International Journal of Health Sciences
Available online at www.sciencescholar.us
Vol. 9 No. 2, August 2025, pages: 699-714
E-ISSN: 2550-696X
https://doi.org/10.53730/ijhs.v9n2.15594

Genetic Alterations of *CYP17A1* in the Occurrence of Colorectal Cancer in Senegal

Anna NDONG [a], Bineta KENEME [b], Yacine SEYE [c], Fatimata MBAYE [d], Mbacké SEMBENE [e]

Manuscript submitted: 27 April 2025, Manuscript revised: 18 May 2025, Accepted for publication: 09 June 2025

Corresponding Author [a]

Keywords

c.-34T>C;
colorectal cancer;
CYP17A1;
mutation;
Senegal;

Abstract

A cancer is said to be hormone-dependent when hormones influence the carcinogenesis process. Studies suggest a possible protective effect of hormone treatment in colorectal cancer. Given the hormone treatment, a thorough study of *CYP17A1* is necessary. This study aims to assess the penetrance of *CYP17A1* in the development of colorectal cancer. This study involved 24 colorectal cancer patients and 24 controls. For each sample, DNA extraction was performed, followed by *CYP17A1* gene amplification and Sanger sequencing. The nature and position of mutations were identified using Mutation Surveyor version 5.1.2. DNASP version 5.10, MEGA version 7.014, and the Arlequin program version 3.1 were used to highlight the parameters of variability, differentiation, and demogenetic evolution of the study population. The results revealed heterozygous mutations in the *CYP17A1* gene and a substitution at the c.-34T>C promoter region. The absence of nonsynonymous mutations was revealed by very low genetic variability in the cancer population compared with controls. There was also little genetic differentiation and distance between the two populations. The frequent c.-34T>C mutation suggests that this polymorphism may modulate the transcriptional activity of *CYP17A1* and, consequently, the hormones under this gene's control, and therefore colorectal tumor growth.

[a] Cheikh Anta Diop University, Dakar, Senegal
[b] Department of Animal Biology, Faculty of Science and Technology (FST), Laboratoire of Genomic, Cheikh Anta Diop University, Animal Biology, Dakar-Fann, Senegal
[c] Cheikh Anta Diop University of Dakar (UCAD), DES, Residency in General Surgery
[d] Department of Animal Biology, Faculty of Science and Technology (FST), Laboratoire of Genomic, Cheikh Anta Diop University, Animal Biology, Dakar-Fann, Senegal
[e] Department of Animal Biology, Faculty of Science and Technology (FST); Laboratoire of Genomic, Cheikh Anta Diop University, Animal Biology, Dakar-Fann, Senegal

Contents

1 Introduction

Colorectal tumors present different clinical, histological, and molecular characteristics, offering a variety of treatment modalities. Among these, laparoscopic surgery is performed in the early stages of the disease and is technically feasible with acceptable morbidity and low mortality (Hahnloser et al., 2002). Tumour resection by open surgery is performed for metastases, and adjuvant radiotherapy for unresectable cases. Other techniques are also involved in the treatment of colorectal cancer (CRC), such as neoadjuvant chemotherapy, aimed at reducing tumor size before surgery, particularly in rectal cancers, and palliative chemotherapy, applied to stage IV and combined with targeted treatments (Heemskerk-Gerritsen et al., 2015). Tyrosine kinase inhibitor (TKI) therapy has also been shown to be effective in the treatment of colorectal cancer (Kircher et al., 2016), while checkpoint inhibitors consisting of monoclonal antibodies (mAbs) have taken their place in the treatment of many forms of cancer (Forese & Ochsenbien, 2008). The complex relationship between the immune system and cancer favors the design of tumor immunotherapy (Kalyan et al., 2018). Hormone therapy is a treatment involving the administration of hormones or, on the contrary, the blocking of hormonal effects, to slow down or interrupt the growth of cancer cells whose growth is influenced by hormonal activity.

Hormones play a major role in the etiology of several cancers, notably breast, ovarian, prostate, and colon cancer (Greenlee et al., 2001). Conjugated equine estrogens offer multiple benefits such as reduced risk of fracture, heart attack, colorectal cancer, and possibly breast cancer (Demers, 2011). Research suggests a potentially protective relationship between conjugated equine estrogens and colorectal cancer risk. Indeed, studies by the Women's Health Initiative (WHI) have shown that women using conjugated equine estrogens as monotherapy (without progesterone) had a reduced risk of developing colorectal cancer compared with non-hormonal treatments (Chlebowski et al., 2004). Thus, estrogens can influence cell proliferation, apoptosis, and differentiation in the colon.

CYP17A1 (cytochrome P450 17) is expressed in the adrenal cortex and gonads and plays a central role in the metabolic pathways leading to the synthesis of androgens, precursors of estrogens (Singh et al., 2022). The *CYP17A1* gene is a regulator of steroidogenesis, on the one hand via 17α-hydroxylase activity, which catalyzes the production of glucocorticoid precursors, and on the other via 17, 20-lyase activity, which participates in the synthesis of androgens (testosterone and dehydroepiandrosterone (DHEA)) (Nakajin & Hall, 1981). This gene is of major interest in cancer due to its key role in steroid synthesis, which strongly influences the development and progression of hormone-dependent cancers (Gilep et al., 2011). Precursor steroid hormones, in particular androgens and estrogens, play a critical role in the growth of hormone-dependent tumors (Edwards et al., 2014). This enzyme has therefore attracted considerable interest as a relatively new drug target, validated by the successful use of the *CYP17A1* inhibitor abiraterone in men with castration-resistant prostate cancer (Murez et al., 2013), and its current evaluation in breast cancer patients.

Overexpression or mutations in *CYP17A1* may contribute to increased production of tumor androgens, promoting resistance to treatment. Studying the *CYP17A1* gene provides a better understanding of resistance mechanisms and enables the development of more effective targeted therapies. However, the regulation of *CYP17A1* expression involves elements located in both promoter and intronic regions, and genetic variations in these regions can affect gene expression and, consequently, steroid hormone production. *CYP17A1* exon 1 is

the first coding region of the gene, corresponding to the beginning of the mRNA sequence transcribed into protein. It encodes essential segments of the enzyme, notably in binding to the endoplasmic reticulum membrane, where steroidogenesis takes place (Bhan et al., 2014). In this study, the mutational penetrance as well as the polymorphism of *CYP17A1* in the occurrence of colorectal cancer were highlighted.

2 Materials and Methods

Study population

The study involved 24 CRC patients, and 24 control cases were used for comparison. These patients were recruited in the general surgery and oncology departments of the Aristide le Dantec hospital, the Principale hospital of Dakar, and Grand-Yoff. For each patient undergoing surgery, a fresh surgical specimen was taken from the middle of the tumour, collected in a dry tube, and stored at 20°C, together with the patient's clinical information sheet. Blood samples were taken from healthy individuals with consent, serving as controls. Samples were sent directly to the Genomics Laboratory of the Animal Biology Department of the Faculty of Science and Technology at the University of Dakar, where the tissues were preserved in 96% alcohol for the various molecular analyses.

Extraction, Amplification, and sequencing of the promoter region_exon 1 of CYP17A1

DNA extraction was performed using the standard protocol of the Zymo Research kit. The region spanning promoter region to exon 1 of the *CYP17A1* gene was amplified in a 25 µl reaction volume containing: 2 µl DNA, 18.4 µl nuclease-free water, 2.5 µl 10X standard buffer, 0.5 µl dNTP, 0.5 µl forward primer (5'-TCCTGAGCCCAGATACCAT-3'), 0.5 µl reverse primer (5'-CCGCCCAGAGAAGTCCT-3'), 0.5 µl MgCl2, 0.1 µl Taq. PCR was performed in a thermal cycler under conditions of initial denaturation at 94°C for 12 minutes, followed by a 35-cycle repeat with denaturation at 94°C for 30 seconds, primer hybridization at 60°C for 30 seconds, elongation of complementary DNA strands at 72°C for 1 minute, and final elongation at 72°C for 7 minutes. The presence of amplicons was verified by electrophoretic migration on a 2% agarose gel. Sanger sequencing was performed to determine the nucleotide sequence of the gene of interest.

Mutation detection

The presence of point mutations and their genomic position was investigated by submitting raw sequencing data to Mutation Surveyor software version 5.0.1 (Minton et al., 2011). Mutation Surveyor software uses anti-correlation technology to compare sample sequences with a reference sequence based on chromatograms. The analysis uses a proprietary algorithm that detects mutations and calculates a mutation confidence score based on peak intensity, drop factor (vertical), overlap factor (horizontal base spacing), and the signal-to-noise ratio of each peak. Variants with a Phred score ≥ 20 were listed and considered significant, a value for which the accuracy is 99%. The *CYP17A1* heterozygosity rate was determined for the whole population, according to the two regions studied, as well as the nature of the mutation.

Genetic and statistical analysis
Sequence alignment and cleaning.

The data set obtained after sequencing was aligned using BioEdit version 8.0.5 (Hall, 1999), which performs a multiple alignment using the ClustalW algorithm (Thompson et al., 1994). This sequence alignment is an essential step before any genetic analysis and is important in determining whether or not sites are similar (Swofford et al., 1996). Nucleotide sequences were carefully checked and corrected concerning the electrophoregram. This sequence alignment and cleaning was also the subject of a BLAST (Basic Local Alignment Search Tool) to check the similarity of our dataset with the reference sequence in GenBank[f].

[f] (https://blast.ncbi.nlm.nih.gov)

Genetic polymorphism in the *CYP17A1* region was highlighted using indices of genetic variability and diversity and amino acid variability in both cancer and control study populations.

The standard indices of genetic variability, including the number of polymorphic sites and the nucleotide composition of the sequences, were extracted using MEGA 7 software (Kumar et al., 2016). Nucleotide frequency, nature and rate of mutation, and types of substitution were calculated with the same software, using the Pattern substitution test. The total number of synonymous and non-synonymous sites (Eta), the average number of nucleotide differences between sequences (K), and the number of haplotypes (H) were calculated with DNAsp software version 5.10.01 (Rozas et al., 2017).

Genetic diversity translates into local variations in the DNA sequence that can form different variants of the same sequence, called alleles. Diversity indices such as haplotypic diversity (Hd) and nucleotide diversity (π) (Nei, 1987), as well as standard deviations, have been evaluated using the DNAsp program. Haplotypic diversity corresponds to the probability that two alleles or haplotypes drawn at random from a sample are different (Nei, 1987). It is therefore a measure of the frequencies and number of haplotypes among individuals with a value between 0 and 1 (Grant et al., 1998). Nucleotide diversity π measures the average nucleotide divergence between all pairs of sequences in a sample (Tajima, 1983); it defines the probability that two haplotypes drawn at random from a sample are different at a given site. A threshold of 5% for nucleotide diversity (π) and 50% for haplotype diversity was considered significant.

For each population, the amino acid frequencies of *CYP17A1* exon 1 were determined using MEGA7 software (Kumar et al., 2016), choosing the universal genetic code and the best reading frame (the one offering the fewest stop codons). A reading frame represents one of the three possibilities for chaining triplets along a portion of mRNA. For a given sequence, there are 3 different reading frames. An open reading frame (ORF) is a DNA sequence beginning with an initiation codon and ending with a stop codon. Between these two codons, the open reading frame contains many codons that potentially encode a protein. A reading frame that cannot be read into protein because stop codons occur frequently is said to be blocked or closed.

Genetic structuring analysis

The genetic differentiation index (F_{ST}) provides information on the effect of subdivision between populations. For DNA sequences, the estimation of F_{ST} is based on genetic distances between haplotypes, i.e., from haplotype frequencies or polymorphic site frequencies, treating each site as a distinct locus (Hudson, 2000). According to Wright (1950), 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. On the other hand, there is no difference between the haplotype frequencies of sub-populations if the F_{ST} is null. F_{ST} values between populations were evaluated using the ARLEQUIN V3.1 program (Excoffier et al., 2006).

Genetic distance is a measure of the genetic links between population samples. Measuring the remaining differences between two populations amounts to showing how they are genetically different. Thus, genetic distances are used either to estimate divergence times or to reconstruct phylogenies, which in turn can be used to decide which populations should be conserved (Takezaki & Nei, 1994). Several measures of distance exist; however, Nei's (1972) standard genetic distance (SD) or Nei's index of dissimilarity, widely used in natural population genetics studies, has been used to assess the divergence that remains between the cancer population and the control population. Indeed, these distances are based on the probability of gene identity by measuring the average number of substitutions occurring after the divergence of two populations, and their value is expected to increase linearly with time (Takezaki & Nei, 1994). Moreover, when genetic distance is large, genetic similarity is lower, and divergence time is greater. On the other hand, when genetic distance is smaller, similarity is higher, and divergence time is shorter. DS varies from 0 (identity of samples compared) to infinity. Genetic distance was determined within each population (intra-population genetic distances) and between populations taken in pairs (inter-population genetic distances) using MEGA 7 software (Kumar *et al.*, 2016). A value of $P < 0.05$ was considered significant for both parameters.

Demo-genetic analysis

These tests are more accurately known as selective neutrality tests. Tests based on the allelic frequency spectrum determine whether the mutation frequency spectrum conforms to the expectations of the standard

neutrality model. Tajima's D (Tajima, 1989) 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 (Fu, 1997) compares the average number of differences observed between pairs of sequences (K) with the number of haplotypes (H) in a population and the R2 (Ramos-Onsin & Rozas, 2002) is a complementary statistic based on the differences between the number of singleton mutations (SS) and the average number of observed differences between pairs of sequences (K).

Mismatch distribution analysis 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. Two models have been defined to infer the demographic evolution of a population from the Mismatch distribution graph:
Population of constant size: this model measures the distribution of observed nucleotide differences per site and the distribution of expected values (in equilibrium and no recombination) in a stable population, i.e. a population of constant size (Slatkin & Hudson, 1991).
Growing-declining population: This model measures the distribution of observed site-pair nucleotide differences and expected values (no recombination) in a growing-declining population (Rogers & Harpending, 1992).
Mismatch analysis combines two indices that test the goodness of fit of the distribution. These are the SSD (sum of squares of deviations) and the Rag (Harpending's Raggedness index), which quantifies the smoothness of the distribution of observed pairwise differences. These indices take on wider values for a multimodal distribution (stationary population) than for a unimodal distribution (expanding population).

Analysis of phylogenetic relationships

Phylogenetic analysis highlights the relationships existing within the individuals of a population, or between two populations, on the assumption that the phylogeny of genes reflects that of organisms. To trace the relationships between individuals in the study population, the haplotype network was constructed with NETWORK ver. 5.0.0.0 using the Median-Joining method (Bandelt et al., 1999).

3 Results and Discussions

3.1 Results
Nature and position of mutations detected

Intronic and coding region mutations were observed through comparison of the chromatograms to that of the reference (NT_030059.13) using the Mutation Surveyor software (Table 1). In the IVS1 region, 19 variants were identified, 15 of which were heterozygous mutations and 4 nucleotide substitutions (Figure 1). Exon 1 of *CYP17A1* revealed 15 heterozygous mutations, that is, a mutation on one of the alleles of the *CYP17A1* gene, and 10 synonymous mutations. Heterozygous mutations (IVS1-34T>TC) and (c.138C>CT) were more frequent (Figure 2). Mutations (IVS1-34T>TC) and (c.138C>CT) (IVS1-281G>GT) and (c.195G>GT) have been already identified in dbSNP.

Table 1
Characteristics of mutations in the studied regions of *CYP17A1*

Pos. Chromosomal	DNA position	Protein position	Nature of mutation	Scores	dbSNPs	Frequencies
10:104597399G	IVS1-281G>GT		het	30		27.3%
10:104597308C	IVS1-190C>CT		het	38	dbSNP:17115144	8.3%
10:104597222G	IVS1-104G>GT		het	21		4.2%
10:104597152T	IVS1-34T>TC		het	144	dbSNP:743572	37.5%
	IVS1-34T>C		Sub	148		16.7%
10:104596981C	c.138C>CT	p.H46HH	het	146	dbSNP:6162	33.33%
	c.138C>T	p.H46H	hom	148		20.83%
10:104596924G	c.195G>GT	p.S65SS	het	95	dbSNP:6163	29.2%
	c.195G>T	p.S65S	hom	133		20.83%

het= heterozygous mutation, hom = homozygous mutation, sub = substitution

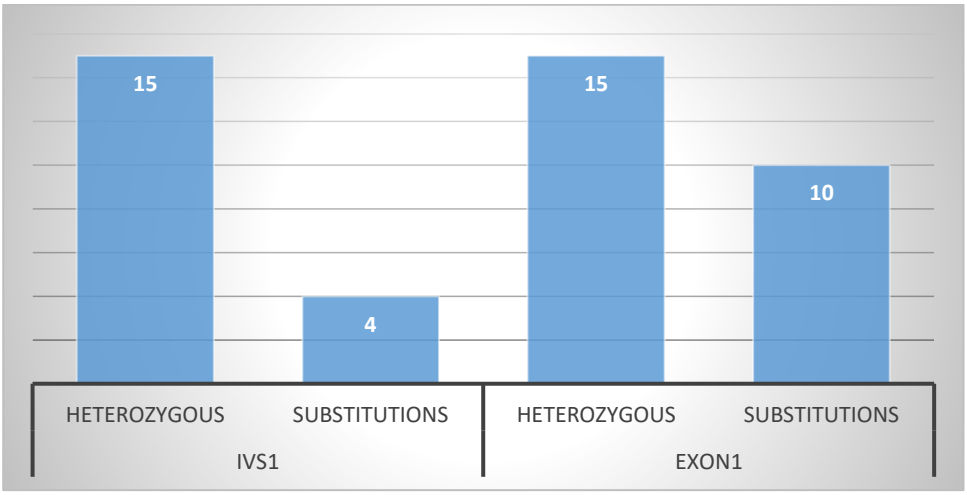


Figure 1. Frequency of mutations according to genomic location and mutation type

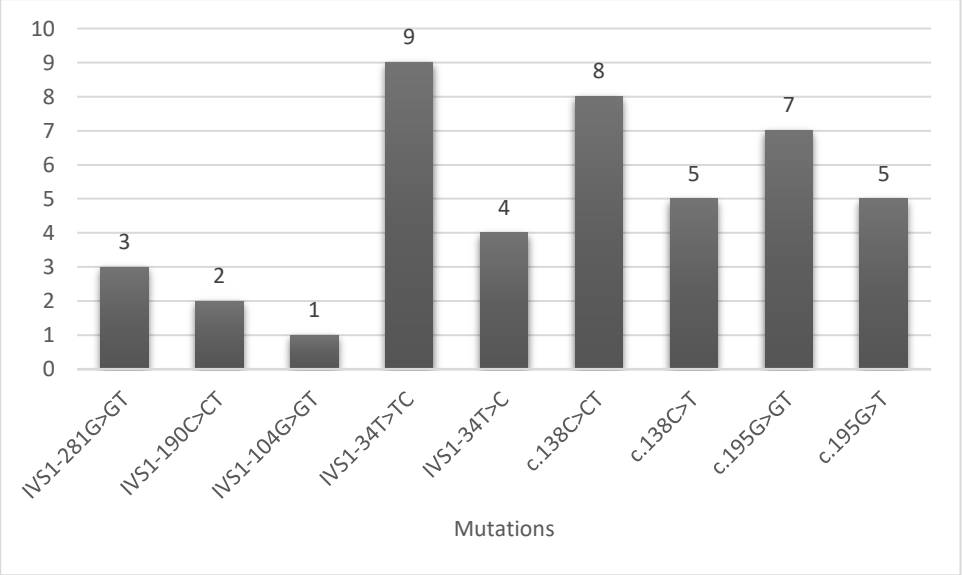


Figure 2. Frequency of each mutation in the total cancer population

CYP17A1 polymorphism
Genetic variability of CYP17A1

Genetic analysis of the *CYP17A1* exon 1 promoter region indicates low genetic variability, with a polymorphism percentage of 0.7% (4/526) of the cancer population. This low polymorphism results in a very low average number of nucleotide differences (K), dividing the 24 cancer patients into 3 haplotypes (H). High haplotype diversity (Hd greater than 50%) and low nucleotide diversity (π less than 5%) were observed in the cancer population, with a predominance of cytosine and guanine nucleotides in both populations. No nucleotide variability was found in the controls. These parameters are shown in Table 2.

Table 2
Genetic polymorphism of the 5'UTR region and exon 1 of *CYP17A1*

	Number of sites	Polymorphic Sites	K	H	Nucleotide frequency		π	Hd
					A+G	C+T		
Controls	526	0	0	1	44.3	55.7	0	0
Cancer patients	526	4	1.717	4	44.4	55.6	0.003	0.627

The total number of Eta mutations is very low and equals 4 in the cancer population, with no mutations found in the control group. The mutation rate (R= 2.01) generated by variations in cancer sequences shows that transitions (S) were more frequent than transversion substitutions (V). Substitutions with amino acid changes proved to be more important than synonymous substitutions in the cancer population. These results are presented in Table 3.

Table 3
Mutational parameters of the 5'UTR region and exon 1 of *CYP17A1*

	Eta	Mutation rate R	Transition transversion bias		Synonymous/Non-synonymous substitutions	
			S	V	dS	dN
Controls	0	0.50	33.32	66.72	0.000	0.000
Cancer patients	4	2.01	66.76	33.28	0.001	0.004

Amino acid variability in CYP17A1 exon 1

Amino acid variability in *CYP17A1* exon 1 shows a difference in the expression of seven protein monomers - cysteine, glycine, leucine, proline, glutamine, arginine, and tyrosine - between cancer cases and controls. All six amino acids showed a significant difference in expression between the two populations compared with cysteine, or a small difference was obtained, marked by a p-value (0.3379) above the 0.05 threshold. These results are presented in Table 4.

Table 4
Amino acid frequency of *CYP17A1* exon 1

Amino acids	Controls	Cancer patients	*p-value*
Ala	10.119	10.119	-
Cys	3.571	3.596	0.3379
Asp	1.190	1.190	-
Glu	2.380	2.380	-
Phe	2.380	2.380	-
Gly	7.738	7.415	3.116e-05*
His	4.761	4.761	-
Ile	3.571	3.571	-
Lys	4.166	4.166	-
Leu	10.714	11.904	7.715e-12*
Met	0	0	-
Asn	0.595	0.595	-
Pro	14.285	13.095	7.715e-12*
Gln	6.547	5.952	7.715e-12*
Arg	5.952	6.522	5.245e-11*
Ser	9.523	9.523	-
Thr	1.190	1.190	-
Val	5.952	5.952	-
Trp	2.976	3.29	3.116e-05*
Tyr	2.380	2.380	-

*= Significant

Genetic structure

Analysis of the genetic distance between controls and cancer sequences revealed a relatively low genetic distance of (0.012) between the population and within the cancer population. No genetic distance was observed within the controls. No genetic differentiation between the two populations was obtained. These results are presented in Table 5.

Table 5
Genetic distance and differentiation factor

	Distance within population	Distance between Tissulaire	F_{ST}
Controls	0	0.016	0
Cancer patients	0.004		

Demo-genetic parameters

The detection of deviations from the mutation-selection balance shows values of Tajima's D zero in the controls and close to zero and positive for the cancer population, as well as *Fu FS* is positive and non-significant. The value of R2 is also very low and positive for cancer tissues. Zero values were obtained for these parameters in all controls (Table 6).

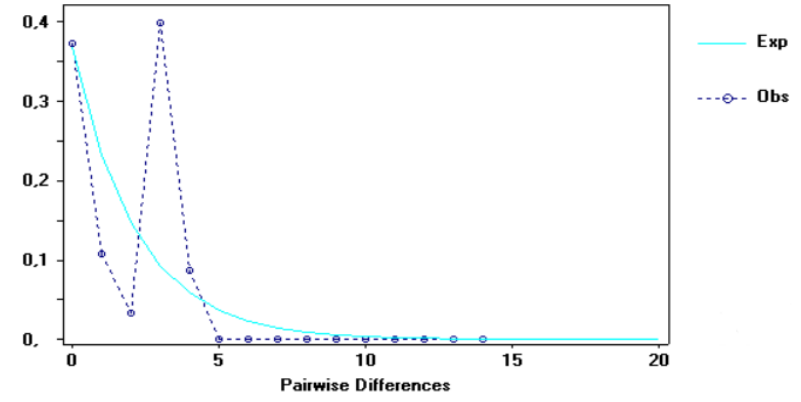
Table 6
Parameters of the test of selective neutrality and population balance

	R2	Tajima's *D*	*Fu's Fs*
Controls	0	0.000 (1)	-
Cancer patients	0.238	1.64998 (0.949)	1.608 (0.815)

Distribution disparities parameter

The qualitative graphical representation of the distribution of *CYP17A1* genetic distances between individuals of a population taken two by two shows a bimodal distribution of the cancer population (Figure 3). This graphical representation was not made for the controls due to the absence of distance or genetic difference between the individuals.

The measurement of the distribution of nucleotide differences by observed pairs of sites and that of the expected values are associated with two indices that test the quality of adjustment of the distribution. These indices are the SSD (sum of squares of deviations) and the Ragg (Harpending's Raggedness index) irregularity index that quantifies the fineness of the distribution of the observed pairwise differences. The graphs obtained are collated with the SSD non-significant and Ragg significant indices for the cancer population.



SSD : 0.110 (0.180)
Ragg : 0.314 (0.120)

Figure 3. Mismatch distribution CCR *CYP17A1*

Signs of irregularities

Evolution parameters

The haplotype network shows the phylogenetic relationships between the different haplotypes. The four nodes that form the network correspond to the different haplotypes of the overall population (Figure 11). The size of each node is proportional to the number of individuals it groups together. The largest node constitutes the majority haplotype and groups together only the controls. This majority haplotype is differentiated from the cancer population by mutational steps at sites 8 then 11 to 15. These mutational steps correspond to the different characteristics established over time that make it possible to differentiate these individuals from those of the majority haplotype. In total, four mutational steps detected at sites 9; 262; 433 and 490 distinguish the cancer population into four haplotypes.

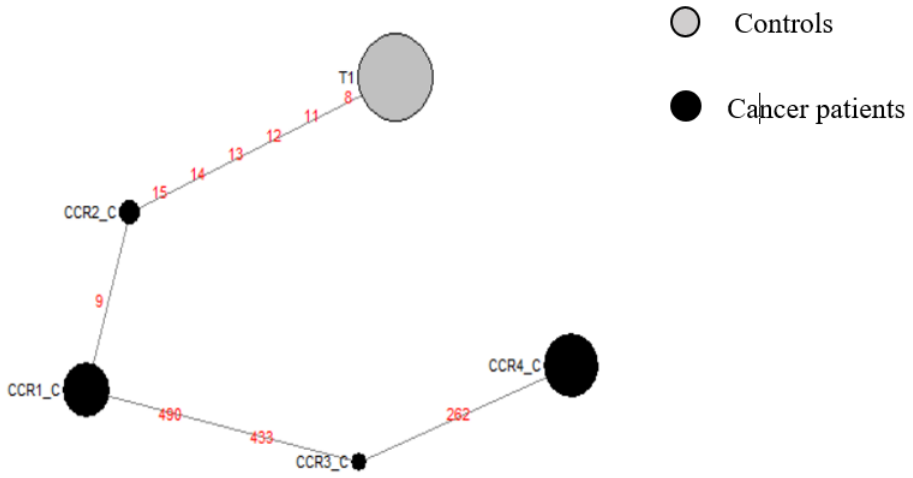


Figure 11. *CYP17A1* haplotype network

3.2 Discussion

The study of the *CYP17A1* gene, also called cytochrome P450, family 17, subfamily, Polypeptide 1, in colorectal cancers is part of the research on hormonal and genetic mechanisms involved in tumorigenesis. This gene encodes a key enzyme involved in the production of steroid hormones. Steroids can influence cell proliferation and angiogenesis, critical processes in tumor development. The objective of this study was to see how the *CYP17A1* gene can influence the risk, progression, or prognosis of colorectal cancers by exploring genetic variations from exon 1 to intron 1 of this gene.

Intron 1 of the *CYP17A1* gene plays a crucial role in regulating gene expression (Hwang et al., 2011). DNA methylation in this region is associated with decreased expression (Missaghian et al., 2009). The polymorphism located in the promoter region of the *CYP17A1* gene 34 bases upstream of the transcription start site, known as rs743572, has been found in cancer tissue sequences and has been reported to regulate increased transcription of this gene as well as androgen production (Tüzüner et al., 2010). According to Keneme & Sembene (2021), this c.-34T>C mutation creates a novel CCACC box site. This novel site is similar to an additional promoter with a potential Sp1-like binding site (Carey et al., 1994), which was predicted to increase *CYP17A1* gene expression. This gene overexpression is likely to increase serum hormone levels, including androstenedione and estradiol (Rai et al., 2014). Moreover, studies by Li et al. (2002) and Sata et al. (2003) hypothesized that the mutant C allele (c.- 34T>C) of *CYP17A1* could be a marker of increased steroidogenesis. Due to its potential role in modulating hormone levels, the mutant C allele of this rs743572 polymorphism (c.- 34T>C) has been associated with multiple cancers. Thus, Gaudet et al. (2008) suggest that the C allele may be associated with an increased risk of multiple hormone-dependent cancers. According to Datkhile et al. (2021), this polymorphism increases the risk of cervical cancer, but also breast cancer in postmenopausal women (Sun et al., 2018) and is one of the few factors that can increase the risk of prostate cancer (Effah et al., 2020)9. Also, an increased risk of gallbladder cancer in tobacco users has been reported for this polymorphism (Rai et al., 2014). An altered mechanism of *CYP17A1* transcriptional regulation control has been correlated with the presence of mutations in the promoter region of this gene in both uterine

fibroids and breast fibroadenomas (Kénémé, 2020). The C allele of the *CYP17A1* c.-34T>C polymorphism was diagnosed in patients at an early age of 18 years and with colorectal tumors compared to T allele carriers, suggesting an association between this polymorphism and the age of onset of colorectal cancer in Lynch syndrome mutation carriers (Campbell et al., 2007). This may explain the association of steroid hormones in colorectal carcinogenesis and suggests that increased levels of androgens or estrogens may be influenced by variations in the *CYP17A1* gene. These genetic variations could modulate cell proliferation and susceptibility to colorectal cancer. It is important to note that these associations are complex and may vary depending on genetic and environmental factors.

Loci with a heterozygosity advantage, previously considered to be a minority of all loci in a species (Hedrick, 2012), are found to be very common in the studied regions of *CYP17A1*. Intronic mutations, often silent, can influence gene transcription and translation. Heterozygosity refers to a genetic variation at a site where only one of the two alleles carries the mutation or genetic change, while the other allele remains "normal" or unmutated. These types of mutations can influence cancer predisposition, particularly when they affect genes involved in cell growth regulation, DNA repair, or tumor suppression (Nikkilä et al., 2013). Steroids are essential hormones in the regulation of many biological processes, including cell growth and differentiation. Mutations in the steroid precursor gene *CYP17A1*, particularly heterozygous mutations, can impact hormone production and thus influence the progression of many types of cancer (Miller & Auchus, 2011), particularly those that are hormone-dependent, such as breast cancer or prostate cancer (Setiawan et al., 2007). Indeed, these mutations can affect enzyme functionality and, consequently, disrupt the hormonal balance, which can influence cancer susceptibility, particularly hormone-dependent cancers (Miller & Auchus, 2011). In this sense, the high frequency of heterozygous mutations in the promoter region of the *CYP17A1* gene could disrupt the synthesis of these hormones and, consequently, influence the growth of colorectal tumor cells. Indeed, the disruption of circulating steroid levels (androgens, estrogens) could affect hormone receptors in colon cells and influence tumor growth. These receptors, although traditionally associated with other tissue types, play a role in colonic tissue. Estrogen receptors (ER) have two main isoforms, ERα and ERβ, with distinct functions. ERβ appears to have a protective role, activating pro-apoptotic pathways, while ERα is associated with anti-apoptotic pathways. A reduction in ERβ expression in advanced adenomas and CRC has been observed, while ERα may be overexpressed in these conditions (Ditonno et al., 2021). Although their role is well understood, further research is needed to clarify the function of androgen receptors (ARs) in colorectal cancer. This heterozygosity of *CYP17A1* could affect the activation and signaling pathways of hormone receptors due to differences in genotypic expression. These effects may largely depend on interactions with other mutations or environmental factors.

Compared to controls, the *CYP17A1* study region was found to be very poorly polymorphic with a very low haplotype number, suggesting the existence of reduced nucleotide diversity within the cancer population. This very reduced polymorphism generates mutations with a fairly high rate within the cancer population. The mutation rate, or the frequency at which genetic changes occur in a given genome over time, is influenced by intrinsic factors, such as the fidelity of DNA replication, and extrinsic factors, such as environmental mutagens (Gao et al., 2016). This rate, quite high compared to controls, relates to a mutational heterogeneity from exon1 to intron1 of *CYP17A1* in favor of the speed of evolution of cancer cells and a predominance of non-synonymous mutations.

Transitions were more frequent than transversions in cancer sequences due to biological mechanisms such as DNA replication errors. Indeed, they often result from spontaneous errors due to the incorrect incorporation of bases similar in size and structure, thus making their recognition more difficult by repair enzymes (Hanna et al., 2005). Some mutations result in amino acid substitutions or even alter protein structure or function, conferring a selective advantage to tumor cells (Mendiratta et al., 2021). The high frequency of leucine, arginine, and tryptophan in cancer populations suggests that they influence cancer cell metabolism, including cell growth, lipid synthesis, gene expression regulation, and immune invasion (Vettore et al., 2020). Leucine is an essential amino acid that serves as a signal for the activation of the mTORC1 pathway, thereby regulating cell growth and protein synthesis. In cancer cells, excessive activation of mTORC1 promotes tumor proliferation. Additionally, leucine can be metabolized to acetyl-CoA, a precursor for lipid synthesis, thereby contributing to the biogenesis of cell membranes required for cell division (Lieu et al., 2020).

Arginine is a precursor to the synthesis of polyamines, molecules involved in the regulation of gene expression and cell proliferation. In cancer cells, arginine-derived polyamines modulate chromatin structure and promote tumor growth (Lieu et al., 2020). Tryptophan is mainly metabolized via the kynurenine pathway, producing metabolites such as kynurenine. These metabolites induce immunosuppression by activating the aryl hydrocarbon receptor (AhR), which impairs the function of dendritic cells and regulatory T cells, allowing cancer cells to escape immune surveillance (Lieu et al., 2020).

The two populations presented a low degree of similarity inferring on the evolution of cancer pathology. The higher intra-tissue genetic distance within cancer tissues provides information on the diversity of the cancer population, referring to the distinction of tumor growth between individuals. The low diversity observed also did not allow for establishing a genetic differentiation, hence the null value of Fst, suggesting the absence of evolutionary force on this diversity between controls and cancer patients (Holsinger & Weir, 2009).

The bimodal or flattened discordance distribution means that the majority of individuals will have few differences between their haplotypes (indicating recent mutations), and a few individuals will have more distant haplotypes (indicating older mutations). Such distributions are observed when a population increases rapidly in size, leading to a concentration of closely related haplotypes between individuals. These data are consistent with Tajima's D values close to zero, suggesting that mutations are neutral or balanced by positive selection, referring to the presence of a polymorphism in connection with a recent population expansion or a low rate of positive selection (Nielsen, 2005). Compared to controls, *CYP17A1* exon 1 and intron 1 result from an accumulation of evolutionary events at the molecular level during colorectal carcinogenesis. This recent expansion of cancer cells was demonstrated by the low R2 value of Ramos and Rozas within the cancer population. In this case, a concentration of recent mutations and an excess of close haplotypes (i.e. haplotypes that differ by a small number of mutations) are expected (Ramos-Onsins & Rozas, 2002). This excess of close haplotypes was revealed by the haplotype network, which advocates that cancer sequences differ at only four sites and therefore are assimilated to a rapid expansion.

4 Conclusion

The human P450 (17 alpha) gene (*CYP17A1*) is suspected to be linked to the risk of multiple cancers due to its role in regulating the biosynthesis of steroidal and non-steroidal hormones. Investigations of genetic variation from the promoter region to exon 1 of *CYP17A1* in colorectal carcinoma revealed variants, most of which were heterozygous mutations. The high frequency of the variant (c.- 34T>C) suggests that the presence of the mutant C allele modifies the sequence of the binding site for transcription factors, which may influence gene expression. The low genetic variability of tumor sequences compared to controls expresses the low penetrance of *CYP17A1* in this pathology. *CYP17A1* polymorphisms could be used as biomarkers to identify individuals or populations at increased risk of hormone-dependent cancer. They could also be used to predict response to therapies targeting hormonal pathways.

Acknowledgments
We are grateful to two anonymous reviewers for their valuable comments on the earlier version of this paper.

References

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

Bhan, A., Hussain, I., Ansari, K. I., Bobzean, S. A., Perrotti, L. I., & Mandal, S. S. (2014). Bisphenol-A and diethylstilbestrol exposure induces the expression of breast cancer associated long noncoding RNA HOTAIR in vitro and in vivo. *The Journal of steroid biochemistry and molecular biology*, *141*, 160-170. https://doi.org/10.1016/j.jsbmb.2014.02.002

Campbell, P. T., Edwards, L., McLaughlin, J. R., Green, J., Younghusband, H. B., & Woods, M. O. (2007). Cytochrome P450 17A1 and catechol O-methyltransferase polymorphisms and age at Lynch syndrome colon cancer onset in Newfoundland. *Clinical cancer research*, *13*(13), 3783-3788.

Carey, A. H., Waterworth, D., Patel, K., White, D., Little, J., Novelli, P., ... & Williamson, R. (1994). Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Human molecular genetics*, *3*(10), 1873-1876.

Chlebowski, R. T., Wactawski-Wende, J., Ritenbaugh, C., Hubbell, F. A., Ascensao, J., Rodabough, R. J., ... & White, E. (2004). Estrogen plus progestin and colorectal cancer in postmenopausal women. *New England Journal of Medicine*, *350*(10), 991-1004.

Datkhile, K. D., Gudur, R. A., Gudur, A. K., Patil, M. N., Durgawale, P. P., Jagdale, N. J., ... & More, A. L. (2021). Cytochrome P450 17 (CYP17) gene polymorphism (rs743572) and cervical cancer risk in women of rural Maharashtra. *Gene Reports*, *24*, 101219. https://doi.org/10.1016/j.genrep.2021.101219

Demers, S. (2011). Female hormone therapy. *Le Médecin du Québec*, volume 46, numéro 5

Ditonno, I., Losurdo, G., Rendina, M., Pricci, M., Girardi, B., Ierardi, E., & Di Leo, A. (2021). Estrogen receptors in colorectal cancer: Facts, novelties and perspectives. *Current Oncology*, *28*(6), 4256-4263.

Edwards, B. K., Noone, A. M., Mariotto, A. B., Simard, E. P., Boscoe, F. P., Henley, S. J., ... & Ward, E. M. (2014). Annual Report to the Nation on the status of cancer, 1975-2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. *Cancer*, *120*(9), 1290-1314.

Effah, C. Y., Wang, L., Agboyibor, C., Drokow, E. K., Yu, S., Wang, W., & Wu, Y. (2020). Polymorphism in the androgen biosynthesis gene (CYP17), a risk for prostate cancer: a meta-analysis. *American Journal of Men's Health*, *14*(5), 1557988320959984.

Excoffier, L., & Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, *7*(10), 745-758.

Farese, S. A., & Ochsenbein, A. F. (2008). Monoclonal antibodies in the treatment of cancer. *In Forum Med Suisse* (Vol. 8, No. 9, pp. 160-165).

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

Gao, Z., Wyman, M. J., Sella, G., & Przeworski, M. (2016). Interpreting the dependence of mutation rates on age and time. *PLoS biology*, *14*(1), e1002355.

Gaudet, M. M., Lacey, J. V., Lissowska, J., Peplonska, B., Brinton, L. A., Chanock, S., & Garcia-Closas, M. (2008). Genetic variation in CYP17 and endometrial cancer risk. *Human genetics*, *123*, 155-162.

Gilep, A. A., Sushko, T. A., & Usanov, S. A. (2011). At the crossroads of steroid hormone biosynthesis: the role, substrate specificity and evolutionary development of CYP17. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1814*(1), 200-209. https://doi.org/10.1016/j.bbapap.2010.06.021

Grant, W. S., Clark, A. M., & Bowen, B. W. (1998). Why restriction fragment length polymorphism analysis of mitochondrial DNA failed to resolve sardine (Sardinops) biogeography: insights from mitochondrial DNA cytochrome b sequences. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(12), 2539-2547.

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.

Hahnloser, D., Ch, C., Nassiopoulos, K., Wahl, P., & Petropoulos, P. (2002). Laparoscopic surgery of colorectal cancer. Apropos of 103 interventions. *Swiss Surgery= Schweizer Chirurgie= Chirurgie Suisse= Chirurgia Svizzera*, *8*(5), 203-208.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series* (Vol. 41, No. 41, pp. 95-98).

Hanna, N., Parfait, B., Vidaud, D., & Vidaud, M. (2005). Mechanisms and consequences of mutations. *M/S: medicine sciences* , *21* (11), 969-980.

Hedrick, P. W. (2012). What is the evidence for heterozygote advantage selection?. *Trends in ecology & evolution*, *27*(12), 698-704.

Heemskerk-Gerritsen, B. A., Rookus, M. A., Aalfs, C. M., Ausems, M. G., Collée, J. M., Jansen, L., ... & Seynaeve, C. (2015). Improved overall survival after contralateral risk-reducing mastectomy in BRCA1/2 mutation carriers with a history of unilateral breast cancer: a prospective analysis. *International journal of cancer*, *136*(3), 668-677.

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F ST. *Nature Reviews Genetics*, *10*(9), 639-650.

Hudson, R. R. (2000). A new statistic for detecting genetic differentiation. *Genetics*, *155*(4), 2011-2014.

Hwang, D. Y., Hung, C. C., Riepe, F. G., Auchus, R. J., Kulle, A. E., Holterhus, P. M., ... & Chen, H. C. (2011). CYP17A1 intron mutation causing cryptic splicing in 17α-hydroxylase deficiency. *PloS one*, *6*(9), e25492.

Kalyan, A., Kircher, S., Shah, H., Mulcahy, M., & Benson, A. (2018). Updates on immunotherapy for colorectal cancer. *Journal of gastrointestinal oncology*, *9*(1), 160.

Keneme, B., & Sembene, P. M. (2021). Pathogenicity of CYP17α mutations in the occurrence of uterine fibroids and breast fibroadenomas. *International Journal of the Archive of Biological and Pharmaceutical Sciences*, *1*(2), 085-099.

Kircher, S. M., Nimeiri, H. S., & Benson III, A. B. (2016). Targeting angiogenesis in colorectal cancer: tyrosine kinase inhibitors. *The Cancer Journal*, *22*(3), 182-189.

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

Li, T. C., Tuckerman, E. M., & Laird, S. M. (2002). Endometrial factors in recurrent miscarriage. *Human Reproduction Update*, *8*(1), 43-52.

Lieu, E. L., Nguyen, T., Rhyne, S., & Kim, J. (2020). Amino acids in cancer. *Experimental & molecular medicine*, *52*(1), 15-30.

Mendiratta, G., Ke, E., Aziz, M., Liarakos, D., Tong, M., & Stites, E. C. (2021). Cancer gene mutation frequencies for the US population. *Nature communications*, *12*(1), 5961.

Miller, W. L., & Auchus, R. J. (2011). The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocrine reviews*, *32*(1), 81-151.

Minton, J. A., Flanagan, S. E., & Ellard, S. (2011). Mutation surveyor: software for DNA sequence analysis. *PCR Mutation Detection Protocols*, 143-153.

Missaghian, E., Kempná, P., Dick, B., Hirsch, A., Alikhani-Koupaei, R., Jégou, B., ... & Fluck, C. E. (2009). Role of DNA methylation in the tissue-specific expression of the *CYP17A1* gene for steroidogenesis in rodents. *Journal of Endocrinology, 202*(1), 99.

Murez, T., Basset, V., Audenet, F., Lebret, T., & Branchereau, J. (2013). Qualitative and quantitative hormonal regulation in castration-resistant prostate cancer. *Progress in urology*, 23, S1-S8.

Nakajin, S., & Hall, P. F. (1981). Microsomal cytochrome P-450 from neonatal pig testis. Purification and properties of A C21 steroid side-chain cleavage system (17 alpha-hydroxylase-C17, 20 lyase). *Journal of Biological Chemistry*, *256*(8), 3871-3876. https://doi.org/10.1016/S0021-9258(19)69538-4

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*(949), 283-292.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia university press.

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

Nikkilä, J., Christin Parplys, A., Pylkäs, K., Bose, M., Huo, Y., Borgmann, K., ... & Winqvist, R. (2013). Heterozygous mutations in PALB2 cause DNA replication and damage response defects. *Nature communications, 4*(1), 2578.

Rai, R., Sharma, K. L., Misra, S., Kumar, A., & Mittal, B. (2014). CYP17 polymorphism (rs743572) is associated with increased risk of gallbladder cancer in tobacco users. *Tumor Biology*, *35*, 6531-6537.

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

Rogers, A. R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular biology and evolution, 9*(3), 552-569.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution*, *34*(12), 3299-3302.

Sata, F., Yamada, H., Yamada, A., Kato, E. H., Kataoka, S., Saijo, Y., ... & Kishi, R. (2003). A polymorphism in the CYP17 gene relates to the risk of recurrent pregnancy loss. *MHR: Basic science of reproductive medicine*, *9*(11), 725-728.

Setiawan, V. W., Schumacher, F. R., Haiman, C. A., Stram, D. O., Albanes, D., Altshuler, D., ... & Chanock, S. J. (2007). CYP17 genetic variation and risk of breast and prostate cancer from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiology Biomarkers & Prevention*, *16*(11), 2237-2246.

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.

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

Sun, X., Li, M., Wang, Y., & Bi, Y. (2018). Two Novel Heterozygous Mutations in the CYP17A1 Gene in a Chinese Patient with 17alpha-Hydroxylase 17, 20-Lyase Deficiency. *Discovery Medicine, 26*(145), 243-249.

Swofford, D. L., Thorne, J. L., Felsenstein, J., & Wiegmann, B. M. (1996). The topology-dependent permutation test for monophyly does not test for monophyly. *Systematic Biology, 45*(4), 575-579.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), 437-460.

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

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.

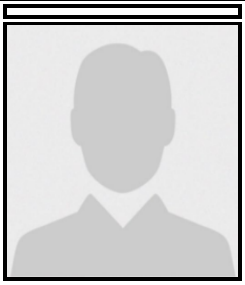
Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, *22*(22), 4673-4680.

Tüzüner, B. M., Öztürk, T., Kisakesen, H. I., Ilvan, Ş., Zerrin, C., Öztürk, O., & Isbir, T. (2010). CYP17 (T-34C) and CYP19 (Trp39Arg) polymorphisms and their cooperative effects on breast cancer susceptibility. *in vivo*, *24*(1), 71-74.

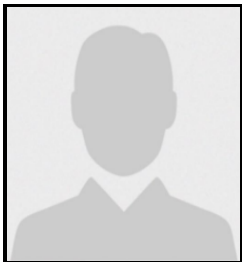
Vettore, L., Westbrook, R. L., & Tennant, D. A. (2020). New aspects of amino acid metabolism in cancer. *British journal of cancer, 122*(2), 150-156.

Wright, S. (1950). Genetical structure of populations. *Nature*, 1950, Vol. 166, 247-49

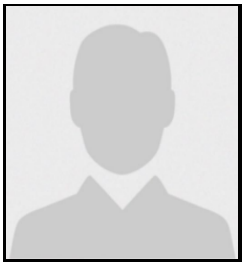
Biography of Authors



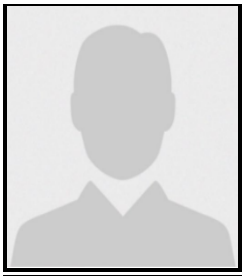
Anna NDONG
Cheikh Anta Diop University, Dakar, Senegal
Email: anna.ndong@ucad.edu.sn



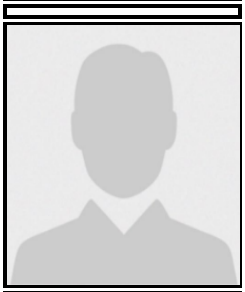
Bineta KENEME
Department of Animal Biology, Faculty of Science and Technology (FST); Laboratoire of Genomic, Cheikh Anta Diop University, Animal Biology, Dakar-Fann, Senegal
Email: bineta.keneme@ucad.edu.sn



Yacine SEYE
Cheikh Anta Diop University of Dakar (UCAD), DES, Residency in General Surgery
Email: yacine87@yahoo.fr



Fatimata MBAYE
Department of Animal Biology, Faculty of Science and Technology (FST); Laboratoire of Genomic, Cheikh Anta Diop University, Animal Biology, Dakar-Fann, Senegal
Email: fatimata2.mbaye@ucad.edu.sn



Mbacké SEMBENE
Department of Animal Biology, Faculty of Science and Technology (FST); Laboratoire of Genomic, Cheikh Anta Diop University, Animal Biology, Dakar-Fann, Senegal
Email: mbacke.sembene@ucad.edu.sn