ORIGINAL ARTICLE VIROLOGY

Epstein–Barr virus BZLF1 gene polymorphisms: malignancy related or geographically distributed variants?

M. A. Lorenzetti[1], M. Gantuz[1], J. Altcheh[2], E. De Matteo[3], P. A. Chabay[1] and M. V. Preciado[1]

1) *Molecular Biology Laboratory, Pathology Division, Ricardo Gutiérrez Children Hospital,* 2) *Parasitology and Chagas Disease Laboratory, Ricardo Gutiérrez Children Hospital and* 3) *Pathology Division, Ricardo Gutiérrez Children Hospital, Buenos Aires, Argentina*

Abstract

The ubiquitous Epstein–Barr virus (EBV) is related to the development of several lymphoid and epithelial malignancies and is also the aetiological agent for infectious mononucleosis (IM). BZLF1, an immediate early gene, plays a key role in modulating the switch from latency to lytic replication, hence enabling viral propagation. Polymorphic variations in the coded protein have been studied in other geographical regions in a search for viral factors that are inherent to malignancies and differ from those present in benign infections. In the present study, in samples of paediatric patients with benign IM and paediatric patients with malignant lymphomas, we detected previously described sequence variations as well as distinctive sequence polymorphisms from our region. By means of phylogenetic reconstruction, we characterized new phylogenetically distinct variants. Moreover, we described an association between specific variants and the studied pathologies in our region, particularly variant BZLF1-*A2* with lymphomas and BZLF1-*C* with IM. Additionally, length polymorphisms within intron 1 were also assessed and compared between pathologies resulting in an association between 29-bp repeated units and lymphomas. In conclusion, this is the first report to characterize BZLF1 gene polymorphisms in paediatric patients from our geographical region and to suggest the association of these polymorphisms with malignant lymphomas.

Keywords: BZLF1 variants, Epstein–Barr virus variation, infectious mononucleosis, paediatric lymphomas, Zta variants
Original Submission: 9 January 2014; **Revised Submission:** 21 March 2014; **Accepted:** 21 March 2014
Editor: G. Antonelli
Article published online: 25 March 2014
Clin Microbiol Infect 2014; **20**: O861–O869
10.1111/1469-0691.12631

Corresponding author: M. A. Lorenzetti, Laboratorio de Biología Molecular, División Patología, Hospital de Niños Ricardo Gutiérrez, Gallo 1330, C1425EFD Buenos Aires, Argentina
E-mail: marioloren@yahoo.com.ar

GenBank accession numbers to the sequences in this manuscript are KF826537–KF826610.

Introduction

The Epstein–Barr virus (EBV) is a ubiquitous herpesvirus that latently infects over 90% of the world's population. In developing regions, primary infection usually occurs during early childhood and is not associated with severe clinical symptoms, although mild cases of infectious mononucleosis (IM) may occur. On the other hand, in developed regions, where primary infection is usually delayed until adolescence or early childhood, it can cause a more severe case of IM [1].

Epstein–Barr virus is transmitted among individuals through saliva, and following primary infection, latent infection is established in memory B cells. Although latency supposes no risk for the immunocompetent individual, it is this type of infection that is usually associated with neoplastic pathologies [2]; but the involvement of EBV in the aetiology, progression and/or outcome of these malignancies is not yet fully understood. Association of EBV with these pathologies varies dramatically in different geographic regions (extensively reviewed in [3]). Even though differences in viral association may be accounted for by other factors, there has long been interest in the possibility that disease-specific genetic variation in EBV in different parts of the world might account for them [4].

Based on polymorphisms in nuclear antigens EBNA3A, -3B and -3C, two different EBV types, EBV1 and EBV2, can be distinguished. Given that these polymorphisms are not

enough to describe the entire EBV natural variation, variants and subvariants have been described based on polymorphisms in other viral antigens, such as BZLF1, EBNA1 and LMP1 [3,4].

Particularly, the BZLF1 gene encodes for an immediate early protein termed Zta, also called ZEBRA or EB1, which is a master regulator for the expression of several early viral genes that are critical for productive replication of EBV. In this way Zta is responsible for the switch between latent and lytic infection because it can initiate the entire lytic cascade [5]. One of the earliest and most critical functions of Zta is to activate its own expression (acting in *cis* over its own promoter) and the expression of Rta, the product of the BRLF1 gene [6,7]. Zta also mediates viral DNA replication by specific binding to the origin of lytic cycle replication (oriLyt) and recruiting viral proteins [8].

The BZLF1 gene, which consists of three exons and two introns with variable tandem repeats in the first intron [3], renders a 245 amino acid (aa) protein [8].

Though the functions of Zta are well established, there is still very little information about naturally occurring polymorphisms present in different clinical samples from different geographic regions [4]. To date, few studies have assessed Zta polymorphisms, but mainly refer to EBV-associated gastric carcinoma or nasopharyngeal carcinoma from Asia or Africa [8–10].

As the incidence of EBV-associated malignancies differs between geographical regions and EBV strains from different regions or populations differ in their gene sequences, the existence of tumour-specific variants is still a matter of debate. Moreover, given the lack of information about natural variation in the Zta protein from other geographical regions, definitive conclusions regarding disease or geographical associations are hampered.

In the present study, BZLF1 gene sequence variations from our geographical region were characterized in peripheral blood mononuclear cells (PBMCs) and oral secretions (OS) from paediatric patients with IM as representative of a replicative yet benign condition and in EBV^+ lymphoma biopsies as representative of latent infection in a malignant condition.

Materials and Methods

Patients and samples

This study was conducted on 51 paediatric patients from Argentina. Twenty-three children with IM, a median age of 4 years (range 1–17 years) and 52% male. Twenty-eight patients had EBV-positive lymphomas (23 Hodgkin and 5 non-Hodgkin), median age 9 years (range 3–18 years) and 77% male. Hospital ethics committees reviewed and approved this study, which is in accordance with the human experimentation guidelines of our institution. Written informed consent was obtained from all patients' parents.

Lymph node biopsy specimens were collected for lymphoma diagnosis before therapy. Presence of EBV was assessed on formalin-fixed, paraffin-embedded tissue sections by *in situ* hybridization for EBV-encoded RNAs according to the manufacturer's instructions (Dako, Carpinteria, CA, USA). Those patients who showed positive nuclear staining in tumour cells without staining in infiltrating lymphocytes were included.

Peripheral blood (6 mL) and OS samples were obtained from patients with presumptive acute IM at the time of diagnosis. Blood samples were obtained by venopuncture and OS samples were obtained by pharyngeal swabbing. The IM was identified on clinical grounds and confirmed by indirect immunofluorescence assay, and those patients with IgM and with or without IgG antibodies against virus capsid antigen were included in the study.

DNA extraction

EDTA-PBMCs were separated from whole blood (6 mL) with Ficoll-Paque plus (GE Healthcare, Little Chalfont, UK). Genomic DNA was extracted from PBMCs, OS samples and lymph node biopsy specimens using a QIAamp DNA minikit (Qiagen, Hilden, Germany).

Epstein–Barr virus typing

The EBV type was determined based on polymorphisms in EBNA3C gene as previously described [11].

Complete BZLF1 gene amplification and sequencing

The BZLF1 gene was amplified by nested PCR of each sample. Primers used in the first round were 5′-tcaaagagagccgacaggaagata-3′ (positions 102127–102150 of prototype EBV genome strain B95.8, GenBank accession number V01555.2) and 5′-cattgcaccttgccggccacc-3′ (103175–103195). PCR was performed in 25 μL using 200 ng of genomic DNA, 1× amplification buffer, 1 mM $MgSO_4$, 0.2 mM dNTPs, 0.6 μM of each primer, and 0.375 U Platinum pfx DNA polymerase, with proof-reading activity (Invitrogen, Carlsbad, CA, USA). One microlitre of this product was re-amplified with primers 5′-aaacgaggcgtgaagcaggcgtgg-3′ (102171–102194) and 5′-ccacctttgctatctttgctg-3′ (103159–103179), yielding a 1008-bp fragment for B95.8 EBV. The reaction was performed under the same conditions as those for the first round. These sets of primers suppose no selection of a particular BZLF1

variant against any other because the targeted sequence is conserved. P3HR1 cell line DNA was used as a positive control and water was used as a non-template control. PCR products were subjected to electrophoresis in 2% agarose gel and purified with a QIAEXII gel extraction kit (Qiagen). These purified PCR products were directly sequenced using a BigDye Terminator 3.1 kit (Applied Biosystems, Foster City, CA, USA) in an automated Genetic Analyzer 3130xl (Applied Biosystems). At least two independent sequencing reactions (sense and anti-sense) were performed with the inner primers to confirm each sequence.

Sequence analysis

The complete BZLF1 sequence was amplified and both intronic sequences were then removed. Sequences were aligned, translated and analysed with BIOEDIT v7.0.9.0 [12]. According to the pattern of amino acid substitutions detected, Zta sequence variants were classified according to the scheme introduced by Luo *et al.* [8], because this scheme best fitted our data since it was the only one to consider the deletion of aa 127, also found in our sequences. Briefly, group A consists of two variants (BZLF1-A1 and A2), which can be distinguished based on specific signature amino acid substitutions compared with prototype B95.8; Group B (B1, B2, B3, B4, B5) is defined by the deletion of codon 127 (alanine deletion) and other signature amino acid substitutions. Finally, BZLF1-C variant is conserved and equal to prototype B95.8 and also to the new reference sequence for EBV genome (NC_007605.1). For phylogenetic reconstruction, the most appropriate model of evolution for this 245-aa protein was inferred and the phylogenetic tree was calculated using maximum likelihood under the previously defined evolutionary model (Jones–Taylor–Thomson model with discrete γ distributions with five rate categories, JTT + G for Fig. 1a and JTT for Fig. 1b,c) using MEGA 5 [13]. Zta protein sequence from Macacine herpesvirus 4 (RHV4, YP_067969.1), a related lymphocryptovirus from monkeys, was used as an out-group. Bootstrap values >70 were considered to merit as valid support for a tree branch.

Multi-region gene phylogeny was also inferred (promoter region variants and coding region variants) from previously characterized variants in the promoter region of the BZLF1 gene [11]. To test statistically for benign versus malignant association of these variants, the association index statistic, parsimony score and monophyletic clade were computed using a Bayesian tip-association significance testing (BaTS) program [14]. The BaTS program examines a posterior sample of trees generated by a Bayesian Markov Chain Monte Carlo approach implemented in BEAST v1.6.2 (Bayesian Evolutionary Analysis Sampling Trees) [15].

Sequence accession numbers

All sequences have been deposited in GenBank under accession numbers KF826537–KF826610.

Statistical analysis

Statistical analysis was performed using GRAPH-PAD INSTAT software, version 3.05 (Graphpad, La Jolla, CA, USA). Fisher's exact test was used to assess the association between categorical variables. All tests were two sided, and $p < 0.05$ was considered statistically significant.

Results

DNA could be purified and amplified from all samples included in the study.

EBV1 was the predominant viral type in both pathologies, in 22/23 patients with IM and 19/28 patients with lymphomas (Tables 1 and 2). There were no differences in the viral type detected in OS samples or PBMCs in patients with IM. Unfortunately, due to the institution's ethics committee guidelines, PBMCs and OS samples were not available from lymphoma patients.

Under the classification scheme introduced by Luo *et al.* [8], BZLF1-A2 variant was detected in one IM sample and ten lymphoma samples. BZLF1-B4 variant was detected in two IM samples and four tumour samples and BZLF1-C variant in 15 IM samples and five lymphoma samples (Tables 1 and 2). Additionally, in every individual patient with IM, sequences retrieved from PBMCs and OS from the same patient possessed 100% identity between each other.

On the other hand, although with lesser frequency, sequences with specific and not previously described mutations were observed in more than one sample (Fig. 1). Hence, sequences from samples IM15, T4 and T8 clustered separately with a bootstrap support of 97 (Fig. 2a). Given that these sequences differed significantly from BZLF1-A, -B or -C variants, they were characterized as a new phylogenetic group termed BZLF1-D. This variant was characterized by several substitutions, namely a tyrosine for an alanine at position 68 (T68A), S76P, Q105L, V129M, V141A, K161R and A206T together with the absence of aa 127 deletion (Fig. 1). Furthermore, this newly defined variant was still found to cluster independently from all other variants with no deletion of aa 127 (Fig. 2b).

In a similar way, patients IM5, IM8, T13, T27 and T28, all harboured substitutions L168V, A204S and the deletion of aa 127, characteristic of BZLF1-B variants, but without a significant bootstrap support to exclude them from the entire B group. As the bootstrap value for this cluster was

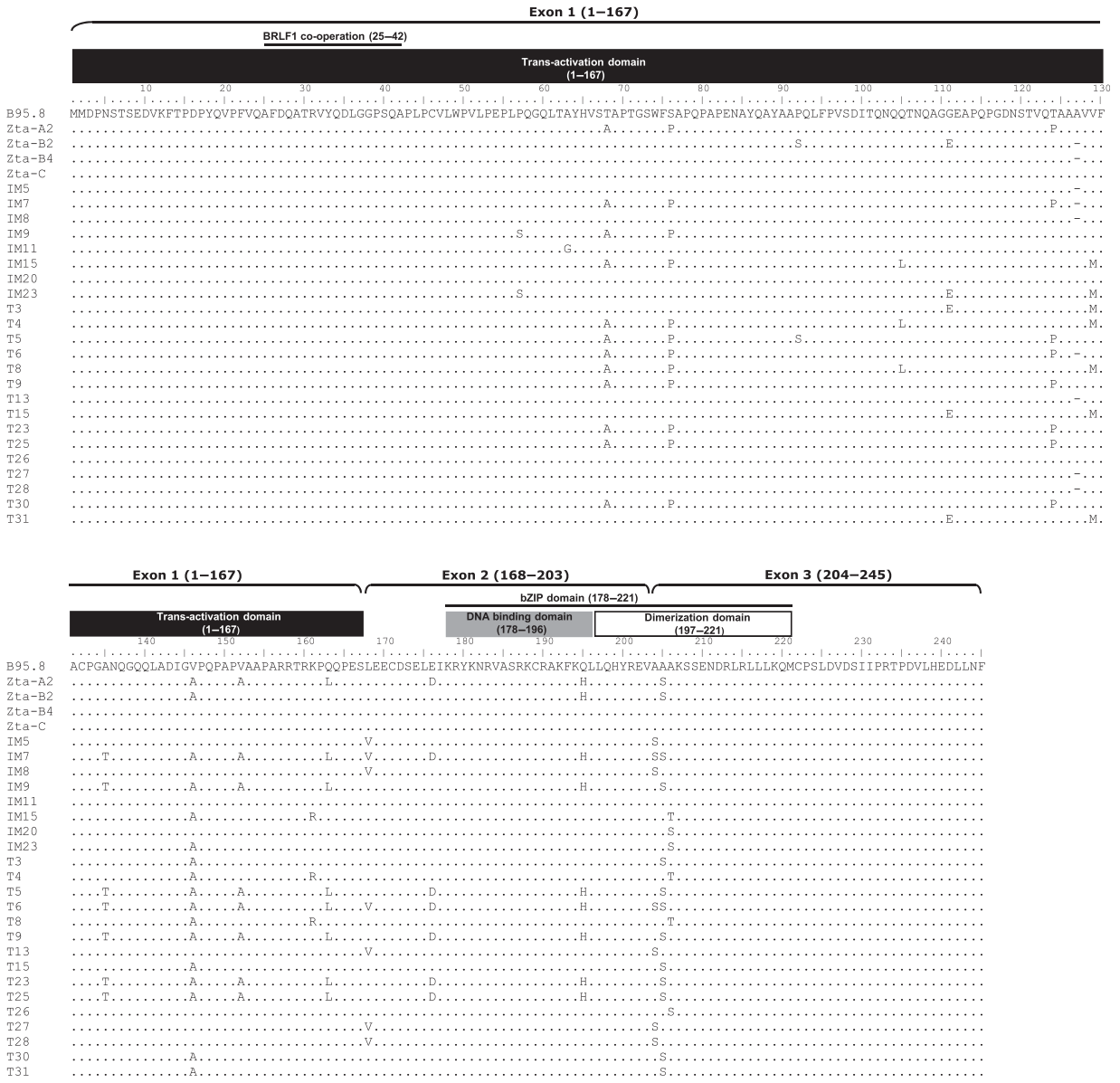

FIG. 1. Schematic figure showing the complete Zta protein sequence alignment obtained from clinical samples and variant reference sequences. IM indicates infectious mononucleosis samples and T indicates tumour samples. B95.8, Zta-A2, Zta-B2, Zta-B4 and Zta-C are the reference sequences. Black, grey and white boxes delimit Zta antigen domains and brackets delimit the three coding exons in BZLF1 gene. Ruler shows amino acid position. Only sequences with additional mutations to those described in the reference sequence are shown in the alignment.

87, they were characterized as a new BZLF1-B subgroup termed BZLF1-B6 (Fig. 2a). When the phylogenetic reconstruction was inferred taking only into account those variants with the deletion of aa 127, BZLF1-B6 sequences also segregated within the BZLF1-B group, but in an independent manner (Fig. 2c).

Sequence variation in EBV proved to be even greater than expected; variants with a mixed pattern of polymorphisms were also detected. Even though sequences from patients IM23, T3, T15 and T31 were very similar to BZLF1-B2 and clustered in close relation in the original tree (Fig. 2a), none of them contained the deletion of aa 127. Moreover, when these sequences were tested against all other sequences without the deletion (groups A and D), they clustered independently and were therefore characterized as new BZLF1-E variants (Fig. 2b). This new group harboured G111E, V129M and V146A; however, they also displayed minor differences. While samples from patients T3, T15 and T31 also contained a A205S substitution, sample IM23 did not, but contained P57S and A206S (Fig. 1).

TABLE 1. Epstein–Barr virus (EBV) type and Zta variant distribution in patients with infectious mononucleosis (IM)

Patient code	Sample	EBV type	Polymorphisms in Zta: No repeats in intron 1	Polymorphisms in Zta: aa 127 deletion	Polymorphisms in Zta: Variant
IM1	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM2	PBMC/OS	1	2 rep 30 bp	Yes	BZLF1-B4
IM3	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM4	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM5	PBMC/OS	1	2 rep 30 bp	Yes	BZLF1-B6
IM6	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM7	PBMC/OS	2	3 rep 29 bp	Yes	BZLF1-F
IM8	PBMC/OS	1	2 rep 30 bp	Yes	BZLF1-B6
IM9	PBMC/OS	1	4 rep 29 bp	No	BZLF1-A2[a]
IM10	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM11	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C[a]
IM12	PBMC/OS	1	4 rep 30 bp	No	BZLF1-C
IM13	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM14	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM15	PBMC/OS	1	10 rep 30 bp	No	BZLF1-D
IM16	PBMC/OS	1	3 rep 30 bp	Yes	BZLF1-B4
IM17	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM18	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM19	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM20	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C[a]
IM21	PBMC/OS	1	3 rep 30 bp	No	BZLF1-C
IM22	PBMC/OS	1	2 rep 30 bp	No	BZLF1-C
IM23	PBMC/OS	1	3 rep 30 bp	No	BZLF1-E

Abbreviations: aa, amino acid; bp, base pairs; OS, oral secretions; PBMC, peripheral blood mononuclear cells; rep, repeat.
GenBank Accession numbers to the sequences in the table are KF826537–KF826582.
[a]Variants with additional substitutions with respect to the reference variant.

TABLE 2. Epstein–Barr virus (EBV) type and Zta variant distribution in lymphoma patients

Patient code	Lymphoma type	EBV Type	Polymorphisms in Zta: No repeats in intron 1	Polymorphisms in Zta: aa 127 deletion	Polymorphisms in Zta: Variant
T1	NHL	2	3 rep 29 bp	No	BZLF1-A2
T2	NHL	2	4 rep 29 bp	No	BZLF1-A2
T3	NHL	1	3 rep 30 bp	No	BZLF1-E
T4	HL	1	8 rep 30 bp	No	BZLF1-D
T5	HL	2	2 rep 29 bp	No	BZLF1-A2[a]
T6	HL	2	3 rep 29 bp	Yes	BZLF1-F
T7	HL	1	2 rep 30 bp	No	BZLF1-C
T8	HL	1	8 rep 30 bp	No	BZLF1-D
T9	HL	2	2 rep 29 bp	No	BZLF1-A2[a]
T10	HL	1	2 rep 30 bp	Yes	BZLF1-B4
T11	HL	1	2 rep 30 bp	No	BZLF1-C
T12	HL	1	2 rep 30 bp	Yes	BZLF1-B4
T13	HL	1	2 rep 30 bp	Yes	BZLF1-B6
T14	HL	1	2 rep 30 bp	Yes	BZLF1-B4
T15	NHL	1	3 rep 30 bp	No	BZLF1-E
T16	NHL	2	3 rep 29 bp	No	BZLF1-A2
T18	HL	1	2 rep 29 bp	No	BZLF1-A2
T19	HL	1	2 rep 30 bp	No	BZLF1-C
T23	HL	2	2 rep 29 bp	No	BZLF1-A2[a]
T24	HL	1	3 rep 29 bp	No	BZLF1-C
T25	HL	2	2 rep 29 bp	No	BZLF1-A2[a]
T26	HL	1	2 rep 30 bp	No	BZLF1-C[a]
T27	HL	1	2 rep 30 bp	Yes	BZLF1-B6
T28	HL	1	2 rep 30 bp	Yes	BZLF1-B6
T29	HL	1	2 rep 30 bp	Yes	BZLF1-B4
T30	HL	1	4 rep 30 bp	No	BZLF1-A2[a]
T31	HL	1	3 rep 30 bp	No	BZLF1-E
T32	HL	2	3 rep 29 bp	No	BZLF1-A2[a]

Abbreviations: aa, amino acid; bp, base pairs; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; rep, repeat.
All samples are lymph node biopsies.
GenBank Accession numbers to the sequences in the table are KF826583–KF826610.
[a]Variants with additional substitutions with respect to the reference variant.

Finally, another set of samples, namely samples IM7 and T6, presented a mixed substitution pattern containing all signature substitutions of variant BZLF-A2 but also the deletion of aa 127 along with A135T and A204S substitutions (Fig. 1). Sequences from these samples clustered together with A2 variants in the original tree, but segregated as an independent group among all variants with deleted aa 127 (Fig. 2a,c). Bootstrap of 100 supports the characterization of a new group of BZLF1 variants, termed BZLF1-F, which also presented aa 127 deletion.

It is worth mentioning that samples from patients IM9, IM11 and IM20 and T5, T9, T23, T25, T26 and T30 contained additional mutations to those originally described in the reference variants. The most frequent were V129M, A135T, K161R, L168V and A204S (Fig. 1). Even so, they did not merit being characterized as new subvariants because no reproducible patterns of substitution were observed (Fig. 1). Moreover, bootstrap values obtained after the phylogenetic reconstruction did not support, in any of these cases, their classification as different variants (Fig. 2a).

Regarding variant association with malignancy, variants BZLF1-B4, -D, -E and -F were evenly distributed in both pathologies ($p > 0.05$ in all cases). On the other hand, while BZLF1-C variant preferentially occurred in IM (p 0.0012), BZLF1-A2 variant was statistically associated with its occurrence in lymphomas (p 0.0134). Given that some degree of variation was observed within different samples with BZLF1-A2 variant, and that certain substitutions are common to other variants (e.g. V146A), tumour association for each substitution detected in this series was assessed independently. Curiously, only those substitutions described in the original BZLF1-A2 variant showed a statistical association with lymphomas and are shown in Table 3. Additionally, a multi-region gene phylogenetic approach was used to assess the association of variants in the coding region of Zta and previously described variants in its promoter region [11] with lymphomas. Interestingly, viral genomes containing the BZLF1-A2 variant in the coding region and Zp-V3 variant in the promoter region of Zta were statistically associated with lymphomas (p 0.009).

Although not a coding region, length polymorphism within the first intron of BZLF1 gene was also described to be associated with malignancy and/or compartmentalization [10]; so we also assessed this sequence variation in our IM and lymphoma samples.

Length polymorphisms were detected both in the number of repeated units and also in its length. Repeated units consisted of 29 or 30 nucleotides that were repeated between two and ten times (Tables 1 and 2). The most frequent

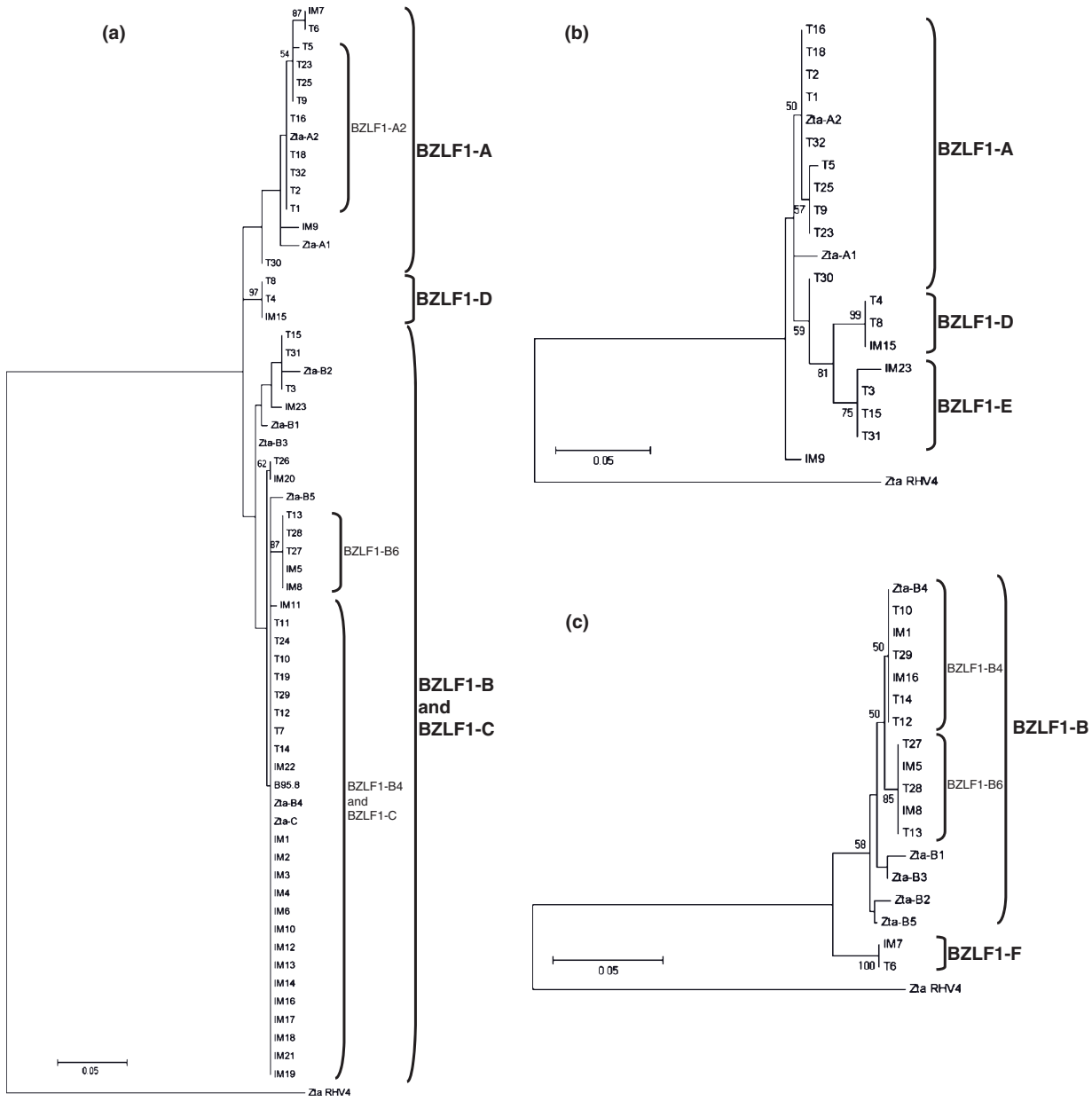

FIG. 2. Phylogenetic reconstruction. Trees obtained from the alignment of the complete protein sequence of the BZLF1 gene from patients with infectious mononucleosis (IM) and Epstein–Barr virus (EBV) -associated lymphomas. Only oral secretions (OS) samples were included from IM patients because both compartments harboured the same variant. Prototype sequences for each BZLF1 variant (Zta-A1, A2, B1 to B5 and Zta-C) were characterized in ref. [8]. Bootstrap values obtained after 1000 resamplings are indicated. The scale bars represent a genetic distance calculated by the number of base substitutions per site of 0.05. (a) Sequences from all patients were included in the original alignment. Variants BZLF1-B4 and BZLF1-C appear as a single taxa in the tree because they only differ in the deletion of amino acid (aa) 127 and deleted positions are not compared by phylogenetic software. (b) Tree obtained from the alignment of those sequences which did not contain the deletion of aa 127. (c) Tree obtained from the alignment of those sequences which contained the deletion of aa 127.

polymorphism in our series was two 30-bp unit repeats, in 16/23 patients with IM and 11/28 patients with lymphomas. As previously described, sequences from OS and PBMCs from IM patients showed 100% identity. Interestingly, the 29-bp unit, no matter the amount of repeats, preferentially occurred in lymphomas and in BZLF1-A2 variants (p 0.022 and p 0.0001, respectively).

Finally, an interesting finding was that the three sequences characterized here as a new variant, BZLF1-D, all contained high numbers of 30-bp unit repeats (Tables 1 and 2).

TABLE 3. Lymphoma association for individual amino acid substitutions

Substitution	Frequency in lymphoma (%)	Frequency in infectious mononucleosis (%)	p value
(T68A)	13/28 (46)	3/23 (13)	0.0152
(S76P)	13/28 (46)	3/23 (13)	0.0152
(T124P)	11/28 (39)	1/23 (4)	0.0063
(V146A)	16/28 (57)	4/23 (17)	0.0047
(V152A)	10/28 (36)	2/23 (7)	0.0441
(Q163L)	10/28 (36)	2/23 (7)	0.0441
(E176D)	10/28 (36)	1/23 (4)	0.0075
(Q195H)	10/28 (36)	2/23 (7)	0.0441
(A205S)	14/28 (50)	2/23 (7)	0.0021

Fisher's Exact Test was applied in every case.

Discussion

Epstein–Barr virus sequence polymorphisms and their association with a wide variety of tumours have been widely studied in different geographical locations. Even though, definitive conclusions regarding geographical and disease associations are still lacking and particularly, information on EBV polymorphisms in South America is scarce. Moreover, much of the existing literature focuses on EBV latency genes while it disregards lytic genes and the notion that EBV reactivation capability could be involved in lymphoma progression. To date only a few studies have addressed sequence variation in the coding sequence of BZLF1, an immediately early gene whose product is responsible for the lytic phase switch (extensively reviewed in refs [3,4]).

Based on phylogenetic analysis and following the terminology introduced by Luo *et al.* [8], we characterized sequence variation in the coding region of the BZLF1 gene in IM and lymphoma patients from our geographic region as representatives of the lytic and latent EBV states, respectively.

In contrast to what was observed in Asia, where a BZLF1-A1 variant (also called variant 5 in ref. [9]) was the most prevalent in all types of samples analysed [8], in our region BZLF1-C variant was the most frequently detected, but still occurred preferentially in IM patients. Meanwhile, only 1/89 samples presented this variant in Asia [8]. On the other hand, variant BZLF1-A2, which was not associated with nasopharyngeal or gastric carcinomas in Asia [8], was statistically associated with lymphomas in our region. Moreover, the presence of a 29-bp unit internal repeat, instead of a 30-bp unit repeat, was also associated with lymphomas and in turn with BZLF1-A2 variant. Hence, in our geographic region BZLF1-A2 variants with 29-bp repeated units could be among the factors involved in lymphomagenesis.

With regards to length polymorphism, published data provide conflicting results but most refer to the amount of repeated units. While Ikeda *et al.* [16] described no differences in compartmentalization of variants with two or three repeats in a healthy population in Japan, in China Chen *et al.* [17] reported that three 29-bp repeats were preferentially found in tumours while two repeats were found in PBMCs. In contrast, Sacaze *et al.* [10] reported three 30-bp units in PBMCs and two 30-bp units in tumours in Tunisian patients with nasopharyngeal carcinomas. In our study, no significant differences were observed concerning the number of repeated units between pathologies or in distribution between compartments in IM patients. On the other hand, we reported a much greater diversity in this aspect concerning both the number of repeated units, ranging from two to ten repeats, as well as in the length of each repeated unit, where both 29-bp and 30-bp units were observed. Moreover, previously characterized sequence variation in the promoter region of BZLF1 gene, showed an association of Zp-V3 variant with tumours (M. A. Lorenzetti, P. A. Chabay & M. V. Preciado, personal communication). This association was also described by other authors regardless of the geographic location [18–21]. Moreover, the association between the BZLF1-A2 and Zp-V3 variants with lymphomas was also significant by means of multi-region gene phylogenetic analysis followed by BaTS. In addition, when comparing all polymorphisms associated with malignancies, variants characterized as Zp-V3 in the promoter region and BZLF1-A2 in the coding region with 29-bp repeated units in intron 1, occurred in nine lymphoma cases but only in one IM case ([11] and M. A. Lorenzetti, P. A. Chabay & M. V. Preciado, personal communication). Hence, this polymorphism combination appears to be associated with lymphomas in our region.

Our present results suggest not only the possible existence of tumour-associated variants, but also the existence of geographically distributed ones. The new BZLF1-D, -E and -F variants appear to circulate preferentially in our region, although definitive conclusions are precluded because reports on this EBV genomic region are scarce and this is the first report to assess BZLF1 coding region polymorphisms from South America. Furthermore, samples harbouring the newly characterized BZLF1-D variant were the same as those that were previously characterized as Zp-V3+51 variants in the promoter region of the BZLF1 gene, another variant of exclusive distribution in our region [11]. As previously mentioned, given that only a few BZLF1 sequences are available on GenBank, deeper phylogeographic studies are not yet possible. However, the notion of geographical distribution of EBV variants is not new and is mostly exemplified by *EBNA1* gene, in which we previously characterized a high proportion of V-leu variants, circulating in the whole American continent, and V-ala variants, restricted to South America [22].

No sequences, from either IM patients or lymphoma patients, presented mutations in any of the specific amino acids, namely N182, A185, S186, C189 and R190, within the DNA-binding domain (aa 178–196), previously described as essential for DNA binding and in turn, the activation of downstream genes required for lytic switch [23,24]. Only substitution Q195H was detected in BZLF1-A2 and BZLF1-F variants but this was not assessed by Heston *et al.* [23]. Likewise, the present sequences had no mutations in the N-terminal region of the protein, specifically in the BRLF1 cooperation domain (aa 25–42). With regards to the dimerization domain (aa 197–221) in the C-terminal region, aa substitutions were detected in positions 204, 205 and 206. Previous work indicated that substitutions A205S and A206S had no effect on Zta protein dimerization capability [25], but there is no information about the possible effects of the naturally occurring substitutions A204S and A206T. Furthermore, no mutations were detected in the conserved coiled-coil dimerization region (aa 208–215) required for EBV DNA replication [26].

Functional analysis of natural occurring variants, particularly those with a high number of mutations within the *trans*-activation domain (aa 1–167), like BZLF1-A2 variant, are still lacking and could provide a clear insight into the ability of the virus to enter the lytic cycle and so help to confirm if this variant is associated with lymphoma development.

In a recent study, Yu *et al.* [27] demonstrated that a mutant virus incapable of repressing BZLF1 transcription was unable to establish latency and was consequently unable to transform B cells *in vitro*, given that spontaneous reactivation caused infected cells to die. However, this same mutant virus was able to transform and induce B-cell lymphoma in an *in vivo* humanized mouse model because most infected B cells appear to have an abortive form of lytic gene expression that does not result in the release of infectious virus particles or killing host cells [28]. It could be speculated that a Zta protein with altered capabilities in *trans*-activating lytic viral genes could produce a similar phenotype, either by over-activating lytic genes, leading to an abortive lytic cycle, or by failing to do so and preventing transformed lymphoma cells from dying if they never undergo spontaneous lytic reactivation. However, when viral load was assessed in plasma (for IM cases) or tissue biopsy (for lymphoma cases) in a proportion of the present cases, no correlation was found between viral copy number and BZLF1 polymorphisms (data not shown).

In a similar way, mutations within the DNA binding domain, especially in position 186, have been shown to disturb the recognition of methylated Zta responsive elements within promoter of lytic cycle genes [24]. Understanding differences in the capability of natural variants concerning lytic cycle induction could be most important, particularly since tumour imaging following induction of the viral lytic cycle was also suggested as a possible follow-up procedure in a preliminary animal model [29]. Furthermore, this same animal model of viral lytic cycle induction, combined with targeted radiation therapy, proved to be an efficient strategy to regress EBV-related tumour growth *in vivo* [30].

In conclusion, we assessed for the first time polymorphisms in the BZLF1 gene in a series of paediatric patients in Argentina. The evaluation of sequence variation in samples from paediatric patients with IM at the time of diagnosis allowed for the identification of the original infecting variant in contrast to co-infection with multiple variants often seen in adult healthy carriers [31]. In addition to previously described variants, we characterized three new phylogenetic groups circulating in our region. Moreover, we suggest the association of particular variants either with benign IM or with malignant lymphomas. Taken together these observations could provide some expansion of the existing knowledge about EBV sequence variation and its association with malignancy and the geographical distribution of EBV variants.

Acknowledgements

We thank G. Moscatelli and S. Moroni (Medical Staff at the Parasitology and Chagas Disease Laboratory, Ricardo Gutiérrez Children Hospital) for providing assistance in reviewing clinical records. This study was supported in part by a grant from the National Agency for Science and Technology Promotion (PICT 2012 no. 0291 and PICT 2008 no. 0052). M.A.L., P.A.C. and M.V.P. are members of CONICET, Research Career Programme. M.G. is supported by a fellowship from the National Agency for Science and Technology Promotion (ANPCyT). J.A. and E.D.M. are members of the CABA Research Career Programme.

Transparency Declaration

None of the authors has any competing interest.

References

1. Luzuriaga K, Sullivan JL. Infectious mononucleosis. *N Engl J Med* 2010; 362: 1993–2000.
2. Young LS, Rickinson AB. Epstein–Barr virus: 40 years on. *Nat Rev Cancer* 2004; 4: 757–768.
3. Chang CM, Yu KJ, Mbulaiteye SM, Hildesheim A, Bhatia K. The extent of genetic diversity of Epstein–Barr virus and its geographic

and disease patterns: a need for reappraisal. *Virus Res* 2009; 143: 209–221.
4. Tzellos S, Farrell PJ. Epstein–Barr virus sequence variation – biology and disease. *Pathogens* 2012; 1: 156–175.
5. Iempridee T, Das S, Xu I, Mertz JE. Transforming growth factor ββ-induced reactivation of Epstein–Barr virus involves multiple SMAD-binding elements cooperatively activating expression of the latent-lytic switch *BZLF1* gene. *J Virol* 2011; 85: 7836–7848.
6. McDonald C, Karstegl CE, Kellam P, Farrell PJ. Regulation of the Epstein–Barr virus zp promoter in B lymphocytes during reactivation from latency. *J Gen Virol* 2010; 91: 622–629.
7. El-Guindy A, Ghiassi-Nejad M, Golden S, Delecluse HJ, Miller G. Essential role of RTA in lytic DNA replication of Epstein–Barr virus. *J Virol* 2013; 87: 208–223.
8. Luo B, Tang X, Jia Y, Wang Y, Chao Y, Zhao C. Sequence variation of Epstein–Barr virus (Ebv) *BZLF1* gene in EBV-associated gastric carcinomas and nasopharyngeal carcinomas in northern China. *Microbes Infect* 2011; 13: 776–782.
9. Ji KM, Li CL, Meng G, Han AD, Wu XL. New *BZLF1* sequence variations in EBV-associated undifferentiated nasopharyngeal carcinoma in southern China. *Arch Virol* 2008; 153: 1949–1953.
10. Sacaze C, Henry S, Icart J, Mariame B. Tissue specific distribution of Epstein–Barr virus (EBV) *BZLF1* gene variants in nasopharyngeal carcinoma (NPC) bearing patients. *Virus Res* 2001; 81: 133–142.
11. Lorenzetti MA, Gutierrez MI, Altcheh J *et al.* Epstein–Barr virus *BZLF1* gene promoter variants in pediatric patients with acute infectious mononucleosis: its comparison with pediatric lymphomas. *J Med Virol* 2009; 81: 1912–1917.
12. Hall TA. Bioedit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/nt. *Nucleic Acids Symp Ser* 1999; 41: 95–98.
13. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011; 28: 2731–2739.
14. Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol* 2008; 8: 239–246.
15. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007; 7: 214.
16. Ikeda T, Kobayashi R, Kogashiwa Y, Matsuda T, Kohno N. Epstein–Barr virus diversity in immunocompetent healthy persons: reassessment of the distribution of genetic variants. *J Med Virol* 2013; 86: 301–305.
17. Chen HL, Lung ML, Chan KH, Griffin BE, Ng MH. Tissue distribution of Epstein–Barr virus genotypes. *J Virol* 1996; 70: 7301–7305.
18. Gutiérrez MI, Ibrahim MM, Dale JK, Greiner TC, Straus SE, Bhatia K. Discrete alterations in the *BZLF1* promoter in tumor and non-tumor-associated Epstein–Barr virus. *J Natl Cancer Inst* 2002; 94: 1757–1763.
19. Tong JH, Lo KW, Au FW, Huang DP, To KF. Re: discrete alterations in the *BZLF1* promoter in tumor and non-tumor-associated Epstein–Barr virus. *J Natl Cancer Inst* 2003; 95: 1008–1009; author reply 1009.
20. Martini M, Capello D, Serraino D *et al.* Characterization of variants in the promoter of EBV gene BZLF1 in normal donors, HIV-positive patients and in AIDS-related lymphomas. *J Infect* 2007; 54: 298–306.
21. Jin Y, Xie Z, Lu G, Yang S, Shen K. Characterization of variants in the promoter of *BZLF1* gene of EBV in nonmalignant EBV-associated diseases in Chinese children. *Virol J* 2010; 7: 92.
22. Lorenzetti MA, Altcheh J, Moroni S, Moscatelli G, Chabay PA, Preciado MV. *EBNA1* sequences in Argentinean pediatric acute and latent Epstein–Barr virus infection reflect circulation of novel South American variants. *J Med Virol* 2010; 82: 1730–1738.
23. Heston L, El-Guindy A, Countryman J, Dela Cruz C, Delecluse HJ, Miller G. Amino acids in the basic domain of Epstein–Barr virus zebra protein play distinct roles in DNA binding, activation of early lytic gene expression, and promotion of viral DNA replication. *J Virol* 2006; 80: 9115–9133.
24. Wille CK, Nawandar DM, Panfil AR, Ko MM, Hagemeier SR, Kenney SC. Viral genome methylation differentially affects the ability of *BZLF1* versus *BRLF1* to activate Epstein–Barr virus lytic gene expression and viral replication. *J Virol* 2013; 87: 935–950.
25. Hicks MR, Balesaria S, Medina-Palazon C, Pandya MJ, Woolfson DN, Sinclair AJ. Biophysical analysis of natural variants of the multimerization region of Epstein–Barr virus lytic-switch protein *BZLF1*. *J Virol* 2001; 75: 5381–5384.
26. McDonald CM, Petosa C, Farrell PJ. Interaction of Epstein–Barr virus *BZLF1* c-terminal tail structure and core zipper is required for DNA replication but not for promoter transactivation. *J Virol* 2009; 83: 3397–3401.
27. Yu X, McCarthy PJ, Wang Z, Gorlen DA, Mertz JE. Shutoff of *BZLF1* gene expression is necessary for immortalization of primary B cells by Epstein–Barr virus. *J Virol* 2012; 86: 8086–8096.
28. Ma SD, Yu X, Mertz JE *et al.* An Epstein–Barr virus (EBV) mutant with enhanced *BZLF1* expression causes lymphomas with abortive lytic EBV infection in a humanized mouse model. *J Virol* 2012; 86: 7976–7987.
29. Fu DX, Tanhehco YC, Chen J *et al.* Virus-associated tumor imaging by induction of viral gene expression. *Clin Cancer Res* 2007; 13: 1453–1458.
30. Fu DX, Tanhehco Y, Chen J *et al.* Bortezomib-induced enzyme-targeted radiation therapy in herpesvirus-associated tumors. *Nat Med* 2008; 14: 1118–1122.
31. Sitki-Green DL, Edwards RH, Covington MM, Raab-Traub N. Biology of Epstein–Barr virus during infectious mononucleosis. *J Infect Dis* 2004; 189: 483–492.