IBBJ

Autumn 2017, Vol 3, No 4

Original Article

Polymorphism and Genetic Diversity of *BAT25* Marker in Colorectal Cancer

Arame Ndiaye[1*], Fatimata Mbaye[2], Bineta Kénémé[2], Fatou Diallo[1], Abdourahmane Samba[1], Fatou Cisse[1], Souleymane Thiam[1], Dominique Doupa[3], Papa saloum Diop[4], Mbacke Sembene[2], Niama Diop Sall[1, 3]

1. Laboratory of Molecular Biochemistry –Biology, Cheikh Anta Diop Universityof Dakar, Senegal.

2. Department of Animal Biology, Faculty of Science, Cheikh Anta Diop University of Dakar, Senegal.

3. Laboratory of Molecular Biochemistry-Biology, Training and Research Unit, Faculty of Medicine (UFR), Gaston Berger University, Saint Louis, Senegal.

4. Department of digestive surgery and hepatobiliary-Cancerology, Grand Yoff Hospital, Dakar, Senegal.

Submitted 2 May 2017; Accepted 26 Jun 2017; Published 30 Jul 2017

By its frequency, colorectal cancer (CRC) has become a serious public health problem. *BAT25* marker is included in the panel of five markers dedicated to the diagnosis of CCR microsatellite instability. The objective of this study was to evaluate the polymorphism and genetic diversity of the *BAT25* marker in CRC cases in Senegal. This prospective study was performed on 22 CRC patients. After DNA extraction, the polymorphism and genetic diversity of *BAT25* was determined by PCR and sequencing. The alignment of the sequences was carried out using Bio Edit software. The parameters of the polymorphism and genetic variability were determined using the Dnasp, Mega and the Arelquin programs. The results showed a genetic variability of *BAT25*. This variability is represented by mutations observed in the tumor tissues. The most frequent mutation was the deletion of a thymine at position 72 (T72d). This deletion was absent in healthy tissues and controls. From this study, it can be concluded that the mutations found in the tumor tissues could have a role in the onset, development and progression of CRC.

Key words: Colorectal cancer, microsatellite, *BAT25*, polymorphism

By its frequency, colorectal cancer (CRC) has become a serious public health problem. It is the third most common malignant disease in terms of incidence and mortality in developed countries (1). In Africa, CRC is relatively uncommon and represents 1 to 3% of cancers, and diagnosis is usually late (2). In Senegal, despite the absence of a cancer registry and the virtual absence of studies on colorectal cancers, the prevalence seems to begin to reach alarming proportions. This phenomenon is observed in most African countries.

This disease exists in several types of tumors, including those due to replication error and microsatellite instability (MSI). At each replication of the DNA, DNA polymerase makes mistakes that are minimal. These errors are repaired by a system known as the Mut HLS system (3). When this system works poorly, DNA replication errors are not repaired. Because of their repeated structure, microsatellites are the target of these errors, and in this case a defect in the DNA repair system could cause the microsatellite instability observed in these

*Correspondence: Arame Ndiaye, Laboratory of Biochimistry and Molecular Biology Cheikh Anta Diop University of Dakar, Senegal.
E-mail: yssndiay@yahoo.fr

types of cancer.

For the diagnosis of CRC, although numerous and contradictory results are obtained, but like during the international Bethesda meeting (Roland et al. 1998), they all include *BAT25* as a key marker in the diagnosis of unstable colorectal cancer (MSI) (4-7).

This study was conducted within this framework and aimed at analyzing the *BAT25* marker in CRC cases in Senegal. More specifically, it aimed to assess the frequency of *BAT25* mutations, the variability, and genetic diversity of *BAT25* in CRC cases in Senegal. Also, genetic differentiation of CRC based on pathological characteristics was investigated.

Materials and methods

Patients and samples

This is a prospective study on 22 patients with colorectal cancer. Patients were selected at the health care facilities such as Hôpital Principal, Grand Yoff hospital and the Aristide Le Dantec University Hospital in Dakar. The distribution of the study population according to the gender, age and location of the tumor is specified in Table 1. After surgery, the surgical specimens were sent to the laboratory of pathology of the health care center. For each patient, a sample of healthy tissue and a tumoral tissue were prepared and then stored in 96% alcohol for genetic analysis. Blood samples from healthy subjects were used as controls.

DNA extraction, amplification and sequencing

The total DNA was extracted from tissues using the QIAGENs kit (QIAmp DNA FFPE Tissue) and the Promega kit. The quality of the DNA was checked by electrophoretic migration on a 1.5% agarose gel. The amplification of the *BAT25* marker was performed in a reaction volume of 50 µl. PCR was performed with 34.9 µl of MilliQ water, 5 µl of 10 × buffer, 1 µl of MgCl2, 2 µl of dNTPs, 2.5 µl of each primer, 0.1 µl of Taq (5U/ µl) and 2 µl of DNA. The primers used were described by Buhard et al. (8). The amplified products were sequenced with a 3730xl ABI type sequencer. The sequences obtained have been corrected and aligned with the Bio Edit software version 8.0.5 (9).

Polymorphism analysis and genetic diversity of *BAT25*

The parameters of the polymorphism, such as the length of the sequences, the number of variable sites (polymorphs), the number of informative sites in parsimony, the number of mutations, the number of haplotypes and the standard parameters of genetic diversity such as haplotype diversity (Hd) and nucleotide diversity (Pi) were determined using the Dnasp version 5.10 software (10). The Hd defines

Table 1. Clinical and anatomo-pathological characteristics of the population

Variables	Numbers	Percentage (%)
Gender		
Male	10	45
Female	12	55
Age (years)		
< 30	0	
30-60	10	45
> 60	12	55
Localisation		
Colon	18	82
Rectum	4	18

the probability that two genes randomly drawn in a sample are similar (11) and considers the number of individuals and the frequency of haplotypes. As a result, the higher Hd, the greater are the chances of observing different haplotypes in two randomly selected individuals. (Pi) is also a measure of genetic diversity, but it also integrates the actual degree of difference between the sequences studied. Thus, the more the haplotypes present in a population are distinct from each other in terms of the number of polymorphic sites differentiating them, the higher (Pi) will be. The nucleotide frequencies and the nature of the mutations were determined using MEGA 6.06 software (12). The average number of nucleotide differences (K) was found under Dnasp version 5.10.01 (10). The polymorphism and genetic diversity of *BAT25* were evaluated at the healthy intra-tissue, tumor intra-tissue, healthy inter-tissue and tumor tissue levels.

Genetic differentiation tests

The intra- and inter-group Nei genetic distance was calculated under MEGA 6.06 (12) using the two -parameter Kimura model. The Fst (genetic differentiation factor) values and the associated probabilities were determined using the Arelequin 3.5.1.3 program (13). A P value < 0.05 was considered as significant. The groups were related to tumor tissues, healthy tissues and controls. To see if there was a genetic differentiation according to the clinical and pathological parameters, we carried out the genetic differentiation test (Fst) under Harlequin. Here groups were related to the clinical and pathological parameters (gender, age, location).

Results

Variability and genetic diversity

Studies were performed on a 222 bp fragment. Controls and healthy tissues showed no mutations in contrast to tumor tissues which revealed 4 polymorphic sites (Table 2). Seven haplotypes were found. Controls and healthy tissues showed a single haplotype. The values for nucleotide frequencies for all samples showed a predominance of A + T nucleotides (67.26%) compared with C + G (32.74%). In the control group and healthy tissues, transitions accounted for 33.33% and transversions

Table 2. Indices of genetic variability of *BAT25* sequences

Parameters	Controls	Healthy tissues	Tumoral tissues	Total
Sample size	10	7	22	39
Number of sites	222	222	222	222
Number of monomorphic sites	222	222	218	222
Number of polymorphic sites	0	0	4	0
Number of haplotypes	1	1	6	7
Transitions (%)	33,33	33,33	0	0
Transversions (%)	66,66	66,66	100	100
Mutation rate (R)	0,41	0,41	0	0
Hd	0	0	0,6407	0,694
Pi	0	0	0,0015	0,00087
K	0	0	0,338	0,192
Healthy and tumoral tissues				
Pi	0,00119			
K	0,26			

Hd: haplotypic diversity; Pi: nucleotide diversity; K: the average number of nucleotide differences.

for 66.66% of mutation cases. Our results showed a predominance (100%) of the transversions in the cancerous tissues (Table 2). Compared with controls and healthy tissues, tumor tissues revealed a high haplotype diversity (0.064 ± 0.117) and a low nucleotide diversity of 0.0015 ± 0.0005. The average number of nucleotide differences (k) in the tumor tissues was 0.338.

The haplotype h1 was present in controls and healthy tissues. The other haplotypes were represented by tumoral tissues. According to Table 3, mutations were at positions 32T>A, 70T>G, 71T>G, 71delT, and 72delT.

Differentiation and genetic structuring

The determination of intra and inter-group genetic distances, showed that there was no difference between controls and healthy tissues. On the other hand, there was a difference (distance= 0.000829 ± 0.6279) between healthy and tumor tissues, as well as between controls and tumor tissues (Table 4).

Table 3. *BAT25* haplotypes in controls, healthy, and colorectal tissues

Haplotypes	Number of sequences	Number of mutations	Mutations
h1	18	0	Controls and healthy tissues
h2	1	2	T32A ; T72d
h3	13	1	T72d
h4	2	3	T32A ; T71G ; T72d
h5	2	2	T71d ; T72d
h6	3	2	T71G ; T72d
h7	1	3	T70G ; T71G ; T72d

Table 4. Genetic differentiation parameters

Nei's genetic distance

Groups	Intra-group	Inter-group
Controls Healthy tissues Tumoral tissues	0 0 0.0015± 0.0011	
Controls Tumoral tissues		0.0008±0.627
Healthy tissues Tumoral tissues		0.0008±0.627

Genetic differentiation factors (Fst)

Groups	Inter-group	P-value
Controls Healthy tissues	0	< 0.0001
Controls Tumoral tissues	0.639	< 0.0001
Healthy tissues Tumoral tissues	0.610	< 0.0001

Discussion

The age range of our study population was between 35 and 92 years, with a majority (55%) being above 60 years. The age of CRC in sub-Saharan Africa is generally below 50 years (46 years on average) (14) compared to the age range observed in developed countries 65 ± 10 years. These results can be explained by the difference in lifestyle. The occurrence of CRC is strongly influenced by environmental factors, particularly the diet.

The study of the polymorphism of *BAT25* showed that there was no variability between the reference sequence, controls and healthy tissues that were monomorphic in our study population. On the other hand, the literature shows a great allelic variability at the *BAT25* locus. According to a study conducted by Pyatt et al. (15) on a population of 103 healthy African American subjects, 21.3% had variations of the *BAT25* locus. Moreover, the polymorphic nature of this locus was also confirmed in a Nigerian population. Zheng et al. (16) showed that the *BAT25* locus is nearly monomorphic. They attempted to establish a relationship between the length of the fragment and poly thymidine sequences in *BAT25* in a Chinese population, and found that *BAT25* alleles length were variable in the population.

Tumor tissues showed variability regarding two sites that have simple substitutions and two sites with deletions. These mutations could be incriminated to the disease because we find them only in the tumor tissues. According to our results, the *BAT25* marker showed 6 haplotypes in the CRC. Controls and healthy tissues were grouped within the same haplotype (h1) as there was no variability between them. The other 6 haplotypes corresponded to the tumor tissues. All individuals with CRC had a T72d mutation. Individuals with the T72d deletion were grouped into haplotype 3. Patients in this group were all at the pT2 or pT3 stage of the disease. This deletion in the tumor tissues could be implicated in the onset and the evolution of the disease. An analysis of 19 MSI tumors performed by Pyatt et al. (15) showed a shortening of the poly (T) tract from 4 to 10 base pairs for all tumor tissues.

The values relating to the nucleotide frequency for all the samples show a predominance of A (25.49%) and T (41.77%) with respect to C (11.82) and G (20.92%). For controls and healthy tissues the transitions accounted for 33.33% of mutations. Whereas for the tumor tissues all the mutations were transversion (100%). According to the literature, mutations found in healthy individuals are often transitions. In the case of tumor tissues, transversions are more frequent. Our results regarding predominance of transversions in tumor tissues are in line with the results obtained by previous researches on different tumor tissues (17-18).

The Hd (haplotype diversity) and Pi (nucleotide diversity) values were determined for all the sequences studied in different groups. The indices of haplotype and nucleotide diversity are parameters for estimating demographic history in phylogeny. A high Hd for tumor tissue is the signal of very rapid growth of tumor cells. These results are in line with the literature in which tumor cell expansion is an important and frequent phenomenon in digestive tumors (19).

The determination of the genetic distances (intra and inter) confirmed that there was no difference between controls and healthy tissues. On the other hand, there was a difference (distance = 0.0008 ± 0.627) between healthy and tumor tissues. The present results showed a variability within the tumor tissues reflecting the heterogeneity of the CCRs. With Fst, the difference between healthy and tumor tissues was statistically significant ($P< 0.0001$). This result reinforces the hypothesis already announced; namely that there is a molecular difference between healthy and tumor tissues.

No genetic differentiation was detected by gender, age, and location. This result is in line with the study carried out by Mbaye et al. on *BAT25* instability in Senegalese women with breast cancer

(unpublished data).

In conclusion, this study showed the presence of mutations in the tumor tissues of patients with CRC. The T72d mutation found only in tumor tissues can be incriminated to disease development.

Conflict of interest

The authors declared no conflict of interest.

References

1. Haggar F A, Boushey R P. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg. 2009;22:191-7.

2. Torre L A, Bray F, Siegel R L, et al. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65:87-108.

3. Duval A, Hamelin R. Réparation des erreurs de réplication, microsatellites et cancer. médecine/sciences. 2003;19:55-62 (Article in French).

4. Iachetta F, Domati F, Reggiani-Bonetti L, et al. Prognostic relevance of microsatellite instability in pT3N0M0 colon cancer: a population-based study. Intern Emerg Med. 2016;11:41-6.

5. Pagin A, Zerimech F, Leclerc J, et al. Evaluation of a new panel of six mononucleotide repeat markers for the detection of DNA mismatch repair-deficient tumours. Br J Cancer. 2013;108:2079-87.

6. Wang Y, Shi C, Eisenberg R, et al. Differences in Microsatellite Instability Profiles between Endometrioid and Colorectal Cancers: A Potential Cause for False-Negative Results? J Mol Diagn. 2017;19:57-64.

7. Ziadi S, Ksiaa F, Ben Gacem R, et al. Clinicopathologic characteristics of colorectal cancer with microsatellite instability. Pathol Res Pract. 2014;210:98-104.

8. Buhard O, Cattaneo F, Wong Y F, et al. Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors. J Clin Oncol. 2006;24:241-51.

9. Hall T A. Bio Edit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser.1999. 41 : 95-98.

10. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451-2.

11. Nei M. Molecular evolutionary genetics: Columbia university press; 1987.

12. Tamura K, Stecher G, Peterson D, et al. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30:2725-9.

13. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online. 2007;1:47-50.

14. Graham A, Adeloye D, Grant L, et al. Estimating the incidence of colorectal cancer in Sub-Saharan Africa: A systematic analysis. J Glob Health. 2012;2:020404.

15. Pyatt R, Chadwick R B, Johnson C K, et al. Polymorphic variation at the BAT-25 and BAT-26 loci in individuals of African origin. Implications for microsatellite instability testing. Am J Pathol. 1999;155:349-53.

16. Zheng Y Y, Xie L, Liu L, et al. BAT-25 polymorphism in Chinese from Jiangsu province and its implication for locus microsatellite instability screening. Int J Biol Markers. 2012;27:e227-31.

17. Gao W, Mady H H, Melhem M F, et al. Analysis of p53 mutations in histologically normal lung tissues and lung tumors from non-small cell lung cancer patients. Mol Carcinog. 2009;48:633-41.

18. Liu S, Liu W, Jakubczak J L, et al. Genetic instability favoring transversions associated with ErbB2-induced mammary tumorigenesis. Proc Natl Acad Sci U S A. 2002;99:3770-5.

19. Vasseur S, Guillaumond F. Pancreatic adenocarcinoma: a tumor highly dependent on lipoproteins. Med Sci (Paris). 2015;31:728-30.