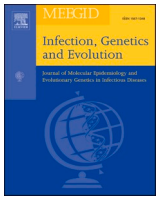
Infection, Genetics and Evolution 125 (2024) 105680

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Unveiling ancestral threads: Exploring CCR5 Δ32 mutation frequencies in Colombian populations for HIV/AIDS therapeutics

Alejandro Barrios-Navas [a], Thanh Long Nguyen [b], Juan Esteban Gallo [c], Leonardo Mariño-Ramírez [c], José María Satizabal Soto [d], Adalberto Sánchez [d], I. King Jordan [b], Augusto Valderrama-Aguirre [a,*]

[a] Grupo Instituto de Investigaciones Biomédicas, Departamento de Ciencias Biológicas, Facultad de Ciencias, Universidad de Los Andes, Bogotá, DC, Colombia
[b] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[c] National Institute on Minority Health and Health Disparities, National Institutes of Health, Rockville, MD, USA
[d] Universidad del Valle, Faculty of Health, School of Basic Sciences, Cali, Valle del Cauca, Colombia

ARTICLE INFO

Keywords:
HIV
AIDS
CCR5
Delta 32
Ancestry
Stem cell

ABSTRACT

AIDS remains a significant global health challenge since its emergence in 1981, with millions of deaths and new cases every year. The CCR5 Δ32 genetic deletion confers immunity to HIV infection by altering a cell membrane protein crucial for viral entry. Stem cell transplants from homozygous carriers of this mutation to HIV-infected individuals have resulted in viral load reduction and disease remission, suggesting a potential therapeutic avenue. This study aims to investigate the relationship between genetic ancestry and the frequency of the CCR5 Δ32 mutation in Colombian populations, exploring the feasibility of targeted donor searches based on ancestry composition. Utilizing genomic data from the CÓDIGO-Colombia consortium, comprising 532 individuals, the study assessed the presence of the CCR5 Δ32 mutation and examined if the population was on Hardy-Weinberg equilibrium. Individuals were stratified into clusters based on African, American, and European ancestry percentages, with logistic regression analysis performed to evaluate the association between ancestry and mutation frequency. Additionally, global genomic databases were utilized to visualize the worldwide distribution of the mutation. The findings revealed a significant positive association between European ancestry and the CCR5 Δ32 mutation frequency, underscoring its relevance in donor selection. African and American ancestry showed negative but non-significant associations with CCR5 Δ32 frequency, which may be attributed to the study's limitations. These results emphasize the potential importance of considering ancestry in donor selection strategies, reveal the scarcity of potential donors in Colombia, and underscore the need to consider donors from other populations with mainly European ancestry if the CCR5 Δ32 stem cell transplant becomes a routine treatment for HIV/AIDS in Colombia.

1. Introduction

The human immunodeficiency virus (HIV) stands as a formidable challenge in modern medicine, being the causative agent behind acquired immunodeficiency syndrome (AIDS), a globally significant disease. As of 2022, the number of individuals living with HIV reached 39 million, with 40.4 million AIDS-related deaths recorded since the identification of the first case in 1981 (*Global HIV and AIDS Statistics — Fact Sheet,* 2024).

HIV infection leads to a progressive decline in CD4+ T-lymphocytes (Cloyd et al., 2001), crucial components of the immune system. This decline ultimately results in AIDS, characterized by increased susceptibility to opportunistic infections and malignancies, often leading to death (HIV and AIDS, 2024). Despite advances in antiretroviral therapy, which have extended the lifespan of HIV-infected individuals to an average of 54.9 years (Gueler et al., 2017), AIDS remains an incurable condition, as declared by the World Health Organization (HIV and AIDS, 2024).

The process of cell infection occurs through protein-protein interactions between cell surface proteins and viral envelope

Abbreviations: CCR5, C-C chemokine receptor type 5; CD4, Cluster of differentiation 4; WT, Wild type.

* Corresponding author at: Cra 1 Nº 18A - 12, Bogotá, DC 111711, Colombia.
E-mail address: a.valderramaa@uniandes.edu.co (A. Valderrama-Aguirre).

https://doi.org/10.1016/j.meegid.2024.105680
Received 25 July 2024; Received in revised form 24 September 2024; Accepted 3 October 2024
Available online 5 October 2024

glycoproteins Env, composed of the subunits gp120 and gp41. Initially, gp120 must bind to the primary receptor, which is the cluster of differentiation 4 (CD4), and to the secondary co-receptor. X4 strain viruses use the protein CXCR4 as a co-receptor, while R5 strain viruses, the focus of this study and the most frequently transmitted (Sarker, 2023), use the protein CCR5. This generates a conformational change in the gp120/gp41 oligomer resulting in the approximation of the cell and viral membranes, allowing fusion (Wang et al., 2023). The Δ32 deletion in the CCR5 gene, results in the deformation of the CCR5 co-receptor (Berkowitz et al., 1998). This mutation makes heterozygotes (Δ32/Normal) less susceptible to infection and slower to progress to AIDS, while homozygotes (Δ32/Δ32) are resistant to infection by R5 type HIV strains (Novembre et al., 2005).

In 2007, a group of researchers from the Charité hospital in Berlin, an institution belonging to the Faculty of Medicine of the Free University of Berlin and Humboldt University of Berlin (Hütter et al., 2009), performed a stem cell transplant in an HIV positive subject with leukemia. The donor's bone marrow was not only compatible with the recipient but also homozygous for the CCR5 Δ32 mutation. The patient stopped antiretroviral therapy and received a second transplant from the same donor in 2008. Over the next three years, several biopsies and blood tests were performed, indicating a reduction in viral load to undetectable levels. The fact that these undetectable levels were maintained until the patient's death, 13 years later, made the now-called “Berlin Patient” the first cured case of HIV in history. Since then, six other patients who underwent similar processes have been announced and are considered cured of HIV: the London Patient, the New York Patient, the City of Hope Patient, the Düsseldorf Patient, the Geneva patient and the recently announced next Berlin patient (Payra et al., 2023). (Beaney, 2024) (NBC News, 2024) (IA Society, 2024) These cases suggest that this approach could become a formally accepted practice when dealing with HIV infection and currently there are at least 4 clinical trials related to this concept reported at ClinicalTrial.gov.

The CCR5 Δ32 mutation, while being documented in diverse ethnic groups worldwide, is found mainly in European populations. Its prevalence has a pronounced north-south trend, with frequencies ranging from 16 % in Nordic countries to 6 % and 4 % in Italy and Greece, respectively (Novembre et al., 2005). The mutation's higher presence in some populations makes it more likely to find carriers among individuals with certain ancestry frequencies.

Against this setting, in a prospective scenario wherein the quest for bone marrow donors exhibiting CCR5 Δ32 homozygosity evolves into a routine aspect of therapeutic protocols against HIV infection, a comprehensive understanding of the mutation's prevalence across diverse global subpopulations and its linkage with genetic ancestry becomes imperative, particularly within the framework of personalized medicine. Colombia, characterized by its intricate genetic landscape comprising African, American, and European genetic constituents, emerges as a pertinent case study for the prevalence and distribution of CCR5 Δ32 mutations. We hypothesized that the frequency of the Δ32 mutation is likely to vary widely among different Colombian populations, making it essential to determine if the mutation is more prevalent in individuals with specific ancestries. We tested this hypothesis using participant genomic variant data and genetic ancestry profiles taken from the CÓDIGO-Colombia consortium.

2. Materials and methods

2.1. Reception and handling of Colombian data

The genomic data used in this study were obtained from the CÓDIGO-Colombia database (https://codigo.biosci.gatech.edu/) (Chande et al., 2021; Chande et al., 2020), a collection of de-identified genomic variant data from diverse Colombian populations, contributed by collaborating investigators and laboratories, under the terms of data use agreements. Data were generated in previous studies using whole genome sequencing, whole exome sequencing, or genome-wide genotyping techniques, with all studies conducted under specific IRB approvals in consequence study participant inclusion and exclusion criteria, and ethical review board approval have been previously published (Conley et al., 2017; The 1000 Genomes Project Consortium, 2015; Li et al., 2008; Mathias et al., 2016; Medina Rivas et al., 2016; Nagar et al., 2019; Nievergelt et al., 2019; Reich et al., 2012; Wojcik et al., 2019). As sample collections were conducted independently by CÓDIGO-Colombia contributors and were not tailored to this study's design, the sampling proportion relative to the Colombian population is not fully representative. Since the CÓDIGO genomic data were not collected specifically for this study, and our team does not have access to subject identifiers linked to the primary data, this study is not considered human subjects research, in line with the NIH Revised Common Rule for the Protection of Human Subjects (https://grants.nih.gov/policy/humansubjects/hs-decision.htm).

Specific datasets about CCR5 Δ32 mutation included two population samplings in Antioquia, encompassing a total of 416 individuals, as well as one population sampling of 116 individuals from Valle del Cauca. Antioquia and Valle del Cauca are politically divided regions in Colombia, known as departments, and the samples analyzed here were collected mainly from their capital cities, Medellín and Cali respectively. The combined dataset represented a total of 532 individuals. These datasets provide insights into the presence or absence of mutations, their occurrence in homozygous or heterozygous states, and the individual proportions of genetic ancestry, expressed as proportions of African, European, and American ancestries. Utilizing this information, we computed the deletion frequency within each population and conducted the HWExact() test of the R package HardyWeinberg (Graffelman, 2015; Graffelman and Camarena, 2008; Graffelman and Weir, 2016) to assess the adherence of frequencies to Hardy-Weinberg equilibrium. All participants signed informed consent indicating their understanding of the study risks and their willingness to participate. Approval for studies was obtained by ethics committees from the universities that participated in the CÓDIGO-Colombia consortium.

2.2. Comparison of ancestry percentages

Doughnut charts were generated to visually convey broad differences in ancestry distribution between the sampled populations. Each doughnut represents central tendencies (percentages) of African, European, and American ancestry in each sample. The significance of those disparities was verified with a Mann-Whitney test. Subsequently, individuals were stratified into four clusters based on their ancestral lineage using the k-means clustering algorithm. These clusters delineated groups primarily characterized by African, American, European, and admixed ancestries. Following this, a bar graph was constructed to illustrate the incidence of the deletion and its absence across each cluster. Moreover, logistic regression analyses were conducted on clusters representing African, American, and European ancestries to observe whether there is a positive or negative association between ancestry percentages and the CCR5 Δ32 mutation frequency. Additionally, the statistical significance of the analyzed ancestral component as a variable affecting mutation presence was assessed. The findings were visually depicted using box plots featuring two columns: one denoting individuals lacking the mutation and the other individuals possessing it.

2.3. Acquisition, processing and frequency visualization of global data

We obtained allelic frequency data concerning the mutation from diverse populations worldwide. This information was sourced from four openly accessible databases: the National Center for Biotechnology Information (NCBI) (National Center for Biotechnology Information, 2024), the Genome Aggregation Database (GnomAD) (gnomAD, 2024), the Ensembl Genome Browser for Vertebrates (Ensembl Genome Browser, 2024), and the Japanese Multi Omics Reference Panel (JMorp)

Table 1
Deletion frequencies and Hardy-Weinberg equilibrium test.

Population sampling	Ancestry group	n	Del %	Hom %	Het %	P-Value
Antioquia 1	Admixed	94	4.26	0.00	8.51	0.1949
Antioquia 2	Admixed	322	2.64	0.31	4.66	1.000
Valle del Cauca	Admixed	116	1.29	0.00	2.59	1.000
Total	–	532	2.63	0.19	4.88	0.3056

Del %: Deletion percentage; Hom %: Homozygous percentage; Het %: Heterozygous percentage; P-Value: P-Value of the exact test.
Antioquia and Valle del Cauca are political divisions of territories in Colombia

(jMorp, 2024). In NCBI, the query 'rs333' was used in the SNP database, selecting *'[Homo sapiens]'* from the search results, and the data corresponding to the ALFA project were downloaded. In Ensembl, the 'rs333' query was used after selecting 'Human' as the species, and data from the 1000 Genomes Project were retrieved from the 'Population Genetics' section. In GnomAD, the same query was run using the GRCh38 human reference genome assembly. In JMorp, the query was applied to the 54KJPN dataset. Data originating from the same geographical region were aggregated and amalgamated into a unified variable. Following the acquisition and organization of genomic data, three heatmaps were constructed to elucidate the distribution of mutation frequencies worldwide. The initial heatmap integrated data from cities, regions, and countries, while the second encompassed data from continents and subcontinents, including information of the United States of America, due to its expansive territorial surface. Lastly, the third heatmap focused on data from groups not confined to specific cities or regions, such as migrant populations.

3. Results

3.1. Allelic and genotypic frequencies and hardy-Weinberg equilibrium

Data from CÓDIGO-Colombia comprise a total of 416 individuals from the department of Antioquia (94 from the first sampling and 322 from the second) and 116 individuals from Valle del Cauca. The frequencies of deletion, homozygotes, and heterozygotes are presented in Table 1. The same table includes the exact test's *p*-value, used to verify if the populations are in Hardy-Weinberg equilibrium. None of the three populations showed evidence for deviation from Hardy-Weinberg equilibrium.

3.2. Ancestry analysis

Doughnut plots in Fig. 2 unveil consistent shifts in ancestral proportions: across all three populations, the prevalence of European ancestry is higher among individuals with the CCR5 Δ32 mutation, whereas American ancestry is consistently lower among individuals with the mutation. African ancestry is lower among individuals with the mutation in two of the three populations. A Mann-Whitney test was conducted in each sample to assess the statistical significance of these trends. The resulting *p*-values from these tests are reported in Table 1S. Employing a significance level of 0.05, the difference in African ancestry is statistically significant in the second Antioquia sample, while the difference in American ancestry is significant solely in the first Antioquia sample. Conversely, the difference in European ancestry is significant in both Antioquia samples.

Utilizing the k-means algorithm in R, the sampled individuals were categorized into 4 clusters based on their ancestry: those predominantly of African descent, those primarily of European lineage, those predominantly of American origin, and those lacking a predominant lineage, indicative of admixture. This clustering is illustrated in the ternary plot shown in Fig. 2.

The frequency of the deletion, as well as the frequency of homozygous and heterozygous individuals in each of the four remaining clusters, are shown in Table 2.

Table 2
Deletion frequencies for every cluster formed by the k-means function.

Cluster	n	Del%	95 % CI	Hom%	Het%
American	59	3.39	[0.12, 6.66]	1.69	3.39
African	24	0.00	[0.00, 0.00]	0.00	0.00
European	139	4.32	[1.93, 6.71]	0.00	8.63
Admixed	310	1.94	[0.85, 3.02]	0.00	3.87

Del %: Deletion percentage; 95 % CI: 95 % Confidence interval; Hom %: Homozygous percentage; Het %: Heterozygous percentage.

In Fig. 3, the binomial logistic regression coefficients and *p*-values are presented. While the association between mutation frequency and the percentage of African and American ancestry exhibits a negative trend (indicated by coefficients below zero), this relationship fails to attain statistical significance, as the respective p-values exceed 0.05. Conversely, for European ancestry, where the coefficient surpasses zero, signifying a positive association, the p-value falls below 0.05, indicating statistical significance.

3.3. Global distribution

Allelic frequencies from various cities, regions, countries, continents, and other populations, as shown in Table 2S, were used to generate the heatmap depicted in Fig. 4.

Fig. 4 highlights the mutation's concentration in European populations, where no population falls below a 5 % frequency threshold. Notably, certain European cities exhibit frequencies soaring to 25 % (refer to Table 2S). The American continent also exhibits a significant presence of the mutation, with multiple samples demonstrating frequencies ranging between 2 % and 5 %. In contrast, Africa and Asia manifest lower prevalence rates, with Lahore, Pakistan representing the sole region surpassing a 1 % frequency (see Table 2S).

The same figure accentuates the continental disparities, with Europe boasting the highest frequency (10.8 %), followed by Latin America and the United States of America with frequencies of 3.4 % and 3.2 %, respectively. Africa, South Asia, and the Middle East exhibit frequencies ranging between 1 % and 2 %, while East Asia registers the lowest frequency (0.04 %).

The right half of Fig. 4, which contains populations without a specific geographical location, four populations stand out with frequencies exceeding 5 %: the Amish (18 %), Ashkenazi Jews (13 %), residents of Utah with ancestry from northern and western Europe (10.6 %), and Americans in the United States (7.5 %).

4. Discussion

A significant positive association was observed between the European ancestry component and the frequency of the CCR5 Δ32 mutation in Colombian populations. In every sample presented in Fig. 1, the fraction with the mutation exhibited a higher percentage of European ancestry compared to the fraction without the mutation. This difference was statistically significant in two of the three samples. Additionally, Fig. 3 indicates that the European cluster was the only one to show a significant result in logistic regression analysis. This observation aligns with the global distribution of the mutation, as depicted in Fig. 4. The European continent exhibits the highest mutation frequency, reaching 10.8 %, which significantly surpasses frequencies observed in other continents and subcontinents analyzed in Fig. 4, where frequencies do not exceed 3.5 % (Table 2S). Examination of cities and regions in Fig. 4 reveals that those with allelic frequencies exceeding 5 % are European populations, including Danes (25 %), Finns (13.5 %), Estonians (10.2 %), Swedes (10.7 %), Iberian populations in Spain (8.9 %), British (8.5 %), and Tuscans in Italy (6.5 %). Regarding the populations without a specific territory, the ancestry of three out of the four groups with

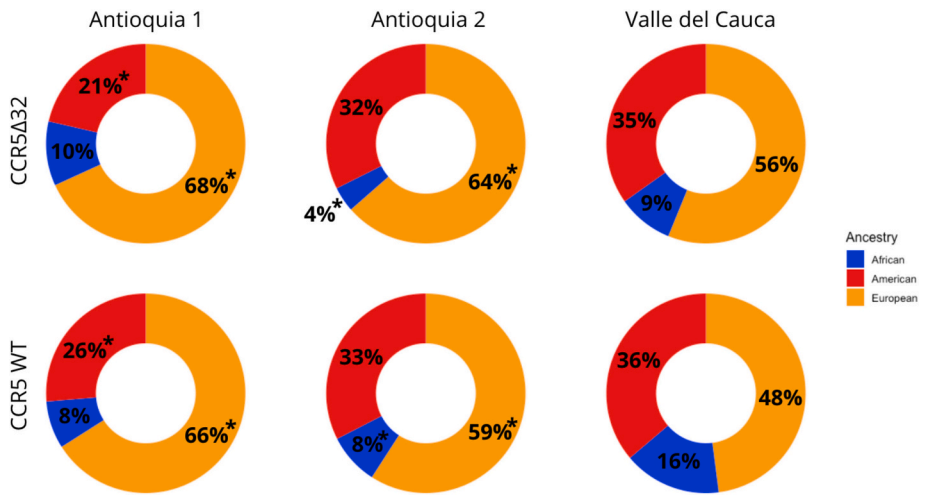


Fig. 1. Doughnut charts illustrating the frequencies of ancestry components for each sample. Asterisks denote statistically significant differences in ancestry percentages between individuals carrying the mutation and those with the wild-type genotype.

mutation frequencies above 5 % (Amish, Ashkenazi Jews, residents of Utah with ancestry from northern and western Europe, and Native Americans from the United States of America) is predominantly European. The Amish, originating from Switzerland, arrived in America between 1717 and 1750 with only 500 individuals, maintaining notable isolation from other American populations (Crowley, 1978). This founder effect suggests that even today, most of their ancestry remains European. The Ashkenazi Jews, with recent ancestry in Central and Eastern Europe, emerged as an ethno-religious group in the German region of Rhineland in the 10th century (Waldman et al., 2022). According to a 2013 study, approximately 81 % of Ashkenazi ancestry is European (Costa et al., 2013). Finally, the sampled residents of Utah are explicitly descended from northern European origins. This demographic composition makes Native Americans the only group without predominantly European ancestry to exhibit such high allelic frequencies. The global distribution in Fig. 4 is also consistent with previous studies, that observed the highest frequencies of CCR5Δ32 in Northern Europe and a gradual decline Southward (Solloch et al., 2017).

On the other hand, the associations with African and American ancestries in Colombia, though potentially negative, were not statistically significant according to our results. In Fig. 1, these associations were significant in only one of the three samples. While American ancestry was consistently higher in the non-mutated fraction, African ancestry varied more irregularly, being higher in two of the non-mutated fractions and one of the mutated fractions. Fig. 3 shows that the logistic regressions for the African and American clusters have negative coefficients, indicating a negative association. The African cluster had the largest coefficient among the three regressions, suggesting a greater impact on the deletion's frequency; however, neither regression was statistically significant. In Fig. 4, the Americas exhibited the second-highest mutation frequencies after Europe, with regions like Cuba, the Dominican Republic, Mexico, and Puerto Rico showing percentages between 2 % and 5 % (Table 2S). Nevertheless, the extensive migratory influx from Europe and Africa since 1492 has resulted in significant genetic admixture over 15 generations (Ongaro et al., 2021), complicating the inference of the specific impact of American ancestry on mutation frequencies. In contrast, Africa had the lowest mutation frequencies among the sampled regions, with none of the individuals in the five sampled African populations carrying the mutation (Table 2S). Unlike the American samples, four of the African samples were from specific ethnic groups (the Esan in Nigeria, the Luhya in Kenya, the Mende in Sierra Leone, and the Yoruba in Nigeria), reinforcing the potential existence of a negative association between African ancestry and the presence of the deletion.

Regarding public health measures, the low number of CCR5 Δ32 carriers in Colombia indicates that finding homozygous donors within the country may be extremely challenging. This suggests that importing stem cells from countries with higher mutation frequencies might be a more effective approach. The low CCR5 Δ32 mutation frequencies in Antioquia and Valle del Cauca are particularly discouraging. In Antioquia, despite its substantial European ancestry, only one homozygous individual was found in two samples totaling 416 individuals, resulting in a frequency of just 0.24 %. The frequency of heterozygotes is higher, at 4.88 % across three samples, as shown in Table 1. This frequency increases when focusing on individuals with higher levels of European ancestry, with the European ancestry cluster having a heterozygous frequency of 8.63 %, as shown in Table 2. Therefore, finding heterozygous donors in Colombia may be somewhat easier. However, the effectiveness of transplants using heterozygous donors remains uncertain. In two cases, referred to as the Boston patients, HIV carriers were treated with transplants from heterozygous donors. In both instances, the virus re-emerged weeks after discontinuing antiretroviral therapy (Payra et al., 2023). Conversely, in a more recent case, known as the next Berlin patient, a heterozygous carrier of the mutation received a transplant from a heterozygous donor. The transplant was performed in 2015, antiretroviral therapy was halted in 2018, and as of July 2024, the virus has not resurfaced, marking this as the seventh successful case of this therapy (Beaney, 2024). Further information is needed regarding the effectiveness of therapy with heterozygous donors, but if proven to be more effective in the future, it would facilitate finding donors in Colombian territory. The case of the Geneva patient, the sixth individual to receive a successful transplant, is particularly noteworthy since the donor who was not a carrier of the mutation (Payra et al., 2023). Nevertheless, like procedures involving heterozygous donors, further research is required to confirm the viability of this approach.

On the one hand, if Colombia considers adopting this therapeutic intervention for managing HIV patients, sourcing hematopoietic stem cells from countries with higher CCR5 Δ32 mutation frequencies might be a more viable solution. On the other hand, if there is a preference to rely on local donors within Colombia, focusing on individuals with high European ancestry should increase the likelihood of finding CCR5 Δ32 homozygotes and heterozygotes.

It is essential to acknowledge the limitations of the sampling. Our dataset comprises data from only 532 individuals, with 78 % of the sampling conducted in Antioquia, a department characterized by a high European component and a reduced American component (Healy et al., 2017; Wang et al., 2008), especially when compared to the country's overall ancestry percentages, which include a European percentage of

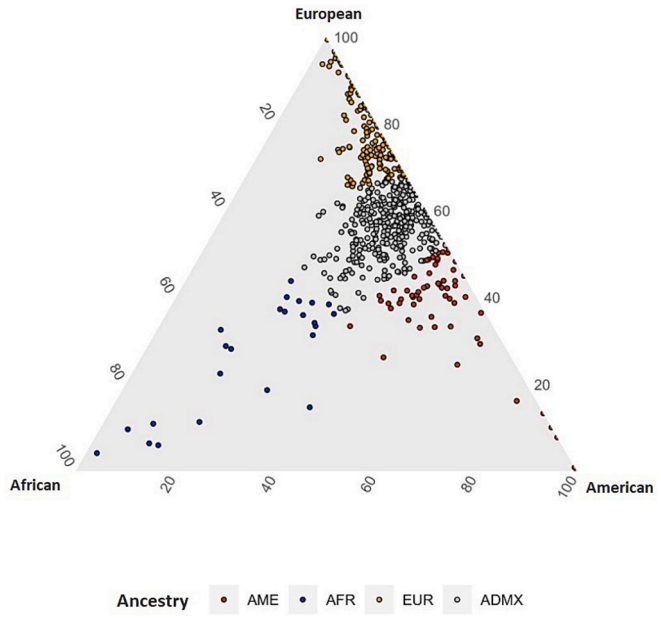


Fig. 2. Ternary plot with the 4 clusters of individuals, segregated based on their ancestry.

40.2 %, an African percentage of 11 %, and an American component of 48.8 % (Rojas et al., 2010). This imbalance becomes evident when categorizing individuals into four clusters based on ancestry, where European ancestry continues to predominate in individuals classified as mestizo and is also present in significant proportions in individuals belonging to the African ancestry cluster, as illustrated in Fig. 2.

5. Conclusion

In conclusion, our study highlights a significant positive association between European ancestry and the frequency of the CCR5Δ32 mutation in Colombia, with European ancestry subgroups consistently exhibiting higher mutation frequencies. This observation aligns with previous research indicating the highest prevalence of this mutation in Northern Europe. In contrast, the negative associations with African and American ancestries were not statistically significant, indicating a complex and less straightforward relationship. The low prevalence of the mutation in Colombia, particularly among individuals with non-European ancestry, poses challenges for finding suitable homozygous donors for therapeutic purposes. Consequently, importing stem cells from regions with higher mutation frequencies may be a more effective strategy for HIV treatment in Colombia.

Declaration of generative AI

During the preparation of this work the author used ChatGPT in order to review the syntax and grammar of the document. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This work was supported by a FAPA project (PVI0122029) from Universidad de Los Andes, granted to Prof. Augusto Valderrama-Aguirre. This work was also supported by the Division of Intramural Research of the National Institute on Minority Health and Health Disparities at the National Institutes of Health to Leonardo Mariño-Ramirez (award numbers: 1ZIAMD000016 and 1ZIAMD000018); and the National Institutes of Health Distinguished Scholars Program to Leonardo Mariño-Ramirez. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CRediT authorship contribution statement

Alejandro Barrios-Navas: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Thanh Long Nguyen:** Writing – review & editing, Visualization, Methodology, Data curation. **Juan Esteban Gallo:** Writing – review & editing, Methodology, Investigation, Data curation. **Leonardo Mariño-Ramírez:** Writing – review & editing, Software, Resources, Project administration, Funding acquisition. **José María Satizabal Soto:** Writing – review & editing, Resources. **Adalberto Sánchez:** Writing – review & editing, Resources. **I. King Jordan:** Writing – review & editing, Software, Resources, Project

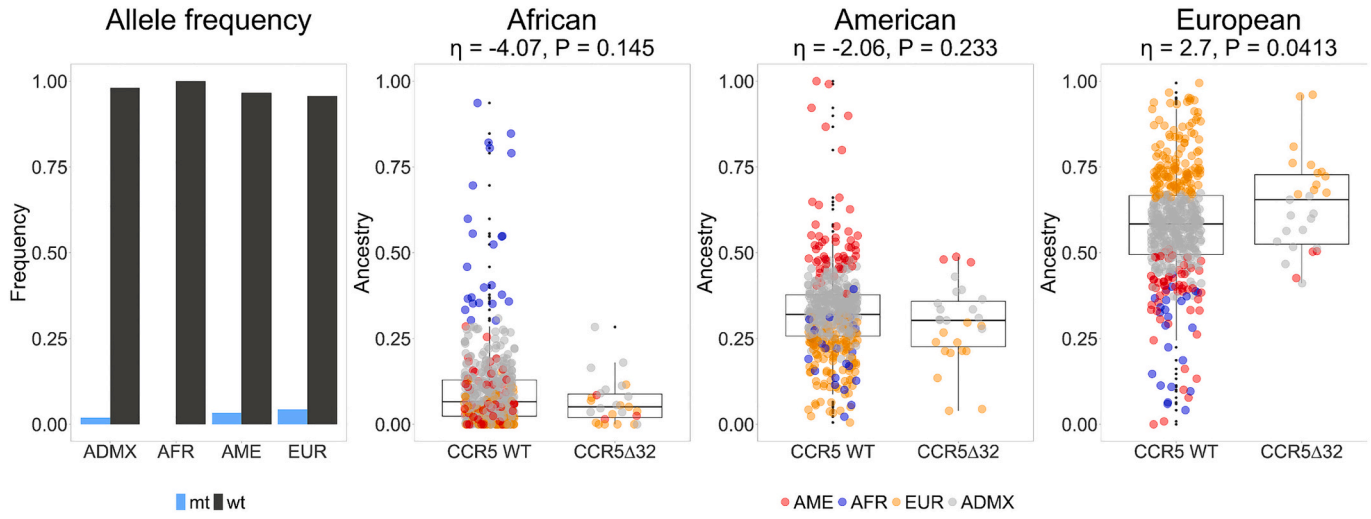


Fig. 3. Bar plot (left) with ancestral associations for the mutation in each of the clusters created by the kmeans function, frequencies of the deletion (in blue) and its absence (in black) are depicted. The bar plot is followed by box plots featuring two columns: the first representing individuals with the mutation and the second those without. The Y-axis on the box plots denotes percentages of African ancestry (left), American ancestry (center), and European ancestry (right). Red dots represent the American cluster, blue denotes the African cluster, yellow signifies the European cluster, and gray indicates the Admixed cluster. Each box plot shows the slope (η) and the *P*-value (P) of the logistic regression. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

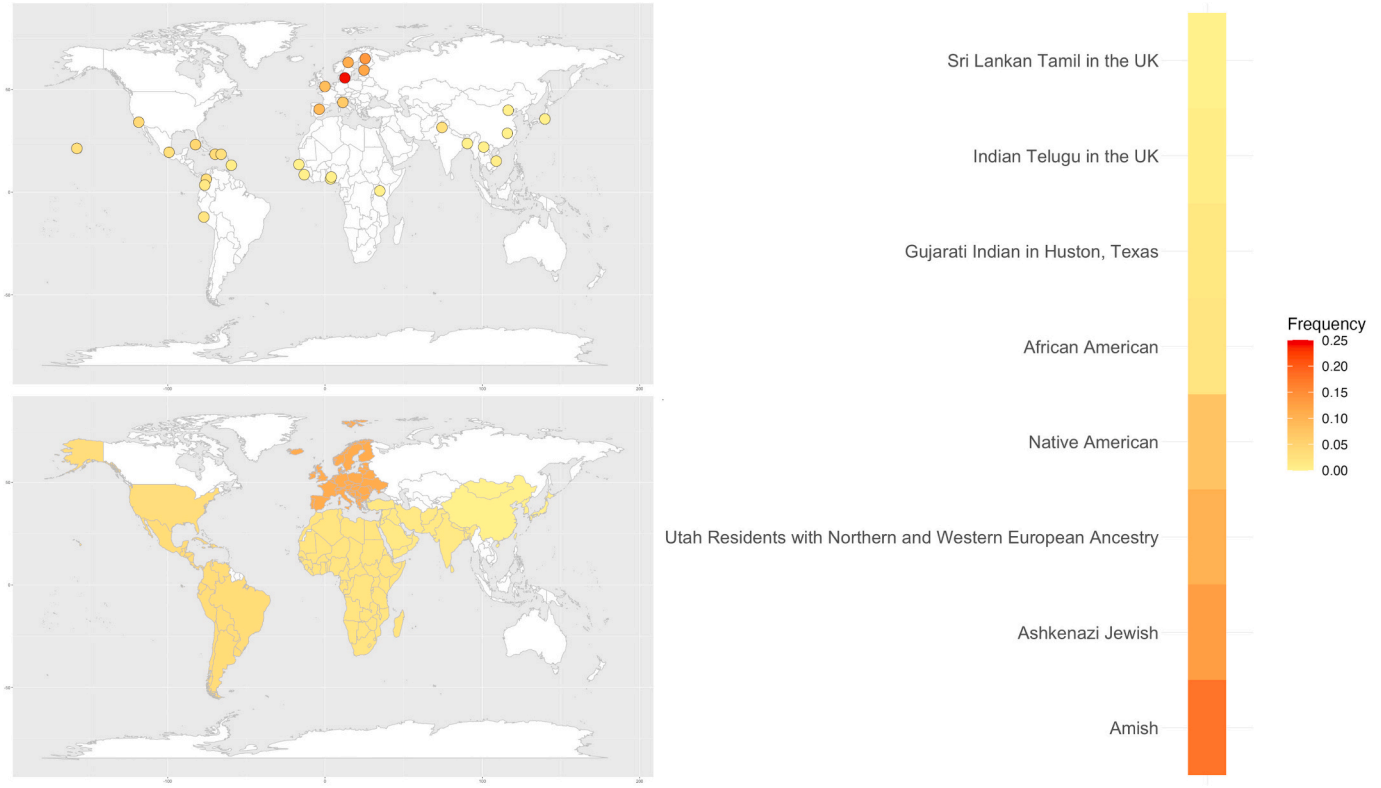


Fig. 4. Heatmap depicting the frequency of the CCR5 Δ32 mutation in cities, countries, and regions worldwide (up left), heatmap depicting the frequency of the CCR5 Δ32 mutation in continents and subcontinents (down left) and heatmap displaying the frequencies of the mutation in groups without a specific territory (right).

administration, Methodology, Investigation. **Augusto Valderrama-Aguirre:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The raw data required to reproduce the above findings cannot be shared at this time due to legal and ethical reasons. The data belongs to a consortium of scientists dedicated to population genomics in Colombia. Every collection has been deposited in CÓDIGO-Colombia web server and summary statistics are available at the website: https://codigo.biosci.gatech.edu/.

Acknowledgements

We extend our gratitude to Alejandro Matamoros for his invaluable assistance in retrieving information from CÓDIOGO-Colombia. Additionally, we thank Jarrison Campos for his expertise and support in utilizing R for data analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2024.105680.

References

gnomAD, 2024. 3–46373452-TACAGTCAGTATCAATTCTGGAAGAATTTCCAG-T | gnomAD v4.1.0. https://gnomad.broadinstitute.org/variant/3-46373452-TACAGTCAGTATCAATTCTGGAAGAATTTCCAG-T?dataset=gnomad_r4.
Beaney, A. Seventh patient 'cured' of HIV among highlights at AIDS 2024. Clin. Trials Arena https://www.clinicaltrialsarena.com/news/hiv-aids-conference-july-2024/ (2024).
Berkowitz, R.D., et al., 1998. CCR5- and CXCR4-utilizing strains of human immunodeficiency virus type 1 exhibit differential tropism and pathogenesis in vivo. J. Virol. 72, 10108–10117.
Chande, A.T., et al., 2020. The phenotypic consequences of genetic divergence between admixed Latin American populations: antioquia and Chocó, Colombia. Genome Biol. Evol. 12, 1516–1527.
Chande, A.T., et al., 2021. The impact of ethnicity and genetic ancestry on disease prevalence and risk in Colombia. Front. Genet. 12, 690366.
Cloyd, Miles W., Chen, Jenny J.Y., Adeqboyega, Patrick, Wang, Liqiang, 2001. How does HIV cause depletion of CD4 lymphocytes? A mechanism involving virus signaling through its cellular receptors. CMM 1, 545–550.
Conley, A.B., et al., 2017. A comparative analysis of genetic ancestry and admixture in the Colombian populations of Chocó and Medellín. G3 7, 3435–3447.
Costa, M.D., et al., 2013. A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. Nat. Commun. 4, 2543.
Crowley, W.K., 1978. Old order amish settlement: diffusion and growth. Ann. Assoc. Am. Geogr. 68, 249–264.
Ensembl Genome Browser, 2024. Ensembl Genome Browser 112. https://www.ensembl.org/index.html.
Evans, D. J. Viral receptors. In Encyclopedia of Virology 319–324 (Elsevier, 2008). doi: https://doi.org/10.1016/B978-012374410-4.00531-8.
Global HIV & AIDS Statistics — Fact Sheet. https://www.unaids.org/en/resources/fact-sheet, 2024.
Graffelman, J., 2015. Exploring Diallelic genetic markers: the HardyWeinberg package. J. Stat. Softw. 64.
Graffelman, J., Camarena, J.M., 2008. Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. Hum. Hered. 65, 77–84.
Graffelman, J., Weir, B.S., 2016. Testing for Hardy–Weinberg equilibrium at biallelic genetic markers on the X chromosome. Heredity 116, 558–568.
Gueler, A., et al., 2017. Life expectancy in HIV-positive persons in Switzerland: matched comparison with general population. AIDS 31, 427–436.
Healy, M.E., et al., 2017. Social-group identity and population substructure in admixed populations in New Mexico and Latin America. PLoS ONE 12, e0185503.
HIV and AIDS, 2024. https://www.who.int/news-room/fact-sheets/detail/hiv-aids.

Hütter, G., et al., 2009. Long-term control of HIV by *CCR5* Delta32/Delta32 stem-cell transplantation. N. Engl. J. Med. 360, 692–698.
IA Society, 2024. World's Seventh HIV Cure Case Following Stem Cell Transplant Among Scientific Highlights at AIDS. https://www.iasociety.org/news-release/worlds-seventh-hiv-cure-case-following-stem-cell-transplant-among-scientific.
jMorp, 2024. jMorp > chr3:46373452 TACAGTCAGTATCAATTCTGGAAGAATTTCCAG/T. https://jmorp.megabank.tohoku.ac.jp/genome-variations/sr-snvindel/01961cd26609dc41f.
Li, J.Z., et al., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100–1104.
Mathias, R.A., et al., 2016. A continuum of admixture in the Western hemisphere revealed by the African diaspora genome. Nat. Commun. 7, 12522.
Medina Rivas, M.A., et al., 2016. El Chocó, Colombia: un hotspot de la biodiversidad humana. Rev. Biodivers. Neotrop. 6, 45.
Nagar, S.D., et al., 2019. Population pharmacogenomics for precision public health in Colombia. Front. Genet. 10, 241.
National Center for Biotechnology Information, 2024. https://www.ncbi.nlm.nih.gov/.
NBC News, 2024. A 7th Person With HIV is Probably Cured After Stem Cell Transplant for Leukemia, Scientists Say. https://www.nbcnews.com/health/health-news/7th-person-hiv-cured-stem-cell-transplant-leukemia-scientists-say-rcna161897.
NCI, 2011. Definition of enveloped virus - NCI Dictionary of Cancer Terms. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/enveloped-virus.
Nievergelt, C.M., et al., 2019. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. Nat. Commun. 10, 4558.
NIH, 2024. CD4 T Lymphocyte. https://clinicalinfo.hiv.gov/en/glossary/cd4-t-lymphocyte.
Novembre, J., Galvani, A.P., Slatkin, M., 2005. The geographic spread of the CCR5 Δ32 HIV-resistance allele. PLoS Biol. 3, e339.
Ongaro, L., et al., 2021. Continental-scale genomic analysis suggests shared post-admixture adaptation in the Americas. Hum. Mol. Genet. 30, 2123–2134.
Payra, S., et al., 2023. HIV cure: are we going to make history? HIV Med. hiv.13557. https://doi.org/10.1111/hiv.13557.
Reich, D., et al., 2012. Reconstructing native American population history. Nature 488, 370–374.
Rojas, W., et al., 2010. Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. Am. J. Phys. Anthropol. 143, 13–20.
Sarker, Md.S., 2023. Prevalence of R5 and X4-tropic strains in HIV-1 infected bangladeshi population. Indian J. Med. Microbiol. 45, 100377.
Solloch, U.V., et al., 2017. Frequencies of gene variant CCR5-Δ32 in 87 countries based on next-generation sequencing of 1.3 million individuals sampled from 3 national DKMS donor centers. Hum. Immunol. 78, 710–717.
The 1000 Genomes Project Consortium, et al., 2015. A global reference for human genetic variation. Nature 526, 68–74.
Waldman, S., et al., 2022. Genome-wide data from medieval German Jews show that the Ashkenazi founder event pre-dated the 14th century. Cell 185, 4703–4716.e16.
Wang, S., et al., 2008. Geographic patterns of genome admixture in Latin American mestizos. PLoS Genet. 4, e1000037.
Wang, C., et al., 2023. Multitargeted drug design strategy for discovery of short-peptide-based HIV-1 entry inhibitors with high potency. Eur. J. Med. Chem. 252, 115294.
What are stem cells. In: Stem Cells, 2014. Elsevier, pp. 53–68. https://doi.org/10.1016/B978-0-12-411551-4.00003-9.
Whitlock, M., Schluter, D., 2009. The Analysis of Biological Data. Roberts and Co. Publishers, Greenwood Village, Colo.
Wojcik, G.L., et al., 2019. Genetic analyses of diverse populations improves discovery for complex traits. Nature 570, 514–518.

Glossary

T CD4 lymphocyte: Immune cell that coordinates the immune response by stimulating other immune cells (NIH, 2024).
Viral envelope: Is a wrapping that covers the viral genome. This envelope comes from the infected cell's plasma membrane and is acquired during the "budding off" process (NCI, 2011).
Virus receptor: A host cell surface component that allows virus attachment and cell entry (Evans, 2008).
Co-receptor: An accessory cell surface component used in cell entry (Evans, 2008).
Stem Cell: A cell that can replicate into an exact copy of itself or a specialized, differentiated, cell (*What are stem cells*, 2014).
Mann-Whitney U test: Statistical test that verifies if the frequency distribution of two groups are the same. If they are, the test compares the means of the two groups. It does not require as many assumptions as the *t*-test (Whitlock and Schluter, 2009).
Logistic regression: A special type of nonlinear regression developed for a binary response variable (Whitlock and Schluter, 2009).