Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

HHi-FiVe: A high-fidelity genetic engineering pipeline for construction of herpesvirus-based vaccines

Michael Jarvis (michael.jarvis@plymouth.ac.uk)
University of Plymouth https://orcid.org/0000-0002-0124-4061
Thekla Mauch
The Vaccine Group Ltd
Eleonore Ostermann
Heinrich Pette Institute
Yvonne Wezel
The Vaccine Group Ltd
Jenna Nichols
MRC-University of Glasgow Centre for Virus Research
Ana da Silva Filipe
MRC-University of Glasgow Centre for Virus Research
Matej Vucak
MRC-University of Glasgow Centre for Virus Research
Joseph Hughes
MRC-University of Glasgow Centre for Virus Research
Summer Henderson
The Vaccine Group Ltd
Kimberli Schmidt
UC Davis
Amitinder Kaur
University of California at Davis
Hester Nichols
The Vaccine Group Ltd
Robin Antrobus
University of Cambridge
Peter Barry
UC Davis
Andrew Davison
MRC-University of Glasgow Centre for Virus Research
Wolfram Brune
Heinrich Pette Institute

Article

Keywords:

Posted Date: October 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-993639/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Herpesvirus-based vectors are attractive for use both as conventional and as transmissible vaccines against emerging zoonoses in hard-to-reach animal populations. However, the threat of off-site mutations during genetic manipulation of vector genomes poses a significant challenge to vaccine construction. Herein, we present the HHi-FiVe (herpesvirus high-fidelity vector) construction pipeline for generating herpesvirus-based vectors by modifying bacterial artificial chromosomes (BACs) and monitoring integrity at each stage by complete genome sequencing. We used this pipeline to repair a highly mutated rhesus cytomegalovirus BAC containing an Ebola virus transgene. The vector derived from this BAC had been shown previously to protect rhesus macaques from lethal Ebola virus challenge by conventional vaccination. Repair of this BAC restored wild-type cellular tropism to the vector, which is essential for transmissible vaccination. Construction of this candidate transmissible vaccine against Ebola virus demonstrates the utility of the HHi-FiVe pipeline for creating precision-made herpesvirus-based vectors.

Introduction

Herpesvirus-based vectors show considerable promise for use as vaccines against infectious diseases. Several such vaccines have been approved for commercial use in agricultural animals, in which they are highly effective. These include a live vaccine based on bovine herpesvirus 1 (Bovilis® IBR Marker Live) to combat infectious bovine rhinotracheitis in cattle[1] and a live recombinant vaccine (VAXXITEK® HVT+IBD) based on turkey herpesvirus to counter infectious bursal disease virus in chickens[2]. Experimental herpesvirus-based vaccines have similarly shown an ability to produce substantial levels of immunity with protection against a range of targeted pathogens, including viruses such as simian immunodeficiency virus[3,4] and Ebola virus (EBOV)[5–7], bacteria such as *Mycobacterium tuberculosis*[8,9], and protozoa (*Plasmodium knowlesi*)[10].

Herpesvirus-based vectors have several key features that have encouraged their development as vaccines[11]. These include an inherently low pathogenic potential, an ability to induce durable levels of antibody-based and T cell-mediated immunity, and a potential for administration via mucosal (i.e. oral and nasal) routes[11]. The vaccines are also amenable to reuse, as prior vector-specific immunity does not prevent reinfection[3,5,11,12]. These features, combined with high host species restriction and the ability to spread among individuals, have motivated the development of transmissible herpesvirus-based vaccines for targeting emerging zoonotic pathogens in the inaccessible wildlife animal populations from which they frequently arise[13,14]. Advances in bacterial artificial chromosome (BAC)-based genetic engineering have played a large part in the development of technology for manipulating the vectors[15]. Nonetheless, compared to other vaccine modalities, the large genome sizes of herpesviruses and the potential for off-site mutation during manipulation present significant challenges to the widespread use of herpesvirus-based vectors as vaccines, especially in emerging zoonotic disease scenarios, where it is critical to respond rapidly while ensuring the accuracy of vaccine construction.

We have established a robust approach for iterative, high-fidelity genetic engineering of herpesvirus-based vectors. This approach was named the HHi-FiVe (herpesvirus high-fidelity vector) pipeline and was used to restore a total of 13 mutated or missing open-reading frames (ORFs) in a BAC containing a cytomegalovirus (CMV) genome from rhesus CMV (RhCMV) bearing a transgene expressing an EBOV antigen. This BAC (RhCMV68-1/EBOV BAC) was chosen as a starting point because the vector derived from it by transfection has been shown to protect rhesus macaques that were vaccinated subcutaneously and then challenged with normally lethal EBOV doses[5]. As anticipated, the vector reconstituted from the repaired BAC exhibited a phenotype characterized by restored epithelial cell tropism and sustained expression of the transgene (EBOV-GP). This work generated a repaired vector suitable for future model studies of animal-to-animal transmission and demonstrated the practicality of the HHi-FiVe pipeline for producing herpesvirus-based vectors for potential use as vaccines.

Results

RhCMV BACs and ORF nomenclature

Following isolation from the urine of a rhesus macaque in 1968, a parental virus (RhCMV strain 68-1; $RhCMV_{68\text{–}1}$) was subjected to extensive and largely undocumented passage in cultured fibroblasts of human or rhesus macaque origin[16]. A stock of the resulting virus was used to construct a primary BAC ($RhCMV_{68\text{–}1}$ BAC)[17], from which all available RhCMV BACs are derived. A succession of

studies has shown that $RhCMV_{68-1}$ and $RhCMV_{68-1}$ BAC are highly mutated[18–23], with the detail having been revealed progressively by the genome sequences of $RhCMV_{68-1}$, $RhCMV_{68-1}$ BAC, derivatives of $RhCMV_{68-1}$ BAC, viruses generated from $RhCMV_{68-1}$-based BACs, and other RhCMV strains (Table 1). As a result, $RhCMV_{68-1}$ and $RhCMV_{68-1}$ BAC lack the functions of many genes required for cellular tropism and fitness *in vivo*. It was necessary to repair these mutations in order to create a candidate for testing as a transmissible vaccine. Achieving this involved making multiple small- and large-scale repairs to $RhCMV_{68-1}$/EBOV BAC and carrying out Illumina-based complete genome sequencing at each stage to monitor fidelity.

Table 1

RhCMV genome sequences[a].

GenBank accession no.	Parental strain	Source of sequence	Reference
AY186194.1	68-1	Isolated virus	Hansen et al., 2003[24]
JQ795930.1	68-1	BAC	Malouli et al., 2012[20]
MF468139.1	68-1	BAC-derived virus	Hansen et al., 2018[8]
MF468140.1	68-1	BAC	Hansen et al., 2018
MF468141.1	68-1	BAC-derived virus	Hansen et al., 2018
MF468142.1	68-1	BAC-derived virus	Hansen et al., 2018
MF468143.1	68-1	BAC-derived virus	Hansen et al., 2018
MF468144.1	68-1	BAC-derived virus	Hansen et al., 2018
MF468145.1	68-1	BAC-derived virus	Hansen et al., 2018
MF468146.1	68-1	BAC-derived virus	Hansen et al., 2018
MF468147.1	68-1	BAC	Hansen et al., 2018
MK937070.1	68-1	BAC	Marshall et al., 2019[42]
MN437483.1	68-1	BAC	Hansen et al., 2009[4]
MT157325.1	68-1	BAC	Taher et al., 2020[23]
MT157326.1	68-1	BAC	Taher et al., 2020
MT157327.1	68-1	BAC	Taher et al., 2020
MZ517252.1[b]	68-1	BAC	Present study
MZ517253.1[c]	68-1	BAC	Present study
DQ120516.1	180.92	Isolated virus	Rivailler et al., 2006[18]
KX689267.1	19262	Isolated virus	Burwitz et al., 2016[21]
KX689268.1	19936	Isolated virus	Burwitz et al., 2016
KX689269.1	24514	Isolated virus	Burwitz et al., 2016
MT157328.1	34844	Isolated virus	Taher et al., 2020
MT157329.1	KF03	Isolated virus	Taher et al., 2020
MT157330.1	UCD52	Isolated virus	Taher et al., 2020
MT157331.1	UCD59	Isolated virus	Taher et al., 2020
MZ517254.1[d]	180.92	Isolated virus	Present study

[a]Sequences are listed in order of parental strain and then GenBank accession no.

[b]Parental $RhCMV_{68-1}$ BAC used in the present study.

[c]$RhCMV_{68-1}$/EBOV/RL11G^{+} BAC generated in the present study.

[d]Full-length sequence generated in the present study; DQ120516.1 has a large deletion.

The original nomenclature for $RhCMV_{68-1}$ ORFs was established in 2003 and consisted of the prefix rh followed by a number (GenBank accession no. AY186194.1)[24]. This nomenclature was modified and extended in 2012 by comparison with the sequence of $RhCMV_{68-1}$ BAC (GenBank accession no. JQ795930.1)[20]. As this nomenclature related only to RhCMV and not other CMVs, a comparative analysis in 2006 of the sequence of RhCMV strain 180.92 (GenBank accession no. DQ120516.1)[18] was used to develop a partially inclusive system in which RhCMV ORFs conserved in human CMV (HCMV) were given names corresponding to those in HCMV. A fully inclusive system applying across sequenced primate CMVs was developed in 2011 (GenBank accession no. FJ483968.2)[22], when the RhCMV genome annotation was improved further and orthologous ORFs in different CMVs were denoted by the same name. The principal names were those of HCMV ORFs, supplemented by those of ORFs specific to Old World monkey CMVs, which are prefixed by the letter O. This nomenclature is used below and in the genetic map of the final product of the HHi-FiVe pipeline (Figure 1). In addition, when available, the alternative names are provided in Table 2, and the 2012 names are specified below in parentheses after first use of an inclusive name. Nucleotide descriptions are given in relation to the genome sequence regardless of ORF orientation.

Table 2
Steps in repairing inactivated ORFs in RhCMV$_{68-1}$/EBOV BAC.

Step	ORF	2003 ORF[a]	2006 ORF[b]	2012 ORF[c]	Mutation[d]	Repair[d]	Location[e]
1	UL36	rh61&rh60	rhUL36	Rh61/Rh60	Frameshifted in T_8	Replaced by T_7	48669-48675
2	UL146C	NP	NP	NP	Wholly deleted	Replaced	167033-171891
	UL146D	NP	NP	NP	Wholly deleted	Replaced	
	UL146F	NP	NP	NP	Wholly deleted	Replaced	
	UL146H	rh161	NP	Rh161	Partially deleted	Replaced	
3	RL11D	rh08	rh8	Rh08	Frameshifted in C_{11}	Replaced by CAC_8	6341-6350
4	UL119	rh152&rh151[f]	rhUL119	Rh152/Rh151	Terminated by stop codon (TCA)	Replaced by CCA	154761-154763
5	US12E	rh197	rh197	Rh197	Terminated by stop codon (CTA)	Replaced by CCA	208625-208627
6	RL11E	NA	rh8.1	Rh08.1	Frameshifted in $CCAC_{10}$	Replaced by C_{12}	6991-7002
7	RL11B	rh06	rh6	Rh06	Frameshifted in C_{11}	Replaced by TCCACCTCC	5197-5208
8	UL128	NP	rhUL128	NP	Wholly deleted	Replaced	161705-167032
	UL130	NA	rhUL130	Rh157.4	Partially deleted	Replaced	
9	RL11G	rh14	rh13.1&rh14	Rh13.1	Frameshifted due to insertion (CT)	Deleted	Between 12847 and 12848
					Frameshifted in A_8	Replaced by A_7	13059-13065

[a]Name of ORF in RhCMV$_{68-1}$, GenBank accession no. AY186194.1 (Hansen et al., 2003)[24]; NA, not annotated; NP, not present; &, separate ORFs; some ORFs are partial because of lack of recognition of errors, mutations or splicing.

[b]Name of ORF in RhCMV strain 180.92, GenBank accession no. DQ120516.1 (Rivailler et al., 2006)[18]; NP, not present; &, separate ORFs; some ORFs are partial because of lack of recognition of mutations.

[c]Name of ORF in RhCMV$_{68-1}$ BAC, GenBank accession no. JQ795930.1 (Malouli et al., 2012)[20]; NP, not present; /, spliced ORFs; some ORFs are partial because of lack of recognition of mutations.

[d]Sequences correspond to the genome sequence regardless of ORF orientation.

[e]In RhCMV$_{68-1}$/EBOV/RL11G$^+$ BAC, GenBank accession no. MZ517253.1 (present study).

[f]Apparently intended to be named rh151 but annotated rh141.

Identification of mutated ORFs in RhCMV$_{68-1}$/EBOV BAC

As noted previously, one of the complications with identifying mutations in RhCMV$_{68-1}$ and RhCMV$_{68-1}$ BAC is that the original RhCMV$_{68-1}$ sequence appears to contain numerous errors[20]. Thus, some of the differences between RhCMV$_{68-1}$ and RhCMV$_{68-1}$ BAC are due to these errors rather than to mutations generated during the construction of RhCMV$_{68-1}$ BAC. We estimated the number of such errors at 20. They include substitutions and insertions or deletions (indels) in noncoding regions, substitutions in ORFs [RL1 (Rh01), RL11A (Rh05), RL11C (Rh07), RL11D (Rh08) and UL55 (Rh89); and RL11G (Rh13.1)], introducing an in-frame stop codon, although this may have been due to a subpopulation of mutants in RhCMV$_{68-1}$ rather than an error), and frameshifts in ORFs [RL11D, COX2 (Rh10), UL34 (Rh57), UL71 (Rh100.1), US18 (Rh199) and US27D (Rh216)].

In order to ensure that the RhCMV component of the repaired $RhCMV_{68-1}$/EBOV BAC was as close in sequence as possible to the original $RhCMV_{68-1}$ genome as perceived to have existed prior to isolation and serial passage in cell culture[25], it was necessary to identify mutations in $RhCMV_{68-1}$ BAC (and hence in $RhCMV_{68-1}$/EBOV BAC) that have resulted in inactivated ORFs. This involved detailed examination of an alignment of all available RhCMV genome sequences, which at the time did not include several reported since by Taher et al (2020)[23]; these recent sequences were incorporated at the end of the study and identified no additional mutations. This comparative exercise revealed a total of 13 putatively inactivated ORFs (Table 2). They fell into two categories: (i) those terminated by in-frame stop codons due to substitutions and those truncated or extended by frameshifts due to small indels (most located within or associated with homopolynucleotide tracts), and (ii) those partly or wholly missing due to large deletions or rearrangements. Seven ORFs [RL11B (Rh06), RL11D, RL11E (Rh08.1), RL11G, UL36 (Rh61/Rh60), UL119 (Rh152/Rh151) and US12E (Rh197)] distributed across the genome were in category (i) and required small-scale repair. Six ORFs [UL128, UL130 (Rh154.7), UL146C, UL146D, UL146F and UL146H] located within a region of the genome called UL/b', which contains ORFs involved in cellular tropism and immunomodulation[19,26] were in category (ii) and required large-scale repair. These ORFs were supplemented by six other ORFs in UL/b' [UL131A (Rh157.6), UL132 (Rh160), UL148 (Rh159), UL147A, UL147 (Rh158) and UL146B (Rh158.1)] that, although intact and therefore probably not inactivated, were inverted as a block. These 19 ORFs were targeted for repair, replacement, or restoration in $RhCMV_{68-1}$/EBOV BAC.

Examination of the sequence alignment also indicated a few additional differences in six $RhCMV_{68-1}$ BAC ORFs [O3, UL41A (Rh67.1), UL45 (Rh72), UL74A, UL141 (Rh164) and US12B (Rh194)] that are not represented in $RhCMV_{68-1}$ or other RhCMV strains but caused predicted amino acid substitutions. Given the error-prone nature of the $RhCMV_{68-1}$ sequence, the reality of these differences was not certain, and they were not targeted for repair.

Pipeline for repairing mutated ORFs

Targeted genetic manipulation of herpesvirus genomes is achieved by BAC-based recombineering followed by reconstitution of virus by transfection of BACs into permissive cells[27]. Off-site mutations are a concern when manipulating such large DNA constructs and reconstituting viruses. In the past, this problem has been addressed by creating viruses from revertant BACs in order to demonstrate that the intended manipulations are genetically and phenotypically reversible. However, this approach is regarded as inadequate because it does not control for off-site mutations that arise during reconstitution of virus; in our experience, this is often when such mutations occur. It is also not practical for vaccine development because of the labor-intensiveness and limited scope of phenotypic assays. To cope with this inherent vulnerability, we coupled BAC-based recombineering with responsive Illumina-based whole genome sequencing to create the HHi-FiVe pipeline for generating and validating BACs and reconstituted viruses (Figure 2).

We set out to use this pipeline to repair the mutations in $RhCMV_{68-1}$/EBOV BAC using BAC-based recombineering[4,28,29]. Recombinant BACs were screened initially by restriction fragment length polymorphism (RFLP) analysis to screen for appropriate changes to fragment mobility (**Supplementary Figure 1**). This was followed by whole genome sequencing of recombinant BACs at each stage. Overall, the complete process was accomplished in nine steps (Table 2).

Small-scale repairs

Six inactivated ORFs (RL11B, RL11D, RL11E, UL36, UL119 and US12E) required small-scale repair (Steps 1 and 3–7). Most mutations were addressed by restoring the perceived original sequence to reinstate the integrity of the ORF. However, an initial attempt at repairing RL11B at Step 7, which consisted of removing two C residues in a C_{11} homopolynucleotide tract to restore a C_9 tract, resulted consistently in a C_{10} tract. Therefore, an alternative strategy was used that involved introducing synonomous substitutions within the tract. Repair of RL11G was also small-scale (see below).

Large-scale repairs

A total of 12 ORFs in UL/b' had undergone extensive deletion or rearrangement during passage of $RhCMV_{68-1}$, and the six ORFs that were completely or partially missing as a result were not amenable to small-scale repair. Instead, the whole region was replaced by a wild type version based on RhCMV strain 19936 (Table 1), using three synthetic DNA segments that together encompassed this region (Steps 2 and 8). The product of Step 8, which still contained two frameshift mutations in RL11G (see below), was denoted $RhCMV_{68-1}$/EBOV/$RL11G^{-}$ BAC.

Repair of RL11G

RL11G contained two separate mutations: a CT insertion in a $(CT)_2$ tract, and further downstream, an A insertion in an A_7 tract. Each mutation resulted in a frameshift, the first removing the transmembrane domain of the encoded protein and the second restoring the correct reading frame near the end of the ORF. The first mutation was predicted to have been sufficient to inactivate RL11G on its own. RL11G is an orthologue of HCMV RL13[20], which has been shown to mutate during viral growth in culture in all cell types tested[30,31]. Therefore, its repair was reserved for the final step (Step 9). This strategy was vindicated by the recent demonstration that a repaired version of RL11G in a BAC-derived version of $RhCMV_{68-1}$ mutates in rhesus fibroblast cell culture[23]. Our approach was to delete RL11G completely and insert a full-length synthetic version in which the two frameshift mutations were repaired and two substitutions unique to $RhCMV_{68-1}$ and $RhCMV_{68-1}$-based BACs were replaced. The final product was denoted $RhCMV_{68-1}$/EBOV/RL11G$^+$ BAC and was repaired in all the genes inactivated in $RhCMV_{68-1}$ BAC and $RhCMV_{68-1}$/EBOV BAC by premature termination, frameshifting, deletion, or rearrangement. As well as the intended manipulations and repairs, both $RhCMV_{68-1}$/EBOV/RL11G$^-$ and $RhCMV_{68-1}$/EBOV/RL11G$^+$ BAC had one inconsequential difference from $RhCMV_{68-1}$/EBOV: an additional G residue in a G_7-tract in one copy of the terminal direct repeat of the viral genome.

Stability of $RhCMV_{68-1}$/EBOV/RL11G$^-$ and $RhCMV_{68-1}$/EBOV/RL11G$^+$

Viruses were reconstituted by transfecting $RhCMV_{68-1}$/EBOV/RL11G$^-$ BAC or $RhCMV_{68-1}$/EBOV/RL11G$^+$ BAC into rhesus fibroblast (Telo-RF) or human epithelial (hTERT RPE-1) cells and passaging further. When the cultures exhibited full cytopathic effect, DNA was extracted from infected cells or infected cell supernatant and sequenced. For each dataset, mutations were identified by visual inspection of an alignment of sequence reads to the anticipated viral genome sequence, and their abundance was calculated by counting the proportion of reads containing the mutation. This approach allowed mutations representing major subpopulations to be identified and quantified, and permitted the prevalence of these mutations to be examined in other samples in the same passage series even if present in minor subpopulations. However, minor subpopulations that did not reach sufficient representation in any sample in the series might not have been detected. Three clones (clones 1–3) of $RhCMV_{68-1}$/EBOV/RL11G$^-$ BAC and one clone (clone 1) of $RhCMV_{68-1}$/EBOV/RL11G$^+$ BAC were transfected, the latter having been derived from one of the former $RhCMV_{68-1}$/EBOV/RL11G$^-$ BAC clones (clone 2). The sequences of the $RhCMV_{68-1}$/EBOV/RL11G$^-$ BAC clones were identical to each other. The scheme for reconstituting and passaging viruses is summarised in Figure 3 and the 16 samples sequenced (Samples A–P) are indicated by red font.

Reconstitution of $RhCMV_{68-1}$/EBOV/RL11G$^-$ BAC clone 1 in Telo-RF cells generated a 1 bp frameshifting deletion in UL128 and a 193 bp frameshifting deletion in UL116 (Rh148). At passages 1, 4 and 8 in Telo-RF cells (Samples A–C), the proportions of the UL128 mutation were 85, 99 and 100 %, respectively, and the proportions of the UL116 mutation were 4, 0 and 0 %, respectively. Virus at passage 1 in Telo-RF cells was also transferred to hTERT RPE-1 cells. At passages 2 and 5 in these cells (Samples D–E), the proportions of the UL128 mutation were 40 and 30 %, respectively, and the proportions of the UL116 mutation were 38 and 66 %, respectively. Thus, both mutations were present in passage 1 in Telo-RF cells, and the UL128 mutation was selected for in Telo-RF cells but selected against in hTERT RPE-1 cells. In contrast, whereas the UL116 mutation was selected against in Telo-RF cells, it was selected for in hTERT RPE-1 cells. Reconstitution of $RhCMV_{68-1}$/EBOV/RL11G$^-$ BAC clone 2 in Telo-RF cells generated a 1022 bp deletion truncating UL128 and UL130 and nine linked C to T substitutions in US12 (Rh190). Among the substitutions, four were synonymous, three were nonsynonymous, and two introduced in-frame stop codons. At passage 1 in Telo-RF cells (Sample F), the percentages of the UL128 and US12 mutations were both 77 %, implying that they were present in the same genome. In contrast, reconstitution of $RhCMV_{68-1}$/EBOV/RL11G$^-$ BAC clone 3 in Telo-RF cells generated no major mutations (Sample G). Selection of mutations in one or more of the three adjacent genes UL128, UL130 and UL131A is a recognised feature of RhCMV and HCMV when passaged in fibroblast cells[18,30–32]. Additional mutations may be carried fortuitously with these mutations when present in the same genome, or they may be selected independently. In contrast, UL128, UL130 and UL131A are essential for growth of HCMV and RhCMV in non-fibroblast cells because they encode a glycoprotein complex that is required for viral entry into these cells[32,33]. Consistent with this, no major mutations were generated by reconstitution of $RhCMV_{68-1}$/EBOV/RL11G$^-$ BAC clone 1 in hTERT RPE-1 cells at passages 1, 5 and 10 (Samples H–J), in a mixture of stocks from passages 3–9 in this series (Sample K) grown in hTERT RPE-1 cells, in an independent stock grown from passage 3 in this series grown in hTERT RPE-1 cells, in a mixture of hTERT RPE-1 and Telo-RF cells or in Telo-RF cells alone (Samples L–N), or in an independent stock grown from passage 1 in this series in HTERT RPE-1 cells (Sample O).

In contrast to the results obtained with the RhCMV$_{68-1}$/EBOV/RL11G$^-$ clones, reconstitution of RhCMV$_{68-1}$/EBOV/RL11G$^+$ BAC clone 1 in RPE-1 cells generated a major mutation at passage 1 in these cells consisting of a 12,778 bp sequence extending from within RL1 to close downstream from RL11H that had been replaced by a 1,786 bp bacterial sequence (Sample P). The proportion of genomes in which RL11G had not been inactivated by this indel was close to 0 %. We conclude that virus reconstituted from the RhCMV$_{68-1}$/EBOV/RL11G$^-$ BACs was genetically unstable when passaged in Telo-RF cells, accumulating mutations not only in UL128, UL130 and UL131A but also in other parts of the genome. In contrast, the genome was stable when virus was passaged in hTERT RPE-1 cells. Virus reconstituted from RhCMV$_{68-1}$/EBOV/RL11G$^+$ BAC was unstable in hTERT RPE-1 cells, in which RL11G was inactivated.

Cellular tropism of RhCMV/EBOV/RL11G$^-$

The purpose of repairing RhCMV$_{68-1}$/EBOV BAC is eventually to examine its potential as a model transmissible vaccine platform for providing protective immunity against EBOV following animal-to-animal dissemination of the vaccine. The extent to which RL11G is required for dissemination remains to be determined, but the use of virus reconstituted from RhCMV$_{68-1}$/EBOV/RL11G$^+$ BAC was precluded because of the instability of RL11G in various cell types tested following reconstitution and passage (Figure 3; data not shown), which is consistent with previous findings for RhCMV[23], and the RL11G orthologue (RL13) in HCMV[30,31]. In contrast, the genome integrity of RhCMV$_{68-1}$/EBOV/RL11G$^-$ BAC was maintained over multiple passages in hTERT RPE-1 cells. To assess cellular tropism, RhCMV$_{68-1}$/EBOV was reconstituted from RhCMV$_{68-1}$/EBOV BAC in Telo-RF cells, and RhCMV$_{68-1}$/EBOV/RL11G$^-$ was reconstituted in hTERT RPE-1 cells. Viral growth was measured in infected Telo-RF cells and hTERT RPE-1 cells. Only RhCMV$_{68-1}$/EBOV/RL11G$^-$ was able to replicate in both cell lines (Figure 4).

EBOV-GP expression by RhCMV$_{68-1}$/EBOV/RL11G$^-$

To examine transcription of EBOV-GP, hTERT RPE-1 cells were infected with RhCMV$_{68-1}$/EBOV/RL11G$^-$, total infected cell RNA was harvested at 21 d p.i., stranded RNAseq data were generated from the polyadenylated RNA fraction, and the relative proportions of sense and antisense RNAs produced from individual coding regions were calculated (**Supplementary Table 1**). Sense transcripts predominated (93.61 % of all sense and antisense transcripts combined), and EBOV-GP was the sixth most highly expressed (2.31 %) sense RNA of the 185 coding regions assessed. Transcripts from the RL11 family were notable by their generally low level of expression.

Finally, to examine translation of EBOV-GP, Telo-RF cells were infected with RhCMV$_{68-1}$ or RhCMV$_{68-1}$/EBOV/RL11G$^-$, and hTERT RPE-1 cells were infected with RhCMV$_{68-1}$/EBOV/RL11G$^-$. Immunoblotting was carried out on infected cell proteins using an EBOV-GP-specific monoclonal antibody (mAb) to detect EBOV-GP, a RhCMV UL44 protein-specific antibody to confirm viral infection, and an anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mAb to monitor cellular protein expression. EBOV-GP was expressed in increasing amounts by RhCMV$_{68-1}$/EBOV/RL11G$^-$ in Telo-RF cells at least until 7 d p.i. (Figure 5A) and in hTERT RPE-1 cells at least until 15 d p.i. (Figure 5B).

Discussion

This work is a proof-of-concept study aimed at establishing the HHi-FiVe pipeline for efficient, high-fidelity genetic manipulation of herpesvirus-based vectors as a means for providing a rapid turnaround platform for developing vaccines. Over the past decade, various recombineering methodologies have been invented that enable precise engineering of large DNA constructs on both the small and large scales. However, even following confirmation of the accuracy of the intended manipulations, off-target mutations remain a concern, especially when, as in this case, multiple iterative changes are made within a single BAC lineage. Added to this is the serious potential for mutation during reconstitution of virus from a BAC. To allay these concerns, we screened BACs initially by RFLP analysis and then assessed their full integrity by complete genome sequencing. We also sequenced various viruses reconstituted from the BACs generated in the final two steps (RhCMV$_{68-1}$/EBOV/RL11G$^-$ BAC and RhCMV$_{68-1}$/EBOV/RL11G$^+$ BAC). Since the case was complex and the pipeline was untested, RFLP screening and genome sequencing were used extensively, often with multiple clones at each step. In total, 974 BACs were subjected to RFLP analysis and 83 BACs and 16 reconstituted viruses were examined by whole genome sequencing. Although most of the repairs were achieved as intended, some were problematic (e.g., the initial attempt at repairing RL11B at Step 7) or because off-site substitutions were introduced. In addition, sequencing the reconstituted viruses provided critical

information on genetic stability, leading to the conclusion that virus reconstituted from $RhCMV_{68-1}$/EBOV/RL11G^{-}, unlike that reconstituted from $RhCMV_{68-1}$/EBOV/RL11G^{+} BAC, was stable in epithelial (hTERT RPE-1) cells.

We chose to establish the HHi-FiVe pipeline by repairing $RhCMV_{68-1}$/EBOV because of the potential of the reconstituted virus to protect rhesus macaques challenged with lethal EBOV infection by transmitted, rather than parenteral subcutaneously administered vaccination. This decision also had the advantage of involving a case in which small- and large-scale repairs were accomplished in several steps, each consisting of multiple manipulations. The success of the pipeline in this situation indicates that it is likely to be broadly applicable. The complexity of this case is inherent in the development of research on $RhCMV_{68-1}$, which, because the virus is highly mutated, lacks key phenotypic properties, including the ability to infect non-fibroblast cells, and in the fact that all available BACs are derived from this strain. This has led to previous attempts to restore wild type properties by repairing $RhCMV_{68-1}$ BAC. The resulting BACs include one in which UL36 and the region containing UL128, UL130 and UL131A were repaired[32], and one containing a full-length genome that has been repaired more extensively[23] but retains the frameshifts in RL11B, RL11D and RL11E. These repaired BACs have formed an important prelude to experimentation on the immunobiology and pathology of RhCMV in its natural host. Our repair of $RhCMV_{68-1}$/EBOV was, by contrast, vaccine-oriented and corrected all inactivated (i.e., prematurely terminated, frameshifted, deleted, or rearranged) ORFs.

In each of the instances described above, repairs were made by identifying mutated and nonmutated sequences from genome alignments of $RhCMV_{68-1}$ and other RhCMV strains. Since all the strains had been isolated in cell culture and were themselves potentially mutated, this involved a degree of interpretation, making it difficult to be sure that all mutations had been identified. A more straightforward approach would be to construct a BAC from a strain that has been sequenced directly from the host and passaged minimally, and then to repair the BAC accordingly, as has been done with HCMV[31]. However, this approach carries the inherent risk that any new BAC may represent a virus with phenotypic differences from those of $RhCMV_{68-1}$, the immunology of which has been characterized extensively during its development as a vaccine platform (further details are below). The RhCMV case also raises the further complexity that the virus reconstituted from a repaired BAC ($RhCMV_{68-1}$/EBOV/RL11G^{+}) is unstable, with mutants in RL11G quickly being selected in cell culture. Alternative solutions are to use a stable virus in which the problematic gene remains inactivated ($RhCMV_{68-1}$/EBOV/RL11G^{-}) or is placed under conditional control, as has been done for both RhCMV[23] and HCMV[31,34]. The former solution is technically the simpler, but requires the phenotype of interest not to depend on the inactivated gene. It may be practical if the mutation is simple and can revert easily under selective pressure.

Herpesvirus-based vectors are showing considerable promise for use as conventional vaccines to control multiple pathogens that heretofore have been difficult to control[3,8]. In this context, herpesvirus-based vectors, in particular those based on CMVs, have been shown to have a distinct immunological profile associated with unique T-cell based antigen recognition based on MHC-E[35]. CMV-based vaccines have also been shown to provide immunological protection when administered via direct parenteral inoculation against lethal EBOV challenge in rhesus macaques, thus providing the basis for the present study[5]. However, most highly pathogenic emerging viruses spill over into human populations from inaccessible wild animal populations[36], which poses a considerable limitation on the use of directly administered vaccines. The features of CMVs thus motivate the development of CMV-based vectors as a transmissible vaccine platform to achieve high immune coverage in such situations[6,13,14]. Our study has provided the means and experience whereby the efficient production of precision-made, genetically validated herpesvirus-based vectors can contribute to the further development of this platform. Towards this goal, the virus reconstituted from $RhCMV_{68-1}$/EBOV/RL11G^{-} BAC using the HHi-FiVe pipeline is currently being tested for its ability to act as a transmissible vaccine against lethal EBOV challenge in the rhesus macaque model.

Methods

Cell lines

Human telomerase-immortalized human retinal pigmented epithelial (hTERT RPE-1) cells (ATCC CRL-4000) and human telomerase-immortalized rhesus fibroblast cells (Telo-RF) cells[37] were maintained in complete Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % (v/v) fetal calf serum (FCS), 100 IU/ml penicillin and 100 µg/ml streptomycin at 37°C in an atmosphere of 5 % (v/v) CO_2.

$RhCMV_{68-1}$-based BACs

The parental BACs used were the original $RhCMV_{68-1}$ BAC[17] and derivative ($RhCMV_{68-1}$/EBOV BAC[5,17]) bearing EBOV-GP in place of RhCMV UL83B (Rh112). The circular sequence of $RhCMV_{68-1}$/EBOV BAC consists of the unique region (U) of the viral genome and three tandem copies of a direct repeat that forms the terminal direct repeat (TR) of the viral genome (one copy at each end), with the BAC vector inserted into a location in U between genes US1 (Rh181) and US2 (Rh182). The transgene consists of a synthetic codon-optimised version of the Zaire ebolavirus/H.sapiens-tc/COD/1976/Yambuku-Mayinga EBOV-GP ORF encoding the EBOV glycoprotein (GenBank accession no. AF086833.2) with its 3' end extended to encode a 14 amino acid residue V5 epitope tag, followed by downstream noncoding sequences. Two unintended but inconsequential differences were noted in the $RhCMV_{68-1}$/EBOV clone used for repair[5]: a non-synonymous substitution in EBOV-GP that results in an A to T amino acid substitution at codon 474 (a T residue is encoded at this position in some EBOV strains), and a noncoding substitution in the viral sequence very close to the right end of the transgene.

$RhCMV_{68-1}$/EBOV-based BAC recombineering

Mutations in $RhCMV_{68-1}$/EBOV-based BACs in *Escherichia coli* GS1783 were repaired by using lambda Red recombination and *en passant* mutagenesis as described previously[38]. For small-scale repairs, a PCR product containing the repaired sequence and a selectable marker with an adjacent I-SceI restriction site was recombined into the BAC. The selectable marker was removed by the *E. coli* GS1783-encoded I-SceI endonuclease, leaving a scarless repair. RL11G was repaired by replacing the mutated ORF with a selectable marker for kanamycin resistance and then replacing this marker by a synthetic, repaired version of RL11G by using a selectable marker for spectinomycin resistance. For large-scale repairs, a 7.3 kbp region (UL/b') [UL124 (Rh156.2) to UL145 (Rh162)] was removed using a Kan marker and then repaired in three stages by *en passant* mutagenesis using three synthetic sequences comprising a wild type version of this region based on strain 19936 (KX689268.1).

RFLP analysis

Correct lambda Red recombination and *en passant* mutagenesis were confirmed using RFLP. Mutated and repaired $RhCMV_{68-1}$/EBOV-based BACs in *E. coli* GS1783 were grown overnight in Luria-Bertani (LB) broth (Thermo Fisher Scientific) containing 17.5 µg/ml chloramphenicol. BAC DNA was extracted, digested with various restriction endonucleases, and subjected to agarose gel electrophoresis. The gels were stained with 0.5 µg/ml ethidium bromide and photographed under ultraviolet illumination to identify differences in DNA fragment mobility.

$RhCMV_{68-1}$-based BAC virus reconstitution

BAC DNA (4 µg) was transfected into resuspended cells ($1x10^6$ hTERT RPE-1 or Telo-RF cells for each well of a 6-well plate) using Genjet (SignaGen Laboratories) according to the manufacturer's instructions. The monolayers were expanded when confluency was reached and all cells showed cytopathic effect. At this time, the cell supernatant was collected, centrifuged at 5000 x g for 10min, aliquoted and store at -80°C. DNA was extracted for complete genome analysis from an aliquot of the infected cell supernatant or a fraction of the cell lysate using an innuPREP DNA Mini kit (Analytik Jena) according to the manufacturer's instructions.

Genome sequencing

Genome sequences were determined by standard techniques. Briefly, DNA (100 ng) from a BAC or a virus in infected cells or infected cell supernatant was sheared in a Covaris S220 sonicator to approximately 450 bp, and a sequencing library was prepared by carrying out seven cycles of PCR with indexed primers (New England Biolabs) using a Kapa LTP library preparation kit (Kapa Biosystems). Libraries were sequenced using MiSeq or NextSeq instruments (Illumina), generating datasets of 611,268–22,429,774 (BAC samples) or 2,767,034–9,488,118 (virus samples) paired-end 150 or 300 nucleotide (nt) reads per sample.

Low-quality reads and sequencing adapters were removed from the datasets using Trim Galore v. 0.4.0 (https://github.com/FelixKrueger/TrimGalore), and the remaining reads were aligned with an appropriate reference genome using Bowtie 2 v. 2.3.1[39]. Alignments were visualized using Tablet v. 1.19.09.03[40]. If necessary, the reference genome was corrected iteratively and fresh alignments were made. The average coverage depth of the final sequences was 136–11,175 (BAC samples) or 19–2,364 (virus samples) reads/nt. Of the BACs and viruses analysed, all but one were derived from $RhCMV_{68-1}$. The exception was sequenced to supplement Table 1, and consisted of a minor population of full-length genomes present in a stock of RhCMV strain

180.92, which consisted mainly of genomes bearing a large deletion in UL/b'[41]. In this case, DNA was isolated from virus generated by transfecting a historical stock of purified virion DNA into Telo-RF cells.

Cellular tropism analysis

Multistep growth curves were conducted in triplicate in hTERT RPE-1 cells and Telo-RF cells in 6-well plates ($5x10^4$ cells/well) and infected at the indicated multiplicity of infection (MOI) based on plaque-forming units (PFU)/cell. At 4 h post infection (p.i.), the cells were washed twice with Dulbecco's phosphate-buffered saline (DPBS), and fresh medium was added. Total supernatant was then collected at various times p.i., followed by washing the monolayers once with DPBS and adding fresh medium. Viral titer in the supernatant was determined by standard plaque assay on Telo-RF cells. Growth curves were performed at least twice.

Viral transcription analysis

Confluent hTERT RPE-1 cells in two 175 cm^2 flasks were infected with $RhCMV_{68-1}$/EBOV/RL11G$^-$ at a MOI of 0.05 PFU/cell. The medium was replaced with fresh medium at 14 d p.i. At 21 d p.i., the medium was removed, and the infected cells were washed with DPBS and trypsinized using 0.05 % (w/v) trypsin in 0.02 % (w/v) ethylenediaminetetraacetic acid (EDTA) in Hank's balanced salt solution (HBSS). The detached cells from each flask were transferred to a RNAse-free 15 ml conical tube and pelleted by centrifugation at 300 x g for 5 min. The supernatant was removed, and the pellets were stored at -80 ^oC. RNA was isolated from the pellets using a RNeasy mini kit (QIAGEN), employing additional steps for virus inactivation. These included incubating at 20 ^oC for 10 min after disrupting the cells by adding buffer RLT (QIAGEN) and incubating at 20 ^oC for 20 min after adding 70 % (v/v) ethanol to the homogenised lysate. RNA was stored at -80 ^oC.

Three separate DNA sequencing libraries were prepared from polyadenylated RNA selected from each sample using a TruSeq stranded mRNA library prep kit (Illumina) with IDT for Illumina TruSeq RNA UD indexes (Illumina). An aliquot of 500 ng of RNA was used for each library, and the TruSeq stranded mRNA protocol was followed with the exception that 12 PCR cycles were performed. The six indexed libraries were pooled and sequenced on a NextSeq 500/550 mid output kit v2.5 (300 cycles) (Illumina), generating approximately 40 million paired-end 150 nt reads per dataset.

Each dataset was quality-filtered using Trim Galore, sorted into sense and antisense reads using Samtools v. 1.13 (http://www.htslib.org) and mapped to the individual $RhCMV_{68-1}$/EBOV/RL11G$^-$ ORFs using Bowtie 2 with the 'local' option. These ORFs included that of mutated RL11G and were supplemented by the sequence encoding one long noncoding RNA (RNA4.9). The number of reads mapping to each coding region was determined by visualising the alignment using Tablet and expressed as the number of reads per kbp per million sense or antisense reads mapping to all coding regions. The relative proportion of each RNA relative to the total was then calculated as a percentage for each dataset and expressed as an average.

EBOV-GP protein expression analysis

Telo-RF cell monolayers in 6-well plates ($5x10^4$ cells/well) were infected with $RhCMV_{68-1}$ (reconstituted from $RhCMV_{68-1}$ BAC) or $RhCMV_{68-1}$/EBOV/RL11G$^-$ at an MOI of 0.2 PFU/cell, and hTERT RPE-1 cells were infected with $RhCMV_{68-1}$/EBOV/RL11G$^-$ at an MOI of 0.4 PFU/cell. At various times p.i., the medium was removed, and the monolayers were lysed in boiling 2 x SDS-PAGE sample buffer (125 mM Tris-HCl pH 6.8, 4 % (w/v) SDS, 20 % (v/v) glycerol and 10% (v/v) 2-mercaptoethanol). Equal volumes of cell lysates were subjected to SDS-PAGE followed by semi-dry transfer to nitrocellulose membranes (GE Healthcare). Primary antibodies were applied at the dilutions indicated: anti-EBOV-GP protein (mAb clone 12/1.1; courtesy of Dr Ayato Takada; 1:10,000), anti-RhCMV UL44 protein (courtesy of Dr Thomas Shenk; 1:2) and anti-GAPDH (mAb clone 14C10, Cell Signalling; 1:1000). Compared to antibodies directed against the V5 epitope tag incorporated into EBOV GP expressed by the $RhCMV_{68-1}$/EBOV vectors, the anti-EBOV-GP mAb gave consistently higher signal and was used for expression analysis. Secondary antibodies (horseradish peroxidase (HRP)-labeled; Dako) were applied at the dilutions indicated: anti-mouse HRP (1:5,000) and anti-rabbit HRP (1:5,000). The membranes were incubated overnight with primary antibody at 4° C, and then for 1 h at room temperature with the secondary antibody in TBS-T (0.1% (v/v) Tween-20, 50 mM Tris and 150 mM NaCl, pH 7.5) containing 5 % (w/v) skimmed milk, and then washed three times for 5 min with TBS-T. The target proteins on the membranes were visualised by enhanced chemiluminescence (GE Healthcare) and imaged using a Fusion Capture Advance FX7 16.15 (Peqlab) instrument.

Declarations

ACKNOWLEDGEMENTS

We thank Dr Thomas Shenk (Princeton University, NJ, USA) for providing the antibody against the RhCMV UL44 protein, Dr Ayato Takada (Hokkaido University, Sapporo, Japan) for providing the mAb against EBOV-GP, Dr Salvatore Camiolo (MRC-University of Glasgow Centre for Virus Research, Glasgow, UK) for advice on analysing transcriptomic data, and Drs Gregory Smith (Northwestern University, IL, USA) and Nikolaus Osterrieder (Institut für Virologie, Berlin, Germany) for providing the *en passant* system. This study was funded by the U.S. Defense Advanced Research Projects Agency (DARPA), under agreement number D18AC00028. The funder had no role in content of the publication or in decision to publish.

AUTHOR CONTRIBUTIONS

M.A.J, A.J.D., W.B. and P.A.B. conceived and designed the study with assistance from members of the UCD PREEMPT Consortium. M.A.J., A.J.D., W.B., and P.A.B., drafted the manuscript with assistance from E.O., T.M., J.N., V.M. and K.A.S. The experiments were performed by T.M., E.O., J.N., Y.W., S.H., A.K., H.N., and R.A. A.J.D. carried out sequence analysis with assistance from J.N., M.V., and J.H. All authors reviewed the manuscript.

DATA AVAILABILITY

The data supporting the findings of this study are available from public databases as stated or from the corresponding author upon reasonable request.

COMPETING INTERESTS

M.A.J., T.M., Y.W., S.H., and H.N are employed, at least in part, by TVG. The remaining authors declare no competing interests.

References

1. Ampe, B., *et al.* Assessment of the long-term effect of vaccination on transmission of infectious bovine rhinotracheitis virus in cattle herds hyperimmunized with glycoprotein E-deleted marker vaccine. *Am J Vet Res* **73**, 1787–1793 (2012).
2. Parker, D., de Wit, S., Houghton, H. & Prandini, F. Assessment of impact of a novel infectious bursal disease (IBD) vaccination programme in breeders on IBD humoral antibody levels through the laying period. *Vet Rec Open* **1**, e000016 (2014).
3. Hansen, S.G., *et al.* Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine. *Nature* **473**, 523–527 (2011).
4. Hansen, S.G., *et al.* Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge. *Nat Med* **15**, 293–299 (2009).
5. Marzi, A., *et al.* Cytomegalovirus-based vaccine expressing Ebola virus glycoprotein protects nonhuman primates from Ebola virus infection. *Sci Rep* **6**, 21674 (2016).
6. Tsuda, Y., *et al.* A replicating cytomegalovirus-based vaccine encoding a single Ebola virus nucleoprotein CTL epitope confers protection against Ebola virus. *PLoS Negl Trop Dis* **5**, e1275 (2011).
7. Tsuda, Y., *et al.* A cytomegalovirus-based vaccine provides long-lasting protection against lethal Ebola virus challenge after a single dose. *Vaccine* **33**, 2261–2266 (2015).
8. Hansen, S.G., *et al.* Prevention of tuberculosis in rhesus macaques by a cytomegalovirus-based vaccine. *Nat Med* **24**, 130–143 (2018).
9. Beverley, P.C., *et al.* A novel murine cytomegalovirus vaccine vector protects against Mycobacterium tuberculosis. *J Immunol* **193**, 2306–2316 (2014).
10. Hansen, S.G., *et al.* Cytomegalovirus vectors expressing Plasmodium knowlesi antigens induce immune responses that delay parasitemia upon sporozoite challenge. *PLoS One* **14**, e0210252 (2019).
11. Jarvis, M.A., Hansen, S.G., Nelson, J.A. & Fruh, K. Cytomegaloviruses: From Molecular Pathogenesis to Intervention. Vol. 2 (ed. Reddehase, M.J.) (Caister Academic Press, Norwich, UK., 2013).
12. Hansen, S.G., *et al.* Evasion of CD8+ T cells is critical for superinfection by cytomegalovirus. *Science* **328**, 102–106 (2010).
13. Murphy, A.A., Redwood, A.J. & Jarvis, M.A. Self-disseminating vaccines for emerging infectious diseases. *Expert Rev Vaccines* **15**, 31–39 (2016).

14. Nuismer, S.L., *et al.* Bayesian estimation of Lassa virus epidemiological parameters: Implications for spillover prevention using wildlife vaccination. *PLoS Negl Trop Dis* **14**, e0007920 (2020).
15. Ruzsics, Z., Borst, E.M., Brune, W. & Messerle, M. Cytomegaloviruses: From Molecular Pathogenesis to Intervention. Vol. 1 (ed. Reddehase, M.J.) 38-58 (Caister Academic Press, Norwich, UK., 2013).
16. Asher, D.M., Gibbs, C.J., Jr., Lang, D.J., Gajdusek, D.C. & Chanock, R.M. Persistent shedding of cytomegalovirus in the urine of healthy Rhesus monkeys. *Proc Soc Exp Biol Med* **145**, 794–801 (1974).
17. Chang, W.L. & Barry, P.A. Cloning of the full-length rhesus cytomegalovirus genome as an infectious and self-excisable bacterial artificial chromosome for analysis of viral pathogenesis. *J Virol* **77**, 5073–5083 (2003).
18. Rivailler, P., Kaur, A., Johnson, R.P. & Wang, F. Genomic sequence of rhesus cytomegalovirus 180.92: insights into the coding potential of rhesus cytomegalovirus. *J Virol* **80**, 4179–4182 (2006).
19. Oxford, K.L., *et al.* Protein coding content of the ULb' region of wild-type rhesus cytomegalovirus. *Virology* **373**, 181–188 (2008).
20. Malouli, D., *et al.* Reevaluation of the coding potential and proteomic analysis of the BAC-derived rhesus cytomegalovirus strain 68-1. *J Virol* **86**, 8959–8973 (2012).
21. Burwitz, B.J., *et al.* Cross-Species Rhesus Cytomegalovirus Infection of Cynomolgus Macaques. *PLoS Pathog* **12**, e1006014 (2016).
22. Davison, A.J., *et al.* Cytomegaloviruses: From Molecular Pathogenesis to Intervention. Vol. 1 (ed. Reddehase, M.J.) (Caister Academic Press, Norwich, UK., 2013).
23. Taher, H., *et al.* In vitro and in vivo characterization of a recombinant rhesus cytomegalovirus containing a complete genome. *PLoS Pathog* **16**, e1008666 (2020).
24. Hansen, S.G., Strelow, L.I., Franchi, D.C., Anders, D.G. & Wong, S.W. Complete sequence and genomic analysis of rhesus cytomegalovirus. *J Virol* **77**, 6620–6636 (2003).
25. Gill, R.B., *et al.* Coding potential of UL/b' from the initial source of rhesus cytomegalovirus Strain 68-1. *Virology* **447**, 208–212 (2013).
26. Oxford, K.L., *et al.* Open reading frames carried on UL/b' are implicated in shedding and horizontal transmission of rhesus cytomegalovirus in rhesus monkeys. *J Virol* **85**, 5105–5114 (2011).
27. Wagner, M., Ruzsics, Z. & Koszinowski, U.H. Herpesvirus genetics has come of age. *Trends Microbiol* **10**, 318–324 (2002).
28. Tischer, B.K., Smith, G.A. & Osterrieder, N. En passant mutagenesis: a two step markerless red recombination system. *Methods Mol Biol* **634**, 421–430 (2010).
29. Britt, W.J., Jarvis, M., Seo, J.Y., Drummond, D. & Nelson, J. Rapid genetic engineering of human cytomegalovirus by using a lambda phage linear recombination system: demonstration that pp28 (UL99) is essential for production of infectious virus. *J Virol* **78**, 539–543 (2004).
30. Dargan, D.J., *et al.* Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. *J Gen Virol* **91**, 1535–1546 (2010).
31. Stanton, R.J., *et al.* Reconstruction of the complete human cytomegalovirus genome in a BAC reveals RL13 to be a potent inhibitor of replication. *J Clin Invest* **120**, 3191–3208 (2010).
32. Lilja, A.E. & Shenk, T. Efficient replication of rhesus cytomegalovirus variants in multiple rhesus and human cell types. *Proc Natl Acad Sci U S A* **105**, 19950–19955 (2008).
33. Wang, D. & Shenk, T. Human cytomegalovirus virion protein complex required for epithelial and endothelial cell tropism. *Proc Natl Acad Sci U S A* **102**, 18153–18158 (2005).
34. Murrell, I., *et al.* Genetic Stability of Bacterial Artificial Chromosome-Derived Human Cytomegalovirus during Culture In Vitro. *J Virol* **90**, 3929–3943 (2016).
35. Hansen, S.G., *et al.* Broadly targeted CD8(+) T cell responses restricted by major histocompatibility complex E. *Science* **351**, 714–720 (2016).
36. Jones, K.E., *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
37. Chang, W.L., Kirchoff, V., Pari, G.S. & Barry, P.A. Replication of rhesus cytomegalovirus in life-expanded rhesus fibroblasts expressing human telomerase. *J Virol Methods* **104**, 135–146 (2002).

38. Tischer, B.K., von Einem, J., Kaufer, B. & Osterrieder, N. Two-step red-mediated recombination for versatile high-efficiency markerless DNA manipulation in Escherichia coli. *Biotechniques* **40**, 191–197 (2006).
39. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
40. Milne, I., *et al.* Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* **14**, 193–202 (2013).
41. Assaf, B.T., *et al.* Limited dissemination and shedding of the UL128 complex-intact, UL/b'-defective rhesus cytomegalovirus strain 180.92. *J Virol* **88**, 9310–9320 (2014).
42. Marshall, E.E., *et al.* Enhancing safety of cytomegalovirus-based vaccine vectors by engaging host intrinsic immunity. *Sci Transl Med* **11**(2019).

Supplementary Information

Supplementary Figure 1 is available in the Supplemental Files section.

Supplementary Figure 1. Example of RFLP analysis of repaired $RhCMV_{68\text{-}1}$/EBOV-based BACs.

RFLP analysis of the $RhCMV_{68\text{-}1}$/EBOV-based BAC in which RL11B was repaired (Step 7 in Table 2). The white dots mark fragments that are present in the parental BAC or the repaired BACs (results of three clones are shown).

Supplementary Table 1. Relative abundance of sense transcripts from $RhCMV_{68\text{-}1}$/EBOV/RL11G^{-} coding regions.

Gene[a]	Protein[b]	% transcripts[c]
UL146D	chemokine vCXCL7	14.99
UL22A	glycoprotein UL22A	11.91
UL146C	chemokine vCXCL6	7.40
RNA4.9	[long noncoding RNA]	4.41
UL132	envelope glycoprotein UL132	3.43
EBOV-GP	EBOV-GP	2.31
UL84	protein UL84	2.21
UL40	membrane glycoprotein UL40	2.00
UL30A	protein UL30A	1.96
UL99	myristylated tegument protein	1.59
UL71	tegument protein UL51	1.47
UL82	tegument protein pp71	1.06
UL80	capsid maturation protease and scaffold protein	1.05
UL148	membrane protein UL148	1.03
UL42	protein UL42	1.03
UL48A	small capsid protein	0.84
UL34	protein UL34	0.83
UL17	protein UL17	0.83
UL26	tegument protein UL26	0.81
UL85	capsid triplex subunit 2	0.74
UL115	envelope glycoprotein L	0.71
UL43	tegument protein UL43	0.70
UL122 (exon 3)	regulatory protein IE2	0.69
UL128	envelope protein UL128	0.67
UL92	protein UL92	0.67
O25	protein O25	0.66
UL94	tegument protein UL16	0.64
UL112	protein UL112	0.63
UL87	protein UL87	0.62
US12	membrane protein US12	0.59
UL32	tegument protein pp150	0.59
UL41A	protein UL41A	0.57
UL88	tegument protein UL88	0.56
RL11I	membrane protein RL11I	0.56
UL98	deoxyribonuclease	0.54
UL91	protein UL91	0.53
UL45	ribonucleotide reductase subunit 1	0.52
UL78	envelope protein UL78	0.50
UL49	protein UL49	0.50
UL74A	envelope glycoprotein 24	0.45
UL144	membrane glycoprotein UL144	0.43
UL111A	interleukin-10	0.43
UL46	capsid triplex subunit 1	0.41
UL100	envelope glycoprotein M	0.40
US30	membrane protein US30	0.39
US6C	protein US6C	0.38
UL14	membrane protein UL14	0.38
UL130	envelope glycoprotein UL130	0.36
UL44	DNA polymerase processivity subunit	0.36
UL75	envelope glycoprotein H	0.36
O12	protein O12	0.35
UL147	chemokine vCXCL2	0.33
US6	membrane glycoprotein US6	0.33
COX2	prostaglandin G/H synthase 2	0.32
UL73	envelope glycoprotein N	0.32
UL38	protein UL38	0.32
US22	tegument protein US22	0.32
UL146B	chemokine vCXCL5	0.32
UL103	tegument protein UL7	0.31
UL116	protein UL116	0.30
UL147A	membrane protein UL147A	0.27
US27A	membrane protein US27A	0.27
UL53	nuclear egress lamina protein	0.27
US27E	membrane protein US27E	0.26
UL83A	protein UL83A	0.25
UL123 (exon 3)	regulatory protein IE1	0.25
UL146F	chemokine vCXCL9	0.25
O19	protein O19	0.24
UL31	protein UL31	0.24
US17	membrane protein US17	0.23
O13	protein O13	0.22

O24	protein O24	0.22
RL11J	membrane protein RL11J	0.21
UL55	envelope glycoprotein B	0.21
UL146H	chemokine vCXCL11	0.21
UL114	uracil-DNA glycosylase	0.20
UL145	protein UL145	0.20
UL30	protein UL30	0.20
O21	protein O21	0.20
UL86	major capsid protein	0.19
UL52	DNA packaging protein UL32	0.19
RS1	tegument protein TRS1	0.19
US18	membrane protein US18	0.18
UL20	membrane protein UL20	0.18
UL33	envelope glycoprotein UL33	0.18
UL131A	envelope protein UL131A	0.18
UL54	DNA polymerase catalytic subunit	0.18
UL25	tegument protein UL25	0.17
UL36	tegument protein vICA	0.17
US27B	membrane protein US27B	0.17
RL1	protein RL1	0.17
O7	protein O7	0.17
UL35	tegument protein UL35	0.17
UL141	membrane glycoprotein UL141	0.16
RL11H	membrane protein RL11H	0.16
UL93	DNA packaging tegument protein UL17	0.16
UL21A	protein UL21A	0.15
O9	protein O9	0.15
RL11A	membrane protein RL11A	0.15
US29	membrane protein US29	0.15
O23	protein O23	0.14
US32	protein US32	0.14
UL50	nuclear egress membrane protein	0.13
UL69	multifunctional expression regulator	0.13
UL13	protein UL13	0.13
US12E	membrane protein US12E	0.13
US12	membrane protein US12A	0.13
UL117	protein UL117	0.12
UL57	single-stranded DNA-binding protein	0.12
UL37	envelope glycoprotein UL37	0.11
O14	protein O14	0.11
UL23	tegument protein UL23	0.11
UL72	deoxyuridine triphosphatase	0.11
US3	membrane glycoprotein US3	0.11
US26	protein US26	0.10
UL119	membrane glycoprotein UL119	0.10
US27D	membrane protein US27D	0.10
UL89	DNA packaging terminase subunit 1	0.10
US28	envelope protein US28	0.10
UL77	DNA packaging tegument protein UL25	0.10
RL11T	membrane protein RL11T	0.09
UL104	capsid portal protein	0.09
US11	membrane glycoprotein US11	0.08
UL27	protein UL27	0.08
UL105	helicase-primase helicase subunit	0.08
US31	protein US31	0.07
US20	membrane protein US20	0.07
US6A	membrane glycoprotein US6A	0.07
UL97	tegument serine/threonine protein kinase	0.06
US1	protein US1	0.06
UL124	membrane protein UL124	0.05
UL19	protein UL19	0.05
O15	protein O15	0.05
UL56	DNA packaging terminase subunit 2	0.05
US19	membrane protein US19	0.05
RL11G	membrane protein RL11G	0.05
O22	protein O22	0.05
UL79	protein UL79	0.05
O10	protein O10	0.05
UL51	DNA packaging protein UL33	0.05
US2	membrane glycoprotein US2	0.04
UL48	large tegument protein	0.04
US6B	membrane glycoprotein US6B	0.04

UL76	nuclear protein UL24	0.04
O4	protein O4	0.04
UL96	tegument protein UL14	0.04
UL24	tegument protein UL24	0.04
UL29	protein UL29	0.04
UL47	tegument protein UL37	0.03
UL74	envelope glycoprotein O	0.03
UL102	helicase-primase subunit	0.03
US12B	membrane protein US12B	0.03
O16	protein O16	0.02
US23	protein US23	0.02
US21	membrane protein US21	0.02
O11	protein O11	0.02
UL95	protein UL95	0.02
US12C	membrane protein US12C	0.02
O8	protein O8	0.02
US12D	membrane protein US12D	0.02
UL70	helicase-primase primase subunit	0.02
O17	protein O17	0.01
O6	protein O6	0.01
RL11S	membrane protein RL11S	0.01
O1	membrane protein O1	0.01
UL120	membrane protein UL120	0.01
UL121	membrane protein UL121	0.01
US24	tegument protein US24	0.01
O20	membrane protein O20	0.01
O5	protein O5	0.01
RL11F	membrane protein RL11F	0.00
O3	protein O3	0.00
RL11R	membrane protein RL11R	0.00
RL11Q	membrane protein RL11Q	0.00
RL11K	membrane protein RL11K	0.00
O18	protein O18	0.00
RL11E	membrane protein RL11E	0.00
RL11P	membrane protein RL11P	0.00
RL11B	membrane protein RL11B	0.00
RL11D	membrane protein RL11D	0.00
RL11C	membrane protein RL11C	0.00
RL11O	membrane protein RL11O	0.00
RL11L	membrane protein RL11L	0.00
RL11M	membrane protein RL11M	0.00
RL11N	membrane protein RL11N	0.00

[a]From Fig. 1 and GenBank accession no. MZ517253.1.

[b]From GenBank accession no. MZ517253.1.

[c]Number of sense transcripts from individual coding regions as a percentage of the number of sense and antisense transcripts from all coding regions. Ranked from the greatest number of transcripts to the least. Accuracy is limited by several factors. First, values for small coding regions may be depressed because of reduced mapping of reads to their ends. For example, this may lead to an underestimate of the levels of expression of UL146D and UL22A (first and second, respectively, in the list). Second, expression of upstream coding regions in a family of overlapping mRNAs sharing a 3'-end will enhance the apparent level of expression of downstream coding regions. For example, this may lead to an overestimate of the level of expression of UL146D (first in the list), which is probably 3'-coterminal with UL146C (third in the list); this would make UL22A first in the list. Third, a proportion of antisense transcripts may have been generated from overlapping noncoding regions of sense transcripts from adjacent genes. For example, UL89 (exon 1) is overlapped by the 3'-end of the more highly expressed UL94 on the opposite strand. This factor may also lead to an overestimate of the overall proportion of antisense trancripts (6.84%).

Supplementary Table 2. Primers used for repair of ORFs in $RhCMV_{68\text{-}1}$/EBOV

Name	Sequence	Purpose
Rh157.4 Rh157.5 Rh157.6 Forward	GTACGATACTGTACGGTTGTTGCATATTTTGAAGTCATAGCAGTCCTGTG TAGGGATAACAGGGTAATCGATTT	Introduction of I-SceI sequences and a kanamycin resistance gene in the shuttle plasmid containing UL124 to UL132
Rh157.4 Rh157.5 Rh157.6 Reverse	GTAGCCAGTGTTACAACCAATTAACC	
Rh157.4 Rh157.6 Forward	AGCCGTCTACATACGGACACCACCTTTATAAGTGCTTCTACAGTGGATTTATGTGCTTAACAATACGGCAAGTAAGCCTG	Amplification of fragment containing UL130 to UL131A and introduction of a kanamycin resistance gene
Rh157.4 Rh157.6 Reverse	TCCCTACTATAAATAACGTGCTG	
Rh157.5 Forward	AGCCGTCTACATACGGACAC	Amplification of fragment containing UL128 and introduction of a kanamycin resistance gene
Rh157.5 Reverse	TGTTTAGGCGGCATCCTTC	
Rh13.1 Forward	AAACAGTGCCATGACTAAGTATACGTGCTTCAGATCGATGTCTGCCTGCGTAGGGATAACAGGGTAATCGATTT	Removal of RL11G and introduction of a kanamycin resistance gene
Rh13.1 Reverse	AAACTCGCAATGGCTTCTCAGGGTAGCTTGTCGATCCGCGCGCACGTTCGGCCAGTGTTACAACCAATTAACC	

Figures

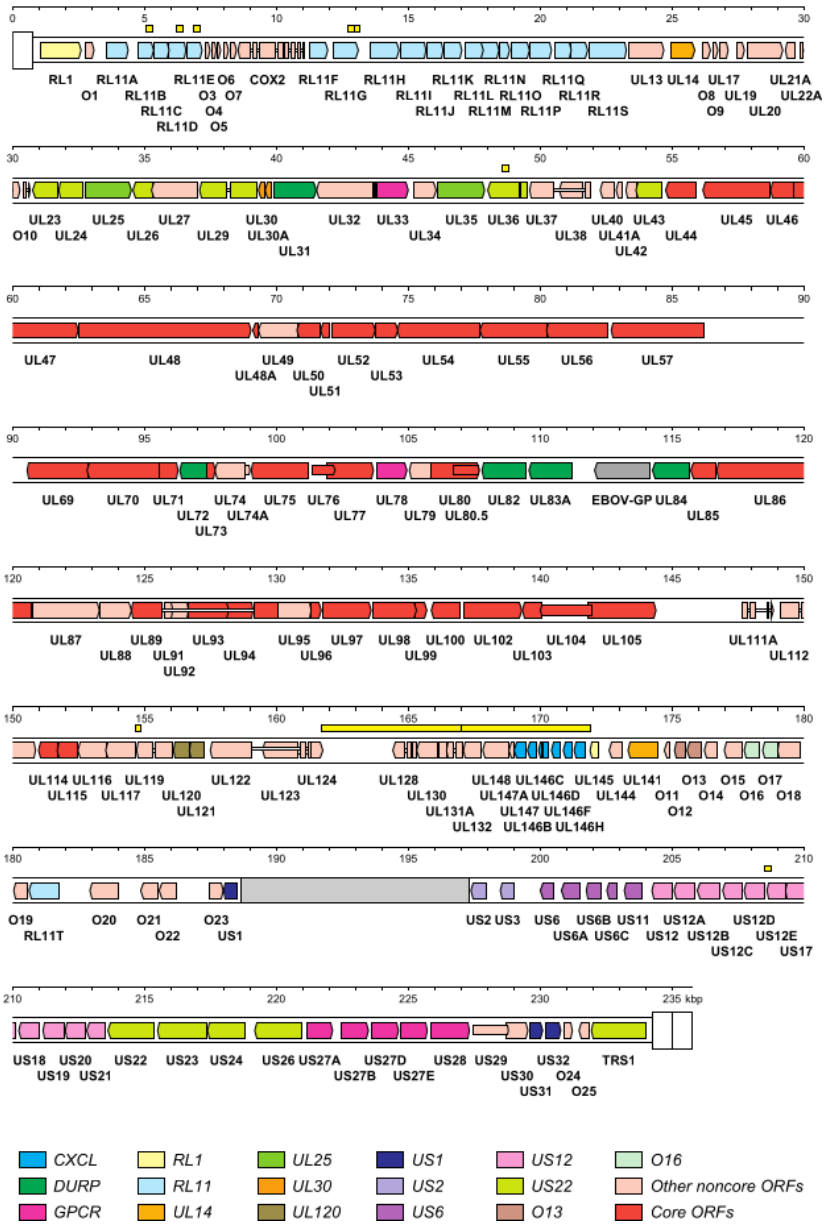

Figure 1

Genetic map of RhCMV68-1/EBOV/RL11G+ BAC. The circular sequence is depicted in linear form, starting at the left end of the viral genome with one copy of the terminal direct repeat (TR), proceeding through the unique region (U), and ending with two more copies of TR. The copies of TR are shown in a thicker format than U. Protein-coding ORFs are indicated by coloured open arrows grouped according to the key at the foot to indicate gene families, other non-core genes that are not conserved among herpesviruses and core genes that are conserved among herpesviruses. Introns connecting protein-coding regions are shown as narrow white bars. UL72 is both a core gene and member of the DURP gene family and is shown as the latter. The BAC vector is shown by the grey-shaded region between US1 and US2. The EBOV-GP ORF encoding a V5-tagged EBOV spike glycoprotein is also grey-shaded and replaced US83B. The locations of small-scale and large-scale repairs (see Table 2) are marked above the genome by yellow squares and bars, respectively.

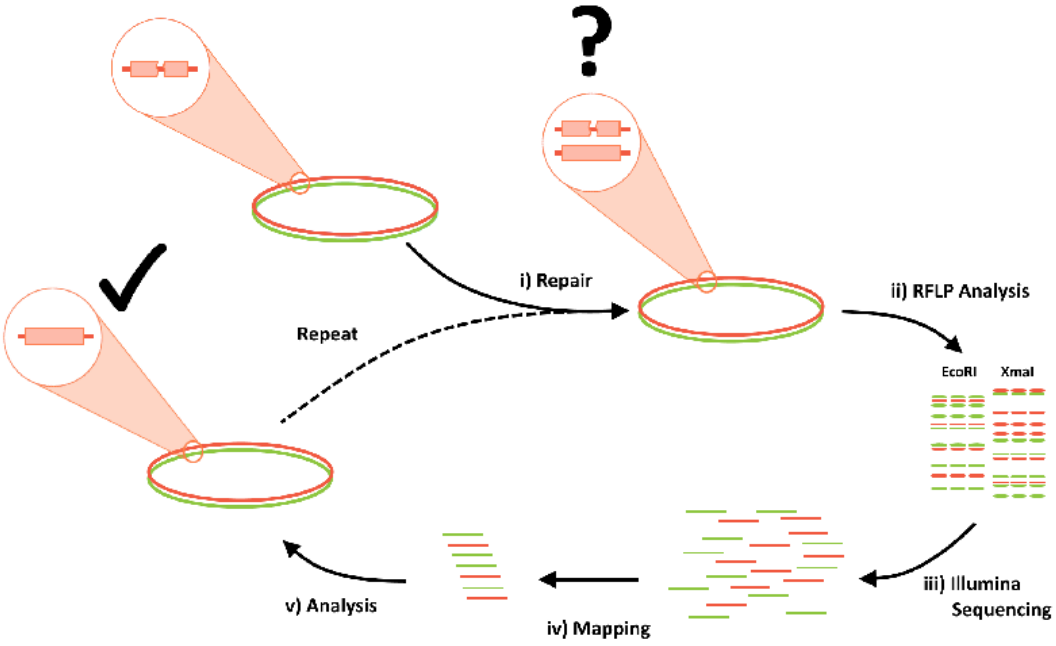

Figure 2

Schematic showing the HHi-FiVe pipeline. i) Lambda Red recombination and en passant mutagenesis are used to repair individual ORFs in the BAC, ii) the repaired BAC is screened by using RFLP, and then iii) Illumina sequencing, iv) read mapping v) accompanied by analysis are used to confirm the intended repair and the absence of off-site mutations. The process is repeated until all repairs are made.

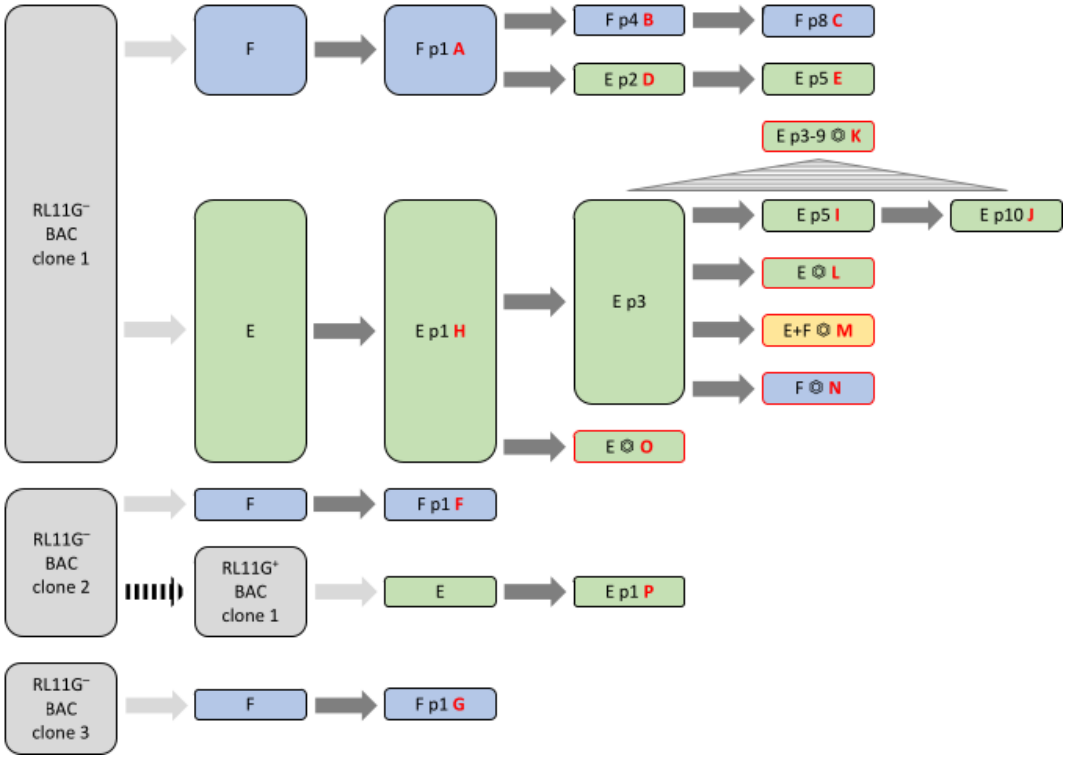

Figure 3

Reconstitution and passaging of viruses for sequencing. Clones of RhCMV68-1/EBOV/RL11G– BAC and RhCMV68-1/EBOV/RL11G+ BAC (denoted RL11G– BAC and RL11G+ BAC, respectively; shaded grey) were transfected into Telo-RF (F; shaded blue) or hTERT RPE-1 (E; shaded green) cells or a 1:1 mixture of these cells (E+F; shaded orange). The transfected cells were passaged further (p, passage number), and, in some instances, virus stocks (⊗) were made. The hatched triangle indicates that a stock was made from a mixture of yields from several passages. Hatched arrows indicate recombineering, light grey arrows indicate transfection, and dark grey arrows indicate passaging and expansion. Samples A–P (red font) were sequenced from DNA isolated from infected cells (black border of the rounded rectangles) or infected cell supernatant (red border of the rounded rectangles).

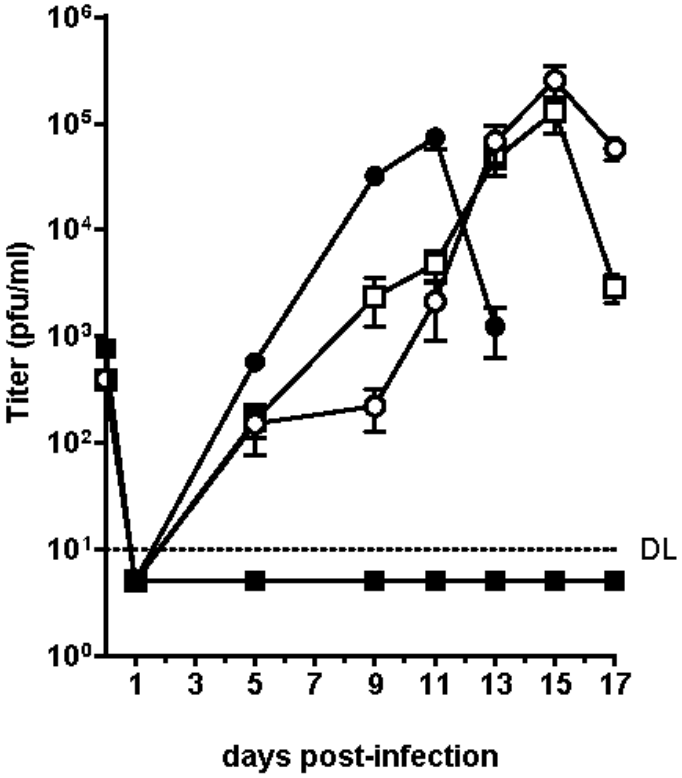

Figure 4

Cellular tropism of RhCMV68-1/EBOV/RL11G– compared to parental RhCMV68-1/EBOV-GP. Multi-step growth analysis was conducted by infecting Telo-RF (circles) or hTERT RPE-1 (squares) cells at a MOI of 0.2 PFU/cell with the reconstituted viruses RhCMV68-1/EBOV-GP (closed symbols) or RhCMV68-1/EBOV/RL11G– (open symbols). Supernatant was collected at the indicated d p.i. and titrated using a plaque assay. Titers are shown as ± SEM. DL, detection limit.

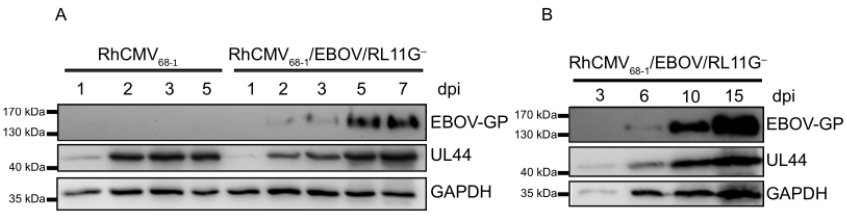

Figure 5

EBOV-GP expression of reconstituted viruses. A) Telo-RF cells were infected at a MOI of 0.2 PFU/cell with either RhCMV68-1 or RhCMV68-1/EBOV/RL11G–. B) hTERT RPE-1 cells were infected at a MOI of 0.4 PFU/cell with RhCMV68-1/EBOV/RL11G–. In both experiments, cell lysates were collected at the indicated d p.i and analyzed by immunoblotting. EBOV-GP was detected by using an EBOV GP-specific mAb. An antibody against the RhCMV UL44 protein was used as a viral infection control, and a mAb against cellular GAPDH was used as a cellular protein loading control.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RL11Brepaired.pdf
- UCDavisPREEMPTConsortiumAuthorListSubmitted19Oct21.docx