Virus Research 243 (2018) 1–9

Contents lists available at ScienceDirect

Virus Research

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

Phylogenetic analysis of Human papillomavirus 16 variants isolated from Indian Breast cancer patients showed difference in genetic diversity with that of cervical cancer isolates

Saimul Islam[a], Dipanjana Mazumder (Indra)[b], Mukta Basu[a], Anirban Roychowdhury[a], Pijush Das[c], Hemantika Dasgupta[a], Anup Roy[d], Neyaz Alam[e], Ranajit Kumar Mondal[f], Susanta Roychoudhury[g], Chinmay Kumar Panda[a,*]

[a] Department of Oncogene Regulation, Chittaranjan National Cancer Institute, 37, S.P. Mukherjee Road, Kolkata, West Bengal 700026, India
[b] Department of Zoology, Siliguri College, Darjeeling, West Bengal, 734001, India
[c] Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Kolkata, West Bengal 700032, India
[d] Department of Pathology, NilRatan Sircar Medical College and Hospital, Sealdah, Raja Bazar, 138, Acharya Jagadish Chandra Bose Rd, Kolkata, West Bengal 700014, India
[e] Department of Surgical Oncology, Chittaranjan National Cancer Institute, 37, S.P. Mukherjee Road, Kolkata, West Bengal 700026, India
[f] Department of Gynaecology Oncology, Chittaranjan National Cancer Institute, 37, S.P. Mukherjee Road, Kolkata, West Bengal 700026, India
[g] Saroj Gupta Cancer Centre and Research Institute, Thakurpukur, Kolkata 700 063, India

ARTICLE INFO

Keywords:
Human papilloma virus (HPV)
Viral phylogeny
Breast cancer (BC)
Cervical cancer (CACX)
Viral pathogenesis

ABSTRACT

The genetic variations of HPV16 in Breast Cancer (BC) are not well studied unlike HPV16 in Cervical Cancer (CACX). In this study, the genetic variations of HPV16 in BC were compared with HPV16 in CACX. In sequencing analysis of LCR, E6 and E7 regions of HPV16 in BC and CACX the A lineage was seen to be 64.2% and 66.6% respectively. The other lineages showed differential frequency in BC and CACX. The mutation frequency index of the regions in BC and CACX was in the following order: LCR > E6 > E7. However, the inter-patient genetic diversity in LCR and E6/E7 regions was high in BC than CACX. The LCR region showed more variations than the E6/E7 region in BC. Apart from some common variations, some unique tissue specific variants in LCR and E6/E7 region were seen in BC and in CACX. Besides the selection of some common variants in both BC and CACX, some unique variants in BC (D98Y; 395 G > T) and CACX (R48W; 245 G > T) were observed. The 7521 G > A variant of LCR showed association with Luminal B subtype of BC and progression of CACX. Whereas, 145 G > T (Q14H) and 335 C > T (H78Y) variants of E6 showed association with either early invasiveness of BC and/or poor outcome of the patients. Thus, this study indicates that there may be a difference in the genetic variation of HPV16 in BC and in CACX.

1. Introduction

Human papillomavirus (HPV) belongs to the family *papillomaviridae* (IARC Monogr Eval Carcinog Risks Hum, 2007). Within the family, the *alpha* genus contains HPVs that are associated with the development of mucosal tumours in human while the genus *beta* contains those that are involved in the development of cutaneous diseases (Bernard et al., 2010). The HPV genome consists of a non-coding long control region (LCR) and eight protein-coding genes like L1, L2, E1, E2, E4, E5, E6 and E7 (Van Doorslaer et al., 2013). Based on the more than 10% nucleotide variability of L1 gene sequence, HPVs are divided into more than 200 genotypes (Burd, 2003; Van Doorslaer et al., 2013). Only 12 mucosal HPV genotypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59 are designated as 'high-risk' HPVs (hrHPVs) (Van Doorslaer et al., 2013). Among these hrHPVs, HPV16, primarily, showed frequent association with different cancers like cervical cancer (CACX), head and neck squamous cell cancer (HNSCC), prostate cancer, breast cancer (BC), etc. (Bae and Kim, 2016; Michaud et al., 2014; Munoz et al., 2003; Shukla et al., 2009).

Now, depending on the nucleotide sequence variability of HPV16 genome, HPV16 genotype was geographically branched into four phylogenetic lineages A, B, C and D and their following sub-lineages A1(European), A2(European), A3(E), A4(Asian), B1(African-1a), B2(African-1b), D1(North American 1), D2 (Asian-American 2) and

* Corresponding author.
E-mail address: ckpanda.cnci@gmail.com (C.K. Panda).

http://dx.doi.org/10.1016/j.virusres.2017.10.004
Received 29 June 2017; Received in revised form 21 September 2017; Accepted 4 October 2017
Available online 06 October 2017

D3(Asian-American 1) (Burk et al., 2013). The clustering of HPV variants was often reported to be associated with the specific geographical location, ethnicity and tumours of different organs (Bernard et al., 2006). The different lineages have diverse selection pressures leading to varied oncogenic potential and pathogenicity of the virus (Chen et al., 2005). It is evident that the selection pressures of the variants of the proteins depend on the changes in ratio of non-synonymous (dn) and synonymous (ds) variation in amino acids in the coding regions (Yang et al., 2000). It is known, that if dn/ds = w > 1 for a particular amino acid in an ORF, indicates that the amino acid has a strong molecular adaptation force, i.e positive selection pressure (Yang et al., 2000). The selection pressure is regulated by several factors including the host's geographical location, ethnicity, immune response, tissue micro-environment (Sichero and Villa, 2006).

Several studies have shown that E31G, L83V and D25E transitions on E6 coding region of HPV16 in the European lineage were significantly associated with the development of CACX in North American, Great Britain's and the Koreas' populations respectively (Bernard et al., 2006; Sichero and Villa, 2006). However, the N29S variant on E7 coding region of HPV16 in European lineage was strongly linked with CACX of Thai patient population (Song et al., 1997). In addition, some variants in the URR regulating the transcription factor binding sites like, TEF-1 (G7193T, C7689A), GRE-1 (A7458T, A7485C, G7489A) and YY1 (G7521A, C7786T, G7826A, A7837C, A7839G) were seen to be prevalent in non-European lineages and were reported to be associated with CACX development (Bernard et al., 2006; Sichero and Villa, 2006). Likewise, the R10G variant of E6 region of HPV16 was relatively frequent (19%) in tonsillar squamous cell carcinoma (TSCC) than CACX in Swedish patients (Du et al., 2012).

In our previous study, frequent (69%) infection of HPV16 has been seen in BC of Indian patients (Islam et al., 2017). To the best of our knowledge, the genetic variations of HPV16 in BC, if any, or other tumours, are not studied in details. The knowledge of the genetic variations of HPV16 in BC will help to understand the natural history and pathogenicity of the virus as well as to develop proper vaccination strategies against the virus.

Therefore, in the present study, nucleotide variations in the LCR and E6/E7 regions of HPV16 in primary BC of Indian patients were analyzed at first, followed by its comparison with the sequences in HPV16 of CACX of Indian patients. Our data showed (i) lineage A was frequent in both BC and CACX with high inter-patients genetic distance in BC and (ii) positive selection of D98Y (395 G > T) variant of E6 ORF in BC while, the R48W (245 G > T) variant of E6 ORF in CACX indicating there may be a difference in the genetic variation of HPV16 of these two tumours.

2. Materials and methods

2.1. Sample collection and demography of patients

The HPV16 positive primary breast cancer (N = 57) and cervical cancer (N = 30) samples from our previous studies (Islam et al., 2017) and (Mazumder (Indra) et al., 2011) were randomly included in this study. Tissue samples of primary Breast carcinoma (BC) and Cervical cancer (CACX) used in the earlier studies (Islam et al., 2017; Mazumder (Indra) et al., 2011) were collected from the hospital section of the Chittaranjan National Cancer Institute, Kolkata after appropriate approval from the Institutional Ethical Committee with written informed consent from patients. Samples were frozen immediately after collection at −80 °C until use. The details of clinicopathological parameters of BC and CACX samples were shown in Table S1a and b.

2.2. Micro-dissection and DNA extraction

Samples were microdissected to enrich tumour cells (more than 60%) for DNA isolation by standard phenol/chloroform method (Sambrook et al., 2006).

2.3. HPV screening and genotyping

The presence of HPV was previously detected by polymerase chain reaction (PCR) using primers (MY09 and MY11) from consensus L1 region (Dutta et al., 2012). Screening of HPV16/18/33 in HPV positive samples was further done by PCR using specific primers from E6, LCR and E1 region (Table S2) (Dutta et al., 2012; Baay et al., 1996) HPV 16/18/33 plasmids were used as positive control. Thus, based on this previous analysis (Islam et al., 2017; Mazumder (Indra) et al., 2011), it was observed that among the HPV16 positive cases (BC n = 57 and CACX n = 30), four BC samples (BC06, BC08, BC18, BC28,) and one CACX (CACX05) samples also have HPV18 co-infection. In addition, one BC (BC49) samples was infected with both HPV18 and HPV33.

2.4. Sequencing analysis of LCR and E6-E7

Sequencing analysis of LCR region was done by polymerase chain reaction (PCR) in 57 BC using AmpliTaq polymerase with a specific set of primers (Table S2). Using these PCR products, LCR were sequenced in both the sense and antisense directions with a BigDye Terminator Cycle Sequencing 3.1 Kit (Applied Biosystems, USA) according to the manufacturer's instructions in a 3130xl Genetic Analyzer (Applied Biosystems, USA) (Islam et al., 2017) and the sequences were submitted to GenBank (KY314399-KY314455). Similarly, amplification and sequencing of E6-E7 region were done in 56 BC and the sequences were submitted to GenBank (KY290891-KY290946). Likewise, sequencing of LCR and E6-E7 regions were analyzed in 30 and 21 CACX samples respectively. The sequences of both LCR and E6-E7 regions were submitted to GenBank with accession number KY978914- KY978943 and KY978893- KY978913 respectively.

2.5. Phylogenetic tree generation based on LCR and E6-E7 regions

Phylogenetic tree with concatenated LCR and E6-E7 sequences was constructed using 56 (out of 57) BC and 21 (out of 30) CACX samples to further find out the lineages (and sub-lineages), by the Maximum Likelihood (ML) method of MEGA v7.0 software (Tamura et al., 2013). Here, 1 BC sample (BC57) and 9 CACX samples (CACX29, CACX22, CACX27, CACX26, CACX24, CACX23, CACX30, CACX28, CACX25) were excluded as they lack sequences for both LCR and E6-E7 region. Lineage specific reference sequences were retrieved from GenBank database (Benson et al., 2017) [http://www.ncbi.nlm.nih.gov] for the analysis. The reference sequences of sub-lineage(s) are as follows: FJ006723, HQ644282, K02718, AY686584 & KU298880 for A1; AF536179 & FJ610152 for A2;HQ644236 for A3; AF534061, HQ644234 & HQ644261 for A4; AF472508, AF536180 & HQ644296 for B1; HQ644298 for B2; AF472509, HQ644243 & AB818690 for C; HQ644285, HQ644289, HQ644276 & AF402678 for D3; AY686579, HQ644241 & HQ644270 for D2; AB818689 and HQ644257 for D1.The reliability of the tree was assessed by the calculation of bootstrap with 1000 replicates. After that, separate sequences of the LCR and E6-E7 regions in both BC and CACX cases were aligned with standard HPV16 reference sequence (K02718.1) (Gostout et al., 2000; Assoumou et al., 2015 and PaVE database, https://pave.niaid.nih.gov, Doorslaer et al., 2017) using BioEdit version v7.0.1 (Hall, 1999) to identify the sequence variation within these regions according to the data published by Yamada et al. (1997).

The pair-wise genetic distance among BC and CACX samples of LCR and E6/E7 regions were calculated by the Bootstrap method of MEGA v7.0 software (Tamura et al., 2013).

The mutation frequency index (MFI) was calculated according to Tuteja et al. (2014) as describe follows:

$$(\text{MFI}) = [X/(Y \times Z)] \times 100$$