Medical Microbiology and Immunology (2019) 208:81–88
https://doi.org/10.1007/s00430-018-0561-2

ORIGINAL INVESTIGATION

The sequence analysis of Epstein–Barr virus EBNA1 gene: could viral screening markers for nasopharyngeal carcinoma be identified?

Ana V. Banko[1] · Ivana B. Lazarevic[1] · Danijela Z. Karalic[1] · Vojko B. Djukic[2] · Maja D. Cupic[1] · Goran Stevanovic[3] · Tanja P. Jovanovic[1]

Received: 16 April 2018 / Accepted: 5 September 2018 / Published online: 10 September 2018

Abstract

Epstein–Barr virus (EBV) has been identified as a group 1 carcinogenic agent, particularly for nasopharyngeal carcinoma (NPC). The sequence diversity of EBV nuclear antigen 1 (EBNA1) reflects region-restricted polymorphisms, which may be associated with the development of certain malignancies. The aims of the present study were to evaluate EBV EBNA1 gene polymorphisms circulating in NPC, infectious mononucleosis, and isolates from patients with transplanted organs to determine if EBNA1 sequence specificities are useful as viral biomarkers for NPC. Forty biopsies of undifferentiated carcinoma of nasopharyngeal type (UCNT), 31 plasma samples from patients with mononucleosis syndrome, and 16 plasma samples from patients after renal transplantation were tested in this study. The EBNA1 gene was amplified by nested PCR. Further investigation included sequencing, phylogenetic, and statistical evaluations. Eighty-seven sequences were identified as one of the four EBNA1 subtypes, P-Ala, P-Thr, V-Val, and V-Ala, with further classification into ten subvariants. Of these, P-Thr-sv-1 and P-Thr-sv-3 have never been identified in Europe, while V-Val-sv-1 was newly discovered. Statistical analysis revealed significant differences in the distribution of EBNA1 P-Thr subvariants between the three groups of patients, with noticeable clustering of P-Thr-sv-5 in NPC isolates ($p < 0.001$). EBV EBNA1 showed no sequence specificity in primary infection. This research revealed a newly discovered EBNA1 subvariant. Importantly, EBNA1 P-Thr-sv-5 showed carcinoma-specific EBNA1 variability. Thus, identification of this subvariant should be considered as a viral screening marker for NPC or UCNT.

Keywords Epstein–Barr virus (EBV) · EBNA1 · Nasopharyngeal carcinoma · Sequence variability · Mononucleosis

Introduction

Epstein–Barr virus (EBV) is an oncogenic herpesvirus that infects over 90% of humans by the time of adulthood. The virus is transmitted through the saliva and primarily infects B-lymphocytes where it establishes a lifelong latent infection. Although primary EBV infection is typically asymptomatic, at times it can result in the benign lymphoproliferative disease infectious mononucleosis (IM), particularly in late childhood or young adulthood in developing countries [1]. The symptomatology of IM can include fever, oropharyngitis, malaise, lymphadenopathy, and hepatosplenomegaly. EBV infection is also associated with a broad range of human lymphoid, and epithelial cell malignancies such as Burkitt's, Hodgkin's, and nasal natural killer/T-cell lymphomas, post-transplantation lymphoproliferative disease, nasopharyngeal carcinoma (NPC), and gastric adenocarcinoma, among other conditions.

Nasopharyngeal carcinoma (NPC) is a rare human malignancy in most populations, and EBV infection has been identified as a group 1 carcinogenic agent by the International Agency for Research and Cancer. In Europe and USA, the incidence of NPC is below 1 per 100,000 persons per year. However, in endemic regions of Asia, the incidence rate is 30 per 100,000 persons per year [2]. In this carcinoma, latent EBV infection is characterized by EBNA1 driven by the Qp

✉ Ana V. Banko
ana.banko@med.bg.ac.rs

1 Virology Department, Institute of Microbiology and Immunology, Faculty of Medicine, University of Belgrade, Dr Subotica, Belgrade 11000, Serbia

2 Clinic of Otorhinolaryngology and Maxillofacial Surgery, Clinical Center of Serbia, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

3 Clinics of Infectious and Tropical Diseases, Clinical Center of Serbia, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

promoter, expression of EBV-encoded RNA (EBER) and BamHI A rightward transcripts, latent membrane protein 2 (LMP2), and variable expression of LMP1 [3]. Numerous unidentified genome specificities contribute to the pathogenesis of EBV-related carcinoma, particularly NPC.

Viral persistence is promoted through mechanisms that reduce antigen presentation to the adaptive immune system, including limited viral protein expression which is necessary for replication [4]. The restricted set of genes includes two EBV-encoded RNAs (EBER1 and EBER2), six EBV nuclear antigens (EBNA1, EBNA2, EBNA3A, EBNA3B, EBNA3C, and leader protein EBNA-LP), and three integral membrane proteins (LMP1, LMP2A, and LMP2B). The most notable expression in cells of all EBV-related tumors and latency programs is the expression of EBNA1. This protein is essential for viral replication, maintenance of extrachromosomal episomes, and transcriptional control of the viral latency programs through sequence-specific binding to its replication origin, OriP. Additionally, EBNA1 may play a critical role in the onset, progression, and persistency of EBV-associated tumors. Studies have shown that EBNA1 influences cellular proteins and signal pathways involved in cell survival and proliferation, such as inhibition of p53-dependent apoptosis in Burkitt's lymphoma [5].

The 641-amino acid EBNA1 protein is composed of unique amino-terminal (residues 1–89) and carboxyl-terminal (residues 327–641) domains linked by a large Gly-Ala repeat (residues 90–326). Most described mutations were identified in the carboxyl-terminal domain, which contains the dimerization, DNA binding, and transactivation domains [6]. In contrast, Gly-Ala repeats reduce the translational efficiency of EBNA1 and inhibit the initiation of translation, reducing the ability of EBV-specific $CD8^+$ T cells to recognize expressed EBNA1 peptides [3]. Sequence variability within EBNA1 is classified into five subtypes: two prototype sequences P-Ala (B95-8 prototype) and P-Thr, and three variant sequences V-Val, V-Leu, and V-Pro. The subtype V-Ala was subsequently identified [7]. Subtypes are defined by the amino acid at position 487, while subvariants are based on amino acid substitutions at loci other than locus 487 [8].

Diversities within the EBNA1 gene sequence have been widely explored by analyzing the unambiguous geographically specific distribution of EBNA1 subtypes. The role of these subtypes in tumor development is unclear, as the results of previous studies are controversial [9, 10]. This may be because of the small sample sizes, limited geographic regions and disease status, and control sample selection used in previous studies [7]. While EBNA1 sequence diversity has been suggested to contribute to different tissue tropisms of EBV and is associated with the development of certain malignancies, other studies showed that this diversity only reflects region-restricted polymorphisms [6]. However, the understanding of this diversity was sufficient for predicting viral epitope targets to determine a treatment strategy [11]. This approach described the therapeutic application of EBV-specific T-cell adoptive immunotherapy against EBV-related malignancies.

Serbia and other countries in southeast Europe are considered as non-endemic regions for most EBV-related malignancies. Previous reports from this region have revealed an association between EBV gene polymorphisms and the clinical characteristics of different EBV-associated diseases only for the LMP1 gene. EBNA1 sequencing studies of NPC isolates were performed previously [12]. Unknown sequence characteristics were identified, but it is necessary to evaluate their significance in nonmalignant EBV-related diseases. Thus, the aims of the present study were to further evaluate EBV EBNA1 gene polymorphisms circulating in IM and isolates from patients with transplanted organs to determine if EBNA1 sequence specificities can be used as viral biomarkers for NPC.

Materials and methods

Patients and samples

In this study, EBV EBNA1-positive plasma samples from 31 patients with mononucleosis syndrome treated at the Clinic of Infectious and Tropical Diseases, Clinical Center of Serbia, and plasma samples from 16 patients after renal transplantation from the Urology Clinic, Clinical Center of Serbia were evaluated. Furthermore, 40 EBV EBNA1-positive archived tissue blocks of histologically confirmed nasopharyngeal carcinoma (type UCNT, undifferentiated nasopharyngeal carcinoma) fixed in formalin and embedded in paraffin were retrieved for this study from the Clinic of Otorhinolaryngology and Maxillofacial Surgery, Clinical Center of Serbia.

This research was approved by the Ethics Committee of the Faculty of Medicine, University of Belgrade (no. 29/VI-12). Informed consent for participating in this study was obtained from all the IM patients and also the patients with transplanted organs. As the collection of biopsies was retrospective, both institutions waived the requirement for written informed consent from the donors (review board of the Clinic of Otorhinolaryngology and Maxillofacial Surgery, Clinical Center of Serbia and Ethics Committee of the Faculty of Medicine, University of Belgrade).

Deparaffinization and DNA isolation

Isolation of viral DNA was carried out using 200 µL plasma collected from IM patients and transplanted patients using the QIamp Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Biopsies from

UCNT patients were processed by deparaffinization with xylene, rehydration in alcohol, and then air-drying of three 10-mm-thick tissue sections from each block in a PCR tube. The tissue sections were resuspended and lysed overnight at 56 °C in 180 µL digestion buffer (Qiagen) and 20 µL proteinase K (Qiagen). Viral DNA was isolated using a QIAamp Mini Kit (Qiagen) according to the manufacturer's instructions. DNA isolates were further used for nested PCR to amplify the EBNA1 gene.

EBNA1 carboxy-terminal region sequencing

Amplification of the C terminus of the EBNA1 gene was performed by nested PCR using the primers reported by Lorenzetti et al. [13]. Both PCRs were carried out in 40 cycles. The first reaction was conducted at 95 °C for 1 min, 57 °C for 2 min, and 72 °C for 90 s; the second reaction was conducted at 95 °C for 1 min, 60 °C for 2 min, and 72 °C for 90 s. PCRs were prepared and handled based on the rules of good laboratory practice. After analyzing the PCR products by gel electrophoresis with ethidium bromide staining, the products were purified using a Qiagen MinElute Purification Kit according to the manufacturer's instructions. For cycle sequencing reactions, internal PCR primers and a Big Dye Terminator v 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) were used. Both sense and antisense strands were sequenced and compared.

Sequence and phylogenetic analysis

The 329-bp nucleotide sequences of EBNA1 were aligned and compared to a reference wild-type sequence using Bioedit 7.0.5.3 software [14]. Using the same software, signature amino acid changes were investigated at positions 471, 475, 476, 479, 487, 492, 499, 500, 502, 517, 520, 524, 525, 528, and 533 to identify and classify EBNA1 subtypes and subvariants. Ten sequences obtained from the GenBank/EMBL7DDBJ database under accession numbers V01555, GU475455, JN986939, AF192742, GU475448, AF192743, GU475431, AF192744, JN986947, and GU475442 were used as representative and reference EBNA1 sequences. Eighty-seven EBNA1 sequences from this study are available in the GenBank/EMBL7DDBJ database with accession numbers MH196405–MH196452 and KT820449–KT820488. Finally, all sequences were aligned pairwise using the ClustalW method implemented in MEGA 6.0 software [15]. The most appropriate model for evolution for the C-terminal region of EBNA1 gene was inferred using jModelTest 2.1.4 [16]. Maximum-likelihood trees were estimated according to the defined best-fit F81+I+G evolutionary model using PhyML 3.0 software [17]. Statistical significance of the phylogeny was estimated by bootstrap analysis with 1000 pseudo-replicate datasets. Graphs and phylogenetic trees were prepared using Fig Tree 1.4.0 [18] and MEGA 6.0 [15] software.

Statistical analyses

Chi-squared or Fisher's exact test and Student's *t* test were performed using SPSS v.21 for Windows (SPSS, Inc., Chicago, IL, USA) for statistical analysis. p value ≤ 0.05 was considered significant.

Results

Eighty-seven sequences of the EBV EBNA1 gene fragment (coordinates 109,261–109,590) were analyzed and compared with the B95-8 prototype strain. Based on amino acid substitutions and clustering of sequences in the phylogenetic tree, isolates were divided into four subtypes: two prototype subtypes, P-Ala and P-Thr, and two variant subtypes, V-Val and V-Ala (Fig. 1). Subtypes V-Pro and V-Leu were not identified. The most frequent subtype was P-Thr (59.8%). All P-Ala, 17 P-Thr, V-Val, and V-Ala isolates were further analyzed because of additional nucleotide variability and subtype-specific amino acid substitutions. This step covered subvariant characterization within the scope of each subtype. Each identified subvariant (sv) (P-Ala-sv-1 and -2; P-Thr-sv-1, -2, -3, -4, -5, and -6; V-Val-sv-1 and V-Ala-sv-1) showed characteristic amino acid changes and clustered separately in the phylogenetic tree (Table 1).

The distributions of the discovered EBNA1 subtypes and subvariants in the three groups of EBV-related diseases are shown in Table 2. There was no significant difference in the distribution of the four EBNA1 subtypes between the three EBV-related diseases ($p = 0.339$). The most common subtype in each defined group of patients was P-Thr, with frequencies of 61.3% in IM patients, 57.5% in NPC patients, and 62.5% in patients with transplanted organs. In contrast, we found a significant difference in the distribution of EBNA1 P-Thr subvariants between the three groups of patients, with noticeable clustering of P-Thr-sv-5 in the group of carcinoma isolates ($p < 0.001$). Three combinations of diseases were formed, IM/T, IM/NPC, and T/NPC for more detailed analysis of these results. Two significant differences in the distribution of EBNA1 P-Thr subvariants were revealed: one between IM and NPC isolates ($p < 0.001$) and the other between T and NPC isolates ($p = 0.013$). The potential association between specific EBNA1 subvariants and the year in which the biopsy was conducted was investigated, but a correlation could not be determined because the most prominent subvariants showed similar prevalence rates throughout the years 2008–2013.

The distribution of EBNA1 P-Ala subvariants did not significantly differ between the three groups of patients

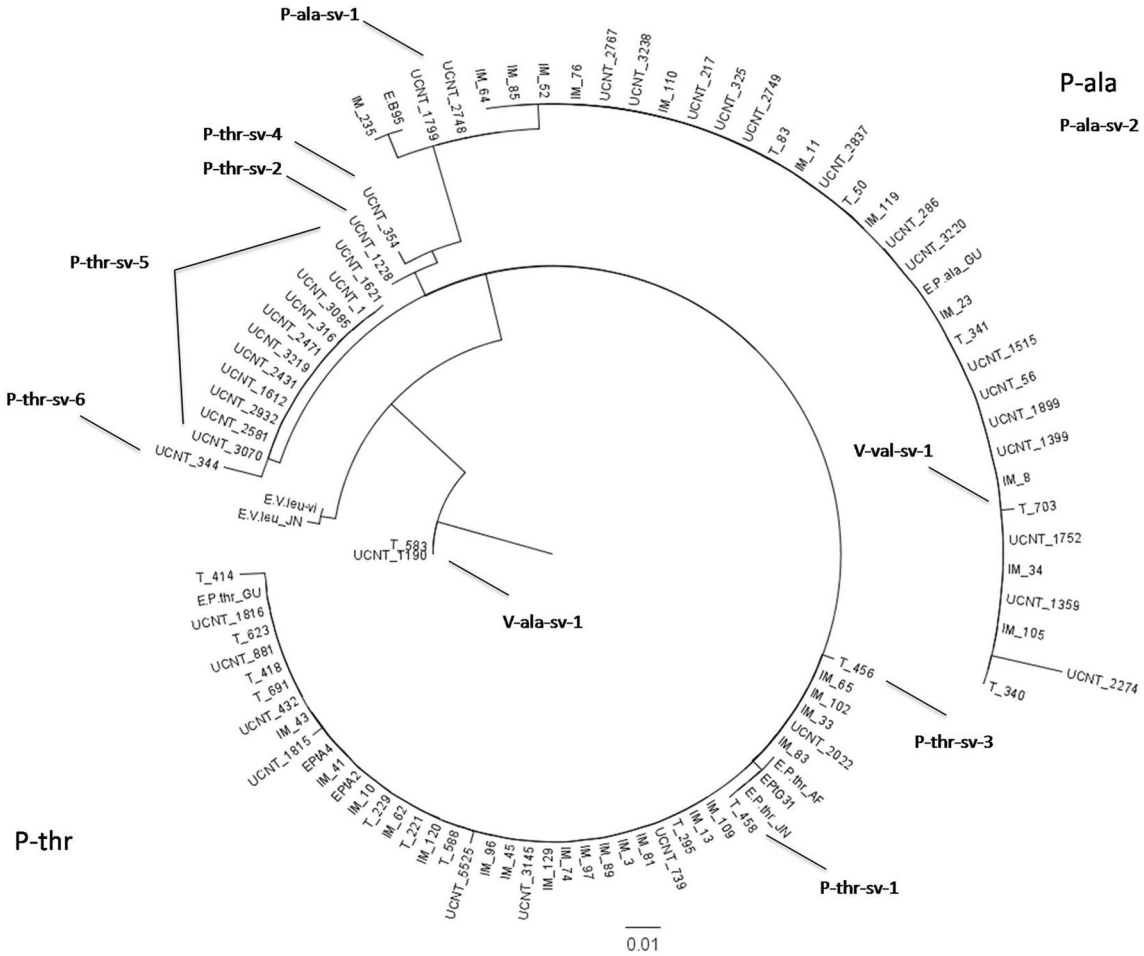

Fig. 1 Phylogenetic tree of the C-termini of EBNA1. Eighty-seven 329-bp fragments of EBNA1 (from coordinates 109,261–109,590) sequences from this research (40 undifferentiated nasopharyngeal carcinoma, UCNT available in the GenBank/EMBL7DDBJ database with accession numbers: KT820449–KT820488; 31 infective mononucleosis, IM available in the GenBank/EMBL7DDBJ database with accession numbers: MH196405–MH196434; 16 patients after transplantation, T available in the GenBank/EMBL7DDBJ database with accession numbers: MH196436–MH196452) and ten sequences obtained from GenBank/EMBL7DDBJ database under the following accession numbers: V01555, GU475455, JN986939, AF192742, GU475448, AF192743, GU475431, AF192744, JN986947 and GU475442

($p = 0.336$). Additionally, no correlation was observed between the characteristic EBNA1 variabilities and sex, age, specific anamnestic (geographical origin of NPC patients and smoking), clinical (hepatosplenomegaly in IM patients, comorbidity, and TNM stages of carcinoma in NPC patients), or biochemical (elevated level of liver enzymes in IM patients) data.

Discussion

EBNA1 gene variations in different EBV-related diseases have been extensively characterized, but the precise conclusions remain unclear. Geographical restriction of EBNA1 subtypes represents one possible theory. Thus, in gastric carcinoma and reactive follicular hyperplasia isolates, Chen et al. reported that V-Val dominated in Japan, while P-Ala, P-Thr, and V-Leu were the most frequent subtypes in North America [19]. Further reports reinforced the association with geographical origin. According to previous studies, the dominant subtypes in Europe and North America are P-Thr and P-Ala. Remaining isolates include V-Leu and rarely V-Val (only in AIDS lymphoma) [7].

Identification of EBNA1 subtypes with the dominance of P-Thr and absence of V-Pro in this study supported previously described European patterns, except for the first identification of V-Ala in European isolates [7]. Despite the sporadic findings of V-Leu in 2.5–8.2% of healthy patients or lymphoma and nonmalignant isolates in Europeans, this subtype was not detected in the present study [7, 9, 20].

Statistical analysis revealed no significant differences in the frequencies of EBNA1 subtypes in the three EBV-related

Table 1 EBNA1 C-terminal nucleotide and amino acid changes found in four subtypes and ten subvariants identified in this study

	B95-8[a] (P-Ala)	P-Ala-sv-1	P-Ala-sv-2	P-Thr	P-Thr-sv-1	P-Thr-sv-2[b]	P-Thr-sv-3	P-Thr-sv-4[b]	P-Thr-sv-5[b]	P-Thr-sv-6[b]	V-Val-sv-1[c]	V-Ala-sv-1
Locus												
471	CAA Gln				CAT His							
476	CCG Pro			CAG Gln	CAG Gln		CAG Gln		CAG Gln	CAG Gln		
483	GAA Glu							GAC Asp	GAC Asp	GAC Asp	GTT Val	GAC Asp
487	GCT Ala			ACT Thr	ACT Thr	ACT Thr	ACT Thr	ACT Thr	ACT Thr	ACT Thr	GTT Val	
492	AGT Ser			TGT Cys	TGT Cys	TGT Cys			TGT Cys	TGT Cys		
499	GAC Asp	GAA Glu	GAA Glu	GAT Asp	GAT Asp	GAT Asp	GAT Asp	GAT Asp	GAT Asp	GAT Asp	GAA Glu	GAG Glu
502	ACT Thr									ATT Ile		AAT Asn
520	CTA Leu			CTC Leu	CTC Leu	CTC Leu	CTC Leu		CTC Leu	CTC Leu	CTC Leu	CTC Leu
524	ACT Thr	ATT Ile	GTT Val	ATT Ile	ATT Ile	ATT Ile	ATT Ile		ATT Ile	ATT Ile	GTT Val	ATT Ile
529	CCA Pro											CAA Gln
533	CTT Leu										GTT Val	
Number of isolates	–	2	30	35	1	1	1	1	12	1	1	2
Total	32 (36.8%)			52 (59.8%)							1 (1.1%)	2 (2.3%)

[a]Prototype sequence (represents the P-Ala subtype)

[b]Unique subvariants for this region: P-Thr-sv-2, P-Thr-sv-4, P-Thr-sv-5 and P-Thr-sv-6

[c]New subvariant: V-Val-sv-1

Table 2 Distribution of different EBNA1 subtypes and subvariants in three groups of EBV-related diseases

EBNA-1 subtype	EBNA-1 subvariant	Groups of EBV-related diseases			Total
		Infective mononucleosis	Transplantation	NPC	
P-Thr	P-Thr	19	8	8	35
	P-Thr-sv-1	–	1	–	1
	P-Thr-sv-2	–	–	1	1
	P-Thr-sv-3	–	1	–	1
	P-Thr-sv-4	–	–	1	1
	P-Thr-sv-5	–	–	12	12
	P-Thr-sv-6	–	–	1	1
P-Ala	P-Ala-sv-1	–	–	2	2
	P-Ala-sv-2	12	4	14	30
V-Val	V-Val-sv-1	–	1	–	1
V-Ala	V-Ala-sv-1	–	1	1	2
Number of isolates		31	16	40	87

diseases. The most frequent subtype, P-Thr, was disease-independent (approximately 60%), as shown in Danish isolates [9]. However, the frequency of P-Thr in this study was lower (60% vs. 72–80%), while the frequency of P-Ala was higher (40% vs. 20%) than in previous European reports [7]. In addition to the similar subtype distributions between patients with different diagnoses, variability in the IM isolates showed the exclusion of variant subtypes (V-Val and V-Ala). This supports an earlier hypothesis that in Europe and North America, prototype types of EBNA1 are found only in nonmalignant isolates [21].

Because 17 P-Thr sequences from this research appeared to differ from the reference sequence, the spectrum of atypical mutations was further analyzed. According to a previous EBNA1 subvariant characterization, 17 P-Thr isolates were identified as 6 subvariants (sv-1 to sv-6), representing substantial variability. Of the six subvariants, P-Thr-sv-1 and P-Thr-sv-3 have been described in isolates from NPC, lymphoma, and mononucleosis [8, 13, 22]. However, based on previous reports, P-Thr-sv-1 and P-Thr-sv-3 have never been detected in Europe. Thus, detection of these subvariants in isolates from transplanted patients represents their newly discovered European distribution. Notably, we identified and characterized four new P-Thr subvariants (sv-2, sv-4, sv-5, and sv-6), which were described previously by the same group [12]. These four subvariants showed additional amino acid substitutions and dissociated in the phylogenetic tree constructed from all sequences in this research. Interestingly, all four subvariants were from the group of UCNT isolates with a prominent quantum of P-Thr-sv-5, resulting in significant differences in the distribution of P-Thr subvariants between different diseases. Therefore, the most crucial result is that this new P-Thr subvariant originated from 12 isolates of the same type of tissue origin and was absent from any other type of sample or disease obtained in this study. Additionally, no studies have revealed a consistent group of new EBNA1 sequences for more than a few isolates. Thus, P-Thr-sv-5 should be recognized as a UCNT-specific EBNA1 subvariant. Because of the lack of support for functional associations between specific EBNA1 sequences and the course of pathogenesis, the results of this study can provide a foundation for further in vitro investigations of specific molecular processes in host cells infected with P-Thr-sv-5. As suggested previously, like the influence of the V-Val variability on NPC oncogenesis [23] or association between EBV gene variability and environmental factors together with genetic predisposition of the infected host [24], P-Thr-sv-5 together with genetic predispositions have a multifactorial influence on the pathogenesis of NPC. Moreover, P-Thr-sv-5 should be included in screening and diagnostic protocols as a specific marker of UCNT evolution.

P-Ala represents the prototype sequence. Although this subtype was detected in more than one-third of isolates in this study, no sequences were identical to the prototype. Therefore, 32 sequences were divided into two groups of disease-independent subvariants (P-Ala-sv-1 and sv-2) described previously in Danish NPC, lymphoma, and healthy controls [9]. P-Ala-sv-1 and sv-2 correspond to European specificities within the scope of the P-Ala subtype and are independent of disease origin.

Although in this study, V-Ala was present only in one patient after transplantation and one UCNT patient, this subtype is rare even in South America, which is the only continent in which V-Ala was discovered. Described frequencies were between 6% in Hodgkin lymphoma from Brazil and 15% in pediatric IM and lymphoma patients in Argentina [13, 20]. However, interestingly, V-Ala sequences in this study were the same as that of one Argentinian sequence from a Hodgkin lymphoma isolate [13]. Thus, V-Ala-sv-1 represents an identical subvariant with Argentinian V-Ala-iii. Given the lack of

data regarding the residence, traveling, or ethnic associations between these three patients, it is not possible to evaluate the geographic specificity of this very rare V-Ala subvariant.

V-Val was found only in one isolate from a patient after transplantation. According to additional mutations (particularly at position 483), the sequence was defined as V-Val-sv-1, which also represents a newly discovered subvariant. V-Val is dominant in Asia, and V-Val was detected in Europe in 2% of isolates from AIDS-related lymphoma [7]. Although the number of V-Val isolates in this study was small, this variant may be associated with an immunocompromised status to enable survival of the virus with EBNA1 V-Val variability.

Based on our results, EBNA1 variability is clearly not determined only by geographic origin. Moreover, these findings support that a critical association exists between EBNA1 variability and specific pathogenesis. EBV EBNA1 showed no sequence specificity in primary infection. Pathogenetic mechanisms, in correlation with environmental and host characteristics, may alter the viral sequence structure. Thus, P-Thr-sv-5 in this study should be recognized as a UCNT-specific subvariant, as it is represented in the carcinoma-specific EBNA1 variability. Additionally, this research revealed a newly discovered EBNA1 subvariant. All described associations require additional investigations to elucidate the molecular mechanisms between specific EBNA1 subvariants and disease or ethnic characteristics of infected host cells.

Acknowledgements The authors are grateful to the laboratory technicians Gabrijela Pavlovic and Marija Zivotic from the Virology Department, Institute of Microbiology and Immunology, Faculty of Medicine, University of Belgrade for their technical assistance.

Funding This study was supported by the Ministry of Education, Science and Technological Development, Republic of Serbia, Research Grant no. 175073.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in study involving human participants were in accordance with the ethical standards of the Ethics Committee of Faculty of Medicine, University of Belgrade no. 29/VI-12 and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For retrospective part of this study, formal consent was not reqired.

References

1. Rickinson AB, Kieff ED (2007) Epstein–Barr virus and its replication. In: Knipe D, Howlry P (eds) Fields virology, 5th edn. Lippincott Williams & Wilkins, Philadelphia, pp 2603–2654
2. Jia WH, Qin HD (2012) Non-viral environmental risk factors for nasopharyngeal carcinoma: a systematic review. Semin Cancer Biol 22:117–126. https://doi.org/10.1016/j.semcancer.2012.01.009
3. Rickinson AB, Kieff ED (2013) Epstein–Barr virus. In: Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA et al (ed) Fields virology, 6th edn. Lippincott Williams & Wilkins, Philadelphia, pp 1898–1959
4. Klenerman P, Hill A (2005) T cells and viral persistence: lessons from diverse infections. Nat Immunol 6:873–879. https://doi.org/10.1038/ni1241
5. Kennedy G, Komano J, Sugden B (2003) Epstein–Barr virus provides a survival factor to Burkitt's lymphomas. Proc Natl Acad Sci USA 100:14269–14274. https://doi.org/10.1073/pnas.2336099100
6. Sun L, Zhao Z, Liu S, Liu X, Sun Z, Luo B (2015) Sequence variation analysis of Epstein–Barr virus nuclear antigen 1 gene in the virus associated lymphomas of northern China. PLoS One e0140529. https://doi.org/10.1371/journal.pone.0140529
7. Chang CM, Yu KJ, Mbulaiteye SM, Hildesheim A, Bhatia K (2009) The extent of genetic diversity of Epstein–Barr virus and its geographic and disease patterns: a need for reappraisal. Virus Res 143:209–221. https://doi.org/10.1016/j.virusres.2009.07.005
8. Bhatia K, Raj A, Guitierrez MI, Judde JG, Spangler G, Venkatesh H et al (1996) Variation in the sequence of Epstein–Barr virus nuclear antigen 1 in normal peripheal blood lymphocytes and in Burkitt's lymphomas. Oncogene 13:177–181
9. Sandvej K, Zhou XG, Hamilton-Dutoit S (2000) EBNA-1 sequence variation in Danish and Chinese EBV-associated tumors: evidence for geographical polymorphism but not for tumor-specific subtype restriction. J Pathol 191:127–131. https://doi.org/10.1002/(SICI)1096-9896(200006)191:2%3C127::AID-PATH614%3E3.0.CO;2-E
10. Puchhammer-Stockl E, Gorzer I (2006) Cytomegalovirus and Epstein–Barr virus subtypes—the search for clinical significance. J Clin Virol 36:239–248. https://doi.org/10.1016/j.jcv.2006.03.004
11. Tschochner M, Leary S, Cooper D, Strautins K, Chopra A, Clark H et al (2016) Identifying patient-specific epstein–barr nuclear antigen-1 genetic variation and potential autoreactive targets relevant to multiple sclerosis pathogenesis. PLoS One. https://doi.org/10.1371/journal.pone.0147567
12. Banko AV, Lazarevic IB, Folic MM, Djukic VB, Cirkovic AM, Karalic DZ et al (2016) Characterization of the variability of Epstein–Barr virus genes in nasopharyngeal biopsies: potential predictors for carcinoma progression. PLoS One. https://doi.org/10.1371/journal.pone.0153498
13. Lorenzetti MA, Altcheh J, Moroni S, Moscatelli G, Chabay PA, Preciado MV (2010) EBNA1 sequences in Argentinean pediatric acute and latent Epstein–Barr virus infection reflect circulation of novel South American variants. J Med Virol 82:1730–1738. https://doi.org/10.1002/jmv.21871
14. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95, 98, NT, 2000, XP. Nucl Acids Symp 41:95–98
15. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197
16. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772. https://doi.org/10.1038/nmeth.2109
17. Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704
18. Rambaut A (2007) Molecular evolution, phylogenetics and epidemiology. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 27 Nov 2015

19. Chen YY, Chang KL, Chen WG, Shibada D, Hayashi K, Weiss LM (1998) Epstein–Barr virus-associated nuclear antigen-1 carboxy-terminal gene sequences in Japanese and American patients with gastric carcinoma. Lab Investig 78:877–882
20. MacKenzie J, Gray D, Pinto-Paes R, Barrezueta LF, Armstrong AA, Alexander FA et al (1999) Analysis of Epstein–Barr virus (EBV) nuclear antigen 1 subtypes in EBV-associated lymphomas from Brasil and the United Kingdom. J Gen Virol 80:2741–2745. https://doi.org/10.1099/0022-1317-80-10-2741
21. Gutierrez MI, Raj A, Spangler G, Sharma A, Hussain A, Judde JG et al (1997) Sequence variations in EBNA-1 may dictate restriction of tissue distribution of Epstein–Barr virus in normal and tumour cells. J Gen Virol 78:1663–1670. https://doi.org/10.1099/0022-1317-78-7-1663
22. Fassone L, Bhatia K, Gutierrez M, Capello D, Gloghini A, Dolcetti R et al (2000) Molecular profile of Epstein–Barr virus infection in HHV-8-positive primary effusion lymphoma. Leukemia 14:271–277. https://doi.org/10.1038/sj.leu.2401651
23. Wang JT, Sheeng TS, Su IJ, Chen JY, Chen MR (2003) EBNA-1 sequence variations reflect active EBV replication and disease status or quiescent latency in lymphocytes. J Med Virol 69:417–425. https://doi.org/10.1002/jmv.10305
24. Chen JN, Zhang NN, Jiang Y, Hui DY, Wen ZJ, Li HG et al (2012) Variations of Epstein–Barr virus nuclear antigen 1 in Epstein–Barr virus-associated gastric carcinomas from Guangzhou, southern China. PLoS One. https://doi.org/10.1371/journal.pone.0050084