In silico restriction site analysis for characterization of Toxoplasma gondii isolate

F Ekawasti1,2, U Cahyaningsih1,*, S Sadi’ah1, N L P I Dharmayanti2 and D T Subekti2

1Department of Veterinary Medicine, IPB University, Bogor, Indonesia
2Indonesia Research Center for Veterinary Science, Bogor, Indonesia
*Corresponding author e-mail: umi-ch@apps.ipb.ac.id

Abstract. Toxoplasma gondii infection is a serious major public health concern. Toxoplasmosis is a disease that affects humans and warm-blooded animals all over the world. The virulence and severity of the sickness may be influenced by parasite load. The three clonal lineages for biological research are T. gondii genotypes I, II, and III. Using primer genes for gra1, rop1, and mic3, which have been identified as essential proteins for tachyzoite invasion and replication in host cells. PCR was used to determine the genotype of a T. gondii isolate from the Indonesian research center for veterinary science. In silico restriction site analysis was performed on T. gondii isolates using CLC sequence viewer 8.0 software to assess sequence data for the existence of restriction enzyme patterns. Predictable restriction fragment length polymorphism using In silico analysis. T. gondii sequencing genes are In silico digested with different restriction enzymes. The findings show that they can easily distinguish archetypal parasites from biotypes and estimate the genetic diversity of the parasites. According to the interpretation of the data, isolate T. gondii belongs to strain RH genotype I.

Keywords: fragment, genotype, marker, polymorphism, toxoplasma.

1. Introduction

Toxoplasma gondii is one of the most prevalent zoonotic illnesses, affecting almost one-third of the global population [1]. T. gondii was formerly thought to be a single species of the genus Toxoplasma. Early research on parasite strains from North America and Europe revealed a limited genetic variety that was divided into genetic categories I, II, and III [2]. The implications of T. gondii infection may vary depending on parasite genotype and host species [3]. Disease symptoms in humans range from mild to severe acute toxoplasmosis. Severe toxoplasmic retinochoroiditis is more likely to be linked with genotype I as a virulent type [4]. Outbred mice are universally fatal to type I isolates, but type II and III isolates are substantially less virulent [5]. T. gondii infection has nonspecific clinical signs that make diagnosis difficult. Traditional method for T. gondii diagnosis relies on bioassays and serological testing, which have limitations in detecting and distinguishing parasite strains [6]. Because of their great sensitivity and specificity, molecular techniques for detecting T. gondii infection are tempting. Based on a genetic study using polymerase chain reaction-restriction fragment length polymorphism, T. gondii has a clonal population. DNA fragmentation methods commonly employed for the genetic characterization of T. gondii by PCR-RFLP [7].

T. gondii virulence may also be determined in cell culture by monitoring the rate of development and capacity to pass the biological barrier [8]. Other researchers devised a number of methods for predicting
of *T. gondii*. It has been shown that PCR-RFLP can distinguish between virulent and avirulent *T. gondii* strains. In the meantime, the genotyping method of choice is PCR-RFLP. Unfortunately, the PCR-RFLP method requires a significant amount of effort, time, and money. Alternative approaches must be sought in these cases. As a result, a different technique is required, such as restriction site analysis using sequence encoded by *gra1*, *rop1*, and *mic3*. The genes are the most specific and sensitive for identifying *T. gondii*, and have been identified as important proteins during tachyzoite invasion and reproduction in host cells [9]. In silico virulence prediction studies based on gene sequences are fast and can decrease identification mistakes in *T. gondii* strain classification. In silico gene sequence studies can predict the overlapping reactions that will be caused by restriction enzymes in PCR-RFLP [9, 10].

The "digest DNA" function may be used to perform In silico restriction digestion of provided DNA sequences to forecast the predicted DNA pieces. In silico analysis is a method of analyzing the restriction fragment length distribution using a computer. Comparison of restriction patterns for two or more DNA sequences might be performed using In silico restriction digestion sites. When In silico restriction gene program comes into contact with a DNA sequence having a structure that matches part of the numerous restriction enzymes routinely utilized in laboratories to build recognition sites [11], it searches for restriction sites. Based on this information, *gra1*, *rop1*, and *mic3* were selected as genetic markers for predicting genotype of *T. gondii* using In silico restriction site. Based on sequences gene of PCR products compared with sequences gene from National Center for Biotechnology Information Search database (NCBI) with various restriction enzymes. This study is the first attempt to explore the possibility of using a restriction site analysis method based on a sequence of *gra1*, *rop1*, and *mic3* for predicting the genotype of *T. gondii*.

2. Materials and Methods

2.1. DNA isolation

*T. gondii* tachyzoites (virulent strain RH) were isolated from culture cells at the Indonesia Research Center for Veterinary Sciences (IRCVS). *T. gondii* DNA was isolated from a liquid nitrogen stored isolation. The PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific–Invitrogen, Carlsbad, CA) was used to extract the DNA template. Thermo Scientific NanoDrop Products, Inc. was used to purify and concentrate DNA extracts for DNA quantification, and extracts were kept at -20°C before PCR testing [9].

2.2. Genetic markers and DNA sequencing

The three genetic markers used in this study were *gra1*, *rop1*, and *mic3*. The primers used are presented as follows *gra1* were: F: 5’-CGGTTCCTTTGTTGTGTTTTG-3’, R: 5’-CATGGGTAAGTATACACAACA-3’ [12]. *Rop1* were: F: 5’-CGTGCATACTAATGCACTGAC-3’, R: 5’-ATCCTCAAATCAGTAC -3’ [13] and *mic3* were F: 5’-GTTGTGATATCCTTTGCTCAATCTGGAATCG-3’. R: 5’-CACGAAGCTTTGCCAATGCGGCG -3’ [14]. Amplification of DNA in PCR for *gra1*, *rop1*, and *mic3* revealed *T. gondii*. In a Thermal Cycler, PCR was performed (Bio-Rad). 25 L reactions using 1 L DNA template, 0.25 M primer, 1.5 mM MgCl2, 0.01 U Taq DNA polymerase, and 0.2 mM dNTP were used in the PCR tests. 95°C for 1 minute, 58°C for 1 minute, 72°C for 2 minutes, followed by 35 cycles of amplification. PCR product electrophoresis was carried out using agarose 1.5% and SYBRTM safe staining at a voltage of 100 volts. A UV transilluminator was used to see the bands [9]. The sequenced and recognized PCR products from the amplification of a gene's target region. The Bioneer Sequencing technique was used to accomplish the work. CLC Sequence Viewer Version 8.0 software (https://clc-sequence-viewer.software.informer.com/8.0/) was used to examine DNA sequences.

2.3. In silico restriction site analysis

The three *T. gondii* sequence genes *gra1*, *rop1*, and *mic3* were retrieved using PCR product sequences and the National Center of Biotechnology Information's (NCBI, US National Library of Medicine, https://www.ncbi.nlm.nih.gov/genbank/) GenBank database. Each strain's genes are retrieved from the
accessible database. The chosen strain is one whose virulence status and genotype have been determined. Based on their virulence, *T. gondii* strains were classified as virulent (type I) or avirulent (types II and III). The genes of *gra1, rop1*, and *mic3* in virulent and avirulent strains were retrieved from the GenBank database, respectively, from reference strains with well-known virulence and genotype. All gene sequences in fasta format were imported into CLC Sequence Viewer Version 8.0 software (Qiagen, Denmark) (https://clc-sequence-viewer.software.informer.com/8.0/) to perform the restriction site analysis. Enzymes are selected from the list of enzymes that must be considered in the calculation, then processed by making a boundary map, a list of cutting enzymes and adding restriction sites as annotations to the sequence to determine the number of cutting sites and DNA patterns recognized by the enzyme.

### 3. Results and Discussion

PCR-RFLP (polymerase chain reaction-restriction fragment length polymorphism) is a technique for detecting genetic strain among species by amplification of a conserved area of DNA sequence using PCR and digestion of the generated fragment with certain restriction enzymes. We did not do RFLP PCR in the lab for this investigation. After performing PCR and receiving the findings of gene sequencing, move to the *In silico* restriction site by utilizing several different restriction enzymes from the *In silico* software's enzyme list for each of *T. gondii* *gra1, rop1*, and *mic3* genes.

Based on the *In silico* analysis, The findings revealed that the *gra1* gene of *T. gondii* sequences derived from the results of the PCR product sequencing in this study and the gene sequence derived from the GenBank database, namely ID. HM067753.1 (RH), ID. EU983103.1 (RH) and ID. XM_002365660.2 (ME 49).

| No | Genes/Accession number | Strain | Type | Size (bp) | Restriction site |
|----|------------------------|--------|------|----------|-----------------|
|    |                        |        |      |          | DdeI<sup>C</sup>*TNAG | HinfI<sup>G</sup>*ANTC | HindIII<sup>A</sup>*AGCCTT | Sau96I<sup>G</sup>*GNCC | EcoRI<sup>G</sup>*AAGTTC | XhoI<sup>C</sup>*TCGAG |
| 1  | GRA1                   | RH     | I    | 783      | 3            | 3               | 2               | -               | -               | -               |
|    | HM067753.1             | RH     | I    | 639      | 2            | 2               | 1               | -               | -               | -               |
|    | EU983103.1             | RH     | I    | 573      | 2            | 2               | 2               | -               | -               | -               |
|    | XM_002365660.2         | ME49   | II   | 1959     | 5            | 6               | -               | 3               | -               | -               |
| 2  | ROP1                   | RH     | I    | 1218     | 1            | 10              | -               | 11              | -               | -               |
|    | AY661790.1             | RH     | I    | 1249     | 1            | 9               | -               | 11              | -               | -               |
|    | AF350261.1             | RH     | I    | 1191     | 1            | 10              | -               | 11              | -               | -               |
|    | XM_002364175.2         | ME49   | II   | 4246     | 9            | 17              | -               | 12              | -               | 4               |
| 3  | MIC3                   | RH     | I    | 1129     | 4            | 8               | -               | 5               | 2               | 1               |
|    | AF509564.1             | RH     | I    | 1130     | 3            | 8               | -               | 5               | 2               | 1               |
|    | EU572718.1             | RH     | I    | 1080     | 3            | 7               | -               | 6               | 2               | 1               |
|    | GU121670.1             | RH     | I    | 1075     | 3            | 7               | -               | 4               | 2               | 1               |
|    | AJ132530.1             | ME49   | II   | 2247     | 5            | 15              | -               | 6               | 2               | 3               |
|    | XM_018782955.1         | ME49   | II   | 3837     | 9            | 24              | -               | 7               | 3               | 4               |

As shown in Figure 1 and Table 1 was show the pattern of restriction sites formed using *DdeI, HinfI*, and *HindIII* restriction enzymes. Based on Figure 1, it can be seen that there is a similar pattern formed between *T. gondii* strains RH (a), (b), and (c), while *T. gondii* strain ME49 (d) has a different pattern. This indicates that the restriction site formed on the *gra1* gene marker can differentiate between the two strains of *T. gondii*. Restriction site analysis can also be used to analyze the diversity of a group of isolates to the level of strain, in addition to DNA sequence analysis testing [9].
Figure 1. GRA1 restriction site by CLC Sequence Viewer Ver. 8, (a). Sequence gene from PCR product in the present study, (b) HM067753.1 (RH/virulent), (c) EU983103.1 (RH/virulent) and (d) XM_002365660.2 (ME49/avirulent).

Figure 2. The rop1 restriction site by CLC Sequence Viewer Ver. 8, (a) Sequence gene from PCR product in the present study, (b) AY661790.1(RH/virulent), (c) AF350261.1 (RH/virulent) and (d) XM_002364175.2 (ME49/avirulent).

The rop1 gene of T. gondii sequences derived from the results of the PCR product sequencing in this study and the gene sequence derived from the GenBank database, namely ID AY661790.1 (RH), ID. AF350261.1 (RH) and ID. AF350261.1 (ME 49). As shown in Figure 2 and Table 1 was shown the pattern of restriction sites formed using DdeI, HinfI, HindIII, and Sau961 restriction enzymes. It produced by enzyme digestion shows in Figure 2. that there are similarities between samples (a) and (b), (c) and different from a pattern (d) which has a pattern shape and number of restriction sites with a higher number of restriction enzymes. Based on these data, the genotyping rop1 gene of T. gondii can be determined as a virulent genotype (RH/I).
The $mic3$ gene of $T. gondii$ sequences derived from the results of the PCR product sequencing in this study and the gene sequence derived from the GenBank database, namely ID. AF509564.1 (RH), ID. EU572718.1 (RH), ID. GU121670.1 (RH), ID. AJ132530.1 (ME49) and ID. XM_018782955.1 (ME 49), as shown in Figure 3 and Table 1, showed the pattern of restriction sites formed using $DdeI$, $HinFI$, $HindIII$, $Sau96I$, $EcoRI$, $XhoI$ restriction enzymes. The pattern produced by enzyme digestion is shown in Figure 3 which has a restriction pattern that is almost similar between the pattern of the $T. gondii$ RH strain and the ME 49 strain. It may happen because the gene sequences produced in PCR are not perfect and specific so that there can be overlapping restriction sites between virulent and avirulent strains. Based on the research and discussion above, two features of this technique might be deemed limits. First, the gene sequences must be compared in this study. Second, if the $T. gondii$ being studied is an unusual genotype, prediction errors may occur. The prevalence of these mistakes, however, is relatively low and is restricted to genotype grouping. We believe that this is not a severe issue because it may be mitigated by running a combination study using many genes and enzymes [9].

The pattern of fragments formed by virulent and avirulent strains can be distinguished with detailed analysis, namely the total number of sites produced and the type of enzyme used. According to Figure 3 shows that the restriction pattern has similarities between (a) and (b), (c) and different from (d), which has a pattern shape and number of restriction sites with a higher number of restriction enzymes. Based on these data, the genotyping $mic3$ gene of $T. gondii$ can be determined as a virulent genotype (RH/I).
Based on the results of the In silico restriction site in this study, the best restriction pattern for genotyping *T. gondii* was using the *gra1* and *rop1* genes. Restriction sites with *gra1* and *rop1* genes using several restriction enzymes showed good results. The restriction enzymes produced a polymorphic digestion pattern from the PCR product that virulent RH strain of *T. gondii* (type I) as acutely virulent phenotype by the genotype PCR-RFLP characterization [9, 10]. These genes are cut with restriction enzyme according to the cutting site then it will produce differences in the size of the DNA fragments (Table 1). A whole-genome examination of restriction sites can give more information to assist in enzyme selection decisions. Massive volumes of data are generated by high-throughput sequencing technology. Genotyping and mapping bias in genomic sequencing can be reduced using an In silico restriction enzyme [11]. In general, however, the analysis of In silico does not allow for evaluating the number of obtained DNA fragments of the given length to predict their visualization in the experiment [15]. In silico analysis does not allow for the estimation of the number of DNA fragments collected in order to forecast their visibility in the experiment [15]. Restriction enzyme-based molecular markers are commonly used in genetics applications. The user must examine the usage of enzymes ahead of time to verify that the enzymes chosen are appropriate for the study and kind of DNA processing to be performed [16].

In silico preliminary testing can be used to identify and select the appropriate restriction enzymes for use in the laboratory (wet lab). It is feasible to forecast the restriction sites that will develop on the agarose gel by using In silico restriction analysis. As a result, this In silico technology can reduce the amount of time spent in the laboratory performing the restriction fragment length polymorphism test. Due to the reduced requirement for laboratory supplies, this technique based on In silico restriction site analysis is very simple, broadly applicable, repeatable, and cost-effective. When the study involves iterative interactions, extra analysis, or the comparison of two pieces of research, this information might save time. Depending on the quantity of input sequences or enzymes being processed, computer processing rates can be modified. In silico study is based on the FASTA format base sequence, which lacks any sequence-specific methylation type or methylation susceptible restriction enzyme site identification. At this stage, methylation can be identified more easily based on any methylation-induced change in restriction enzyme activity, such as needing methylation, requiring no-methylation, or any partial specificity. Overall, this strategy has several advantages. When compared to PCR-RFLP, the findings of which are often ambiguous since the amount of DNA used is insufficient for the needs of PCR analysis, this approach can better predict *T. gondii* genotype and is easier to conclude if the study uses a combination of genes. Apart from that, this approach does not rely on live bacteria [10].

4. Conclusion

In silico restriction sites were able to distinguish archetypal from genotypic parasites and estimate the genetic diversity of the parasites fast and easily. According to the interpretation of the data, isolate *T. gondii* belongs to strain RH virulent genotype I. In silico analysis with multiple restriction enzymes was used to analyze the sequences and genes. In silico restriction site analysis is a stand-alone, open-source front-end tool for making the most of next-generation sequencing data. It is intended to assist scientists in focusing on how much data can be acquired as well as how to effectively use it in the future.

References

[1] Saadatnia G, Golkar M A 2012 Scand J Infect Dis 44 805-814
[2] Liu Q, Wang ZD, Huang SY, and Zhu XQ 2015 Vectors 292 1-14
[3] Su C, Shwab EK, Zhou P, Zhu XQ, and Dubey JP 2010 Parasitology 137 1–11
[4] Grigg ME, Ganatra J, Boothroyd JC, and Margolis TP 2001 J Infect Dis 184 633–9
[5] Boothroyd JC, Grigg ME 2002 Curr Opin Microbiol 5 438–42
[6] Kotresha D, Noordin R 2010 Recombinant proteins in the diagnosis of toxoplasmosis APMIS 118 529–42
[7] Vijaykumar BR, Kant RS, Rajendran C, Lekshmi SU, Keerthana S, Mahadevan A, Shankar SK, and Jayshree RS 2018 Ann Indian Acad Neurol 21 250-255
[8] Barragan A, Sibley LD 2003 Trends Microbiol. 11 426-430
[9] Ekawasti F, Cahyaningsih U, Dharmayanti NLPI, Sa'diah S, Subekti DT, Azmi Z, and Desem MI 2021 Int. J. One Health 7 196-203
[10] Subekti DT, Ekawasti F, Desem MI, and Azmi Z 2021 J Vet Sci. 22 6 e88
[11] Roszik J, Fenyőfalvi G, Halász L, Karányi Z, and Székvölgyi L 2017 Methods Clin. Dev 6 (9) 66-67
[12] Subekti DT, Artama WT, Sulistyaningsih E, Poerwanto SH, Sari Y, and Bagaskoro F 2008 JITV 13 43-51
[13] Kusumaningsih P 2018 Teknologi, Sains, dan Sosial Humaniora
[14] Indrasanti D, Haryanto A, and Artama W 2011 Molekul 6 10
[15] Chernukhin VA, Abdurashitov MA, Tomilov V, Gonchar DA, and Degtyarev S 2006 Transl. Ovchinnikov Bull. Biotechnol.Phys. Chem. Biol. 2 (3) 39-46
[16] Sharpe RM, Koepke T, Harper A, Grimes J, Galli M, Satoh-Cruz M, Kalyanaraman A, Evans K, Kramer D, Dhingra A 2016 PLoS One 11 4 e0152404