BMC Genomics

Comparative genomic, transcriptomic, and proteomic reannotation of human herpesvirus 6

Alexander L. Greninger[1,2]* , Giselle M. Knudsen[3], Pavitra Roychoudhury[1,2], Derek J. Hanson[1], Ruth Hall Sedlak[1], Hong Xie[1], Jon Guan[1], Thuy Nguyen[1], Vikas Peddu[1], Michael Boeckh[2], Meei-Li Huang[1], Linda Cook[1], Daniel P. Depledge[4], Danielle M. Zerr[5], David M. Koelle[1], Soren Gantt[6], Tetsushi Yoshikawa[7], Mary Caserta[8], Joshua A. Hill[2] and Keith R. Jerome[1,2]

Abstract

Background: Human herpesvirus-6A and -6B (HHV-6) are betaherpesviruses that reach > 90% seroprevalence in the adult population. Unique among human herpesviruses, HHV-6 can integrate into the subtelomeric regions of human chromosomes; when this occurs in germ line cells it causes a condition called inherited chromosomally integrated HHV-6 (iciHHV-6). Only two complete genomes are available for replicating HHV-6B, leading to numerous conflicting annotations and little known about the global genomic diversity of this ubiquitous virus.

Results: Using a custom capture panel for HHV-6B, we report complete genomes from 61 isolates of HHV-6B from active infections (20 from Japan, 35 from New York state, and 6 from Uganda), and 64 strains of iciHHV-6B (mostly from North America). HHV-6B sequence clustered by geography and illustrated extensive recombination. Multiple iciHHV-6B sequences from unrelated individuals across the United States were found to be completely identical, consistent with a founder effect. Several iciHHV-6B strains clustered with strains from recent active pediatric infection. Combining our genomic analysis with the first RNA-Seq and shotgun proteomics studies of HHV-6B, we completely reannotated the HHV-6B genome, altering annotations for more than 10% of existing genes, with multiple instances of novel splicing and genes that hitherto had gone unannotated.

Conclusion: Our results are consistent with a model of intermittent de novo integration of HHV-6B into host germline cells during active infection with a large contribution of founder effect in iciHHV-6B. Our data provide a significant advance in the genomic annotation of HHV-6B, which will contribute to the detection, diversity, and control of this virus.

Keywords: Herpesvirus, Human herpesvirus 6, HHV-6B, HHV-6A, Betaherpesvirus, Viral genomics, Herpesvirus genomics, Comparative genomics, Genomic annotation, iciHHV-6

Background

HHV-6 is a ubiquitous betaherpesvirus that is divided into two species (HHV-6A and -6B) [1]. HHV-6B infects > 90% of children by 2 years of age, causing roseola, also called exanthem subitem or sixth disease, which is the leading cause of febrile seizures among children [2–5]. The virus persists in multiple cell types with consistent detectable viral DNA in saliva. HHV-6B reactivates in approximately 50% of allogeneic hematopoietic cell transplant (HCT) patients and is the most common cause of encephalitis in this setting. HHV-6B has also been associated with graft-versus-host disease, hepatitis, pneumonitis, and mortality after HCT, although causality remains to be proven [2].

Like other human herpesviruses, HHV-6A and -6B establish lifelong latency, but unique among human herpesviruses, they have the ability to integrate into human chromosomes. When this integration occurs in a germ cell, the virus can be passed to offspring and

* Correspondence: agrening@uw.edu
[1]Department of Laboratory Medicine, University of Washington, Seattle, WA, USA
[2]Fred Hutchinson Cancer Research Center, Seattle, WA, USA
Full list of author information is available at the end of the article

Greninger *et al. BMC Genomics* (2018) 19:204

Page 2 of 17

results in inherited chromosomally integrated HHV6 (iciHHV6). Affected individuals have a copy of the virus in each of their cells and the ability to pass on the integrated state to 50% of their offspring. IciHHV-6 is present in 0.5-2% of the population, constituting almost 70 million people worldwide, with the majority of these being iciHHV-6B [6]. iciHHV-6 can also be passed between individuals via transplantation [7, 8]. IciHHV-6 was recently associated with an increased risk of acute graft versus host disease and CMV viremia in HCT patients [9]. Integrated virus has been shown to reactivate both in vitro and in vivo and can confound assays for active HHV-6 infection [10]. The mechanism of integration and viral proteins required for integration are unclear [11].

Clinical testing for HHV-6 has hitherto been reserved to large academic medical centers and reference labs due to concerns over reactivation in HCT patients. Unbiased metagenomic sequencing has uncovered HHV-6 infection in a number of cases of encephalitis and febrile illness that were previously "unsolved" [12–14]. Of note, HHV-6 has recently been included in new, rapid, point-of-care multiplex PCR panels for meningitis/encephalitis and febrile illness [15]. Given the ease of use and extraordinarily rapid turn-around time of these multiplex PCR panels, they have already been adopted by thousands of hospitals across the world [15–17]. Because it is not uncommon for children to have HHV-6B in their cerebrospinal fluid around the time of primary infection, we expect the coming years to see hundreds of thousands of HHV-6 infections detected that previously would have gone undetected based on the sheer number of samples that will be tested for HHV-6 [18]. With so many new infections detected, there is an increasing need to understand the clinical associations, sequence diversity, and basic biology of this virus.

To date, only two complete genomes from replicating HHV-6B are available – the Z29 type strain from Zaire and the HST strain from Japan – and limited comparative genomics studies have been conducted for HHV-6B [19–21]. These two genomes have multiple conflicting annotations for gene and protein products. In addition, the annotated gene functions are mostly based on homology from cytomegalovirus, another human betaherpesvirus. Gene boundaries, protein sequences, and diversity of strains across time, place, and iciHHV-6 status are relatively unknown. These factors are critical for being able to perform molecular mechanism studies of viral pathogenesis [22].

Given that so little is known about HHV-6 genome diversity, gene/protein annotation, and gene/protein function despite its clinical disease associations, there is an opportunity to use agnostic technologies to rapidly annotate the HHV-6 genome. Large scale genome sequencing, RNA-Seq, and ribosome profiling have previously been conducted in other human herpesviruses to discover new genes and proteins and to ascribe novel functions to known genes of these obligate intracellular parasites [23–27]. Here we report the results of the first large-scale genome sequencing effort for HHV-6B with 125 near complete genomes along with reannotation of the genome with comparative genomics, transcriptomics, and proteomics. The results reveal limited sequence diversity among HHV-6B sequences with geographical clustering of HHV-6B sequences from acute infections and identical iciHHV-6B sequences among individuals without known recent common ancestry. RNA sequencing and shotgun proteomics combined with comparative genomic analysis enabled a consensus re-annotation of HHV-6B gene products that will serve as a resource for future clinical and basic science studies of HHV-6B.

Results

Global genomic diversity of HHV-6

In order to understand the genomic diversity of HHV-6, we performed capture sequencing of 125 strains of HHV-6B, comprised of 20 viral isolates from Japan, 35 isolates from New York, 6 strains from Uganda, and 74 strains of iciHHV-6 (64 species B, 10 species A) from HCT recipients or donors in Seattle (Fig. 1a, Table 1). The HHV-6B oligonucleotides designed for capture sequencing could retrieve > 99% of the HHV-6B genome, with less than 1% unresolved due to repetitive elements. The same panel was able to retrieve approximately 80% of the HHV-6A genome, again due to repetitive elements and in this case the reduced sequence identity with the HHV-6B oligonucleotide set (Fig. 1b). Across the HHV-6B strains, the recoverable contiguous HHV-6B unique (U) region measured 119.6 kb, the N-terminal U86 contig measured 3.1 kb, the U90/91 contig measured 6.0 kb, and the U94-U100 contig measured 10.2 kb. The 10 HHV-6A strains assembled ranged from lengths of 60 kb to 119 kb with a median length of 118 kb.

Demographic characteristics of cohorts

The median age of the roseola cohort from Japan was 12 months [8 - 24 months], the New York febrile infant cohort was 11 months [1 – 25 months], and the Uganda cohort was 25 months. All 20 patients from the two cohorts from Japan were of Japanese ancestry and all 6 patients from the Uganda cohort were Black Africans. In the New York cohort 16/35 (45.7%) of patients were Caucasian, 8/35 (22.9%) were African-American, 4/35 (11.4) were Hispanic, 1/35 (2.9%) were Asian, and 6/35 (17.1%) were of unknown ethnicity. Of the iciHHV-6 samples sequenced, 68/74 (91.9%) of patients came from the United States, while 2 patients
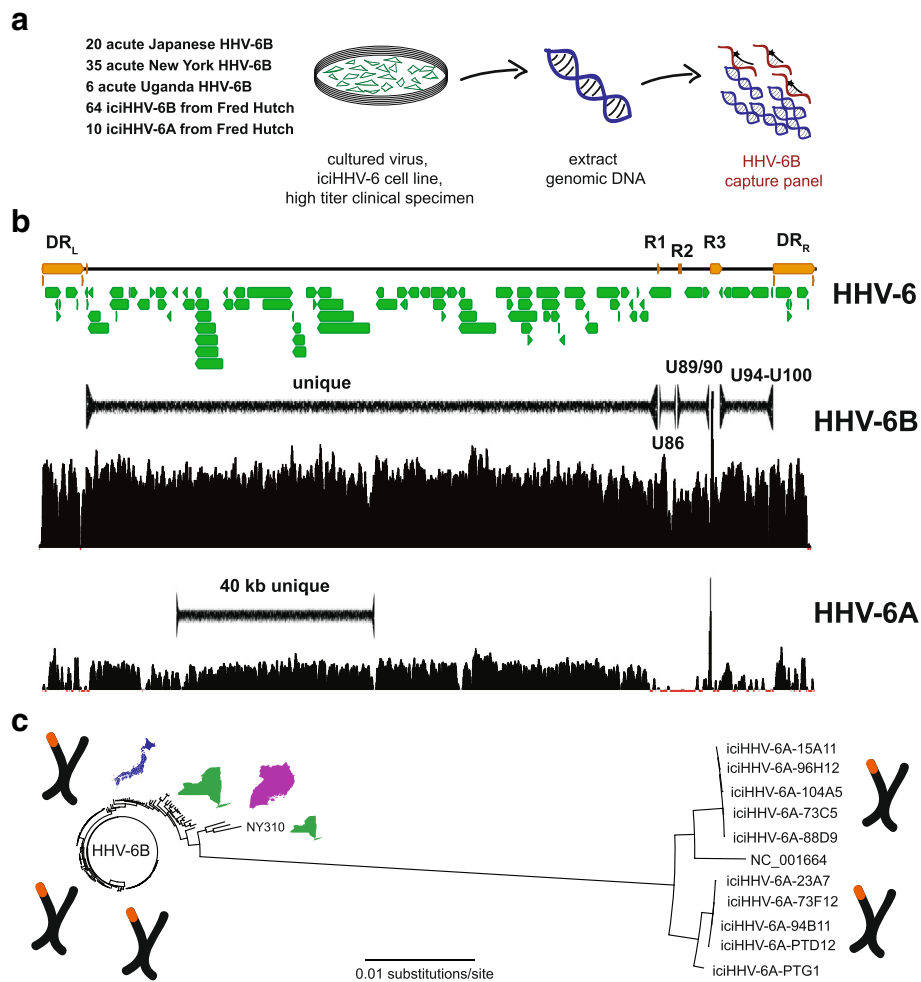
Greninger *et al. BMC Genomics* (2018) 19:204

Page 3 of 17

Fig. 1 Experimental set up and HHV-6 genome calling mock up. **a** A total of 129 HHV-6 specimens comprised of 55 cultured HHV-6B strains from acute infections, 6 clinical samples from acute infections, and 64 iciHHV-6B and 10 iciHHV-6A cell lines were sequenced using a capture panel based on the HHV-6B reference genome (NC_000898). **b** Consensus genomes used for phylogenetic analysis were called for regions outside of the repeat regions for the HHV-6B specimens, including the unique long region (119 kb), U90/91 region (6 kb, between R2 and R3 repeats), and U94-100 (10 kb, between R3 and DR-R repeat). The HHV-6B capture panel recovered much of the unique long region from HHV-6A specimens. **c** Overall phylogeny of 40.2 kb sequence that was recovered from HHV-6A and HHV-6B strains sequenced reveals separation of HHV-6A and HHV-6B as separate herpesvirus species. Location images purchased from Adobe Stock

Table 1 Summary of Samples Sequenced in This Study

Location	Number	Species	Clinical	Median Age
Japan	10	HHV-6B	Acute, BMT	28 yrs [3 - 64 yrs]
Japan	10	HHV-6B	Acute, Exanthem subitum	1 yr [8 mo - 3 yrs]
New York	35	HHV-6B	Acute, fever	11 mo [1-25 mo]
Fred Hutch	64	HHV-6B	iciHHV-6, BMT	40 yrs [1 - 68 yrs]
Fred Hutch	10	HHV-6A	iciHHV-6, BMT	57 yrs [21 - 63 yrs]
UW Virology	11	HHV-6B	reactivation, acute	unknown
Uganda	6	HHV-6B	Primary	unknown

Abbreviations: *BMT* bone marrow transplant

came from the United Kingdom, 2 patients from Germany, and 1 from Australia (Additional file 1: Table S1). The median age of iciHHV-6B individuals sequenced was 40 years [1 − 68 years] and 57 years [21 − 63 years] for iciHHV-6A individuals (Table 1).

Comparison of HHV-6A and HHV-6B

Phylogenetic analysis of a 40.2 kb segment ranging from U18 to U41 that could be captured in both the HHV-6A and HHV-6B strains sequenced in this study demonstrated separate clustering of the HHV-6A and HHV-6B strains, consistent with their designation as unique species of human herpesviruses (Fig. 1c). Recombination analyses using all 10 iciHHV-6A partial sequences, HHV-6A type strain, and 14 selected HHV-6B sequences

Greninger *et al. BMC Genomics* (2018) 19:204

Page 4 of 17

revealed no recombination sites between HHV-6A and HHV-6B sequences. Two individuals from Germany and the United States who shared no relations were found to have identical iciHHV-6A sequences. HHV-6A sequences showed little divergence in this 40.2 kb region with 98.4% of sites having no nucleotide variants. When just comparing the maximum divergence between iciHHV-6A and ici-HHV6B sequences across the 40.2 kb region, iciHHV-6A strains showed greater maximal pairwise divergence than iciHHV-6B strains (354 versus 68 SNPs).

Sequence divergence in HHV-6B

Phylogenetic analysis of the unique long region revealed a cluster of two viruses from Uganda and one from New York NY310 that comprise the most divergent HHV-6B viruses sequenced to date (Fig. 1c). NY310 most closely aligned to the Z29 strain, differing from the Z29 strain by 644 of 119,635 sites (0.54%). NY310 showed greater genetic distance to the next closest American strain NY434 (703 sites, 0.59%). This strain had no obvious unique demographic or clinical characteristics, having been derived from an 18-month old white male with fever after only 2 passages in culture (Additional file 1: Table S1). NY310 served as the outgroup for all subsequent phylogenetic analyses of HHV-6B genomes.

Overall, the 119.6 kb HHV-6B unique long contig showed remarkably little sequence divergence with 98.1% of sites being identical among all 127 HHV-6B genomes and an overall pairwise identity of > 99.9%

between strains. The prototypical typing gene U90 was the most divergent with changes at 6.14% of total sites, while U15 and B6 repeat genes had the least amount of divergence with only 0.52% and 0.41% of sites being divergent (Fig. 2). Average nucleotide diversity of the Ugandan HHV-6B U sequences was 7 times that of the iciHHV-6B strains, which had the least amount of diversity of all strains profiled (Table 2). The Japanese isolates were approximately 40% less diverse than New York isolates. Both Achaz and Tajima neutrality tests to test for non-random sequence evolution across the entire U region were strongly negative in all cohorts, likely due to the population structure and demographic history of the samples analyzed [28, 29].

Phylogenetic analysis of HHV-6B sequences from acute infections

The continual replication of HHV-6B virus present in acute infections suggests a different evolutionary history than that of iciHHV-6B sequences, which are potentially preserved over time due to the high fidelity of human genomic replication. Phylogenies of the 119.6 kb unique long region of the HHV-6B genomes revealed clustering by geography as well as by sample type (i.e. acute HHV-6B infection versus iciHHV-6B) (Fig. 3). Strains from active New York infections demonstrated the greatest amount of sequence divergence with at least two different clusters while Japanese strains all clustered together, including the HST strain reference sequence. Three of the
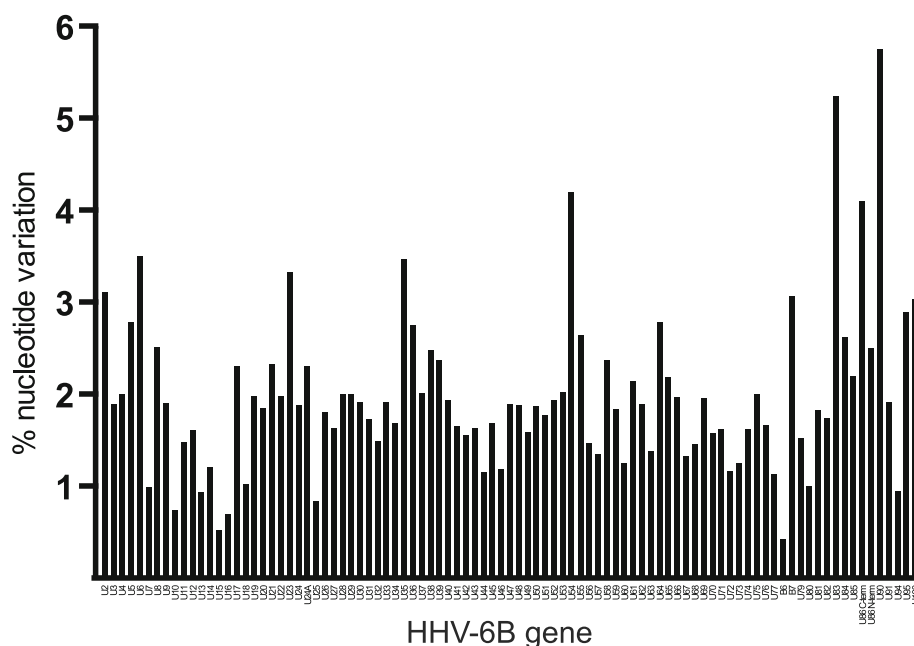

Fig. 2 Nucleotide diversity by gene. Non-identical sites listed by gene by percent of sites with any variance. The prototypical typing gene U90 is the most divergent with changes at 6.14% of total sites, while U15 and B6 repeat genes had the least amount of divergence with only 0.52% and 0.41% of sites being divergent

Greninger *et al. BMC Genomics* (2018) 19:204

Page 5 of 17

Table 2 Population genomics statistics for U region by cohort sequenced

Locus	Cohort	Samples	Nucleotide diversity	Achaz Y	Tajima's D	H-K sites
119 kb U	Japan	20	109.5	−0.75	−0.85	25
	New York	35	177.3	−0.33	−1.72	103
	Uganda	6	357.3	−0.39	−0.11	34
	iciHHV-6B	64	51.1	−0.94	−1.86	20
	All Us	127	144.9	−1.59	−2.02	162
40 kb U	11 HHV-6A 129 HHV-6B	140	256.0	−1.59	−1.47	144

Uganda sequences clustered among one of the New York strain clusters, while three others formed a unique clade (Fig. 3). Sequence from the only known patient of Asian descent from New York (NY379) fell into the Japanese cluster along with two additional New York HHV-6B strains.

Identical iciHHV-6B strains in unrelated individuals

IciHHV-6B strains showed remarkable relatedness among unrelated individuals. Across 62 of the 64 iciHHV-6B U regions sequenced here, only 334 of 119,635 (0.28%) sites had polymorphisms. Among iciHHV-6B HCT recipients whose donors were first-degree relatives, all 6 pairs had iciHHV-6B strains that were found to be identical (Fig. 3). Identical iciHHV-6B strains were also found between unrelated individuals from Germany and the United States. Notably, resequencing of several of the identical iciHHV-6B strains from unrelated individuals gave identical sequence in 11 of 12 samples, controlling for possibility of laboratory contamination or sample mix-up (Additional file 2: Figure S1). The lone outlier was a strain with a singular base with a variant allele frequency of almost exactly 50% in each sequencing replicate. Analysis of the off-target human mitochondrial reads from unrelated individuals revealed unique mitochondrial SNPs, confirming that these are from unrelated individuals (data not shown). IciHHV-6B strains were found intersperse among a New York cluster of acute infections as well as the Japanese cluster of acute infections, and several New York acute infection strains fell in the iciHHV-6B clusters. Branch lengths were generally longer for most of the acute infection strains indicating greater sequence divergence from common ancestor compared to iciHHV-6B strains.

Sequence diversity of HHV-6B in non-U regions, U90-91 and U94-100

Based on our data showing U90 to be the most divergent gene in HHV-6B, we sequenced an additional 11 U90 sequences from HHV-6 positive clinical specimens present in the UW Virology clinical lab. Phylogenies from the U90-91 and U94-100 regions revealed a similar topology

to that of the unique long phylogeny with a few notable exceptions. New York strains again showed the greatest diversity and the Japanese strains again clustered together. The U90-91 phylogeny showed two Japanese strains (B1 and B4), a New York strain (NY40), a UW clinical isolate (UW_AH1), and one iciHHV-6B strain (61C11) that clustered with the Z29 type strain and four strains with U90 sequence present in Genbank (Fig. 4, Additional file 3: Figure S2B). Additional recent UW clinical strains for which U90 sequence was available clustered throughout the HHV-6B U90 tree recovered from the whole genome sequence, with one additional prominent outgroup (UW_BF2). The Japanese and New York strains were each located in their respective unique long cluster, while the iciHHV-6B-61C11 strain fell in the Japanese cluster. The disparity between the U and U90 phylogenies is evidence of potential recombination in these strains with HHV-6B that is closer to the Zairian Z29 strain or the NY310 outgroup. Of note, the Japan-B1 strain also fell in a unique position in the U94-100 region among an iciHHV-6B cluster, while the NY40 strain was located in the U94-100 Japanese cluster (Additional file 3: Figure S2B). In addition to the phylogenetic analysis indicative of recombination, Hudson-Kaplan RM estimates of parsimonious recombination events across the U region ranged from 20 recombination sites for iciHHV-6B strains to 103 sites for New York strains (Table 2), suggesting widespread recombination within HHV-6B species. No interspecies HHV-6A x HHV-6B recombinants were observed [30].

Annotation of HHV-6B via comparative genomics

Multiple gene annotation discrepancies exist between the published HHV-6B Z29 and HST genomes. With the availability of 125 new HHV-6B genomes, we next examined the sites of these annotation discrepancies in our new HHV-6B genome sequences. For instance, the U91 gene contains an annotated splice site in the Z29 while no such annotation is found in the HST assembly. Sanger sequencing of U91 cDNA from our lab's cultured Z29 strain (Z29-1) revealed a different splice site 13 bp away from the annotated Z29 splice site, adding 5 additional amino acids to the middle of the U91 protein
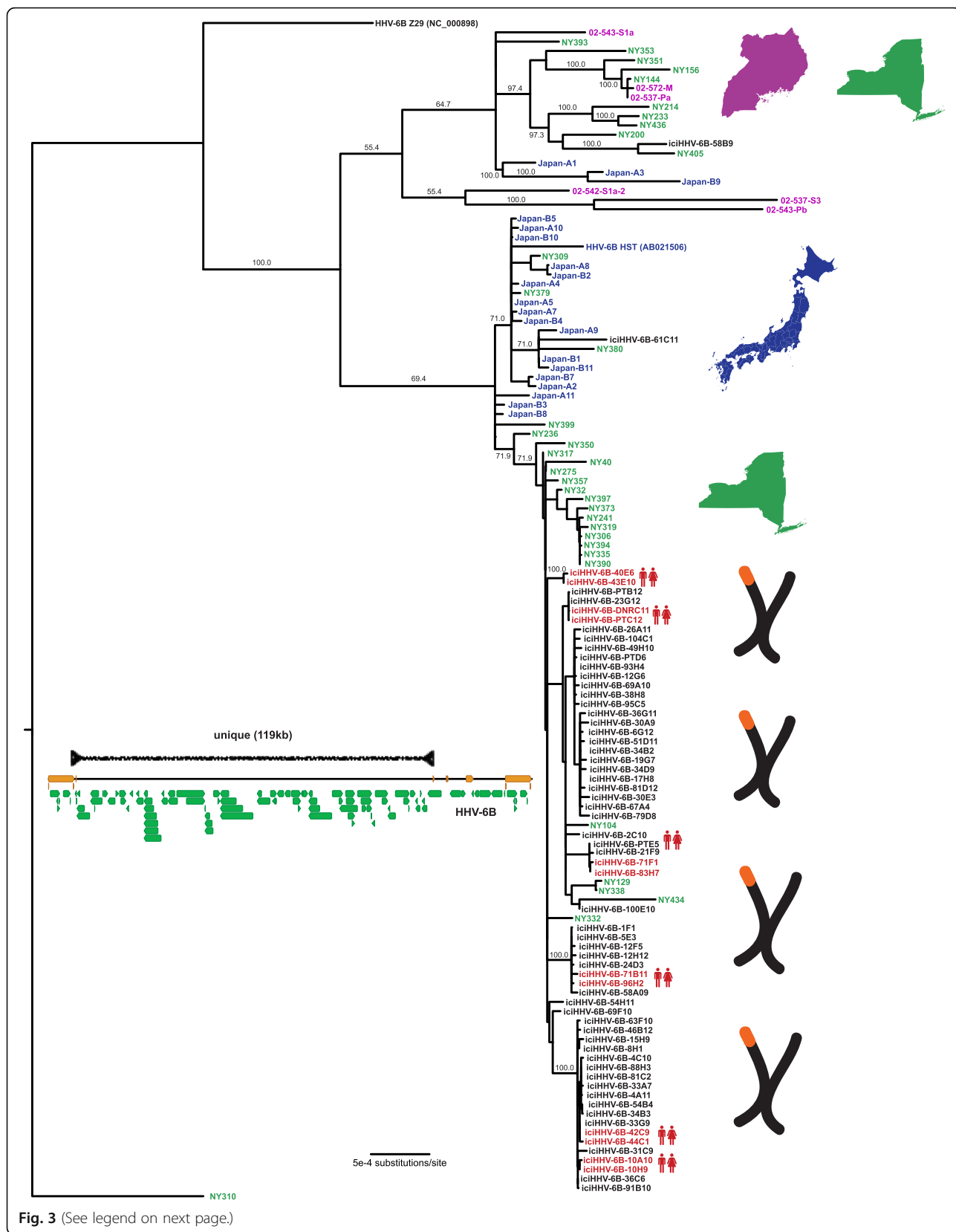
Greninger *et al. BMC Genomics* (2018) 19:204

Page 6 of 17

Fig. 3 (See legend on next page.)

Greninger *et al. BMC Genomics* (2018) 19:204

Page 7 of 17

(See figure on previous page.)
Fig. 3 Phylogenetic tree of unique long region from HHV-6B samples. HHV-6B genomes were aligned using MAFFT, curated for sequence outside of repeat regions, and phylogenetic trees were constructed using MrBayes along the 119 kb unique long region. HHV6-6B NY310 was used as an outgroup. Samples are colored and labeled for origin based on New York (green), Japan (blue), Uganda (purple), or iciHHV6 from HSCT recipients or their donors in Seattle (black), as well as whether two genomes were recovered from first-degree relatives (red). Location images purchased from Adobe Stock

(Fig. 5a). Both the cloned splice site and the annotated splice site contained canonical intronic splice sequencing (GU…AG). Cloning of the Z29-1 cDNA with the new splice site revealed an early stop codon that would disrupt the annotated C-terminal half of the protein in Z29 strains. Shotgun genomic sequencing of the cultured HHV-6B Z29-1 strain matched the Z29-1 cDNA sequence. Of note, Z29 is the only HHV-6B strain in our genomic sequencing with a single adenine insertion near the start of the second exon. Using the cloned splice site, all other U91 genes sequenced in this study would be in-frame to the end of the annotated U91, revealing that Z29 is likely unique among HHV-6B strains in missing the C-terminal half of U91.

Several other annotation discrepancies between existing HHV-6B sequences could be reconciled with our new genome sequences. The U12 gene in the Z29 strain is interrupted by a stop codon while the HST strain contains one long ORF (Fig. 5b). Comparison with the 126 U genome sequences in this study show that for U12, the HST CDS should be considered the more representative of the original two genomes. Alternatively for U27 and U52, homopolymeric SNPs in HST creates abnormally long and short annotated ORFs, respectively, that are not reflected in the newly sequenced genomes (Fig. 5c/d). Homopolymeric SNPs are also found in the U83 gene resulting in a polymorphic annotation across many of the sequence genomes (Fig. 5e).
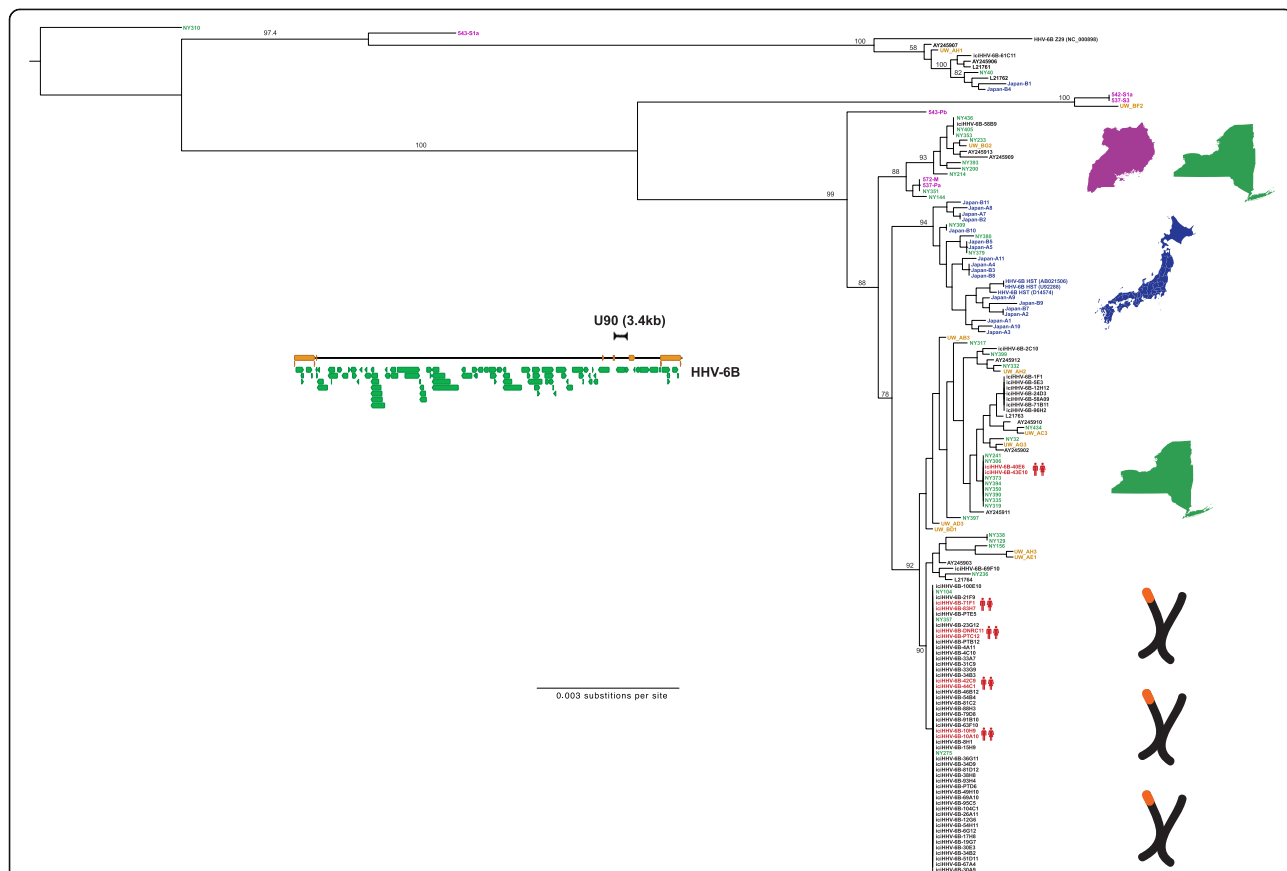

Fig. 4 Phylogenetic trees of HHV-6B samples in U90 region including UW clinical isolates. HHV-6B U90 sequence captured from the 125 complete genomes and directly PCR-amplified from the UW cohort specimens were aligned using MAFFT and phylogenetic trees were constructed using MrBayes. Samples are colored and labeled for origin based on New York (green), Japan (blue), Uganda (purple), UW Virology clinical specimens (gold), or iciHHV6/FHCRC (black), as well as whether two genomes were recovered from first-degree relatives (red). Location images purchased from Adobe Stock
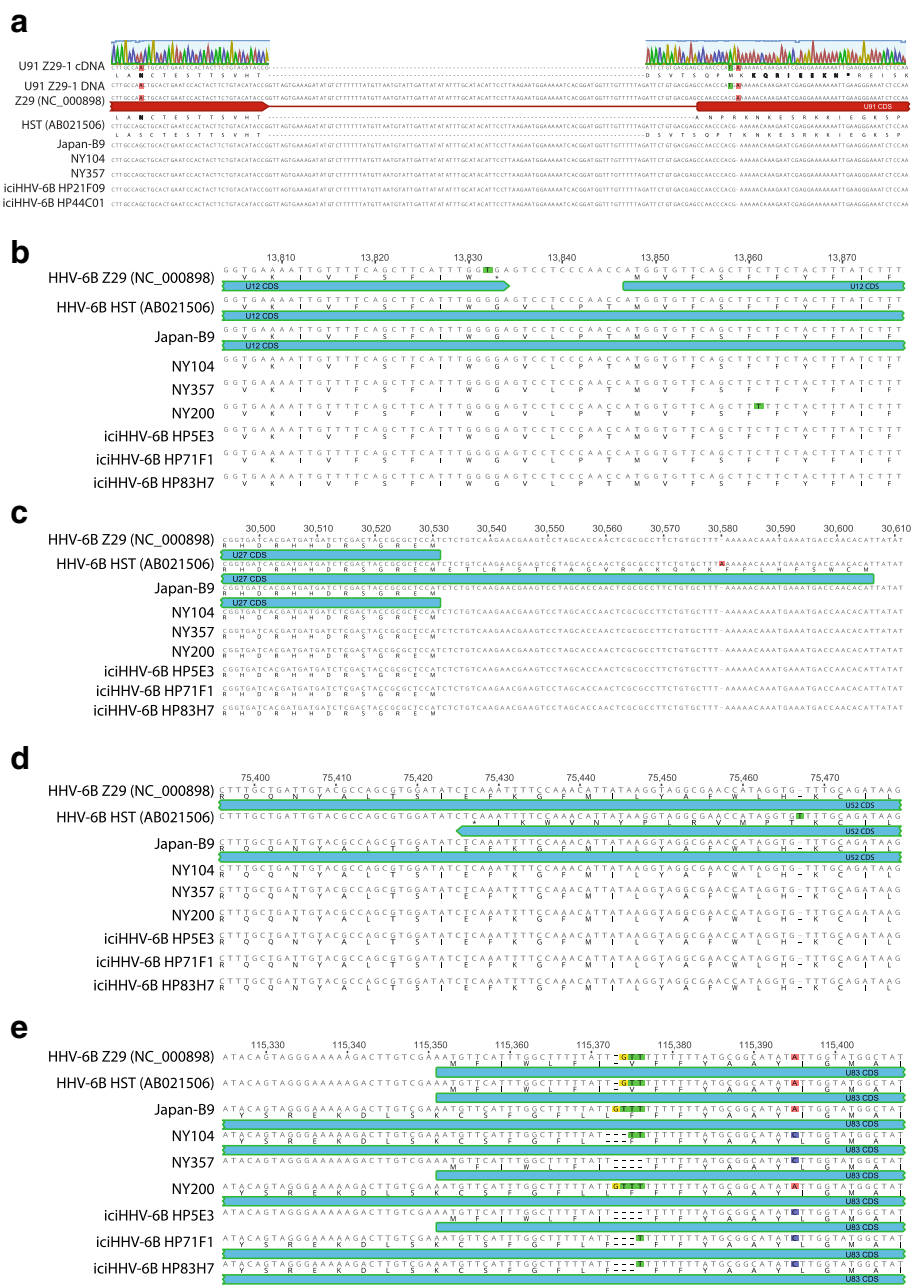
Greninger *et al. BMC Genomics* (2018) 19:204

Page 8 of 17

Fig. 5 HHV-6B annotation based on comparative genomics. Differences in annotation between HHV-6B Z29 and HST sequences are compared with a subset of the 119 genomes sequenced in this study. **a** Sanger sequencing of U91 cDNA revealed a different splice site 13-bp upstream than that which is annotated in the reference Z29 strain. The aberrant splice-site annotation in Z29 is likely due to a single base insertion found only in Z29 that alters the reading frame in the second exon. Genome sequencing of our cultured HHV-6B strain (Z29-1) confirmed the Z29-1 cDNA sequence. The reading frame depicted for Z29 is as annotated in the NCBI reference genome (NC_000898). Based on the newly discovered splice site, the Z29 U91 would contain an early stop codon while all other U91 sequences obtained in this study would continue the reading frame to the end of U91 as annotated in Z29. Several key different loci in U12 (**b**), U27 (**c**), and US52 genes (**d**) that alter the length of open reading frames in Z29 and HST are depicted. **e** A homopolymeric polymorphism in U83 changes in expected length and sequence of its open reading frame between different strains

Reannotation of HHV-6B genome through RNA-sequencing and shotgun proteomics

Based on the number of discrepancies between HST and Z29 strain annotation that could be resolved by comparative genomics, we pursued RNA sequencing of the transcriptome of the HHV-6B Z29 type strain to more exhaustively reannotate the HHV-6B genome. Two biological replicates were prepared for the HHV-6B Z29

Greninger *et al. BMC Genomics* (2018) 19:204

Page 9 of 17

RNA-Seq library in SupT1 cells and were sequenced at coverages of 266X and 3600X, while one strand-specific library was prepared for a HHV-6B Z29 infected MOLT3 cells at an average coverage of 5751X. RPKM values for HHV-6 genes from SupT1 replicates were highly reproducible (r^2 = 0.92) (Fig. 6a). Compared to the Z29 transcriptome in SupT1 cells, the Z29 transcriptome in MOLT3 cells demonstrated significantly less correlation (r^2 = 0.66) (Fig. 6b). While only 3/104 (2.9%) HHV-6B CDS had 2-fold higher expression in in SupT1 cells compared to MOLT3 cells, 19/104 (18.2%) CDS had greater expression in MOLT3 cell lines (Fig. 6c).

Analysis of the mapped reads revealed a number of novel spliceoforms that were present. All splice sites mapped were perfectly conserved in the 127 HHV-6B genomes analyzed. Five of 43 (11.6%) total splice sites recovered were non-canonical with 4/5 (80%) non-canonical splice sites occurring in U7-U9 transcripts. To validate these novel spliceoforms and extensions that affected coding sequences, we performed shotgun mass spectometry on 1D gel-separated proteins from HHV-6B Z29 cultured in SupT1 cells (Additional file 4: Figure S3, Additional file 5: Figure S4). Shotgun proteomic analysis produced 350 unique spectra covering 39 different HHV-6 proteins that may be viewed in MS Viewer (Additional file 6: Table S2 and Additional file 7: Table S3).
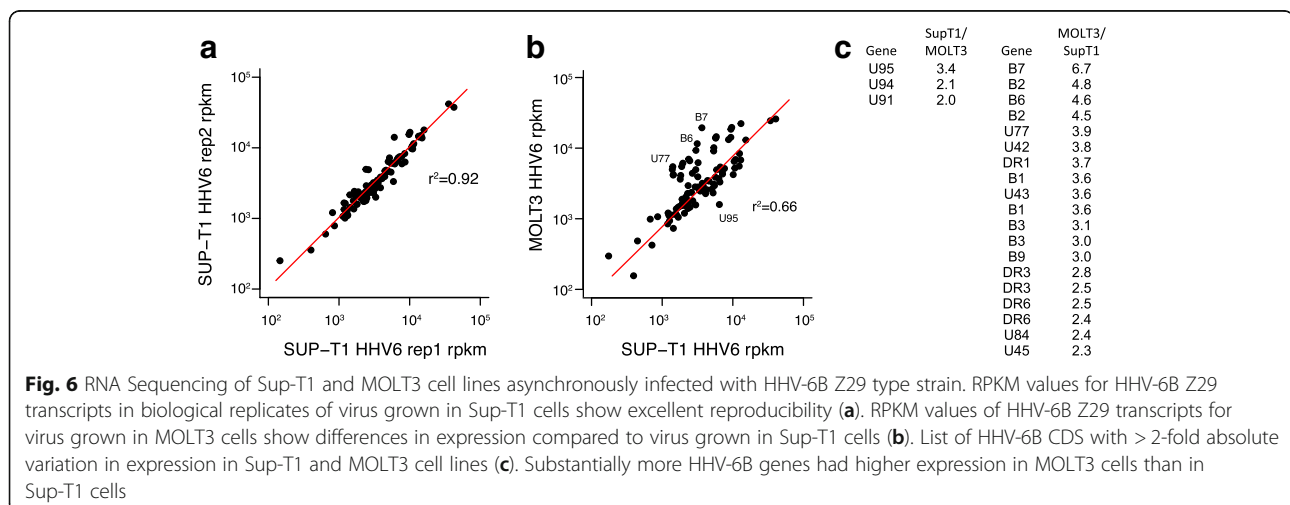
Intriguingly, three novel U79 mRNA isoforms were found, one of which also demonstrated divergent splicing based on culture in SupT1 versus MOLT3 cell lines (Fig. 7). Peptide confirmation of the novel U79 spliceoform present in SupT1 cells was confirmed with two peptides – LSTCEYLK with m/z 507.25 (2+), and YLCVR 355.68 (2+) – from shotgun proteomics analysis (Additional file 7: Table S3). The U19 gene demonstrated an unannotated splice junction just prior to the annotated stop codon, extending the C-terminus of the protein by 13 amino acids (Fig. 8). Peptides immediately before and after

the splice junction were recovered, confirming the expression of the C-terminal extension (DFLEEIAN 475.72 (2+) and SPENAVHESAAVLR 493.92 (3+) in Additional file 7: Table S3). Antisense reads along with a novel stop codon were recovered to the existing U83 annotation (Fig. 9).

Discussion

In this study we sequenced 125 HHV-6B genomes and 10 partial HHV-6A genomes, increasing the full genome data available for HHV-6 by more than an order of magnitude. We found remarkably little sequence diversity among HHV-6B strains sampled from New York, Seattle, and Japan, with the average strain having fewer than 150 differences across the 119 kb unique long region relative to any other strain sequenced here. IciHHV-6B from across the United States had considerably less diversity than other cohorts of HHV-6 sampled. HHV-6A and HHV-6B strains sequenced here showed no overlap or recombination between species and the most divergent HHV-6B strain identified to date was isolated and sequenced. Viral sequences clustered by geographical origin and identical iciHHV-6B strains were found among many apparently unrelated individuals.

These results suggest that HHV-6B integration is a relatively infrequent event, that iciHHV-6B does not general reflect strains circulating in community causing acute infection, and that sequence diversity may be driven by a founder effect. Alternatively, certain strains could be prone to integration. At the same time, iciHHV-6B sequences were found admixed with HHV-6B strains from acute infection, suggesting that integration events are not uncommon. The hypothesis that HHV-6 integration into the germline is an infrequent event, however, would be consistent with a founder effect for each clade of identical iciHHV-6B found across our North American patients and account for the identical iciHHV-6 sequences found between two pairs of
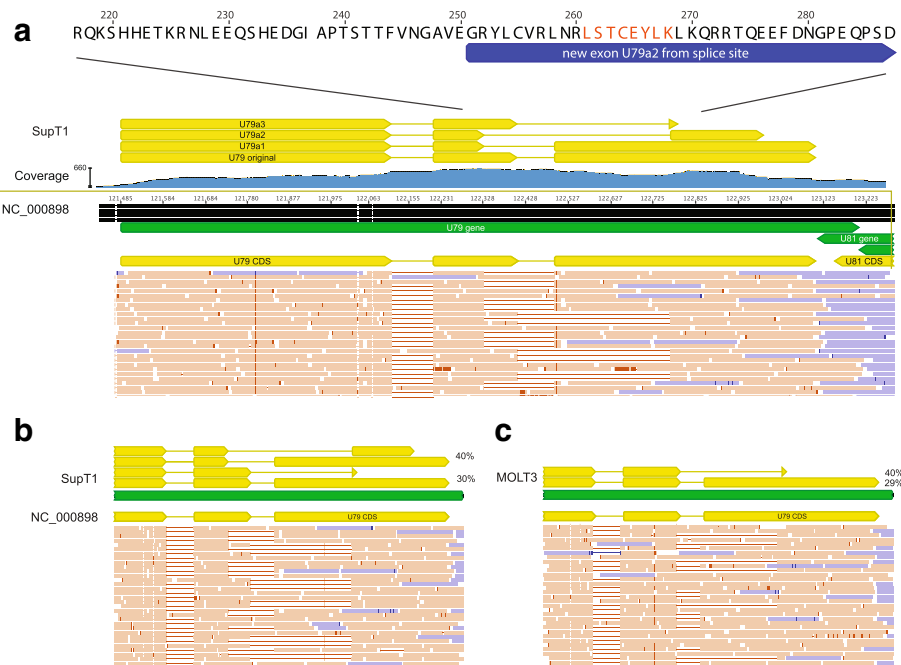

Fig. 6 RNA Sequencing of Sup-T1 and MOLT3 cell lines asynchronously infected with HHV-6B Z29 type strain. RPKM values for HHV-6B Z29 transcripts in biological replicates of virus grown in Sup-T1 cells show excellent reproducibility (**a**). RPKM values of HHV-6B Z29 transcripts for virus grown in MOLT3 cells show differences in expression compared to virus grown in Sup-T1 cells (**b**). List of HHV-6B CDS with > 2-fold absolute variation in expression in Sup-T1 and MOLT3 cell lines (**c**). Substantially more HHV-6B genes had higher expression in MOLT3 cells than in Sup-T1 cells

Greninger *et al. BMC Genomics* (2018) 19:204

Page 10 of 17

Fig. 7 Alternative and differential splicing of HHV-6B U79 transcripts in Sup-T1 versus MOLT3 cells. Strand-specific RNA sequencing reveals three additional spliceoforms of the U79 gene in HHV-6B Z29 strain cultured in Sup-T1 cells compared with the Z29 reference annotation in NC_000898 (**a**). Reads depicted in orange are positive-sense reads, while negative-sense reads are shown in blue. The highlighted peptide from the U79a2 transcript in red was confirmed by shotgun proteomics of the Sup-T1 cultured HHV-6B. While in SupT1 four total spliceoforms are found (**b**), in MOLT3 cells, only two forms of splicing in U79 are detected (**c**)

individuals from different sides of the Atlantic Ocean. It would also suggest that chromosomal integration of HHV-6 into the germ line is an extraordinarily rare event and most iciHHV-6 individuals acquired their virus from a remote integration event [31]. More sequencing of both circulating HHV-6 strains and iciHHV-6 individuals is needed to test this hypothesis and will no doubt become available as more human genomes are sequenced. The

hypothesis that integration bias due to viral sequence is the cause of the degeneracy of iciHHV-6 genomes is difficult to separate from founder effect and would only be testable in vitro or by following many individuals acutely infected with different strains of HHV-6B.

Despite widespread recombination, phylogenetic analyses demonstrated geographical clustering of HHV-6B strains with unique clades for Japanese strains and for
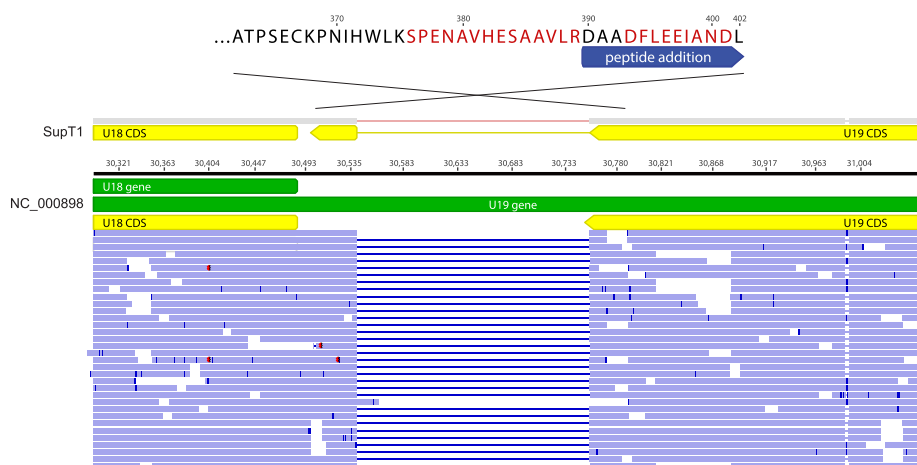

Fig. 8 Unannotated splicing leading to C-terminal extension of HHV-6B U19 protein. Strand-specific RNA sequencing of HHV-6B cultured in Sup-T1 cells demonstrated a novel splice site at the 3′ end of the U19 transcript in the codon immediately before the annotated stop codon. The new splice site leads to a 13 amino acid C-terminal extension, which was confirmed by shotgun proteomics
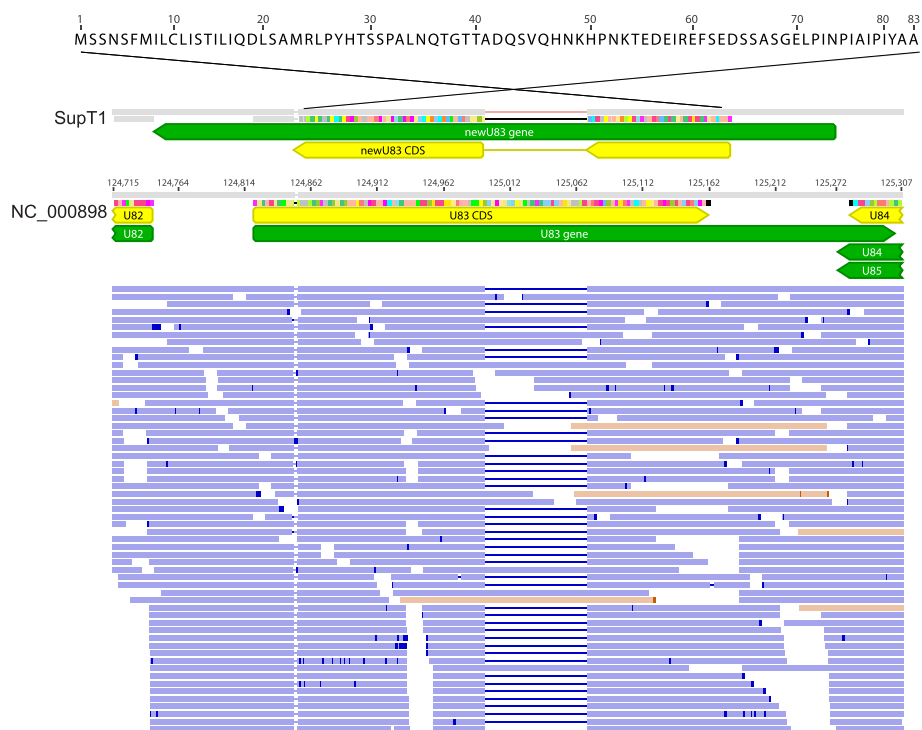
Greninger *et al. BMC Genomics* (2018) 19:204

Page 11 of 17

Fig. 9 Antisense transcription and novel splicing of HHV-6B U83 gene. Nearly all of the strand-specific RNA-seq reads from Sup-T1 cells at the annotated HHV-6B Z29 U83 gene were antisense to the existing annotation and included a novel splice site. The same splice site in the context of antisense transcript predominance was recovered from virus cultured in MOLT3 cells. No high-confidence peptides were recovered to this alternatively spliced antisense transcript by shotgun proteomics

several of the New York strains. Of note, the only patient of Asian descent in the New York cohort aligned best to the Japanese strains. These data would be consistent with the hypothesis of a familial source of transmission of acute HHV-6B. Because of the clustering of New York and Japan HHV-6 sequences, we are unable to ascertain whether strain differences can account for the striking differences in reported rates of encephalitis in infants with primary HHV-6 infection between Japan and the United States [32].

The geographical cluster of HHV-6B is similar to that seen for HSV-1 and HSV-2 genome sequences, which also show high degrees of interspecies recombination [25, 26, 33]. The limited diversity of HHV-6B as measured by average pairwise nucleotide diversity is comparable to that found in HSV-2 in contrast to that identified in HSV-1 strains [25, 26]. Of note, the diversity seen in HHV-6B is substantially less than that seen for the phylogenetically related human betaherpesvirus CMV (HHV-5) [27]. No comparative genomics have been performed to date on the other human betaherpesvirus HHV-7.

Limitations of our approach include the limited worldwide sampling of HHV-6B strains, which included the Uganda, Japan, and the United States (with samples in the iciHHV-6 Fred Hutchinson cohort coming from

several northern European individuals and only one Australian individual). Of note, our North American iciHHV-6 sequencing included individuals from at least 25 different states. More strains from both acutely infected and iciHHV-6 individuals are needed from Asia, the Middle East, South America, and Africa. Given the diversity seen in a limited subset of Ugandan strains and the limited diversity seen in iciHHV-6 in our study, it would be worthwhile to sequence iciHHV-6 from African populations to test hypotheses on the contribution of founder effect and strain sequence effects on HHV-6 integration. Sequencing of the U90 gene from reactivated HHV-6B strains from our clinical lab revealed additional lineages of HHV-6, which were subsequently confirmed by sequencing Ugandan HHV-6 isolates. Our clinical U90 sequences indicate even more lineages exist that we have not sampled on a genome-wide basis.

We also were not able to sequence through every repeat in the virus and thus our estimates of diversity would be biased to the null given that the repetitive elements may be one of the first sites of genome evolution. We also were not able to recover near-complete genomes of HHV-6A due to the use of a HHV-6B capture panel for sequencing. Future studies should be focused on continuing to probe the global diversity of HHV-6 sequences, understanding the degree of admixture between acute

Greninger *et al. BMC Genomics* (2018) 19:204

Page 12 of 17

infections and iciHHV-6 strains, and whether genotypes identified here are associated with different clinical outcomes. Based on the results presented here, there was no clear association between viral sequence and clinical phenotypes such as CNS symptoms, although our power to detect such differences was limited. Future studies will also be required to test the contribution of human SNPs and genetic diversity to any associations found between iciHHV-6 sequences and clinical phenotypes.

Our RNA-sequencing data found novel spliceoforms and antisense transcripts in 10% of the genes currently annotated in HHV-6B Z29. These data were limited by the use of a single transcriptome replicate for MOLT3 cells, although we note biological replicates were highly correlated in SupT1 cells. Shotgun proteomic analysis recovered peptides for three changes in HHV-6B coding sequences and confirmed expression of 39 existing proteins in lytic HHV-6B infection. We also discovered differential splicing of U79 in SupT1 versus MOLT3 cells. These data allow for the most comprehensive annotation of an HHV-6 genome to date and will allow for confident study of HHV-6 protein-protein interactions [22, 34]. Certainly, more work is also required to characterize how the novel spliceoforms, extensions, and transcripts discovered here affect viral replication and gene function, and whether they are present in the many strains sequenced here.

Conclusions

The sequences recovered here represent by far the largest HHV-6 sequencing effort conducted to date and significantly increases the number of available genomes for HHV-6B. Using these data, we propose a model of intermittent de novo integration of HHV-6B into host germline cells during active infection with a large contribution of founder effect in iciHHV-6B. Our data provide a significant advance in the genomic annotation of HHV-6B, which will contribute to the detection, diversity, and control of this virus. By building consensus gene and protein annotations, immediate outcomes informed by the experiments detailed here have included the development of a HHV-6B ORFeome that will enable downstream studies in gene function and T-cell epitope and antigen discovery and the design of RT-PCR primers and RNA-ISH probes to target highly expressed gene to test clinical samples for HHV-6 reactivation in situ. These data also underscore the continual need for genome sequences to achieve consensus annotation for understanding microbial biology [35].

Methods
Collection of specimens
New York cohort
Thirty five HHV-6B viral isolates were obtained from peripheral blood samples from children under 3 years of age with acute febrile illnesses or seizures presenting to the University of Rochester Medical Center Emergency Department or ambulatory settings in Rochester, NY, as previously described [4, 18, 36, 37]. Samples from children with a known abnormality of immune function were excluded.

Peripheral blood mononuclear cells (PBMCs) were separated from EDTA anticoagulated blood samples via density gradient centrifugation (Histopaque 1077;Sigma Diagnostics, St. Louis, Mo.), and co-cultivated with stimulated cord blood mononuclear cells. Positive cultures were identified by characteristic cytopathic effect (CPE), confirmed by indirect immunofluorescent staining with monoclonal antibodies directed against HHV-6A and HHV-6B, and polymerase chain reaction, as previously described [4, 38].

Japanese cohort
HHV-6B was isolated from PBMCs obtained from 10 ES patients and 10 HSCT recipients by co-cultivation with stimulated cord blood mononuclear cells. Infected cultures were identified on the basis of cytopathic effect (i.e., characteristics of pleomorphic, balloon-like large cells). The presence of virus was confirmed by immunofluorescence staining of the co-cultures with a specific HHV-6B monoclonal antibody (OHV-3; provided by T. Okuno, Department of Microbiology, Hyogo College of Medicine, Hyogo, Japan). Co-cultivated cord blood mononuclear cells infected with the clinical isolates were stored after several passages at – 80 °C until assayed.

Uganda cohort
Saliva samples were obtained from infants in a previously described birth cohort study of primary herpesvirus infection [39]. Acute HHV-6B infection determined by weekly PCR testing of oral swabs. Whole saliva was collected every 4 months using the Salivette® collection system (Sarstedt), transferred to cryovials, and frozen at – 80 °C until assayed. The samples used for this study were from 2 infants (both 3 months old at the time of sampling), 3 older children (ages 2.1 years, 2.8 years, and 4.2 years), and 1 adult (age unknown).

IciHHV-6 cohort
Seventy four individuals with iciHHV-6A or -6B were identified as part of a continuation of a previously described study [40]. DNA was extracted from beta lymphoblastoid cell lines (LCLs) generated from Epstein-Barr virus infected peripheral blood mononuclear cells (PBMCs) obtained from hematopoietic cell transplant recipients and donors. Patients received HSCTs at Fred Hutchinson Cancer Research Center (FHCRC) in Seattle, WA. Donors were sourced from patient relatives and international bone marrow donor registries. We then used a pooling testing

Greninger *et al. BMC Genomics* (2018) 19:204

Page 13 of 17

strategy as previously described using quantitative PCR [41] and droplet digital PCR [42] to identify individuals with iciHHV-6. A conserved region of the U94 gene was amplified to distinguish between species HHV-6A and HHV-6B.

University of Washington Virology patient cohort

Samples from 21 different individuals previously found to be HHV-6 PCR positive were randomly selected from plasma submitted for testing in the Clinical Molecular Virology Laboratory at the University of Washington in 2014-2015. The majority of samples were from post-transplant-associated testing for suspected HHV-6 systemic infections. Eleven samples were from children < 16 years of age (3-16 years old), and 10 were from adults (17-51 years old). Samples were from 13 males and 8 females. Five samples had viral loads < 1000 copies/mL (910, 740, 720, 550, and 480) while the remaining viral loads ranged from 1000 to 53,000 c/mL. Of these 11 gave sufficient sequence after nested PCR to be included in downstream analyses.

DNA extraction and quantitative PCR and U90 sequencing

Approximately 5 μg of DNA were extracted from B-LCLs with iciHHV-6 and aliquoted at concentrations of ~ 200 ηg/μL. DNA from the Japan and New York strains was extracted from 200 ul of viral culture using QIAamp 96 DNA kit (Qiagen) and eluted into 100 ul of AE buffer (Qiagen). To quantify the amount of HHV-6 and human DNA, 10ul of purified DNA was used to perform real-time quantitative PCR as described previously [40]. Plasma samples from the University of Washington patient cohort were extracted using a MagnaPure LC (Roche) and MagnaPure LC DNA Isolate Kit with a starting volume of 200uL and elution volume of 100uL. The U90 locus was amplified using a nested PCR protocol as described previously [43]. Amplicons from the same patient were pooled, diluted, and next-generation sequencing libraries were created using the Nextera XT kit.

Sequencing of U91 RNA transcript

Seven million HHV6B (Z29)-infected SupT1 cells (from NIH AIDS Reagent Program) were used as starting material to create an RNA library with the Qiagen RNeasy Mini Kit according to manufacturer's instructions. Total RNA was treated with TURBO DNase I (Thermo Fisher Scientific) and then used to create a cDNA library with SuperScript II Reverse Transcriptase (Thermo Fisher Scientific) according to manufacturer's instructions. Using this cDNA as template and Platinum *Taq* DNA Polymerase High Fidelity (Thermo Fisher Scientific), the *U91* transcript was amplified by PCR with annealing temperature of 55.5 °C for 30 cycles with primers that included cloning recognition sequences

as follows: *U91* sense, 5'-GGGGACAAGTTTGTACAA AAAAGCAGGCTTCTCTGTAACACTGATCATGATG GGATATGAGGA-3'; *U91* antisense, 5'-GGGGACCACT TTGTACAAGAAAGCTGGGTCTTACACATTCATTTC AGTTTTCGGTATAATAGCCTC-3'. This PCR product was inserted into the pDONR221 cloning vector (Thermo Fisher Scientific) and Sanger sequenced using the M13F (– 21) and M13R primers.

Capture sequencing

Sequencing libraries for the New York, Japan, and iciHHV-6 cohorts were prepared using 100 ng of genomic DNA using either NEB fragmentase, end repair/dA tailing, Y-adapter ligation, and dual-index Truseq PCR based or via the Kapa HyperPlus kit, following manufacturer's protocol [44]. Approximately 60 ng of cleaned, amplified DNA library was pooled into sets of seven or eight samples based on relative viral qPCR to human beta-globin qPCR ratio, so that samples with similar relative concentrations of virus were pooled together [45]. Capture sequencing was performed following the IDT xGen protocol with the use of half the amount of blocking adapter and at least 4 h of 65C hybridization with a tiling biotinylated oligo capture library based on the reference HHV6-B genome (NC_000898). Post-capture libraries were sequenced to achieve at least 200,000 reads per sample library (at least 100X coverage based on at least 50% on-target) on a 1x180bp single-end run or on a 300x300bp paired-end run on an Illumina MiSeq.

Capture sequencing for the Uganda cohort ($n = 6$ samples) was performed using a custom-designed SureSelect[XT] oligonucleotide panel covering HHV-6 and HHV-7 genomes and sequenced using an Illumina NextSeq using a v2 300 cycle mid-output kit (2x150bp paired end) [46, 47]. Libraries were prepared as outlined in the SureSelect[XT] Automated Target Enrichment protocol version J0 (December 2016) with two minor modifications. 20 ng of total DNA was sheared prior to end-repair, A-tailing and adapter ligation (1:100 dilution). Two extra cycles of PCR were performed during library amplification prior to hybridization while four extra cycles of PCR were added to the post-hybridization amplification / indexing step.

RNA-Seq of HHV-6B Z29 strain

Total RNA was extracted from MOLT3 and Sup-T1 cells asynchronously infected with HHV-6B Z29 strain with > 10^6 copies/mL of virus in the supernatant. 3μg of total RNA was used as input for polyA-purification and strand-specific RNA-Seq libraries were prepared from using the NEBNext Ultra Directional RNA Library Prep Kit. Two libraries were prepared from infected SupT1 cells and one from infected MOLT3 cells. Transcriptome libraries were sequenced on an Illumina MiSeq using

Greninger *et al. BMC Genomics* (2018) 19:204

Page 14 of 17

multiple runs types (2x94bp, 1x188bp). RPKM values for HHV-6B genes in both SupT1 and MOLT3 cell lines are available in Additional file 6: Table S2.

Shotgun proteomics

Proteomic samples were prepared from soluble cell lysates or serum-free conditioned media from HHV6-infected Sup-T1 cells. HHV-6B quantitation in lysates was 23,683,766 copies per PCR reaction with a corresponding beta-globin copy number of 12,900 copies per reaction; HHV-6B quantitation in the serum-free media was 3,122,307 copies per reaction with a corresponding beta-globin copy number of 10,115 copies per reaction. Approximately 2-20 micrograms of protein were separated on two 10–20% Criterion Tris-HCl run in MOPS or one 4-12% Criterion Tris-HCl run in MES SDS-PAGE gels (Bio-Rad), silver stained, and gel bands were excised for mass spectrometry-based peptide sequencing as described previously [48, 49] (Additional file 4: Figure S3). Samples were digested with sequencing grade trypsin (Promega) only, or with trypsin followed by AspN (Roche) following the standard UCSF MS facility protocol (http://msf.ucsf.edu/protocols.html) [50].

Peptide sequencing was performed using an LTQ-Orbitrap Velos (Thermo) mass spectrometer, equipped with a 10,000 psi nanoACUITY (Waters) UPLC. Reversed phase liquid chromatography was performed using an EasySpray C18 column (Thermo, ES800, PepMap, 3 μm bead size, 75 μm × 15 cm). The LC was operated at 600 nL/min flow rate for loading and 300 nL/min for peptide separation over a linear gradient over 60 min from 2% to 30% acetonitrile in 0.1% formic acid. For MS/MS analysis on the LTQ Orbitrap Velos, survey scans were recorded over 350-1400 m/z range, and MS/MS HCD scans were performed on the six most intense precursor ions, with a minimum of 2000 counts. For HCD scans, isolation width was 3.0 amu, with 30% normalized collision energy. Internal recalibration to a polydimethylcyclosiloxane (PCM) ion with m/z = 445.120025 was used for both MS and MS/MS scans [51].

Mass spectrometry centroid peak lists were generated using in-house software called PAVA, and data were searched using Protein Prospector software v. 5.19.1 [52]. Data were searched with carbamidomethylation of Cys as a fixed modification, and as variable modifications, oxidation of methionine, N-terminal pyroglutamate from glutamine, start methionine processing, and protein N-terminal acetylation. Trypsin, or trypsin plus AspN specificity was chosen as appropriate for each experiment. Mass accuracy tolerance was set to 20 ppm for parent and 30 ppm for fragment masses. For protein identification, searches were performed against a 9874 entry database containing all protein sequences longer than or equal to 8 amino acids derived from

HHV-6 Z29 strain genomic sequence translated in all six reading frames combined with translated splice junctions derived from RNA-Seq data. Searches were also performed with the SwissProt human database (*downloaded September 6, 2016*) containing 20,198 entries, and fetal bovine serum (P02769) as a cell culture supplement. Databases were concatenated with matched, fully randomized versions of each database to estimate false discovery rate (FDR) [53].

The HHV-6B protein database was searched initially allowing for two missed and one non-specific cleavage to allow for peptides with alternative splicing or unpredicted start/stop sites. Standard Protein Prospector scores (minimum protein score 22, minimum peptide score 15, maximum protein expectation value 0.01 and maximum peptide expectation value 0.001) produced a 5% FDR for protein identifications. All matched HHV6 peptide spectra were manually de novo sequenced, and may be viewed with the freely available software MS-Viewer, accessible through the Protein Prospector suite of software at the following URL: http://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=ms-viewer, with the search key: 7awn6ehwzd. Raw mass spectrometry data files and peak list files have been deposited at ProteoSAFE (http://massive.ucsd.edu) with accession number MSV000081332 (Additional file 7: Table S3 Additional file 8: Table S4).

Sequence analysis

DNA Sequencing reads were quality and adapter-trimmed using Trimmomatic v0.36 and Cutadapt, de novo assembled using SPAdes v3.7 and mapped to reference genomes NC_000898 and NC_001664 using Bowtie2 [54–56]. Contigs were aligned to reference genomes using the multiple alignment program Mugsy v1.2.3 and resolved against consensus sequences from mapped reads using custom scripts in R/Bioconductor [57–59]. Final assemblies were generated after discarding any contigs with mapq <= 5. Assembled genomes were annotated using Prokka and deposited to Genbank (accession numbers in Additional file 1: Table S1).

As the sequencing length was not sufficient to regularly discern sequence in the direct repeats and across several of the smaller repeats present in the HHV-6B genome, analysis was performed on aligned sequences that were pruned to keep four non-repeat-containing regions: between R0 and R1 repeats (U), between R1 and R2A repeats (upstream and N-terminal U86 region), between R2B and R3 repeats (containing U90/91 genes), and between U94-U100 genes (Fig. 1). Population genomics analyses including nucleotide diversity estimates, Tajima's D, Achaz's Y, and Hudson-Kaplan recombination estimates were executed using the PopGenome R package [28, 29, 60]. Recombination detection analyses

Greninger *et al. BMC Genomics* (2018) 19:204

Page 15 of 17

were performed using the DualBrothers package using a window length of 800 bp and a step size of 100 bp [61].

RNA sequencing reads were trimmed using cutadapt and mapped to the HHV-6B Z29 reference genome using Geneious v9.1 read aligner with structural variant discovery (decreased gap penalty) [62]. RPKM values were calculated based on HHV-6B Z29 reference genome annotations and displayed using custom scripts in R/Bioconductor.

Additional files

Additional file 1: Table S1. List of samples sequenced in this study and associated accession numbers. (DOCX 72 kb)

Additional file 2: Figure S1. Resequencing of select iciHHV-6B specimens confirms identical sequences among unrelated patients. Samples from select iciHHV-6B specimens with identical sequences were re-extracted, re-prepared and re-sequenced from original patient material to rule out contamination or a sample specimen switch during the sequencing process. 11/12 of specimens gave identical sequence throughout the unique long region directly from de novo assembly. One specimen (iciHHV-6B-30E3) had one nucleotide change (G77564 T) upon resequencing at a base that had a G/T variant allele frequency of approximately 50% each time the sample was sequenced. (PDF 145 kb)

Additional file 3: Figure S2. Phylogenetic tree of HHV-6B complete U90/91 and U94/100 loci. HHV-6B genomes were aligned using MAFFT, curated for sequence outside of repeat regions, and phylogenetic trees were constructed using MrBayes along the 6 kb U90/91 (A), and 10 kb U94-100 (B) regions. HHV6-6B NY310 was used as an outgroup. Samples are colored and labeled for origin based on New York (green), Japan (blue), or iciHHV6-B from HSCT recipients or their donors in Seattle (black), as well as whether two genomes were recovered from first-degree relatives (red). Location images purchased from Adobe Stock. (ZIP 656 kb)

Additional file 4: Figure S3. Non-contiguous gel images of silver stain of HHV-6B Z29 lysates in SupT1 cells or serum-free supernatant run on 10-20% TrisHCl gels in MOPS buffer. (PDF 3011 kb)

Additional file 5: Figure S4. Gel image of silver stain of HHV-6B Z29 lysate in SupT1 cells or serum-free supernatant run on 4-12% TrisHCl gel in MES buffer. (PDF 1335 kb)

Additional file 6: Table S2. RPKM values for RNA-Seq data. (XLSX 51 kb)

Additional file 7: Table S3. HHV-6 Proteins Identified by Shotgun Proteomics. Mass spectrometry database search results are shown for HHV6 proteins identified using Protein Prospector v 5.19.1 as described in Methods. Data were scored at the 5% FDR with Protein and Peptide minimum scores of 22 and 15, and maximum expectation values for proteins and peptides of 0.01 and 0.001, respectively. The number of unique peptides, the peptide (or spectral) count, the percent sequence coverage and the best peptide expectation value are given for each protein identification, merged from all samples. (XLSX 53 kb)

Additional file 8: Table S4. HHV-6 Peptides Identified by Shotgun Proteomics. Mass spectrometry database search results are shown for HHV6 peptides identified using Protein Prospector v 5.19.1 described in Materials and Methods. The table reports the best matched peptide spectra. Provided are the mass to charge ratio (m/z), charge (z), mass error in ppm, the peptide sequence with previous and next amino acids in the sequence, variable modification, the fraction and retention time as spectrum identifiers. The start and end sequence numbers are given, along with Protein Prospector peptide score and peptide expectation value. (XLSX 88 kb)

Abbreviations
BMT: Bone marrow transplant; CMV: Cytomegalovirus; CPE: Cytopathic effect; DNA: Deoxyribonucleic acid; HCT: Hematopoietic cell transplant; HHV-6: Human herpesvirus 6; HSCT: Hematopoietic stem cell transplant; iciHHV-6: Inherited chromosomally integrated HHV-6; IRB: Institutional review board; PBMC: Peripheral blood mononuclear cells; PCR: Polymerase chain reaction; RNA: Ribonucleic acid; RPKM: Reads per kilobase of transcript per million reads; U: Unique

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Laboratory Medicine, University of Washington, Seattle, WA, USA. [2]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [3]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, USA. [4]Division of Infection and Immunity, University College London, London, UK. [5]Department of Pediatrics, University of Washington, Seattle, WA, USA. [6]University of British Columbia, BC Children's Hospital Research Institute, Vancouver, Canada. [7]Department of Pediatrics, Fujita Health University, Fujita, Toyoake, Japan. [8]University of Rochester Medical Center School of Medicine, Rochester, New York, USA.

Greninger *et al. BMC Genomics* (2018) 19:204

Page 16 of 17

References

1. Ablashi D, Agut H, Alvarez-Lafuente R, Clark DA, Dewhurst S, DiLuca D, Flamand L, Frenkel N, Gallo R, Gompels UA, Höllsberg P, Jacobson S, Luppi M, Lusso P, Malnati M, Medveczky P, Mori Y, Pellett PE, Pritchett JC, Yamanishi K, Yoshikawa T. Classification of HHV-6A and HHV-6B as distinct viruses. Arch Virol. 2014;159:863–70.
2. Braun DK, Dominguez G, Pellett PE. Human herpesvirus 6. Clin Microbiol Rev. 1997;10:521–67.
3. Millichap JG, Millichap JJ. Role of viral infections in the etiology of febrile seizures. Pediatr Neurol. 2006;35:165–72.
4. Hall CB, Long CE, Schnabel KC, Caserta MT, McIntyre KM, Costanzo MA, Knott A, Dewhurst S, Insel RA, Epstein LG. Human herpesvirus-6 infection in children. A prospective study of complications and reactivation. N Engl J Med. 1994;331:432–8.
5. Zerr DM, Meier AS, Selke SS, Frenkel LM, Huang M-L, Wald A, Rhoads MP, Nguy L, Bornemann R, Morrow RA, Corey L. A population-based study of primary human herpesvirus 6 infection. N Engl J Med. 2005;352:768–76.
6. Clark DA. Clinical and laboratory features of human herpesvirus 6 chromosomal integration. Clin Microbiol infect off Publ Eur soc Clin Microbiol. Infect Dis. 2016;22:333–9.
7. Arbuckle JH, Medveczky MM, Luka J, Hadley SH, Luegmayr A, Ablashi D, Lund TC, Tolar J, De Meirleir K, Montoya JG, Komaroff AL, Ambros PF, Medveczky PG. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. Proc Natl Acad Sci U S A. 2010;107:5563–8.
8. Arbuckle JH, Pantry SN, Medveczky MM, Prichett J, Loomis KS, Ablashi D, Medveczky PG. Mapping the telomere integrated genome of human herpesvirus 6A and 6B. Virology. 2013;442:3–11.
9. Hill JA, Magaret AS, Hall-Sedlak R, Mikhaylova A, Huang M-L, Sandmaier BM, Hansen JA, Jerome KR, Zerr DM, Boeckh M. Outcomes of hematopoietic cell transplantation using donors or recipients with inherited chromosomally integrated HHV-6. Blood. 2017;130:1062–9.
10. Sedlak RH, Hill JA, Nguyen T, Cho M, Levin G, Cook L, Huang M-L, Flamand L, Zerr DM, Boeckh M, Jerome KR. Detection of human Herpesvirus 6B (HHV-6B) reactivation in hematopoietic cell transplant recipients with inherited chromosomally integrated HHV-6A by droplet digital PCR. J Clin Microbiol. 2016;54:1223–7.
11. Wallaschek N, Gravel A, Flamand L, Kaufer BB. The putative U94 integrase is dispensable for human herpesvirus 6 (HHV-6) chromosomal integration. J Gen Virol. 2016;97:1899–903.
12. Seo S, Renaud C, Kuypers JM, Chiu CY, Huang M-L, Samayoa E, Xie H, Yu G, Fisher CE, Gooley TA, Miller S, Hackman RC, Myerson D, Sedlak RH, Kim Y-J, Fukuda T, Fredricks DN, Madtes DK, Jerome KR, Boeckh M. Idiopathic pneumonia syndrome after hematopoietic cell transplantation: evidence of occult infectious etiologies. Blood. 2015;125:3789–97.
13. Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL. Virus identification in unknown tropical febrile illness cases using deep sequencing. PLoS Negl Trop Dis. 2012;6:e1485.
14. Kawada J-I, Okuno Y, Torii Y, Okada R, Hayano S, Ando S, Kamiya Y, Kojima S, Ito Y. Identification of viruses in cases of pediatric acute encephalitis and encephalopathy using next-generation sequencing. Sci Rep. 2016;6:33452.
15. Leber AL, Everhart K, Balada-Llasat J-M, Cullison J, Daly J, Holt S, Lephart P, Salimnia H, Schreckenberger PC, DesJarlais S, Reed SL, Chapin KC, LeBlanc L, Johnson JK, Soliven NL, Carroll KC, Miller J-A, Dien Bard J, Mestas J, Bankowski M, Enomoto T, Hemmert AC, Bourzac KM. Multicenter evaluation of BioFire FilmArray meningitis/encephalitis panel for detection of Bacteria, viruses, and yeast in cerebrospinal fluid specimens. J Clin Microbiol. 2016;54:2251–61.
16. Salimnia H, Fairfax MR, Lephart PR, Schreckenberger P, DesJarlais SM, Johnson JK, Robinson G, Carroll KC, Greer A, Morgan M, Chan R, Loeffelholz M, Valencia-Shelton F, Jenkins S, Schuetz AN, Daly JA, Barney T, Hemmert A, Kanack KJ. Evaluation of the FilmArray blood culture identification panel: results of a multicenter controlled trial. J Clin Microbiol. 2016;54:687–98.
17. Green DA, Hitoaliaj L, Kotansky B, Campbell SM, Peaper DR. Clinical utility of on-demand multiplex respiratory pathogen testing among adult outpatients. J Clin Microbiol. 2016;54:2950–5.
18. Caserta MT, Hall CB, Schnabel K, McIntyre K, Long C, Costanzo M, Dewhurst S, Insel R, Epstein LG. Neuroinvasion and persistence of human herpesvirus 6 in children. J Infect Dis. 1994;170:1586–9.

19. Gompels UA, Nicholas J, Lawrence G, Jones M, Thomson BJ, Martin MED, Efstathiou S, Craxton M, Macaulay HA. The DNA sequence of human Herpesvirus-6: structure, coding content, and genome evolution. Virology. 1995;209:29–51.
20. Isegawa Y, Mukai T, Nakano K, Kagawa M, Chen J, Mori Y, Sunagawa T, Kawanishi K, Sashihara J, Hata A, Zou P, Kosuge H, Yamanishi K. Comparison of the complete DNA sequences of human Herpesvirus 6 variants a and B. J Virol. 1999;73:8053–63.
21. Dominguez G, Dambaugh TR, Stamey FR, Dewhurst S, Inoue N, Pellett PE. Human Herpesvirus 6B genome sequence: coding content and comparison with human Herpesvirus 6A. J Virol. 1999;73:8040–52.
22. Nakatsu F, Baskin JM, Chung J, Tanner LB, Shui G, Lee SY, Pirruccello M, Hao M, Ingolia NT, Wenk MR, De Camilli P. PtdIns4P synthesis by PI4KIIIα at the plasma membrane and its impact on plasma membrane identity. J Cell Biol. 2012;199:1003–16.
23. Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, Holdorf M, Weissman JS, Ganem D. KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. PLoS Pathog. 2014;10:e1003847.
24. Stern-Ginossar N, Weisburd B, Michalski A, Le VTK, Hein MY, Huang S-X, Ma M, Shen B, Qian S-B, Hengel H, Mann M, Ingolia NT, Weissman JS. Decoding human cytomegalovirus. Science. 2012;338:1088–93.
25. Szpara ML, Gatherer D, Ochoa A, Greenbaum B, Dolan A, Bowden RJ, Enquist LW, Legendre M, Davison AJ. Evolution and diversity in human herpes simplex virus genomes. J Virol. 2014;88:1209–27.
26. Newman RM, Lamers SL, Weiner B, Ray SC, Colgrove RC, Diaz F, Jing L, Wang K, Saif S, Young S, Henn M, Laeyendecker O, Tobian AAR, Cohen JI, Koelle DM, Quinn TC, Knipe DM. Genome sequencing and analysis of geographically diverse clinical isolates of herpes simplex virus 2. J Virol. 2015;89:8219–32.
27. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. PLoS Pathog. 2011;7:e1001344.
28. Achaz G. Frequency spectrum neutrality tests: one for all and all for one. Genetics. 2009;183:249–58.
29. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95.
30. Greninger A, Roychoudhury P, Makhsous N, Hanson D, Chase J, Krueger G, Xie H, Huang M-L, Saunders L, Ablashi D, Koelle DM, Cook L, Jerome KR. Copy number heterogeneity, large origin tandem repeats, and interspecies recombination in HHV-6A and HHV-6B reference strains. bioRxiv. 2017: 193805. https://www.ncbi.nlm.nih.gov/pubmed/29491155.
31. Zhang E, Bell AJ, Wilkie GS, Suárez NM, Batini C, Veal CD, Armendáriz-Castillo I, Neumann R, Cotton VE, Huang Y, Porteous DJ, Jarrett RF, Davison AJ, Royle NJ. Inherited chromosomally integrated human herpesvirus 6 genomes are ancient, intact and potentially able to reactivate from telomeres. J Virol JVI. 2017;91:01137–17.
32. Tesini BL, Epstein LG, Caserta MT. Clinical impact of primary infection with Roseoloviruses. Curr Opin Virol. 2014;9:91–6.
33. Koelle DM, Norberg P, Fitzgibbon MP, Russell RM, Greninger AL, Huang M-L, Stensland L, Jing L, Magaret AS, Diem K, Selke S, Xie H, Celum C, Lingappa JR, Jerome KR, Wald A, Johnston C. Worldwide circulation of HSV-2 × HSV-1 recombinant strains. Sci Rep. 2017;7:44084.
34. Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimermancic P, Greninger AL, Pico AR. Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions. Nat Protoc. 2014;9:2539–54.
35. Greninger AL, Messacar K, Dunnebacke T, Naccache SN, Federman S, Bouquet J, Mirsky D, Nomura Y, Yagi S, Glaser C, Vollmer M, Press CA, Kleinschmidt-DeMasters BK, Klenschmidt-DeMasters BK, Dominguez SR, Chiu CY. Clinical metagenomic identification of Balamuthia mandrillaris encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. Genome Med. 2015;7:113.
36. Hall CB, Caserta MT, Schnabel KC, Long C, Epstein LG, Insel RA, Dewhurst S. Persistence of human herpesvirus 6 according to site and variant: possible greater neurotropism of variant a. Clin Infect Dis Off Publ Infect Dis Soc Am. 1998;26:132–7.
37. Norton RA, Caserta MT, Hall CB, Schnabel K, Hocknell P, Dewhurst S. Detection of human herpesvirus 6 by reverse transcription-PCR. J Clin Microbiol. 1999;37:3672–5.

Greninger *et al. BMC Genomics* (2018) 19:204

Page 17 of 17

38. Pruksananonda P, Hall CB, Insel RA, McIntyre K, Pellett PE, Long CE, Schnabel KC, Pincus PH, Stamey FR, Dambaugh TR. Primary human herpesvirus 6 infection in young children. N Engl J Med. 1992;326:1445–50.

39. Gantt S, Orem J, Krantz EM, Morrow RA, Selke S, Huang M-L, Schiffer JT, Jerome KR, Nakaganda A, Wald A, Casper C, Corey L. Prospective characterization of the risk factors for transmission and symptoms of primary human Herpesvirus infections among Ugandan infants. J Infect Dis. 2016;214:36–44.

40. Hill JA, HallSedlak R, Magaret A, Huang M-L, Zerr DM, Jerome KR, Boeckh M. Efficient identification of inherited chromosomally integrated human herpesvirus 6 using specimen pooling. J Clin Virol Off Publ Pan Am Soc Clin Virol. 2016;77:71–6.

41. Zerr DM, Gupta D, Huang M-L, Carter R, Corey L. Effect of antivirals on human herpesvirus 6 replication in hematopoietic stem cell transplant recipients. Clin Infect Dis Off Publ Infect Dis Soc Am. 2002;34:309–17.

42. Sedlak RH, Cook L, Huang M-L, Magaret A, Zerr DM, Boeckh M, Jerome KR. Identification of chromosomally integrated human herpesvirus 6 by droplet digital PCR. Clin Chem. 2014;60:765–72.

43. Stanton R, Wilkinson GWG, Fox JD. Analysis of human herpesvirus-6 IE1 sequence variation in clinical samples. J Med Virol. 2003;71:578–84.

44. Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. J Clin Microbiol. 2015;53:1072–9.

45. Johnston C, Magaret A, Roychoudhury P, Greninger AL, Reeves D, Schiffer J, Jerome KR, Sather C, Diem K, Lingappa JR, Celum C, Koelle DM, Wald A. Dual-strain genital herpes simplex virus type 2 (HSV-2) infection in the US, Peru, and 8 countries in sub-Saharan Africa: a nested cross-sectional viral genotyping study. PLoS Med. 2017;14:e1002475.

46. Depledge DP, Palser AL, Watson SJ, Lai IY-C, Gray ER, Grant P, Kanda RK, Leproust E, Kellam P, Breuer J. Specific capture and whole-genome sequencing of viruses from clinical samples. PLoS One. 2011;6:e27805.

47. Tweedy J, Spyrou MA, Donaldson CD, Depledge D, Breuer J, Gompels UA. Complete genome sequence of the human Herpesvirus 6A strain AJ from Africa resembles strain GS from North America. Genome Announc. 2015;3. https://doi.org/10.1128/genomeA.01498-14.

48. Greninger AL, Knudsen GM, Betegon M, Burlingame AL, Derisi JL. The 3A protein from multiple picornaviruses utilizes the golgi adaptor protein ACBD3 to recruit PI4KIIIβ. J Virol. 2012;86:3605–16.

49. Greninger AL, Knudsen GM, Betegon M, Burlingame AL, DeRisi JL. ACBD3 interaction with TBC1 domain 22 protein is differentially affected by enteroviral and kobuviral 3A protein binding. MBio. 2013;4:e00098–13.

50. Hellman U, Wernstedt C, Góñez J, Heldin CH. Improvement of an "in-gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. Anal Biochem. 1995;224:451–5.

51. Olsen JV, de LMF G, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics MCP. 2005;4:2010–21.

52. Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL. In-depth analysis of tandem mass spectrometry data from disparate instrument types. Mol Cell Proteomics MCP. 2008;7:2386–98.

53. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007;4:207–14.

54. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. Journal. 2011;17:10–2.

55. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.

56. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014;30:2114–20.

57. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl. 2014;30:2068–9.

58. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80.

59. Johnston C, Magaret A, Roychoudhury P, Greninger AL, Cheng A, Diem K, Fitzgibbon MP, Huang M-L, Selke S, Lingappa JR, Celum C, Jerome KR, Wald A, Koelle DM. Highly conserved intragenic HSV-2 sequences: results from next-generation sequencing of HSV-2 UL and US regions from genital swabs collected from 3 continents. Virology. 2017;510:90–8.

60. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 1985;111:147–64.

61. Minin VN, Dorman KS, Fang F, Suchard MA. Dual multiple change-point model leads to more accurate recombination detection. Bioinforma Oxf Engl. 2005;21:3034–42.

62. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinforma Oxf Engl. 2012;28:1647–9.