ISBN: 978-623-92207-1-6

Metagenomic Sequencing of P30-gene-positive Using MinION Sequencer for Apicomplexan Species Identification

1st Dela Ria Nesti
Departement of BioresourcesTechnology and Veterinary, Vocational College, Universitas Gadjah Mada
Yogyakarta, Indonesia
dela.riaresti@ugm.ac.id

2nd Ahmad Baidlowi
Departement of BioresourcesTechnology and Veterinary, Vocational College, Universitas Gadjah Mada
Yogyakarta, Indonesia
ahmad.baidlowi@ugm.ac.id

3rd Lucky Ronald Runtuwene
AIDS Research Center National Institute of Infectious Dieseases
Tokyo, Japan
luckyron@niid.go.jp

4th Ari Indrawati
Departement of BioresourcesTechnology and Veterinary, Vocational College, Universitas Gadjah Mada
Yogyakarta, Indonesia
ari.indrawati@mail.ugm.ac.id

5th Wayan Tunas Artama
Departement of Biochemistr, Faculty of Veterinary Medicine, and OHCC Universitas Gadjah Mada
Yogyakarta, Indonesia
artama@ugm.ac.id

6th Yutaka Suzuki
Departement of Computational Biology and Medical Science
Tokyo, Japan
ysuzuki@k.u-tokyo.ac.jp

***Abstract*—Oxford Nanopore Technologies MinION sequencer is a recent technology of portable, real-time, single molecule long-read DNA/RNA sequencing which has the potential to analyze clinical and environmental samples in the field. Here, we describe the result of metagenomic sequencing from samples with P30-gene-positive using MinION sequencer. One hundred cattle blood samples which were confirmed positive of blood parasites using blood smear method were collected from Diseases Investigation Centre (DIC), Yogyakarta province. After DNA extraction, each sample was amplified by conventional PCR using primers targeting the P30 gene. PCR products were visualized under UV illumination after electrophoresis on 1.5% TBE/agarose gels and stained with SYBR Safe™ DNA gel stain. Twenty-five samples showed the desired 400-bp DNA fragment, which was inferred as P30 gene positive. Four whole cattle blood samples with the highest DNA concentration were sequenced separately using MinION sequence for 24 hours. Sequencing data was then analyzed using bioinformatics tools. Three samples resulted in low sequencing yield and could not be analyzed, but we were able to identify Theileria orientalis as the major pathogen from one sample. We were also able to construct 95.8% of its genome. To our knowledge, this is the first publication of a parasite genome reconstruction from metagenomic sample using MinION sequencer in Indonesia.**

Keywords— MinION, sequencer, P30 gene, apicomplexan

I. Introduction

Oxford Nanopore Technologies MinION sequencer is a new technology allowing the direct sequencing of long DNA/RNA fragments read in real time. The rapid sequencing speed and low upfront instrument cost are some of the features drawing interest in this device for the genomics community. Its portable size will also allow sequencing in the field with only a standard laptop equipped with a standard USB 3.0 port for a quick genetic assessment [1,2]. The sequencer has various purported advantages over conventional next generation sequencing platforms, which include direct sequencing of nucleic acids allowing modification detection of DNA and the perform of direct RNA sequencing, real time data result and analysis, and the potentiality to generate long reads [2].

The sequencing principle relies on ultrasmall pore proteins with nanometer size which are embedded on a polymer membrane. Electricity will flow through the nanopores and bringing the DNA/RNA molecules with it. A sensor inside the nanopores will detect the electrical current disturbance and through a computer algorithm the specific electrical current patterns can be associated with specific sequence of nucleic acids [3].

Apicomplexa are protozoan parasites responsible for numerous human and livestock animal diseases. They are obligate parasites and have a complex life cycle, including intra and extracellular stages as well as multispecies [4]. Plasmodium spp., Theileria, Babesia, and *Toxoplasma gondii* have a high global negative impact on animal health [5,15]. Plasmodium have an intraerythrocytic stage where the organisms split by merogony and sexual stages occurring in the midgut of the arthropod vector [7]. *Theileria orientalis* is the causative agent of oriental theileriosis which infecting cattle [20]. *T. orientalis* has been reported to cause mortality in up to 5% of infected cattle [6,8]. Clinical outbreaks commonly occur when naive cattle are introduced into endemic herds, when animals undergo stress through transportation or are immunosuppressed [19]. *Babesia bovis* occurs in tropical and subtropical areas worldwide, causing fevers which leads to important economic losses [17]. *Toxoplasma gondii* has an exceptionally broad host range of the phyla Apicomplexa. It is the most successful parasites to infect, persists in most warm-blooded animal [16,18,22,23], and causes disseminated tissue infection as well as congenital disease [9,21]. It is also problematic in immunocompromised

 ISBN: 978-623-92207-1-6

individuals and has been known to cause abortion in animals and humans [10,11].

The immunodominant surface antigen of *T. gondii*, surface antigen 1 (SAG1) or P30, is generally expressed in *Escherichia coli* as a fusion protein containing a majority of the SAG1 protein supplied with six histidyl residues in the N-terminal end [10]. P30 is specifically expressed on the surface of the parasite during the tachyzoite stage, but not during the sporozoite or bradyzoite stage. P30 is one of the major surface antigens in the tachyzoite stage of *T. gondii*. It elicits high antibody titers in infected individuals. P30 is a very large protein [12,13,14] and less specific [24]. In this study, we performed metagenomic sequencing to identify the species from P30-gene-positive samples using MinION sequencer.

II. METHOD

A. Sample Collection

A hundred of cattle blood samples which were confirmed positive of blood parasites using the blood smear method were collected from Diseases Investigation Centre (DIC), Yogyakarta province. Blood samples source from 3 districts (Galur, Lendah and Panjatan) in Kulon Progo regency Yogyakarta province.

DNeasy® Blood & Tissue Kit (Qiagen, Hilden., Germany) and DNA-XT™ DNA extraction kit (QuantuMDx) were used to isolate genomic DNA following the manufacturer's instructions. The concentration and volume of extraction product determined by fluorescence using a Qubit™ fluorimeters (Thermo Fisher Scientific).

B. Preparing P30 Gene Amplicons

Polymerase chain reaction (PCR) was used to amplify a fragment of P30 gene. GoTaq® Green Master Mix (Promega), forward primer (5'-CAC ACG GTT GTA TGT CGG TTT CGC-3'), and reverse primer (5'-TCA AGG AGC TCA ATG TTA CAG CCT-3') were mixed in a reaction tube together with DNA sample. Reactions were cycled 40 times using a pre-denaturation step at 94ºC for 5 minutes, denaturation step of 94ºC for 30 seconds, annealing at 55ºC for 30 seconds, elongations at 72ºC for 25 seconds and extension at 72ºC for 5 minutes. Amplicon of P30 gen was visualized under UV illumination following electrophoresis on 1.5% TBE/agarose gels and stained with SYBR Safe™ DNA gel stain.

C. Sequencing Library Preparation

One microgram in 48 µl distilled water of a whole blood DNA extract was used for library preparation. First, DNA was repaired with 3.5 µl NEBNext FFPE DNA Repair Buffer, 2 µl NEBNext FFPE DNA Repair Mix, 3.5 µl Ultra II End-prep buffer and 3 µl Ultra II End-prep enzyme, incubated for 5 minutes at 20ºC and 5 minutes at 65ºC. After clean-up with AMPure XP, the repaired DNA was subjected to ligation using SQK-LSK109 sequencing kit from Oxford Nanopore Technologies according to the manufacturer's protocol. The final concentration of 300 ng of repaired and ligated DNA in 12 µl elution buffer was used for sequencing after priming the flow cell according to manufacturer's protocol.

D. Sequencing Process

MinION device, flow cell (FLO-MIN106) and MinKNOW software were set up and connected to a regular laptop. Prior to sequencing, flow cell check had to be performed to obtain the most optimal number of pores for sequencing. Oxford Nanopore Technologies suggests a minimal of 800 pores for sequencing. We loaded the DNA library to the loading port and sequencing was performed for 24 hours. Due to the relatively low computational power of a regular laptop used for sequencing, we opted to convert the raw data after the sequencing was finished.

E. Bioinformatics Analysis

Raw data obtained from sequencing process was converted to FASTQ file using Guppy v3.6.1 (Oxford Nanopore Technologies). After removal of reads mapped to host (*Bos taurus*) with in-house Python script, the remaining reads were subjected to BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and OneCodex (https://www.onecodex.com/) analyses to identify the pathogen species. Reads mapped to *T. orientalis* were selected for whole genome construction. Reads correction with Miniscrub v0.4 and trimming with Porechop v0.2.4 were performed before genome assembly. Reference-guided whole genome construction was performed with Pomoxis v0.3.6 using *T. orientalis* strain Shintoku reference genome (GenBank: GCA_000740895.1). Quast (v5.0.2) and D-GENIES (http://dgenies.toulouse.inra.fr/) were used for quality control of the assembly.

III. RESULTS

F. P30 Gene Amplification

After 40 cycles amplification using primers specific for P30 gene, positive samples were confirmed if they had a 400-bp DNA fragment (Fig. 1) with electrophoresis.

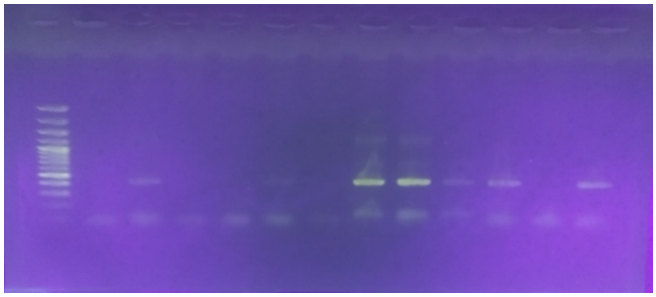

Fig.1 Amplification of P30 gene with PCR which showed the 400-bp DNA fragment.

G. Evaluation of MinION Process

The DNA from the whole cattle blood with positive results was extracted and four samples with the highest DNA concentrations were continued to sequencing using MinION sequencer in separate running for 24 hours.

Sequencing with MinION was successful in one run and failed in three others. Less concentration after sequencing library preparation was one of the problems. The library

ISBN: 978-623-92207-1-6

preparation needs relatively high concentration, high quality, and non-fragmented DNA together with some specialized laboratory tools and the competency of technicians and staffs to obtain the optimum condition for successful sequencing. To resolve the requirement of DNA concentration necessity, some protocol modifications can be performed [24]. In addition, a low input MinION library preparation kit is also available.

The quality of the successful MinION sequencing process was proper but still not optimum. It yielded 3.93 Gb with total reads of 347.680 (Table 1.). About 121.189 reads with more than 10 kb and 40 reads >100 kb. It showed that MinION sequencer can sequence long DNA fragments. Nanopores can read the length of DNA or RNA presented to them — from short to ultra-long (longest >1 Mb) [25]. Maximum length was 184.990 but with a mean length of 11.302, it showed the DNA was generally fragmented. The mean quality of individual read was 7.6 which was significantly lower than second-generation sequencer. However, the quality of the individual reads can be improved using bioinformatics tools.

TABLE I. STATISTICS OF MINION SEQUENCING

Parameter	Result
All reads	**Total.gigabases: 3.9297593** **Total.reads: 347680** Mean.length: 11302.8 Median.length: 5522.0 **Max.length: 184990.0** **Mean.q: 7.6** Median.q: 8.2
Reads	>10kb: 121189 >20kb: 67473 >50kb: 8134 >100kb: 40 >200kb: 0 >500kb: 0 >1m: 0 Ultralong: 88.0
Gigabases	>10kb: 3.1191909 >20kb: 2.3475811 >50kb: 0.4842819 >100kb: 0.0046033 >200kb: 0.0 >500kb: 0.0 >1m': 0.0 Ultralong: 0.0091204

H. Species Identification

P30 gene is commonly found in tachyzoite stadium of *Toxoplasma gondii* live cycle [12,13,24]. However, BLAST, identified *T. orientalis* as the major pathogen. We also confirmed the results using another species identification program, OneCodex. In agreement to BLAST, OneCodex also showed *T. orientalis* as the major pathogen (Fig.2). Although only 2.85% of 66,112 reads were classified using the OneCodex database, the software was confident with the result.

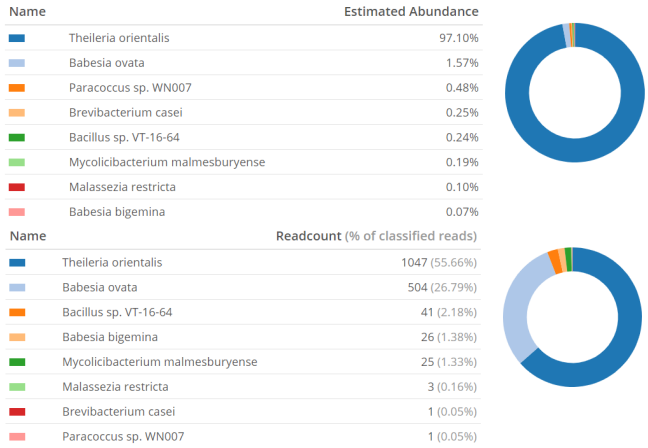

Fig.2 OneCodex shows *T. orientalis* as the major pathogen. Only 2.85% of 66,112 reads were classified using the OneCodex Database.

Primers used to amplify P30 gene was not species specific. Therefore, although we suspected of a Toxoplasmosis case, apparently sequencing confidently pointed out this was a case of *T. orientalis* infection. More study using specific gene for *T. orientalis* is needed to prove the result molecularly.

I. Whole genome Construction (De Novo)

As MinION is now very capable to sequence whole genome of pathogens from isolates or clinical samples directly, we tried to construct *T. orientalis* genome from this metagenomic sequencing. There are two methods to construct whole genome: without any reference (de novo) or reference-guided. De novo whole genome construction is preferable so no bias from the reference is introduced, but it needs a high coverage for successful construction.

After filtering only reads mapped to *T. orientalis,* we were left with only 17,443,121 bp to construct a 9 Mb *T. orientalis* genome. With a coverage of 1.9x, Flye and Canu long-read genome assembler could not run.

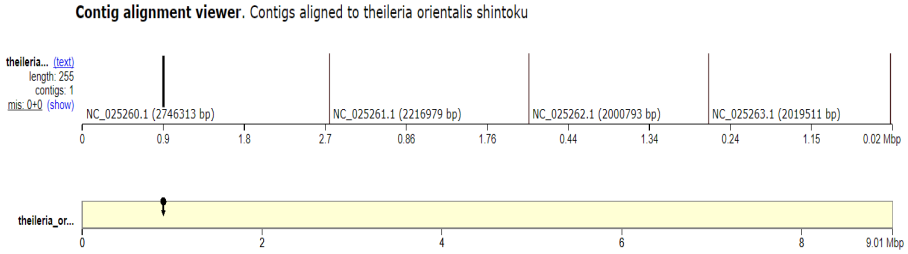

Fig.3 De-novo-constructed contig mapped to *T. orientalis*. The yellow bar shows the reference genome and the bold black line shows the location where the assembly was mapped in the genome. The thin black lines show the separation of chromosomes in the genome sequence.

ISBN: 978-623-92207-1-6

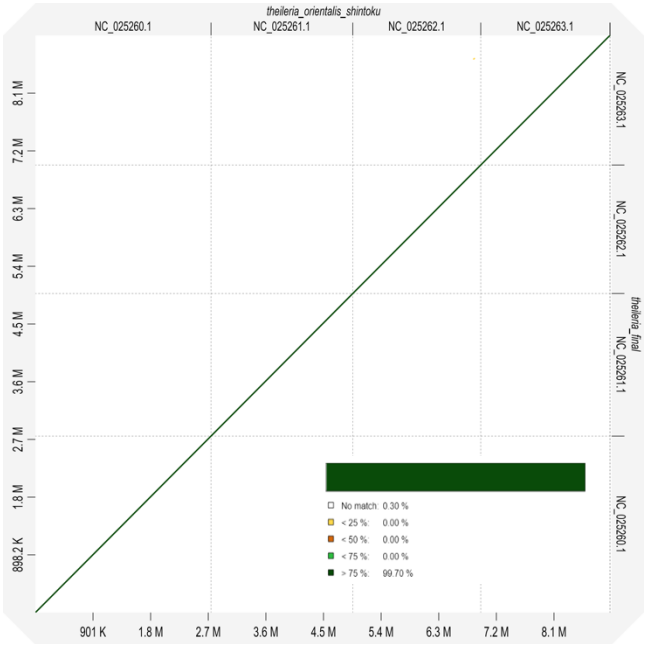

Fig.4. Visualization of complete the *T. orientalis* genome construction using D-GENIES.

Instead, we constructed whole genome using all of the reads with quality score > 7 (this includes reads mapped to the host *Bos taurus*). Around 249,337 reads were subjected to assembly with Flye v2.5. The software was able to create 41 contigs. The largest contig was contig 7 with the length of 409,347 bp. The total length of all contigs was 2,292,900 bp. The total length suggested that the constructed genome was incomplete either for host and *T. orientalis*. Mapping of the contigs to *T. orientalis* genome showed that contig 15 was mapped to the genome. We then isolated only this contig and run QUAST to visualize the assembly. Analysis of this contig showed that only 0.03% (255 bp) of the contig was able to be mapped to the genome, suggesting non-assembly (Fig.3).

J. Whole genome construction (reference-guided)

We then performed reference-guided assembly using Pomoxis. We opted to correct and trim our data manually instead of relying on Pomoxis modules. For correction we used Miniscrub and trimming of suspicious sequence with PoreChop. Assembly yielded 4 contigs with a total length of 8,981,984 bp. The total length suggests that the constructed genome was near complete. Indeed, when we visualized the result with D-GENIES, we found that 95.8% of the genome was constructed with high identity to the reference genome (Fig.4).

IV. CONCLUSION

P30-gene-positive sample which were sequenced using MinION sequencer confidently showed that *T. orientalis* was the major pathogen. Although we could not construct the whole genome of the parasite de novo, reference-guided construction was successful to construct 95,8% of the genome with high identity.

ACKNOWLEDGMENT

The authors thank the following institutions for their advice and support: One Health Collaborative Center UGM, Diseases Investigation Centre Wates District, Laboratory of Systems Genomic The University of Tokyo and Biotechnology Study Center Laboratory.

REFERENCES

[1] K. Brookes, T. Patel, G. Z. Erazo, I. Barber, A. Braae, N. Clement, T. G. Baranes, S. Chappell and K. Morgan "Identifying Polymorphisms in the Alzheimer's Related APP Gene Using the Minion Sequencer" , Journal of Next Generation: Sequencing & Applications, 3:1, 2016.

[2] A. D. Tyler, L. Mataseje, C. J. Urfano, L. Schmidt, K. S. Antonation, M. R. Mulvey and C. R. Corbett, "Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Application. Sci Rep 8, 10931, 2018.

[3] A. Patel, E. Belykh, E. J. Miller, L. L. George, N. L. Martirosyan, V. A. Byvaltsev dan M. C. Preul "MinION rapid sequencing: Review of potential applications in neurosurgery", Surgical Neurology International, 9:157, 2018.

[4] M. Preller, D.J. Manstein, "Molecular motors and motility", Comprehensive Biophysics, Vol. 4, pp. 118-150 , 2012.

[5] C. E. Suarez, R. P. Bishop, H. Z. Alzan, W. A. Poole, and B. M. Cooke, "Advances in the application of genetic manipulation methods to apicomplexan parasites", International Journal for Parasitologyjournal Vol 47 pp. 701-710, 2017.

[6] J. Yam, D.R. Bogema and C. Jenkins, "Oriental Theileriosis", Intechopen, 8 November 5 th 2018.

[7] M. G. Silva, M. W. Ueti, J. Norimine, M. F. Christensen, R. G. Bastos, W. L. Goff, W. C. Brown and A. Oliva, C. E. Suarez, "Babesia bovis expresses Bbo-6cys-E, a member of a novel gene family that is homologous to the 6-cys family of Plasmodium" Parasitology International, vol. 60, Issue 1, pp. 13-18, January 2011,

[8] H. Hakimi, T. Ishizaki, Y. Kegawa, O. Kaneko, S. I. Kawazu and M. Asada "Genome editing of Babesia bovis Using the CRISPR/Cas9 System", Msphere Vol. 4, Issue 3, May/June 2019.

[9] J. G. Montoya and O. Liesenfeld. 2004. Toxoplasmosis. Lancet 363 : 1965.

[10] R. Abdizadeh,S. Maraghi, A. A. Ghadiri, M. Tavalla, and S. Shojaee, "Cloning and Expression of Major Surface Antigen 1 Gene of Toxoplasma gondii RH Strain Using the Expression Vector pVAX1 in Chinese Hamster Ovary Cells", Jundishapur J Mikrobiol, Vol. 3 (8) March 2015.

[11] Wan. O. Abdullah "Toxoplasma gondii Recombinant Surface Antigen (SAG 1; P 30) and its Applications in Serodiagnosis and Vaccine Development for Toxoplasmosis", Malaysian Journal of Medicine and Health Sciences Vol. 6 (1) January 2010.

[12] A. Makioka And A. Kobayashi "Toxoplasmacidal Activity of Macrophages Activated by Recombinant Major Surface Antigen (P30) of Toxoplasma gondii", Infection And Immunity, pp. 2851-2852, August 1991.

[13] L. H. Kasper, "Identification of stage-specific antigens of Toxoplasma gondii", Infect. Immun. 57:668-672, 1989.

[14] L. H. Kasper, M. S. Bradley, and E. R. Pfefferkorn, "Identification of stage-specific sporozoite antigens of Toxoplasma gondii by monoclonal antibodies", J. Immunol, 132:443- 449, 1984

[15] N. S. Morrissette and L. David Sibley "Cytoskeleton of Apicomplexan Parasites", Microbiology And Molecular Biology Reviews, pp. 21–38, Mar. 2002.

[16] O. A. Mendez and A. A. Koshy "Toxoplasma gondii: Entry, association, and physiological influence on the central nervous system", Plos Pathogens, July 2017.

[17] H. Mehlhorn and E. Schein "The Piroplasms: Life Cycle and Sexual Stages" Vol. 23, pp. 37-103. 1985.

ISBN: 978-623-92207-1-6

[18] F. Emmanuel G. Vieira, J. P. Sasse, A. F. Minutti, A. C. Miura, L. D. Barros, S. T. Cardim, T. A. Martins, M. Seixas, M. I. Yamamura, C. S. and J. L. Garcia, "Toxoplasma gondii prevalence and characterization of new genotypes in free-range chickens from south Brazil", Parasitology Research, 117:681–688, 2018.

[19] J.G. Watts, M.C. Playford and K.L. Hickey, "heileria orientalis: a review", New Zealand Veterinary Journal, 2015

[20] H. Gebrekidana, R. B.Gassera, G. Baneth, D. Y. Landau, Y. N. Biala, A. Hailu and A. Jabbar, "Molecular characterization of Theileria orientalis from cattle in Ethiopia". in ticks and tick-borne diseases, Vol. 7, Issue 5, July 2016, pp. 742-747.

[21] F. Nogareda, Y. Le Strat, I. Villena, H. De Valk and V. Goulet. Incidence and prevalence of Toxoplasma gondii infection in women in France, 1980–2020: model-based estimation. Epidemiol. Infect 142, 2014.

[22] C. Lekutis, D. J. P. Ferguson, M. E. Grigg, M. Camps, and J. C. Boothroyd,"Surface antigens of Toxoplasma gondii: variations on a theme", International Journal for Parasitology, Vol. 31, Issue, pp. 1285-1292. October 2001.

[23] A. Bergmann, K. Floyd, M. Key, C. Dameron, K. C. Rees, L. B. Thornton, D. C. Whitehead, I. Hamza and Z. Dou "Toxoplasma gondii requires its plant-like heme biosynthesis pathway for infection", Plos Pathogens. May 2020.

[24] C. D. Jones, N. Okhravi,P. Adamson, S. Tasker and S. Lightman "Comparison of PCR Detection Methods for B1, P30, and 18S rDNA Genes of T. Gondii in Aqueous Humor", IOVS, Vol. 41, No. 3, March 2000.

[25] https://nanoporetech.com/products/minion. Why MinION?. Accesss date March 2021.