuroscience B.Sc (Hons) course at the University of Central Lancashire, England. She is currently exploring interests in the genetics underlying neurological disorders as well as neuropharmacological developments. She hopes to contribute further to this field in the future.

DOI: 10.36838/v6i6.4