The translational landscape of Zika virus during infection of mammalian and insect cells

Nerea Irigoyen[1], Adam M. Dinan[1], Luke W. Meredith, Ian Goodfellow, Ian Brierley, Andrew E. Firth[*]

Division of Virology, Department of Pathology, University of Cambridge, Cambridge CB2 1QP, U.K.

[1] These authors contributed equally to this work

[*] Correspondence: aef24@cam.ac.uk (A.E.F.)

Abstract

Zika virus is a single-stranded, positive-sense RNA virus of the family *Flaviviridae*, which has recently undergone a rapid expansion among humans in the Western Hemisphere. Here, we report a high-resolution map of ribosomal occupancy of the Zika virus genome during infection of mammalian and insect cells, obtained by ribosome profiling. In contrast to some other flaviviruses such as West Nile, we find no evidence for substantial frameshift-induced ribosomal drop-off during translation of the viral polyprotein, indicating that Zika virus must use alternative mechanisms to downregulate levels of catalytically active viral polymerase. We also show that high levels of ribosome-protected fragments map in-frame to two previously overlooked upstream open reading frames (uORFs) initiating at CUG and UUG codons, with likely consequences for the efficiency of polyprotein expression. Curiously, in African isolates of Zika virus, the two uORFs are fused in-frame into a single uORF. A parallel RNA-Seq analysis reveals the 5' end

position of the subgenomic flavivirus RNA in mammalian and insect cells. Together, these provide the first analysis of flavivirus gene expression by ribosome profiling.

Author Summary

Recent Zika virus outbreaks have been associated with congenital diseases and neurological complications. An enhanced understanding of the molecular biology of this pathogen may contribute towards the development of improved treatment and control methods. We present a single-codon resolution analysis of Zika virus translation in mammalian and mosquito cells using ribosome profiling. The analysis revealed two hitherto uncharacterized uORFs in the 5' leader of Zika virus Brazilian isolate PE243, both of which are occupied by ribosomes during infection. In contrast, these two uORFs are fused into a single uORF in African isolates. This observation provides a new avenue for further investigations into potential factors involved in the emergence of Zika virus from a rarely detected pathogen into a major epidemic.

Introduction

Zika virus (ZIKV) is an emerging *Aedes* mosquito-borne flavivirus, initially isolated in Uganda in 1947 (**1**). The first large epidemic was reported from Yap Island in the Western Pacific in 2007, and the virus has since spread through Oceania, with Brazil the centre of the current epidemic. Until recently, ZIKV was not viewed as a particularly important pathogen, as the majority of infections are asymptomatic, and symptomatic infections resemble mild cases of dengue fever (**2**). It is now apparent that infection can cause substantial neurological disease, such as Guillain-Barré syndrome (**3**) and microcephaly in neonates (**4**). More than 200 genomic sequences for ZIKV isolates are now publicly available (**5**); however, there are limited functional genomics data for this pathogen.

Like other flaviviruses, ZIKV possesses a positive-sense, single-stranded RNA genome (gRNA) of ~11 kb, which contains a long open reading frame (ORF) flanked by 5′ and 3′ "untranslated" regions (UTRs) of approximately 100 and 400 nucleotides, respectively (**Fig. 1A**). The ORF encodes a polyprotein which is cleaved by host and viral proteases to yield three structural proteins located in the N-terminal region (capsid - C, precursor/membrane - Pr/M, and envelope - E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) (**6**). Consequently, these proteins are expected to be produced in equimolar amounts during infection, although the former set are presumed to be required in higher quantities for the production of progeny virus particles (**7**). Indeed, in the flavivirus lineage that includes West Nile virus, efficient ribosomal frameshifting in the NS2A region diverts up to 50% of translating ribosomes into an overlapping ORF, resulting in a substantial reduction in synthesis of the 3′-encoded replicative proteins relative to the 5′-encoded structural proteins (**8**). A similar phenomenon is also seen in a group of insect-specific flaviviruses (**9**).

Results/Discussion

We utilized sequencing of ribosome protected fragments (RPFs), known as ribosome profiling (Ribo-Seq), in combination with whole transcriptome sequencing (RNA-Seq), to investigate translation of the ZIKV genome at sub-codon resolution. African green monkey (Vero E6) cells and *Aedes albopictus* (C6/36) cells were infected with ZIKV Brazilian isolate PE243 (**10**). Cells were harvested at 24 h post infection (p.i.) and Ribo-Seq and RNA-Seq libraries were prepared and deep sequenced. The resulting reads were mapped to viral and host genomes (**S1 Table**). **Fig. 1B** illustrates Ribo-Seq and RNA-Seq densities on the viral genome for infected Vero and C6/36 cells. Localized variations in RPF density may be attributable partly to technical biases (ligation, PCR and nuclease biases; **11**) and partly to ribosomes pausing at specific sites during translation (**Fig. 1B**) (**12**). On a larger scale (**Fig 2A**), we observed a gradual decline in mean RPF density in the 5′ to 3′ direction in Vero cell infections, but no evidence of a sharp reduction at any single location as

would arise from frameshifting at levels comparable to those seen in West Nile virus. Similarly, no evidence for ribosomal drop-off was observed in insect cells. Consistent with this, a survey of ZIKV genomic sequences revealed few conserved ORFs of significant length in either of the two alternative reading frames (**Fig. 1C**), and no statistically significant evidence for overlapping elements (e.g. ribosomal frameshift signals) embedded within the polyprotein coding sequence (cf. West Nile virus; **13**) except near the 5′ end (**Fig. 1D**). Note, however, that these comparative analyses are restricted by the low sequence diversity available in the ZIKV clade.

Strikingly, we found significant ribosome occupancy within the ZIKV 5′ UTR (**Fig. 2A**). The length distribution of Ribo-Seq reads mapping to this region mirrored that of polyprotein-mapping RPFs, indicating that they are *bona fide* ribosome footprints (**S1 Fig**). A prominent peak of RPF density occurred at nucleotide 25 of the 5′ UTR in Vero cells, coinciding with a non-canonical (CUG) initiation codon **(Fig. 2B)**. RPFs mapped in-frame along the length of the associated 29-codon upstream ORF (uORF1), which ends at nucleotide 111 (4th nucleotide of the polyprotein ORF) (**Fig. 2B-C**). Additionally, RPFs mapped in-frame to a second uORF (uORF2), associated with a non-canonical (UUG) initiation codon at nucleotide 80 (**Fig. 2B-C**). This uORF2, 77 codons in length, extends 202 nucleotides into the polyprotein ORF, and is generally conserved across ZIKV isolates (**Fig. 1C; Frame +2**). Both uORFs were also occupied by ribosomes during infection of C6/36 cells, although uORF2 was more prominent in this case (**Fig. 2B**).

An analysis of available ZIKV 5′ UTR sequences showed that the CUG and UUG codons are conserved in all sequenced isolates. Moreover, both codons have strong initiation contexts (**Fig. 2B**) (**14**). Notably, in all African ZIKV isolates with available sequence data, uORF1 and uORF2 exist as a single ORF, which appears to have been split in two by the insertion of a uracil residue at position 81 in the Malaysian lineage that gave rise to the American strain. Translation of uORFs has the potential to regulate expression of downstream genes, even though the peptides encoded by

uORFs are often non-functional (**15**). In some instances, ribosomes reinitiate at downstream AUG codons after translating short uORFs (**16**); however this appears unlikely in ZIKV given that both uORFs extend beyond the polyprotein AUG codon. Additional work in animal infection models is required to assess the functional significance of these uORFs.

Few RPFs were present in the 3′ UTR, consistent with a lack of translation of this region. In Vero cell infections, 3′ UTR RNA-Seq density was on average 68% higher than in the rest of the gRNA. Structured flavivirus 3′ UTRs resist degradation by 5′-3′ Xrn1 host exonuclease, giving rise to noncoding subgenomic flavivirus RNAs (sfRNAs) that accumulate during infection (**17**). These sfRNAs are linked to cytopathic and pathologic effects and previous work on ZIKV strain PE243 has shown that they can modulate host type-I interferon signaling by interacting with the RIG-I nucleic acid pattern recognition receptor (**10**). A sharp peak in read 5′ end mappings occurred in all libraries at nucleotide position 10478 consistent with the presence of a nuclease-resistant RNA structure at this position (**Fig. 2D**). This location is one nucleotide upstream of the predicted 5′ end of RNA “stem-loop 2” (SL2). In contrast, while Xrn1 halts preferentially at the adjacent SL1 in ZIKV strain PRVABC59 (**17**), we found only a much more modest peak in read 5′ end mappings at the SL1 site (nucleotide position 10394).

Using RNA sequencing and ribosome profiling, we have provided the first high-resolution map of flavivirus translation in mammalian and mosquito cells. In contrast to some other flaviviruses, we find no evidence for efficient ribosomal frameshifting during Zika virus translation. The observation of ribosomal occupancy within two non-AUG uORFs, which exist as a single uORF in Old World isolates, provides a starting point for further investigations into potential factors involved in the emergence of ZIKV from a rarely detected pathogen into a major epidemic.

Materials and Methods

Cells and virus

Cell lines were obtained from the European Collection of Authenticated Cell Cultures (ECACC) and are certified mycoplasma-free (tested by PCR, culturing and Hoechst stain). In addition, the sequenced libraries were queried for mycoplasma sequences. ZIKV isolate PE243 (GenBank accession number KX197192) stocks were produced and titred on Vero cells.

Ribosomal profiling and RNA-Seq

Vero E6 (*Chlorocebus sabaeus*) and C6/36 (*Aedes albopictus*) cells were maintained in Dulbecco's modification of Eagle's medium (DMEM) and L-15 medium respectively, supplemented with 10% (vol/vol) foetal calf serum (FCS). 10^7 cells were plated in 10-cm dishes and infected with 5 focus forming units (FFU) per cell of virus. After 1 h at 37 ºC for Vero cells and 28 ºC for C6/36 cells, the inoculum was removed and cells were incubated in DMEM or L-15 containing 10% FCS, 100 U/ml penicillin and 100 mg/ml streptomycin at 37 ºC or 28 ºC as previously indicated.

At 24 h p.i., cells were treated with cycloheximide (Sigma-Aldrich; to 100 μg/ml; 3 min). Cells were rinsed with 5 ml of ice-cold PBS, transferred to ice, and 400 μl of lysis buffer [20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 1% Triton X-100, 100 μg/ml cycloheximide and 25 U/ml TURBO DNase (Life Technologies)] dripped onto the cells. The cells were scraped extensively to ensure lysis, collected and triturated with a 26-G needle ten times. Lysates were clarified by centrifugation for 20 min at 13,000 *g* at 4 °C. Cell lysates were subjected to Ribo-Seq and RNA-Seq using methodologies based on the original protocols of Ingolia and colleagues (**18, 19**), except library amplicons were constructed using a small RNA cloning strategy (**20**) adapted to Illumina smallRNA v2 to allow multiplexing, as described previously (**21, 22**).

Ribosomal RNA was depleted using the "Human, Mouse, Rat" Ribo-Zero kit (Illumina) which is also recommended for insect rRNA depletion. Due to poor Ribo-Zero depletion of rRNA in the C6/36 Ribo-Seq samples, additional aliquots of the two biological repeats were further treated with duplex-specific nuclease (DSN; Evrogen) as described previously (**21**). Amplicon libraries were sequenced using the Illumina NextSeq platform at the Department of Biochemistry, University of Cambridge. Sequencing data have been deposited in ArrayExpress (http://www.ebi.ac.uk/arrayexpress) under the accession number E-MTAB-5418.

Computational analyses of sequence data

Sequencing results for Ribo-Zero-treated and DSN-treated libraries for each C6/36 biological repeat were combined. Adaptor sequences were trimmed using the FASTX-Toolkit, and reads shorter than 25 nt following adaptor trimming were discarded. Mapping steps were performed as described in (**22**), using Bowtie version 1 (**23**). Trimmed reads were mapped first to host rRNA, followed by ZIKV PE243 sequence KX197192.1, followed by the complete set of NCBI RefSeq mRNAs for the relevant host organism (*Chlorocebus sabaeus* for Vero cells and *Aedes albopictus* for C6/36 cells). The order of mapping was tested to check that host-derived reads were not accidentally mis-mapped to the virus genome, or *vice versa*.

For analyses of viral gene expression and for visualizing Ribo-Seq coverage of the viral genome, a +12 nt offset was applied to the 5′ end mapping positions of RPFs, to approximate the P-site position of the ribosome (**22**). To normalize for different library sizes, reads per million mapped reads (RPM) values were calculated using the sum of total virus RNA plus total positive-sense host RefSeq mRNA as the denominator. Host mRNA Ribo-Seq and RNA-Seq phasing distributions were calculated as described in (**22**). When visualising viral RNA coverage in both Ribo-Seq and RNA-Seq libraries (i.e. Figures 1B, 2B and 2D), reads from biological repeats were pooled together.

To compare the 5′ UTRs of ZIKV strains, the complete set of available ZIKV genomes was retrieved from NCBI via a tblastn search, using the ZIKV PE243 polyprotein sequence as the query. The subset of genomic sequences which included 5′ UTRs was identified, and these sequences were aligned against one another using Kalign (**24**). The synonymous site variability analysis was performed as described in (**25**). All Zika and Spondweni virus full-length polyprotein coding sequences available in GenBank as of 5 Dec 2016 (155 and 1 sequence, respectively) were included in the analysis.

References

1. **Dick, G.W., Kitchen, S.F., Haddow, A.J.** 1952. Zika Virus (I). Isolations and serological specificity. Trans R Soc Trop Med Hyg **46**:509-520.

2. **Lazear, H.M., Diamond, M.S.** 2016. Zika virus: new clinical syndromes and its emergence in the Western Hemisphere. J Virol **90**:4864-4875.

3. **Cao-Lormeau VM, Blake A, Mons S, Lastère S, Roche C, Vanhomwegen J, Dub T, Baudouin L, Teissier A, Larre P, Vial AL.** 2016. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. Lancet **387**:1531-1539.

4. **Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, Kolenc M, Resman Rus K, Vesnaver Vipotnik T, Fabjan Vodušek V, Vizjak A.** 2016. Zika virus associated with microcephaly. N Engl J Med **374**:951-958.

5. **Lanciotti RS, Lambert AJ, Holodniy M, Saavedra S, Signor LD.** 2016. Phylogeny of Zika virus in western hemisphere, 2015. Emerg Infect Dis **22**:933-935.

6. **Rice CM, Lenches EM, Eddy SR, Shin SJ, Sheets RL, Strauss JH.** 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science **229**:726-733.

7. **Ahlquist, P.** 2006. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. Nat Rev Microbiol **4**:371-382.

8. **Melian EB, Hall-Mendelin S, Du F, Owens N, Bosco-Lauth AM, Nagasaki T, Rudd S, Brault AC, Bowen RA, Hall RA, van den Hurk AF.** 2014. Programmed ribosomal frameshift alters expression of west nile virus genes and facilitates virus replication in birds and mosquitoes. PLoS Pathog **10**:e1004447.

9. **Firth, A.E., Blitvich, B.J., Wills, N.M., Miller, C.L., Atkins, J.F.** 2010. Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. Virology **399**:153-166.

10. **Donald CL, Brennan B, Cumberworth SL, Rezelj VV, Clark JJ, Cordeiro MT, de Oliveira França RF, Pena LJ, Wilkie GS, Filipe AD, Davis C.** 2016. Full genome sequence and sfRNA interferon antagonist activity of Zika virus from Recife, Brazil. PLoS Negl Trop Dis **10**:e0005048.

11. **Artieri CG, Fraser HB.** 2014. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. Genome Res **24**:2011-2021.

12. **Ingolia NT, Lareau LF, Weissman JS.** 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell **147**:789-802.

13. **Firth AE, Atkins JF.** 2009. A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1'may derive from ribosomal frameshifting. Virol J **6**:14.

14. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283-292.

15. **Somers J, Pöyry T, Willis AE.** 2013. A perspective on mammalian upstream open reading frame function. The international journal of biochemistry & cell biology. Int J Biochem Cell Biol **45**:1690-1700.

16. **Jackson RJ, Hellen CU, Pestova TV.** 2012. Termination and post-termination events in eukaryotic translation. Adv Protein Chem Struct Biol **86**:45-93.

17. **Akiyama BM, Laurence HM, Massey AR, Costantino DA, Xie X, Yang Y, Shi PY, Nix JC, Beckham JD, Kieft JS.** 2016. Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. Science **354**:1148-1152.

18. **Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS.** 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science **324**:218-223.

19. **Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman, JS.** 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc **7**:1534-1550.

20. **Guo H, Ingolia NT, Weissman JS, Bartel DP.** 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature **466**:835-840.

21. **Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC, Brierley I.** 2015. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. RNA **21**:1731-1745.

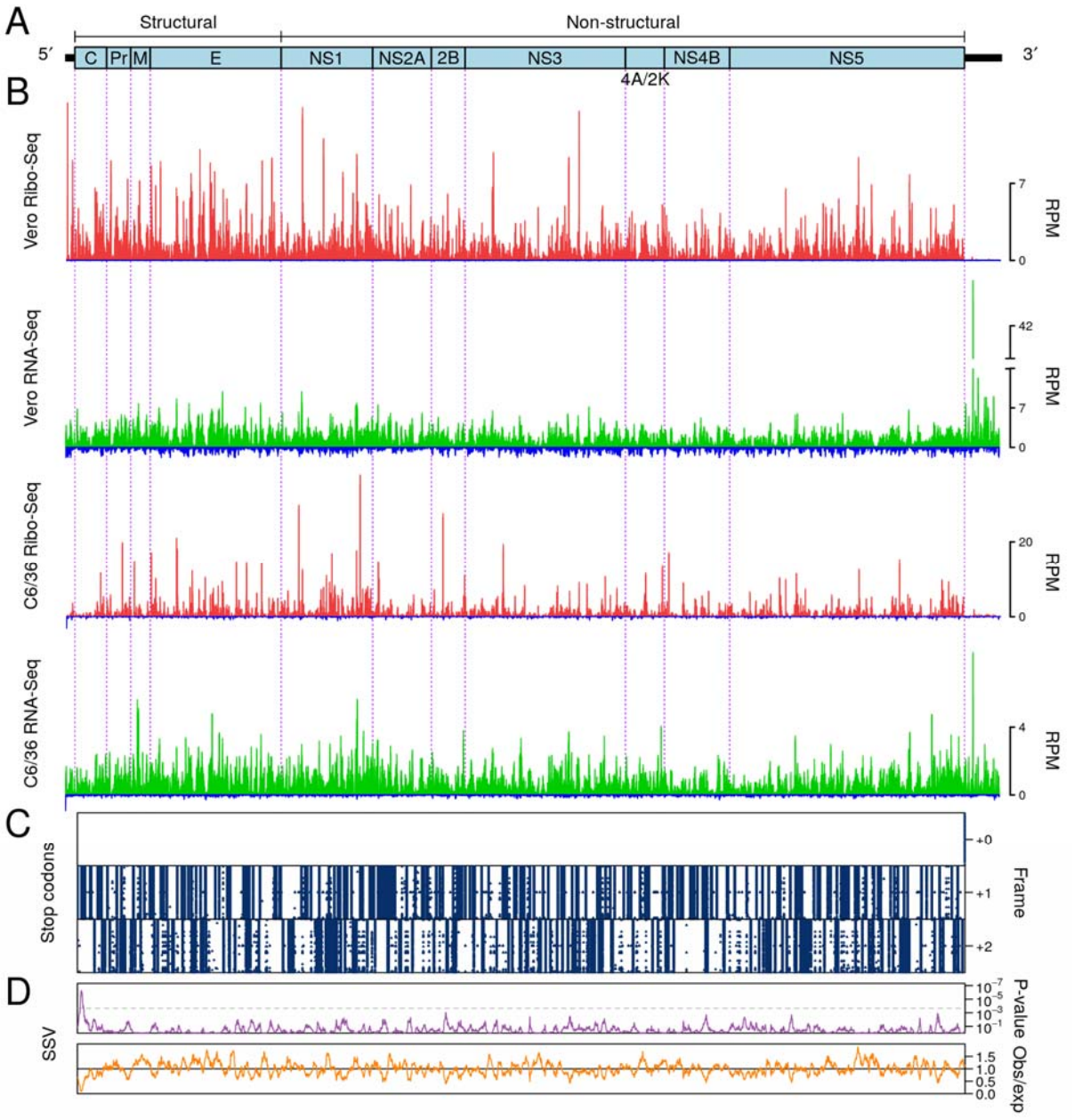
22. **Irigoyen N, Firth AE, Jones JD, Chung BY, Siddell SG, Brierley I.** 2016. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. PLoS Pathog **12**:e1005473.

23. **Langmead B, Trapnell C, Pop M, Salzberg SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **10**: R25.

24. **Lassmann, T., Sonnhammer, E.L.** 2005. Kalign–an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics **6**:298.

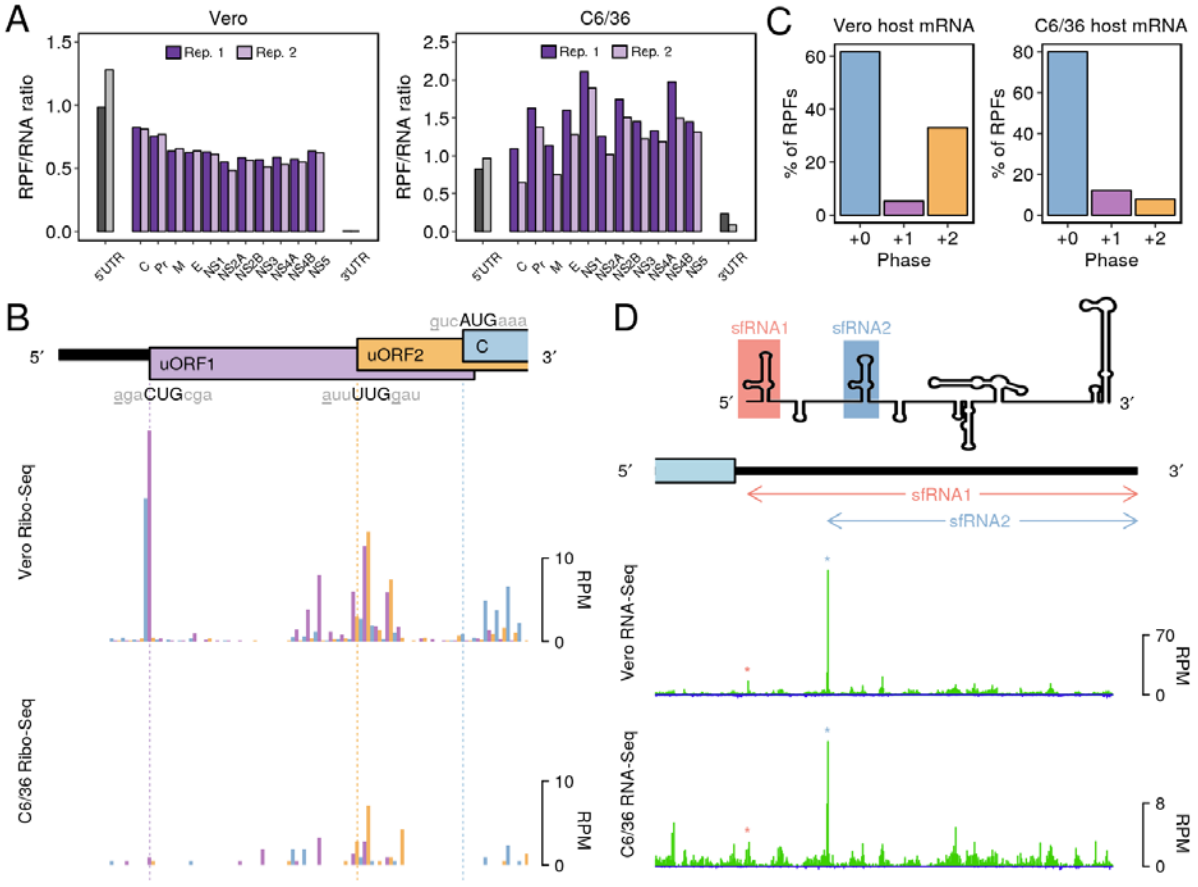
25. **Firth, A.E.** 2014. Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. Nucleic Acids Res **42**:12425-12439.

Figure 1. ZIKV RNA synthesis and translation. (**A**) Map of the 10807-nt ZIKV/Brazil/PE243/2015 genome. The 5′ and 3′ UTRs are indicated in black and the polyprotein ORF is indicated in pale blue with subdivisions showing mature cleavage products. (**B**) Ribo-Seq (red) and RNA-Seq (green) reads per million mapped reads (RPM), smoothed with a 3-nt sliding window, for Vero cells (upper panels) and C6/36 cells (lower panels). Histograms show the positions of the 5′ ends of reads with a +12 nt offset to map (for RPFs) approximate P-site positions. Negative-sense reads are shown in dark blue below the horizontal axis. (**C**) Positions of +0, +1 and +2 frame stop codons (blue points) in all available ZIKV and Spondweni virus polyprotein coding sequences. (**D**) Synonymous site variability (SSV) in a 15-codon window for the same sequences. The grey dashed line indicates a $P = 0.05$ threshold corrected for multiple testing.



267
268

Figure 2. Analysis of novel translational and transcriptional features of ZIKV. (**A**) Ratios of Ribo-Seq to RNA-Seq read density in different regions of the ZIKV genome for Vero cells (left) and C6/36 cells (right). **(B)** The 5′ region of the ZIKV genome showing two non-AUG uORFs (upper). Ribo-Seq counts at 24 h p.i. for Vero cells (middle) and C6/36 cells (lower). Histograms show the positions of the 5′ ends of reads with a +12 nt offset to map the approximate P-site. Reads whose 5′ ends map to the first, second or third phases relative to codons in the polyprotein reading frame are indicated in blue, purple or orange, respectively. **(C)** Mean phasing of host mRNA RPF 5′ end positions where the translated reading frame is indicated in blue. The expected phasing of RPFs in out-of-frame uORFs corresponds to a cyclic permutation of the three colours. **(D)** Predicted RNA structures (upper) and RNA-Seq densities at 24 h p.i. for Vero cells (middle) and C6/36 cells (lower) in the ZIKV 3′ UTR. Negative-sense reads are shown in dark blue.



Acknowledgements

The authors are indebted to Alain Kohl (Centre for Virus Research, University of Glasgow) and Lindomar J. Pena and Rafael Oliveira de Freitas França (Fiocruz Recife, Pernambuco, Brazil) for the provision of PE243 ZIKV RNA used to generate the virus stock.

Author Information

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.E.F. (aef24@cam.ac.uk).