



METHOD ARTICLE

REVISED **Rapid in-country sequencing of whole virus genomes to inform rabies elimination programmes [version 2; peer review: 3 approved]**

Kirstyn Brunker ^{1,2}, Gurdeep Jaswant^{1,3}, S.M. Thumbi ³⁻⁵, Kennedy Lushasi ⁶, Ahmed Lugelo⁷, Anna M. Czupryna¹, Fred Ade⁴, Gati Wambura ⁴, Veroniah Chuchu⁴, Rachel Steenson ¹, Chanasa Ngeleja⁸, Criselda Bautista⁹, Daria L. Manalo ⁹, Ma. Ricci R. Gomez⁹, Maria Yna Joyce V. Chu ⁹, Mary Elizabeth Miranda^{9,10}, Maya Kamat ¹, Kristyna Rysava ¹¹, Jason Espineda¹², Eva Angelica V. Silo¹², Ariane Mae Aringo¹², Rona P. Bernales¹², Florencio F. Adonay¹³, Michael J. Tildesley ¹¹, Denise A. Marston¹⁴, Daisy L. Jennings¹⁴, Anthony R. Fooks ^{14,15}, Wenlong Zhu¹, Luke W. Meredith ¹⁶, Sarah C. Hill¹⁷, Radoslaw Poplawski ^{18,19}, Robert J. Gifford ²⁰, Joshua B. Singer ²⁰, Mathew Maturi²¹, Athman Mwatondo²¹, Roman Biek ^{1,2}, Katie Hampson^{1,2}

¹Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, G12 8QQ, UK

²The Boyd Orr Centre for Population and Ecosystem Health, University of Glasgow, Glasgow, G12 8QQ, UK

³University of Nairobi Institute of Tropical and Infectious Diseases (UNITID), Nairobi, Kenya

⁴Center for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

⁵Paul G. Allen School for Global Animal Health, Washington State University, Pullman, WA, USA

⁶Ifakara Health Institute, Ifakara, Tanzania

⁷Department of Veterinary Medicine and Public Health, Sokoine University of Agriculture, Morogoro, Tanzania

⁸Tanzania Veterinary Laboratory Agency, Ministry of Livestock and Fisheries Development, Dar es Salaam, Tanzania

⁹Research Institute for Tropical Medicine (RITM), Manilla, Philippines

¹⁰Field Epidemiology Training Program Alumni Foundation (FETPAFI), Manilla, Philippines

¹¹The Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research, School of Life Sciences and Mathematical Institute, University of Warwick, Coventry, UK

¹²Department of Agriculture Regional Field Office 5, Regional Animal Disease, Diagnostic Laboratory, Cabangan, Camalig, Albay, Philippines

¹³Albay Veterinary Office, Provincial Government of Albay, Albay Farmers' Bounty Village, Cabangan, Camalig, Albay, Philippines

¹⁴Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal and Plant Health Agency (APHA), Weybridge, UK

¹⁵Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

¹⁶Department of Pathology, University of Cambridge, Cambridge, UK

¹⁷University of Oxford, Oxford, UK

¹⁸Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK

¹⁹Advanced Research Computing, University of Birmingham, Birmingham, B15 2TT, UK

²⁰MRC-University of Glasgow Centre for Virus Research (CVR), University of Glasgow, Glasgow, UK

²¹Zoonotic Disease Unit, Ministry of Health, Ministry of Agriculture, Livestock and Fisheries, Nairobi, Kenya

v2

First published: 07 Jan 2020, 5:3
<https://doi.org/10.12688/wellcomeopenres.15518.1>

Latest published: 19 May 2020, 5:3
<https://doi.org/10.12688/wellcomeopenres.15518.2>

Abstract

Genomic surveillance is an important aspect of contemporary disease management but has yet to be used routinely to monitor endemic disease transmission and control in low- and middle-income countries. Rabies is an almost invariably fatal viral disease that causes a large public health and economic burden in Asia and Africa, despite being entirely vaccine preventable. With policy efforts now directed towards achieving a global goal of zero dog-mediated human rabies deaths by 2030, establishing effective surveillance tools is critical. Genomic data can provide important and unique insights into rabies spread and persistence that can direct control efforts. However, capacity for genomic research in low- and middle-income countries is held back by limited laboratory infrastructure, cost, supply chains and other logistical challenges. Here we present and validate an end-to-end workflow to facilitate affordable whole genome sequencing for rabies surveillance utilising nanopore technology. We used this workflow in Kenya, Tanzania and the Philippines to generate rabies virus genomes in two to three days, reducing costs to approximately £60 per genome. This is over half the cost of metagenomic sequencing previously conducted for Tanzanian samples, which involved exporting samples to the UK and a three- to six-month lag time. Ongoing optimization of workflows are likely to reduce these costs further. We also present tools to support routine whole genome sequencing and interpretation for genomic surveillance. Moreover, combined with training workshops to empower scientists in-country, we show that local sequencing capacity can be readily established and sustainable, negating the common misperception that cutting-edge genomic research can only be conducted in high resource laboratories. More generally, we argue that the capacity to harness genomic data is a game-changer for endemic disease surveillance and should precipitate a new wave of researchers from low- and middle-income countries.

Keywords

dog-mediated rabies, field sequencing, lyssavirus, MinION, nanopore, neglected tropical diseases, phylogenetic, rabies virus, whole genome sequencing, zoonoses, surveillance

Open Peer Review

Reviewer Status ✓✓✓

	Invited Reviewers		
	1	2	3
version 2 (revision) 19 May 2020			
version 1 07 Jan 2020	✓ report	✓ report	✓ report

- Claude Sabeta**, Agricultural Research Council-Onderstepoort Veterinary Institute, Onderstepoort, South Africa
- Herve Bourhy** , Institut Pasteur, Paris, France
- Christine Carrington** , The University of the West Indies, St. Augustine, Trinidad and Tobago

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Kirstyn Brunker (kirstyn.brunker@glasgow.ac.uk)

Author roles: **Brunker K:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Jaswant G:** Investigation; **Thumbi SM:** Funding Acquisition, Project Administration, Resources, Supervision; **Lushasi K:** Investigation; **Lugelo A:** Investigation; **Czupryna AM:** Investigation, Project Administration; **Ade F:** Resources; **Wambura G:** Resources; **Chuchu V:** Investigation; **Stenson R:** Visualization; **Ngeleja C:** Resources; **Bautista C:** Investigation; **Manalo DL:** Resources; **Gomez MRR:** Investigation; **Chu MYJV:** Investigation; **Miranda ME:** Resources; **Kamat M:** Investigation; **Rysava K:** Resources; **Espineda J:** Investigation; **Silo EAV:** Investigation; **Aringo AM:** Investigation; **Bernales RP:** Data Curation; **Adonay FF:** Data Curation; **Tildesley MJ:** Resources; **Marston DA:** Resources; **Jennings DL:** Investigation; **Fooks AR:** Resources; **Zhu W:** Investigation; **Meredith LW:** Resources; **Hill SC:** Investigation; **Poplawski R:** Resources, Software; **Gifford RJ:** Data Curation, Methodology, Resources, Software; **Singer JB:** Data Curation, Resources, Software, Writing – Review & Editing; **Maturi M:** Investigation; **Mwatondo A:** Investigation; **Biek R:** Writing – Review & Editing; **Hampson K:** Conceptualization, Data Curation, Funding Acquisition, Project Administration, Resources, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust through a Senior Research Fellowship to KH [207569], a Training Fellowship in Public Health and Tropical Medicine to SMT [110330] and Institutional Strategic Support Fund grants to the University of Glasgow [204820]. KB

was supported by a Royal Society of Tropical Medicine and Hygiene Small Grant 2017 [GR000892] and the GCRF/Scottish Funding Council through a grant awarded to the University of Glasgow [EP/S51584X/1]. KL is supported by the DELTAS Africa Initiative [Afrique One-ASPIRE/DEL-15-008]. Afrique One-ASPIRE is funded by a consortium of donors including the African Academy of Sciences (AAS) Alliance for Accelerating Excellence in Science in Africa (AESAs), the New Partnership for Africa's Development Planning and Coordinating (NEPAD) Agency, the Wellcome Trust [107753] and the UK government. The work at APHA was supported by funding from Defra, the Scottish Government and Welsh Government [SE0431].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Brunner K *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Brunner K, Jaswant G, Thumbi SM *et al.* **Rapid in-country sequencing of whole virus genomes to inform rabies elimination programmes [version 2; peer review: 3 approved]** Wellcome Open Research 2020, 5:3 <https://doi.org/10.12688/wellcomeopenres.15518.2>

First published: 07 Jan 2020, 5:3 <https://doi.org/10.12688/wellcomeopenres.15518.1>

REVISED Amendments from Version 1

Version 2 of the manuscripts has been revised based on three reviewer's comments. We also provide more detailed phylogenetic trees in [Figure 6](#) and [Figure 7](#).

Any further responses from the reviewers can be found at the end of the article

Abbreviations

RABV, rabies virus; LMICs, low- and middle-income countries; ONT, Oxford Nanopore Technology; APHA, Animal and Plant Health Agency; RITM, Research Institute for Tropical Medicine; UNITID, University of Nairobi Institute of Tropical and Infectious Diseases; WGS, whole genome sequencing; RIDT, rapid immunodiagnostic test.

Introduction

Surveillance is critical to inform infectious disease management and control programmes ([Klepac et al., 2013](#); [Molyneux et al., 2004](#)) and genomic data is emerging as a powerful new surveillance tool ([Cotton et al., 2018](#)). Pathogen sequencing has become increasingly routine in well-resourced public health laboratories, but in low- and middle-income countries (LMICs) where most emerging and endemic zoonoses occur, sequencing capacity is more limited and the direct application of genomics-informed surveillance is still at an early stage ([Balloux et al., 2018](#); [Folarin et al., 2014](#); [Gardy & Loman, 2018](#)).

Oxford Nanopore Technologies' (ONT) MinION machine has revolutionized genomic sequencing capacity ([Johnson et al., 2017](#)). This portable, low-cost sequencer has supported innovative "lab-in-a-suitcase" platforms to diagnose crop pathogens *in situ* ([Boykin et al., 2018](#); [Boykin et al., 2019](#)) and enable real-time outbreak surveillance of emerging pathogens, such as Ebola, Yellow Fever and Zika viruses ([Faria et al., 2018](#); [Faria et al., 2016](#); [Quick et al., 2015](#); [Quick et al., 2016](#); [Quick et al., 2017](#)). There is potential to extend genomic technologies to endemic pathogens such as rabies, foot-and-mouth disease, brucellosis and bovine tuberculosis, which are responsible for major impacts on people's health and livelihoods on a daily basis ([Cleaveland et al., 2017](#); [Danielsen et al., 2019](#); [Halliday et al., 2015](#)). Such an approach could reduce the lag time from sample to sequence and epidemiological action that is typical for these neglected diseases, while simultaneously building genomic capacity in LMICs. Investing in the surveillance of these endemic pathogens builds core capacities necessary for the detection of unusual disease emergence events that precede epidemics ([Cleaveland et al., 2017](#); [Halliday et al., 2017](#)).

Genomic surveillance has potential applications for management of a range of endemic zoonoses. Here we focus on the development and deployment of genomic surveillance tools to support rabies control and elimination programmes. Rabies is a devastating, invariably fatal zoonotic disease caused by a single stranded negative-sense RNA virus ([Brunker & Mollentze, 2018](#)). Rabies is globally widespread and can infect

and be transmitted by all mammal species, but the vast majority of human deaths worldwide result from bites by infected domestic dogs ([WHO, 2018a](#)). Canine rabies elimination is feasible through sustained well-implemented mass dog vaccination programmes ([Abela-Ridder et al., 2016](#); [Lankester et al., 2014](#); [WHO, 2018b](#)). A global initiative to end human deaths from dog-mediated rabies by 2030 is now underway ([WHO, 2018b](#)) and practical, effective surveillance tools are required to support this campaign ([Benavides et al., 2019](#); [Bitek et al., 2019](#); [Broban et al., 2018](#); [Hampson et al., 2016](#); [Mtema et al., 2016](#)).

Genomic surveillance can inform global efforts to eliminate dog-mediated rabies and may be particularly valuable during the endgame, when the disease can circulate for extended periods at low levels ([Klepac et al., 2013](#)). Endemic rabies typically circulates at low levels in dog populations (e.g. 58 to 384 cases per 100,000 dogs in low to high incidence settings in Tanzania, [Hampson et al., 2016](#)) and during endgame stages incidence can drop below detectable levels with traditional surveillance, often based on laboratory-based confirmation of submitted suspect cases. Enhanced surveillance including integrated bite case management can increase case detection up to ten-fold ([Suseno et al., 2019](#)) and genetic data from cases detected during this period is critical in determining their role in maintaining circulation and dispersing infection. For example: differentiating dog rabies variants from other wildlife variants ([Lembo et al., 2007](#)); distinguishing ongoing endemic transmission from incursions ([Bourhy et al., 2016](#); [Mollentze et al., 2014b](#); [Zinsstag et al., 2017](#)); and identifying sources of incursions ([Mollentze et al., 2013](#); [Tohma et al., 2016](#)). Furthermore, the capacity for an informed rapid response to newly detected cases will be pivotal for control efforts to succeed ([Henderson, 2011](#)).

Here we establish methods for rapid whole genome sequencing (WGS) of rabies viruses (RABV) using the MinION sequencer, including a workflow from sample to sequence and the associated bioinformatics pipeline. We also incorporate procedures for downstream analysis, interpretation, and management of genomic data in a surveillance context. Through testing and replication of methods across different locations in Kenya, Tanzania and the Philippines we demonstrate the feasibility of real-time sequencing of RABV to rapidly inform policy decisions and disease management. We further highlight the potential for cost savings that could make routine genomic surveillance of rabies affordable in low-resource endemic settings.

Methods

Rabies positive samples were processed under a range of conditions to generate near whole genome sequences for viral characterization and phylogenetic analysis. Here we describe the sequencing and bioinformatics pipelines and validation process, detailing adaptations required for different settings. Our aim was to establish a protocol that could be used under minimal laboratory infrastructure; therefore, the approach is supported by a lab-in-a-suitcase platform: a portable toolkit to support the steps required to get from sample to sequence ([Figure 1](#)). A comprehensive,



Figure 1. Rabies “lab-in-a-suitcase” setup for sequencing in low-resource settings. A) Lab-in-a-suitcase being used at the Nelson Mandela African Institute of Science and Technology, Tanzania; **B)** RNA extractions performed in the back of a vehicle using a portable glove box and battery-powered centrifuge in Makueni District, Kenya; **C)** Laboratory setup for sample inactivations in a district livestock office in Mugumu, Serengeti District, Tanzania.

up-to-date sample-to-sequence laboratory protocol is available on protocols.io and supporting pipelines can be found on [GitHub](https://github.com) will be updated with improvements as methods develop. The ARTIC network, a Wellcome funded project to develop the application of genomic surveillance for viral outbreak response, provides comprehensive open-source resources for laboratory and sequencing work, bioinformatics, phylogenetics and analysis. We take advantage of these resources for our rabies workflow. They can be found here on the [ARTIC network website](https://artic.network) and are published in [Quick *et al.*, 2017](#). We also utilise [RABV-GLUE](#) ([Singer *et al.*, 2018](#)), a general-purpose resource for analysis, interpretation, and management of RABV genome data, to organise and interpret RABV consensus sequences.

Samples

Brain tissue samples were collected by field officers and veterinarians from suspect rabid animals as part of rabies surveillance and research projects across Kenya ($n = 9$), Tanzania ($n = 26$, with 2 archived samples processed in the UK) and the Philippines ($n = 52$). These samples were accessible through existing collaborations and time allowed for their processing whilst optimizing our protocols and delivering training workshops, therefore numbers do not necessarily represent the surveillance capacity in each locality. Samples were obtained predominantly from domestic dogs but also included livestock and wildlife samples; details are provided as underlying data ([Brunker, 2019b](#)). Samples were processed in settings with varying infrastructure and resources, from central diagnostic laboratories to field sites. Details of each site can be found in [Table 1](#) and [Figure 2](#). Rapid immunodiagnostic tests (RIDT), manufactured by BioNote, Inc. (RG18.01, Gyeonggi-do, Republic of Korea), were used to confirm rabies positive cases in the field in Tanzania. These kits have been validated as an effective rabies field diagnostic tool ([Léchenne *et al.*, 2016](#)). In Kenya, samples were confirmed positive by the Kenya Medical

Research Institute (KEMRI) using the direct rapid immunohistochemical test (dRIT) and tested by RIDT prior to sequencing. In the Philippines samples were tested by the Research Institute for Tropical Medicine (RITM) using the fluorescent antibody test (FAT) ([Abela-Ridder, 2018](#)).

Equipment

A full suite of portable, battery-powered equipment and laboratory consumables was put together based on the ARTIC network’s published kit list. Reagents that required cold chain transportation were carried in polystyrene temperature control boxes packed with ice packs frozen at -20 or -80°C prior to use. This equipment, including reagent boxes, could be packed into standard airline sized checked luggage (lab-in-a-suitcase) for transportation ([Figure 1](#)).

Sample preparation and RNA extraction

Methods to perform sample processing and RNA extraction varied according to resources available on site. In research or diagnostic laboratories, samples were handled in containment level (CL) 2+ or CL3 conditions until sample inactivation. At field sites, samples were handled in a portable glove box (UY-33666-50, Cole Parmer). RNA extraction was performed using the Zymo Quick RNA Miniprep kit (R1054, Zymo Research) according to the manufacturer’s protocol. Sample preparation varied according to the state of the sample (fresh/frozen, homogenised/non-homogenised) and its preservation (RNAlater [Invitrogen], DNA/RNA Shield [Zymo Research], 50% glycerol-saline solution). Recommended methods for different sample types can be found in the manufacturer’s instructions but generally the following protocol was used. A small amount of brain tissue (50–100mg) was transferred from the original sample tube, using a wooden applicator stick or toothpick, to a 2ml reinforced homogeniser tube containing 1.4mm ceramic beads (Qiagen) and ~1ml of Zymo kit RNA lysis buffer. The

Table 1. Locations across Kenya, Tanzania, the Philippines and UK where rabies virus sequencing was conducted. The number of samples sequenced, number of MinION sequencing runs and resources available at each location are detailed. (*six out of the 11 samples were repeats of FIELD2 samples).

Site	Abbreviation	Location (City, Country)	# Samples sequenced (# MinION runs)	Year	Classification	Resources					
						Containment level 2	Freezer -80°C	Freezer -20°C	Laboratory	Back-up Generator	Standard lab equipment ¹
Tanzania Veterinary Laboratories Agency	TVLA	Dares Salaam, Tanzania	9 (2)	2017	Research/ diagnostic lab	✓	✓	✓	✓	✓	✓
Research Institute for Tropical Medicine	RITM	Manilla, Philippines	52 (3)	2019	Research/ diagnostic lab	✓	✓	✓	✓	✓	✓
Mugumu District field office	FIELD1	Mugumu, Tanzania	4 (2)	2017	Field site			✓			
University of Nairobi Institute of Tropical and Infectious Diseases	UNITID	Nairobi, Kenya	11* (1)	2019	Research laboratory	✓	✓	✓	✓	✓	
Makeuni Field site	FIELD2	Makueni District, Kenya	4 (1)	2019	Field site (RNA extractions only)						
Institute of Biodiversity, Animal Health & Comparative Medicine, University of Glasgow	IBAHCM	Glasgow, UK	1 (1)	2017	Research laboratory (from RNA stage)	✓	✓	✓	✓	✓	✓
Animal & Plant Health Agency	APHA	Surrey, UK	1 (1)	2016	Research/ diagnostic lab	✓	✓	✓	✓	✓	✓
Nelson Mandela Institute of Biodiversity, Animal Health & Comparative Medicine	NMAIST	Arusha, Tanzania	12 (3)	2017, 2018	Research laboratory	✓	✓	✓	✓	✓	✓

[1] Centrifuge and thermocycler available for use

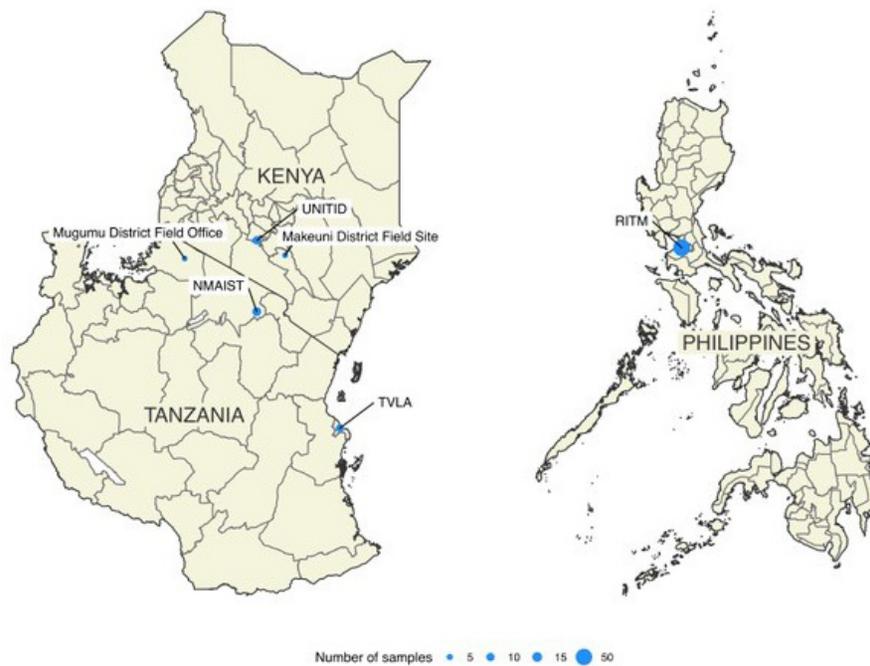


Figure 2. Locations in East Africa where sequencing of rabies viruses was performed. Circles are scaled according to numbers of samples sequenced. Details of laboratory capacity at each site are provided as underlying data (Brunker, 2019b).

sample was homogenised using a Terralyzer (Zymo Research) with 30 second pulses until fully homogenised and left to inactivate for 2 minutes. When a Terralyzer was unavailable, samples were manually homogenised using a toothpick. Lysed samples were processed according to manufacturer's column-based RNA purification protocol with in-column DNase I treatment to remove genomic DNA. In brief, lysed samples were added to filter columns and subject to a series of centrifugation steps to selectively bind RNA and purify (ethanol washes) and finally eluted. Purified viral RNA was eluted in 50µl DNase/RNase-Free Water. Aliquots of four samples stored in RNA later were shipped on dry ice from Tanzania to the Animal & Plant Health Agency (APHA, Weybridge UK) in 2017. These, and two samples used for validation (see below), were extracted according to a previously described Trizol-based method at APHA (Brunker *et al.*, 2015). Briefly, 50–100mg of brain tissue was homogenised in 1ml Trizol and 200µl chloroform added to enable phase separation. RNA separates into the upper aqueous phase, which is removed and RNA precipitated using isopropanol. This is followed by centrifugation to pellet RNA and ethanol washes to purify, with elution in DNase/RNase-Free Water.

Multiplex primer design and validation

Multiplex primer schemes were designed for RABV using [Primal Scheme](#). For Tanzania, a set of 10 reference sequences from datasets curated and filtered in RABV-GLUE were chosen to represent the known RABV diversity in Tanzania (in file order submitted to Primal Scheme: KR906748.1, KR906747.1, KY210263.1, KY210264.1, KY210227.1, KR906774.1, KR906763.1, KR906742.1, KR534218.2, KY210305.1). Settings were applied to generate primers for 400bp amplicons across

the entire genome (~11923bp), with an overlap of 50bp. A short amplicon length was chosen due to previous experience of poor sample quality from field collected and stored samples resulting in RNA fragmentation and a low viral copy number (Brunker *et al.*, 2015; Brunker *et al.*, 2018). The first scheme created, rabvTanzDg, was tested on archived Tanzanian RABV RNA samples at APHA following the protocol described in Quick *et al.* (2016). Two sequencing validations were performed: 1) Amplicons from one sample, RV3227, were sequenced on an Illumina NextSeq to confirm the amplification of individual products across the genome. This sample had already been sequenced using a PCR-free metagenomic approach, briefly described in section "Additional Sequencing" (described in Brunker *et al.*, 2015), providing a comparison between library preparations; 2) Subsequently, amplicons from another sample (RAB16025) were sequenced on both the MinION and NextSeq (fragmented before sequencing on NextSeq). In addition, metagenomic sequencing was conducted on the NextSeq for this sample. A second RABV scheme was produced for the Philippines (rabvSEasia) following the same approach and used directly for MinION sequencing (reference sequences: KX148260, KX148263, N-gene 99% consensus sequence of [AB116581, AB116582, AB683592:AB683635], AB981664, KX148255, JN786878, KX148250, KX148254, EU293111, KX148248, KX148266).

DNA preparation

Protocols and reagents varied slightly due to availability and as protocols evolved; the latest version of the protocol is available on the GitHub resource and Figshare (Brunker, 2019d). Initially, extracted RNA was reverse transcribed using

Proscript II First Strand Sequencing kit (E6560, NEB) under manufacturer's instructions. At later stages this was changed to Lunascript RT Supermix (E3010, NEB) in a reduced reaction volume of 10µl (2µl Lunascript + 3µl nuclease-free water +5µl RNA). Multiplex PCR was conducted with 2 to 2.5 µL of cDNA, the RABV primer scheme and Q5 High Fidelity Hot-Start DNA Polymerase (M0493, NEB) following the two-step PCR protocol (Quick *et al.*, 2017), with the following PCR cycling conditions: 98°C heat inactivation for 30sec, followed by 30–40 cycles of denaturation at 98°C for 15sec and a combined extension/annealing step of 65°C for 5min. PCR products were purified using a 1x solid-phase reversible immobilization bead (SPRI) cleanup (Ampure XP, Beckman-Coulter or PCRclean DX, Aline Biosciences) and concentrations measured using a Qubit dsDNA High Sensitivity kit (Q32851, ThermoFisher) on a Qubit 3.0 fluorimeter (ThermoFisher). In field sites, or when standard thermocyclers in laboratories were unavailable, PCR was performed using a portable MiniPCR machine (Amplyus). This was powered by a lithium battery (PowerAdd Pilot Pro2 85Wh LiION 9-20V) for the full PCR cycle when power was intermittent or unavailable. A negative control (water) was included at the cDNA synthesis stage for each batch of samples and as part of the sequencing library. Some of the first runs in Tanzania did not include a negative control as part of the sequencing library as the protocol was still being finalised. However, a negative control up to the PCR stage was always included as part of sample preparation. If the negative control was >1ng/µl by Qubit quantification, sample preparation was repeated from RNA extraction.

MinION library preparation and sequencing

Library preparation for the MinION was conducted using Ligation Sequencing 1D (SQK-LSK108/SQK-LSK109, ONT) and Native Barcoding kits (EXP-NBD103/EXP-NBD104/EXP-NBD114, ONT) on FLO-MIN106 (R9.4/R9.4.1, ONT) flowcells according to the manufacturer's instructions and modifications as per Quick *et al.* (2017). Briefly, DNA prepared in the previous step (PCR amplicon pools for each sample were pooled) was end-repaired and dA-tailed using an UltraII End Prep Reaction Module (E7442, NEB) followed by ligation of barcodes using the NEBNext UltraII Ligation module (E7595, NEB) in a “one-pot” reaction. Following barcode ligation, samples were purified using a SPRI bead cleanup and pooled together into a single tube before ONT adaptor ligation with the NEBNext UltraII Ligation module. Ligated DNA was then cleaned using a SPRI bead clean-up with ABB/SFB (SQK-LSK108/SQK-LSK109, ONT) washes and final library elution in ONT's elution buffer (SQK-LSK108/SQK-LSK109, ONT). With the latest kit versions, up to 24 native barcodes are available, but multiplexing varied by run according to available samples and barcodes, with multiplexed MinION runs of three to 24 samples conducted. Sequencing libraries were prepared with a negative control (non-template control from cDNA synthesis), which was sequenced as a barcoded sample. Sequencing was performed on a Dell Latitude E5470 laptop (CPU: 6th Generation Intel Core i7-6820HQ [Quad Core, 2.7GHz, 8MB cache]; Memory: 16GB RAM; Storage: 1TB solid state drive) using MinKNOW (version available at the time, v1.7.3 to 18.12.9) with live basecalling turned

off. The MinION functions as a “read-until” platform, whereby the length of sequencing runs is determined by the user when sufficient data has been produced (maximum 48h); therefore, this varied between runs. An offline version of MinKNOW provided by ONT was used due to varying internet quality at different locations. To ease the installation and use of bioinformatic pipelines in the field, a lab-on-an-SSD approach was used to ensure a working bioinformatic environment was always available. The lab-on-an-SSD is a bootable USB3 drive containing a linux installation, ONT software and bioinformatic analysis software, available from the ARTIC network resources, that offers a plug-and-play bioinformatic environment for most laptops. SSD image available as an open-source resource here: <https://github.com/artic-network/fieldbioinformatics/tree/master/lab-on-an-ssd>),

Additional sequencing

Four samples shipped to the UK and extracted at APHA were sequenced using the metagenomic approach previously described (Brunker *et al.*, 2015). Briefly, this involved host genomic DNA depletion of extracted total RNA using the on-column DNase treatment in RNeasy plus mini kit (74134, Qiagen) followed by first- and second-strand cDNA synthesis with a cDNA synthesis system kit (04379012001, Roche) and library preparation with a Nextera XT sample preparation kit (FC-131-1024, Illumina).

Cost estimates

The costs per sequenced sample were estimated for the MinION-based multiplex PCR approach and the previously used Illumina-based metagenomic approach (Brunker *et al.*, 2015; Brunker *et al.*, 2018). These estimates include processing from sample to sequence, i.e. incorporating all costs from the workflow shown in Figure 3 and any transportation costs associated with the shipment of samples to laboratory facilities. Reagents and consumable costings assume a bulk buy for at least 100 samples. For metagenomic sequencing, costs were based on previous quotes from UK sequencing facilities to process RABV samples, including sample shipment from Tanzania to the UK.

Bioinformatics and data management

i) Consensus sequence generation. Bioinformatic pipelines were run on the same laptop used for sequencing. A RABV version of the ARTIC network open-source bioinformatics pipeline (Quick *et al.*, 2017) was used to generate consensus sequences for each barcoded sample (Figure 3B). Briefly, raw files were basecalled with the latest ONT software at the time (Albacore or Guppy) and demultiplexed with Porechop using stringent settings to select reads with barcodes at both ends of the read. Demultiplexed reads were then mapped with BWA (Li & Durbin, 2009) to a reference genome chosen according to the country of origin (Tanzania/Kenya: Genbank accession no. KF155002; Philippines: Genbank accession no. KX148260). Prior to variant calling, primer sequences were trimmed using their genome coordinates and their genome coverage normalised. Variants were called using Nanopolish (Loman *et al.*, 2015), which uses MinION signal-level data to “polish” the genome

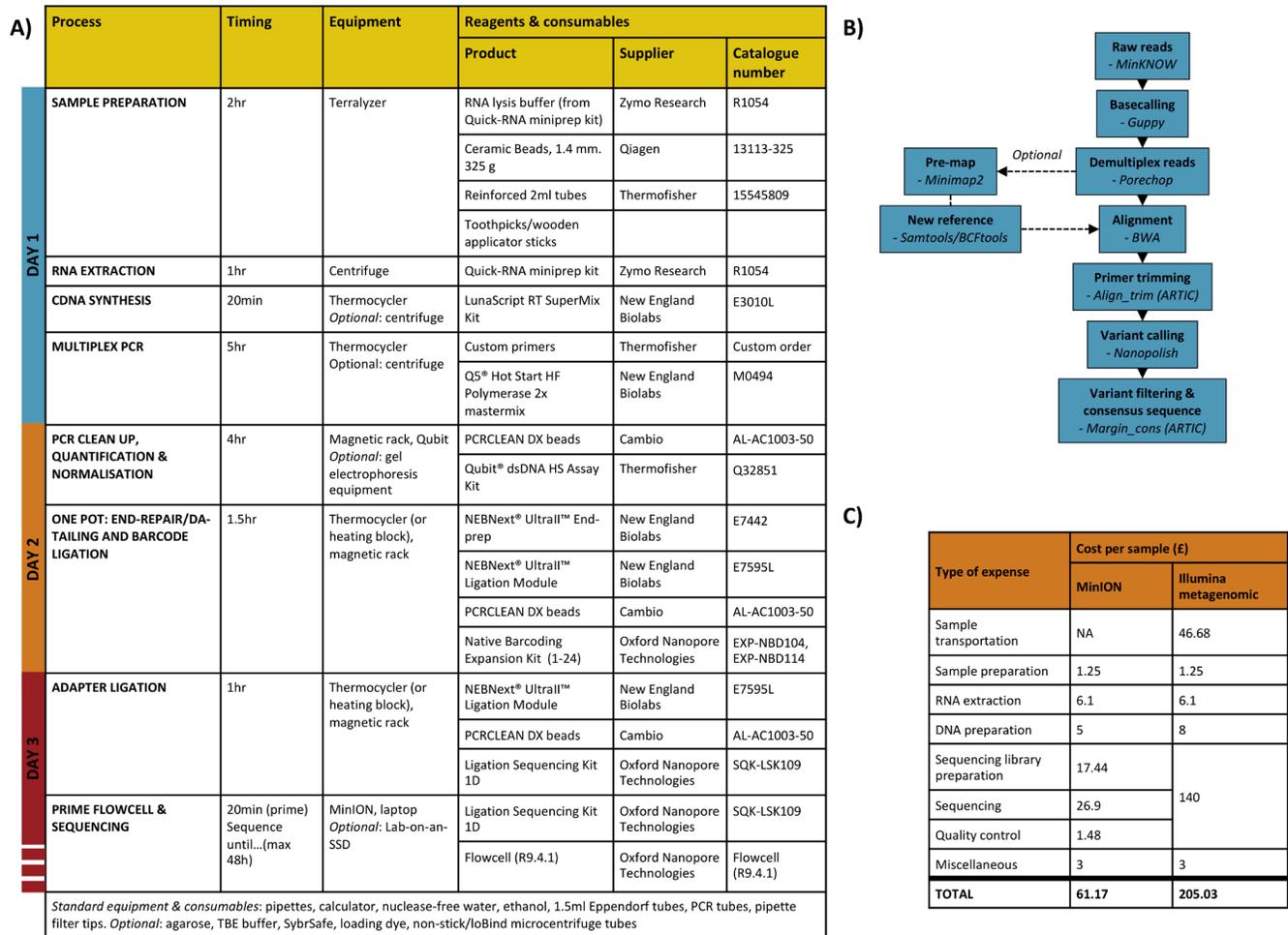


Figure 3. Workflow from sample to sequence. A) Table of laboratory workflow steps with estimated timings and details of important reagents/consumables; **B)** Bioinformatics workflow diagram; **C)** Estimated cost per sample for end-to-end rabies virus sequencing in-country using a PCR-based MinION approach versus the PCR-free metagenomic approach previously used in the UK. Costs assume a bulk buy of reagents for 100 reactions and a multiplex capacity of 24 samples per flowcell for MinION sequencing.

assembly. Consensus sequences were generated under the condition that non-overlapping primer binding sites and sites for which coverage was $<20X$ were masked by ambiguity code N. Illumina reads were processed according to a previously described method (Brunker *et al.*, 2015). Briefly, Genome Analysis Toolkit (Poplin *et al.*, 2017) was used to identify high quality SNPs using filters on strand bias (FS >60 , SOR >4), mapping quality (MQ <40 , MQRankSum ≤ 12.5), read position (ReadPosRankSum ≤ 8), and depth of coverage (DP <5). Consensus sequences were produced with a 75% majority base-call rule. Raw and processed sequence data files were stored locally on the laptop's internal hard drive or external SSDs. Data was transferred to University of Glasgow servers for storage and backup on return to the UK.

ii) Analysis/interpretation. Major and minor clade assignments were obtained for newly generated RABV sequences using Maximum Likelihood Clade Assignment (MLCA) as implemented in RABV-GLUE (Singer *et al.*, 2018). MCLA uses RAXML's Evolutionary Placement Algorithm (Stamatakis, 2014)

to select high-likelihood branch placements for a query virus sequence in a previously calculated phylogeny of virus reference strains. The reference phylogeny in RABV-GLUE was reconstructed using maximum likelihood, and contains a representative set of RABV sequences obtained from recent studies of RABV diversity (Kuzmin *et al.*, 2012; Troupin *et al.*, 2016). Following MLCA, RABV-GLUE was used to import relevant sequences from GenBank to compare to the new sequence data from each area i.e. East Africa and the Philippines. GenBank sequences were curated from the same minor clade associated with each area according to MLCA and both GenBank and MinION sequences were filtered to include only sequences with $>90\%$ genome coverage. For sequences produced for this study that was 24 out of 34 sequences from East Africa and 44 out of 52 sequences from Philippines. Alignments were produced using MAFFT (Katoh & Standley, 2013) implemented in RABV-GLUE. Maximum likelihood phylogenies were produced for each of the alignment sets using RAXML implemented in RABV-GLUE with a General Time Reversible model of nucleotide substitution with the gamma model of

rate heterogeneity and a proportion of invariant sites and 1000 bootstraps. Outgroups consisting of RABV genomes from other minor clades were included in tree building.

Training in MinION sequencing

Workshops were conducted in Kenya (11 participants) and the Philippines (10 participants) to provide hands-on training in MinION library preparation, sequencing and basic bioinformatics. Participants typically had a background in molecular biology but most had never prepared sequencing libraries. RABV RNA from local archived samples was used for sequencing following the sample-to-sequence method described above. Basic bioinformatic training included an introduction to the command line and use of the lab-on-an-SSD resource with the RABV version of the pipeline. In addition to workshops, training of local researchers was conducted on a one-to-one or small group basis at every opportunity. Consent from participants was requested as part of a questionnaire to give feedback on training workshops. Only those that consented provided feedback (75% of participants). Given that no identifying or

sensitive information was collected, no ethical approval was sought for this questionnaire.

Results

Validation of pipeline in UK laboratory

Amplicons from sample RV3227 sequenced on the NextSeq platform produced over 22million mapped reads and an average read depth of 6971 with >99% genome coverage. The same sample sequenced using the PCR-free metagenomic approach on the NextSeq produced 9498 mapped reads (0.4% of the total reads sequenced) with >99% genome coverage. Consensus sequences from each approach were identical. Amplicons from sample RAB16025 were sequenced on both NextSeq and MinION platforms, and metagenomic sequencing was also performed. Comparison between the three consensus sequences revealed 100% identity between Illumina consensus sequences (metagenomic vs PCR) and 99.98% identity between Illumina and MinION sequences. There were two SNPs between Illumina and MinION consensus sequences, highlighted in [Figure 4](#).

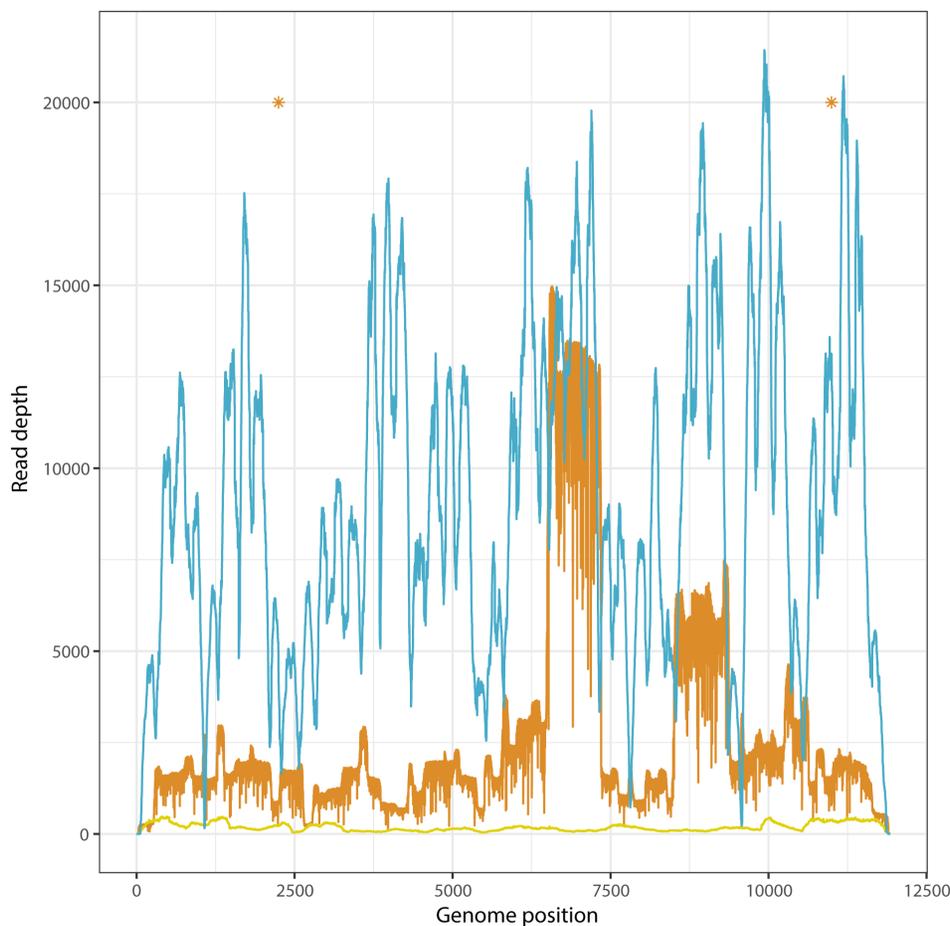


Figure 4. Genome coverage profiles for a Tanzanian dog rabies virus sample sequenced by different methods. Blue: PCR amplicons (fragmented by Nextera library preparation) sequenced on Illumina NextSeq; orange: PCR amplicons sequenced on MinION sequencer; yellow: metagenomic sequencing on Illumina NextSeq. Asterisks indicate locations of single nucleotide polymorphisms between Illumina and MinION consensus sequences.

In-country genome sequencing

RABV samples were sequenced across a range of settings in Kenya, Tanzania and the Philippines between 2017–2019 (Table 1 and Figure 2). Using a lab-in-a-suitcase approach provided the flexibility to ensure that the necessary resources were available for sequencing in resource-limited settings but also in well-resourced laboratories if particular equipment was not available. In total, 84 brain tissue samples and one salivary gland sample were processed from sample to sequence in their respective country of origin. For all samples, a consensus was generated for all sites across the genome where coverage was $\geq 20\times$ (Brunker, 2019a). The majority of samples (~80%) achieved >90% genome coverage at a depth of at least 20x (Brunker, 2019b). Due to practical constraints, sequencing run times varied; therefore, the total throughput from each run was different. Instances where coverage dropped below 20x may have been improved with longer sequencing times.

The total processing time from sample to sequence-ready library for a batch of 24 samples was ~15 hours (Figure 3A). Sequencing run length varied batch to batch and was determined by practical constraints and the expected output required for 100x read coverage for n samples in a run (and assuming loss of reads during demultiplexing). Typically, we found 5–8 hours was sufficient to get adequate coverage for multiplexed runs. For most samples, the time taken to process them end-to-end was 2–3 days when accounting for reasonable working hours (which are often restricted in laboratories) and processing a large number of samples (up to 24) for each run. The most time-consuming parts of the protocol were sample preparation for RNA extraction due to variation in the quality of brain tissue; PCR setup for a large number of samples (24 samples = 48 PCR reactions); and the normalisation of PCR products for sequencing libraries - samples with high concentrations of DNA often required an additional dilution step to provide suitable volumes (>1 μ l) for pooling after normalisation (Figure 3). Workflow times can be reduced with multiple workers (estimated timings assume only one person is doing the work) and good preparation e.g. prior tube labelling/setup, the preparation of next steps during incubations, preparation of stock solutions ahead of time etc, but careful work is required to follow workflow steps to avoid mistakes and introducing contamination.

Negative controls were included in each batch of samples processed and sequenced in most runs. Contamination was found in some batches, particularly during group training exercises (e.g. Philippines workshop batch), highlighting the importance of this quality control step. When contamination was detected during PCR quantification steps (concentration >1ng/ μ l), sample preparation was repeated from RNA extraction. However, some cross-contamination was not apparent until sequencing; therefore, a negative control should always be included in the sequencing library. A low level of cross-contamination was noted in all sequencing runs (a few hundred reads), often caused by amplicons with particularly high levels of amplification.

Costings

We estimated the full cost of sample-to-sequence preparation for RABV samples to give a better understanding of the

costs involved to process field collected samples (Brunker, 2019c). The cost reduced as improvements to the protocol and multiplexing capacity were made. Early runs, which had a low number of samples per flowcell and were used mainly for validation, were approximately £160 per sample. This reduced to ~£60 per sample at full multiplexing capacity, i.e. 24 samples per flowcell (Figure 3C). Furthermore, the development of an effective washing procedure for flowcell reuse (“Nuclease-flush”, Oxford Nanopore) means that, potentially, multiple runs per flowcell could be achieved. Therefore, we anticipate that this will reduce even further to ~£35 to £45 per sample (based on 2–6 reuses of one MinION flowcell). For example, a similar PCR-based sequencing approach for Zika virus in Brazil was able to yield ~150 samples per flowcell by making use of the nuclease-flush between runs (S. Hill, University of Oxford, personal communication). In comparison, sample shipment to the UK followed by sample-to-sequence costs for an Illumina-based metagenomic approach used previously for RABV WGS costs an estimated ~£205 per sample (Figure 3C). In terms of bioinformatic/data management costs, the SSDs used were typically purchased for £75–80 but are an optional component of the workflow and therefore not included in the overall estimate.

Training

21 participants from veterinary and public health laboratories across Tanzania, Kenya and the Philippines took part in four-day sequencing workshops held in-country. Each participant prepared one sample during group laboratory sessions, providing hands-on experience of end-to-end preparation for MinION sequencing. Participants typically had experience in molecular laboratory methods such as PCR but most had never prepared sequencing libraries before. Feedback questionnaires revealed that participants were overall very satisfied with training workshops but indicated a desire for more comprehensive bioinformatics training. At the end of each workshop, participants were asked to submit a mini research proposal and a commitment from their research institute to support MinION genomic work in order to obtain a MinION starter pack and additional reagents. In total, four MinION starter kits were provided to groups of researchers from the workshops to kick-start the application of their new training skills for genomics research. At the time of writing none of the groups had conducted sequencing mostly due to problems procuring additional reagents (either due to funding or supplier issues). However, we have offered ongoing support e.g. provision of extra reagents, linking up with local expertise to ensure projects can be completed and all are expected to sequence within the next few months.

Consensus sequence analysis

Major and minor clade assignments were obtained for all sequences using RABV-GLUE (see example in Figure 5), even for sequences with very poor consensus genome coverage of coding regions (as little as 15.1%; see underlying data; Brunker, 2019b). Philippines sequences were all identified as Asian SEA4, a minor clade so far detected exclusively in the Philippines (Saito *et al.*, 2013; Tohma *et al.*, 2016, source: RABV-GLUE). Subdivision within the Philippines subclade indicates

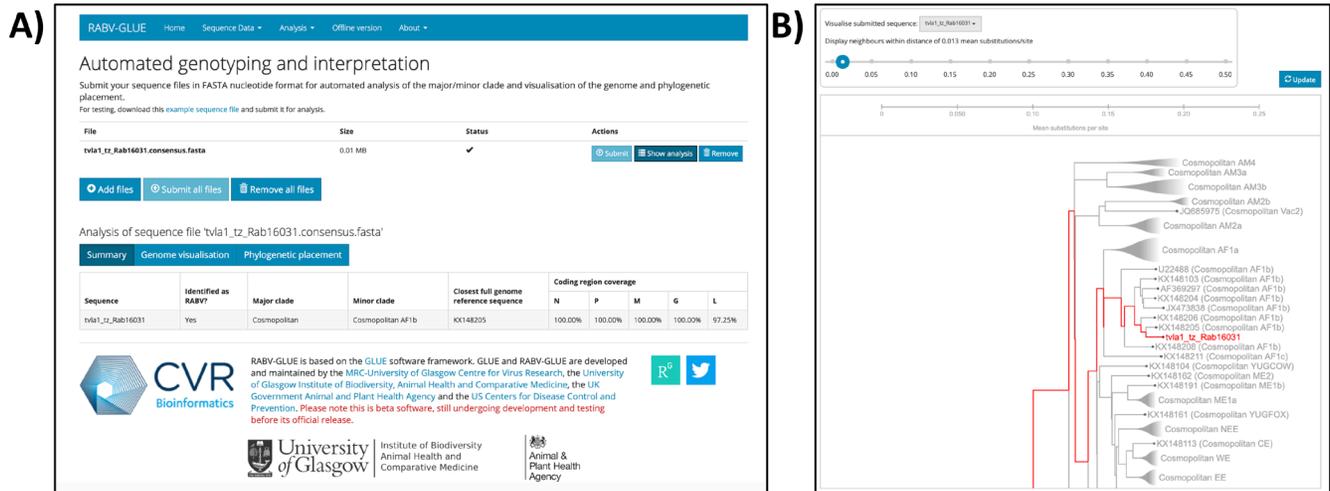


Figure 5. Example output from the online RABV-GLUE: an online tool for automated genotyping and interpretation of rabies virus (RABV) sequence data. A) Major and minor clade assignments for a Tanzanian RABV whole genome sequence (sample id=Rab16031, GenBank accession=MN726824); **B)** Phylogenetic placement of the sequence within the RABV-GLUE rabies virus reference phylogeny. This figure has been reproduced with permission from the University of Glasgow.

regional phylogeographic structure, as found previously (Saito *et al.*, 2013; Tohma *et al.*, 2016) but also further subdivision that may indicate structure at a finer spatial resolution (Figure 6). Tanzanian sequences were all from the Africa 1b lineage, as found previously (Brunker *et al.*, 2015; Brunker *et al.*, 2018). Interestingly, we detected two minor RABV subclades in Kenya; Africa 1a and Africa 1b (Figure 7). The limited RABV genetic data available for Kenya prior to our work (one WGS, nine complete/partial nucleoprotein sequences) are all Africa 1b subclade.

Discussion

Through a series of sequencing applications and trainings conducted in different settings between 2017–2019, we demonstrate the feasibility of a cost-effective, in-country pipeline for WGS of RABV using the MinION. Although challenges exist, we show they are not insurmountable. Our pipeline offers both a scalable, low cost workflow that can be particularly beneficial for sequencing a large set of archived samples and a rapid sample-to-sequence turnaround for priority samples to inform decision-making for disease management. We share our protocols and guidance to facilitate WGS of RABV, even in resource-limited settings, discuss its limitations and prospects for ongoing application.

Genomic data has become an important aspect of contemporary infectious disease surveillance (Gardy & Loman, 2018). Yet for rabies and other neglected endemic diseases it is underused - largely limited to academic studies and retrospective analyses with more limited actionable outputs. Even well-funded research projects can take up to six months from sample collection to dissemination of results (e.g. Brunker *et al.*, 2018;

Boykin *et al.*, 2019), impeding the quick decision-making necessary for timely disease control. Using our MinION-based sequencing protocol, RABV samples can be sequenced in-country in a matter of days and at a cost much lower than third-party sequencing facilities (Figure 3C), which typically requires the export of samples. We envision two broad applications of our protocols, for: 1) cost-effective batch sequencing of archived samples to characterize circulating RABV diversity, and 2) rapid *in situ* genomic surveillance of new cases, particularly during the endgame when such information can target appropriate responses. Many countries have repositories of archived tissue samples from rabid animals, which could be sequenced in bulk and therefore at low cost. These samples are potentially very valuable to characterise historical RABV diversity in a locality from which to compare to contemporary sequences. This comparison can show whether lineages continue to circulate or have disappeared, perhaps as a result of rabies control efforts. The latter option of rapid *in-situ* sequencing makes less use of cost-cutting measures such as multiplex sequencing, but the benefit of genomic surveillance to gain time-critical insights into persisting or re-emerging RABV far outweighs the expense of re-emergence. Further options to reduce costs when sample numbers are low include combining runs with other pathogens or using ONT's low-cost, lower throughput Flongle flowcells (e.g. Grädel *et al.*, 2019).

To facilitate replicable sequencing in different settings, we provide detailed instructions for each stage of sample processing. Each stage independently is relatively simple, requiring standard molecular techniques, such as PCR, or adherence to step-by-step protocols. However, as an end-to-end protocol, it is intensive and ideally undertaken by experienced laboratory

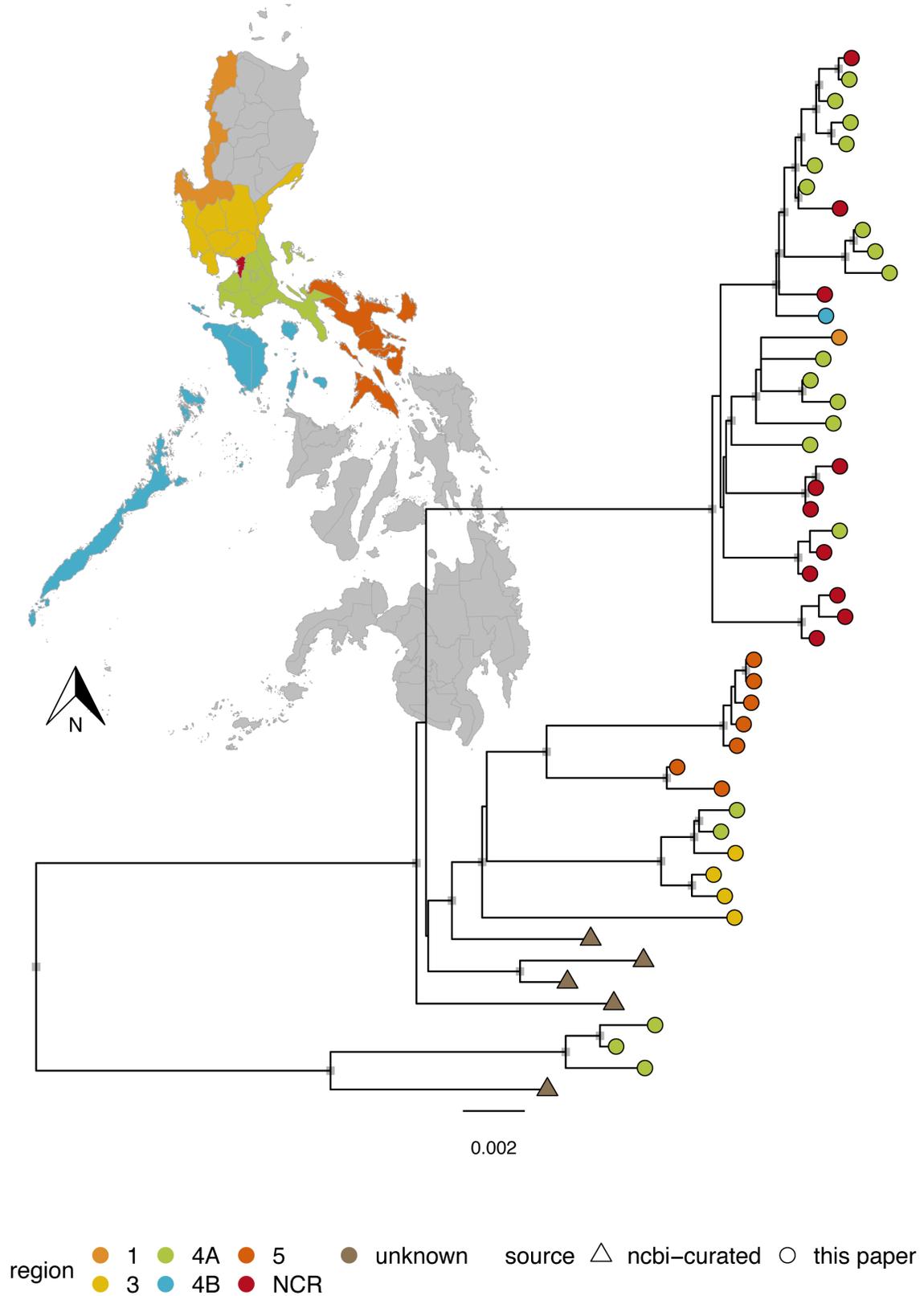


Figure 6. Maximum likelihood tree of rabies virus genomes from the Philippines. Genomes from the rabies virus Asian minor clade SEA4, which were sequenced in this study (n=44) and curated from GenBank (n=5). Colours indicate the administrative region associated with each sequence as shown on the map and internal node symbols indicate bootstrap support ≥ 80 . Genbank sequences GU358653 and GU647092 representing Asian minor clades SEA2a and SEA2b were used as an outgroup (branch not shown). Administrative shapefiles were obtained from <https://www.diva-gis.org/datadown> and plotted in R.

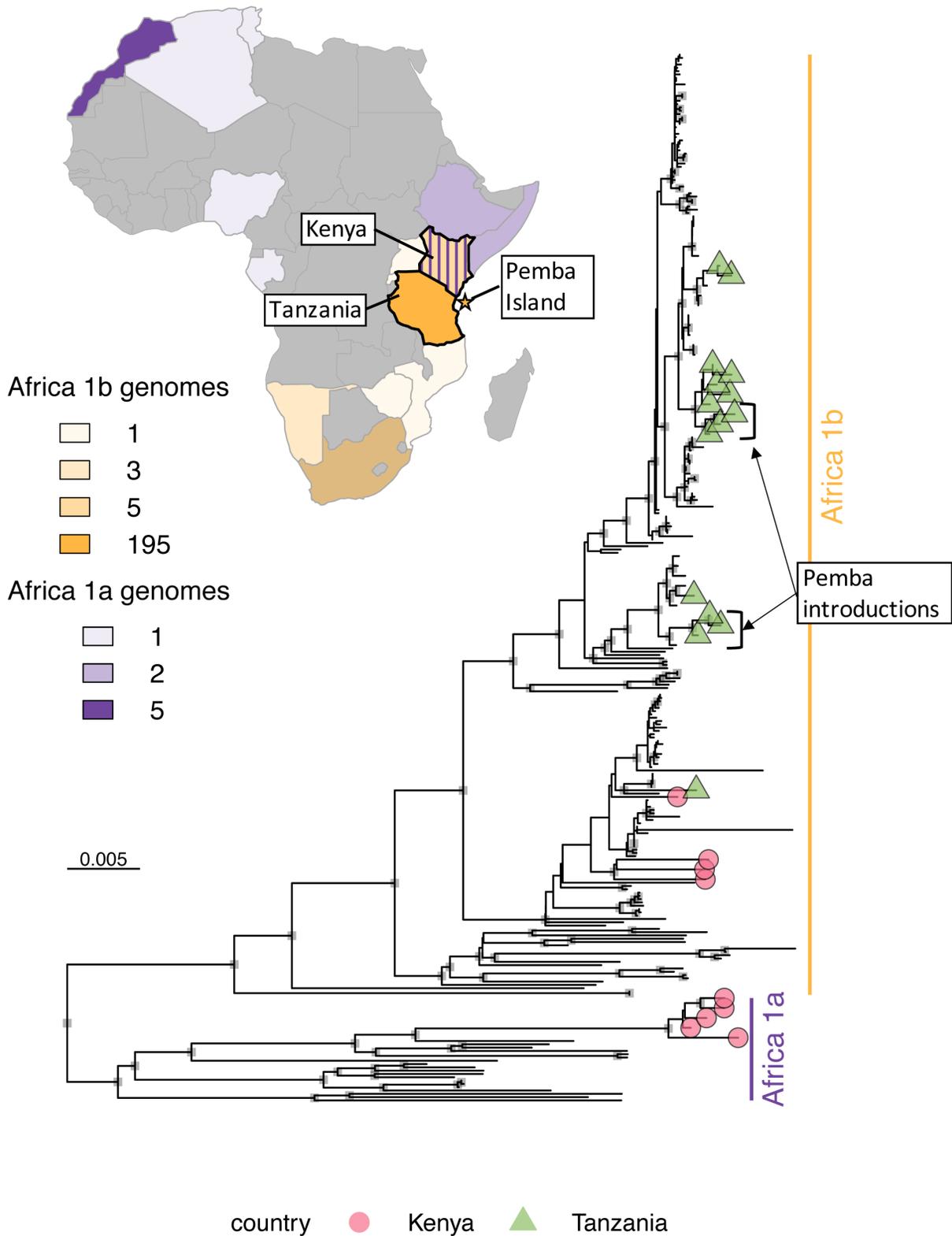


Figure 7. Maximum likelihood tree of rabies virus genomes from Kenya and Tanzania. Genomes belonging to two minor clades in the rabies virus Cosmopolitan clade are highlighted, Africa 1a and Africa 1b, which were sequenced in this study (tip shapes, n=24) and curated from GenBank (n=220). The country of origin and number of sequences in each minor clade are annotated on a map of Africa and the location of Pemba sequences from a 2016/17 outbreak is annotated on the tree, indicating multiple introductions from the mainland. Genbank sequences representing other minor clades ME1a, ME1b and ME2 (KX148162, KX148190, KX148191) were used as an outgroup (branch not shown). Maps were generated with package rWorldMaps in R.

workers. The workflow needs to accommodate constraints on working hours and fit with other workloads to enable routine implementation. Realistically, the estimated 15-hour workflow for sample-to-sequence (for a batch of 24 samples) should be conducted over a 2–3 day period, but could be performed rapidly when required. Feedback from our training workshops indicated that hands-on training was valued and provided an appreciation of what it takes to produce quality sequence data. However, while useful as an introduction to MinION sequencing, workshops should be followed by extended one-on-one training and continued use to achieve full competency in performing the workflow.

Challenges

Sub-optimal storage due to frequent power cuts and limited local storage facilities are a challenge in LMICs. Previously we have encountered difficulties in WGS of RABV samples archived in laboratories under variable conditions for up to a decade due to RNA damage (Brunker *et al.*, 2015). We found that the amplicon-based approach ensured not only efficient viral read throughput on the MinION but also that samples varying in quality could be sequenced in the same way and with the best chance of success (a worst-case scenario approach). Generating longer amplicons for RABV sequencing (Nadin-davis *et al.*, 2017) has proven difficult for samples typical of these conditions (K. Brunker, unpublished report) but may be an option for *in situ* sequencing when samples are fresh. We were unable to perform quantitative reverse-transcription PCR (qRT-PCR) on samples during field trips due to lack of equipment, but if available, this could inform the sequencing approach and aid as a diagnostic tool (Marston *et al.*, 2019). In addition, sequencing can usefully validate diagnostic results, e.g. a cat sample in the Philippines was determined positive by dRIT but shown to be negative during sequencing preparation (K. Brunker, unpublished report). Notes from the dRIT on this sample indicated it had been difficult to interpret, therefore the sequence information provided vital data to confirm the sample had been misdiagnosed as a false positive.

More generally, logistical challenges impact on infectious disease research in LMICs, from sample collection difficulties to limited internet and reliable electricity supplies and sourcing of reagents and consumables. A recent study found only two thirds of hospitals providing surgical care in 21 LMICs have a continuous electricity source or an available generator (Chawla *et al.*, 2018), problems that only worsen in more remote areas. To overcome these issues, we used a lab-in-a-suitcase approach, with portable equipment to enable quick setup in different locations using battery power if required. However, maintaining a cold chain was problematic. Access to generator backed -20°C (for reagents) and -80°C (RNA storage) freezers was limited, even in research facilities, and minor issues often taken for granted in UK labs (e.g. access to ice for benchwork with reagents) were common. ONT have recently developed a field version of some of their sequencing kits, containing lyophilised reagents to mitigate cold chain constraints. Lyophilised

reagents could be an option for future work but are expensive and would require significant bulk purchasing to be affordable. Moreover, few local suppliers can provide sequencing-specific reagents and a considerable mark-up is added to already monopolised prices in research-rich countries like the UK/USA. Ordering directly from the company comes with its own difficulties, including expensive shipment costs, import taxes and airway bills for the delivery of reagents under cold-chain conditions. Delays at customs can compromise reagent viability and occurred frequently on our sequencing trips. To promote in-country scientists and genomic research, better solutions are needed for sustaining a reliable, affordable supply chain for reagents and consumables.

Bioinformatics and data management proved to be a bottleneck in our pipeline and are a primary target for optimisation. The lab-on-an-SSD method (ARTIC network) simplified the installation of bioinformatic software and available pipelines, but as this is a developing resource, we did encounter issues that stalled analysis. This was largely due to limited experience and lack of access to a dedicated bioinformatician to troubleshoot problems. In general there is a critical need for bioinformatics training in LMICs, which has been recognised as a priority area for development, e.g. the [Pan African Bioinformatics Network](#) for the Human Heredity and Health in Africa consortium and NIH-funded Application of Genomics and Modelling to the Control of Virus Pathogens (GeMVi). Future work will focus on support for bioinformatic pipeline development, including optimising barcoding assignment and consensus building thresholds for RABV samples. This may involve the integration of Read Assignment, Mapping, and Phylogenetic Analysis in Real Time (RAMPART), a promising new tool to assess genome coverage and reference matching for each barcode concurrently as sequencing proceeds. We sequenced completely offline to avoid problems with internet connection (normally necessary to initialise MinION sequencing) and limit reliance on data transfer to servers for data storage. Sequence data was stored locally on either internal or external hard drives, with subsequent transfer to university servers in the UK at a later date. However, data transfer was problematic with the large number of files produced (in one instance corrupting a hard drive) and using one laptop for all sequencing and subsequent analysis was slow. Identifying local or cloud-based data solutions is advised to ensure data backup and take advantage of high-performance computing resources when available, but will come at additional cost.

RNA extraction from brain tissue was time-consuming and typically yields a low viral proportion after depletion of host genomic DNA (Marston *et al.*, 2013). It also remains one of the most costly aspects of the workflow at ~£6 per sample. Goharriz *et al.* (2017) have shown that complete viral genome sequence can be obtained from Whatman Flinders Technology Associates (FTA) card samples without RNA extraction, which opens avenues for refining this part of the protocol and simplifying sample field collection and storage. Efforts to

sequence from FTA card samples in Tanzania have shown promise, producing partial genome data (K. Brunker, unpublished report), and may be an option in future for field surveillance.

There is potential for cross-sample contamination, particularly when working with amplicon sequencing. This can occur both during library preparation and the bioinformatic stage of barcode demultiplexing following sequencing (Xu *et al.*, 2018). As advised in Quick *et al.* (2016), we took precautions to minimise amplicon contamination in pre-PCR areas, reagents and equipment, such as physical separation between pre- and post-PCR areas, regular decontamination and aliquoting of reagents. We also processed and sequenced negative controls (in most runs) to assess the degree of contamination, finding a low level of amplicon contamination in most sequencing runs. However, the proportion of reads in the negative control relative to the total number of reads per run was low (median [range]=0.03 [0-1.5]%) for runs prepared outside of training workshops i.e. in normal working conditions. Contamination was most problematic in regions where products overamplified in certain samples, increasing the risk of cross-over to other samples. This could be mitigated by limiting the number of PCR cycles for high viral titre samples, if they can be assessed by qPCR prior to sequencing. Unsurprisingly, contamination was higher during training exercises (see underlying data; Brunker, 2019b) and any sequencing results from training should therefore be repeated for validation.

PCR bias may preferentially amplify specific haplotypes in a sample (Grubaugh *et al.*, 2019). This has implications for the use of this type of data in intra-host diversity studies and our results indicate it can contribute to consensus sequence variation between platforms/approaches. Our comparisons between samples sequenced on different platforms (MinION vs Illumina) and with different methodology (PCR vs metagenomic) indicate that intra single nucleotide variation (iSNV) can change the nucleotides that hold the majority frequency in RABV consensus sequences. The extent to which this occurs may be influenced by the frequency of primer mismatches, depth of coverage and initial viral input (Grubaugh *et al.*, 2019). Minor consensus variation has no impact on our objective of using phylogenetic-based methods to identify variants and spatio-temporal-genetic patterns. However, it does indicate that this sequencing approach will require some adjustment to accommodate in-depth analysis, such as inferring who-infected-whom and measuring intra-host diversity. Grubaugh *et al.*'s experimental approach to "correct" such biases in amplicon-generated sequence data could be applied to limit false positive consensus assignments and accurately estimate within-host diversity in future.

The main limitation of an amplicon-based approach is the requirement for prior knowledge of the pathogen for targeted sequencing. This can be an issue if little is known about circulating RABV diversity in an area e.g. following an emergence

event, an incursion of a new variant or where multiple divergent lineages co-circulate. We have found that primer design can accommodate a reasonable level of RABV diversity, but its application is still limited. Recent discoveries of novel lyssaviruses (Shiple *et al.*, 2019), cross-species transmission events (Kuzmin *et al.*, 2012; Mollentze *et al.*, 2014a) and the emergence of new variants (Kotait *et al.*, 2019) indicate that when an amplicon-based approach is insufficient, a less targeted method would be useful to monitor threats as dog-mediated rabies is eliminated. Metagenomic sequencing is considered the gold standard for viral pathogen sequencing, offering an unbiased, untargeted approach but suffers from sensitivity issues when working with low viral titres. Success with RABV metagenomic sequencing varies according to sample quality and viral titre, with often <1% of reads mapping to a RABV reference sequence (Brunker *et al.*, 2015). Therefore, a high sequencing throughput per sample is required. The MinION is not considered a high-throughput platform but recent and ongoing improvements to flowcell chemistry and data output from the MinION (10–20Gb in 48hours) have advanced the feasibility of conducting whole genome viral metagenomic sequencing directly from clinical samples, even with low viral titre samples (Kafetzopoulou *et al.*, 2018).

Opportunities

The value of genomic data cannot be understated and has yet to reach its full potential for RABV. While the emphasis of this paper is on the application of the workflow and not phylogenetic analysis, the genomic data has already provided new insights. For example, uncovering the presence of two RABV subclades in Kenya, including one that is closely related to the lineage found in Tanzania. The extent of cross-border transmission and the scale over which persisting RABV lineages circulate in East Africa will have important implications for national and regional control efforts. In Box 1, we suggest how genetic data can inform control measures at different stages in the pathway to elimination. In addition, during the application of our workflow in Tanzania we were able to show directly how genomic data enhanced an investigation of a rabies outbreak on the island of Pemba (K. Lushasi, Ifakara Health Institute, unpublished report). Pemba is a small island (988 km²) off the coast of Tanzania. After four years of sustained vaccination campaigns, rabies appeared to have been eliminated from Pemba, with the last case detected mid-2014. However, after a lapse in vaccination efforts there was a rabies outbreak in 2016–2017. Samples collected in the early stages of the outbreak were exported to the UK as a priority for sequencing but still experienced a lag time of around three months. During our first field trip to validate in-country MinION sequencing in Tanzania (2017), we had the opportunity to sequence more samples from the outbreak. Eight samples were sequenced with a turnaround of two days, yielding whole genome consensus sequences that provided evidence of multiple incursions from mainland Tanzania (see Figure 7). This emphasises the importance of maintaining vaccination coverage and continued surveillance on Pemba to contain incursions and sustain freedom from rabies.

Box 1. Surveillance insights from genomic data

Here we highlight the current and prospective ways in which genetic data can be used to answer questions at different stages along the complex pathway towards elimination of rabies (Figure 8).

1. Characterize circulating variants and lineages:
 - a. Identify host associations/biotypes to inform targeted vaccination, such as vampire bat versus dog rabies in Latin America (Kotait *et al.*, 2019) or mongoose versus dog rabies in parts of Southern Africa (Nel *et al.*, 2005), and the identification of domestic dogs as the reservoir host in a wildlife-rich ecosystem (Lembo *et al.*, 2008)
 - b. Assess lineage distributions in the context of political/administrative borders, such as provinces in the Philippines (Tohma *et al.*, 2016), states in North Africa (Talbi *et al.*, 2010) and East Africa (Brunker *et al.*, 2015); and quantifying transboundary risks as shown by Trewby *et al.*, 2017.
 - c. Determine the socio-ecological processes underlying local and long-distance transmission, for example the identification of long-distance human-mediated movement of RABV lineages (Brunker *et al.*, 2015; Denduangboripant *et al.*, 2005; Trewby *et al.*, 2017) or associations with human density and landscape accessibility (Dellicour *et al.*, 2017; Dellicour *et al.*, 2019).
2. Monitor progress of control:
 - a. Report emergence and extinction of lineages as evidence of control failures or successes (Dudas *et al.*, 2017)
 - b. Integrate with data streams (e.g. case detection, bite patients, vaccination coverage) to resolve complexities in dynamics (Bourhy *et al.*, 2016; Lembo *et al.*, 2008; Viana *et al.*, 2016)
 - c. Identify epidemiologically connected areas that RABV lineages circulate between once control underway e.g. RABV spread associated with road networks, (Brunker *et al.*, 2018; Talbi *et al.*, 2010), spread between peri-urban and rural areas (Bourhy *et al.*, 2016).
3. Enhance surveillance during endgame:
 - a. Differentiate undetected endemic transmission versus incursions (Bourhy *et al.*, 2016; Faria *et al.*, 2017; Zinsstag *et al.*, 2017)
 - b. Discriminate variants to verify freedom from dog rabies (Velasco-Villa *et al.*, 2017) or identify emergence of new reservoirs, such as those seen in Latin America (primate/marmoset associated variants (Kotait *et al.*, 2019), circulation in crab-eating foxes (Carnieli *et al.*, 2009), terrestrial carnivores associations in Mexico (Velasco-Villa *et al.*, 2005))
 - c. Exclude possibility that cases are due to vaccine reversion, given vaccine quality concerns in some settings (Suseno *et al.*, 2019)
4. Post-elimination surveillance:
 - a. Identify the origin of imported cases (animal or human) or new outbreaks to direct control and prevention efforts and limit spread (Mollentze *et al.*, 2013; Zhang *et al.*, 2019)

WGS using portable nanopore technology is a rapidly moving field, with protocols and pipeline developments evolving quickly. Although this comes with the caveat that protocols have to evolve to keep up with changes, these developments bring cost savings and improve accuracy. New refinements are expected to reduce costs further from a current estimate of ~£60 to £35 per sample, according to projections with flowcell re-use or the smaller low cost Flongle flowcell. Moreover, ongoing optimisation to reduce reagent use (K. Brunker, unpublished report) indicate that this can be reduced even further to the point where sample preparation/RNA extraction, rather than sequencing itself, become the most costly aspect of the protocol. Improvements to pore chemistry, kits and basecalling software have significantly improved the error rate, often considered the MinION's Achilles heel. Single read accuracy has improved from 20% error rate of early phase 2014 releases to <5% error rate with R9.4.1 pore chemistry (Plesivkova *et al.*, 2019). The latest version shows further improvements with 1D² sequencing technology and error correcting algorithms reaching ~99.3-99.9% consensus accuracy (Wick *et al.*, 2019). A major benefit of nanopore sequencing is the ability to convert raw data to basecalled sequence data in real-time. Although we did not have a GPU powered laptop to take advantage of this function, it offers an even speedier resolution for time-critical sample analysis to inform decision-making. ONT's MinIT, a portable data processing unit, offers similar functionality but with the flexibility to use with different laptops.

Conclusions

Genomic research has the potential to leap-frog LMICs to the forefront of infectious diseases surveillance and control, and facing challenges head-on will hugely benefit the future of public and veterinary health. Most countries with endemic dog-mediated rabies currently have limited local sequencing capacity. These countries also suffer a heavy burden of other neglected endemic diseases, that could benefit from genomic research. The approach that we present is affordable and readily deployable to such settings with limited infrastructure, overcoming issues that typically constrain routine genomics surveillance, including cold chain, electricity and internet. Moreover, the approach is scalable and low-cost (~£60 per sample) for both sequencing large numbers of samples and for rapid turnaround to inform disease management. We predict further improvements to sequencing performance including cost-effectiveness, speed and accuracy, with further optimization underway and protocols available to researchers (https://github.com/kirstyn/rabies_minion; Brunker, 2019d; Brunker, 2019e). Nonetheless, improvement to the supply and affordability of reagents for LMICs' scientists is an area that needs addressing, as is the development of bioinformatics resources and translation of genomic research findings to digestible recommendations for policy-makers. Overall, this in-country application can support capacity building and the sustainability of routine genomic surveillance in LMICs, empowering local scientists to conduct cutting-edge research in their own backyard.

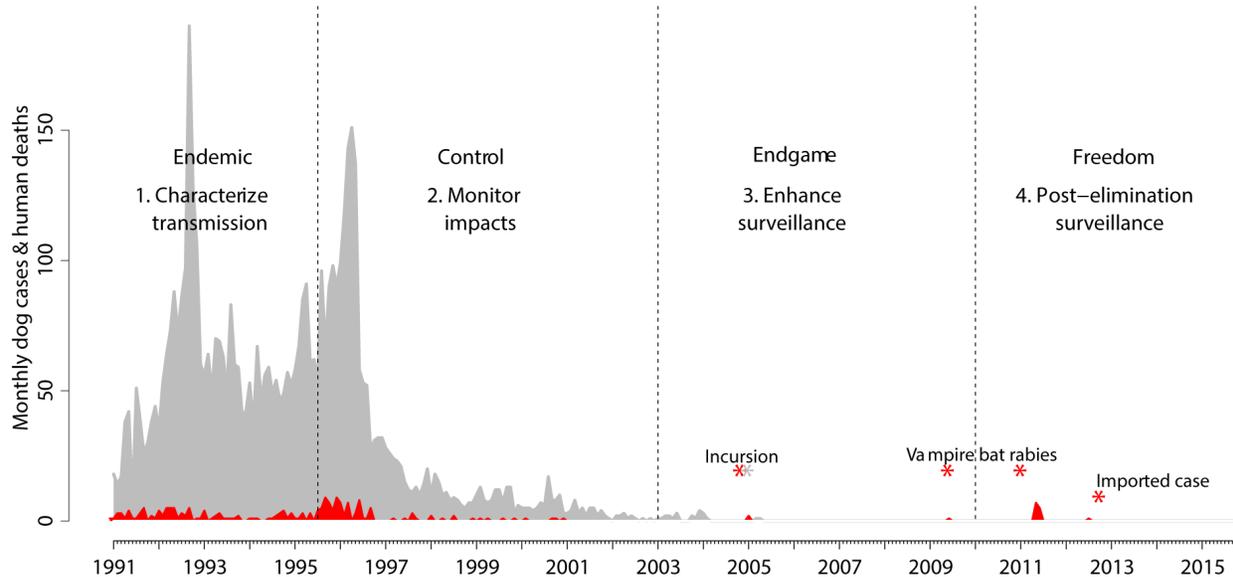


Figure 8. Progress towards elimination of dog rabies and role for genomic surveillance. Detected animal rabies cases are shown in grey and human rabies cases in red, based on surveillance data from Ecuador (from SIRVERA). Data were used to inform hypothetical scenarios relevant to rabies circulation in Latin America to illustrate examples of how genomic surveillance can add value. Stages of genomic surveillance corresponding to [Box 1](#) are shown with examples of inferences from genomic surveillance indicated by asterisks.

Data availability

Underlying data

Consensus sequences on Genbank, Accession numbers [MN726802 - MN726882](#)

Figshare: [rabv_minionSeqConsensus.fasta](#). <https://doi.org/10.6084/m9.figshare.10282859.v1> (Brunker, 2019a)

Figshare: Table 1. Rabies virus samples sequenced on the MinION machine 2016–2019 in East Africa, Philippines and UK. <https://doi.org/10.6084/m9.figshare.11111771.v1> (Brunker, 2019b)

Figshare: Breakdown of costs to sequence rabies virus samples on MinION (spreadsheet). <https://doi.org/10.6084/m9.figshare.11120405.v1> (Brunker, 2019c)

Extended data

Figshare: RabiesVirus_sampleToSequence_SOP. <https://doi.org/10.6084/m9.figshare.10742090.v1> (Brunker, 2019d)

Figshare: Rabies virus primer schemes. <https://doi.org/10.6084/m9.figshare.11112134.v1> (Brunker, 2019e)

Figshare: Extended Figure 6 phylogeny including tip labels: <https://doi.org/10.6084/m9.figshare.12284006> (Brunker, 2020a)

Figshare: Extended Figure 7 phylogeny including tip labels: <https://doi.org/10.6084/m9.figshare.12284009> (Brunker, 2020b)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Consent

Written consent to publish the images included in this paper was obtained from the individuals in these images.

Acknowledgements

We are grateful to the governments of the Philippines, Kenya and the United Republic of Tanzania and local collaborating research institutions for hosting this research. These partners include Ifakara Health Institute, Nelson Mandela African Institution of Science and Technology, the National Institute for Medical Research, the Tanzania Commission for Science and Technology, Sokoine University of Agriculture, and the Tanzanian Veterinary Laboratories Agencies in Tanzania; the University of Nairobi Institute for Tropical Infectious Diseases, the Kenya Zoonotic Disease Unit, Kenya Medical Research Institute, Central Veterinary Laboratory and the regional veterinary investigation laboratories in Kenya; and the Research Institute for Tropical Medicine and the Field Epidemiology Training Programme Alumni Foundation, Inc in the Philippines. We thank the Tanzanian Ministry of Livestock Development and Fisheries, the One Health Coordination Desk under the prime minister's office and all district authorities and veterinary workers who undertook sample collection. Thanks to all workshop participants for taking part in training. Many thanks to Beatus Lyimo for his help to facilitate lab work at NMAIST. This work benefited enormously from the support of the [ARTIC network](#) for generously donated time and expertise, in particular Josh Quick. Luke Meredith and Radoslaw Poplawski work as part of the ARTIC Network funded by Wellcome Trust Collaborative Award grant 206298/Z/17/Z. Many thanks to ONT for streamlining the setup of MinION sequencing in East Africa/Philippines and helping with logistical challenges. ONT provided some reagents for training workshops.

References

- Abela-Ridder B: **Laboratory techniques in rabies, fifth edition. Volume 2.** Geneva: World Health Organization; Licence: CC BY-NC-SA 3.0 IGO. (Rupprecht CE, Fooks AR, Abela-Ridder B: Eds.) (5th ed., Vol. 2). Geneva: Geneva: World Health Organization. 2018.
[Reference Source](#)
- Abela-Ridder B, Knopf L, Martin S, *et al.*: **2016: the beginning of the end of rabies?** *Lancet Glob Health.* 2016; **4**(11): e780–e781.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Balloux F, Brønstad Brynildsrud O, van Dorp L, *et al.*: **From Theory to Practice: Translating Whole-Genome Sequencing (WGS) into the Clinic.** *Trends Microbiol.* 2018; **26**(12): 1035–1048.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Benavides JA, Megid J, Campos A, *et al.*: **An evaluation of Brazil's surveillance and prophylaxis of canine rabies between 2008 and 2017.** *PLoS Negl Trop Dis.* 2019; **13**(8): e0007564.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bitek AO, Osoro E, Munyai PM, *et al.*: **A hundred years of rabies in Kenya and the strategy for eliminating dog-mediated rabies by 2030 [version 2; peer review: 4 approved].** *AAS Open Res.* 2019; **1**: 23.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bourhy H, Nakouné E, Hall M, *et al.*: **Revealing the Micro-scale Signature of Endemic Zoonotic Disease Transmission in an African Urban Setting.** *PLoS Pathogens.* 2016; **12**(4): e1005525.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Boykin L, Ghalab A, Rossitto De Marchi B, *et al.*: **Real time portable genome sequencing for global food security [version 1; peer review: 1 approved, 1 approved with reservations].** *F1000Res.* 2018; **7**: 1101.
[Publisher Full Text](#)
- Boykin LM, Sseruwagi P, Alicai T, *et al.*: **Tree Lab: Portable genomics for early detection of plant viruses and pests in Sub-Saharan Africa.** *BioRxiv.* 2019; 702613.
[Publisher Full Text](#)
- Broban A, Tejiokem MC, Tiembré I, *et al.*: **Bolstering human rabies surveillance in Africa is crucial to eliminating canine-mediated rabies.** *PLoS Negl Trop Dis.* 2018; **12**(9): e0006367.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brunker K: **rabv_minionSeqConsensus.fasta.** *figshare.* Dataset. 2019a.
<http://www.doi.org/10.6084/m9.figshare.10282859.v1>
- Brunker K: **Table 1. Rabies virus samples sequenced on the MinION machine 2016-2019 in East Africa, Philippines and UK.** *figshare.* Journal contribution. 2019b.
<http://www.doi.org/10.6084/m9.figshare.11111771.v1>
- Brunker K: **Breakdown of costs to sequence rabies virus samples on MinION (spreadsheet).** *figshare.* Journal contribution. 2019c.
<http://www.doi.org/10.6084/m9.figshare.1112134.v1>
- Brunker K: **Figure 6 extended.** *figshare.* Figure. 2020a.
<http://www.doi.org/10.6084/m9.figshare.12284006.v1>
- Brunker K: **Figure 7 extended.** *figshare.* Figure. 2020b.
<http://www.doi.org/10.6084/m9.figshare.12284009.v1>
- Brunker K, Lemey P, Marston DA, *et al.*: **Landscape attributes governing local transmission of an endemic zoonosis: Rabies virus in domestic dogs.** *Mol Ecol.* 2018; **27**(3): 773–788.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brunker K, Marston DA, Horton DL, *et al.*: **Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing.** *Virus Evol.* 2015; **1**(1): vev011.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brunker K, Mollentze N: **Rabies Virus.** *Trends Microbiol.* 2018; **26**(10): 886–887.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Carnieli P, Castilho JG, Fahl Wde O, *et al.*: **Molecular characterization of Rabies Virus isolates from dogs and crab-eating foxes in Northeastern Brazil.** *Virus Res.* 2009; **141**(1): 81–89.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chawla S, Kurani S, Wren SM, *et al.*: **Electricity and generator availability in LMIC hospitals: improving access to safe surgery.** *J Surg Res.* 2018; **223**: 136–141.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cleaveland S, Sharp J, Abela-Ridder B, *et al.*: **One Health contributions towards more effective and equitable approaches to health in low- and middle-income countries.** *Philos Trans R Soc B Biol Sci.* 2017; **372**(1725): pii: 20160168.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cotton JA, Berriman M, Dalén L, *et al.*: **Eradication genomics-lessons for parasite control.** *Science.* 2018; **361**(6398): 130–131.
[PubMed Abstract](#)
- Danielsen S, Kajura C, Mulema J, *et al.*: **Reaching for the low hanging fruits: One health benefits of joint crop-livestock services for small-scale farmers.** *One Health.* 2019; **7**: 100082.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dellicour S, Rose R, Faria NR, *et al.*: **Using Viral Gene Sequences to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics.** *Mol Biol Evol.* 2017; **34**(10): 2563–2571.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dellicour S, Troupin C, Jahanbakhsh F, *et al.*: **Using phylogeographic approaches to analyse the dispersal history, velocity and direction of viral lineages - Application to rabies virus spread in Iran.** *Mol Ecol.* 2019; **28**(18): 4335–4350.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Denduangboripant J, Wacharapluesadee S, Lumlerdacha B, *et al.*: **Transmission dynamics of rabies virus in Thailand: implications for disease control.** *BMC Infect Dis.* 2005; **5**: 52.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dudas G, Carvalho LM, Bedford T, *et al.*: **Virus genomes reveal factors that spread and sustained the Ebola epidemic.** *Nature.* 2017; **544**(7650): 309–315.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Faria NR, Kraemer MUG, Hill SC, *et al.*: **Genomic and epidemiological monitoring of yellow fever virus transmission potential.** *Science.* 2018; **361**(6405): 894–899.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Faria NR, Quick J, Claro IM, *et al.*: **Establishment and cryptic transmission of Zika virus in Brazil and the Americas.** *Nature.* 2017; **546**(7658): 406–410.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Faria NR, Sabino EC, Nunes MR, *et al.*: **Mobile real-time surveillance of Zika virus in Brazil.** *Genome Med.* 2016; **8**(1): 97.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Folarin OA, Happi AN, Happi CT: **Empowering African genomics for infectious disease control.** *Genome Biol.* 2014; **15**(11): 515.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gardy JL, Loman NJ: **Towards a genomics-informed, real-time, global pathogen surveillance system.** *Nat Rev Genet.* 2018; **19**(1): 9–20.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Goharriz H, Marston DA, Sharifzoda F, *et al.*: **First Complete Genomic Sequence of a Rabies Virus from the Republic of Tajikistan Obtained Directly from a Flinders Technology Associates Card.** *Genome Announc.* 2017; **5**(27): pii: e00515-17.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grädel C, Terrazos Miani MA, Barbani MT, *et al.*: **Rapid and Cost-Efficient Enterovirus Genotyping from Clinical Samples Using Flongle Flow Cells.** *Genes (Basel).* 2019; **10**(9): pii: E659.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grubaugh ND, Gangavarapu K, Quick J, *et al.*: **An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar.** *Genome Biol.* 2019; **20**(1): 8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Halliday JE, Allan KJ, Ekwem D, *et al.*: **Endemic zoonoses in the tropics: a public health problem hiding in plain sight.** *Vet Rec.* 2015; **176**(9): 220–225.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Halliday JEB, Hampson K, Hanley N, *et al.*: **Driving improvements in emerging disease surveillance through locally relevant capacity strengthening.** *Science.* 2017; **357**(6347): 146–148.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hampson K, Abela-ridder B, Brunker K, *et al.*: **Surveillance to establish elimination of transmission and freedom from Dog-mediated rabies.** *bioRxiv.* BioRxiv Preprint, 2016; 1–22.
[Publisher Full Text](#)
- Henderson DA: **The eradication of smallpox—an overview of the past, present, and future.** *Vaccine.* 2011; **29**(Suppl 4): D7–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Johnson SS, Zaikova E, Goerlitz DS, *et al.*: **Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer.** *J Biomol Tech.* 2017; **28**(1): 2–7.
[PubMed Abstract](#) | [Free Full Text](#)
- Kafetzopoulou LE, Efthymiadis K, Lewandowski K, *et al.*: **Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples.** *Euro Surveill.* 2018; **23**(50): 1–13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol Biol Evol.* 2013; **30**(4): 772–80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Klepac P, Metcalf CJ, McLean AR, *et al.*: **Towards the endgame and beyond: complexities and challenges for the elimination of infectious diseases.** *Philos Trans R Soc Lond B Biol Sci.* 2013; **368**(1623): 20120137.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Kotait I, Oliveira RN, Carrieri ML, *et al.*: **Non-human primates as a reservoir for rabies virus in Brazil.** *Zoonoses Public Health*. 2019; **66**(1): 47–59.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kuzmin IV, Shi M, Orciari LA, *et al.*: **Molecular inferences suggest multiple host shifts of rabies viruses from bats to mesocarnivores in Arizona during 2001–2009.** *PLoS Pathog*. 2012; **8**(6): e1002786.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lankester F, Hampson K, Lembo T, *et al.*: **Infectious Disease. Implementing Pasteur's vision for rabies elimination.** *Science*. 2014; **345**(6204): 1562–1564.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Léchenne M, Naïssengar K, Lepelletier A, *et al.*: **Validation of a Rapid Rabies Diagnostic Tool for Field Surveillance in Developing Countries.** *PLoS Negl Trop Dis*. 2016; **10**(10): e0005010.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lembo T, Hampson K, Haydon DT, *et al.*: **Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem.** *J Appl Ecol*. 2008; **45**(4): 1246–1257.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lembo T, Haydon DT, Velasco-Villa A, *et al.*: **Molecular epidemiology identifies only a single rabies virus variant circulating in complex carnivore communities of the Serengeti.** *Proc Biol Sci*. 2007; **274**(1622): 2123–30.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics*. 2009; **25**(14): 1754–60.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Loman NJ, Quick J, Simpson JT: **A complete bacterial genome assembled *de novo* using only nanopore sequencing data.** *Nat Methods*. 2015; **12**(8): 733–735.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Marston DA, Jennings DL, MacLaren NC, *et al.*: **Pan-lyssavirus Real Time RT-PCR for Rabies Diagnosis.** *J Vis Exp*. 2019; 149: 1–10.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Marston DA, McElhinney LM, Ellis RJ, *et al.*: **Next generation sequencing of viral RNA genomes.** *BMC Genomics*. 2013; **14**(1): 444.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mollentze N, Biek R, Streicker DG: **The role of viral evolution in rabies host shifts and emergence.** *Curr Opin Virol*. 2014a; **8**: 68–72.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mollentze N, Nel LH, Townsend S, *et al.*: **A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data.** *Proc Biol Sci*. 2014b; **281**(1782): 20133251.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mollentze N, Weyer J, Markotter W, *et al.*: **Dog rabies in southern Africa: regional surveillance and phylogeographical analyses are an important component of control and elimination strategies.** *Virus Genes*. 2013; **47**(3): 569–73.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Molyneux DH, Hopkins DR, Zagaria N: **Disease eradication, elimination and control: the need for accurate and consistent usage.** *Trends Parasitol*. 2004; **20**(8): 347–351.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mtema Z, Chungalucha J, Cleaveland S, *et al.*: **Mobile Phones As Surveillance Tools: Implementing and Evaluating a Large-Scale Intersectoral Surveillance System for Rabies in Tanzania.** *PLoS Med*. 2016; **13**(4): e1002002.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nadin-Davis SA, Colville A, Trewby H, *et al.*: **Application of high-throughput sequencing to whole rabies viral genome characterisation and its use for phylogenetic re-evaluation of a raccoon strain incursion into the province of Ontario.** *Virus Res*. 2017; **232**: 123–133.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nel LH, Sabetta CT, von Teichman B, *et al.*: **Mongoose rabies in southern Africa: a re-evaluation based on molecular epidemiology.** *Virus Res*. 2005; **109**(2): 165–73.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Plesivkova D, Richards R, Harbison S: **A review of the potential of the MinION™ single-molecule sequencing system for forensic applications.** *Wiley Interdiscip Rev: Forensic Sci*. 2019; **1**(1): e1323.
[Publisher Full Text](#)
- Poplin R, Ruano-Rubio V, DePristo MA, *et al.*: **Scaling accurate genetic variant discovery to tens of thousands of samples.** *bioRxiv*. 2017.
[Publisher Full Text](#)
- Quick J, Ashton P, Calus S, *et al.*: **Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*.** *Genome Biol*. 2015; **16**(1): 114.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quick J, Grubaugh ND, Pullan ST, *et al.*: **Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples.** *Nat Protoc*. 2017; **12**(6): 1261–1276.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quick J, Loman NJ, Duraffour S, *et al.*: **Real-time, portable genome sequencing for Ebola surveillance.** *Nature*. 2016; **530**(7589): 228–232.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Saito M, Oshitani H, Orbina JR, *et al.*: **Genetic diversity and geographic distribution of genetically distinct rabies viruses in the Philippines.** *PLoS Negl Trop Dis*. 2013; **7**(4): e2144.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shiple R, Wright E, Selden D, *et al.*: **Bats and Viruses: Emergence of Novel Lyssaviruses and Association of Bats with Viral Zoonoses in the EU.** *Trop Med Infect Dis*. 2019; **4**(1): pii: E31.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Singer JB, Thomson EC, Mclauchlan J, *et al.*: **GLUE: a flexible software system for virus sequence data.** *BMC Bioinformatics*. 2018; **19**(1): 532.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Stamatakis A: **RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.** *Bioinformatics*. 2014; **30**(9): 1312–1313.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Suseno PP, Rysava K, Brum E, *et al.*: **Lessons for rabies control and elimination programmes – a decade of One Health experience from Bali, Indonesia.** *Rev Sci Tech*. 2019; **38**(1): 213–224.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Talbi C, Lemey P, Suchard MA, *et al.*: **Phylogenetics and human-mediated dispersal of a zoonotic virus.** *PLoS Pathog*. 2010; **6**(10): e1001166.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tohma K, Saito M, Demetria CS, *et al.*: **Molecular and mathematical modeling analyses of inter-island transmission of rabies into a previously rabies-free island in the Philippines.** *Infect Genet Evol*. 2016; **38**: 22–28.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Trewby H, Nadin-Davis SA, Real LA, *et al.*: **Processes Underlying Rabies Virus Incursions across US-Canada Border as Revealed by Whole-Genome Phylogeography.** *Emerg Infect Dis*. 2017; **23**(9): 1454–1461.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Troupin C, Dacheux L, Tanguy M, *et al.*: **Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts.** *PLoS Pathog*. 2016; **12**(12): e1006041.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Velasco-Villa A, Escobar LE, Sanchez A, *et al.*: **Successful strategies implemented towards the elimination of canine rabies in the Western Hemisphere.** *Antiviral Res*. 2017; **143**: 1–12.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Velasco-Villa A, Orciari LA, Souza V, *et al.*: **Molecular epizootiology of rabies associated with terrestrial carnivores in Mexico.** *Virus Res*. 2005; **111**(1): 13–27.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Viana M, Shirima GM, John KS, *et al.*: **Integrating serological and genetic data to quantify cross-species transmission: brucellosis as a case study.** *Parasitology*. 2016; **143**(7): 821–834.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- WHO: **WHO Expert Consultation on Rabies, third report.** World Health Organization technical report series. Geneva: World Health Organization. 2018a; **1012**.
[Reference Source](#)
- WHO: **Zero by 30: The global strategic plan to end human deaths from dog-mediated rabies by 2030.** Abela-Ridder B, Editor: WHO. 2018b.
[Reference Source](#)
- Wick RR, Judd LM, Holt KE: **Performance of neural network basecalling tools for Oxford Nanopore sequencing.** *Genome Biol*. 2019; **20**(1): 129.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Xu Y, Lewandowski K, Lumley S, *et al.*: **Detection of viral pathogens with multiplex nanopore MinION sequencing: Be careful with cross-Talk.** *Front Microbiol*. 2018; **9**: 2225.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zhang S, Li S, Gu W, *et al.*: **Zoonotic Source Attribution of *Salmonella enterica* Serotype Typhimurium Using Genomic Surveillance Data, United States.** *Emerg Infect Dis*. 2019; **25**(1): 82–91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zinsstag J, Lechenne M, Laager M, *et al.*: **Vaccination of dogs in an African city interrupts rabies transmission and reduces human exposure.** *Sci Transl Med*. 2017; **9**(421): pii: eaaf6984.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 10 February 2020

<https://doi.org/10.21956/wellcomeopenres.16987.r37504>

© 2020 Carrington C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Christine Carrington 

Department of Preclinical Sciences, Faculty of Medical Sciences, The University of the West Indies, St. Augustine, Trinidad and Tobago

Brunker *et al.* have presented a very detailed and thorough description of an end-to-end workflow for generation and bioinformatic analysis of rabies virus whole genome sequences using Nanopore technology and its validation in resource-limited settings. By including in-country training and involving local scientists in their validation they demonstrated the feasibility of the approach under realistic conditions and were able to identify typical challenges. The manuscript is very well written and the level of detail provided in describing the methods (e.g. the specifications for the computer used) is appreciated. The work presents a strong case for the use of in-country sequencing to inform rabies elimination programmes. I have only a few suggestions / comments:

1. Page 11 - The authors mention that workshop participants (presumably all 21) were asked to submit research proposals and commitment from their institutions to support MinION genomic work in order to obtain MinION starter kits. They then state that four MinION starter kits were provided to groups "to kickstart the application of their new training skills for genomic research". To provide readers with a better idea of the impact of the training, it would be helpful to know how many of the participants actually submitted proposals (with institutional commitment) as well as how many "groups" received kits. Also, is there information on whether any of the proposed projects have been initiated?
2. Figure 5 - Although it is indicated in the table within Figure 5a and highlighted in the phylogeny in Figure 5b, I would suggest including the name of the Tanzanian whole genome sequence in the figure legend i.e. "...for a Tanzanian RABV whole genome sequence (tvla1_tz_Rab16031);"
3. Figure 6 - It would be useful to know the regions from which the Genbank sequences in the phylogeny originated (do they overlap with any of those sampled in the study?). I would also like to see a fully annotated tree (perhaps as extended data) that included bootstrap values, sequence IDs and the regions and years associated with all the sequences.
4. Figure 7 - As for Figure 6 a fully annotated tree (as extended data) would be a helpful addition.

5. Figure 8 - It is unclear whether the data in this figure is hypothetical or real data from a specific country? Please clarify in the Figure title or legend.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: virology, RNA viruses, vector-borne viruses, virus evolution

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 03 February 2020

<https://doi.org/10.21956/wellcomeopenres.16987.r37507>

© 2020 Bourhy H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Herve Bourhy 

Unit Lyssavirus Epidemiology and Neuropathology, WHO Collaborating Centre for Reference and Research on Rabies, Institut Pasteur, Paris, France

Brunker *et al.* present an interesting paper describing an approach that can be used under minimal laboratory infrastructure and that could reduce the lag time from sample to sequence and epidemiological action in LMICs. They propose an in-country pipeline for rapid whole genome sequencing (WGS) of rabies viruses (RABV), one of the major neglected tropical disease causing a large public health and economic burden in Asia and Africa. The approach is based on the MinION sequencing technology, including a workflow from sample to sequence and the associated bioinformatics pipeline.

The paper is well written and documented. The authors provided a lot of information and incorporate procedures for downstream analysis, interpretation, and management of genomic data in a surveillance context. The feasibility of this methods was tested in different locations in Kenya, Tanzania and the Philippines using the concept of a lab-in-a-suitcase platform:

To this aim, samples from Kenya (n = 10), Tanzania (n = 24) and the Philippines (n = 52) were sequenced using three distinct technologies and 92 of them were sequenced in LMICs using the MinION sequencing technology. The protocol included RNA extraction and PCR amplification of 400bp amplicons across the entire genome (~11923bp), with an overlap of 50bp. Therefore, at least 30 different amplicons were needed to cover the whole RABV genome. The costs per sequenced sample were estimated for the MinION-based multiplex PCR approach and compared favorably with the previously used Illumina-based metagenomic approach. Therefore, the authors claimed that they demonstrate the feasibility of real-time sequencing of RABV to rapidly inform policy decisions and disease management. I would easily agree that the first objective is reached and I really consider that as a great achievement which deserves publication. But this does not seem to be fully the case for the second one (inform policy decision and disease management). Most of the limitations of the techniques are presented in the discussion. However, some of the limitations would require more discussion and others are simply not addressed at all although some of these dramatically weaken the ability of this approach to accurately inform epidemiological surveillance and complex epidemiological questions.

There are many studies published by other authors showing that samples collected in a defined geographic area reflecting an endemic situation and during a limited period of time (a few months) differ only by a very low number of SNPs. Any errors concerning the distribution of these SNPs along the genomes and between the samples will have a dramatic effect on the phylogenetic tree reconstruction and the derived inferences.

As mentioned by the authors, the MinION is so far not considered a high-throughput platform. One of the major risks is the potential for cross-sample contamination, particularly when working with amplicon sequencing. This was noticed by the authors “A low level of cross-contamination was noted in all sequencing runs (a few hundred reads), often caused by amplicons with particularly high levels of amplification”. This can occur both during library preparation and the bioinformatic stage of barcode demultiplexing following sequencing as written by the authors but the major risk lies in the preparation of the samples and their amplification. The authors stressed the fact that they took precautions to minimise amplicon contamination in pre-PCR areas, reagents and equipment. However, all clinical biology laboratories even in the developed countries have shifted to single tube amplifications to avoid these contaminations. It appears unrealistic to think that such contaminations will not happen in a significant percentage when performed in field conditions when many samples will be received and processed every day and at least 30 different amplicons generated for each of them. This will unfortunately happen independently of all authors’ efforts to train local people that are documented in this manuscript. These cross contaminations and mislabelling of tubes will have a strong effect on the final whole genome sequences generated and consequently on some of the conclusions drawn from their phylogenetic analysis.

I fully agree with the fact that improvements to pore chemistry, kits and basecalling software have significantly improved the error rate of the MinION. However, despite these improvements every user will agree that when used in the field condition the consensus accuracy reach only 98-99% which seems to be quite low for many applications.

The authors are fair writing “Our comparisons between samples sequenced on different platforms (MinION vs Illumina) and with different methodology (PCR vs metagenomic) indicate that intra single nucleotide variation (iSNV) can change the nucleotides that hold the majority frequency in RABV consensus sequences. The extent to which this occurs may be influenced by the frequency of primer mismatches, depth of coverage and initial viral input”. This is further exemplified in Figure 4 where

locations of single nucleotide polymorphisms between Illumina and MinION consensus sequences are shown (at least two on this example). However, I disagree with the rather superficial conclusion saying “Minor consensus variation has no impact on our objective of using phylogenetic-based methods to identify variants and spatio-temporal-genetic patterns”. This has to be demonstrated and is probably not the case for many of the applications of genomic data on surveillance. I would bet that the longer internal branches of the maximum likelihood tree linking the rabies virus genomes from Kenya and Tanzania to those collected from Genbank sequences illustrate these points. Further the authors should comment on the impact of the sites masked by ambiguity code N (coverage was <20X) and their final proportion on the full genome sequences.

Two broad applications of the presented protocol are claimed by the authors such as characterization of circulating RABV diversity, and rapid in situ genomic surveillance of new cases, particularly during the endgame and I will agree that the error rate of the MinION will generally not interfere too much with such applications. However later in the discussion they presented many more applications of genetic data to inform surveillance (see Box 1, page 15: Current and prospective ways in which genetic data can be used to answer questions at different stages along the complex pathway towards elimination of rabies). Among the four main applications listed by the authors and their secondary objectives (10 are cited), at least 6 of them (listed below) will not tolerate cross contamination, ambiguities in the sequencing, and sequencing errors even in a very low percentage and therefore renders the proposed pipeline inapplicable for such studies.

- Identify host associations to inform targeted vaccination.
- Assess lineage distribution in the context of political/administrative borders, quantifying transboundary risks.
- Determine the socio-ecological processes underlying local and long-distance transmission.
- Resolve complexities in dynamics.
- Identify epidemiologically connected areas that RABV lineages circulate between once control underway.
- Discriminate variants to verify freedom from dog rabies or identify emergence of new reservoirs.

This should be made very clear in the discussion even if we may all share the author’s hope that further development in the technology will in the next years minimize the risk of sequencing errors.

In conclusion, the MinION based pipeline proposed is interesting to enhance surveillance during the endgame and the post-elimination phases, stages that are not reached so far by many LIMCs, but is not appropriate for addressing complex epidemiological questions related to RABV dynamics and drivers of persistence and diffusion in an endemic or epidemic context.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: I have co-authored at least one paper with one of the authors of this paper

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 11 May 2020

Kirstyn Brunner, University of Glasgow, Glasgow, UK

We are grateful for this positive review and critical insights to improve the manuscript. To address the reviewers concerns we have emphasized the potential that we believe real-time sequencing has to inform policy-decisions and management, whilst also highlighting the further work required to put this into practice. As suggested, we now discuss limitations cautioned by the reviewer. In response, we also provide examples of questions that WGS derived from our MinION protocol would be sufficiently resolved and accurate to address.

We agree that appropriate tube labelling is a concern but argue that this is so for all genomic studies and not specific to our study. Furthermore, we emphasise that our approach is a multiplex PCR, which involves closed tube amplification using combined pools of primers (two tubes per sample), not 30+ individual reactions, therefore minimising the risk of cross-contamination and risk of mislabelling when working with many tubes. By including a negative control (processed from cDNA stage) in sequencing runs we are able to monitor the level of cross-contamination and identify high levels (based on read coverage across the sample set and how that compares to the controls) that could impact results. While we did note some reads in negative controls the relative proportion compared to the overall mapped reads in runs was low (median 0.03% of reads in the negative, range 0-1.5%) for sequencing runs that were not part of training exercises.

Cross-contamination is a general concern for any type of sequencing and we do agree that it will be increasingly hard to control in field situations. We now emphasize that in an endgame scenario, where samples are increasingly valuable and informative and need to be processed rapidly in-situ (i.e. in the field rather than a clinical biology laboratory), sequencing would likely be conducted individually, i.e. one sample per flow cell. In such cases, cross-contamination would not be expected and higher depth of coverage would address basecall ambiguity and mitigate concerns (hopefully including tube labelling!). Nonetheless we appreciate the reviewers apprehension and have therefore reduced our emphasis on how this approach can address complex epidemiological questions.

The error rate for nanopore sequences following consensus correction is around 0.3-0.7% (for the

sequencing chemistry versions we used), where quality scores are Q30. Although higher than Illumina (~0.1%) corrected nanopore sequences should be accurate enough for many uses aside from perhaps within-host variation (although Nathan Grubaugh's recent paper suggests that further biases can be corrected to enable characterisation of within-host diversity even with amplicon data). We have also further limited consensus miscalls by masking regions with low depth of coverage. Multiple studies have now demonstrated the utility of nanopore sequences for viral surveillance as noted by the references cited- ebola (Dudas et al), yellow fever (Hill et al), zika (Faria et al), dengue (Kafetzopoulou et al) etc and how these sequences can be used for applications such as we have suggested in box 1. We disagree that the nanopore error rate will impact sequences to the extent that they cannot discriminate different lineages or host-associated variants, particularly at whole genome resolution, even with minor consensus errors. Comparisons between Illumina and nanopore generated sequences on the phylogenetic tree show that they group tightly together and therefore the same conclusion would have been reached regardless of platform. Furthermore, nanopore consensus sequencing accuracy continues to improve and has now reached Q50 with R.10.3 flowcells.

More specifically, we agree that some questions, such as identification of host adaptation (e.g. emergence in foxes from dogs in Turkey) and the mechanisms underpinning persistence in alternative hosts (see *Nardus kudu*/dogs/foxes) may require ultradeep sequencing (e.g. Marston et al). These however are questions that we are struggling to resolve no matter what sequencing approach is used, and are indicative of areas of biology for which we require more fundamental understanding. Since MinION approaches will not be key to breakthroughs in these areas, we have removed them from Box 1. Similarly, socio-ecological processes will be challenging to address, but we argue that epidemiological and contextual data will be a greater limitation than genomic data.

For those questions where we believe a MinION approach could be valuable, we have edited Box 1 to give examples:

- For host associations, such as vampire bat versus dog rabies in Latin America or mongoose (Kotait et al) vs dog rabies in parts of Southern Africa (Nel et al)
- Lineage distributions in the context of political/ administrative borders, such as provinces in the Philippines (Tohoma et al), states in North Africa (Talbi et al.) and East Africa (Tanzania and Pemba mainland); and assessing transboundary risk as shown by e.g. Trewby et al)
- Confirming epidemiologically connected areas is a clear question for areas in Latin America now, e.g. Peru/ Bolivia and a logical and feasible extension from studies by Talbi et al. (localities connected by road networks), Bourhy et al. (connectivity between peri-urban and rural areas), Bruncker et al. (connectivity between pastoralist areas)
- Identification of long-distance human-mediated movement of lineages (Thailand paper, rural to urban transmission, TZ examples).
- We argue that resolving complexities in dynamics using genomic data integrated with epidemiological data, such as distinguishing genetic lineages and their association with spatio-temporal patterns, would be possible with nanopore consensus sequences e.g. subtypes reported in Bourhy et al. (where divergence between subtypes was ~0.2-14%) should be possible with MinION data even with minor consensus inaccuracies.
- Discriminate variants to verify freedom from dog rabies or identify emergence of new reservoirs, such as those seen in Latin America (Crab-eating fox, Carnieli et al, primate/marmoset associated variants, Kotait et al). For example, Carnieli et al found a divergence of ~9% in the G gene between RABV isolates from domestic v wild hosts, variants which nanopore WGS are more than capable of discriminating.

Generally studies from other pathogens demonstrate this potential (previously mentioned above), and although yet to be achieved for rabies, provide optimism for the future.

Competing Interests: No competing interests were disclosed.

Reviewer Report 27 January 2020

<https://doi.org/10.21956/wellcomeopenres.16987.r37505>

© 2020 Sabeta C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Claude Sabeta

OIE Rabies Reference Laboratory, Agricultural Research Council-Onderstepoort Veterinary Institute, Onderstepoort, South Africa

The authors Brunker *et al.* have described and advocate a rapid in-country sequencing of whole virus genomes to inform rabies elimination programmes using Tanzania, Kenya and the Philippines as sites of testing of this technology. Whereas genomic surveillance is not being used in Africa and Asia and as we know this is due to limited infrastructure, costs, supply chains and other logistical challenges. This may change as the global scientific community has adopted the zero human deaths by 30. This may result in a shift in funding for this elimination programme as regional blocks may support this programme to be in line with global goals for the elimination of rabies. With this surveillance approach, regional members will be able to differentiate rabies variants from other wildlife, distinguish ongoing transmissions from incursions as well as identifying sources of incursions. The authors described sequencing and bioinformatics pipelines and validation process. Protocols for sample-to-sequence and a general purpose for analysis, interpretation and management of rabies virus genome data are available making the approach described here appealing to regional laboratories and members states. The concept of a portable, low-cost sequencer (lab-in-a suitcase) is interesting. Overall, the manuscript is well written and I have very few issues that require clarification:

1. Para 4 - please explain what is meant by disease can circulate for extended periods at low levels in the context of rabies.
2. In many of these countries, there are rabies virus repositories that date to many years back. What would the authors such laboratories in terms of sequencing to inform rabies elimination programmes?
3. The samples were collected by field officers, in actual fact a varied number, Kenya=10, Tanzania=24 and the Philippines=52. Do these numbers in any way indicate how good the different countries are in collecting and submitting samples? If so, how can this be improved so that whole genome sequencing of viruses truly reflects the rabies situation in the countries?
4. On page 5, the authors indicated that samples were processed in settings with varying infrastructure and resources - samples were confirmed using different methods from RIDT, to dRIT and the FAT. Can the authors shed more light on the antibody preparation used in the dRIT and

also the FAT in the Philippines? What was the correlation between the results obtained by these laboratories and the sequencing?

5. In the description of the Trizol method on page 5 - 50-100g should read 50-100mg.
6. Page 7 - please indicate the number of archived and tested samples to give the reader a sense of the sample numbers.
7. On page 7, the authors indicated that extracted RNA was reverse transcribed first with the Protoscript II kit and later Lunascript RT Supermix - was the rationale only to use reduced volumes or they were other considerations?
8. I think the technology described in this manuscript is useful for the Asia and Africa, particularly as the global community move towards rabies elimination human mediated dog rabies. Some of the outputs have been taken up by Kenya, Tanzania and the Philippines. How do the authors intend to disseminate such important technologies (particularly on minimising cross-contamination in sequencing runs, preparation of sequencing libraries and bioinformatics training) for this purpose to other regions of the two continents (if any)?
9. On page 10 - one salivary gland sample (amend).
10. Page 11 - Variations in quality of brain tissues was observed and generally influenced sample preparation times - what would the authors advise on sample collection in general to overcome this problem?
11. How do the authors plan to overcome the bottlenecks experienced with bioinformatics and data management?
12. The authors indicated that prior knowledge about circulating RABV diversity in an area is important - and we know many countries on the African and Asian continents do not have this knowledge. How do the authors hope that member states on these continents can circumvent this issue?

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epidemiology, virus evolution, virus diagnostics, method development and validation.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 11 May 2020

Kirstyn Brunker, University of Glasgow, Glasgow, UK

We thank the reviewer for their positive remarks and in our revised manuscript we address each of the minor issues as indicated below.

- 1. We now describe how endemic rabies circulates at low incidence and refer to examples from Latin America where rabies control was largely achieved early this century, but persistent foci have remained a challenge for elimination.*
- 2. Archived tissue samples from rabid animals and publicly available sequences are potentially very valuable. Archived samples could be sequenced in bulk and therefore at low cost. Together with existing sequences they can characterize historical RABV diversity in a locality from which to compare to contemporary sequences. This comparison can show whether lineages continue to circulate or have disappeared, perhaps as a result of rabies control efforts. We now explain this in the main text.*
- 3. We have clarified that the number of sequenced samples in our study do not represent the surveillance capacity of these countries. These samples were accessible through existing collaborations and time allowed for their processing whilst optimizing our protocols and delivering training workshops.*
- 4. In the Philippines, dRIT antibody was prepared at RITM using plasmid antigen from National Institute for Infectious Diseases, Tokyo Japan. Rabbit was used for hyperimmunization, purify and biontinylate (a paper has been submitted by RITM collaborators for publication of this research). FAT antibody was commercial (Fujirebio). We were able to sequence virus from all positively diagnosed samples (from all locations), aside from the one sample from a cat in the Philippines (mentioned in the text) which was diagnosed positive by dRIT but proved negative during preparation for sequencing (notes from the initial diagnosis had indicated that this was a difficult to diagnose case and it was possibly negative).*
- 5. We have corrected typographical errors (trizol units)*
- 6. We tested n=11 archived samples but are unable to reference these in the methods as we don't have unedited gel images. However, details and annotated gel images can be found here: <https://doi.org/10.6084/m9.figshare.10282865.v1>. In the text we refer only to samples taken right through to sequencing stage (n=2) prior to field trips.*
- 7. Extracted RNA first transcribed with Protoscript II kit and subsequently with Lunascript RT supermix, because of the availability of this kit, its ease of application in the lab and its reduced cost (we use reduced reaction volumes to allow 2.5x the normal recommended reactions per kit).*
- 8. We are very keen to disseminate the technology and protocols to as wide an audience as possible- note a protocol has now also been added to protocols.io and is referenced in the paper. We have therefore made our protocols openly available and will update them with any improvements as they are discovered (see protocols.io etc etc). We also emphasize*

caveats of the method (such as cross-contamination) and how they can be best avoided. We are working broadly with laboratories in East Africa (G. Kour) and in the Philippines (C. Bautista) as part of collaborative PhD studentships. We hope to conduct more training workshops, including one later this year for participants from 3 East African countries and funding permitting we will also conduct further training. For example we are currently carrying out training in Peru and through the ARTIC network have been supporting training in this technology elsewhere, including Brazil and several West African countries.

- 9. We have corrected typographical errors (salivary gland samples)*
- 10. In general for sample collection, we advise collection of brainstem with homogenisation and storage in RNA shield. This reagent has the advantage of both inactivating the sample and protecting the RNA for future use. RNA later may also be used to preserve RNA but does not fully inactivate the virus. Excellent protocols are available for this (CDC, Blueprint etc) which we refer to. Importantly once collected, samples should be frozen as soon as possible in a secure setting not subject to freeze-thaw cycles.*
- 11. Our experience has been that training in bioinformatics is lagging in some LMICs, however we are pleased to see several initiatives aiming to address capacity development in this area. We are involved in the development and testing of publicly available support tools, like lab-on-an-SSD, RAMPART and GLUE, for bioinformatics and data management, and we are consolidating our workflow using Snakemake- a workflow management system to create reproducible and scalable data analyses. This process and accompanying tools will make it easier for laboratory scientists learning these methods and we now mention these in the discussion.*

Competing Interests: No competing interests were disclosed.