Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome

Citation for published version:

Digital Object Identifier (DOI):
10.1093/nar/gkt840

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Nucleic Acids Research

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome

Yan Huang1, Alberto Hidalgo-Bravo1, Enjie Zhang1, Victoria E. Cotton1, Aaron Mendez-Bermudez1, Gunjan Wig1, Zahara Medina-Calzada1, Rita Neumann1, Alec J. Jeffreys1, Bruce Winney2, James F. Wilson3, Duncan A. Clark4, Martin J. Dyer5 and Nicola J. Royle1,*

1Department of Genetics, University of Leicester, University Road, Leicester, LE1 7RH, UK, 2Department of Oncology, ORCRB, University of Oxford, Old Road Campus, Headington, Oxford, OX3 7DQ, UK, 3Centre for Population Health Sciences, College of Medicine and Veterinary Medicine, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland, 4Department of Virology, Barts Health NHS Trust, Pathology and Pharmacy Building, 80 Newark St, London E1 2ES, UK and 5Department of Cancer Studies and Molecular Medicine, University of Leicester, University Road, Leicester, LE1 7RH, UK

Received July 23, 2013; Revised and Accepted August 28, 2013

ABSTRACT

Linear chromosomes are stabilized by telomeres, but the presence of short dysfunctional telomeres triggers cellular senescence in human somatic tissues, thus contributing to ageing. Approximately 1% of the population inherits a chromosomally integrated copy of human herpesvirus 6 (CI-HHV-6), but the consequences of integration for the virus and for the telomere with the insertion are unknown. Here we show that the telomere on the distal end of the integrated virus is frequently the shortest measured in somatic cells but not the germline. The telomere carrying the CI-HHV-6 is also prone to truncations that result in the formation of a short telomere at a novel location within the viral genome. We detected extra-chromosomal circular HHV-6 molecules, some surprisingly comprising the entire viral genome with a single fully reconstituted direct repeat region (DR) with both terminal cleavage and packaging elements (PAC1 and PAC2). Truncated CI-HHV-6 and extra-chromosomal circular molecules are likely reciprocal products that arise through excision of a telomere-loop (t-loop) formed within the CI-HHV-6 genome. In summary, we show that the CI-HHV-6 genome disrupts stability of the associated telomere and this facilitates the release of viral sequences as circular molecules, some of which have the potential to become fully functioning viruses.

INTRODUCTION

The ends of linear chromosomes are distinguished from double-strand breaks within the genome by specialized nucleoprotein structures known as telomeres. Human telomeres comprise a variable-length double-stranded DNA molecule that is predominantly 5’-TTAGGG)n-3’ (1) but includes, at the start of the repeat array, sequence-variant or degenerate repeats interspersed with TTAGGGG repeats (2–4). Each telomere is terminated by an essential single-strand overhang of 50–300 nt (5). The telomeric DNA binds to the six-protein complex, known as Shelterin, via the double-strand binding TRF1 and TRF2 proteins and the single-strand binding protein POT1 (6). Telomeric DNA bound to the Shelterin complex forms looped structures (t-loops) in which the single-strand overhang invades the upstream duplex telomeric DNA forming a D-loop (7). The Shelterin complex on the capped telomere prevents inappropriate repair by non-homologous end-joining or by homologous recombination (HR) (8) and, in conjunction with other protein complexes, it regulates telomerase activity by controlling access to the single-strand overhang. However, telomerase is inactive in differentiated human cells and shows limited activity in stem cells; consequently in most cells, telomeric...
DNA is eroded as a result of incomplete lagging-strand synthesis and end-processing that restores single-strand overhangs (6). Disturbance of the telomere capping function or telomere length regulation can result in genome instability leading to tumourigenesis, and the presence of a few short telomeres (9,10) induces senescence, which has known roles in ageing.

HHV-6A and B belong to the Roseolovirus genus of the \( \beta \)-herpesvirus subfamily, and though closely related, they present diverse biological, epidemiological, pathological and molecular properties. HHV-6B is prevalent in most populations and primary infection usually occurs in early childhood (<2 years) causing an intense fever and rash, whereas infections by HHV-6A tend to be less common (11–13). As for most herpesviruses, HHV-6A and B can remain latent lifelong, with the potential to reactivate that can have severe consequences. For example, HHV-6B reactivation in immunocompromised transplant patients can cause encephalitis and has been linked to bone marrow suppression and to graft failure (14). The HHV-6A and B genomes are double-stranded DNA molecules comprising \( \approx 145 \) kb of unique sequence encoding U1 to U100 ORFs, flanked by identical left and right direct repeats (DR\(_{L}\) and DR\(_{R}\)), each \( \approx 8 \) kb (15–17). Each DR is terminated by the packaging and cleavage sequences, PAC1 and 2 (18), and includes two arrays of telomere-like repeats (T1 and T2) and a variety of ORFs (DR1–8). T2 is known to be a short array of pure (TTA\(_n\))GGG\(_n\) that varies in length between viral isolates, whereas T1 comprises an array of degenerate telomere-like repeats (19,20).

HHV-6A and B are distinct among herpesviruses, as they can integrate into the human telomeric DNA, probably via HR with the terminal T2 region of DR\(_{R}\) (21–25). Once integrated in the germline, the chromosomally integrated virus (CI-HHV-6) can be inherited. Approximately 0.8% of the UK population are CI-HHV-6 carriers (26), but it is not known whether integration is a natural biological form of HHV-6 latency or whether it affects normal telomere function and length regulation. We show that the telomere carrying the integrated virus is often the shortest measured and that the presence of the full-length viral genome disrupts stability of the telomere, resulting in frequent truncations. Moreover, we detected extra-chromosomal circular DNA comprising viral sequences and showed that the viral genome can be released from the telomere, possibly via a t-loop excision mechanism. Altogether, our data are consistent with the proposal that telomeric integration is a form of HHV-6 latency.

**MATERIALS AND METHODS**

**Oligonucleotides**

Oligonucleotides for amplification of HHV6 sequences were primarily designed based on the GenBank consensus sequences for HHV-6A (U1102 accession no. X83413.1) or HHV-6B (HST accession no. AB021506.1; Z29 accession no. AF157706.1).

**DNA samples and lymphoblastoid cell lines**

To identify CI-HHV-6 carriers, we screened the HapMap Phase I (27), the CEPH-HGDP (28) and parental DNA samples from the CEPH family panel (29). Therefore, we screened 1178 samples from unrelated donors in these three panels (duplicated samples between the panels were removed). We also screened 528 samples in the People of the British Isles panel (30) and 2153 samples from the Orkney Complex Disease Study (31). The panel of 92 sperm donors, a gift from Alec J. Jeffreys (32), was also screened for the presence of CI-HHV-6.

Lymphoblastoid cell lines (LCLs) from one CI-HHV-6A (designated 3-10q26.3) and five CI-HHV-6B carriers (designated 1-9q34.3, 2-9q34.3, 4-11p15.5, 5-17p13.3 and 6-17p13.3) reported previously (24) were donated by Duncan Clark. Other LCLs were from Bruce Winney: NWA090 (CI-HHV-6A), BAN519, COR264, CUM082, DER512, NWA008, YOR546 (all CI-HHV-6B); from David Gurwitz CEPH-HGDP00628 (CI-HHV6A); from the cell bank at the Coriell Institute for Medical Research: GM18999 (CI-HHV-6A), GM07022 (CI-HHV-6B), CEPH 1375.02 [also known as GM10863 (CI-HHV-6B)] and from other members of the CEPH 1375 family. All the LCLs were grown in RPMI1640 medium with 10–15% foetal calf serum at 37°C in a humidified incubator with 5% CO\(_2\) and subcultured using standard methods. To measure the telomere shortening rate in the GM18999 (CI-HHV-6A) and CEPH1375.02 (CI-HHV-6B) cell lines, the cells were cultured over a longer period. The number of population doublings was calculated from cell counts obtained using a haemocytometer, and the percentage of dead cells was determined using Trypan blue staining at each subculturing. Cell pellets were obtained throughout the time course.

**Identification of HHV-6-positive DNA samples**

Panels of DNA samples in 96-well format were screened by polymerase chain reaction (PCR) for the presence of U11 or U18 sequences. Standard 10 μl PCRs were prepared containing 5 ng genomic DNA, 0.3 μM of each primer, 0.05 U/μl of Taq polymerase (Kapa Biosystems) and a buffer supplied with the enzyme. Thermal cycling conditions were 96°C for 1 min then 35 cycles of 96°C for 15 s; 62–64°C for 30 s; and 68°C for 1 min). Forward and reverse primers: U11F 5’-TTTTTTATCATTACGACGCGATC; U11R 5’-GGGACGCGAATCGGAGGAAGC; U18F 5’-CATATCGATCAACCTTGCGATG; and U18R 5’-ATTTCGCGTCAACCTTGCGATG; and U18R 5’-AT AACACGATCTGAAAATGCAC. Identification of the virus type was achieved by amplification with primers DR5F(A): 5’-CCGTCGACCTTCTGTTCTTATGC; DR5R(A): 5’-CACATACATGAAACCCACAC for HHV-6A and DR7F(B): 5’-AGGCC; DR7R(B): 5’-CCGGATACGTCGACTTCTGTTCTTATGC for HHV-6B. Amplicons of the expected sizes were detected by agarose gel electrophoresis.

**Single telomere length analysis**

Single telomere length analysis (STELA) was conducted as described (33) using the specialized PCR buffer and other
modifications reported previously (34,35). The primer concentrations in the 10μl STELA reactions were flanking primer 0.3μM, Telorett2 0.225μM and Teltail 0.05μM. The Taq polymerase (Kapa Biosystems) was used at 0.04 U/μl and Pwo (Genaxxon Bioscience) at 0.025 U/μl. The primers located adjacent to the telomere repeat and used to amplify the double-stranded portion of the 12q, 17p and XpYp telomeres were 12q-STE LA: 5′-GAAGCACTTCCTCTCAG; 17p6: 5′-GGCTGAACT TATAGCCTCTGC; XpYpE2: 5′-TTGTCTCAGGTTGTC TTAGTG. Primers for use for STELA for the integrated HHV-6A or B were DR1R: 5′-GAAGAAGATGCGGTT GTCTTGTT, DR8F: 5′-GGATTACCGAGTGGAAT GTTGC and DR8R(HHV-6B): 5′-CGCCCCCGCAGCTGC CATAGAG. Control primers in the opposite orientation that did not yield STELA amplicons were: DR1F: 5′-TA TGCCCGAGCAAGATGC; DR3F: 5′-TCCGTTC CCTCATCGGCATCT and DR8F: 5′-CATAGATCGG GACTGCTTGA. DNA in each STELA reaction ranged between 100 and 1000 pg. Agarose gels (0.8%) were used for telomere length analysis with size markers (GeneRuler 1 kb and GeneRuler High Range DNA ladder, Fermentas) run in every gel. Amplicon length analysis was conducted on the phosphor-images (minimum two gels per sample) using the Imagequant software (Typhoon 9400, GE Healthcare). The length of the flanking DNA was subtracted from each amplicon, and the median telomere length and interquartile range of the telomeres molecules were determined. For each donor, the data from four telomeres are presented as scatter plots (GraphPad Prism). The shortest telomere frequencies were compared using the non-parametric Kruskal–Wallis analysis (GraphPad Prism). For single Gaussian distributions, outliers were then defined as those telomere lengths of greater or less than three standard deviations from the mean (which includes >99.7% of values from that population). Where the data showed two Gaussian distributions, outliers were then defined as those telomere lengths of greater or less than three standard deviations from the mean (which includes >99.7% of values from that population). Where the data showed two Gaussian distributions, the distribution of the major population was used to determine the number of outliers. The frequency of short outliers was then determined as a percentage of the total number of telomere molecules analysed and plotted against the mean telomere length. The data were tested for linear correlation using Pearson’s correlation coefficient (GraphPad Prism).

Sperm DNA analysis

The panel of sperm donors (32) was screened for the presence of HHV-6 as described. There was sufficient DNA to conduct telomere length analysis by STELA on one CI-HHV-6A and three CI-HHV-6B carriers. Telomere length measurements in the sperm DNA samples were made from phosphor-images of STELA blots and scatter plots produced (as above). Frequency histograms were generated by sizing individual molecules and allocating them to 1 kb bins. Distributions were then fitted using a model with either one or two Gaussian distributions (GraphPad Prism). For single Gaussian distributions, outliers were then defined as those telomere lengths of greater or less than three standard deviations from the mean (which includes >99.7% of values from that population). Where the data showed two Gaussian distributions, the distribution of the major population was used to determine the number of outliers. The frequency of short outliers was then determined as a percentage of the total number of telomere molecules analysed and plotted against the mean telomere length. The data were tested for linear correlation using Pearson’s correlation coefficient (GraphPad Prism).

Analysis of the HHV-6 DR regions

The length of the DR1-T2 was determined following PCR using primers UDL6R: 5′-TTTCTACAGCAGTGCAGT CT and DR8F. The presence of PAC2 in these amplicons was determined by PCR with UDL6R and a primer that anneals to PAC2, PAC2F: 5′-TTGTACACGCGCCGCTTTGTT TTTC. The absence of the PAC1 sequence in STELA amplicons was determined by digestion of STELA products (DRL) and DR8F. The presence of PAC1 in these amplicons was determined by PCR with UDL6R and a primer that anneals to PAC2, PAC1B: 5′-TGGAAGGCGCCCGAGCTCTGAA. The absence of PAC1 was confirmed by digestion of STELA products (DR1) or U100F2-DR1R amplicons (DRR) with SmaI (cuts at PAC1) or Sall as a control for digestion.
Sequencing the CI-HHV-6B genome using ion semiconductor technology

Overlapping amplicons across the unique region of HHV6 and the DRs (without T1 and T2) were generated from CEPH1375.02 CI-HHV-6B carrier, using primers based on the HST sequence (AB021506.1) (37). The amplicons were pooled and sonicated using a Bioruptor (Diagenode Inc.). The subsequent preparation of the sonicated DNA for sequence analysis using an IonTorrent personal genome sequencer (Life Technologies) was essentially conducted according to the manufacturer’s protocol. Briefly, the sheared DNA was end repaired and the adapters were ligated (according to Life Technologies’ protocol). Size selection of ~170 bp fragment was achieved in a 2.5% NuSieve agarose gel followed by purification using a Zymoclean Gel DNA recovery kit (ZymoResearch, Irvine, CA, USA). The size-selected DNA was then prepared for Ion Torrent sequencing on a 314 chip (100 bp read length) according to Life Technologies’ protocol.

Detection of HHV-6 RNA transcripts

Total RNA was extracted using Tri-reagent (Sigma-Aldrich). The RNA was dissolved in water (100 μl) and treated with RNase-free DNase I (100 U/μl, 20 min, 37°C). The RNA was puriﬁed by phenol–chloroform extraction, precipitated with isopropanol and re-suspended in RNA-free water (100 μl). cDNA was synthesized from 2 μg of total RNA using M-MuLV reverse transcriptase (New England Biolabs) and random primer hexamers (0.2 μl) in a total reaction volume of 20 μl at 42°C for 1 h, followed by enzyme inactivation at 90°C for 10 min. PCRs were then performed in 10 μl, containing 40 ng of synthesized cDNA and 0.05 U/μl of Tag DNA polymerase. Oligonucleotides used were U38F: 5'-TTGTCTTATTCGCGCTATAGA; U73F: 5'-TTACGGAGCGGCAGTAAAC and U73R: 5'-AGACATCGCTTCAAGCGGAA; U90B2F: 5'-CTTGACTAGCGATCTTGTTGCA and U90B2R: 5'-ACGACATCGCTTCAAGCGGAA. Thermal cycling conditions were denaturation at 96°C for 1 min followed by 35 cycles at 96°C for 15 s, annealing at 58–64°C (depending on amplicons) for 20 s, extension at 70°C for 1 min and a ﬁnal extension step at 70°C for 10 min. PCR products were analysed by agarose gel (1%) electrophoresis.

2D gel analysis

Genomic DNA (25 μg) from LCLs was digested with 100 U of EcoRI in a 200-μl reaction volume at 37°C for 2 h, extracted with phenol–chloroform and precipitated with sodium acetate and ethanol. To enrich for circular molecules, the EcoRI-digested genomic DNA was then digested with 100 U of ‘Plasmid-safe’ ATP-dependent DNase (Epicentre), which only digests double-stranded linear DNA, in a total volume of 500 μl at 37°C for 20 h. The DNA was then concentrated by isopropanol precipitation before 2D gel electrophoresis. Briefly, the DNA was separated in 0.4% agarose, 0.5× Tris-borate-ethylenediaminetetraacetic acid buffer, pH 8.3, at 1 V/cm for 18 h at 4°C in the ﬁrst dimension and in 1% HGT agarose in 0.5× Tris-borate-ethylenediaminetetraacetic acid containing 0.5 μg/ml ethidium bromide at 5 V/cm for 5.5 h in the second dimension. A Southern blot was prepared and hybridized to a 32P-labelled HHV6B DR probe, then after stripping to a telomere probe.

Detection of CI-HHV-6B truncations at DR1-T2

Genomic DNA (0.5 μg) was digested with ExoI (New England Biolabs) to remove 3’ single-strand overhangs from telomeres. Following ExoI digestion, the integrity of the double-stranded genomic DNA was veriﬁed by PCR ampliﬁcation of ~6 kb fragment using primers U100FB 5’-CATCTGATCTGTTGCGGCGGCGG and DR1R. STELA was conducted as above using ﬂanking primers DR1R or the UDL6R on genomic DNA with or without ExoI digestion.

Detection of extra-chromosomal circular HHV-6 DNA

Genomic DNA (90 ng/reaction) was ampliﬁed with primers UDL6R and U100F at 0.2 μM with Taq (0.04U/μl) and Pwo (0.025U/μl) DNA polymerases in 10 μl reaction prepared in the same PCR buﬀer used for STELA. Control PCRs were also performed with primers UDL6R and DR3F or U100F2 and DR3R. The PCR products were resolved in 0.8% agarose gels, and Southern blots were prepared and hybridized to a DR3 probe generated using DR3F and DR3R primers. The amplicons generated from extra-chromosomal circular DNA in ﬁve separate PCRs from CEPH1375.11 CI-HHV-6B were reampliﬁed and the composition of the DR region was veriﬁed by secondary PCRs and sequencing.

RESULTS

Identiﬁcation of CI-HHV-6 carriers and veriﬁcation of viral integration using STELA

To determine whether HHV-6 integration is a natural biological form of latency and what effect it has on the associated telomere, we investigated six LCLs from unrelated CI-HHV-6 carriers (24) with integration sites as follows: one CI-HHV-6 A at 10q, two CI-HHV-6B at 9q, one CI-HHV-6B at 11p and two CI-HHV-6B at 17p (24). We used STELA (33) to conﬁrm integration and to orientate the viral genome within the telomere (Figure 1A and B).

To identify more CI-HHV-6 carriers, we screened 3859 DNA samples from unrelated donors (see Materials and Methods) using primers that amplify conserved segments from the viral U11 and U18 genes. Integration was veriﬁed using STELA. As a result, 58 additional CI-HHV-6 positive samples were identiﬁed among the various populations (4 CI-HHV-6 A and 54 CI-HHV-6B). Where available we obtained LCLs or additional high
Figure 1. Characterization of CI-HHV-6 carriers and proof of telomeric integration using STELA. (A) The HHV-6A or 6B genome inserts in the reverse orientation such that DRR and the U100 gene are closest to the subtelomeric sequences. The DRL is expanded to show the reverse order of DR8, 3 and 1 ORFs. Plain arrows represent primers. The telorette 2 oligonucleotide (grey line) and Teltail primer used in STELA are shown. Orientation of DRS is shown by dashed arrows. Cen, centromere. (B) Amplification of the CI-HHV-6-associated telomere using STELA in a CI-HHV-6B carrier [1-9p34.3 (24)]. Amplicons are generated with the DR1, 3 and 8 reverse (R) primers but not the forward (F) primers, confirming the orientation of the integrated virus. The increasing size of the STELA products generated by DR1R, 3R and 8R reflects the distance from the 3' single-strand overhang. The blot was hybridized to a telomere probe. (C–E) Identification of the CEPH1375 family that segregates CI-HHV-6B through three generations. (C) Family members were screened for the presence the HHV-6 U18 gene. An amplicon from the XpYp telomere-adjacent sequence was used as a PCR control. (D) CEPH1375 family members are positive for CI-HHV-6B and showed the presence of the HHV-6B DR7 and absence of the HHV-6A DR5. Φ, size marker; NC, negative control. (E) CEPH1375 family showing transmission of CI-HHV-6B through three generations. (F) Detection of STELA amplicons using the DR1R (R), but not the DR1F (F), primer confirmed the orientation and integration of HHV-6B in family members. All the STELA amplicons contain viral sequences (left) and (TTAGGG)_n repeats (right). The maternal grandfather (CEPH1375.11) shows particularly short telomeres on the end of the CI-HHV-6B.
molecular weight genomic DNA from the newly identified CI-HHV-6 carriers for further analysis. We also identified a family (CEPH1375) that has transmitted CI-HHV-6B over three generations (Figure 1C–F).

Characterization of the DRs in CI-HHV6 carriers

Sequence analysis of single STELA products from seven individuals showed that the terminal PAC1 sequence (at the distal end of DR₄) was absent in CI-HHV-6 A (n = 2; 3-10q26.3 and 1501) and CI-HHV-6 B (n = 5; 1-9q34.3, 2-9q34.3, 5-17p13.3, 6-17p13.3 and HGDP0092) carriers. We isolated the 10q chromosome-HHV-6 junction from one CI-HHV-6 A carrier [3-10q26.3 (24)] using inverse PCR (25). The internal chromosome-HHV-6 junctions from four other carriers were isolated by amplification between a subterminal sequence (found at several chromosome ends, Sub17p-539) and the DR₈F primer in DRR (Figure 2 and Supplementary Figure S1). The isolated CI-HHV6 telomere was often the shortest in carriers (Figure 3A). The median length of the telomere (average 79 bp/cell division) at the virus-associated telomere was compared with the length of the CI-HHV-6-associated telomere in both cell lines (Supplementary Figure S3B), indicating that short virus-associated telomeres do not arise from the establishment or propagation of LCLs. The combined data show that the shortest-telomere frequency differs significantly between the chromosome-ends in somatic cells (Kruskal–Wallis test \( P < 0.0117 \)) and is highest at the CI-HHV-6-associated telomere (shortest in 45% of the carriers).

CI-HHV-6-associated telomeres in the germline

To investigate virus-associated telomere lengths in the germline, we screened sperm DNA samples from 92 men from UK of North European descent (32) and identified five (5.4%) CI-HHV-6 carriers (4 HHV-6B and 1 HHV-6 A), a higher frequency than seen in the British Isles panel but not statistically significant. Telomere length analysis in four men showed that all were considerably longer in sperm DNA, as expected (40). Interestingly, the 17p telomere, not the CI-HHV-6-associated telomere, was the shortest we measured in the germline of all four sperm donors. Moreover, the CI-HHV-6-associated telomere was the longest measure in two of the sperm donors (Table 1 and Supplementary Figure S4).

Further characterization of the left and right DR regions in 15 CI-HHV-6 carriers using a variety of primers (some anchored in the unique region to facilitate specific amplification of DR₉R or DR₉L) confirmed that the sequence organization is the same between the DRs. The length of the T2 region (perfect (TTAGGG)ₙ) in DR₉L is variable between carriers (20). Unexpectedly, the T1 region (degenerate telomere-like repeats) in DR₉R is highly expanded compared with the published strains (HHV-6 A U1102; HHV-6B HST and Z29) ranging between 0.7 and 1.5 kb in CI-HHV-6 A carriers and more strikingly 3–9 kb in CI-HHV-6B carriers (Figure 2C). We also show that all CI-HHV-6 carriers retain a copy of the PAC2 sequence immediately adjacent to DR₉L-T2 and an internal copy of PAC1 immediately distal to DR₉R-T1 (Figure 2C and Supplementary Figure S2).

The CI-HHV6 telomere was often the shortest measured in carriers

To investigate the length of the telomere on the end of the virus, we used STELA with the DR1R primer to amplify telomere molecules from small aliquots of LCL-DNA from CI-HHV-6 carriers (Figure 3A). The median length of the virus-associated telomere was compared with the length of XpYp, 12q and 17p telomeres. In 50% (8/16) of the LCLs, the CI-HHV-6-associated telomere was the shortest measured (Figure 3B, Supplementary Figure S3A) (9,38), even compared with the 17p telomere that has been reported as often being the shortest (36,39). Corresponding analysis of blood DNA samples from 24 carriers (21 CI-HHV-6 B from the Orkney Complex Disease Study study, and 3 CI-HHV-6 A siblings in a British family) again showed that the virus-associated telomere was most often the shortest [42% (10/24); Figure 3B and Supplementary Figure S3B], indicating that short virus-associated telomeres do not arise from the establishment or propagation of LCLs. The combined data show that the shortest-telomere frequency differs significantly between the chromosome-ends in somatic cells (Kruskal–Wallis test \( P < 0.0117 \)) and is highest at the CI-HHV-6-associated telomere (shortest in 45% of the carriers).

The HHV-6-associated 12q, 17p and XpYp telomeres shorten at similar rates in LCLs

The difference in the ranked length of the CI-HHV-6-associated telomere between somatic cells and the germline suggests that the presence of the viral genome affects length regulation. Analysis of telomere shortening rates in two LCLs (GM18999 CI-HHV-6 A and CEPH1375.02 CI-HHV-6B) showed similar attrition rates (average 79 bp/cell division) at the XpYp, 12q, 17p and CI-HHV-6-associated telomere in both cell lines (Supplementary Figure S5). This suggests that the disparity between the length of the CI-HHV-6-associated telomere in somatic cells and the germline is not explained by a higher rate of telomere erosion at the virus-associated telomere. However, we detected low-level telomerase expression in the CEPH1375.02 CI-HHV-6B at all time points and in some time points for the GM18999 CI-HHV-6A cell line. Clearly, the level of telomerase is insufficient to maintain telomere length in these cell lines but...
we cannot exclude the possibility that it may target one telomere more than another.

**Somatic truncation of the integrated HHV-6 genome and viral excision**

The detection of short telomere molecules in the germline (Supplementary Figure S4 and Figure 3C) prompted us to look for extra-chromosomal circular molecules. We detected extra-chromosomal circles containing (TTAGGG)_n repeats in the CEPH1375.02 LCL, which arise from low-frequency intra-telomere t-loop excision at all telomeres (41–43). We also detected extra-chromosomal circular molecules containing HHV-6 DR sequences (Supplementary Figure S6). These may arise from...
Figure 3. Telomere length analysis in CI-HHV-6 carriers. (A) STELA was used to measure telomere length at 12q, 17p, XpYp and on the end of the virus (DR1R) in CI-HHV-6 carriers. It shows a representative STELA-Southern blot on blood DNA from one CI-HHV-6 A carrier (1501) and below the scatter plots of the data obtained from several blots. The median and inter-quartile ranges are shown as red lines. (B) The median values for each of the four telomeres measured were ranked by length in DNAs from 16 LCLs and 24 blood DNAs from CI-HHV-6 carriers. The proportion of the samples in which each telomere is the shortest is shown in pie charts for the LCLs and blood DNA samples. The histogram of the combined data shows that the shortest telomere frequency (+SEM) is different between the chromosome ends (Kruskal–Wallis test $P < 0.0117$). (C) Relationship between telomere length and the frequency of short telomere molecules in the male germline. The short (outlier) telomere molecules, shorter than 3 standard deviations from the mean (<3SD), were counted as a proportion of the total molecules analysed for each telomere in four sperm donors and plotted against mean telomere length (Table1).
Table 1. Telomere length analysis in sperm DNA from CI-HHV-6 carriers

<table>
<thead>
<tr>
<th>Donor and telomere</th>
<th>Mean length ± SD (bp)</th>
<th>Total number of bands</th>
<th>Number of outliers</th>
<th>Percentage outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1-DR1R</td>
<td>13 767 ± 1590</td>
<td>385</td>
<td>28</td>
<td>7.3</td>
</tr>
<tr>
<td>D1-12q</td>
<td>14 428 ± 4111</td>
<td>357</td>
<td>22</td>
<td>6.2</td>
</tr>
<tr>
<td>D1-17p</td>
<td>9 996 ± 18 256</td>
<td>277</td>
<td>7</td>
<td>2.5</td>
</tr>
<tr>
<td>D1-XpYp</td>
<td>14 311 ± 2931</td>
<td>259</td>
<td>15</td>
<td>5.8</td>
</tr>
<tr>
<td>D2-DR1R</td>
<td>16 625 ± 2627</td>
<td>191</td>
<td>26</td>
<td>13.6</td>
</tr>
<tr>
<td>D2-12q</td>
<td>13 160 ± 2970</td>
<td>261</td>
<td>13</td>
<td>5.0</td>
</tr>
<tr>
<td>D2-17p</td>
<td>8 424 ± 1719</td>
<td>256</td>
<td>4</td>
<td>1.6</td>
</tr>
<tr>
<td>D32-DR1R</td>
<td>13 803 ± 1418</td>
<td>153</td>
<td>20</td>
<td>13.1</td>
</tr>
<tr>
<td>D44-DR1R</td>
<td>12 680 ± 2172</td>
<td>255</td>
<td>24</td>
<td>9.4</td>
</tr>
<tr>
<td>D44-12q</td>
<td>11 304 ± 3085</td>
<td>482</td>
<td>21</td>
<td>4.4</td>
</tr>
<tr>
<td>D44-17p</td>
<td>8 259 ± 1535</td>
<td>187</td>
<td>4</td>
<td>2.1</td>
</tr>
<tr>
<td>D44-XpYp</td>
<td>14 122 ± 1210</td>
<td>141</td>
<td>17</td>
<td>12.1</td>
</tr>
<tr>
<td>D56-DR1R</td>
<td>16 018 ± 1600</td>
<td>141</td>
<td>21</td>
<td>14.9</td>
</tr>
<tr>
<td>D56-12q</td>
<td>13 470 ± 3569</td>
<td>358</td>
<td>26</td>
<td>7.3</td>
</tr>
<tr>
<td>D56-17p</td>
<td>9 742 ± 2427</td>
<td>243</td>
<td>4</td>
<td>1.6</td>
</tr>
<tr>
<td>D56-XpYp</td>
<td>14 825 ± 1201</td>
<td>156</td>
<td>17</td>
<td>10.9</td>
</tr>
</tbody>
</table>

excision of a quasi t-loop formed by invasion of the telomeric 3’ single-strand overhang into the DR<sub>L-T2</sub> (Figure 4A). The reciprocal product of excisions should be a truncated CI-HHV-6 with a novel telomere at DR<sub>L-T2</sub>. Therefore, we conducted STELA using a flanking primer at the U1 gene and detected very short telomere molecules (140–200 bp) in five CI-HHV6B cell lines (two shown in Figure 4B). Amplification of the short telomere molecules is dependent on the presence of a telomeric 3’ single-strand overhang, as Exo I digestion before STELA abolished the products. Sequence analysis showed these amplicons contain DR<sub>L-T</sub>PAC2.

We hypothesized that the entire viral genome could be released from the telomere through telomeric 3’ single-strand overhang (eroded into the DR<sub>L-T1</sub> region) invasion into the internal DR<sub>R-T1</sub> region or between a viral genome already truncated at DR<sub>L-T2</sub> into the internal DR<sub>R-T2</sub> (Figure 4A). Resolution of the quasi t-loop, which would include the whole viral genome, could release the HHV-6 genome as a large circular t-loop, which would include the whole viral genome, and one LCL showed only the longer product. Sequence analysis showed that the shorter product had undergone a correct splicing to remove an intron (Figure 5). The HHV-6 U90 gene is an immediate early transactivator that produces different transcripts during productive infection and latency owing to alternative splicing and the use of alternative start sites (46). Detection of the spliced U90 transcripts is further evidence that CI-HHV-6 is a form of latency.

DISCUSSION

Among the five internal CI-HHV-6 junction fragments we isolated (2 CI-HHV-6A and 3 CI-HHV-6B carriers), four were different, thus representing independent integration events, and all shared a similar sequence organization (25,47). The absence of DR<sub>R-PAC2</sub> and the presence of telomere sequence-variant repeats (3) in each junction support the proposition that HHV-6 integrates by HR between DR<sub>R-T2</sub> and the proximal end of a telomere. The absence of PAC1 at the terminal end of DR<sub>L</sub> could also be a consequence of HR at the time of integration. However, the curious expansion of the T1 region,
identified through analysis of the internal DR<sub>R</sub>-T1, and the variable truncation of the terminal DR<sub>T</sub>-T1 in carriers suggest that the degenerate repeat region underwent erosion. Therefore, the expanded T1 region may serve a role during integration by acting as a buffer for replication-driven erosion that results in loss of PAC1 and part of the DR<sub>T</sub>-T1 region until (TTAGGG)<sub>n</sub> repeats are added by telomerase in the germline.

The CI-HHV-6-associated telomere was the shortest measured in somatic cells from 45% of the 40 carriers investigated but not the shortest in the four CI-HHV-6 sperm donors showing that the virus-associated telomere...
is readily lengthened by telomerase in the germline. The disparity between CI-HHV-6-associated telomere length in somatic cells and the germline suggests that the integrated virus perturbs an aspect of telomere function. We showed that the rate of telomere erosion through replication-driven processes was not higher at the CI-HHV-6-associated telomere in LCLs from two CI-HHV-6 carriers. However, as these cell lines showed a low level of telomerase activity, further analysis will be required to determine whether or not telomerase is targeted to the CI-HHV-6 telomere more frequently.

We surmised that the presence of the HHV-6 genome might interfere with telomere capping by perturbing of t-loop formation and excision giving rise to telomere rapid deletion events that were first described in yeast (48,49). We showed that LCLs from CI-HHV-6B carriers are mixed populations of cells carrying the full-length CI-HHV-6B and a subset with CI-HHV-6B chromosomes truncated at DRL-T2 by the presence of a novel very short telomere with a single-strand overhang (Figure 4). The truncations at DR_L-T2 could arise through a double-strand break at DR_L-T2, followed by end processing to generate a 3' single-strand overhang. However, as we also detected extra-chromosomal circular molecules containing HHV-6 DR sequences, we favour a model in which the truncated chromosomes arise through processing of a t-loop (41–43) formed by strand invasion of the telomeric 3' single-strand overhang within DR_L-T2. Excision of the quasi t-loop would result in a sudden truncation and formation of a telomere at DR_L-T2. The presence of the novel short telomere may contribute to cell cycle arrest and the onset of senescence.

Sequence analysis of the CI-HHV-6B genome in the CEPH1375.02 showed that the viral genome is intact in this family; moreover, we detected molecules that contain a reconstituted DR region with both PAC1 and PAC2 in the grandfather CEPH1375.11. The reconstituted DR region could only arise from a recombination event between the terminal DRL (that contains a full-length T2 and PAC2) and the internal DRR (that contains a full-length T1 and PAC1) resulting in the release of a circular molecule containing the viral genome with a single DR (Figure 4). Therefore, consistent with our model invoking t-loop formation including the viral genome, we propose that occasional single-strand invasion of the telomeric 3' overhang into the internal HHV-6 DR_R-T1 facilitates release of the entire viral genome from the chromosome with a single reconstituted DR (Figure 4A). These large extra-chromosomal circular molecules arise from an intact viral genome (as shown by the next-generation sequence analysis), and so they have the potential to undergo rolling circle replication that could regenerate full-length HHV-6B genomes with two identical terminal DRs. This and the evidence that spliced
transcripts from the U90 gene, which is involved in latency, were detected in the cell lines investigated support the hypothesis that CI-HHV-6 is an alternative form of viral latency.

The reciprocal product of a large excision event would be retention of a DR, lacking both PAC sequences, within the telomere. Interestingly, we have identified two unrelated individuals (2/3859; population frequency 0.05%) that lack the unique portion of the HHV-6B genome but carry a single HHV-6B-DR (without PAC1 and 2) integrated into a telomere. These individuals may have inherited chromosomes that had already undergone a viral excision event, although partial insertion during integration cannot be excluded. The released viral genome with a single reconstituted DR in CEPH1375.11 could also be retention of a DR, lacking both PAC sequences, within the telomere. Interestingly, we have identified two forms of viral latency.

In summary, we have shown that the CI-HHV-6-associated telomere is often one of the shortest in somatic cells and prone to sudden deletions that create a critically short telomere at a new location in the viral genome. The presence of a short telomere associated with the CI-HHV-6 will increase the chance that the cell will become senescent, thus affecting tissue homeostasis (9,10). Moreover, cells carrying critically short telomeres associated with CI-HHV-6 may be prone to telomere fusion events that can drive instability. We also show that integrated copies of HHV-6B can be excised from the chromosome and we propose that this is achieved through use of the t-loop excision mechanism. This may facilitate spreading to another telomere within the cell or represent the first step towards viral reactivation.

ACCESSION NUMBERS
GenBank: KF366418, KF366419, KF366420.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
The authors thank Sir Walter Bodmer for generously sharing the Population of the British Isles DNA panel; Mark Jobling, Howard Cann and David Gurwitz for sharing the HapMap, CEPH-HGDP DNA panels and for providing some cell lines. They thank Jenny Foxon and Dipti Vashi for help with cell culture and Farzaneh Atrain-Afyani for some technical assistance. They thank Dzanan Osmanovic and Pierpaolo Maisano-Delser for their assistance with IonTorrent sequence analysis.

FUNDING
The Medical Research Council [G0901657 to N.J.R.]; the Wellcome Trust Institutional Strategic Support Fund [WT097828MF] and by Conacyt, Mexico [to A.H.-B.]. Funding for open access: University of Leicester RCUK.

Conflict of interest statement. None declared.

REFERENCES


