J Vet Sci. 2021 Nov;22(6):e88
https://doi.org/10.4142/jvs.2021.22.e88
pISSN 1229-845X·eISSN 1976-555X

Journal of Veterinary Science

Rapid Communication
Parasitology

Check for updates

Toxoplasma gondii virulence prediction using hierarchical cluster analysis based on coding sequences (CDS) of *sag1*, *gra7* and *rop18*

Didik T Subekti [1,*], **Fitrine Ekawasti** [1], **Muhammad Ibrahim Desem** [1], **Zul Azmi** [2]

[1]Indonesia Research Center for Veterinary Science, Bogor, West Java 16124, Indonesia
[2]Goats Research Station, Sei Putih, North Sumatra 20585, Indonesia

OPEN ACCESS

Received: Apr 26, 2021
Revised: Sep 9, 2021
Accepted: Sep 22, 2021
Published online: Oct 1, 2021

***Corresponding author:**
Didik T Subekti, DVM., M.Med.Sc.
Indonesia Research Center for Veterinary Science, JL. R.E. Martadinata 30, Bogor, West, Java 16124, Indonesia.
E-mail: subektididik96@yahoo.com
didiktulus@pertanian.go.id

ORCID iDs
Didik T Subekti
https://orcid.org/0000-0002-2649-0242
Fitrine Ekawasti
https://orcid.org/0000-0001-6299-4539
Muhammad Ibrahim Desem
https://orcid.org/0000-0003-4921-2031
Zul Azmi
https://orcid.org/0000-0002-3006-2814

Conflict of Interest
The authors declare no conflicts of interest.

ABSTRACT

Toxoplasma gondii consists of three genotypes, namely genotype I, II and III. Based on its virulence, *T. gondii* can be divided into virulent and avirulent strains. This study intends to evaluate an alternative method for predicting *T. gondii* virulence using hierarchical cluster analysis based on complete coding sequences (CDS) of *sag1*, *gra7* and *rop18* genes. Dendrogram was constructed using UPGMA with a Kimura 80 nucleotide distance measurement. The results showed that the prediction errors of *T. gondii* virulence using *sag1*, *gra7* and *rop18* were 7.41%, 6.89% and 9.1%, respectively. Analysis based on CDS of *gra7* and *rop18* was able to differentiate avirulent strains into genotypes II and III, whereas *sag1* failed to differentiate.

Keywords: *Toxoplasma gondii*; genotype; virulence; cluster analysis

INTRODUCTION

Toxoplasma gondii is one of the most common zoonotic diseases and is reported to infect nearly one third of the world's population [1]. The successful isolation of various *T. gondii* strains has led to the knowledge of genetic variations in many isolates. Variations in *T. gondii* isolates showed distinct pathogenicity in mice and, therefore, *T. gondii* virulence determination is based on its pathogenicity in mice [2]. *T. gondii* is categorized as virulent if $LD_{100} = 1$ and avirulent if $LD_{50} \geq 10^3$ when its pathogenicity is tested on mice [2]. Other author divided avirulent into two types, namely intermediate avirulent with $LD_{50} \geq 10^3$ and avirulent with $LD_{50} \geq 10^5$ [3] while the other one classified as low virulence with $LD_{50} \geq 10^3$ and non virulent with $LD_{50} \geq 10^5$ [4]. However, hereinafter both opinions are only called avirulent strains, referring to the previously well-established categorization. [2,5].

T. gondii is known to have a clonal population based on genetic analysis by polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP) and microsatellite markers associated with its virulence. Based on its genetic analysis, the clonal population of *T. gondii* was divided into genotypes I, II and II [5]. Genotype I corresponds to a virulent strain, while genotypes II and III correspond to avirulent strains with $LD_{50} \geq 10^3$ and $LD_{50} \geq 10^5$, respectively

Author Contributions

Conceptualization: Subekti DT, Ekawasti F; Data curation: Subekti DT, Ekawasti F; Investigation: Subekti DT, Ekawasti F; Methodology: Subekti DT; Resources: Ekawasti F, Subekti DT, Azmi Z; Writing - original draft: Subekti DT, Desem MI; Writing - review & editing: Subekti DT, Desem MI.

[2-4]. Some virulent strains, such as CAST and GPHT, cannot actually be categorized into three existing genotypes (archetypal lineages), so they are called atypical (non-archetypal lineage) [6]. Virulence corresponds to the genotype of *T. gondii* that has been reported to be associated with clinical manifestations and severity of congenital cases [4,7,8].

T. gondii virulence can also be predicted through cell culture by observing their growth rate and their ability to penetrate through the biological barrier [9]. Other researchers also developed various approaches to predicting *T. gondii* virulence. These approaches include isoenzyme analysis using a six variable enzyme system [10] and serotyping using peptide membrane arrays [11]. However, all of them have produced unsatisfactory results. The mouse bioassay has proven to be the most accurate in establishing virulence, while PCR-RFLP is the preferred method for genotyping. Unfortunately, the PCR-RFLP method requires much effort and is often inconclusive, especially if the parasite strain is atypical.

However, there are circumstances where tachyzoite isolation fails to be achieved or cannot be grown due to tachyzoites dying, whereupon a mouse bioassay cannot be implemented. These situations require looking for alternative methods. Therefore, another approach is needed, including hierarchical cluster analysis, which is based on a complete coding sequence (CDS) of *sag1*, *gra7* and *rop18* that encode a virulence-associated protein of *T. gondii*. SAG1 and GRA7 proteins contain different structures between virulent and avirulent strains and are thought to be used to predict *T. gondii* virulence [10]. PCR-RFLP on *sag1* locus has been reported as being able to differentiate between virulent and avirulent strains of *T. gondii* [12]. Other proteins that have been shown to be virulence factors for *T. gondii* are GRA7 [13] and ROP18 [14]. Deletion of *gra7* or *rop18* genes results in partial attenuation of *T. gondii* [14]. However, if deletion is carried out in both genes, there is a complete attenuation that makes *T. gondii* become avirulent [14].

Based on this information, *sag1*, *gra7* and *rop18* were selected as genetic markers for predicting *T. gondii* virulence. This study is the first attempt to explore the possibility of using a hierarchical cluster analysis method based on a complete coding sequence of *sag1*, *gra7* and *rop18* for predicting *T. gondii* virulence.

MATERIALS AND METHODS

Virulent and avirulent strain

The virulence and genotypes of each *T. gondii* strain are obtained from the available genome database. The selected strain is one that has had its virulence status and genotype has been recognized. The genes of each strain are extracted from the available database, namely *sag1*, *gra7* and *rop18* (**Table 1**). Based on its virulence, *T. gondii* is grouped into virulent and avirulent strains. The genotype is divided into four groups, namely genotype I, II, III and atypical.

Selection of *sag1*, *gra7* and *rop18* genes

The three desired genes were obtained from the GenBank database provided by National Center of Biotechnology Information (NCBI, US National Library of Medicine, https://www.ncbi.nlm.nih.gov/genbank/). The *sag1*, *gra7* and *rop18* genes used must meet two criteria. First, the genes must be in a full-length sequence of CDS, which becomes a template for mature protein translation. Therefore, all partial sequences will be ignored. Second, all of these genes come from reference strains whose virulence and genotype are well known.

Table 1. Summary of *Toxoplasma gondii* virulence and genotypes used in this study

sag1 CDS				*gra7* CDS				*rop18* CDS			
Strains	Accession number	Vir.	Gen.	Strains	Accession number	Vir.	Gen.	Strains	Accession number	Vir.	Gen.
RH	JX045360	V	I	RH	JX045573.1	V	I	RH	JX045330.1	V	I
GT1	JX045361	V	I	GT1	JX045574.1	V	I	GT1	GQ243202.1	V	I
ENT	JX045418	V	I	ENT	JX045578.1	V	I	ENT	JX045328.1	V	I
VEL	JX045412	V	I	VEL	JX045580.1	V	I	VEL	JX045327.1	V	I
MOR	JX045394	V	I	MOR	JX045579.1	V	I	MOR	JX045326.1	V	I
FOU	JX045357	V	I	FOU	JX045576.1	V	I	FOU	JX045332.1	V	I
ME49	JX045362	AV	II	ME49	JX045583.1	AV	II	OH3	JX045324.1	V	I
DEG	JX045364	AV	II	DEG	JX045584.1	AV	II	ME49	JX045319.1	AV	II
PIH	JX045374	AV	II	PIH	JX045586.1	AV	II	DEG	JX045318.1	AV	II
WTD1	JX045375	AV	II	BEV	EU157141.1	AV	II	PTG	GQ243204.1	AV	II
WTD3	JX045376	AV	II	WTD1	JX045589.1	AV	II	PIH	JX045320.1	AV	II
RAY	JX045379	AV	II	WTD3	JX045595.1	AV	II	ARI	JX045322.1	AV	II
ARI	JX045387	AV	II	RAY	JX045590.1	AV	II	QHO	GQ243205.1	AV	II
R961	JX045389	AV	II	ARI	JX045611.1	AV	II	CTG	JX045346.1	AV	III
B41	JX045382	AV	II	R961	JX045594.1	AV	II	VEG	JX045348.1	AV	III
H44	JX045373	AV	II	B41	JX045592.1	AV	II	NED	JX045325.1	AV	III
CTG	JX045386	AV	III	H44	JX045588.1	AV	II	STRL	JX045351.1	AV	III
VEG	JX045384	AV	III	CTG	JX045617.1	AV	III	M7741	JX045353.1	AV	III
STRL	JX045385	AV	III	VEG	JX045618.1	AV	III	P89	JX045347.1	AV	Atyp.
M7741	JX045366	AV	III	NED	DQ459455.2	AV	III	RUB	JX045336.1	V	Atyp.
B73	JX045365	AV	III	STRL	JX045621.1	AV	III	BOF	JX045331.1	V	Atyp.
ROD	JX045405	AV	III	M7741	JX045623.1	AV	III	CAST	JX045345.1	V	Atyp.
GPHT	JX045415.1	V	Atyp.	B73	JX045622.1	AV	III				
BOF	JX045400	V	Atyp.	ROD	JX045619.1	AV	III				
CAST	JX045358	V	Atyp.	P89	JX045616.1	AV	Atyp.				
RUB	JX045356	V	Atyp.	RUB	DQ459450.2	V	Atyp.				
P89	JX045409	AV	Atyp.	GPHT	JX045582.1	V	Atyp.				
				BOF	JX045575.1	V	Atyp.				
				CAST	JX045613.1	V	Atyp.				

CDS, coding sequence; Atyp., atypical strains; Vir., virulence; Gen., genotypes; V, virulent; AV, avirulent.

Data analysis

Cluster analysis was performed under hierarchical cluster analysis by CLC Sequence Viewer Ver. 8. Dendrogram was constructed by UPGMA (unweighted pair group method with arithmetic mean) with a Kimura 80 nucleotide distance measurement. Successful grouping in dendrograms was evaluated based on the lowest number of individual strains misplaced into the wrong cluster.

RESULTS

The *sag1* coding sequence only differentiates *T. gondii* population into two clusters: virulent and avirulent (Fis. 1A and 2A). Misplacement occurs in a ROD strain that is categorized into the virulent category under genotype I (**Fig. 1A**), with an error rate of 4.3% (1/23). ROD has been recognized as an avirulent strain [10] and was assigned to genotype III [15]. In a further analysis involving atypical strains, the prediction error rate increased to 7.4% (2/27) due to misplacement of ROD and P89 strains (**Fig. 2A**). The dendrogram pattern in this study is similar to a previous study that used neighbour joining under Neil's genetic distance measurement [15]. In general, *sag1* is only suitable for predicting *T. gondii* virulence, has an error rate of 4.3%–7.4%, and has even failed to distinguish genotypes in avirulent strains.

The *gra7* coding sequence has been used successfully to classify archetypal isolates from the *T. gondii* population into three clusters without error (**Fig. 1B**). Virulent clusters correspond

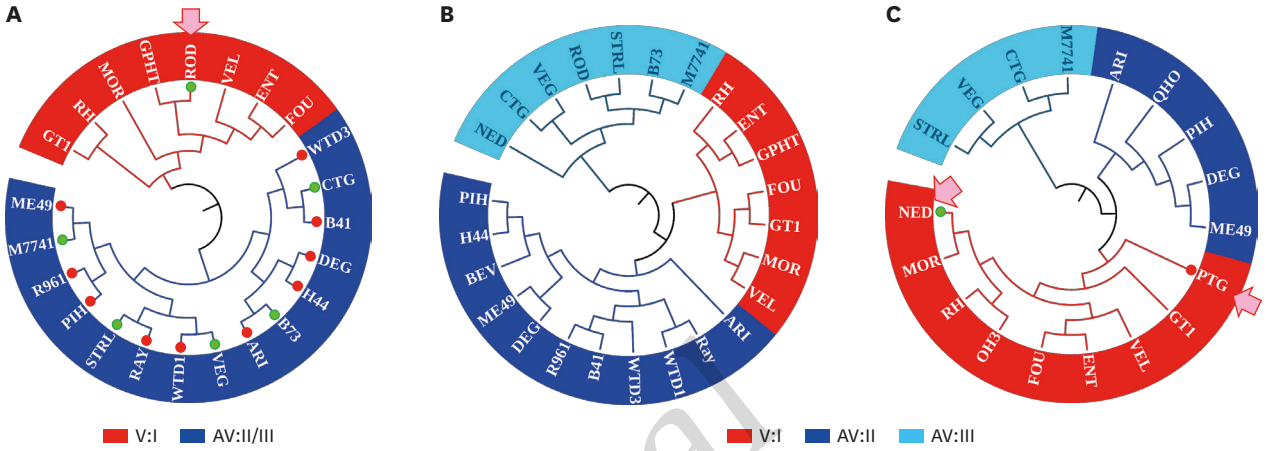

Fig. 1. Clustering of archetypal *Toxoplasma gondii* isolates as constructed by the UPGMA method. (A) *sag1* CDS from 23 *T. gondii* isolates consist of genotypes I, II and III. (B) *gra7* CDS from 25 *T. gondii* isolates consist of genotypes I, II and III. (C) *rop18* CDS from 18 *T. gondii* isolates consist of genotypes I, II and III (red dot = avirulent strain, genotype II; green dot = avirulent strain, genotype III; small arrow = misplacement based on its virulence).
V, virulent; AV, avirulent; I, genotype I; II, genotype II; III, genotype III.

to genotype I, while 2 avirulent clusters correspond to genotypes II and III. These results indirectly support the evidence that GRA7 relates to *T. gondii* virulence. The addition of several atypical strains in hierarchical cluster analysis using UPGMA led to the misplacement of some isolates (**Fig. 2B**). Two misplacements (6.9%) occurred, putting virulent strains CAST and RUB into avirulent clusters. One genotype (3.5%) misidentification, i.e., RUB strains that should be atypical (I/III), are grouped into genotype II. In general, *gra7* can predict *T. gondii* virulence with an error rate of about 0%–6.9%. Overall, *gra7* is better than *sag1* because it can also distinguish genotypes with a prediction error of about 0%–3.5%.

The *rop18* coding sequence has been used to classify archetypal isolates from the *T. gondii* population into three clusters with an 11.1% (2/18) error rate due to misplacement of NED (genotype III) and PTG (genotype II) into virulent clusters (**Fig. 1C**). The addition of several atypical strains in hierarchical cluster analysis using UPGMA did not change the error rate for

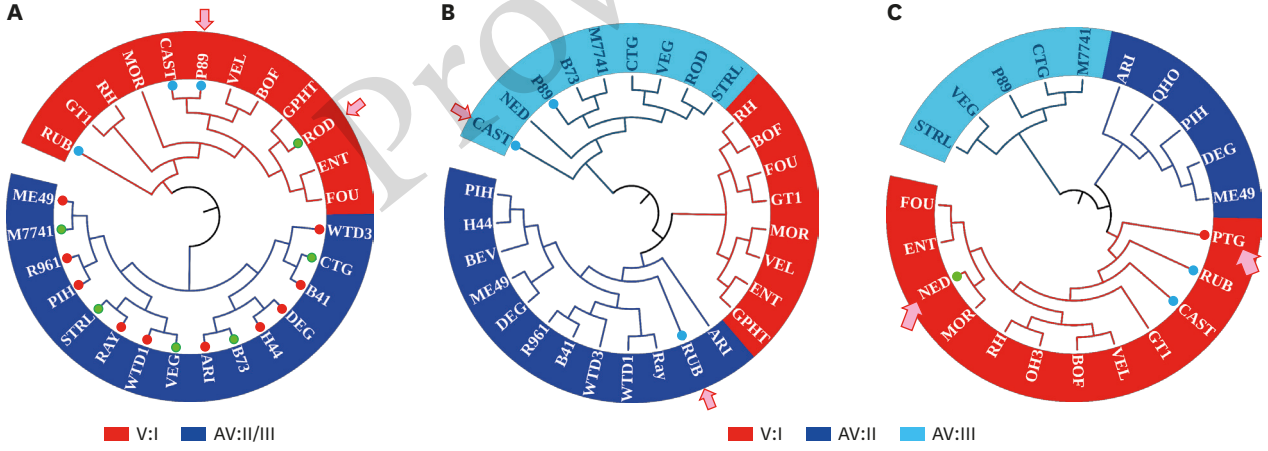

Fig. 2. Clustering of archetypal and non-archetypal *Toxoplasma gondii* isolates as constructed by the UPGMA method. (A) *sag1* CDS from 27 isolates of *T. gondii* consist of 23 archetypal and 4 atypical. (B) *gra7* CDS from 29 isolates of *T. gondii* consist of 25 archetypal and 4 atypical. (C) *rop18* CDS from 22 isolates of *T. gondii* consist of 18 archetypal and 4 atypical (red dot = avirulent strain, genotype II; green dot = avirulent strain, genotype III; blue dot = atypical strains and there is a dispute over virulence or genotype; small arrow = misplacement based on its virulence).
V, virulent; AV, avirulent; I, genotype I; II, genotype II; III, genotype III.

https://doi.org/10.4142/jvs.2021.22.e88

classifying virulence and genotyping of *T. gondii* (**Fig. 2C**). In general, *rop18* is able to distinguish the virulence and genotypes of *T. gondii* with a prediction error rate of about 11.1%.

DISCUSSION

There is still a dispute about the virulence and genotype of RUB, CAST, and P89 strains. RUB and CAST were classified as virulent strains according to an analysis based on CDS of *sag1* and *rop18* but were identified as avirulent strains based on CDS of *gra7*. Meanwhile, P89 has been identified as a virulent strain based on CDS of *sag1*, but *gra7* and *rop18* have been identified as avirulent strains.

The actual status of the P89 strain is still disputed by some researchers, some suggesting P89 is a virulent strain with genotype I [15], but other researchers claim that it is an intermediate virulent strain [10] with an atypical genotype [16]. The intermediate virulent strain is actually an avirulent strain with $LD_{50} \geq 10^3$ [3]. P89 was identified as genotype III based on multilocus sequence analysis under *sag2A* and *sag3* genetic markers [17]. This evidence supports several reports suggesting that P89 is atypical with the possibility of genetic recombination (type I/III). Therefore, classifying P89 as an avirulent strain and genotype III in this study was normally acceptable as it is in line with the analysis based on CDS of *gra7* and *rop18*.

RUB is categorized as a virulent strain [10] with an atypical genotype [16]. RUB can be grouped differently when analyzed using multilocus markers on a PCR-RFLP. RUB is grouped with genotype I when using *sag1* and *sag2* as locus markers [18]. Conversely, if using *sag4* and *bsr4* as locus markers, RUB will be in one cluster of genotype III [18]. Similarly, CAST is categorized as a virulent strain and genotype I [15] but other researchers consider its genotype atypical [6,16]. Based on a multilocus genotyping analysis using several genetic markers, CAST is an atypical genotype grouped into I/III recombinant isolate [6]. Therefore, in this study RUB and CAST were grouped into a virulent strain of genotype I as this result is also in line with the analysis based on CDS of *sag1* and *rop18*.

Based on the analysis and discussion above, two aspects can be considered limitations of this approach. First, this analysis requires the complete CDS of the gene. Second, prediction errors may occur if the *T. gondii* being analyzed is an atypical genotype. However, the frequency of these errors is quite low and is limited to genotype grouping. We consider that this is not a serious problem because it can be reduced by performing a combination analysis using *sag1*, *gra7* and *rop18*.

Overall, this approach has several advantages. First, it requires fewer genes, namely *sag1*, *gra7* and *rop18* when compared to PCR-RFLP, which requires multilocus genes, and the results are often inconclusive. Second, this approach is able to predict the virulence of *T. gondii* satisfactorily and it is easier to conclude if the analysis is based on a combination of both or all three genes. Third, this approach does not depend on live parasites as used in the mouse bioassay.

In conclusion, the average misplacement of *T. gondii* strains according to its virulence by *sag1, gra7* and *rop18* were 7.41%, 6.89% and 9.1%, respectively. Analysis based on CDS of *gra7* and *rop18* were able to categorize avirulent strains into genotypes II and III, whereas *sag1* failed to differentiate. Conclusions drawn from the analysis based on the CDS of both or all three genes will reduce prediction errors in classifying atypical *T. gondii*.

REFERENCES

1. Saadatnia G, Golkar M. A review on human toxoplasmosis. Scand J Infect Dis. 2012;44(11):805-814.
PUBMED | CROSSREF
2. Dardé ML. *Toxoplasma gondii*, "new" genotypes and virulence. Parasite. 2008;15(3):366-371.
PUBMED | CROSSREF
3. Simon S, de Thoisy B, Mercier A, Nacher M, Demar M. Virulence of atypical *Toxoplasma gondii* strains isolated in French Guiana in a murine model. Parasite. 2019;26:60.
PUBMED | CROSSREF
4. Rico-Torres CP, Vargas-Villavicencio JA, Correa D. Is *Toxoplasma gondii* type related to clinical outcome in human congenital infection? Systematic and critical review. Eur J Clin Microbiol Infect Dis. 2016;35(7):1079-1088.
PUBMED | CROSSREF
5. Zhou P, Zhang H, Lin RQ, Zhang DL, Song HQ, Su C, et al. Genetic characterization of *Toxoplasma gondii* isolates from China. Parasitol Int. 2009;58(2):193-195.
PUBMED | CROSSREF
6. Su C, Zhang X, Dubey JP. Genotyping of *Toxoplasma gondii* by multilocus PCR-RFLP markers: a high resolution and simple method for identification of parasites. Int J Parasitol. 2006;36(7):841-848.
PUBMED | CROSSREF
7. Cunha MM, Carneiro AC, Costa JG, Vitor RW. Genotyping of *Toxoplasma gondii* directly from human and animal biological samples: from partial genotypes to a new genotype. J Parasitol. 2016;102(1):157-160.
PUBMED | CROSSREF
8. Ajzenberg D, Cogné N, Paris L, Bessières MH, Thulliez P, Filisetti D, et al. Genotype of 86 *Toxoplasma gondii* isolates associated with human congenital toxoplasmosis, and correlation with clinical findings. J Infect Dis. 2002;186(5):684-689.
PUBMED | CROSSREF
9. Barragan A, Sibley LD. Migration of *Toxoplasma gondii* across biological barriers. Trends Microbiol. 2003;11(9):426-430.
PUBMED | CROSSREF
10. Dardé ML. Biodiversity in *Toxoplasma gondii*. Curr Top Microbiol Immunol. 1996;219:27-41.
PUBMED | CROSSREF
11. Arranz-Solís D, Cordeiro C, Young LH, Dardé ML, Commodaro AG, Grigg ME, et al. Serotyping of *Toxoplasma gondii* infection using peptide membrane arrays. Front Cell Infect Microbiol. 2019;9:408.
PUBMED | CROSSREF
12. Sibley LD, Boothroyd JC. Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. Nature. 1992;359(6390):82-85.
PUBMED | CROSSREF
13. Alaganan A, Fentress SJ, Tang K, Wang Q, Sibley LD. Toxoplasma GRA7 effector increases turnover of immunity-related GTPases and contributes to acute virulence in the mouse. Proc Natl Acad Sci U S A. 2014;111(3):1126-1131.
PUBMED | CROSSREF
14. Lei T, Wang H, Liu J, Nan H, Liu Q. ROP18 is a key factor responsible for virulence difference between *Toxoplasma gondii* and *Neospora caninum*. PLoS One. 2014;9(6):e99744.
PUBMED | CROSSREF
15. Sibley LD, Ajioka JW. Population structure of *Toxoplasma gondii*: clonal expansion driven by infrequent recombination and selective sweeps. Annu Rev Microbiol. 2008;62(1):329-351.
PUBMED | CROSSREF
16. Ajzenberg D, Collinet F, Mercier A, Vignoles P, Dardé ML. Genotyping of *Toxoplasma gondii* isolates with 15 microsatellite markers in a single multiplex PCR assay. J Clin Microbiol. 2010;48(12):4641-4645.
PUBMED | CROSSREF
17. Grigg ME, Suzuki Y. Sexual recombination and clonal evolution of virulence in *Toxoplasma*. Microbes Infect. 2003;5(7):685-690.
PUBMED | CROSSREF
18. Grigg ME, Bonnefoy S, Hehl AB, Suzuki Y, Boothroyd JC. Success and virulence in Toxoplasma as the result of sexual recombination between two distinct ancestries. Science. 2001;294(5540):161-165.
PUBMED | CROSSREF