Virus Research 221 (2016) 1–7

Contents lists available at ScienceDirect

Virus Research

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

Phylogenetic and structural analysis of merkel cell polyomavirus VP1 in Brazilian samples

Camila F. Baez[a], Nuria C. Diaz[b], Marianna T. Venceslau[c], Flávio B. Luz[d], Maria Angelica A.M. Guimarães[a], Mariano G. Zalis[a], Rafael B. Varella[c,*]

[a] Department of Preventive Medicine, Universidade Federal do Rio do Janeiro, Brazil
[b] Department of Pharmaceutical Biotechnology, Universidade Federal do Rio do Janeiro, Brazil
[c] Department of Microbiology and Parasitology, Universidade Federal Fluminense, Brazil
[d] Sector of Dermatology, University Hospital Antonio Pedro, Universidade Federal Fluminense, Brazil

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form 3 May 2016
Accepted 6 May 2016
Available online 10 May 2016

Keywords:
Merkel cell polyomavirus
VP1
Phylogeny
Protein structure
South-America

ABSTRACT

Our understanding of the phylogenetic and structural characteristics of the Merkel Cell Polyomavirus (MCPyV) is increasing but still scarce, especially in samples originating from South America. In order to investigate the properties of MCPyV circulating in the continent in more detail, MCPyV *Viral Protein 1* (VP1) sequences from five basal cell carcinoma (BCC) and four saliva samples from Brazilian individuals were evaluated from the phylogenetic and structural standpoint, along with all complete MCPyV VP1 sequences available at Genbank database so far. The VP1 phylogenetic analysis confirmed the previously reported pattern of geographic distribution of MCPyV genotypes and the complexity of the South-American clade. The nine Brazilian samples were equally distributed in the South-American (3 saliva samples); North American/European (2 BCC and 1 saliva sample); and in the African clades (3 BCC). The classification of mutations according to the functional regions of VP1 protein revealed a differentiated pattern for South-American sequences, with higher number of mutations on the neutralizing epitope loops and lower on the region of C-terminus, responsible for capsid formation, when compared to other continents. In conclusion, the phylogenetic analysis showed that the distribution of Brazilian VP1 sequences agrees with the ethnic composition of the country, indicating that VP1 can be successfully used for MCPyV phylogenetic studies. Finally, the structural analysis suggests that some mutations could have impact on the protein folding, membrane binding or antibody escape, and therefore they should be further studied.

1. Introduction

Merkel cell polyomavirus (MCPyV) is a small (~45 nm of diameter) non-enveloped dsDNA circular virus of approximately 5400 pb (Feng et al., 2008). The genome can be divided into three functional regions: the early region encodes regulatory proteins involved in viral DNA replication and gene expression [large T antigen (LT-ag), small T antigen (ST-ag), and an alternative large T antigen (ALTO)], the late region encodes the capsid proteins (VP1, 2 and 3), and the non-coding region encompasses the origin of replication and transcription control elements (NCCR) (Van Ghelue et al., 2012; Shuda et al., 2008; Carter et al., 2013).

MCPyV was the first and so far the only polyomavirus etiologically related to a human neoplasia, i.e., Merkel cell carcinoma (MCC), a neuroendocrine cancer of homonymous cells (Munde et al., 2013; Feng et al., 2008). Although rare, the incidence of this aggressive skin tumor has been increasing in the last two decades (Hodgson, 2005; Goedert, 2009), with strong evidence pointing towards ultra-violet radiation exposure, Caucasian ethnicity, older age and immunosuppression as important risk factors (Hodgson, 2005; Schrama et al., 2012; Robertson et al., 2015; Clarke et al., 2015). It is still not clear how MCC develops, but nearly all MCPyV genomes isolated from tumors are integrated to cellular genome and have mutations that truncate the C terminus of the LT antigen, in a way that the natural ability of the protein to activate viral DNA replication is lost (Chang and Moore, 2012; Shuda et al., 2008). There is a strong selective pressure within tumors to eliminate viral replication capacity and, in addition to LT truncation, certain mutations in VP1, the main capsid protein responsible for virus-

* Corresponding author at: Laboratory of Virology. Department of Microbiology and Parasitology—Biomedical Institute-UFF, Niteroi-RJ, 24210-130, Brazil.
E-mail address: rvarella@id.uff.br (R.B. Varella).

http://dx.doi.org/10.1016/j.virusres.2016.05.004
0168-1702/

cell interaction, prevent the formation of active viruses (Chang and Moore, 2012; Kwun et al., 2009; Kassem et al., 2008).

From the epidemiological perspective, MCPyV prevalence in MCC can vary from 24 to 100% among studies from different locations (Garneski et al., 2009; Sastre-Garau et al., 2009; Bhatia et al., 2011; Leitz et al., 2014). Moreover, MCPyV can be found in diverse tissues samples from MCC and non-MCC patients, as non-lesional skin, blood, hair bulb, saliva, urine, respiratory and digestive tracts, among others (Laude et al., 2010; Loyo et al., 2010; Foulongne et al., 2010a,b, 2012; Murakami et al., 2011; Peretti et al., 2014). Interestingly, MCPyV DNA has been isolated from a variety of non-MCC cancers, as chronic lymphocytic leukemia (Pantulu et al., 2010; Teman et al., 2011; Comar et al., 2012), non-small cell lung cancer (Andres et al., 2009; Gheit et al., 2012; Joh et al., 2010; Hashida et al., 2013), malignant tonsilar tissues (Salakova et al., 2015), cervical carcinomas (Imajoh et al., 2012; Salehi-Vaziri et al., 2015) and non-melanoma skin cancers (Dworkin et al., 2009; Kassem et al., 2009; Murakami et al., 2011; Andres et al., 2010; Rollison et al., 2012; Imajoh et al., 2013; Mertz et al., 2013; Peretti et al., 2014; Du-Thanh et al., 2015).

Despite the increasing number of reports on MCPyV DNA detection in diverse samples from neoplastic and non-cancerous tissues around the world, data regarding to its genetic variability among different regions has just started to be considered. In fact, there is evidence of a geographical distribution for MCPyV, as already seen for other polyomavirus (Zheng et al., 2007; Zhong et al., 2009; Martel-Jantin et al., 2012; Martel-Jantin et al., 2014; Hashida et al., 2014). However, published data are mainly from European, Asiatic and North American populations, which may not be representative of the South-American population. Hence, we aimed to evaluate the genetic variability and phylogeny of MCPyV in samples from Brazil, the largest Latin-American country with a long history of racial mixing, using complete VP1 gene sequences. Moreover, those and previously published mutations are here described in the context of the VP1 structure, in an attempt to enlighten their effects in protein function.

2. Materials and methods

2.1. Samples

For this study, 9 MCPyV-positive samples, being 4 saliva samples (S) from asymptomatic individuals and 5 skin cancer biopsies (basal cell carcinomas or BCC), were used for sequencing. MCPyV DNA was detected according to previously established conditions (Baez et al., 2013). Samples were collected through spontaneous salivation and skin biopsy, respectively. All samples were obtained from residents of Rio de Janeiro State, Brazil. The study was approved by the Bioethics Committee of Rio de Janeiro State University Hospital and Bioethics Committee of Fluminense Federal University Hospital.

2.2. DNA sequencing and analysis

The whole VP1 gene (1272 nucleotides) was amplified using a semi-nested PCR adapted from Hashida et al. (2013). Originally, the technique used by Hashida et al. were conventional PCRs, using three different primer sets (VP1F1 and VP1R1; VP1F2 and VP1R2; VP1F3 and VP1R3) which results in three fragments of 490 bp, 563 bp and 543 bp, respectively. The whole VP1 gene, which comprises 1272 nucleotides, was covered by the amplification of 1413 bp fragment, from nucleotide 3081 to 4493 (Hashida et al., 2013). The methodology was adapted into a semi-nested PCR by using VP1F1/VP1R2 (891pb) previously to the VP1F1/VP1R1 and VP1F2/VP1R2 amplifications, and VP1F2/VP1R3 (987 bp) previously to the VP1F3/VP1R3 reaction. The same reaction conditions described by Hashida et al. were used in the semi-nested steps. The three resulting fragments were purified by GE Healthcare commercial kit. Only full-length VP1 samples were sequenced (sense and anti-sense) using the BigDye TM® terminator chemistry kit (Applied Biosystems, Foster City, Calif., USA) according to previously established conditions (Baez et al., 2014). In addition to the nine Brazilian samples, 38 MCPyV VP1 sequences, corresponding to all complete sequences available at Genbank database up to November 2015, were used for the alignment in order to infer the phylogeny of the MCPyV. Only human-derived and complete VP1 gene sequences were considered, hence sequences from either environmental samples or clones were not used. However, in order to enrich the study with a more representative sampling of South-American and African sequences, partial VP1 sequences obtained of human samples originated from French Guiana and Africa, as well as environmental-derived sequences from Argentina were included in the analysis, totalizing 63 sequences.

Sequence alignments were conducted in ClustalW (Larkin et al., 2007) with default settings. Phylogenetic analysis was carried out using MEGA 7.0 software by the maximum likelihood (ML) algorithm using Tamura 3-parameter model with Gamma distribution (T92 + G), as chosen by the ML model selection feature of Mega 7.0. Statistical significance of clades was measured by 2000 bootstrap replicates.

2.3. VP1 structural analysis

In order to access the position of the mutated residues on the protein structure, and therefore get insight into their effects on protein function, a visual inspection was performed using the published structure of Merkel VP1 protein (PDB code 4FMH) (Neu et al., 2012). Being the N- and C-terminus of VP1 Merkel not resolved, a comparison was made by superposition with the structure of the SV40 virus VP1 protein (PDB code 1SVA) (Stehle et al., 1996). The protein structure was divided into six functional regions, using the following nomenclature: N-terminus (residues 1–51), β-sandwich core (residues 52–59, 102–132, 161–169, 241–248, 254–295, 298–316), apical or membrane binding loops (residues 133–160, 249–253, 296–297), side or neutralizing epitope loops (residues 60–101, 170–240), conserved region of the C-terminus (residues 316–370), and non-conserved region of the C-terminus (residues 370–423) (For more information about those regions, so as its location on the protein structure, see Supplementary info).

3. Results

3.1. MCPyV VP1 phylogenetic analysis

The phylogenetic tree of the VP1 gene generated by the Maximum-likelihood method (Fig. 1) was very similar to the Neighbor Joining method (data not shown). It showed five distinct clades according to geographical distribution: USA/Europe (Caucasian); Asia; Africa; Oceania and South-America.

The phylogenetic analysis of the Brazilian MCPyV VP1 sequences showed heterogeneity of genotypes: 3/9 (33%) sequences (BCC 72, 3 and S13) were grouped within the Caucasian clade; 3/9 (33%) sequences (BCC 4, 9 and 51) within the African clade; and 3/9 (33%) sequences (S10, S22, S19) within the South-American clade. Interestingly, the two skin BCC (BCC3 and 72) sequences grouped in the Caucasian clade were closely related to the MCC350 strain, and one sequence obtained from saliva (S22) was related to a urine-derived Argentinean sequence. Also, two out of three sequences from skin cancer (BCC 9 and 4) were related to sequences from the African clade. Finally, two out of three sequences grouped