Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples

Joshua Quick[1], Nathan D Grubaugh[2], Steven T Pullan[3], Ingra M Claro[4], Andrew D Smith[1], Karthik Gangavarapu[2], Glenn Oliveira[5], Refugio Robles-Sikisaka[2], Thomas F Rogers[2,6], Nathan A Beutler[2], Dennis R Burton[2], Lia Laura Lewis-Ximenez[7], Jaqueline Goes de Jesus[8], Marta Giovanetti[8,9], Sarah C Hill[10], Allison Black[11,12], Trevor Bedford[11], Miles W Carroll[3,13], Marcio Nunes[14], Luiz Carlos Alcantara Jr.[8], Ester C Sabino[4], Sally A Baylis[15], Nuno R Faria[10], Matthew Loose[16], Jared T Simpson[17], Oliver G Pybus[10], Kristian G Andersen[2,5] & Nicholas J Loman[1]

[1]Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham, UK. [2]The Scripps Research Institute, La Jolla, California, USA. [3]Public Health England, National Infection Service, Porton Down, Salisbury, UK. [4]Department of Infectious Disease and Institute of Tropical Medicine, University of Saõ Paulo, Saõ Paulo, Brazil. [5]Scripps Translational Science Institute, La Jolla, California, USA. [6]Massachusetts General Hospital, Boston, Massachusetts, USA. [7]Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil. [8]Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Brazil. [9]University of Rome, Tor Vergata, Italy. [10]Department of Zoology, University of Oxford, Oxford, UK. [11]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [12]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [13]University of Southampton, South General Hospital, Southampton, UK. [14]Instituto Evandro Chagas, Belem, Brazil. [15]Paul-Ehrlich-Institut, Langen, Germany. [16]DeepSeq, School of Life Sciences, University of Nottingham, Nottingham, UK. [17]OICR, Toronto, Canada. Correspondence should be addressed to N.J.L. (n.j.loman@bham.ac.uk).

Published online 24 May 2017; doi:10.1038/nprot.2017.066

Genome sequencing has become a powerful tool for studying emerging infectious diseases; however, genome sequencing directly from clinical samples (i.e., without isolation and culture) remains challenging for viruses such as Zika, for which metagenomic sequencing methods may generate insufficient numbers of viral reads. Here we present a protocol for generating coding-sequence-complete genomes, comprising an online primer design tool, a novel multiplex PCR enrichment protocol, optimized library preparation methods for the portable MinION sequencer (Oxford Nanopore Technologies) and the Illumina range of instruments, and a bioinformatics pipeline for generating consensus sequences. The MinION protocol does not require an Internet connection for analysis, making it suitable for field applications with limited connectivity. Our method relies on multiplex PCR for targeted enrichment of viral genomes from samples containing as few as 50 genome copies per reaction. Viral consensus sequences can be achieved in 1–2 d by starting with clinical samples and following a simple laboratory workflow. This method has been successfully used by several groups studying Zika virus evolution and is facilitating an understanding of the spread of the virus in the Americas. The protocol can be used to sequence other viral genomes using the online Primal Scheme primer designer software. It is suitable for sequencing either RNA or DNA viruses in the field during outbreaks or as an inexpensive, convenient method for use in the lab.

INTRODUCTION

Genome sequencing of viruses has been used to study the spread of disease in outbreaks[1]. Real-time genomic surveillance is important in managing viral outbreaks, as it can provide insights into how viruses transmit, spread and evolve[1–4]. Such work depends on rapid sequencing of viral material directly from clinical samples—i.e., without the need to isolate the virus in pure culture. During the Ebola virus epidemic of 2013–2016, prospective viral genome sequencing was able to provide critical information on virus evolution and help inform epidemiological investigations[3–6]. Sequencing directly from clinical samples is faster, less laborious and more amenable to near-patient work than time-consuming culture-based methods. Metagenomics, the process of sequencing the total nucleic acid content in a sample (typically cDNA or DNA), has been successfully applied to both virus discovery and diagnostics[7–9]. Metagenomic approaches have seen rapid adoption over the past decade, fueled by relentless improvements in the yield of high-throughput sequencing instruments[5,10–12]. Whole-genome sequencing of Ebola virus directly from clinical samples without amplification was possible because of the extremely high virus copy numbers found in acute cases[13–15]. However, direct metagenomic sequencing from clinical samples poses challenges with regard to sensitivity: genome coverage may be low or absent when attempting to sequence viruses that are present at low abundance in a sample with high levels of host nucleic acid background.

Development of the protocol

During recent work on the Zika virus epidemic[16], we found that it was difficult to generate whole-genome sequences directly from clinical samples using metagenomic approaches (**Table 1**). These samples had cycle threshold (Ct) values between 33.9 and 35.9 (equivalent to 10–48 genome copies per microliter). Before sequencing, these samples were depleted of human rRNA and prepared for metagenomic sequencing on the Illumina MiSeq platform as previously described[2,17]. In these cases, sequences from Zika virus comprised <0.01% of the data set, resulting in incomplete coverage. Greater coverage and depth are critical for accurate genome reconstruction and subsequent phylogenetic inference. In addition, there are substantial sequencing, analysis and storage costs associated with generating large sequencing data sets; therefore, metagenomic approaches currently do not lend themselves to the cost-effective use of lower-throughput portable sequencing devices such as the Oxford Nanopore MinION.

TABLE 1 | Results of metagenomic sequencing of five Zika-positive clinical samples collected from Colombia in January 2016 (unpublished data, K.G.A. and N.D.G.).

Sample	Ct[a]	GEs/μl RNA[b]	No. of Reads[c]	No. of ZIKV reads[d]	% ZIKV[e]	% Coverage[f]	Depth[g]
ZC188	34	36	1,929,030	60	0.003	23.7	1.4
ZC192	33.9	39	2,073,344	223	0.011	79.2	5.0
ZC199	35.9	10	1,612,686	0	0	0	0
ZC204	33.6	48	1,966,651	28	0.001	25.3	0.5
ZC207	35.2	16	1,445,414	4	0.000	3.4	0.1

[a]Quantitative reverse-transcription PCR (qRT-PCR) cycle threshold (Ct) value. [b]Genome copy equivalents (GE) per microliter calculated from Ct value and standard dilutions. [c]Number of reads generated for this sample. [d]Number of reads aligning to the Zika virus reference genome. [e]Percentage of reads mapping to the Zika virus reference genome. [f]Percentage of Zika virus genome covered by at least one read. [g]Mean depth of coverage for the Zika virus genome.

To generate complete viral genome coverage from clinical samples in an economic manner, target enrichment is often required[18]. Enrichment can be achieved directly through isolation in culture or the use of oligonucleotide bait probes targeting the virus of interest, or indirectly via host nucleic acid depletion. Amplification may also be required to generate sufficient material for sequencing (>5 ng for typical Illumina protocols and 100–1,000 ng for MinION). PCR can provide both target enrichment and amplification in a single step, and is relatively cheap, available and fast as compared with other methods. To generate coding-sequence complete coverage, a tiling amplicon scheme is commonly used[19–21]. During our work with Ebola virus, we were able to reliably recover >95% of the genome by sequencing 11 long amplicons (1–2.5 kb in length) on the MinION[5].

The likelihood of long fragments being present in the sample, however, reduces with lower virus abundance. Therefore, we anticipated that, for viruses such as Zika that are present at low abundance in clinical samples, we would be more likely to amplify shorter fragments. As an extreme example of this approach, a recent approach termed 'jackhammering' was used to amplify degraded HIV-1 samples stored for >40 years; this approach used 200–300 nt amplicons to help maximize sequence recovery[22]. Using shorter amplicons necessitates a larger number of products to generate a tiling path across a target genome. Doing this in individual reactions requires a large number of manual pipetting steps and therefore increases the potential for mistakes, with a heightened risk of cross-contamination, as well as a greater cost in time and consumables. To solve these problems, we designed a multiplex assay to carry out tens of reactions in an individual tube. This method has been subsequently used to perform Zika sequencing in order to understand the spread of Zika virus in the Americas[16,23–26]. Our resulting step-by-step protocol, described here, allows any researcher to successfully amplify and sequence viruses of low abundance directly from clinical samples. The method also has other potential uses that are not demonstrated here. One potential application is multilocus sequencing typing approaches, which could be carried out by amplifying conserved genes from bacteria, fungi and yeasts. Simultaneously, antibiotic-resistance-determining genes or key virulence genes could also be targeted in the same assay. The scheme could also be used to sequence chloroplast and mitochondrial genomes.

Comparison with other approaches

The three most common approaches for sequencing viruses are metagenomic sequencing, PCR amplicon sequencing and target enrichment sequencing, recently reviewed in detail by Houldcroft *et al.*[27]. The main benefits of the PCR-based approach described here are cost and sensitivity. In theory, both PCR and cell culture require only one viral copy, making them both exquisitely sensitive. In practice, however, the reaction conditions do not allow single-genome amplification, and, typically, multiple starting molecules are required. PCR also has limited sensitivity in cases in which the template sequence is divergent from the expected because of primer-binding kinetics. However, in an outbreak situation in which isolates are highly related, and low cost per sample and rapid turnaround time are required, PCR is particularly suitable. Sequencing amplicons on the Oxford Nanopore MinION is a popular method for determining viral genomes and has been used for diverse viruses, including Ebola, influenza and poxvirus, using either single primer pair reactions generating long amplicons (>1 kb) or multiple reactions that are pooled before sequencing[5,28–30]. However, these approaches are laborious to scale up when many small amplicons are required (because of low viral copy numbers), or when multiple samples are sequenced on a single sequencing run, as in this protocol.

The most similar alternative approach to the one described here is AmpliSeq (Life Technologies), which was previously used for Ebola sequencing on the Ion Torrent PGM[6]. However, this method is specific to the Ion Torrent platform, and primer schemes must be ordered directly from the manufacturer; thus, it may consequently be more expensive per sample. Alternative software packages for designing primer schemes are available, some of which cater specifically to multiplex or tiling amplicon schemes[20,21,31,32], and these may perform better when dealing with divergent genomes because of an increased emphasis on oligonucleotide degeneracy. Primers generated with such software may also be compatible with this protocol, although PCR conditions may require optimization, as the Primal Scheme software used in this protocol is designed with an emphasis on monitoring short-term evolution of known lineages, and primer conditions have been optimized for multiplex PCR amplification efficiency.

Propagation in cell culture is another method that has been widely used for virus enrichment[33–35]. This process is time-consuming, and requires specialist expertise and high containment laboratories for especially dangerous pathogens. There is also concern that viral passage can introduce mutations that are not present in the original clinical sample, potentially confounding analysis[36,37].

Oligonucleotide bait probes have also shown promise as an alternative to metagenomics and amplicon sequencing[38–42]. These isolate viral nucleic acid sequences by hybridizing target-specific biotinylated probes to the DNA/RNA sample and then separating them using magnetic streptavidin-coated beads. Such methods, however, are limited by the efficiency of the capture step because of the kinetics of nucleic acid hybridization in complex samples such as those containing the human genome. The complete hybridization of all probes to targets can take hours (typical protocols suggest a 24-h incubation, although shorter times may be possible) and may never be achieved because of competitive binding by the host DNA. These methods suffer from a coverage bias, which worsens at lower viral abundances, resulting in increasingly incomplete genomes, as demonstrated by recent work on the Zika virus[43]. They work best on samples with higher viral abundances and may not have the sensitivity to generate near-complete genomes for the majority of isolates in an outbreak. Probes for hybridization capture are also more expensive than PCR primers because they are usually designed in a fully overlapping 75-nt scheme, which can run to hundreds of probes per virus and thousands for panels of viruses.

Direct sequencing of RNA has been recently demonstrated on the Oxford Nanopore MinION[44,45]. This method is attractive because it eliminates the need for reverse transcription, and so potentially may reduce biases resulting from nonrandom priming and copying errors introduced by reverse transcriptase. However, this method currently requires 500 ng of RNA as starting material and would suffer from the same sensitivity issues associated with cDNA metagenomics approaches when applied to samples containing very low viral copy numbers.

Limitations of tiling amplicon sequencing

Our method is not suitable for the discovery of new viruses or for sequencing highly diverse or recombinant viruses because primer schemes are virus-genome-specific. This protocol has not been validated for discovery of intra-host nucleotide variants, and we expect that minor allele frequencies will not be reliably recovered when amplifying from very small amounts of starting virus, as shown by Metsky *et al.*[25]. We expect that this method will work for larger virus genomes, but we have not tested this protocol with viral genomes longer than 12 kb. The protocol is designed for infections resulting from single clones, and may not perform well with mixed infections of diverse viruses. We have not tested performance of the method in chronic infections in which large amounts of diversity may have evolved within a patient (for example, viral quasispecies during HIV infection). Amplicon sequencing is prone to coverage dropouts that may result in incomplete genome coverage, especially at lower abundances, and the loss of both 5′ and 3′ regions that fall in regions not covered by primer pairs. Sequencing of complete 5′- and 3′-UTR regions may require alternative techniques such as RACE[46]. Targeted methods are also highly sensitive to amplicon contamination from previous experiments. Extreme caution should be taken to keep pre-PCR areas, reagents and equipment free of contaminating amplicons.

Experimental design

Description of the protocol. We describe a fully integrated end-to-end protocol for rapid sequencing of viral genomes directly from clinical samples. The protocol proceeds in four stages: (i) multiplex primer pool design, (ii) multiplex PCR, (iii) sequencing on MinION or Illumina instruments and (iv) bioinformatic analysis and quality control (QC) (**Fig. 1**).

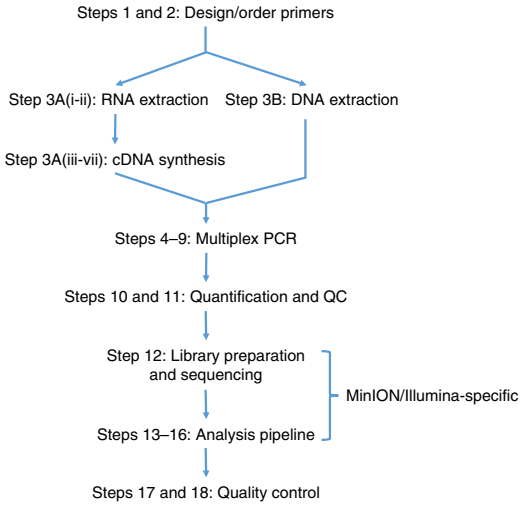

Figure 1 | Workflow for tiling amplicon sequencing on MinION/Illumina platforms, with associated Procedure step numbers indicated.

Primer design. We developed a web-based primer design tool called Primal Scheme (http://primal.zibraproject.org), which provides a complete pipeline for the development of efficient multiplex primer schemes. Each scheme is a set of oligonucleotide primer pairs that generate overlapping products, the size of which is determined by the target genome length, amplicon length and overlap required, as discussed below. For Zika, we use 35 primer pairs, amplifying products of ~400-nt length with a 100-nt overlap for the ~11-kb viral genome. Together, the amplicons generated by the pairs span the target genome or region of interest (**Fig. 2**).

As input, Primal Scheme requires a FASTA file containing one or more reference genomes. The user specifies a desired PCR amplicon length (default = 400 nt, suggested values between 200 and 2,000 nt) and the desired length of overlap between neighboring amplicons (default = 75 nt). Using a shorter amplicon length may be useful for samples in which longer products fail to amplify (e.g., when the virus nucleic acid is highly degraded). However, if amplicon lengths become too short (e.g., <300 nt), it may not be possible to find suitable primer pairs; reducing the overlap parameter may help with this.

The Primal Scheme software performs the following processes:

Generation of candidate primers: The first sequence listed in the FASTA file should be the most representative genome, with further sequences spanning the expected interhost diversity. Primal Scheme uses the Primer3 software to generate candidate primer pairs (five, by default)[47]. It selects primers based on thermodynamic modeling, which takes into account length, annealing temperature, %GC, 3′ stability, estimated secondary structure and likelihood of primer–dimer formation, maximizing the chance of a successful PCR reaction. Primers are designed with a high annealing temperature within a narrow range (65–68 °C) that allows PCR to be performed as a 2-step protocol (95 °C denaturation, 65 °C combined annealing and extension) for highly specific amplification from clinical samples without the need for nested primers.

Testing of candidate primers: Subsequent reference genomes in the file are used to help choose primer pairs that maximize the likelihood of successful amplification of known virus diversity. A semi-global alignment score between each candidate primer and all supplied references is calculated to ensure that the most 'universal' candidate primers are picked for the scheme. Mismatches at the 3′ end are severely penalized, as they have a disproportionate effect on the likelihood of successful extension[48,49]. The alignment scores are summed, and the single best-scoring pair for each region is selected. If no candidates are returned by Primer3 for a region, most likely because all primers had insufficient annealing temperature, an error message prompting you to adjust the amplicon length or the overlap parameter will appear.

Output of primer pairs: Output files include a table of primer sequences to be ordered, a BED file of primer locations that can be used subsequently for primer trimming and a diagram of the primer scheme.

Choice of amplicon length. The choice of amplicon length when designing primer pools for sequencing is important. There is an inverse relationship between amplicon length and the number of primer pairs. It is believed that increasing the number of primer

Figure 2 | Overview of multiplex primer design using Primal Scheme online primer design tool. (**a**) Submission box for online primer design tool. (**b**) Primer table of results. (**c**) Schematic showing expected amplicon products for each pool in genomic context for the ZikaAsian and ChikAsianECSA schemes.

TABLE 2 | Results of MinION R9.4 2D sequencing after barcode demultiplexing for an isolate of Zika and for a clinical sample of chikungunya virus.

Sample	Scheme[a]	Ct[b]	No. of reads (2D pass)[c]	No. of aligned reads[d]	% Aligned[e]	% Coverage (25×)[f]	Depth[g]
Zika WHO Control Reference (11474/16)	ZikaAsian	18–20	9,606	9,528	99.2	97.7	392
Chikungunya PEI N11602	ChikAsianECSA	20	89,365	88,527	99.1	88	2,779

Zika data from Faria *et al.*[23]; chikungunya unpublished data contributed by S.A.B., J.Q. and N.J.L. [a]The primer set generated by Primal Scheme that was used for sequencing. [b]Cycle threshold for qRT-PCR. [c]Number of reads of type 2D pass. [d]Number of reads aligning to the reference genome. [e]Percentage of reads aligning to the reference genome. [f]Percentage of reads aligning to the reference genome with at least 25× coverage. [g]Mean depth of coverage.

pairs reduces the likelihood of successful amplification of each region, owing to interaction between primers[18]. It is plausible that as the number of primer pairs increases, competitive inhibition may decrease PCR efficiency, although the high annealing temperature used in this protocol should reduce this risk. Longer amplicons are preferred, as they mean fewer primer pairs are needed per reaction. They also increase the amount of linkage information that can be recovered as haplotypes, which is of importance for investigation of within-host diversity. On the Illumina platform, 600 bases is the maximum size of amplicon that can be obtained using this protocol without an additional fragmentation step (using 600 cycle kits in paired-end mode—i.e., paired 300 nucleotides without any overlap), although read accuracy may degrade during the last 50 cycles. On the Oxford Nanopore MinION, there is no limit to the maximum amplicon length that can be sequenced; the maximum length is effectively limited by the performance of the reverse transcription and PCR (practically to ~5 kb). However, longer amplicons are less likely to amplify successfully when viral copy number is low or there is sample degradation (e.g., because of inadequate storage).

Optimization of primer schemes. The majority of primers are expected to work even when pooled in equimolar amounts, meaning largely complete genomes can be recovered without optimization. For example, the chikungunya virus data shown in **Table 2** were generated without any optimization. However, to achieve coding-sequence-complete genomes, problem primers causing inefficient amplification of certain regions may need to be replaced or their concentrations adjusted relative to other primers in an iterative manner. Complete coverage of the genome covered by the scheme—i.e., all amplicons successfully amplified—should be achievable for the majority of samples using this protocol; however, coverage is still expected to correlate with viral abundance (**Table 3**).

Multiplex PCR Protocol. Next, we developed a multiplex PCR protocol using novel reaction conditions: specifically low individual primer concentrations, high primer annealing temperatures (>65 °C) and long annealing times, which allows amplification of products covering the whole genome in two reactions (**Fig. 3**). In comparison with single-plex methods, this markedly reduces the cost of reagents and minimizes potential sources of laboratory error. We assign alternate target genome regions to one of two primer pools, so that neighboring amplicons do not overlap within the same pool (which would result in a short overlap product being generated preferentially). By screening reaction conditions based on the concentration of cleaned-up PCR products and specificity as determined by gel electrophoresis, we determined that lower primer concentrations and a longer annealing/extension time were optimal. Given the low cost of the assay, this step could also be performed alongside standard diagnostic quantitative PCR as a quality control measure to help reveal potential false positives[50].

Sequencing protocol optimizations. Optimized library preparation methods for both the MinION and Illumina MiSeq platforms are provided and should be readily adaptable to other sequencing platforms, if required. The MinION system is preferred when portability and ease of setup in harsh environments are important[5]. The Illumina platform is more suited to sequencing very large number of samples, because of greater sequence yields,

TABLE 3 | Results of amplicon scheme sequencing of five Zika-positive clinical samples collected from Colombia in January 2016 using the ZikaAsian scheme on the Illumina MiSeq (unpublished data, contributed by K.G.A. and N.D.G.).

Sample	Ct[a]	GEs/μl RNA[b]	No. of Reads[c]	No. of ZIKV reads[d]	% ZIKV[e]	% Coverage[f]	Depth[g]	GenBank ID[h]
ZC188	34	36	1,114,568	760,976	68.3	98.3	29,395	KY317936
ZC192	33.9	39	1,246,644	795,474	63.8	98.4	30,396	KY317937
ZC199	35.9	10	2,772,457	379,064	13.7	93.4	14,848	KY317938
ZC204	33.6	48	1,065,517	751,872	70.6	99.7	29,003	KY317939
ZC207	35.2	16	939,820	506,821	53.9	96.7	19,478	KY317940

[a]qRT-PCR cycle threshold (Ct) value. [b]Genome copy equivalents (GEs) per microliter calculated from Ct value and standard dilutions. [c]Number of reads generated for this sample. [d]Number of reads aligning to the Zika virus reference genome. [e]Percentage of reads mapping to the Zika virus reference genome. [f]Percentage of Zika virus genome covered by at least one read. [g]Mean depth of coverage for the Zika virus genome. [h]GenBank accession number for deposited sequence.

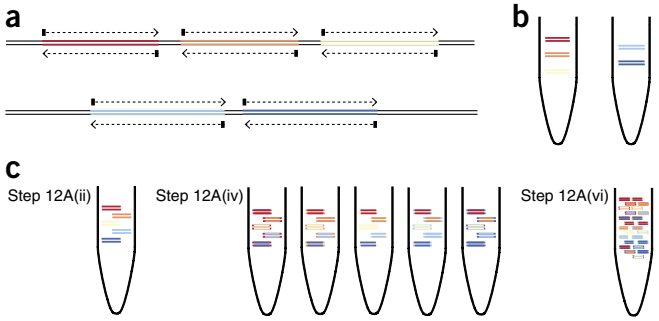

Figure 3 | Overview of multiplex tiling PCR and pooling. (**a**) Schematic showing the regions amplified in pools 1 (upper track) and 2 (lower track), and the intended overlap between pools (as determined in Step 1). (**b**) Products generated by PCR in Step 9 from pools 1 (left tube) and 2 (right tube) for the hypothetical scheme shown in **a**. (**c**) In Step 12A(ii), the input amount is normalized based on the number of samples and the scheme length; pool 1 and 2 products can be pooled at this stage (shown) or kept separate if you wish to barcode them individually. In Step 12A(iv), products for each sample are then barcoded by ligation of a unique barcode. In Step 12A(vi), all barcoded products are pooled together before sequencing adaptor ligation, yielding a sequenceable library.

and the ability to barcode and accurately demultiplex hundreds of samples. Both platforms use ligation-based methods to add the required sequencing adaptors and barcodes.

For the MinION, we used the native barcoding kit (Oxford Nanopore Technologies) to allow up to 12 samples to be sequenced per flow cell. As the manufacturer's protocol is designed for 6–8 kb of fragmented genomic DNA, we have adjusted the input mass to achieve an equivalent number of moles of DNA ends; this improves the efficiency of barcode/adaptor ligation and improves run yields. In the development of the protocol, we used R9 or R9.4 flow cells (FLO-MIN105/FLO-MIN106) and the 2D barcoded library preparation kit (EXP-NBD002/SQK-LSK208). The protocol is also compatible with the current 1D barcoded library preparation kit (EXP-NBD103/SQK-LSK108). Because of the regular revisions of the kits, we have avoided including any specific component names or volumes; be sure to follow the appropriate protocol for your chosen kit version. Depending on the number of reads required, the number of samples multiplexed and the performance of the flow cell, sequencing on the MinION can take from a few minutes up to 72 h. Typically, 2–4 h of sequencing is sufficient for 12 samples. For the MiSeq platform, we used the Agilent SureSelect^{xt2} adaptors and the KAPA Hyper library preparation kit, allowing up to 96 samples to be sequenced per MiSeq run. Other library prep kits (e.g., Illumina TruSeq) and dual-indexed adaptors could also be used on the MiSeq. For the MiSeq, we recommend using the 2 × 250-nt read-length for 400-nt amplicons, which takes 48 h to complete.

Bioinformatics workflow; MinION pipeline. We developed bioinformatic pipelines consisting of primer trimming, alignment, variant calling and consensus generation for both the Oxford Nanopore and Illumina platforms. The MinION pipeline was developed by building upon tools previously developed for Ebola virus sequencing in Guinea and is freely available with components developed under the permissive MIT open source license at https://github.com/zibraproject/zika-pipeline. The pipeline runs under the Linux operating system and is available as a Docker image, which means that it can also be run on Mac and Windows operating systems. The MinION version of the pipeline can process the data from basecalled reads to consensus sequences on the instrument laptop, given the correct primer scheme (a BED file).

FAST5 reads containing raw nanopore signal data may be basecalled in real time using MinKNOW (accessible via the MinION Community Portal for registered users at http://community.nanoporetech.com) or off-line using Albacore. Albacore is a recurrent neural network (RNN) basecaller developed by Oxford Nanopore Technologies and also made available through the MinION Community Portal. Reads are extracted into a FASTA file using the `poretools fasta` command. This FASTA file may be demultiplexed by a script, demultiplex.py, into separate FASTA files for each barcode, as specified in a config file. By default, these are set to the barcodes NB01–12 from the native barcoding kit. Alternatively, the Metrichor online service (https://www.metrichor.com) and versions of Albacore 1.0.1 or later may be used to basecall read files and demultiplex samples. Each file is then mapped to the reference genome using `bwa mem` using the `-x ont2d` flag and converted to BAM format using `samtools view`. Alignments are preprocessed using a script (`align_trim.py`) that performs primer trimming and coverage normalization. Primer trimming is performed by reference to the expected coordinates of sequenced amplicons, and therefore requires no knowledge of the sequencing adaptor (**Fig. 3**). Signal-level events are aligned and variants are called using `nanopolish variants`. Low-quality or low-coverage variants are filtered out and consensus sequences are generated using a script, `margin_cons.py`. Variant calls and frequencies can be visualized using `vcfextract.py` and `pdf_tree.py`.

Bioinformatics workflow; Illumina pipeline. First, we use Trimmomatic[51] to remove primer sequences (first 22 nt from the 5′ end of the reads) and bases at both ends with Phred quality scores <20. Reads are aligned to the genome of a Zika virus isolate from the Dominican Republic, 2016 (GenBank: KU853012), using Novoalign v3.04.04 (http://www.novocraft.com/support/download/). SAMtools is used to sort the aligned BAM files and to generate alignment statistics[52]. The code and reference indexes for the pipeline can be found at https://github.com/andersen-lab/zika-pipeline. Snakemake is used as the workflow management system[53].

Alignment-based consensus generation. We have used an alignment-based consensus approach to generate genomes as opposed to *de novo* assembly. Although *de novo* assembly could in theory be used with this protocol, the use of a tiling amplicon scheme already assumes that the viral genome is present in a particular fixed order. This assumption may be violated in the presence of large-scale recombination. Some *de novo* assemblers, such as SPAdes, use a frequency-based error correction preprocessing stage, and this may result in primer sequences being artificially introduced into the reference if primer sequences are not removed in advance[54]. Importantly, when we compared alignment with *de novo*-based analysis methods for our generated Zika virus genomes, we found that we always obtained the same consensus sequences.

Preparing sequencing controls. We recommend that positive sample controls be included in each sequencing run. To check that

the protocol is generating the expected results, we recommend choosing a positive sample with an established, trusted reference sequence. For the Zika virus, we used the previously sequenced World Health Organization reference strain PF13/251013-18 (GenBank accession: KX369547), which can be obtained on request from the Paul-Ehrlich-Institut[55,56]. Sample archives such as the National Collection of Pathogenic Viruses in the United Kingdom can provide high-quality reference materials for other viruses. Positive controls should have viral copy numbers similar to those of the clinical samples on the same run. This may require the positive control to be heavily diluted until the Ct values are comparable. Negative sequencing controls should be processed in a manner as similar as possible to that used for clinical samples and should not be simply water controls; for example, if samples are collected by swabs, then the same type of unused swab should be subjected to RNA extraction and PCR. Additional negative water controls may be added at each step (e.g., reverse transcription, PCR and library preparation) to detect the sources of contaminants. Even if amplification is not detected (e.g., by gel electrophoresis) or DNA quantity is low or undetectable by fluorimetry, a sequencing library should still be prepared as normal using the total available amount, as contamination may still be detectable by sequencing.

Contamination. Cross-contamination is a serious potential problem when working with amplicon sequencing. Contamination risk is minimized by maintaining physical separation between pre- and post-PCR areas, and performing regular decontamination of work surfaces and equipment—e.g., by UV exposure or with 1% (vol/vol) sodium hypochlorite solution. Contamination becomes harder to mitigate with decreasing viral copy numbers. Processing high-viral-count samples can lead to overamplification during PCR (e.g., generation of unnecessarily high numbers of amplicons), which can increase the risk of amplicon contamination in subsequently processed samples with low viral counts. Such 'between-sample amplification' can occur during sequencing library preparation, or may result from barcode misidentification or 'barcode hopping' (incorporation of incorrect barcode sequences during sequence library preparation) during sequencing. When determining how many PCR cycles to use, begin with a lower number and increase gradually to minimize this contamination risk.

The best safeguard for helping to detect contamination is the use of negative controls. These controls should be sequenced even if no DNA is detected by quantification or no visible band is present on a gel. Negative control samples should be analyzed through the same software pipeline as is used for the other samples, and you should assume that any contaminating amplicons in the negative control will also be present in your other samples. The relative number of reads as compared with positive samples gives a simple guide to the extent of contamination, and inspection of coverage plots can help identify any specific region involved.

MATERIALS

REAGENTS

Tiling amplicon generation

- Clinical sample (serum, plasma, urine) or isolate ! CAUTION Any potentially infectious clinical samples should be handled and made safe in accordance with biosafety regulations. If unsure, contact your local safety officer. ! CAUTION Please follow local institutional review board guidelines covering the collection and storage of clinical samples for research purposes. Our study was evaluated and approved by institutional review boards (IRBs) at The Scripps Research Institute and relevant local IRBs in Colombia and Brazil for Zika and chikungunya sample collection and sequencing.
- QIAamp Viral RNA Mini Kit (Qiagen, cat. no. 52906) ! CAUTION Please consult the MSDS document for safety information on specific kits.
- Random hexamers (50 µM; Thermo Fisher Scientific, cat. no. N8080127)
- Protoscript II First Strand cDNA Synthesis Kit (NEB, cat. no. E6560)
- dNTP solution mix (NEB, cat. no. N0447)
- Q5 Hot Start High-Fidelity DNA Polymerase (NEB, cat. no. M0493) ▲ CRITICAL The primer annealing temperatures are optimized for these reagents. While others may work, thermocycling conditions may need to be optimized.
- PCR primers (listed in **Supplementary Tables 1** and **2** (Integrated DNA Technologies))
- Agencourt AMPure XP (Beckman Coulter, cat. no. A63881)
- Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32854)
- HyPure Molecular Biology Grade Water (GE Life Sciences, cat. no. SH30538.01)
- EB buffer (10 mM Tris-Cl, pH 8.5; Qiagen, cat. no. 19086)
- TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0; Sigma-Aldrich, cat. no. 93283-500ML)
- Ethanol, absolute (Thermo Fisher Scientific, cat. no. BP28184)

Gel electrophoresis

- Gel Loading Dye, Purple (6×) (NEB, cat. no. B7024)
- 100-bp DNA Ladder (NEB, cat. no. N3231)
- SeaKem LE Agarose (Lonza, cat. no. 5000)
- 10× TBE Buffer (Lonza, cat. no. 50843)
- SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, cat. no. S33102)

MinION sequencing

- MinION Flow Cell (Oxford Nanopore Technologies, cat. no. FLO-MIN106)
- Nanopore Sequencing Kit (Oxford Nanopore Technologies, cat. no. SQK-LSK108)
- Native Barcoding Kit (Oxford Nanopore Technologies, cat. no. EXP-NBD103)
- NEB Next Ultra II End-repair/dA-tailing Module (NEB, cat. no. E7546)
- NEB Blunt/TA Ligase Master Mix (NEB, cat. no. M0367)

MiSeq sequencing

- KAPA Hyper Prep Kit (Roche, cat. no. 07962363001)
- SureSelect^{xt2} indexes, MSQ, 16 (Agilent, cat. no. G9622A)
- MiSeq Reagent Kit v2 (500 cycle) (Illumina, cat. no. MS-102-2003)
- D1000 ScreenTape (Agilent, cat. no. 5067-5582)
- D1000 Reagents (Agilent, cat. no. 5067-5583)
- KAPA Library Quantification Kit for Illumina platforms (Roche, cat. no. 07960140001)

EQUIPMENT

Standard equipment

- Filtered pipette tips
- 1.5-ml microcentrifuge tube (Eppendorf, cat. no. 0030 108.051)
- 0.2-ml strip tubes with attached caps (Thermo Fisher Scientific, cat. no. AB2000)
- UV spectrophotometer (Thermo Fisher Scientific NanoDrop 2000, cat. no. ND-2000)
- 96-well thermocycler (Applied Biosystems Veriti, cat. no. 4375786)
- Benchtop microcentrifuge (Thermo Fisher Scientific mySPIN 6, cat. no. 75004061)
- Benchtop heater/shaker (Eppendorf ThermoMixer C)
- Magnetic rack (Thermo Fisher Scientific DynaMag-2, cat. no. 12321D)
- PCR cabinet or pre-PCR room

MinION sequencing

- MinION (Oxford Nanopore Technologies, cat. no. MinION Mk1B)
- Laptop with solid-state disk (SSD) drive

MiSeq sequencing

- MiSeq (Illumina)
- TapeStation 2200 (Agilent)

Gel electrophoresis

- Mini-Sub Cell GT (Bio-Rad, cat. no. 1704406)
- PowerPac Universal Power Supply (Bio-Rad, cat. no. 1645070)

PROCEDURE

Design and ordering of primers ● TIMING 1 h

1| (Optional) Identify representative reference sequences (e.g., from public databases such as GenBank) and generate a primer scheme using Primal Scheme by visiting http://primal.zibraproject.org; see EXPERIMENTAL DESIGN section for further information. Alternatively, predesigned primer schemes are provided in **Supplementary Tables 1** and **2**.
▲ CRITICAL STEP Choose an amplicon length that is suitable for your sequencing platform, and the likely viral copy number of your sample—e.g., 300–500 nt for Zika on MinION/Illumina.
? TROUBLESHOOTING

2| Order the primers generated in Step 1 by the online tool or from the predesigned schemes provided in **Supplementary Tables 1** and **2** from your oligonucleotide supplier.
▲ CRITICAL STEP Primers can be ordered prediluted in TE buffer to 100 µM (usually at additional cost) to avoid manually resuspending a large number of primers.

Extraction and preparation of nucleic acid template

3| Tiling amplification is a general technique that can be applied to DNA or to cDNA generated from RNA by reverse transcription. Use option A to extract viral RNA from samples and prepare cDNA by reverse transcription for analysis of RNA viruses, or use option B to extract viral DNA from samples for analysis of DNA viruses:

(A) RNA extraction and preparation of cDNA for analysis of RNA viruses ● TIMING 2 h

(i) Extract RNA from 200 µl of serum, plasma or urine using the QIAamp Viral RNA Mini Kit according to the manufacturer's instructions, eluting in 50 µl of EB buffer.

(ii) Measure the absorption spectra using a spectrophotometer. Pure RNA should have a 260/280 ratio of 2.0 and a 260/230 ratio of 2.0–2.2.
? TROUBLESHOOTING

(iii) Wash all surfaces with 1% (vol/vol) sodium hypochlorite solution and irradiate labware with UV light for at least 10 min.
▲ CRITICAL STEP Perform the following three steps in a hood or dedicated pre-PCR area.

(iv) Mix the following components in a 0.2-ml tube:

Component	Amount (µl)	Final concentration
Template RNA (from Step 3A(i))	7	
Random hexamers, 50 µM	1	2.5 µM

(v) Denature the template RNA by incubating it on a heat block at 65 °C for 5 min before promptly placing it on ice.
▲ CRITICAL STEP This denaturation step minimizes the secondary structure in the RNA before cDNA synthesis.

(vi) Complete the cDNA synthesis reaction preparation by adding the following to the tube:

Component	Amount (µl)	Final concentration
ProtoScript II Reaction mix (2×)	10	1×
ProtoScript II Enzyme Mix (10×)	2	1×

(vii) Place the tube in a thermocycler and run the following program:

Cycle number	Condition
1	25 °C, 5 min
1	48 °C, 15 min
1	80 °C, 5 min

■ PAUSE POINT cDNA can be stored at –20 °C for a month.

(B) DNA extraction and preparation for analysis of DNA viruses ● TIMING 1 h

(i) Extract DNA from 200 µl of serum, plasma or urine using the QIAamp MinElute Virus Spin Kit, according to the manufacturer's instructions.

(ii) Measure the absorption spectra using a spectrophotometer. Pure DNA should have a 260/280 ratio of 1.8 and a 260/230 ratio of 2.0–2.2.

? TROUBLESHOOTING

Preparation of the primer pools ● TIMING 1 h

4| (Optional) Resuspend lyophilized primers by prespinning tubes to make sure that the pellet is at the bottom of the tube and adding TE buffer to a concentration of 100 µM. If primers were ordered prediluted to 100 µM, continue to the next step.

▲ CRITICAL STEP The volume of TE buffer needed to yield a 100 µM solution is often given on the QC document supplied with the primer.

5| Label two 1.5-ml Eppendorf tubes using the scheme and pool, which are numbered as either '1' or '2'; primers for adjacent regions are added to alternate pools so that individual reactions overlap between pools but not within. Add an equal volume of each 100 µM primer stock such that both the forward and reverse primers for alternate regions are pooled together. For example, Pool '1' for a ZikaAsian scheme would contain ZIKA_400_1_LEFT, ZIKA_400_1_RIGHT, ZIKA_400_3_LEFT, ZIKA_400_3_RIGHT, ZIKA_400_5_LEFT, ZIKA_400_5_RIGHT and so on. Dilute these at a ratio of 1:10 with nuclease-free water to a working concentration of 10 µM.

Performing of multiplex tiling PCR ● TIMING 5 h

6| In Eppendorf tubes, prepare a mastermix for each of the 2 primer pools, as follows.

Component	Amount (µl)	Final concentration
Q5 reaction buffer (5×)	5	1×
dNTPs, 10 mM	0.5	200 µM
Q5 DNA polymerase	0.25	
Primer pool 1 or 2 (10 µM)	Variable	0.015 µM per primer
PCR-grade water	Up to 22.5 µl (assuming 2.5 µl of cDNA template will be added in Step 7)	

Mix thoroughly by vortexing and spin down in a microcentrifuge.

▲ CRITICAL STEP The total volume of mastermix should be 22.5 µl multiplied by the number of samples plus 10% excess volume; this is done to reduce variability between reactions.

▲ CRITICAL STEP The volume of primers to use will depend on the number of primers in the pool, as the final concentration should be 0.015 µM per primer. For example, the ZikaAsian scheme from **Supplementary Table 1** has 36 primers in pool 1, so the volume to use would be 1.35 µl.

7| Label 0.2-ml PCR tubes and add 22.5 µl of mastermix from Step 6 to each tube. If using cDNA from Step 3A(vii) as template, add 2.5 µl of cDNA to each tube. If you are using extracted DNA from Step 3B(ii), however, a larger volume of template (up to 10 µl) can be added, if required, and may improve amplification efficiency.

▲ CRITICAL STEP It is recommended that cDNA volume be kept to 10% of the final volume of the PCR reaction to avoid affecting the buffer conditions.

▲ CRITICAL STEP This step should ideally be performed in a cabinet used only for template addition in order to minimize the risk of amplicon contamination.

8| Place in a thermocycler and run the following program:

Cycle number	Denature	Anneal/extend
1	98 °C, 30 s	
2–40	98 °C, 15 s	65 °C, 5 m

Cleanup and quantification of amplicons ● TIMING 2 h

9| Transfer the contents of the tubes to 1.5-ml Eppendorf tubes. Add the volume of AMPure XP beads given in the table below, taking into account amplicon length. Perform washes following the instructions in the 1D barcoding protocol and elute in 30 µl of EB buffer.

Amplicon length (bp)	Ratio	Volume of beads (µl) for a 25-µl PCR reaction
<500	1.0×	25
500–1,000	0.8×	20
>1,000	0.6×	15

10| Quantify 1 µl of the cleaned products using the Qubit instrument with the high-sensitivity assay per the manufacturer's instructions. You should expect concentrations in the range of 5–50 ng/µl for each reaction from the Qubit quantification, except for the PCR negative control, which should be repeated if >1 ng/µl.

? TROUBLESHOOTING

■ **PAUSE POINT** Cleaned-up PCR products can be stored at –20 °C for up to a month.

11| (Optional) Make a gel by melting 1% (wt/vol) agarose powder in 1× TBE buffer and then adding 1× SYBR Safe gel stain before allowing it to set. Place in a gel tank submerged in 1× TBE buffer. Mix 10 µl of cleaned product from Step 9 or a ladder with 2 µl of 6× loading dye and load on the gel. Perform electrophoresis at 6 V/cm until bands are distinguishable by transillumination. A specific band of the correct size for your scheme should be observed.

? TROUBLESHOOTING

Library preparation and sequencing

12| Perform library preparation and sequencing; these procedures are platform specific and have been validated on the MinION from Oxford Nanopore Technologies (option A) and on the MiSeq from Illumina (option B).

(A) Library preparation and sequencing using the MinION ● TIMING 1–2 d

(i) *Determine the number of samples per flow cell*. We recommend using two barcodes per sample or negative control (one barcode per pool per sample) initially. This means that up to five samples and one negative control can be sequenced on each flow cell, and it allows each pool to be barcoded individually, making it easier to detect contamination that may be pool- rather than sample-specific. However, a single barcode per sample can also be used to maximize the number of samples per flow cell.

(ii) *Normalization*. Use the table below to determine the quantity of amplicons to load to achieve a total input of ~0.3 pM per flow cell. Divide the total input quantity by the number of barcodes being used to calculate the quantity per barcode. Keep PCR products separate at this stage; add the appropriate volume of each sample from Step 9 to individual 1.5-ml Eppendorf tubes and then adjust the volume in each Eppendorf to 20 µl with nuclease-free water.

Amplicon length (bp)	Input total (ng)
300	60
400	80
500	100
1,000	200
1,500	300
2,000	400
5,000	1,000

? TROUBLESHOOTING

(iii) *End-repair and dA-tailing*. For each sample, set up the following end-repair/dA-tailing reaction in a 1.5-ml Eppendorf tube and incubate for 5 min at 20 °C, followed by 5 min at 65 °C. Perform SPRI cleanup by repeating Step 9, eluting in 10 µl of EB buffer.

Component	Amount (µl)
Normalized amplicons (from Step 12A(ii))	20
Ultra II End Prep Reaction Buffer	2.8
Ultra II End Prep Enzyme Mix	1.2

(iv) *Barcode ligation*. In a 1.5-ml Eppendorf tube, prepare the following ligation reactions—one reaction per barcode being used.

Component	Amount (µl)
dA-tailed amplicons (from Step 12A(iii))	10
Native barcode NB01-NB12	2.5
Blunt/TA Ligase Master Mix	12.5

(v) Incubate at room temperature (20 °C) for 10 min, followed by 65 °C for 10 min to denature the ligase.

(vi) *Pool barcoded amplicons*. Combine all the barcode ligation reactions into a single 1.5-ml Eppendorf tube. Perform SPRI cleanup by repeating Step 9 and elute in 30 µl of nuclease-free water.
▲ **CRITICAL STEP** If the pellet is large, you can speed up drying by briefly incubating at 50 °C; do not allow the pellet to overdry and crack, or recovery will be reduced.

(vii) *Barcoding adaptor ligation*. In a 1.5-ml Eppendorf tube, prepare the adaptor ligation reaction according to the native barcoding kit protocol. As Blunt/TA ligase is a 2× master mix, we have reduced the elution volume in the previous step to accommodate this.
▲ **CRITICAL STEP** We have found that using Blunt/TA Ligase instead of the NEBNext Quick Ligation Module as described in the protocol improves the efficiency of this step.

(viii) *Library cleanup and elution*. Complete library construction according to the native barcoding kit protocol.

(ix) *Library loading*. Perform library loading according to the native barcoding kit protocol; there is a helpful video guide to loading the library on the MinION Community Portal (http://community.nanoporetech.com).
? TROUBLESHOOTING

(x) *Start sequencing run*. By default, you will need an Internet connection before the sequencing script can be started, although off-line versions of MinKNOW can be requested from the manufacturer if an Internet connection is not available. Once the flow cell is detected, enter an experiment name and the flow cell ID into the blank fields and then choose the appropriate sequencing script for the library preparation kit and flow cell version.
▲ **CRITICAL STEP** Note that if a 'Live' basecalling script is run, reads will be basecalled in real time. If you do this, then you do not need to perform the basecalling step when running the subsequent bioinformatic pipeline.
? TROUBLESHOOTING

(B) Library preparation and sequencing using the MiSeq ● TIMING 3 d

(i) Determine the number of samples per flow cell; we recommend using two barcodes per sample, which means up to 47 samples plus a negative control can be sequenced on each run and allows each pool to be barcoded individually. This makes it easier to detect contamination that may be pool- rather than sample-specific. It also results in greater yield per sample, which improves genome coverage in samples with more uneven amplification.

(ii) *Normalization*. Keep pools in individual 1.5-ml Eppendorf tubes, add 50 ng of sample material from Step 9 and add nuclease-free water to adjust the total volume to 50 µl.

(iii) *End-repair and dA-tailing*. Perform end-repair and dA-tailing according to the Hyper Prep Kit protocol (KAPA).

(iv) *Library preparation*. Complete library construction with the KAPA Hyper Prep Kit according to the manufacturer's instructions, substituting the KAPA adaptors for SureSelect^{xt2} indexing adaptors in the adaptor ligation step. Perform a 0.8× instead of 1× SPRI cleanup during the postamplification cleanup to remove potential adaptor–dimers.

(v) *Library quality control*. Measure the size distribution of the library using the TapeStation 2200 according to the manufacturer's instructions.

(vi) *Library pooling*. Calculate the molarity of each library using the KAPA Library Quantification Kit according to the manufacturer's instructions and pool libraries in an equimolar manner.
(vii) *Library denaturation and dilution*. Prepare library for loading onto the MiSeq according to the manufacturer's instructions.
(viii) *Start sequencing run*. Generate a SampleSheet.csv file with the Illumina Experiment Manager software by entering sample and barcode information. Complete the instrument setup and start the sequencing run according to the manufacturer's recommended instructions.
(ix) *Basecalling and demultiplexing*. Basecalling and demultiplexing will be performed automatically on the instrument using the sample information provided in the SampleSheet.csv file.

Data analysis ● TIMING 1–2 h

13| Download the Docker application for Linux, Macintosh or Windows from https://www.docker.com/products/overview. Run the installer to set up the Docker tools on your machine. You should now be able to open a terminal window and run the command `docker --version` without getting an error.

14| Download the Zika pipeline image from DockerHub by typing `docker pull zibra/zibra` into the terminal window.
▲ **CRITICAL STEP** The source code of the Zika analysis pipeline is also available from https://github.com/zibraproject/zika-pipeline.
▲ **CRITICAL STEP** The Zika pipeline is compatible with both MinION data and Illumina data, yet there are some differences in the data handling required.

15| Start a Docker container with the following command:

```
docker run -t -i zibra/zibra:latest
```

By default, Docker containers do not have access to the file system of the computer they run within. You will need to provide access to a local directory in order to see data files. This is achieved using the `-v` parameter. You may need to grant access to Docker to share the drive via the 'Shared Drives' menu option under 'Settings'. For example, on Windows, if you wish to provide access to the c:\data\reads directory to the Docker container, use the following:

```
docker run -v c:/data/reads:/data -t -i zibra/zibra:latest /bin/bash
```

Then, within the Docker container, the `/data` directory will refer to `c:\data\reads` on the Windows machine.

16| Run the platform-specific pipeline using option A for MinION data or option B for MiSeq data:
(A) Running analysis pipeline on MinION data
(i) Ensure that the reads are basecalled using either Metrichor or an off-line basecaller. Compatible off-line basecallers include Albacore (available as installable packages for Linux, Windows and Macintosh through the MinION Community Portal) or the freely available and open-source nanonet (https://github.com/nanoporetech/nanonet) software. nanonet is compatible with graphics processing unit cards to increase speed.
(ii) Metrichor will perform demultiplexing if a barcoding workflow is selected. For other basecallers you may need to demultiplex reads manually. To do this, run the script that is provided within the Docker image with the command:

```
demultiplex <directory of FAST5 Files> <output directory>
```

(iii) Run the Zika pipeline using the following command:

```
fast5_to_consensus <scheme> <sampleID> <directory>
```

The pipeline takes three required items:

`sample_id` —the sample name (should not contain space characters)

`directory` —the directory containing the FAST5 files for a single sample (e.g., demultiplexed output directory from Step 16A(ii))

`scheme` —the name of the scheme directory—e.g., ZikaAsian

For example:

```
fast5_to_consensus ZikaAsian Zika1 /data/NB08/downloads/pass
```

Output files will be written to the current directory. The final consensus file will be named `<sampleID>.consensus.fasta`

(B) Running analysis pipeline on Illumina data

(i) Download and follow the instructions for the Illumina pipeline by referring to https://github.com/andersen-lab/zika-pipeline and using the following command:

```
illumina_pipeline <sampleID> <fastq1> <fastq2> <scheme>
```

Quality control ● TIMING 1 h

17| Check the coverage of the genomes by reference to the alignment file. Use an alignment viewer such as IGV[57] or Tablet[58] and load the `<sampleID>.primertrimmed.sorted.bam` file in conjunction with the reference sequence. Amplicons should be evenly spread throughout the genome. Deep piles of reads representing amplification of single regions are potential warning signs of contamination. Compare the alignments with the positive and negative control alignments to help indicate problematic samples or regions.

? TROUBLESHOOTING

18| Use the variant frequency plot produced by the Zika pipeline to help determine the allele frequency of mutations in the sample (as compared with the reference). The variant frequency plot is given the name `<sampleID>.variants.png` and is generated from the `<sampleID>.variants.tab` file that can be opened in spreadsheet applications or statistical software. The principle of the variant frequency plot is to identify mutations that occur at lower-than-expected allele frequencies and help decide whether they are a biological phenomenon (e.g., intra-host single-nucleotide variants), potential signs of contamination or sequencing errors (for example, in homopolymeric tracts in MinION data).

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 4**.

TABLE 4 | Troubleshooting table.

Step	Problem	Possible reason	Solution
1	Primal Scheme failed to generate a full scheme	Using a short amplicon length can result in no suitable primers because of local sequence context	Increase the amplicon length and retry
3A(ii) or 3B(ii)	260/280 value lower than 2.0 for RNA or 1.8 for DNA	Carryover of guanidinium thiocyanate from the lysis buffer	Repeat extraction using additional washes to remove unwanted salt from the column
10	No amplification	Sample may be a diagnostic false positive by qRT-PCR	Repeat qRT-PCR to confirm that the sample is positive
	Poor amplification	RNA may be degraded by RNAses in sample (serum/plasma) or incorrect storage	Avoid freezing and thawing of samples; use a shorter amplicon scheme
	Amplification in negative control	Amplicon contamination in PCR setup area	Wash all surfaces with 1% sodium hypochlorite solution and irradiate labware with UV light for at least 10 min
	Amplification in one pool only	Suboptimal primer concentration	Check that the reaction was set up correctly; repeat failed pool and adjust the primer concentration.
11	No specific band on the gel	Nonspecific amplification of host DNA	Attempt to treat the sample with DNase; use serum or plasma

(continued)

TABLE 4 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
12A(ii)	Insufficient amplicons to make library	Amplicon concentration needed is higher per sample when you have small numbers of samples	Amplify more samples to run together; 5–10 is optimal
12A(ix)	Trouble loading library via SpotON port	Siphon has stopped	Wait 10 min for any existing library to tether, and then reprime the flow cell and retry library loading
12A(x)	MinKNOW fails to start script	Problem with MinKNOW installation	Reinstall MinKNOW from scratch and restart the script
17	Insufficient reads	Incorrect molarity of library or insufficient number of active pores on the flow cell	Run another flow cell and combine the data
	Amplification of some regions but not others	Too many mismatched bases in primers	Re-design primers using a less divergent reference genome

● TIMING

Steps 1 and 2, design and ordering of primers: 1 h
Step 3A, RNA extraction and preparation of cDNA: 2 h
Step 3B, DNA extraction: 1 h
Steps 4 and 5, preparation of the primer pools: 1 h
Steps 6–8, performing of multiplex tiling PCR: 5 h
Steps 9–11, cleanup and quantification of amplicons: 1 h
Step 12A, library preparation and sequencing using the MinION: 1–2 d
Step 12B, library preparation and sequencing using the MiSeq: 3 d
Steps 13–16, data analysis: 1–2 h
Steps 17 and 18, quality control of consensus sequences: 1 h

ANTICIPATED RESULTS

This protocol should achieve near-complete genome coverage.

MinION sequencing

As a demonstration of the ZikaAsian scheme on MinION, we sequenced the World Health Organization Zika reference sample 11474/16[55] (**Table 2**) and a chikungunya clinical sample from Brazil, PEI-N11602. The Ct value for the Zika virus sample was between 18 and 20 depending on the RNA extraction method used. The Ct value for the Chikungunya sample was 20, as determined by the RealStar Chikungunya RT-PCR Kit 1.0 from Altona Diagnostics (Hamburg, Germany). The Zika virus sample generated 97.7% coverage of the genome above 25× coverage. Coverage of the genome was reasonably even, with a dropout in the middle of the genome (**Fig. 4**). The WHO Control Reference MinION data set is available from the CLIMB website (https://s3.climb.ac.uk/nanopore/Zika_Control_Material_R9.4_2D.tar).

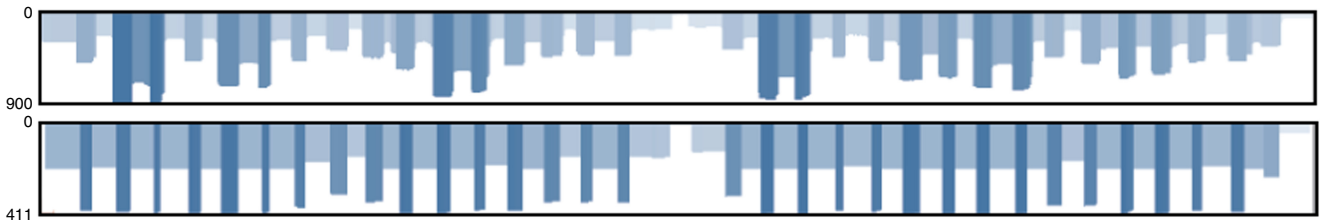

Figure 4 | Coverage plots for ZikaAsian scheme sequenced on MinION before (top panel) and after primer trimming and coverage normalization (bottom panel). During the preprocessing step, reads are trimmed using a BED file containing primer positions, and read coverage is normalized. The coverage plot was produced using the Tablet genome viewer[58] with the Zika reference genome coordinates represented on the *x* axis and depth of coverage on the *y* axis. Alignments are colored by depth of coverage, with darker regions indicating higher depths of coverage—e.g., in overlapping regions.

Illumina sequencing

We compared metagenomic sequencing with the ZikaAsian scheme with the Illumina MiSeq protocol using five clinical samples of Zika from Colombia. Using a previously described method for metagenomics sequencing[2,17], only a small percentage (<0.01%) of our reads aligned to Zika virus and they covered only a fraction of the genome (**Table 1**). Using the ZikaAsian scheme, we were able to generate high coverage of all the genomes (**Table 3**). Illumina sequencing reads are available from BioProject PRJNA358078 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA358078).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS The authors thank the Brazilian Ministry of Health and the Latin American Community Engagement Networks (LACENs) of Natal, João Pessoa, Recife, Maceió and Salvador for their support. We thank T. Fredeking from Antibody Systems for providing the Zika virus samples from Colombia. We thank K. Brunker for testing the Primal Scheme software. The Zika in Brazil Real-time Analysis (ZiBRA) project (http://www.zibraproject.org) is supported by the Medical Research Council/Wellcome Trust/Newton Fund Zika Rapid Response Initiative (grant no, ZK/16-078), which also provides J.Q.'s salary. N.J.L. is supported by a Medical Research Council Bioinformatics Fellowship as part of the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) project. Primal Scheme is hosted on the CLIMB platform, where pipeline development and MinION data analysis was performed[59]. N.D.G. is supported by National Institutes of Health (NIH) training grant 5T32AI007244-33. K.G.A. is a PEW Biomedical Scholar, and his work is supported by NIH National Center for Advancing Translational Studies Clinical and Translational Science Award UL1TR001114 and National Institute of Allergy and Infectious Diseases (NIAID) contract HHSN272201400048C. A.B. and T.B. were supported by NIH awards R35 GM119774 and U54 GM111274. T.B. is a Pew Biomedical Scholar. A.B. is supported by the National Science Foundation Graduate Research Fellowship Program under Grant no. DGE-1256082. N.R.F. was funded by a Sir Henry Dale Fellowship (Wellcome Trust/Royal Society grant 204311/Z/16/Z). Work at the Paul-Ehrlich-Institut was supported by a grant ('Sicherheit von Blut(produkten) und Geweben hinsichtlich der Abwesenheit von Zikaviren') from the German Ministry of Health. This study was supported by USAID Emerging Pandemic Threats Program-2 PREDICT-2 (cooperative agreement AID-OAA-A-14-00102). The contents of this article are the responsibility of the authors and do not necessarily reflect the views of USAID or the US government.

AUTHOR CONTRIBUTIONS J.Q. and N.J.L. conceived the project. J.Q., N.D.G., K.G.A. and N.J.L. designed the experiments and wrote the manuscript. J.Q., A.D.S. and O.G.P. built the online primer design tool. J.T.S. modified nanopolish to support R9/R9.4 data and indels. M.L. wrote the demultiplexing software. N.J.L. designed and implemented the MinION bioinformatics pipeline. N.J.L., T.B. and K.G. built the Docker image. N.D.G., K.G., G.O., R.R.-S. and K.G.A. developed the Illumina sequencing protocol and bioinformatics pipeline. L.L.L.-X. collected the chikungunya sample, performed clinical diagnosis and received local approvals. S.A.B. performed molecular diagnostics and curated Zika and Chikungunya control material. N.D.G., S.T.P., I.M.C., K.G., G.O., R.R.-S., T.F.R., N.A.B., J.G.d.J., M.G., S.H. and A.B. performed the experiments. All other authors tested the protocol and provided feedback. All authors have read and approved the contents of the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details are available in the online version of the paper.

1. Gardy, J., Loman, N.J. & Rambaut, A. Real-time digital pathogen surveillance - the time is now. *Genome Biol.* **16**, 155 (2015).
2. Andersen, K.G. *et al.* Clinical sequencing uncovers origins and evolution of lassa virus. *Cell* **162**, 738–750 (2015).
3. Holmes, E.C., Dudas, G., Rambaut, A. & Andersen, K.G. The evolution of Ebola virus: Insights from the 2013-2016 epidemic. *Nature* **538**, 193–200 (2016).
4. Dudas, G. *et al.* Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **544**, 309–315 (2017).
5. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
6. Arias, A. *et al.* Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evol.* **2**, vew016 (2016).
7. Palacios, G. *et al.* A new arenavirus in a cluster of fatal transplant-associated diseases. *N. Engl. J. Med.* **358**, 991–998 (2008).
8. Nakamura, S. *et al.* Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* **4**, e4219 (2009).
9. Wilson, M.R. *et al.* Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* **370**, 2408–2417 (2014).
10. Metzker, M.L. Sequencing technologies—the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
11. Loman, N.J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012).
12. Goodwin, S., McPherson, J.D. & McCombie, W.R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
13. Gire, S.K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
14. Carroll, M.W. *et al.* Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. *Nature* **524**, 97–101 (2015).
15. Greninger, A.L. *et al.* Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* **29**, 99 (2015).
16. Faria, N.R. *et al.* Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* **8**, 97 (2016).
17. Matranga, C.B. *et al.* Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol.* **15**, 519 (2014).
18. Mamanova, L. *et al.* Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **7**, 111–118 (2010).
19. Salzberg, S.L. *et al.* Genome analysis linking recent European and African influenza (H5N1) viruses. *Emerg. Infect. Dis.* **13**, 713–718 (2007).
20. Li, K. *et al.* Automated degenerate PCR primer design for high-throughput sequencing improves efficiency of viral sequencing. *Virol. J.* **9**, 261 (2012).
21. Yu, Q. *et al.* PriSM: a primer selection and matching tool for amplification and sequencing of viral genomes. *Bioinformatics* **27**, 266–267 (2011).
22. Worobey, M. *et al.* 1970s and 'Patient 0'HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* **539**, 98–101 (2016).
23. Faria, N.R., Quick, J., Morales, I., Theze, J. & de Jesus, J.G. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* http://dx.doi.org/10.1038/nature22401 (2017).
24. Grubaugh, N.D. *et al.* Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* http://dx.doi.org/10.1038/nature22400 (2017).
25. Metsky, H.C. *et al.* Zika virus evolution and spread in the Americas. *Nature* http://dx.doi.org/10.1038/nature22402 (2017).
26. Márquez, S. *et al.* First complete genome sequences of Zika virus isolated from febrile patient sera in Ecuador. *Genome Announc.* **5**, e01673-16 (2017).
27. Houldcroft, C.J., Beale, M.A. & Breuer, J. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* **15**, 183–192 (2017).
28. Kilianski, A. *et al.* Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience* **4**, 12 (2015).
29. Wang, J., Moore, N.E., Deng, Y.-M., Eccles, D.A. & Hall, R.J. MinION nanopore sequencing of an influenza genome. *Front. Microbiol.* **6**, 766 (2015).
30. Hoenen, T. *et al.* Nanopore sequencing as a rapidly deployable Ebola outbreak tool. *Emerg. Infect. Dis.* **22**, 331–334 (2016).
31. Hysom, D.A. *et al.* Skip the alignment: degenerate, multiplex primer and probe design using K-mer matching instead of alignments. *PLoS One* **7**, e34560 (2012).

32. Gardner, S.N. *et al.* Multiplex degenerate primer design for targeted whole genome amplification of many viral genomes. *Adv. Bioinformatics* **2014**, 101894 (2014).
33. Guerbois, M. *et al.* Outbreak of Zika virus infection, Chiapas State, Mexico, 2015, and first confirmed transmission by *Aedes aegypti* mosquitoes in the Americas. *J. Infect. Dis.* **214**, 1349–1356 (2016).
34. Atkinson, B. *et al.* Complete genome sequence of Zika virus isolated from semen. *Genome Announc.* **4**, e01116–16 (2016).
35. Misencik, M.J., Grubaugh, N.D., Andreadis, T.G., Ebel, G.D. & Armstrong, P.M. Isolation of a novel insect-specific Flavivirus from *Culiseta melanura* in the Northeastern United States. *Vector Borne Zoonotic Dis.* **16**, 181–190 (2016).
36. Carrillo, C. *et al.* Genetic and phenotypic variation of foot-and-mouth disease virus during serial passages in a natural host. *J. Virol.* **81**, 11341–11351 (2007).
37. Yuste, E., López-Galíndez, C. & Domingo, E. Unusual distribution of mutations associated with serial bottleneck passages of human immunodeficiency virus type 1. *J. Virol.* **74**, 9546–9552 (2000).
38. Batty, E.M. *et al.* A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. *PLoS One* **8**, e66129 (2013).
39. Brown, J.R. *et al.* SureSelect target enrichment: a robust and sensitive method for sequencing of whole norovirus genomes direct from clinical specimens. *J. Clin. Virol.* **70**, S12–S13 (2015).
40. Brown, J.R. *et al.* Norovirus whole-genome sequencing by SureSelect target enrichment: a robust and sensitive method. *J. Clin. Microbiol.* **54**, 2530–2537 (2016).
41. Thomson, E. *et al.* Comparison of next generation sequencing technologies for the comprehensive assessment of full-length hepatitis C viral genomes. *J. Clin. Microbiol.* **54**, 2470–2484 (2016).
42. Eckert, S.E. *et al.* Enrichment by hybridisation of long DNA fragments for Nanopore sequencing. *Microb. Genomics* **2** http://dx.doi.org/10.1099/mgen.0.000087 (2016).
43. Naccache, S.N. *et al.* Distinct Zika virus lineage in Salvador, Bahia, Brazil. *Emerg. Infect. Dis.* **22**, 1788–1792 (2016).
44. Kilianski, A. *et al.* Use of unamplified RNA/cDNA-hybrid nanopore sequencing for rapid detection and characterization of RNA viruses. *Emerg. Infect. Dis.* **22**, 1448–1451 (2016).
45. Sipos, B., Young, S., Juul, S., Clarke, J. & Turner, D.J. Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv* http://dx.doi.org/10.1101/068809 (2016).
46. Scotto-Lavino, E., Du, G. & Frohman, M.A. 5′ end cDNA amplification using classic RACE. *Nat. Protoc.* **1**, 2555–2562 (2006).
47. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
48. Kwok, S. *et al.* Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* **18**, 999–1005 (1990).
49. Kwok, S., Chang, S.Y., Sninsky, J.J. & Wang, A. A guide to the design and use of mismatched and degenerate primers. *PCR Methods Appl.* **3**, S39–S47 (1994).
50. Chan, M. *et al.* A novel system control for quality control of diagnostic tests based on next-generation sequencing. *J. Appl. Lab. Med.* **1**, 25–35 (2016).
51. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
52. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
54. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
55. Trösemeier, J.-H. *et al.* Genome sequence of a candidate World Health Organization reference strain of Zika virus for nucleic acid testing. *Genome Announc.* **4**, e00917–16 (2016).
56. Baylis, S.A. *et al.* Harmonization of nucleic acid testing for Zika virus: development of the 1st World Health Organization International Standard. *Transfusion* **57**, 748–761 (2017).
57. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
58. Milne, I. *et al.* Tablet—next generation sequence assembly visualization. *Bioinformatics* **26**, 401–402 (2010).
59. Connor, T.R. *et al.* CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microbial Genomics* **2** http://dx.doi.org/10.1099/mgen.0.000086 (2016).