COMBINATION OF RANDOM ISOTHERMAL AMPLIFICATION AND NANOPORE SEQUENCING FOR RAPID IDENTIFICATION OF THE CAUSATIVE AGENT OF AN OUTBREAK

Sören Hansen[1], Oumar Faye[2], Sabri S. Sanabani[3],Martin Faye, Ousmane Faye[2], Susanne Böhlken-Fascher[1], Amadou A. Sall[2], Manfred Weidmann[4], Claus-Peter Czerny[1] and Ahmed Abd El Wahed[1]

[1] Division of Microbiology and Animal Hygiene, University of Goettingen, Goettingen, Germany; [2]Institut Pasteur de Dakar, Dakar, Senegal; [3]Department of Pathology, School of Medicine, University of São Paulo, São Paulo, Brazil; [4]Institute of Aquaculture, University of Stirling, Stirling, Scotland, UK.

Accepted refereed manuscript of:

Hansen S, Faye O, Sanabani SS, Faye M, Böhlken-Fascher S, Faye O, Sall AA, Bekaert M, Weidmann M, Czerny C & Abd El Wahed A (2018) Combination Random Isothermal Amplification and Nanopore Sequencing For Rapid Identification of the Causative Agent of an Outbreak, *Journal of Clinical Virology*, 106, pp. 23-27.

DOI: 10.1016/j.jcv.2018.07.001

Abstract

Introduction: As an outbreak of fever of unknown origin usually starts with nonspecific symptoms and a case definition is only slowly developed and adapted, therefore, identifying the causative agent is crucial to ensure suitable treatment and/or control measures. Polymerase Chain Reaction (PCR) as a gold standard of the molecular diagnostics depends on the previous knowledge of the pathogen genome sequences. Next generation Sequencing is an alternative method, which can be applied to identify the pathogen responsible for the outbreak through sequencing all nucleic acids present in a sample extract. Sequencing data obtained can potentially identify new agents or new variants of known agents.

Aim: In this pilot study, we explored a sequencing protocol relying on multiple displacement isothermal amplification and nanopore sequencing in order to allow the identification of the causative agent in a sample. To develop the procedure for use in a suitcase laboratory, a mock sample consisting of supernatant from a Zika virus tissue culture was used.

Results: The whole procedure took around eight hours including sample preparation and data analysis using BLAST search. In total, 63,678 sequence files covering around 10,000 bases were extracted. BLAST search revealed the presence of Zika virus, which was close to an isolate from Senegal.

Conclusion: In conclusion, the protocol has potential for point of need sequencing to identify RNA viruses. The whole procedure was operated in a suitcase laboratory powered by solar power batteries. However, the procedure is cooling chain dependent and the cost per sequencing run is still very high. In addition, sequencing and data analysis pipelines for optimized and rapid subtraction of background information and assembly of relevant virus information are required.

Introduction

Identifying the causative agent implicated in an outbreak is crucial for selecting the suitable treatment and/or control measures (1). For example, around 25 pathogens can cause influenza like symptoms in the acute phase and up to 20 pathogens have to be considered for diarrhoea.

For direct detection of pathogens, polymerase chain reaction (PCR) is a widely used and well-established test for molecular diagnostics. Since specificity of PCR oligonucleotides depends on known sequences of specific target genes, false negative PCR result might be obtained due to a mismatching sequence of a novel variant of a known pathogen or because of a new emerging infectious agent. An alternative promising technology is next generation Sequencing (NGS), which can be applied to identify the pathogen responsible for the outbreak through sequencing of all nucleic acids in a sample allowing generic detection not limited by specific oligonucleotide design. Additionally, NGS data sets on detected infectious agents can be use for phylogenetic and molecular epidemiological analysis to provide insights on strain and origin of the agent. This information can be crucial for organization and distribution of resources during the outbreak control (2, 3).

There are many NGS technologies available such as sequencing by synthesis, using HiSeq and MySeq devices (Illumina, USA). These devices have a high data output, an error rate below 2% and the possibility to sequence several samples in parallel (4). Nevertheless, there is a high logistic demand through weight, size and costs of the equipment. Furthermore, cumbersome and long sample and library preparation protocols are necessary in order to generate results (4, 5). In contrast, nanopore sequencing technology (Oxford Nanopore Technology, UK) uses a pore-protein embedded in a membrane to identify individual nucleotide by the unique change in

electrical conductivity as a DNA molecule passes through the nanopore protein. Recently, Oxford Nanopore Technology developed a pocket sized (10.5 + 3.5 + 2.5 cm) sequencing device (MinION) which has the potential to be applied in the field or rural areas. A flow cell containing the required nanopores is inserted into the MinION in order to operate the sequencing run. The MinION device operates at a constant sequencing temperature (34°C) and translates the measured changes in current to a real-time nucleotide sequence *via* USB connection to a laptop (5-8).

Here we describe the establishment of a protocol for rapid identification of RNA viruses combining. random isothermal amplification and nanopore sequencing using Zika virus (ZIKV) as model virus. The protocol was performed in a mobile suitcase laboratory (figure 1) in order to allow implementation in outbreak situation (9).

Materials and Methods

Sample origin

ZIKV strains were provided by WHO collaborating Center at the Institute Pasteur of Dakar in Senegal. The monkey strain MR766 and the human strain HD78788 were isolated in 1947 (in Uganda) and 1991 (in Senegal) in Africa, respectively, during surveillance. Viral stocks were prepared by inoculating viral strains into Aedes pseudoscutellaris clone 61 (AP61) monolayer. Cells were grown in cell culture flasks (25 cm2) until they reached a confluence of approximately 80%. The medium was discarded, and 150 µl virus solution was added to the cells. The flasks were gently agitated every 15 min during incubation to enhance viral infection. After 1 h, 5 ml of Leibovitz 15 (L-15) growth medium (GibcoBRL, Grand Island, NY, USA) supplemented with 5% heat-inactivated fetal bovine serum (FBS) (GibcoBRL, Grand

Island, NY, USA), 10% Tryptose Phosphate 1% glutamine, 1% penicillin-streptomycin, 0.05% amphotericin B [Fungizone] (Sigma, Gmbh, Germany) was added and the infected cells were incubated at 28°C without CO2 until a cytopathic effect was observable.. Viral infection was confirmed by an indirect immunofluorescence assay (IFA) using specific hyper-immune mouse ascitic fluid, as described previously (Digoutte *et al.*, 1992).

Sample preparation

Zika virus (ZIKV) RNA was extracted from cell culture supernatant using the QIAamp Viral RNA Mini Kit (QIAGEN Hilden, Germany) following the manufacturer`s instructions. The RNA quantity was measured by NanoDrop ND-1000 spectrometer (Thermo Scientific, Waltham, MA, USA). For elimination of genomic DNA and reverse transcription, the QuantiTect Reverse Transcription Kit (QIAGEN Hilden, Germany) was employed using a prolonged incubation time (25 min) for the reverse transcription step. Second strand cDNA Synthesis was performed with the NEBNext mRNA Second Strand Synthesis Module (New England Biolabs, Ipswich, MA, USA). The double-stranded cDNA (ds-cDNA) was purified with the 1.8X Agencourt AMPure XP Beads Kit (Beckman Coulter, Brea, CA, USA), eluted in 55 µl nuclease-free water and quantified (NanoDrop ND-1000). To fragment and increase the amount of DNA, random amplification was done using the REPLI-g UltraFast Mini Kit (QIAGEN Hilden, Germany), Briefly, 1 µl of ds-cDNA, containing at least 10 ng, was incubated with 1 µl denaturation buffer at room temperature. To terminate the denaturation, 2 µl neutralization buffer was added after 3 min. The denatured ds-cDNA was mixed with 16 µl of the master mix containing 15 µl REPLI-g UltraFast reaction buffer and 1 µl REPLI-g UltraFast DNA polymerase and incubated at 30°C for 90 min. The reaction mix was heated to 65°C for 3 min to inactivate the reaction enzymes. Then, the DNA

was purified with the 1.8X Agencourt AMPure XP Beads Kit, eluted in 30 µl nuclease free water and quantified (NanoDrop ND-1000).

Library preparation and sequencing

For library preparation, the protocol for amplicon sequencing, SQK-NSK007, was used as recommended by Oxford Nanopore Technology. Briefly, 45 µl containing at least 1µg ds-cDNA were used for end-repairing and dA-tailing using the NEBNext Ultra II end-repair / dA-tailing module. The end-prepped DNA was purified with the 1.8X Agencourt AMPure XP Beads Kit and eluted in 31 µl nuclease free water. DNA recovery aim was at least 700 ng/µl. Adapter ligation and tethering was carried out with the NEB Blunt/TA Ligase Master Mix. The DNA was purified using the Dynabeads® MyOne™ Streptavidin C1 Kit (Thermo Fisher Scientific, Waltham, MA, USA) and solved in 25 µl of Oxford Nanopores` Elution Buffer. Six microliter of the adapted and tethered DNA was mixed with 31.5 µl nuclease free water and 37.5 µl of Oxford Nanopores` Running Buffer FM1 and then loaded into the flow cells in the MinION device.

Data processing

The MinION device generates data in fast5 format. These reads were processed with the METRICHORE AGENT (Oxford Nanopore Technology, Oxford, UK). Afterwards, the files were transformed to fastq format with PORETOOLS (10). Duplicate reads were deleted and the remaining sequences were loaded in BLAST search using GENEIOUS 9.1.6 (Biomatters Ltd., Auckland, New Zealand). Contigs were aligned to Zika strain KF383115 *via* Map to Reference option in GENEIOUS.

Results

The described procedure took around eight hours as shown in table 1. In total, 63678 sequences were extracted and transformed to fastq format. After uploading the sequences to BLAST, ZIKV sequences were identified in approximately 4% of the reads. The complete original ZIKV sequence (GenBank accession number: KF383115) was recovered with 2454 reads with an average read length of XXX (Max.: 585, Min.: 36, Std. Dev.: 122.6) (figure 2 and 3). The average coverage was x fold the minimum coverage was x-fold. Pairwise identity in BLAST analysis was 67.4%.

Additionally a total of 411 correct ZIKV reads were found in the FAST5 fail sequence file. If included in the assembly they matched correctly to the respective ZIKV sequence (figure 3).

Table 1: Sequencing Workflow

Procedure	**Reagents/Software**	**Time (min)**
RNA extraction	QIAamp Viral RNA Mini Kit	30
DNA digestion and reverse transcription	QuantiTect Reverse Transcription Kit	35
second strand cDNA synthesis	NEBNext mRNA Second Strand Synthesis Module	90
random isothermal amplification	REPLI-g UltraFast Mini Kit	120
library preparation	Nanopore sequencing kits: SQK-NSK007	70
sequencing	MinION device and R9 flow cell	20
data analysis and BLAST search	PORETOOLS and Geneious 9.1.6	120
Total		**485**

Discussion

Identifying the causative agent of an outbreak using sequencing instead of molecular techniques like PCR could have a high impact on selecting and implementing the right patient management and control measures.

The most widespread sequencing device is the MiSeq, as Illumina`s smallest device, which has still a size of 68.6 + 52.3 + 56.5 cm and a weight of approximately 57 kg. Moreover, it has a higher data output (15 Giga bases) in comparison to the MinION (10 Giga bases). Nevertheless, read length by MiSeq is limited to around 300 bp and a maximum of 22-25 million reads can be produced in a run time between 4h and 56h (11, 12). In contrast, the MinION has through its nanopore technology no limit in read length and number. Moreover, reads are generated in 20-120 min and data are easily accessible on laptop or PC.

We have discovered that 1/5 of the correct ZIKV reads was placed into the “fail” file. The METRICHORE AGENT classifies the reads into pass and fail reads by neuronal network computing assessing definite conductivity readout events at the pore exit for 5-6 mers. This complicated sequence definition needs quality scoring to decide on the statistical trustworthiness of the sequencing result. Fails are defined through the following approach. Initially base calling (1D base calling) of template and complement reads is performed separately. If the resulting sequence length ratio is between 0.5-2.0, all sequences are stacked together for base 2D base calling. If resultant 2D sequences are assessed with a Q-score > 9 they are sorted into a FAST5 fail sequence file (13).

Short Illumina device reads have a 0.1% non-random error rate, which means an error at one site can still dominate the base calling process. The MinION reads have a 10% error rate but sites are distributed at random throughout the sequence which

is compensated for by base calling and which therefore do not dominate at one site reducing the overall error rate in comparison to Illumina reads (14). Our results suggest that the analysis algorithm and the Q-score need to be optimised for viral RNA sequences. At this current development stage therefore a recommended assembly approach would be first to use all pass reads to identify the infectious agent. To improve the result, the fail reads can be included in a 2nd step.

In general however the passed sequence assembly result already produce a robust result with a average coverage of 40.

The MinION was successfully used in the Ebola virus outbreak in Guinea (7) and during the Zika virus outbreak in Brazil (15). In both cases, specific PCR fragment sequencing strategies were used. RT-PCR assays were applied to reverse transcribe RNA and create multiple fragments to increase the sequencing efficacy (7, 16). This strategy limits sequencing output to targeted agents, which is ideal for molecular epidemiological analysis. The use of PCR leads to logistic issues due to heavy devices and requirement of a cold chain for the reagents. In Brazil, this was solved by transporting the whole laboratory in a caravan. The generic sequencing approach described here is intended for diagnostic identification of unknown infectious agents. It uses only random isothermal steps throughout the procedure and PCR cycling is not required which avoids the use of a thermal cycler.

We have already shown that isothermal amplification can be easily implemented in a mobile suitcase laboratory (9, 17, 18) and we successfully adapted this concept for the workflow needed for library preparation for the MinION sequencing procedure (figure 1). The suitcase, contains all materials and reagents needed for sequencing in one box of 56.0 + 45.5 + 26.5 cm in size and less than 23 kg in weight.

All steps of data collection and analysis except the BLAST search were performed offline using MINKNOW and METRICHORE AGENT as well as GENEIOUS. This is a major improvement since during the Ebola outbreak base calling for MinION datasets was only possible through cloud computing which needed internet capacity often not available locally (7). The simple structure and clear layout of these analysis programmes makes it easy for users without bioinformatic background to obtain basic information about origin and phylogeny of the sequenced target. Therefore, a bioinformatician is not necessarily needed for analysis of the datasets obtained. To perform BLAST offline a database of infectious agent sequences only located on the laptop needs to be assembled. It could be replenished with new entries to GenBank whenever online.

Currently, the following challenges have to be solved. In our hands, the sequencing reagents can be kept at 25°C for one day without any changes in their efficacy (confirmed by Oxford Nanopore Technologies, UK). However for long-term storage a -20°C freezer is still required. Moreover, the price per sequencing run is very high (around $1500), as one flow cell costs between $500 and $900 depending on the amount of ordered flow cells. In addition, the shelf life of the flow cells is around 8 weeks at 4°C. One of the biggest drawbacks is that the manufacturer is progressively changing the reagents and flow cells so that it is difficult to match biochemistry to flow cells.

The goal of this pilot study was to establish a protocol for pathogen identification during an outbreak field investigation. In principle this seems possible in a suitcase laboratory setup. The next steps will be to assemble an offline solution to compare identified sequences with preloaded database and to identify cold chain independent reagents.

References

1. **Ni PX, Ding X, Zhang YX, Yao X, Sun RX, Wang P, Gong YP, Zhou JL, Li DF, Wu HL, Yi X, Yang L, Long Y.** 2015. Rapid detection and identification of infectious pathogens based on high-throughput sequencing. Chin Med J (Engl) **128:**877-883.
2. **Robinson ER, Walker TM, Pallen MJ.** 2013. Genomics and outbreak investigation: from sequence to consequence. Genome Med **5:**36.
3. **Croucher NJ, Didelot X.** 2015. The application of genomics to tracing bacterial pathogen transmission. Curr Opin Microbiol **23:**62-67.
4. **Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M.** 2012. Comparison of next-generation sequencing systems. J Biomed Biotechnol **2012:**251364.
5. **Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD, Martellaro C, Falzarano D, Marzi A, Squires RB, Wollenberg KR, de Wit E, Prescott J, Safronetz D, van Doremalen N, Bushmaker T, Feldmann F, McNally K, Bolay FK, Fields B, Sealy T, Rayfield M, Nichol ST, Zoon KC, Massaquoi M, Munster VJ, Feldmann H.** 2016. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. Emerg Infect Dis **22:**331-334.
6. **Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M.** 2012. Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. Nat Biotechnol **30:**344-348.
7. **Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouedraogo N, Afrough B, Bah A, Baum JH, Becker-Ziaja B, Boettcher JP, Cabeza-Cabrerizo M, Camino-Sanchez A, Carter LL, Doerrbecker J, Enkirch T, Garcia-Dorival I, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D, Nitzsche K, Pallasch E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan K, Vitoriano I, Yemanaberhan RL, Zekeng EG, Racine T, Bello A, Sall AA, Faye O, et al.** 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature **530:**228-232.
8. **Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ.** 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif **3:**1-8.
9. **Abd El Wahed A, Weidmann M, Hufert FT.** 2015. Diagnostics-in-a-Suitcase: Development of a portable and rapid assay for the detection of the emerging avian influenza A (H7N9) virus. J Clin Virol **69:**16-21.
10. **Loman NJ, Quinlan AR.** 2014. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics **30:**3399-3401.
11. **Urban JM, Bliss, Jacob, Lawrence, Charles E, Gerbi, Susan A** 2015. Sequencing ultra-long DNA molecules with the Oxford Nanopore MinION. BioRxiv doi:10.1101/019281.
12. **Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J, Harmsen D.** 2013.

Updating benchtop sequencing performance comparison. Nat Biotechnol **31:**294-296.

13. **Lu H, Giordano F, Ning Z.** 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics Proteomics Bioinformatics **14:**265-279.
14. **Goodwin S, McPherson JD, McCombie WR.** 2016. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet **17:**333-351.
15. **Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC, Jr., Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ.** 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc **12:**1261-1276.
16. **Faria NR, Sabino EC, Nunes MR, Alcantara LC, Loman NJ, Pybus OG.** 2016. Mobile real-time surveillance of Zika virus in Brazil. Genome Med **8:**97.
17. **Faye O, Faye O, Soropogui B, Patel P, El Wahed AA, Loucoubar C, Fall G, Kiory D, Magassouba N, Keita S, Konde MK, Diallo AA, Koivogui L, Karlberg H, Mirazimi A, Nentwich O, Piepenburg O, Niedrig M, Weidmann M, Sall AA.** 2015. Development and deployment of a rapid recombinase polymerase amplification Ebola virus detection assay in Guinea in 2015. Euro Surveill **20**.
18. **Mondal D, Ghosh P, Khan MA, Hossain F, Bohlken-Fascher S, Matlashewski G, Kroeger A, Olliaro P, Abd El Wahed A.** 2016. Mobile suitcase laboratory for rapid detection of Leishmania donovani using recombinase polymerase amplification assay. Parasit Vectors **9:**281.

1. Digoutte JP, Calvo-Wilson MA, Mondo M, Traore-Lamizana M, Adam F. **Continuous cell lines and immune ascitic fluid pools in arbovirus detection.** *Res Virol.* 1992, vol. 143: 417–422.