Infection, Genetics and Evolution 65 (2018) 226–230

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

Genotyping of *Toxoplasma gondii*: The advantages of variable number tandem repeats within coding regions

Rosalia Moretta[a,b,*], Vanesa Roxana Sanchez[a,b], Ignacio Martín Fenoy[a,b], Alejandra Goldman[a,b], Paula Ruybal[c,1], Valentina Martin[a,b,1]

[a] Centro de Estudios en Salud y Medio Ambiente (CESyMA). Escuela de Ciencia y Tecnología. Universidad Nacional de San Martín. Av. Gral. Paz 5445, San Martin, Pcia de Bs. As 1650, Argentina

[b] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Godoy Cruz 2290 (C1425FQB) CABA, Argentina

[c] Universidad de Buenos Aires. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Instituto de Investigaciones en Microbiología y Parasitología Médica (IMPaM). Facultad de Medicina. Paraguay 2155 Piso: 12, CABA 1121, Argentina

ARTICLE INFO

Keywords:
Toxoplasmosis
Lineage
Strain
Epidemiology
VNTR

ABSTRACT

Toxoplasma gondii is an intracellular protozoan which is widely distributed. Infection occurs as a result of ingestion of uncooked meat and exposure to cat feces. Immunocompetent individuals are generally asymptomatic, while severe disease may occur in immunocompromised subjects and in congenital toxoplasmosis, which is caused by transplacental acquisition of *T. gondii*.

Genetic diversity of *T. gondii* has often been studied using a PCR-RFLP scheme based on nine molecular markers. These studies led to the description of a clonal population structure with three main lineages (I, II and III) in North America and Europe and higher genetic diversity in South America.

The aim of this study was to develop molecular markers that could allow the discrimination of genetic variants within each clonal lineage.

We analyzed the genome of *T. gondii* to identify genes containing variable number tandem repeats (VNTRs). The coding sequences of *T. gondii* ME49 genome were processed with Tandem Repeat Finder software. A panel of candidate markers was selected based on the following parameters: the repeat unit size (> 9 bp) and composition (to avoid single and dinucleotide runs), the number of copies (< 20), and the absence of introns within the repeat region.

The selected panel of eight molecular markers was analyzed in PRU and RH strains. As a first step, the variability of the sequence size allowed us to differentiate PRU from ME49 (two type II strains) and RH from GT1 (two type I strains). Additionally, amplification products from PRU and RH strains were sequenced to study intra-lineage variability. Aside from size polymorphisms in the amplification products we were able to identify sequence variability in polymorphic markers.

Toxoplasma gondii is an intracellular pathogen of the phylum Apicomplexa that can infect virtually all warm-blooded mammals (Robert-Gangneux, 2014). Infection occurs as a result of ingestion of tissue cyst forms present in meat products from infected animals or vegetables contaminated with oocysts released in the feces of infected cats, or through transplacental passage. In healthy individuals the infection is usually asymptomatic, but can cause severe complications in immunocompromised and in congenitally infected subjects (Robert-Gangneux and Dardé, 2012; Contini, 2008; Bojar and Szymańska, 2010).

Three main methods have been used for *T. gondii* genotyping. Restriction fragment length polymorphism (PCR-RFLP) led to the description of a clonal population structure with three main lineages (I, II and III) in North America and Europe (Su et al., 2006; Dubey,

* Corresponding author at: Edificio 23, INTI, Av. Gral. Paz 5445, San Martin, Pcia de Bs. As 1650, Argentina.
E-mail address: rmoretta@unsam.edu.ar (R. Moretta).

[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.meegid.2018.07.026
Received 21 October 2017; Received in revised form 20 July 2018; Accepted 21 July 2018
Available online 23 July 2018
1567-1348/ © 2018 Elsevier B.V. All rights reserved.

Table 1
Specific primers were designed for each gene selected, from T. gondii ME49 strain genome, as a molecular marker in this study. The optimum annealing temperature for each set of primers is shown (TM).

Gene (ME49 genome locus tag)	Molecular marker name	Primer	Sequence	TM
Cwf21 protein (TGME49_282140)	282140	140For	5´CAGAGCGACATGCGAACACAG 3´	59
		140Rev	5´ATCACCTCGATCATGTCCCCTG 3´	
Uncharacterized protein (TGME49_316650)	316650	650For	5´TGGGGACTCTGAGTCTGGAA 3´	62
		650Rev	5´GGAGTGTCTTTCATGTCTCCCG 3´	
RNA pseudouridine synthase superfamily protein (TGME49_202640)	202640	640For	5´AGATCACCGAGAGCACACAC 3´	59
		640Rev	5´CGCGAGTCCAAGGAACTCTC 3´	
Uncharacterized protein (TGME49_225830)	225830	830For	5´GGGGGCAATATGACCTCAGTGAC 3´	63
		830Rev	5´AGCTTCGGTGACAGTAGGC 3´	
Uncharacterized protein (TGME49_231200)	231200	1200For	5´ACACGCTCCAAAAAGACGCG 3´	62
		1200Rev	5´TTGATTTCCGCGACCTCTGC 3´	
Serine/threonine specific protein phosphatase (TGME49_223985)	223985	985For	5´GGTTCTGCACTTTCGGCGA 3´	63
		985Rev	5´TCTGTGCTCCTCTCTCCTCT 3´	
Folate-binding protein YgfZ protein (TGME49_267560)	267560	560For	5´TGAGCACTTCCGACCTTCTC 3´	59
		560Rev	5´AGAGAAAGATGGCCCTTGTC 3´	
Uncharacterized protein (TGME49_225090)	225090	5090For	5´ AGGCGAGAGACACTGGCAT 3´	63
		5090Rev	5´ATCTCTGTCTCTCAGTCGCGC 3´	

Table 2
Toxoplasma strains included in the study.

Strain	Type	Haplogroup	ToxoDB #	Year	Host	Geographical origin
ME49[a]	II	2	1	1965	sheep	USA
VEG[a]	III	3	2	1989	human	USA
PRU[a]	II	2	3	1963	human	France
ARI	12, atypical	12	5	1992	human	USA
FOU	Africa 1, atypical	6a	6	1988	human	France, imported from sub-Saharan, Africa
P89	BrIII, atypical	9	8	1991	pig	USA
RH [a]	I	1	10	1939	human	USA
GT1 [a]	I	1	10	1978	goat	USA
MAS	BrIV, atypical	4	17	1991	human	France, imported from South America
VAND	Amazonian, atypical	10	60	1997	human	French Guiana
RUB	Amazonian, atypical	5	98	1992	human	French Guiana
GAB2	Africa 1, atypical	14	203	2007	chicken	Gabon

GAB2: GAB2-2007-GAL-DOM2.
[a] Reference strain.

2009). On the other hand, atypical genotypes are often found in other continents where a higher genetic diversity describes a more complex population structure of this pathogen (Pena et al., 2008; Ajzenberg et al., 2002, 2009). Multilocus microsatellite markers have been applied to increase discriminatory power among strains and sequence-based analysis of intron region of housekeeping genes led to the description of 12 haplogroups in North America with local admixture (Ajzenberg et al., 2002; Khan et al., 2011; Lorenzi et al., 2016). Moreover, Su et al. (2012) combined PCR-RFLP and intron data (15 haplogroups) in a worldwide collection (956 isolates) describing a six clades population structure.

Genotyping has proven to be a powerful tool to asses *T. gondii* exposure risk associated with meat products and other routes of infection. In this line of work, Silva et al. (2014) were able to identify overlapping of *T. gondii* PCR-RFLP genotypes in animals and humans from the same geographical region. However, a very low genetic variability has been reported within each lineage (1% or less), making it difficult to develop simple, low-cost methods to differentiate intra-lineage strains (Su et al., 2003). To overcome this constraint, we propose a set of markers based on the description of variable number tandem repeats (VNTRs) within coding regions of the genome. This strategy holds two main advantages: i) variation arises from DNA polymerase slippage and does not depend on sexual replication of the parasite; ii) adjacent regions are usually conserved due to negative selective pressure.

The aim of this study was to develop molecular markers that allow the discrimination of genetic variants within each clonal lineage and/or haplogroup.

In order to identify genes containing VNTRs, the coding regions in the genome of ME49 (Type II) and GT1 (Type I) reference strains (http://toxodb.org/common/downloads/release-30/) were analyzed with tandem repeat finder (TRF) software (Benson, 1999). From the initial 970 coding sequences containing VNTRs, eight candidate markers were selected based on the following parameters: the repeat unit size (> 9 bp) and nucleotide composition (to avoid single and dinucleotide runs), the number of copies (< 20), and the absence of introns within the repeat region. Primers were designed for each VNTR flanking region (Table 1). To study intra-lineage polymorphisms, we obtained the sequences of RH type I and PRU type II strains to compare them with GT1 and ME49.

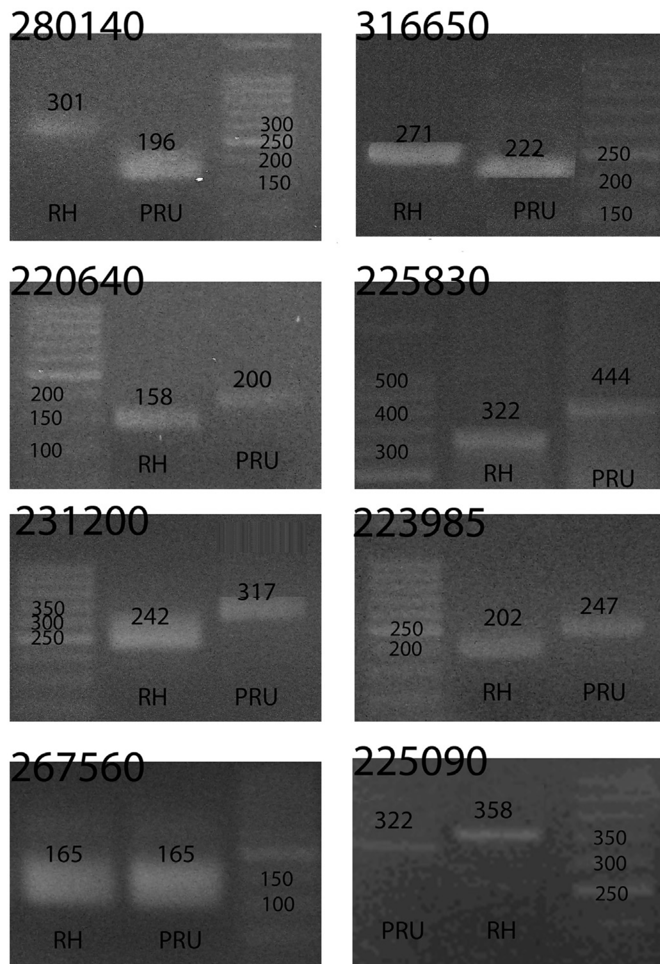

Fig. 1. Agarose gel electrophoresis of the amplification products obtained from PRU and RH strains. Product size was observed by comparison with the standard ladder 50 bp precision (Productos Bio-Lógicos, Buenos Aires. Argentina) and calculated from the sequences.

Variability of the repeat regions in other strains was searched by BLAST tool hosted by the Toxoplasma Genomics Research (http://toxodb.org/toxo/) and EXPASY (www.expasy.org) databases. The reference strains included in this analysis are shown in Table 2. The sequence of each repeat unit was codified to study the number and sequence of the repeats in the different strains (see below). Based on the frequency of each genetic variant, the discriminatory power (DP) of the molecular markers was estimated with the online DP Calculator tool provided by the University of the Basque Country (http://insilico.ehu.es/mini_tools/discriminatory_power/; Hunter, 1990).

RH and PRU tachyzoites were cultured in Vero cells as previously described (Martin et al., 1998). Following Vero cell lysis, *T. gondii* tachyzoites were harvested, centrifuged, washed twice with phosphate-buffered saline (PBS), and stored at −80 °C. To obtain DNA, pellets were resuspended in 20 µl of H_2O, incubated at 100 °C for 5 min and centrifuged at 10.000 ×*g* for 5 min. Polymerase chain reactions were performed with an initial 2 min denaturation step at 95 °C, followed by 35 cycles of 20 s at 95 °C, 30 s at the optimum annealing temperature of each set of primers (Table 1), 20 s at 72 °C and a final extension step at 72 °C for 10 min. The amplification products were visualized by 2% agarose gel electrophoresis and the product size was analyzed. Fig. 1 shows the amplification product of the selected VNTRs from RH and PRU strains with size polymorphisms in all cases except for 267560.

Both strands of the amplification product were sequenced on a Big Dye Terminator v3.1 kit and an ABI3730XL DNA Analyzer (Applied Biosystems, California, US). The nucleotide sequences were deposited in GenBank (MF320226-MF320241).

To improve the analysis of the markers variability, each repeat unit for each VNTR was codified with a letter (Fig. 2, Supplementary Table 1). This means that, for example, repeat unit A represents different sequences in different markers but the same sequence within each marker.

When comparing intra-lineage variation, size polymorphism was evident for 280140 and 316650 between ME49 and PRU strains, and 267560 between GT1 and RH strains (Fig. 2).

Polymorphic markers within type I (267560) and type II (280140 and 316650) strains presented specific arrangements of the VNTR pattern. With the exception of loci 280140 and 225830, markers that didn't present intra-lineage size polymorphisms were also conserved at the sequence level.

Moreover, most related VNTR patterns belonged to the same main lineage. Strains RH and GT1 (Type I/haplogroup I) varied only in markers 267560 and 280140 whereas strains ME49 and PRU (Type II/haplogroup 2) were polymorphic in three of eight VNTRs (280140, 316650 and 225830). The other eight strains varied in six or seven out of eight VNTRs showing high levels of polymorphisms between lineages, although intra-lineage variation could not be evaluated in these cases. It should be noted that combining all eight markers, we reached a maximum DP value (=1).

Finally, there are two main concerns that might be considered for future analysis with a larger set of data. First, size homoplasy should be taken into account when using fragment size variation such as VNTR analysis (Estoup and Angers, 1998). In fact, different marker variants were widespread distributed among different strains. Nonetheless, size homoplasy depends on mutation rate and divergence time, significantly decreasing its impact on intra-lineage follow-up studies (closely related strains). Second, *T. gondii* has a particularly complex life cycle and evolution history (Su et al., 2012; Lorenzi et al., 2016). Events such as recombination during the sexual stage of the parasite can introduce confusing phylogenetic signals. We observed that the set of markers selected in our study is highly incongruent, which makes it unsuitable for describing genetic population structure or evolutionary distances (Fig. 2). However, epidemiological studies rely on these incongruences to distinguish strains in situations such as outbreaks or when assessing *T. gondii* exposure risk associated with different infection sources.

In the present work we analyzed eight polymorphic coding regions of *T. gondii*. This panel of molecular markers displayed a high discriminatory power and was suitable for intra-lineage genotyping. We propose that, given a clinical case, an optimal typing algorithm could include intron data (haplogroup) followed by VNTR determination.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2018.07.026.

STRAIN	280140		316650		202640		225830	
	Repeat unit: 21 bp	(bp)	Repeat unit: 24 bp	(bp)	Repeat unit: 42 bp	(bp)	Repeat unit size: 24-27 bp	(bp)
ME49	ABCAADEFG	259	ABACDECDFGHI	366	AB	200	ABCDEEFFFG	403
PRU	AADEFG	196	ACDFGHI	271	AB	200	ABDDEHFFFI	403
FOU	DADEEFHIGG	280	ND	ND	C	158	CBDDDFG	322
GAB2	DADEEFHIGG	280	ABACDECDFGHI	366	C	158	ABCDEEFFG	376
ARI	AAAADEFG	238	ABACDECDFGHI	366	A	158	CBCDDFFG	349
VAND	DDJDADEEFHFGGG	364	ABAJKLAMNI	318	C	158	CBCCDEEFFFG	430
RUB	DKAADELFHFGGG	322	ABDJKOAJKDAMPI	414	D	158	ND	ND
MAS	DADEMEEFGG	280	ABDJKAAMPI	318	AE	200	ND	ND
VEG	DADEELFGGGGGG	343	ABACDECDFGHI	366	AE	200	CBCDEEFFG	376
P89	DDMMEEFHFHIGGG	364	QBDJDI	222	AE	200	BCCDEEDEG	376
RH	NDDELELFGGG	301	ABDARI	222	C	158	CBDDDFG	322
GT1	ODDELELFGGG	301	ABDARI	222	C	158	CBDDDFG	322
	15 repeat unit variants (11 SNPs) DP: 0.9848		18 repeat unit variants (14 SNPs) DP: 0.8727		5 repeat unit variants (6 SNPs) DP: 0.7879		9 repeat unit variants (12 SNPs) DP: 0.9333	

STRAIN	231200		223985		267560		225090	
	Repeat unit: 15 bp	(bp)	Repeat unit: 24 bp	(bp)	Repeat unit: 18bp	(bp)	Repeat unit: 18bp	(bp)
ME49	ABCDAABBE	317	ABBA	247	ABCCD	165	ABCDAEEF	322
PRU	ABCDAABBE	317	ABBA	247	ABCDD	165	ABCDAEEF	322
FOU	AFBE	242	AAAA	226	ACBAD	165	ABCCCDEEGG	358
GAB2	AFBE	242	ABBA	247	ACBAD	165	ABCDAEEF	322
ARI	ABCDAABE	302	AABBBA	295	ABCD	147	ND	ND
VAND	AGBH	242	AAAA	226	ABBEF	174	ABCCCDEEGG	358
RUB	AGDH	242	AAA	202	ABBEGF	192	ABCCCCDEEGG	376
MAS	ADIE	242	AAA	202	ABBD	147	ABHCCAEIGG	358
VEG	ABCDAABBE	317	AAAAA	250	HIF	138	ABCCCCDEEGG	376
P89	JGBGBK	272	AAA	202	ABBCAD	183	ABCCDEEGG	340
RH	AFBE	242	AAA	202	ACBJK	165	ABCCCDEEGG	358
GT1	AFBE	242	AAA	202	ACAD	147	ABCCCDEEGG	358
	11 repeat unit variants (6 SNPs) DP: 0.8636		2 repeat unit variants (1 SNP) DP: 0.7879		11 repeat unit variants (11 SNPs). DP: 0.9697		9 repeat unit variants (7 SNPs) DP: 0.8182	

Fig. 2. Variability of the VNTR pattern. Each sequence of each repeat unit was codified with a letter with independent coding for each VNTR. Sequence information for each repeat unit/VNTR is specified in the Supplementary Table 1. Band sizes are also specified according to the entire sequence spanning between primers (bp in the figure). ND: Not determined. DP: Discriminatory Power.

Acknowledgements

This work was supported by Bunge and Born Foundation and the National Council for Research and Technology (CONICET).

References

Ajzenberg, D., Cogné, N., Paris, L., Bessières, M.H., Thulliez, P., Filisetti, D., Pelloux, H., Marty, P., Dardé, M.L., 2002. Genotype of 86 *Toxoplasma gondii* isolates associated with human congenital toxoplasmosis, and correlation with clinical findings. J. Infect. Dis. 186 (5), 684–689.

Ajzenberg, D., Year, H., Marty, P., Paris, L., Dalle, F., Menotti, J., Aubert, D., Franck, J., Bessières, M.H., Quinio, D., Pelloux, H., Delhaes, L., Desbois, N., Thulliez, P., Robert-Gangneux, F., Kauffmann-Lacroix, C., Pujol, S., Rabodonirina, M., Bougnoux, M.E., Cuisenier, B., Duhamel, C., Duong, T.H., Filisetti, D., Flori, P., Gay-Andrieu, F., Pratlong, F., Nevez, G., Totet, A., Carme, B., Bonnabau, H., Dardé, M.L., Villena, I., 2009. Genotype of 88 *Toxoplasma gondii* isolates associated with toxoplasmosis in immunocompromised patients and correlation with clinical findings. J. Infect. Dis. 199 (8), 1155–1167.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27 (2), 573–580.

Bojar, I., Szymańska, J., 2010. Environmental exposure of pregnant women to infection with *Toxoplasma gondii*-state of the art. J. Ann. Agric. Environ. Med. 17 (2), 209–214.

Contini, C., 2008. Clinical and diagnostic management of toxoplasmosis in the immunocompromised patient. Parasitologia. 50, 45–50.

Dubey, J.P., 2009. Toxoplasmosis of Animals and Humans, Second Edition. CRC Press (336 p).

Estoup, A., Angers, B., 1998. Microsatellites and minisatellites for molecular ecology: Theoretical and empirical considerations. In: Carvalho, Gary R., Carvalho, G.R. (Eds.), Advances in Molecular Ecology - Page 64. North Atlantic Treaty Organization. Scientific Affairs Division. F OS Press.

Hunter, P., 1990. Reproducibility and indices of discriminatory power of microbial typing methods. J. Clin. Microbiol. 28, 1903–1905.

Khan, A., Dubey, J.P., Su, C., Ajioka, J.W., Rosenthal, B.M., Sibley, L.D., 2011. Genetic analyses of atypical *Toxoplasma gondii* strains reveal a fourth clonal lineage in North America. Int. J. Parasitol. 41 (6), 645–655.

Lorenzi, H., Khan, A., Behnke, M.S., Namasivayam, S., Swapna, L.S., Hadjithomas, M., Karamycheva, S., Pinney, D., Brunk, B.P., Ajioka, J.W., Ajzenberg, D., Boothroyd, J.C., Boyle, J.P., Dardé, M.L., Diaz-Miranda, M.A., Dubey, J.P., Fritz, H.M., Gennari, S.M., Gregory, B.D., Kim, K., Saeij, J.P., Su, C., White, M.W., Zhu, X.Q., Howe, D.K., Rosenthal, B.M., Grigg, M.E., Parkinson, J., Liu, L., Kissinger, J.C., Roos, D.S., Sibley, L.D., 2016. Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. Nat. Commun. 7, 10147.

Martin, V., Arcavi, M., Santillan, G., Amendoeira, M.R., De Souza Neves, E., Griemberg, G., Guarnera, E., Garberi, J.C., Angel, S.O., 1998. Detection of human toxoplasma-specific immunoglobulins a, M, and G with a recombinant *Toxoplasma gondii* rop2 protein. Clin. Diagn. Lab. Immunol. 5 (5), 627–631.

Pena, H.F., Gennari, S.M., Dubey, J.P., Su, C., 2008. Population structure and mouse-

virulence of *Toxoplasma gondii* in Brazil. Int. J. Parasitol. 38 (5), 561–569.
Robert-Gangneux, F., 2014. It is not only the cat that did it: how to prevent and treat congenital toxoplasmosis. Infection 68 (1), S125–S133.
Robert-Gangneux, F., Dardé, M.L., 2012. Epidemiology of and diagnostic strategies for toxoplasmosis. Clin. Microbiol. Rev. 25 (2), 264–296.
Silva, L.A., Andrade, R.O., Carneiro, A.C., Vitor, R.W., 2014. Overlapping *Toxoplasma gondii* genotypes circulating in domestic animals and humans in southeastern Brazil. PLoS One 9 (2), e90237.
Su, C., Evans, D., Cole, R.H., Kissinger, J.C., Ajioka, J.W., Sibley, L.D., 2003. Recent expansion of *Toxoplasma* through enhanced oral transmission. Science 299, 414–416.
Su, C., Zhang, X., Dubey, J.P., 2006. Genotyping of *Toxoplasma gondii* by multilocus PCR-RFLP markers: a high resolution and simple method for identification of parasites. Int. J. Parasitol. 36 (7), 841–848.
Su, C., Khan, A., Zhou, P., Majumdar, D., Ajzenberg, D., Dardé, M.L., Zhu, X.Q., Ajioka, J.W., Rosenthal, B.M., Dubey, J.P., Sibley, L.D., 2012. Globally diverse *Toxoplasma gondii* isolates comprise six major clades originating from a small number of distinct ancestral lineages. Proc. Natl. Acad. Sci. U. S. A. 109 (15), 5844–5849.