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

Quick, Joshua; Grubaugh, Nathan D; Pullan, Steven T; Claro, Ingra M; Smith, Andrew D; Gangavarapu, Karthik; Oliveira, Glenn; Robles-Sikisaka, Refugio; Rogers, Thomas F; Beutler, Nathan A; Burton, Dennis R; Lewis-Ximenez, Lia Laura; de Jesus, Jaqueline Goes; Giovanetti, Marta; Hill, Sarah C; Black, Allison; Bedford, Trevor; Carroll, Miles W; Nunes, Marcio; Alcantara, Luiz Carlos

Document Version
Peer reviewed version

[Link to publication on Research at Birmingham portal](#)

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

Josh 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 Hill[10], Allison Black[11,12], Trevor Bedford[11], Miles W Carroll[3,13], Marcio Nunes[14], Luiz Carlos Jr. Alcantara[7], Ester C Sabino[4], Sally A Baylis[15], Nuno R Faria[9], 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, CA, 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, CA, USA. [6]Massachusetts General Hospital, Boston, USA. [7]Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro [8]Fiocruz Bahia, Salvador, Brazil. [9]University of Rome, Tor Vergata. [10]Department of Zoology, University of Oxford, Oxford, UK. [11]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, USA. [12]Department of Epidemiology, University of Washington, Seattle, USA. [13]University of Southampton, South General Hospital, Southampton, UK. [14]Instituto Evandro Chagas, Belem, Brazil. [15]Paul-Ehrlich-Institut, Langen, Germany. [16]University of Nottingham, Nottingham, UK. [17]OICR, Toronto, Canada.
 Correspondence should be addressed to N.J.L (n.j.loman@bham.ac.uk).

Genome sequencing has become a powerful tool for studying emerging infectious diseases; however, genome sequencing directly from clinical samples (i.e. without isolation) remains challenging for viruses such as Zika, where 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, optimised 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 starting with clinical samples in 1-2 days 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.

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 (without culture isolation) 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, fuelled 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 due to the extremely high virus copy numbers found in acute cases [13–15]. However, direct metagenomic sequencing from clinical samples poses challenges with regards 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. 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 Ct values between 33.9 and 35.9 (equivalent to 10-48 genome copies per microlitre). 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 dataset, resulting in incomplete coverage. Greater coverage and depth is critical for accurate genome reconstruction and subsequent phylogenetic inference. Additionally, there are significant sequencing and storage costs associated with generating large sequencing datasets, and these approaches currently do not lend themselves to lower-throughput portable sequencing devices such as the Oxford Nanopore MinION.

Target enrichment is often required to generate complete viral genome coverage from clinical samples in an economic way [18]. Enrichment can be achieved directly through i) isolation in culture, ii) oligonucleotide bait probes targeting the virus of interest, or iii) indirectly via host nucleic acid depletion. Amplification may also be required to generate sufficient material for sequencing (>5 ng for typical Illumina protocols, and 100-1000 ng for MinION). Polymerase chain reaction (PCR) can provide both target enrichment and amplification in a single step, and is relatively cheap, available, and fast compared to other methods. In order to generate coding-sequence complete coverage, a tiling amplicon scheme is commonly employed [19–21]. During our work with Ebola virus we were able to reliably amplify >95% of the genome by sequencing eleven long amplicons (1-2.5 kb in length) on the MinION [5]. The likelihood of long fragments being present in the sample reduces with lower virus abundance. Therefore, we anticipated that for viruses like 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 over 40 years and used 200-300 nucleotide (nt) amplicons to help maximise sequence recovery [22].

Using shorter amplicons necessitates a larger number of products to generate a tiling path across a target genome. To do this in individual reactions would require a large number of manual pipetting steps and therefore the potential for mistakes and 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 allows any researcher to successfully amplify and sequence low abundance viruses directly from clinical samples. The method has other potential uses that are not demonstrated here. One potential application is to generate a panel for multi-locus sequencing typing genes in 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.

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 number of primer pairs. It is believed that increasing numbers of primer pairs reduces the likelihood of successful amplification of each region due to interaction between primers [18]. 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 into haplotypes, of importance for investigation of within-host diversity. On the Illumina platform, 600 bases is the maximum size of amplicon that can be obtained without an additional fragmentation step using this protocol (using 600 cycle kits in paired-end mode, i.e. paired 300 bases without any overlap), although read accuracy may degrade significantly during the last 50 cycles . On the Oxford Nanopore MinION, there is no limit to the maximum amplicon length that can be sequenced; the effective limitation will relate to the performance of the reverse transcription and polymerase chain reaction step (practically around 5 kb). However, longer amplicons are less likely to amplify successfully when viral copy number is low, or there is significant sample degradation (e.g. due to inadequate storage).

Limitations of tiling amplicon sequencing

Our method is not suitable for the discovery of new viruses or sequencing highly diverse or recombinant viruses as primer schemes are virus genome-specific. Amplicon sequencing is prone to coverage dropouts that may result in incomplete genome coverage, especially at lower abundances, and the loss of both the 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 [27]. 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].

The majority of primers are expected to work even when pooled in equimolar amounts, meaning largely complete genomes can be recovered without optimisation. The chikungunya virus data shown in **Table 2** was generated without any optimisation. However, to achieve coding-sequence complete genomes,

problem primers may need to be replaced or concentration relative to other primers adjusted in an iterative manner. Full coverage should be achievable for the majority of samples, however coverage is expected to correlate with viral abundance (**Table 3**). 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. 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. It is plausible that as the number of primer pairs increases that competitive inhibition may decrease PCR efficiency, although the high annealing temperature employed in this protocol should reduce this risk.

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 steps; i) multiplex primer pool design, ii) multiplex PCR protocol, iii) sequencing on MinION or Illumina instruments and iv) bioinformatics analysis and QC (**Figure 1**).

Primer Design

We developed a web-based primer design tool entitled Primal Scheme (https://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. Together, the amplicons generated by the pairs span the target genome or region of interest (**Figure 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 2000 nt), and the desired length of overlap between neighbouring amplicons (default 75 nt). Using a shorter amplicon length may be useful for samples where 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 steps:

- *Generate candidate primers:* The first sequence listed in the FASTA file should represent the most representative genome, with further sequences spanning the expected interhost diversity. Primal Scheme uses the Primer3 software to generate candidate primers [28]. This chooses primers based on highly accurate thermodynamic modeling to take into account length, annealing temperature, %GC, 3' stability, estimated secondary structure and likelihood of primer-dimer formation maximising the chance of a successful PCR reaction. Primers are designed with a high annealing temperature within a narrow range (65-68°C) which allows PCR to 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.
- *Test candidate primers:* Subsequent reference genomes in the file are used to help choose primer pairs that maximise the likelihood of successful amplification for known virus diversity. A

pairwise semi-global alignment score between each candidate primer and reference is calculated to ensure the most 'universal' candidate primers are picked for the scheme. Mismatches at the 3' end are severely penalised due as they have a disproportionate effect on the likelihood of successful extension [29,30]. The pairwise distances are summed, and the minimum scoring primer pairs are selected.

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

Multiplex PCR Protocol

Next, we developed a multiplex PCR protocol employing 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 (**Figure 3**). In comparison to single-plex methods, this dramatically reduces the cost of reagents and minimises potential sources of laboratory error. We assign alternate target genome regions to one of two primer pools, so that neighbouring 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 concentration of cleaned-up PCR products and specificity as determined by gel electrophoresis, we determined that lower primer concentrations and longer annealing/extension time were optimal. Given the low-cost of the assay this step could also be performed alongside standard diagnostic qRT-PCR as a quality control measure to help reveal potential false positives [31].

Sequencing protocol optimisations

Optimised 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 is important [5]. The Illumina platform is more suited to sequencing very large numbers of samples, due to greater sequence yields and the ability to barcode and accurately demultiplex hundreds of samples. Both platforms employ ligation-based methods to add the required sequencing adapters and barcodes. For the MinION we utilised the native barcoding kit (Oxford Nanopore Technologies, Oxford, UK) to allow up to 12 samples to be sequenced per flowcell. As the manufacturer's protocol is designed for 6-8 kb 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/adapter ligation and improves run yields. For the work presented here we used R9 or R9.4 flowcells (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). Due to 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. For the MiSeq platform the Agilent SureSelect[xt2] adaptors and Kapa Hyper library preparation kit was used and allowed up to 96 samples per MiSeq run. Other library prep kits (e.g. Illumina TruSeq) and dual-indexed adaptors could also be used on the MiSeq. Depending on the number of reads required, the number of samples multiplexed, and the performance of the flowcell, sequencing on the MinION can be run for anything from a few minutes up to 72 hours. Typically 2-4 hours of

sequencing is sufficient for 12 samples. For the MiSeq we recommend using the 2 x 250 nt read-length for 400 nt amplicons, which takes 48 hours to complete.

Bioinformatics workflow

A bioinformatics pipeline consisting of primer trimming, alignment, variant calling, and consensus generation is made available for both Oxford Nanopore and Illumina platforms, the former building upon tools previously developed for Ebola virus sequencing in Guinea. The MinION pipeline 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 meaning that it can also be run on Mac and Windows operating systems. This MinION pipeline can process the data from raw reads to consensus sequences on the instrument laptop given the correct primer scheme as a BED file and sample sheet as required input files.

MinION pipeline

FAST5 reads containing signal-level information may be basecalled in real-time using MinKNOW or offline using Albacore. Albacore is a recurrent neural network (RNN) basecaller developed by Oxford Nanopore Technologies and made available through the MinION Community Portal. The resulting FASTA files are 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 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 that performs primer trimming and coverage normalisation. Primer trimming is performed by reference to the expected coordinates of sequenced amplicons, and therefore requires no knowledge of the sequencing adaptor (**Figure 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 visualised using `vcfextract.py` and `pdf_tree.py`.

Illumina Pipeline

First, we use Trimmomatic [32] to remove primer sequences (first 22 nt from the 5' end of the reads) and bases at both ends with Phred quality score <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 [33]. 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 [34].

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 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, employ 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 [35]. Importantly, when we compared alignment to *de novo*-based analysis methods of our generated Zika virus genomes, we found that we always obtained the same consensus sequences.

Preparing Sequencing Controls

We recommend positive sample controls are included in each sequencing run. In order to check that the protocol is generating expected results we recommend choosing a positive sample with an established, trusted reference sequence. For 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 [36,37]. Sample archives such as the National Collection of Pathogenic Viruses (NCPV) in the United Kingdom can provide high quality reference materials for other viruses. Positive controls should have similar viral copy numbers to 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 as similar way as possible as 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 if 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.

A Note on Contamination

Cross-contamination is a serious potential problem when working with amplicon sequencing. Contamination risk is minimised 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% sodium hypochlorite solution. Contamination becomes harder to mitigate against as the viral load in a sample reduces. Processing high viral count samples (e.g. post-isolation) can increase the risk of amplicon 'bleed-over' into subsequent low viral count samples. When determining how many PCR cycles to use, begin with a lower number and increase gradually to avoid generating excessive quantities of amplicons making physical separation harder.

The best safeguard to help 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 analysed through the same software pipeline as the other samples. The relative number of reads compared with positive samples give a simple guide to the extent of contamination, and inspection of coverage plots demonstrates can help determine whether a specific region is involved, and potentially its origin by reference to the sequence.

Comparison with other approaches

The three most common approaches for sequencing viruses are metagenomic sequencing, PCR amplification sequencing and target enrichment sequencing, recently reviewed in detail by Houldcroft et al. [38]. The main benefit to the PCR based approach described here is cost and sensitivity. In theory both PCR and cell culture require only one starting molecule/virion, making them both exquisitely sensitive. In practice, however, the reaction conditions do not allow single-genome amplification and typically a number of starting molecules are required. PCR also has limited sensitivity in cases where the template sequence is divergent from the expected due to primer binding kinetics. This makes it more suitable for use in an outbreak situation where isolates are highly related and low cost per sample and rapid turnaround time are required. Sequencing amplicons is a popular method to determine viral genomes on the Oxford Nanopore MinION 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 pooled before sequencing [5,39–41]. However, these approaches are laborious to scale up when many small amplicons are required (due to 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). However, this protocol is specific to the Ion Torrent platform and primer schemes must be ordered directly from the manufacturer, and is consequently more expensive per sample. Alternative software to design primer schemes is available, some of which cater specifically for multiplex or tiling amplicon schemes [20,21,42,43], and these may perform better when dealing with divergent genomes due to an increased emphasis on oligonucleotide degeneracy. Primers generated with such software may also be compatible with this protocol, although PCR conditions may require optimisation. In contrast, Primal Scheme is designed with an emphasis on monitoring short-term evolution of known lineages and primer conditions have been optimised for multiplex PCR amplification efficiency.

Propagation with cell culture has been widely used for virus enrichment [44–46]. This process is time-consuming, requires specialist expertise, and requires high containment laboratories for especially dangerous pathogens. There is also concern that viral passage can introduce mutations not present in the original clinical sample, potentially confounding analysis [47,48].

Oligonucleotide bait probes have also shown promise as an alternative to metagenomics and amplicon sequencing [49–53]. This method fishes for viral nucleic acid sequences by hybridising target-specific biotinylated probes to the DNA/RNA sample then pulls them down using magnetic streptavidin beads. These methods, however, are limited by the efficiency of the capture step due to the kinetics of nucleic acid hybridisation in complex samples, such as those containing the human genome. The complete hybridisation 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 due to 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 Zika virus [54]. These methods work best on higher abundance samples, and may not have the sensitivity to generate near-complete genomes for the majority of isolates in an outbreak. Probes for hybridisation capture are 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 [55,56]. This method is attractive because it eliminates the need for reverse transcription, and so potentially may reduce biases resulting from non-random priming, and copying errors introduced by reverse transcriptase. However, currently this method 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.

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. This study has been evaluated and approved by Institutional Review Boards 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 MSDS document for safety information on specific kits
Random Hexamers (50 μM) (Thermo Fisher, cat. no. N8080127)
Protoscript II First Strand cDNA Synthesis Kit (NEB, cat. no. E6560)
Deoxynucleotide (dNTP) Solution Mix (NEB, cat. no. N0447)
Q5 Hot Start High-Fidelity DNA Polymerase (NEB, cat. no. M0493) **CRITICAL** The primer annealing temperatures are optimised for these reagents. Whilst others may work, thermocycling conditions may need to be optimised.
PCR primers are listed in **Supplementary Table 1 and Table 2** (Integrated DNA Technologies)
Agencourt AMPure XP (Beckman Coulter, cat. no. A63881)
Qubit dsDNA HS Assay Kit (Thermo Fisher, cat. no. Q32854)
HyPure Molecular Biology Grade Water (GE Life Sciences, cat. no. SH30538.01)
EB buffer (10 mM Tris-Cl, pH 8.5)
TE buffer (10 mM Tris-Cl pH 8.0, 0.1 mM EDTA)
Ethanol, absolute (Thermo Fisher, cat. no. BP28184)

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

NEBNext 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 Library Prep (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
Filtered pipette tips
1.5 ml microcentrifuge tube (Eppendorf, cat. no. 0030 108.051)
0.2 ml strip tubes with attached caps (Thermo Fisher, cat. no AB2000)
UV spectrophotometer (Thermo Fisher NanoDrop 2000, cat. no. ND-2000)
96-well thermocycler (Applied Biosystems Veriti, cat. no. 4375786)
Benchtop microcentrifuge (Thermo Fisher mySPIN 6, cat. no. 75004061)
Benchtop heater/shaker (Eppendorf ThermoMixer C)
Magnetic rack (ThermoFisher 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 primal.zibraproject.org. Choose an amplicon length that is suitable for your sequencing platform, and likely viral copy number of your sample (e.g. 300-500 nt for Zika on MinION/Illumina).

2| Order the primers generated by the online tool or from the pre-designed schemes provided in the **Supplementary Table 1 and 2** from your oligonucleotide supplier. Primers can be ordered pre-diluted in TE to 100 µM (usually at addition cost) to avoid manually resuspending large number of primers.

Extraction of nucleic acids and preparation of cDNA if required TIMING 1-2 h

3| Tiling amplification is a general technique which can be applied to DNA or cDNA generated from RNA by reverse transcription. Follow the appropriate instructions for your application.

    (A) RNA extraction and preparation of cDNA for RNA viruses

        (i) Extract RNA from 200 µl of either serum, plasma or urine using the QIAamp Viral RNA Mini Kit according to the manufacturer's instructions eluting in 50 µl 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) Perform the following two steps in a hood or dedicated pre-PCR area. Wash all surfaces with 1% sodium hypochlorite solution and irradiate labware with UV light for at least 10 minutes. Sample negative controls for each primer pool should be included on each PCR run.

        CRITICAL STEP

        (iv) Mix the following in a 0.2 ml tube. This denaturation step minimises secondary structure in the RNA prior to cDNA synthesis.

Component	Amount (µl)	Final concentration
Template RNA	7	
Random hexamers, 50 µM	1	2.5 µM

        (v) Denature the template RNA by incubating on a heat block at 65 °C for 5 minutes before placing promptly on ice.

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

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

        (vii) Place on 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

    (B) DNA extraction for DNA viruses

        (i) Extract DNA from 200 µl of either serum, plasma or urine using the QIAamp MinElute Virus Spin Kit.

        (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

Preparing the primer pools TIMING 1 h

4| To resuspend lyophilised primers, pre-spin tubes to make sure pellet is at the bottom of the tube and add TE to a concentration of 100 µM. The volume of TE to give 100 µM is often given on the QC document supplied with the primer. If they are already pre-diluted to 100 µM continue to the next step.

5| Label two 1.5 ml Eppendorf tubes using the scheme and pool, either '1' or '2'. Add an equal volume of each 100 µM primer stock such that both the forward and reverse primer for alternate regions are pooled together. For example. Pool '1' for 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 etc. Dilute these 1:10 to a working concentration of 10 µM.

Performing multiplex tiling PCR TIMING 5 h

6| In Eppendorf tubes prepare a mastermix for each of the primer pools 1 and 2. Mix thoroughly by vortexing and spin down in a microfuge. Label 0.2 mL PCR tubes and add 22.5 µl (if using 2.5 µl cDNA) mastermix to each tube. A master mix including 10% excess volume is used for setting up PCR as it reduces variability between reactions. 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.

Component	Amount (µl)	Final concentration
Q5 reaction buffer (5x)	5	1x
dNTPs, 10 mM	0.5	200 µM
Q5 DNA polymerase	0.25	
Primer pool 1 or 2 (10 µM)	Varies	0.015 µM per primer
PCR grade water	Varies	

7| Add 2.5 ul cDNA from step 3(A)(vii) to each tube. It is recommended that cDNA volume be kept to 10% of the final volume of PCR reaction to avoid significantly affecting the buffer conditions. If you are using extracted DNA from step 3(B), a larger volume be added if required. This step should ideally be performed cabinet used only for template addition to minimise risk of amplicon contamination.

8| Place on a thermocycler 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 1 h

9| Transfer contents of the tube to a 1.5 ml Eppendorf tube. Add the volume of SPRI 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 nuclease-free water or EB.

Amplicon length	Ratio	Volume of beads in µl (assumes 25 µl PCR reaction)
<500	1.0x	25
500-1000	0.8x	20
>1000	0.6x	15

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

11| (Optional) Make a gel by melting 1% (w/v) agarose powder in 1X TBE then adding 1X SYBR Safe gel stain before allowing it to set. Place in a gel tank submerged in 1X TBE. Mix 10 µl cleaned product or ladder with 2 µl 6x 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 TIMING Up to 1-2 d
12| Library preparation and sequencing are platform specific and and have been validated on the MinION from Oxford Nanopore Technologies and on the MiSeq from Illumina.

   (A) Library preparation and sequencing using the MinION

      (i) *Number of samples per flowcell.* We recommend using two barcodes per sample or negative control (one barcode per pool per sample) initially. This makes it easier to detect contamination that may be pool- rather than sample-specific. However a single barcode per sample can also be used to maximise the number of samples per flowcell.

      (ii) *Normalisation*. Use the table below to determine the quantity of amplicons to load to achieve a total input of ~0.3 pM. The total quantity must be divided by the number of barcodes being used to calculate the quantity per barcode. PCR products should be kept separate at this stage, and the appropriate volume added to individual 1.5 ml Eppendorf tubes. The volume in each Eppendorf is then adjusted to 20 µl with nuclease-free water.

Scheme length	Input total (ng)
300	60
400	80
500	100
1000	200
1500	300
2000	400

5000   1000

TROUBLESHOOTING

(iii) *End-repair and dA-tailing.* For each sample setup the following end-repair/dA-tailing reaction in a 1.5 ml Eppendorf tube and incubate for 5 minutes at 20°C followed by 5 minutes at 65°C. Perform a SPRI cleanup as per step 9, eluting each in 10 µl nuclease-free water or EB.

Component	Amount (µl)
Amplicons	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 prepare the following ligation reactions, one reaction per barcode being used.

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

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

(vi) *Pool barcoded amplicons.* Combine all the barcode ligation reactions into a single 1.5 ml Eppendorf tube. Perform cleanup as before using the ratio of SPRI beads used previously and elute in 30 µl nuclease-free water. 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 adapter ligation*. In a 1.5 ml Eppendorf tube prepare the adapter ligation reaction according to the native barcoding protocol. 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. As this is a 2X master mix we have adjusted the elution volume in the previous step to accommodate this.

(viii) *Library cleanup and elution.* Complete library construction according to the native barcoding protocol. In the last step elute the library in 14.5 ul and use the whole volume when preparing the library dilution.

(ix) *Library loading.* Perform library loading and sequencing according to the manufacturer's instructions. If the siphon action of the SpotON flowcell stops while loading library, push a little air into the inlet port using the volume adjustment screw of a pipette then back it out again and resume loading.

TROUBLESHOOTING

(x) *Start sequencing run.* By default you will need an internet connection before the sequencing script can be started, although completely offline versions of MinKNOW may be available from the manufacturer on request. Once the flowcell is detected enter an experiment name and the flowcell into the blank fields then choose the appropriate sequencing script for the library preparation kit and flowcell version. If a 'Live' basecalling script is run, reads will be basecalled in real-time. If you do this then you need to remember to skip this step when running the analysis pipeline later.

(B) Library preparation, sequencing and analysis using the MiSeq.

(i) *Number of samples per flowcell.* 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) *Normalisation*. Keeping pools in individual 1.5 ml Eppendorf tubes add 50 ng material and add nuclease-free water to adjust 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 Adapters for SureSelect^{xt2} Indexing Adaptors in the adapter ligation step. A 0.8X instead of 1X SPRI cleanup should be performed during the post-amplification cleanup to remove potential adaptor-dimers.

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

(vi) *Library pooling.* Calculate the molarity of each library using the KAPA Library Quantification according to the manufacturer's instructions and pool libraries in an equimolar fashion.

(vii) *Library dilution and loading.* Prepare library for loading onto the MiSeq according to the manufacturer's instructions.

(viii) *Basecalling and demultiplexing.* Basecalling and demultiplexing will be performed automatically on the instrument if the sample sheet is provided when starting the run.

Data analysis TIMING 1-2 h

13| Download Docker application for Linux, Mac or Windows from (https://www.docker.com/products/overview). Run the installer to setup 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. The source code of the pipeline is also available via The Zika analysis pipeline is available from https://github.com/zibraproject/zika-pipeline.

14| Download the Zika pipeline image from DockerHub by typing `docker pull zibra/zibra` into the terminal window. 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. To test the container is working, run:
```
docker run -t -i zibra/zibra:latest
```
Docker containers do not have access to the file system of the computer they run within by default. 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 wished 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. The Zika pipeline is compatible with both MinION data and Illumina data yet there are some differences in the data handling required.

> **(A) Running analysis pipeline on MinION data.**
>> (i) Ensure the reads are basecalled using either Metrichor or an offline basecaller. Compatible basecallers including Albacore (available as installable packages for Linux, Windows and Mac through the MinION Community Portal) or and the freely-available and open-source nanonet (https://github.com/nanoporetech/nanonet) software. Nanonet is compatible with graphics processing unit (GPU) 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. A script is provided within the Docker image to do this, which is run with the command:
>> `demultiplex.sh <directory of FAST5 Files> <output directory>`
>> (iii) Run the Zika pipeline. 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 previous step)
>> `scheme` -- the name of the scheme directory, e.g. ZikaAsian
>> `fast5_to_consensus.sh <scheme> <sampleID> <directory>`
>> For example:
>> `fast5_to_consensus.sh ZikaAsian Zika1 /data/NB08/downloads/pass`
>> (iv) 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 reference to https://github.com/andersen-lab/zika-pipeline
>> `illumina_pipeline.sh <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 potentially warning signs of contamination. The alignments should be compared with the positive and negative control alignments to help indicate problematic samples or regions.

18| The Zika pipeline produces a variant frequency plot to help determine the allele frequency of mutations in the sample (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.**

TIMING
Steps 1-2, Design and ordering of primers: 1 h
Step 3(A), RNA extraction, QC and cDNA synthesis: 2 h
Step 3(B), DNA extraction and QC: 1 h
Steps 4-5, Preparing the primer pools: 1 h
Steps 6-8, Performing multiplex tiling PCR: 5 h
Steps 9-11, Cleanup and quantification of amplicons: 1 h
Step 12, Library preparation and sequencing: 1-3 d
Steps 13-17, Data analysis: 1-2 h
Steps 18-19, 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 [36] (**Table 2**) and a chikungunya clinical sample from Brazil PEI-N11602**.** The Ct value for the Zika virus sample was between 18-20 depending on 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 25x coverage. Coverage of the genome was reasonably even, with a drop-out in the middle of the genome (**Figure 4**). The WHO Control Reference MinION dataset is available from the CLIMB website (https://s3.climb.ac.uk/nanopore/Zika_Control_Material_R9.4_2D.tar).

Illumina Sequencing

We compared metagenomic sequencing to 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).

REFERENCES

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 the factors that spread and sustained the West African Ebola epidemic. *bioRxiv* 071779 (2016). doi:10.1101/071779

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 Evolution* **2,** (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.* (2015). doi:10.1101/020420

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. Epidemic establishment and cryptic

transmission of Zika virus in Brazil and the Americas. *bioRxiv* (2017).

24. Grubaugh, N. D. *et al.* Multiple introductions of Zika virus into the United States revealed through genomic epidemiology. *bioRxiv* 104794 (2017).

25. Metsky, H. C., Matranga, C. B., Wohl, S. & Schaffner, S. F. Genome sequencing reveals Zika virus diversity and spread in the Americas. *bioRxiv* (2017).

26. Márquez, S. *et al.* First Complete Genome Sequences of Zika Virus Isolated from Febrile Patient Sera in Ecuador. *Genome Announc.* **5,** (2017).

27. Scotto–Lavino, E., Du, G. & Frohman, M. A. 5' end cDNA amplification using classic RACE. *Nat. Protoc.* **1,** 2555–2562 (2006).

28. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132,** 365–386 (2000).

29. 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).

30. 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,** S39–47 (1994).

31. Chan, M. *et al.* A Novel System Control for Quality Control of Diagnostic Tests Based on Next-Generation Sequencing. *The Journal of Applied Laboratory Medicine: An AACC Publication* **1,** 25–35 (2016).

32. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

33. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

34. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28,** 2520–2522 (2012).

35. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).

36. 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).

37. 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).

38. Houldcroft, C. J., Beale, M. A. & Breuer, J. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* **15,** 183–192 (2017).

39. Kilianski, A. *et al.* Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience* **4,** 12 (2015).

40. 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).

41. Hoenen, T. *et al.* Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg. Infect. Dis.* **22,** 331–334 (2016).

42. 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).

43. Gardner, S. N. *et al.* Multiplex degenerate primer design for targeted whole genome amplification of many viral genomes. *Adv. Bioinformatics* **2014,** 101894 (2014).

44. 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).

45. Atkinson, B. *et al.* Complete Genome Sequence of Zika Virus Isolated from Semen. *Genome Announc.* **4,** (2016).

46. 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).

47. 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).

48. 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).

49. 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).

50. 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).

51. 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).

52. Thomson, E. *et al.* Comparison of next generation sequencing technologies for the comprehensive assessment of full-length hepatitis C viral genomes. *J. Clin. Microbiol.* JCM–00330 (2016).

53. Eckert, S. E. *et al.* Enrichment by hybridisation of long DNA fragments for Nanopore sequencing. *Microbial Genomics* **2,** (2016).

54. Naccache, S. N. *et al.* Distinct Zika Virus Lineage in Salvador, Bahia, Brazil. *Emerg. Infect. Dis.* **22,** 1788–1792 (2016).

55. 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).

56. Sipos, B., Young, S., Juul, S., Clarke, J. & Turner, D. J. Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv* (2016).

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,** (2016).

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.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, did clinical diagnosis and received local approvals. S.B. and J.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.J.G, 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

Table 1| Results of metagenomic sequencing on five Zika-positive clinical samples collected from Columbia in January 2016 (unpublished data, contributed by Kristian Andersen and Nathan Grubaugh).

Sample	Ct [1]	GE/ul RNA [2]	# Reads	# ZIKV reads [3]	% ZIKV	% Coverage	Depth
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

[1] qRT-PCR cycle threshold (Ct) value
[2] Genome copy equivalents (GE) per microlitre calculated from Ct value and standard dilutions.
[3] Number of reads aligning to a Zika virus reference genome

Table 2| Results of MinION sequencing after barcode demultiplexing for an isolate of Zika and for a clinical sample of chikungunya virus (Zika data from Faria et al. [23], chikungunya data unpublished contributed by Sally Baylis, Josh Quick and Nicholas Loman).

Sample	Scheme	Pore	Ct	# reads (2D pass)	# Aligned reads	% aligned	% coverage (>25x)	Depth
Zika WHO Control Reference (11474/16)	ZikaAsian	R9.4	18-20	9,606	9,528	99.2	97.7	392
Chikungunya PEI N11602	ChikAsian ECSA	R9.4	20	89,365	88,527	99.1	88	2,779

Table 3| Results of amplicon scheme sequencing on five Zika-positive clinical samples from Colombia using the ZikaAsian scheme on the Illumina MiSeq (unpublished data, contributed by Kristian Andersen and Nathan Grubaugh).

Sample	Ct [1]	GE/ul RNA [2]	Reads	ZIKV reads [3]	% ZIKV	% Coverage	Depth	Genbank ID
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

[1] qRT-PCR cycle threshold (Ct) value

[2] Genome copy equivalents (GE) per microlitre calculated from Ct value and standard dilutions.

[3] Number of reads aligning to a Zika virus reference genome

Table 4 | Troubleshooting table.

Step	Problem	Possible reason	Possible solution
1	Primal scheme failed to generate a full scheme	Using a short amplicon length can results in no suitable primers due to local sequence context	Increase the amplicon length and retry
3(A)(ii) or 3(B) (ii)	260/280 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 false positive	Repeat qRT-PCR to confirm sample is positive
10	Poor amplification	RNA may be degraded by RNAses in sample (serum/plasma) or incorrect storage	Avoid freeze thawing of samples; use a shorter amplicon scheme
10	Amplification in negative control	Amplicon contamination in PCR set-up area	Wash all surfaces with 1% sodium hypochlorite solution and irradiate labware with UV light for at least 10 minutes.
10	Amplification in one pool only	Suboptimal primer concentration	Check reaction was set up correctly; repeat failed pool

			and adjust the primer concentration.
11	No specific band on gel	Non-specific amplification of host DNA	Attempt to DNase treat the sample; use serum or plasma.
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)	MinKNOW fails to start script	Problem with MinKNOW installation	Reinstall MinKNOW from scratch and restart script
17	Insufficient reads	Incorrect molarity library or insufficient active pores on flowcell	Run another flowcell and combine data
17	Amplification of some regions but not others	Too many mismatched bases in primers	Redesign primers using a less divergent reference genome

SUPPLEMENTARY INFORMATION

Supplementary Table 1: List of PCR primers for ZikaAsian scheme

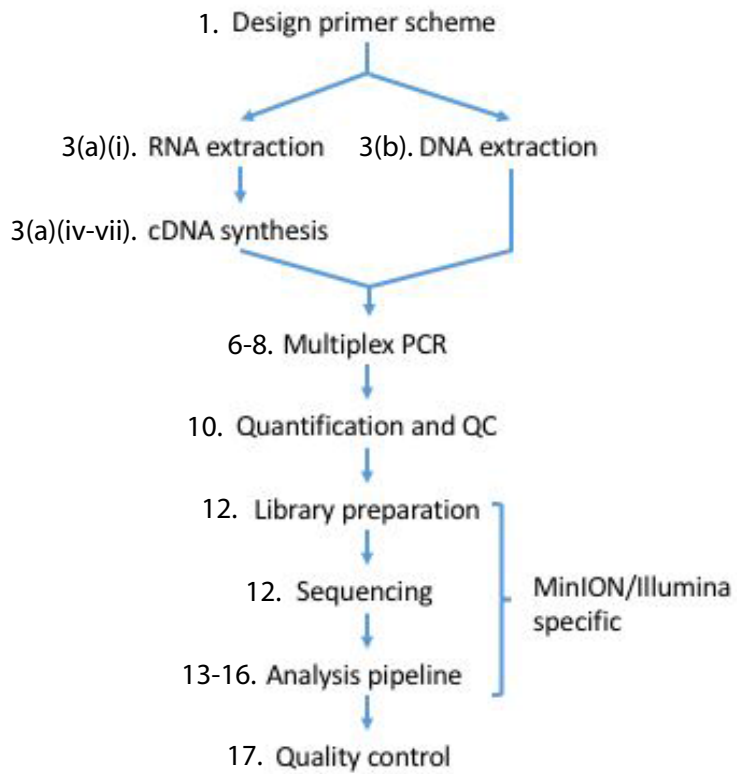**Supplementary Table 2**: List of PCR primers for ChikAsianECSA scheme

Figure 1| Workflow for tiling amplicon sequencing on MinION/Illumina with associated protocol steps.

Figure 2| Overview of multiplex primer design using primal 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.
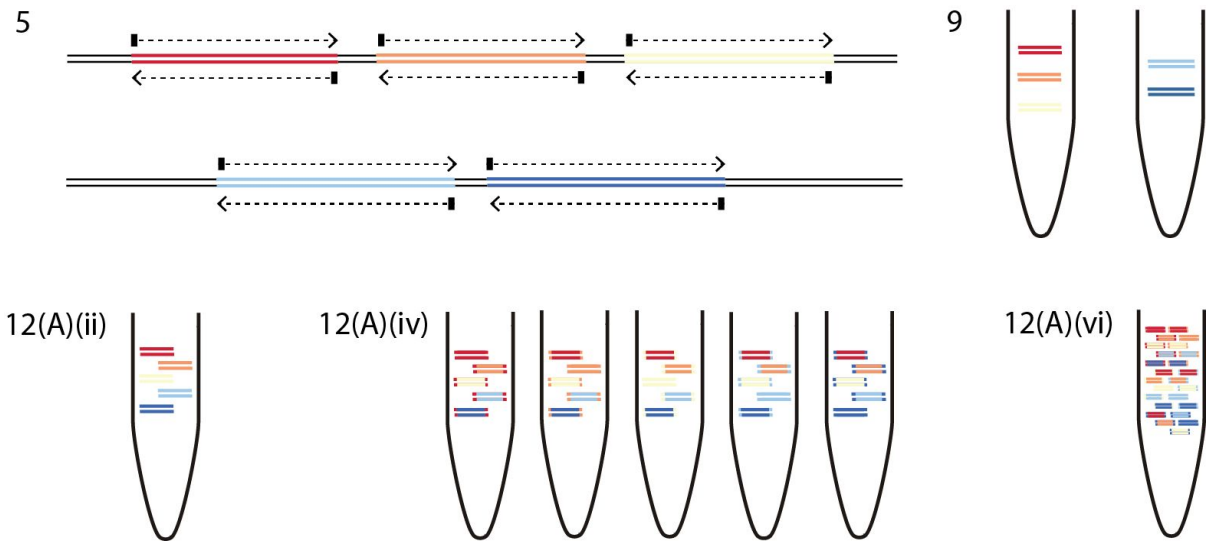
Figure 3| Overview of multiplex tiling PCR and pooling. 5 Schematic showing the regions amplified in pool 1 (upper track) and 2 (lower track) intended overlap between pools. 9 Products generated by pool 1 (left tube) and 2 (right tube) for hypothetical scheme shown in 5. 12(A)(ii) Input amount is normalised based on 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. 12(A)(iv) products for each sample are then barcoded by ligation of a unique barcode. 12(A)(vi) all barcoded products are all pooled together before sequencing adapter ligation yielding a sequenceable library.
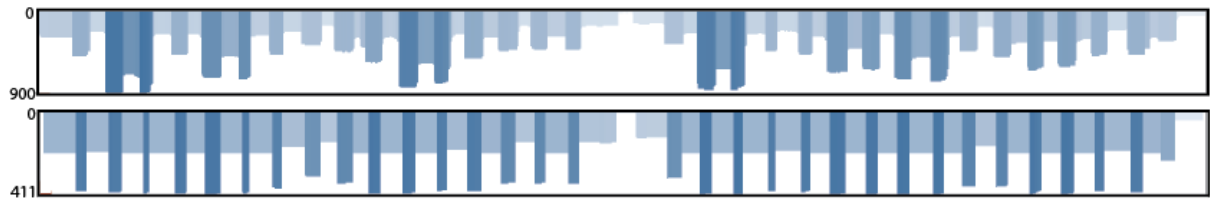
Figure 4| Coverage plots for ZikaAsian scheme sequenced on MinION before (top panel) and after primer trimming and coverage normalisation (bottom panel). During the preprocessing step reads are trimmed using a BED file containing primer positions and read coverage is normalised. The coverage plot was produced using the Tablet genome viewer [30].

SUPPLEMENTARY INFORMATION

Supplementary Table 1 | Zika virus "ZikaAsian" scheme used by ZiBRA project generated by the Primal Scheme software [15].

| Name	Sequence	Pool
ZIKA_400_1_LEFT	GACAGTTCGAGTTTGAAGCGAAAG	1
ZIKA_400_1_RIGHT	AGTATGCACTCCCACGTCTAGT	1
ZIKA_400_2_LEFT	AAGAAAGATCTGGCTGCCATGC	2
ZIKA_400_2_RIGHT	TGATTCCAACCAGGTTTGCGAC	2
ZIKA_400_3_LEFT	AGATGACGTCGATTGTTGGTGC	1
ZIKA_400_3_RIGHT	TACGGTGACACAACCTCCATGT	1
ZIKA_400_4_LEFT	TCAGGTGCATAGGAGTCAGCAA	2
ZIKA_400_4_RIGHT	GGAGCCATGAACTGACAGCATT	2
ZIKA_400_5_LEFT	AGAACGTTAGTGGACAGAGGCT	1
ZIKA_400_5_RIGHT	TGTGCGTCCTTGAACTCTACCA	1
ZIKA_400_6_LEFT	TTGATTGTGAACCGAGGACAGG	2
ZIKA_400_6_RIGHT	CCATCTGTCCCTGCGTACTGTA	2
ZIKA_400_7_LEFT	TGAAGGGCGTGTCATACTCCTT	1
ZIKA_400_7_RIGHT	CGCCTCCAACTGATCCAAAGTC	1
ZIKA_400_8_LEFT	GGGAGAAGAAGATCACCCACCA	2
ZIKA_400_8_RIGHT	TTGACTGCTGCTGCCAATCTAC	2
ZIKA_400_9_LEFT	GCCTTAGGGGGAGTGTTGATCT	1
ZIKA_400_9_RIGHT	GAGTGGGCATTCCTTCAGTGTG	1
ZIKA_400_10_LEFT	ACGGTCGTTGTGGGATCTGTAA	2
ZIKA_400_10_RIGHT	GTGGGACTTTGGCCATTCACAT	2
ZIKA_400_11_LEFT	CAGCCGTTATTGGAACAGCTGT	1
ZIKA_400_11_RIGHT	CCTGGGCCTTATCTCCATTCCA	1
ZIKA_400_12_LEFT	CACTAAGGTCCACGTGGAGGAA	2
ZIKA_400_12_RIGHT	TATCAGCGCCAGATGAGCTACA	2
ZIKA_400_13_LEFT	TGGCAGTGCTGGTAGCTATGAT	1
ZIKA_400_13_RIGHT	AGAGAGAGGAGCATAAACCCCC	1
ZIKA_400_14_LEFT	CAATGGTTTTGCTTTGGCCTGG	2
ZIKA_400_14_RIGHT	TTTCCCATGTGATGTCACCTGC	2
ZIKA_400_15_LEFT	CCCTAGCGAAGTACTCACAGCT	1
ZIKA_400_15_RIGHT	TACACTCCATCTGTGGTCTCCC	1
ZIKA_400_16_LEFT	GTGGCATGAACCCAATAGCCAT	2
ZIKA_400_16_RIGHT	GCTCCAATGTCCCCATCCTTTG	2
ZIKA_400_17_LEFT	GTGGTCCATGGAAGCTAGATGC	1
ZIKA_400_17_RIGHT	CCTCTAAGGGCCTCCTCCATTT	1
ZIKA_400_18_LEFT	CTGTTGAGTGCTTCGAGCCTTC	2

ZIKA_400_18_RIGHT	TGGTGAGTTGGAGTCCGGAAAT	2
ZIKA_400_19_LEFT	TATGGATGAGGCCCACTTCACA	1
ZIKA_400_19_RIGHT	GCCATCAAGTATGACCGGCTTT	1
ZIKA_400_20_LEFT	GGCTGGAAAACGGGTCATACAG	2
ZIKA_400_20_RIGHT	CCTTTGCTCCGTCCTAAGCTTG	2
ZIKA_400_21_LEFT	AGAGACTGACGAAGACCATGCA	1
ZIKA_400_21_RIGHT	CTCCAAAAGCCGCTCCTCTTTT	1
ZIKA_400_22_LEFT	TGGACCAGACACGGAGAGAAAA	2
ZIKA_400_22_RIGHT	ATTCTGGCTGGCTCAATTTCCG	2
ZIKA_400_23_LEFT	CGTCTTGATGAGGAACAAGGGC	1
ZIKA_400_23_RIGHT	AAGTGGTCACTGCATGTTGGAC	1
ZIKA_400_24_LEFT	TAATGGGAAGGAGAGAGGAGGG	2
ZIKA_400_24_RIGHT	TCTCCACTTGGGGGTCAATTGT	2
ZIKA_400_25_LEFT	CCCTGACCCTAATAGTGGCCAT	1
ZIKA_400_25_RIGHT	CCTTCCATTTCTCTCCCAGGGT	1
ZIKA_400_26_LEFT	ACTGGAACTCCTCTACAGCCAC	2
ZIKA_400_26_RIGHT	ACCAGGGCCTCCTTTTGTGTAT	2
ZIKA_400_27_LEFT	AGTGCAAAGCTGAGATGGTTGG	1
ZIKA_400_27_RIGHT	ATGTGTAGAGTTGCGGGAGAGT	1
ZIKA_400_28_LEFT	GGTGGGGGATTGGCTTGAAAAA	2
ZIKA_400_28_RIGHT	GGGCCTCATAGCTTCCATGGTA	2
ZIKA_400_29_LEFT	AGGATGTGAATCTCGGCTCTGG	1
ZIKA_400_29_RIGHT	ATGCTGCATTGCTACGAACCTT	1
ZIKA_400_30_LEFT	AAAAGTGGACACTAGGGTGCCA	2
ZIKA_400_30_RIGHT	TAATCCCAGCCCTTCAACACCA	2
ZIKA_400_31_LEFT	ACAAGGGGAATTTGGAAAGGCC	1
ZIKA_400_31_RIGHT	CGTAAGTGACAACTTGTCCGCT	1
ZIKA_400_32_LEFT	AAATGGAAAAAGGGCACAGGGC	2
ZIKA_400_32_RIGHT	TGTCCCATCCAGTTGAGGGTTT	2
ZIKA_400_33_LEFT	CAAACGAATGGCAGTCAGTGGA	1
ZIKA_400_33_RIGHT	ATCCACACTCTGTTCCACACCA	1
ZIKA_400_34_LEFT	ATTTCCACAGAAGGGACCTCCG	2
ZIKA_400_34_RIGHT	TGACTAGCAGGCCTGACAACAT	2
ZIKA_400_35_LEFT	ACCACCTGGGCTGAGAACATTA	1
ZIKA_400_35_RIGHT	ACCACTAGTCCCTCTTCTGGAG	1

Supplementary Table 2 | Chikungunya virus "ChikAsiaECSA" 400 nt scheme.

Name	Sequence	Pool
CHIK_400_1_LEFT_3	CTTTTTGAAGGCCCTGCAACGT	1
CHIK_400_1_RIGHT_3	ACGGCCATCACCTCTTGTAAGT	1

CHIK_400_2_LEFT_0	TGTCGGACAGGAAGTACCACTG	2
CHIK_400_2_RIGHT_0	CTCATCCGCCCAATTTGTCGAG	2
CHIK_400_3_LEFT_0	GTACTGGGTAGGGTTCGACACA	1
CHIK_400_3_RIGHT_0	CACACGAAACCACTGTGTCACA	1
CHIK_400_4_LEFT_4	TCGGTGTTCCATCTAAAGGGCA	2
CHIK_400_4_RIGHT_4	CTTCATGGTGTTCGTGTTCCGT	2
CHIK_400_5_LEFT_0	GGTGGGGCTGAACCAGAGAATA	1
CHIK_400_5_RIGHT_0	TACAAAGCTGTCGAATTCGGCC	1
CHIK_400_6_LEFT_4	AAGCAGAAAACACACACGGTCT	2
CHIK_400_6_RIGHT_4	GCGCCTCTCGGAGTCTCTATTA	2
CHIK_400_7_LEFT_0	TACAGGCAGCACAGGAAGATGT	1
CHIK_400_7_RIGHT_0	TGGAAGTCTTCGGGCGAAATTG	1
CHIK_400_8_LEFT_1	GGAGCAAGTGAAGACGTGTACG	2
CHIK_400_8_RIGHT_1	TATATGGACAAGCGGGGCGAAT	2
CHIK_400_9_LEFT_0	TGCAAGAAGGAAGAAGCTGCAG	1
CHIK_400_9_RIGHT_0	AAGGCGATCAAGGCAAGTAACG	1
CHIK_400_10_LEFT_0	ATCTGCACGTACGGTTGATTCG	2
CHIK_400_10_RIGHT_0	GGTTTTGTTGAGCCCGTAGTGT	2
CHIK_400_11_LEFT_0	GGCAAAATGCGCACTACGAATG	1
CHIK_400_11_RIGHT_0	CCTTCCGTACGCGTTAGGAGTA	1
CHIK_400_12_LEFT_1	AGGAGTTTACGCAGTTAGGCAAA	2
CHIK_400_12_RIGHT_1	TTCATTCAGGGCTACTTCGGGT	2
CHIK_400_13_LEFT_0	TGGTCCCTATCCTCGAAACAGC	1
CHIK_400_13_RIGHT_0	TCCTGGTAGTCACGCAGATCTG	1
CHIK_400_14_LEFT_0	TGAGGCAGCGTCCATTCTAGAA	2
CHIK_400_14_RIGHT_0	ATACCTACCAAGTGTTGCCGGT	2
CHIK_400_15_LEFT_4	ACCTTGCACTGCCTACTAAGAGA	1
CHIK_400_15_RIGHT_4	ACATGGTGGTTTCAATGCTCTGG	1
CHIK_400_16_LEFT_0	TCAGAGCATACGGTTACGCAGA	2
CHIK_400_16_RIGHT_0	TTTGCGGTTCCTACTGGTGTTG	2
CHIK_400_17_LEFT_0	GACGGTGTTTGCAAGGCAGTAT	1
CHIK_400_17_RIGHT_0	TCCATGGCTGTAAAGAGGTGGT	1
CHIK_400_18_LEFT_0	ACCTCTCCTCTCCACAGGTGTA	2
CHIK_400_18_RIGHT_0	GCCTCTATTTGCTTTGGCCACA	2
CHIK_400_19_LEFT_0	AAGGGACACGTTTTCACCAGAC	1
CHIK_400_19_RIGHT_0	GGAGCATTTGACTTTTTGCACTCC	1
CHIK_400_20_LEFT_0	CGACTTCGCATGAACCATGTCA	2
CHIK_400_20_RIGHT_0	GTATGTATCGCCCCGTCGTCTA	2
CHIK_400_21_LEFT_0	AGTTTGATCTAAGCGCCGATGG	1
CHIK_400_21_RIGHT_0	AAATGTGATGGGGAACGTCTCG	1

CHIK_400_22_LEFT_0	CGCGTGACACAGCTATTTCCTT	2
CHIK_400_22_RIGHT_0	TGGCGTACCGACTTCTGTTGTA	2
CHIK_400_23_LEFT_0	ACACGGACGACGAGTTATGACT	1
CHIK_400_23_RIGHT_0	GACAATCGGACGTTGATCGGAG	1
CHIK_400_24_LEFT_0	TAATGGCAGAGACCCCGAAAGT	2
CHIK_400_24_RIGHT_0	TGTGTTCTGGAATGGGGAAGGT	2
CHIK_400_25_LEFT_0	CGAGCGACATTCAATCCGTCAA	1
CHIK_400_25_RIGHT_0	CGGCAGCAGATTATGGGTTCTT	1
CHIK_400_26_LEFT_3	GAAGAATTTGCTGCCAGCCCTA	2
CHIK_400_26_RIGHT_3	GTGCGGCTATAATGGCATCGAA	2
CHIK_400_27_LEFT_0	TTAGGAGATTGAACGCCGTCCT	1
CHIK_400_27_RIGHT_0	GTCAGACGATCTTCCAACACCC	1
CHIK_400_28_LEFT_2	GACAGGTACGCGCTTCAAGTTC	2
CHIK_400_28_RIGHT_2	CGCTAACGGTTTGCCCAGTTTA	2
CHIK_400_29_LEFT_0	AGCTCCCTACTTTTGTGGAGGG	1
CHIK_400_29_RIGHT_0	GACCGCCGTACAAGGTTATGAC	1
CHIK_400_30_LEFT_0	TAATGTCCATGGCCACCTTTGC	2
CHIK_400_30_RIGHT_0	TATTCTTCCGATTCTTGCGCGG	2
CHIK_400_31_LEFT_0	CAACTTGCCCAGCTGATCTCAG	1
CHIK_400_31_RIGHT_0	ATCGATGGTCCCCTTTACGTGT	1
CHIK_400_32_LEFT_0	TAAGGTAACAGGTTACGCGTGC	2
CHIK_400_32_RIGHT_0	TACGGGCTCCTTCATTAGCTCC	2
CHIK_400_33_LEFT_2	GACCGATCTTCGACAACAAGGG	1
CHIK_400_33_RIGHT_2	GCGTCGGGGAGAACATGTTAAG	1
CHIK_400_34_LEFT_0	GAGAAAACCTTGCGCATGCTTG	2
CHIK_400_34_RIGHT_0	CTGCTGGCATGTGATTGTCCAT	2
CHIK_400_35_LEFT_0	CCAGGTTTCCTTGCAAATCGGA	1
CHIK_400_35_RIGHT_0	GTACGTGCTGCAAGGTAGTTCC	1
CHIK_400_36_LEFT_1	CATGTACGCACCCATTTCACCA	2
CHIK_400_36_RIGHT_1	GGTCCCCGAATTCAGCATTACG	2
CHIK_400_37_LEFT_0	CTGCAAGGTCGATCAATGCCAT	1
CHIK_400_37_RIGHT_0	TCATTGTTACCCCACGTGACCT	1
CHIK_400_38_LEFT_0	AGTGGGTGACGCATAAGAAGGA	2
CHIK_400_38_RIGHT_0	TGGTGTCAGTTCGTACGGTGTA	2
CHIK_400_39_LEFT_0	ATACTCCTGTCGATGGTGGGTG	1
CHIK_400_39_RIGHT_0	CGGGATCACTGTTACGTGTTCG	1
CHIK_400_40_LEFT_0	TTTTTAGCCGTACTGAGCGTCG	2
CHIK_400_40_RIGHT_0	ATTTTCGGTGTCGCAGAAGCAG	2
CHIK_400_41_LEFT_0	GCGGTACAGCAGAGTGTAAGGA	1
CHIK_400_41_RIGHT_0	GTCGCCTTTGTACACCACGATT	1

CHIK_400_42_LEFT_0	CCATGCCGTCACAGTTAAGGAC	2
CHIK_400_42_RIGHT_0	AGTTCATCGCTCTTACCGGGTT	2
CHIK_400_43_LEFT_0	ATTGGCTAAAAGAACGAGGGGC	1
CHIK_400_43_RIGHT_0	TTCCCGAATAGTGACGGCGTTA	1
CHIK_400_44_LEFT_0	GCACCCATTCCTCAGACTTTGG	2
CHIK_400_44_RIGHT_0	CCTGCTAAACGACACGCATAGC	2