ISSN: 2634-9221
European Journal of Applied Sciences; Vol. 14 No. 01 (2026) (131-140)
https://doi.org/10.14738/aivp.1401.19809

# Evolutionary Signature and Genetic Structure of Concatenated *TP53-CYP17A1* Genes in Colorectal Cancer in Senegal

Anna Ndong[1,2], Bineta Keneme[1,2], Mbacké Sembene[1,2]

1. Department of Animal Biology, Faculty of Science and Technology (FST), Cheikh Anta Diop University of Dakar (UCAD), Dakar, Senegal
2. Laboratoire of Genomic, Cheikh Anta Diop University, Animal Biology, Dakar-Fann, Senegal

**Abstract:** This study aims to characterize the mutational profiles of the *TP53* and *CYP17A1* genes in Senegalese patients with colorectal cancer (CRC) in order to better understand the local molecular mechanisms of tumorigenesis. The analysis included 24 patients with CRC and 24 healthy controls. The sequences of exon 4 of *TP53* and the promoter region at exon 1 of *CYP17A1* were concatenated, then subjected to genetic structuring, historical demography, and phylogenetic relationship analyses using DnaSP (v5.10), MEGA (v7.014), and Arlequin (v3.1). The results reveal significant genetic differentiation between cancerous and healthy tissues ($F_{ST}$ = 0.113, p = 0.009), as well as greater genetic diversity within tumors. Neutrality tests (Tajima's D and Fu's FS) indicate recent demographic expansion in the tumor population, with an excess of rare variants. The multimodal distribution of mismatches and the haplotype network confirm this evolutionary dynamic, marked by the emergence of haplotypes specific to cancerous tissues. These data suggest that the accumulation of mutations in *TP53* and *CYP17A1* contributes to genetic heterogeneity and CRC progression in the Senegalese population, opening up prospects for targeted therapeutic approaches.

**Keywords:** colorectal cancer, *TP53*, *CYP17A1*, concatenation, genetic diversity, Senegal, tumor evolution.

## INTRODUCTION

DNA is vulnerable to genotoxic stress (endogenous, environmental, or therapeutic) that induces mutations [1]. Some mutations confer a selective advantage, triggering multi-step carcinogenesis through the gradual accumulation of genetic alterations targeting oncogenes and tumor suppressors, followed by clonal expansion and disruption of the proliferation-apoptosis balance [2, 3]. The tumor suppressor protein p53 plays an important role in maintaining genome integrity, in particular by activating and repressing the expression of certain genes. Alterations in the p53 protein or its regulators play a fundamental role in the resistance of cancer cells to apoptosis. Indeed, p53 is a transcription factor that plays a central role in initiating the apoptotic process, particularly in response to genotoxic stress. Deleterious mutations in the p53 protein are found in more than 50% of human tumors [4, 5] and also in approximately 40 to 50% of colorectal tumors [6]. At the same time, hormones are a major etiological factor in many cancers [7]. The *CYP17A1* enzyme (cytochrome P450) catalyzes the synthesis of steroid hormones, including glucocorticoid precursors such as cortisol, which regulates the immune response, and androgens, such as testosterone, which stimulates the development and maintenance of male characteristics or is converted into estrogen in women, as well as the metabolism of cholesterol, bile acids, and arachidonic acid derivatives [8]. These androgens promote cancer development [9], giving *CYP17A1* major therapeutic interest. This target is validated by the efficacy of abiraterone (a specific inhibitor) in castration-resistant prostate cancer [10], with trials currently underway in breast cancer.

However, the lack of data on *TP53* and *CYP17A1* mutation profiles in colorectal cancer in the Senegalese population limits our understanding of local tumor mechanisms and the adaptation of targeted therapeutic strategies. The simultaneous study of these genes with distinct expressions would help elucidate the role of genetic factors in the tumorigenesis and progression of CRC in Senegal, and the potential of their combined mutation profiles may provide insights into the progression of colorectal cancer.

This part of the study, which aimed to assess the progression and diversity of colorectal cancer in the Senegalese population, was conducted in the context of a better understanding of the molecular mechanisms that govern this disease.

## METHODOLOGY

### Study Population

The study involved 24 patients with colorectal cancer (CRC) and 24 control subjects. These patients were recruited from the general surgery and oncology departments of Aristide le Dantec, Principale de Dakar, and Grand-Yoff hospitals. For each patient who underwent surgery, a sample of the tumor was taken from a fresh surgical specimen, collected in a dry tube, and stored at -20°C, along with their clinical information sheet. After collection, the tumor tissues were sent directly to the Genomics Laboratory of the Department of Animal Biology at the Faculty of Science and Technology of the University of Dakar, where they were fixed in 96% alcohol for various molecular analyses.

### Extraction, Amplification, and Sequencing of Exon 4 of the *TP53* Gene and the Promoter Region at Exon of the *CYP17A1* Gene

DNA extraction was performed on 24 tumor tissue samples and 24 control blood samples using the standard protocol of the Zymo Research kit. Exon 4 of the *TP53* gene was amplified and sequenced according to the protocol described by Keneme *et al*. [11], and the region of the *CYP17A1* gene studied was amplified and sequenced according to the protocol described by Ndong *et al*. [12].

### Genetic Analysis

Gene concatenation, also known as the supermatrix method, is a molecular biology technique that involves assembling gene fragments from the same genome of an individual to form a single continuous DNA molecule. This approach improves the reliability of structural and phylogenetic analyses by integrating information from multiple genes. It overcomes the biases inherent in analyzing a single gene, particularly the different rates of gene evolution, and provides more accurate information on the history of species and their relationships to each other [13]. The BioEdit bioinformatics tool was used to concatenate the two genes *TP53* and *CYP17A1*. Genetic structuring analyses and evolutionary parameters were determined from the concatenated sequences.

### *Genetic Structuring Analyses*

#### *Differentiation and Genetic Distance:*

The genetic differentiation index ($F_{ST}$) provides information on the effect of subdivision between populations. For DNA sequences, the $F_{ST}$ estimate is based on genetic distances between haplotypes, i.e., haplotype frequencies or frequencies of polymorphic sites, treating each site as a distinct locus [14]. According to Wright [15], $F_{ST}$ values range from 0 to 1, and the closer the $F_{ST}$ is to 1, the more isolated the populations are from each other. However, no difference between haplotype frequencies in subpopulations is observed if the $F_{ST}$ is zero. $F_{ST}$ values between populations were evaluated using the ARLEQUIN V3.1 program [16].

Genetic distance is a measure of the genetic links between population samples. Measuring the differences that remain between two populations shows how they are genetically different. Genetic distances are therefore used either to estimate the time of divergence or to reconstruct phylogenies, which can in turn be used to decide which populations should be conserved [17]. When genetic distance is large, genetic similarity is lower and the time of divergence is greater. Conversely, when genetic distance is small, similarity is higher and the time of divergence is shorter [18]. DS varies from 0 (identity of the samples compared) to infinity. Genetic distance was determined within each population (intra-population genetic distances) and between pairs of populations (inter-population genetic distances) using MEGA 7 software [19]. A value of $P < 0.05$ was considered significant for both parameters.

#### *Molecular Variance Analyses:*

Molecular variance analysis (AMOVA) is performed to understand the structure of the study population through hierarchical analysis. This analysis provides an estimate of the total genetic variance attributable to components such as allele content between haplotypes, due to differences between individuals, between individuals within a population, and between populations. The interpretation of the genetic structure of populations using F-statistics is tested using a non-parametric permutation approach [20]. Significance tests are performed after 1023 permutations using the ARLEQUIN V3.1 program [16].

### *Demo-genetic Analyses*

These tests are more accurately described as tests of selective neutrality and population equilibrium. Tests based on the allele frequency spectrum determine whether the mutation frequency spectrum is consistent with the expectations of the standard model of neutrality. Tajima's D [21] is the difference between the total number of polymorphic sites observed (S) and the average number of differences observed between pairs of sequences (K); Fu's FS [22] compares the average number of differences observed between pairs of sequences (K) with the number of haplotypes (H) in a population.

### *Analysis of Mismatch Distribution and Demographic Indices*

Mismatch distribution analysis is the qualitative graphical representation of the distribution of genetic distances between individuals in a population taken in pairs. It was performed under the assumption of a population of constant size. This model measures the distribution of nucleotide differences observed in pairs of sites and that of expected values (in equilibrium and no recombination) in a stable population of constant size [23]. Mismatch analysis combines two indices that test the quality of fit of the distribution. These indices are the SSD (sum of squares of deviations) and the Rag (Harpending's Raggedness index), an irregularity index that quantifies the fineness of the distribution of observed pairwise differences. These indices take larger values for a multimodal distribution (stationary population) than for a unimodal distribution (expanding population).

### *Analysis of Phylogenetic Relationships*

Haplotype networks are an application of the median link method to show phylogenetic relationships between different haplotypes. A minimum haplotype network is characterized by nodes (circles) and branches (links) that connect the nodes. Each node corresponds to a haplotype whose size is proportional to the frequency of the haplotype in the dataset. They are either sequences from the dataset (haplotypes) or median vectors. The links are the differences in characters. A median vector is a hypothetical (often hereditary) sequence that is necessary to connect existing sequences in the network with maximum parsimony. The haplotype network is constructed with NETWORK ver. 5.0.0.0 using the Median-Joining method [24] to show the phylogenetic relationships between different haplotypes.

## RESULTS

### Structuring Parameters

The cancer population shows a significantly higher intra-population genetic distance (0.0075 ± 0.0024) than that observed in the healthy population (0.0041 ± 0.0025). The genetic distance between populations, which is around 0.0068, can be explained in part by the difference observed within cancerous tissues, highlighting the heterogeneity of mutations observed in cancer patients. Healthy tissues appear to be more homogeneous. These results are presented in Table 1.

A strong genetic differentiation ($F_{ST}$ = 0.113) that is statistically robust, as evidenced by the highly significant p-value (p-value = 0.009), is found at the overall population level. The value of the differentiation factor $F_{ST}$, or fixation index, corroborates the AMOVA (Analysis of Molecular Variance) analysis, which indicates that, across the entire dataset, 88.63% contributes to variation at the population level. This value shows that approximately one quarter of the overall genetic diversity is attributable to population subdivision. Meanwhile, 11.37% comes from divergence at the individual level, highlighting that individuals retain significant genetic diversity. These results are presented in Table 1.

**Table 1: Genetic distance and differentiation factors**

<table>
<tr><th colspan="3"></th><th colspan="2">Intra-population distance</th><th>Distance between populations</th></tr>
<tr><td colspan="3">Healthy</td><td colspan="2">0.0041 (0.0025)</td><td rowspan="2">0.0068 (0.0026)</td></tr>
<tr><td colspan="3">Cancerous</td><td colspan="2">0.0075 (0.0024)</td></tr>
<tr><td colspan="6">Source of molecular variance and $F_{ST}$</td></tr>
<tr><th></th><th>d.f</th><th>Sum of squares</th><th>Variation components</th><th>(%) of change</th><th>$F_{ST}$</th></tr>
<tr><td>Intra-population</td><td>1</td><td>1.667</td><td>0.052</td><td>11.37</td><td rowspan="3">0.113 (0.009)</td></tr>
<tr><td>Between populations</td><td>46</td><td>18.792</td><td>0.408</td><td>88.63</td></tr>
<tr><td>Total</td><td>47</td><td>20.458</td><td>0.460</td><td>100</td></tr>
</table>

d.f = degrees of freedom

### Demo-genetic Parameters

Neutrality test analysis reveals distinct demographic dynamics between cancerous and control tissues. In cancerous tissues, negative Tajima's D (-1.062, p = 0.134) and Fu's $F_S$ (-2.342, p = 0.126) values are obtained. This reflects an excess of rare variants conferring a selective advantage to tumor cells. Conversely, control tissues show a positive Tajima's D value (0.530, p = 0.722), reflecting a deficit of deleterious variants (Table 2).

**Table 2: Selective neutrality and population equilibrium test**

| | Tajima's D | $F_S$'s $F_U$ |
|---|---|---|
| Healthy | 0.530 (0.722) | -2.279 (0.062) |
| Cancerous | -1.062 (0.134) | -2.342 (0.126) |

### Distribution Disparity Parameter

The disparity distribution graph constructed under the assumption of a stable population indicates a unimodal distribution, which highlights that there is no difference between the expected and observed evolution for the control population (Figure 1), whereas for cancer cases, the distribution is multimodal (Figure 2), reflecting a population in demographic expansion.

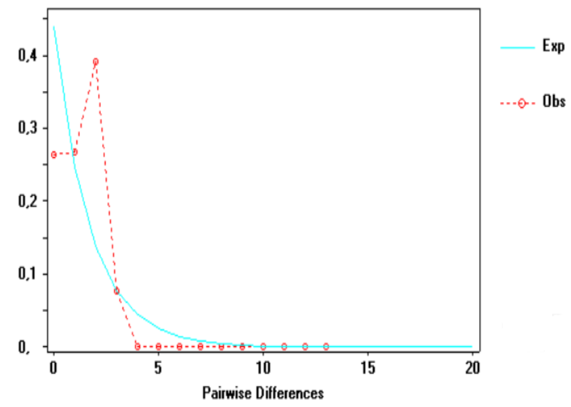

**Figure 1**: Mismatch curve distribution of the control population

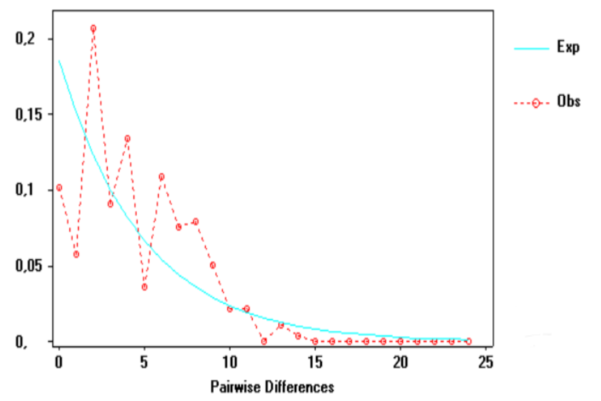

**Figure 2**: Mismatch distribution curve of the cancer population

| | |
|---|---|
| SSD : 0.027 (0.110) ; Ragg : 0.120 (0.230) | SSD : 0.017 (0.46) ; Ragg : 0.057 (0.200) |

SSD (sum of squared deviations) et le Ragg (*Harpending's Raggedness index*)

### Evolutionary Parameter

The haplotype network of concatenated *TP53-CYP17A1* sequences (Figure 3) reveals moderate genetic diversity characterized by 18 haplotypes (Table 3). The topology shows an uneven distribution of haplotype frequencies, represented by nodes sized proportionally to their abundance. The largest node is the majority haplotype (H7), which is predominantly found in healthy tissue (9 individuals versus 5 cancerous individuals) and has distinct phylogenetic connections: it is directly linked to several rare haplotypes from healthy tissue by very small mutational distances, while a median vector connects it to a peripheral cluster of haplotypes specific to cancerous tissue via short branches (mutational distances that are not large enough). Haplotypes (H3) and (H14) are private haplotypes and are found only in cancerous individuals and controls, respectively.

**Table 3: Frequency of haplotypes**

| List of haplotypes (H) | Frequency | Cancerous (%) | controls (%) |
|---|---|---|---|
| H1 | 4.16 (2/48) | 8.33 (2/24) | 0 |
| H2 | 2.08 (1/48) | 4.17 (1/24) | 0 |
| H3 | 12.5 (6/48) | 25 (6/24) | 0 |
| H4 | 2.08 (1/48) | 4.17 (1/24) | 0 |
| H5 | 2.08 (1/48) | 4.17 (1/24) | 0 |
| H6 | 2.08 (1/48) | 4.17 (1/24) | 0 |
| H7 | 29.16 (14/48) | 20.83 (5/24) | 37.5 (9/24) |
| H8 | 2.08 (1/48) | 4.17 (1/24) | 0 |
| H9 | 4.16 (2/48) | 8.33 (2/24) | 0 |
| H10 | 4.16 (2/48) | 8.33 (2/24) | 0 |
| H11 | 2.08 (1/48) | 4.17 (1/24) | 0 |
| H12 | 2.08 (1/48) | 4.17 (1/24) | 0 |
| H13 | 4.16 (2/48) | 0 | 8.33 (2/24) |
| H14 | 18.75 (9/48) | 0 | 37.5 (9/24) |

| | | | |
|---|---|---|---|
| H15 | 2.08 (1/48) | 0 | 4.17 (1/24) |
| H16 | 2.08 (1/48) | 0 | 4.17 (1/24) |
| H17 | 2.08 (1/48) | 0 | 4.17 (1/24) |
| H18 | 2.08 (1/48) | 0 | 4.17 (1/24) |

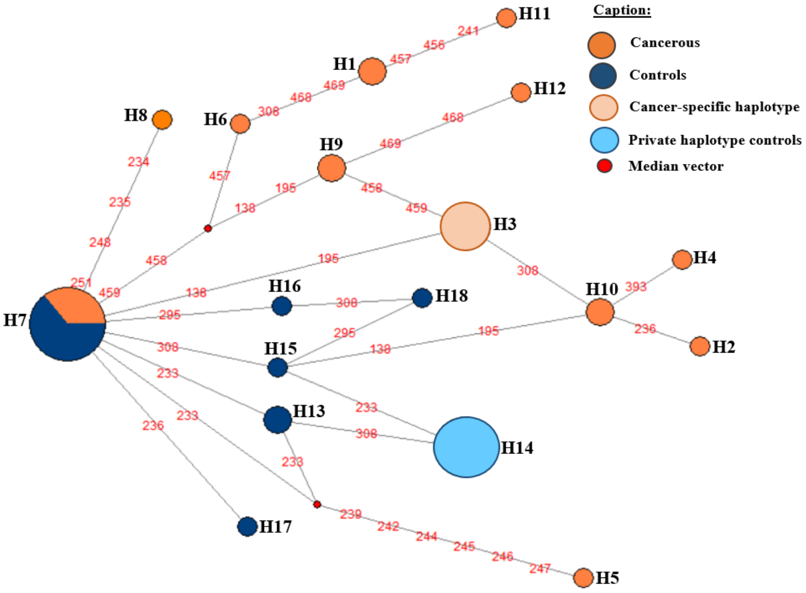

**Figure 3: Haplotype network of concatenated *TP53-CYP17A1* loci**

The size of each node reflects its size, and the number of mutations between nodes is indicated on the branches. The median vector is an evolutionary step between two haplotypes that differ by several mutations.

## DISCUSSION

Colorectal cancer is the third most common cancer, and despite significant and steady progress in medical treatment, its incidence is constantly increasing. This growth is only possible through the accumulation of numerous genetic abnormalities, following different pathways of carcinogenesis. Thus, the accumulation of numerous mutations in a single cell is only possible through the disruption of the cell's genetic stability. Therefore, to understand the extent of genetic alterations in colorectal cancer cells, *TP53-CYP17A1* gene analysis was performed in 24 Senegalese patients.

The estimation of net genetic divergence revealed a genetic distinction between controls and cancerous tissues. Although small, this genetic distance reflects the average number of substitutions or alterations that have occurred since the initiation of the tumor and suggests that the populations share most of their alleles with a relatively short divergence time on an evolutionary scale. The higher intra-tissue genetic distance within cancerous tissues indicates a high diversity of the cancerous population and therefore a high genetic variability of this population, as well as a different evolution at the individual level. The higher genetic distance within cancerous tissue (0.0075) than between the two populations (0.0068) indicates that the cancerous tissue has a part of the body whose DNA has been altered, creating a line of cells that have lost control of their growth and functions. Despite this, this lineage retains profound similarities with the DNA of the original healthy cells.

The measurement of population structuring or subdivision reflects moderate genetic differentiation between control and cancerous tissues. This process can be explained by the monoclonal origin of colorectal tumors [25], whose development is based on a succession of waves of clonal expansion. A tumor is composed of different subpopulations of abnormal cells that share early alterations and whose genesis follows the laws of a spatiotemporal evolutionary continuum [26].

The qualitative graphical representation of the distribution of genetic distances between pairs of individuals within the cancer population revealed a multimodal distribution. This multimodal distribution, with non-significant SSD and Rag indices, which indicates that there is no difference between the observed and simulated values, provides information about a population undergoing demographic expansion. These results are consistent with demographic genetic tests that highlight a molecular signature marked by reduced genetic diversity and the presence of high-frequency derived alleles, and therefore a recent evolutionary event. These tests reveal negative Tajima's D and Fu's Fs values, highlighting recent demographic expansion with an excess of rare alleles [27]. In this case, a concentration of recent mutations and an excess of closely related haplotypes (i.e., haplotypes that differ by a small number of mutations) are expected [28].

These data are consistent with the haplotype network, which reflects heterogeneity, with the mutation profiles of the majority of individuals showing similar haplotypes (associated with recent mutations), while a minority have highly divergent haplotypes (indicating older mutational events). Such distributions are observed when a population increases in size, leading to a concentration of closely related haplotypes. These suggest rapid evolution towards haplotypes that promote tumor growth.

## CONCLUSION

This study provides a characterization of the concatenated mutational profiles of *TP53* and *CYP17A1* in colorectal cancer in the Senegalese population. The results highlight moderate but significant genetic differentiation between healthy and cancerous tissues, as well as greater mutational diversity within tumors, reflecting the clonal heterogeneity characteristic of colorectal tumor progression.

The signatures of demographic expansion detected in cancerous tissues, associated with the presence of specific haplotypes, suggest rapid and adaptive evolution of tumor cells, probably under the effect of local selective pressures. These observations reinforce the hypothesis that the combined alteration of key genes such as *TP53* (apoptosis regulator) and *CYP17A1* (involved in hormone synthesis) could contribute to the aggressiveness and therapeutic resistance of CRC in Senegal.

However, the small sample size and the analysis limited to two genes are limitations. Larger studies, incorporating other candidate genes and clinical data, would be necessary to validate these mutational profiles and evaluate their prognostic or therapeutic value. Ultimately, this work could contribute to personalized medicine that is better adapted to the genomic specificities of African populations.

## REFERENCES

1. Cazaux, C. (2010). Genetic instability, the driving force behind oncogenesis. *Cancer Bulletin*, 97(11), 1241-1251.
2. Kim, H., & Kim, Y. M. (2018). Pan-cancer analysis of somatic mutations and transcriptomes reveals common functional gene clusters shared by multiple cancer types. Scientific reports, 8(1), 6041.
3. Lemaire, J., Larrue, R., Perrais, M., Cauffiez, C., Pottier, N. (2020). " Fundamental aspects of tumor development." *Cancer Bulletin* 107.11: 1148-1160.
4. Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *cell*, 100(1), 57-70.
5. Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *cell*, 144(5), 646-674.
6. Russo, A., Bazan, V., Iacopetta, B., Kerr, D., Soussi, T., & Gebbia, N. (2005). The *TP53* colorectal cancer international collaborative study on the prognostic and predictive significance of p53 mutation: influence of tumor site, type of mutation, and adjuvant treatment. *Journal of clinical oncology*, 23(30), 7518-7528.
7. Greenlee, R. T., Hill-Harmon, M. B., Murray, T., & Thun, M. (2001). Cancer statistics, 2001. CA: *a cancer journal for clinicians*, 51(1), 15–36.
8. Singh, H., Kumar, R., Mazumder, A., Salahuddin, Mazumder, R., & Abdullah, M. M. (2022). Insights into Interactions of Human Cytochrome P450 17A1: *A Review. Current drug metabolism*, 23(3), 172–187.
9. Edwards BK, Noone AM, Mariotto AB, Simard EP, Boscoe FP, Henley SJ, Jemal A., Cho H., Anderson RN, Kohler BA, Eheman CR, Ward EM (2014) Annual report to the nation on the state of cancer, 1975-2010, presenting the prevalence of comorbidity and its impact on survival in people with lung, colorectal, breast, or prostate cancer. *Cancer* 120, 1290-1314
10. Schreyer, M., Sattarov, T., Borth, D., Dengel, A., et Reimer, B. (2017). Detection of anomalies in large-scale accounting data using deep autoencoder networks. *Preprint arXiv: arXiv:1709.05254*.
11. Kénémé, B., Ndong, A., & Mbaye, F. (2025). Prognosis and Functional Analyzes of Missense Mutations in Exon 4 of the *TP53* Gene in Colorectal Cancer in the Senegalese Population. *Journal of Genetics and Genetic Engineering*, *7*(1), 13-29.
12. Ndong, A., Keneme, B., Seye, Y., Mbaye, F., & Sembene, M. (2025). Genetic alterations of *CYP17A1* in the occurrence of colorectal cancer in Senegal. *International Journal of Health Sciences*, 9(2), 699–714.
13. Zou, Y., Zhang, Z., Zeng, Y., Hu, H., Hao, Y., Huang, S., & Li, B. (2024). Common Methods for Phylogenetic Tree Construction and Their Implementation in R. *Bioengineering (Basel, Switzerland)*, 11(5), 480.
14. Hudson, RR, Slatkin, M., et Maddison, WP (1992). Estimation of genetic flow levels from DNA sequence data. *Genetics*, *132* (2), 583-589.
15. Wright, S. (1950). Genetical structure of populations.
16. Excoffier, L., & Heckel, G. (2006). Computer programs for analyzing population genetics data: a survival guide. *Nature Reviews Genetics*, *7* (10), 745-758
17. Takezaki, N., & Nei, M. (1994). Inconsistency of the maximum parsimony method when the rate of nucleotide substitution is constant. *Journal of molecular evolution*, 39, 210-218.

18. Avise J.C. 1994. Molecular markers, Natural history and evolution. *Chapman & Hall. New York, London* 511p.

19. Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7 : Molecular evolutionary genetic analysis version 7.0 for larger datasets. *Molecular biology and evolution*, 33 (7), 1870-1874.

20. Excoffier, L., Smouse, PE, & Quattro, JM (1992). Analysis of molecular variance derived from metric distances between DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131* (2), 479-491

21. Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, 123(3), 597-601.

22. Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915-925.

23. Slatkin, M., & Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, *129*(2), 555-562.

24. Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, *16*(1), 37-48.

25. Braakhuis, BJ, Tabor, MP, Kummer, JA, Leemans, CR, & Brakenhoff, RH. (2003). A genetic explanation for Slaughter's concept of terrain cancerization: evidence and clinical implications. *Cancer research*, *63* (8), 1727-1730.

26. Dubard-Gault, M. (2013). Breast cancer in women under 50 in Réunion between 2005 and 2010.

27. Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39(1), 197-218.

28. Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular biology and evolution*, 19(12), 2092-2100.