

## NUCLEOTIDE SEQUENCE AND SEQUENCE COMPARISON OF THE REPEAT SEQUENCE 529 (R529) OF AN INDONESIAN *TOXOPLASMA GONDII* ISOLATE

Sri Hartati<sup>1</sup>, Asmarani Kusumawati<sup>2,3</sup> Hastari Wuryastuti<sup>1</sup>, Sri Widada,<sup>4\*</sup>

<sup>1</sup>Department of Internal Medicine, Faculty of Medicine Airlangga University

<sup>2</sup>Department of Reproduction, Faculty of Veterinary Medicine,

<sup>3</sup>Study Center of Biotechnology, University Gadjah Mada, Yogyakarta, Indonesia,

<sup>3</sup>UMR 5119, "Pathogens and immunity", CNRS, Université Montpellier II, 34095 Montpellier Cedex 5, France.

<sup>4</sup>e-mail: widada@univ-montp2.fr

\*corresponding author

### ABSTRACT

*Toxoplasmosis is one of the most common parasitic disease in humans and other warm-blooded animals. In many species of animals, severe disease may result in disastrous effects and important economic consequences. In humans, infections are usually asymptomatic but severe disease can occur in immunocompromized individuals and newborns. Owing to the importance of the disease, the availability of efficient and sensitive early diagnosis is of capital importance. Toxoplasma gondii genome contains repeat sequences and due to their multiple copy number they constitute ideal targets for genome-based detection methods. The repeat sequence 529 (R529) exhibits the highest copy number within the genome, ranging from 200 to 300 copies. To develop efficient R529-based diagnosis, it is important to establish its exact nucleotide sequence. We have isolated and amplified the R529 gene from an Indonesian isolate and determined its nucleotide sequence. Sequence comparison was done with two other R529 sequences. Variations were observed.*

**Keywords:** Diagnosis, Indonesian isolate, repeat sequence R529, sequence comparison, *Toxoplasma gondii*.

**Correspondence:** Sri Hartati, Department of Internal Medicine, Faculty of Medicine Airlangga University  
Jl. Prof Dr Moestopo 47, Surabaya 60131, ph. 62-31-5501617

### INTRODUCTION

Infection with *Toxoplasma gondii* is one of the most common parasitic infections in humans and other warm-blooded animals. It has been found worldwide in almost one third of the human population (Dubey & Beattie 1998, Tenter et al. 2000). *Toxoplasma gondii* is an obligate protozoan intracellular parasite that belongs to the phylum Apicomplexa, subclass coccidia, and constitutes one of the most successful protozoan parasites on Earth. There are three infectious stages of *T. gondii* for all hosts: tachyzoites (individually and in groups), bradyzoites (in tissue cysts) and sporozoites (in oocysts). Warm-blooded animals are intermediate hosts while felids are the sole definitive hosts. Parasite transmission occurs by a fecal-oral cycle, carnivorous or transplacentally (Dubey 2004, for a review; Wong & Remington 1994).

In many species of animals, severe disease includes embryonic death and resorption, fetal death and mummification, abortion, stillbirth and neonatal death. In healthy adult humans, infections are usually asymptomatic. Severe disease can occur in immunocompromized individuals and newborns (Dubey

2004, for a review; Dubey & Beattie 1988; Tenter et al. 2000). Encephalitis is the most clinically important manifestation of toxoplasmosis in immunosuppressed patients. This disease is the major cause of death among patients with AIDS. In congenital infections, the severity of the disease may depend upon the stage of pregnancy at the time of infection. Severe effects include miscarriage, retinochoroiditis, mental retardation, hydrocephalus, convulsions, and intracerebral calcification.

In order to take suitable measures, an early diagnosis of the disease is essential and the availability of efficient and sensitive diagnostic tools are of great importance, particularly in pregnant women. A certain number of genes are used as targets in *Toxoplasma* detection by in vitro amplifications (PCR). Among the commonly used gene-targets, repeat sequences are interesting in molecular diagnosis, owing to their multiple copy number within the parasite genome and their potentiality as target has been tested. They are the repeat B1 (Burg et al. 1989, Jalal et al. 2004), the repeat 529 (R529) (Homan et al. 2000; Reischl et al. 2003), the repeat Tg4 (Angel et al. 1991) and the interspersed repeat (IRE) (Echeverria et al. 2000).

Of these repeat sequences, the repeat R529 which is a non-coding sequence, exhibits the highest copy number as it has been evaluated to be 200-300 copies within the parasite genome. Two R529 nucleotide sequences have been referenced up to date in GeneBank data (AF146527, complete sequence; AF487550, partial sequence). Some differences are noticed between the two sequences. We described in this paper the nucleotide sequence of R529 of the Indonesian *T. gondii* isolate IS-1 and the comparison with the two R529 known sequences. This sequence comparison may help to design suitable primers for gene amplification.

## MATERIALS AND METHODS

The Indonesian isolate of *T. gondii*, called IS-1 for convenience in this paper, was isolated from the diaphragm of a goat at the slaughterhouse Cibadak at Sukabumi, West Jawa, Indonesia (Tolibin Iskandar 1998). The complete repeat sequence R529 of *T. gondii* contains 529 base pairs (Homan et al. 2000). The repeat sequence R529 of IS-1 was isolated and amplified from the genome by PCR using primers located at its each end (sense primer S1 and anti-sense primer C1) or by anti-sense primer C1 and internally located sense primer S2. They are as follows and their localization within the repeat sequence is indicated between brackets. S1: ctgcagggaggaagacgaaag (1-21); S2: ttacaggaagctgcctg (151-171); C1: ctgcagacacagtgcattctgg (529-508). The expected length of amplified fragments is 529 bp (couple S1/C1) and 379 bp (couple S2/C1).

The conditions of amplification were the following: prior denaturation at 94 for 2 mm; 30 cycles of denaturation at 94°C/30 sec, annealing at 55°C/30 sec, elongation at 72°C/1 mm; additional elongation at 72°C/5 mm. Reactions were carried out in a final volume of 25 µl, containing 25 moles of each primer and variable amounts of DNA template. Amplification products were analyzed by electrophoresis on 1% agarose gel.

The amplification products obtained with the primer set S2/C1 were then directly cloned in pCR2.1, using TOPO TA cloning system (Invitrogen Life Technologies, France) according to the instruction manual. The topoisomerase reaction mixture contained 2 µl of the amplification products and 1 µl TOPO vector in 6 µl volume, and was incubated for 5 mm at 22°C. Some of the resulting positive clones were analyzed by PCR using M13 universal primers, localized outside the insertion. Sequencing was carried out on R529 fragment cloned in pCR2.1 on the two strands (MilleGen, France). Comparison was done by manually aligning the R529

sequence of IS-1 with two R529 nucleotide sequences available in Gene Bank.

## RESULTS AND DISCUSSION

### Isolation-Amplification and Cloning of the Repeat Sequence R529 of IS-1

We have previously established, using a new SAG1-based methodology we proposed, that the Indonesian isolate IS-1 belongs to the RH strain (Sri Hartati et al. 2006). The R529 repeat sequence of the Indonesian isolate was isolated and amplified by PCR using the primer couple S1/C1 or S2/C1. The analysis of the amplification products is shown in figure 1.

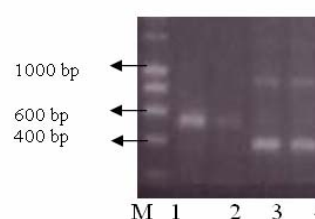


Figure 1. Amplification of R529 repeat sequence by PCR. Using primer couple S 1/C 1 (1, 2) or primer couple S2/C 1 (3, 4); molecular weight markers (M).

As expected, the amplification using S1/C1 primers gave rise to a fragment of the calculated length, i.e. 529 bp. (Figure 1, lanes 1,2). However by amplification using primers S2/C1, two fragments were obtained, respectively of the expected length, i.e. 379 bp and of a much longer length (about 900 bp) (Figure 1, lanes 3, 4). The amplified products obtained with primer couple S2/C1 were further cloned in pCR2.1. Five positive clones were analyzed by PCR using M13 universal primers and results are shown in figure 2.

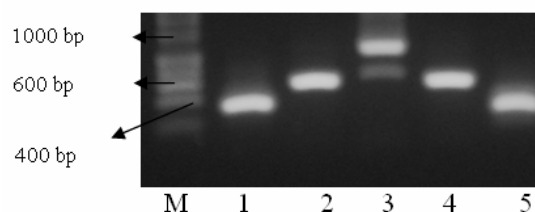


Figure 2. Analysis of positive pCR2.1 clones by PCR. Five independent clones (white colonies) were analyzed by PCR using universal M13 primers. Fragment 379 bp (2,5); fragment 529 bp (2,4);

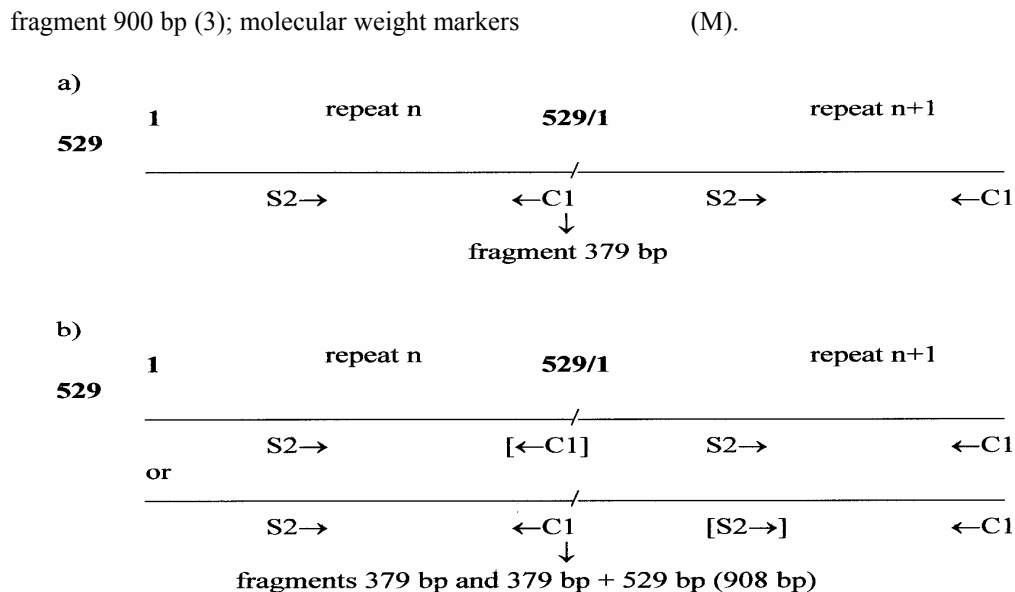


Figure 3. Two possible annealing of primers S2 and C1 with R529 gene during PCR amplification. Annealing within each repeat unit (1<sup>st</sup>-529<sup>th</sup>) (a), giving rise to a 379 bp DNA fragment and within each repeat unit as well as within a tandem repeat (b), resulting in two DNA fragments of respectively 379 bp and of ca 900 bp.

Three types of clone were obtained. Two clones contained a DNA fragment with the expected size, i.e. 379 bp; two contained a DNA fragment of the size of the complete repeat sequence, i.e. 529 bp and one clone contained a DNA fragment of about 900 bp. Taking into account the length of the regions comprised between the primers used in analysis by PCR and the inserted DNA (several tens of base pairs), the amplified fragments shown in figure 2 are so longer than they are actually.

The obtaining of a DNA fragment of ca 900 bp is interesting. This fragment might correspond to a combination of the full-length R529 sequence (529 bp) and the expected amplified fragment (379 bp) giving rise to a total length of 908 bp. This happened as if the primer S2 or C1 annealed to their target once within one repeat (possibility a) and once within a tandem repeat (possibility b) (Figure 3). In the last case, the amplified fragment contained the full-length R529 (529 bp) linked to the normally amplified DNA fragment (379 bp).

### Comparison of the Nucleotide Sequence of R529

The sequencing of the 908 bp DNA fragment was carried out on clone 3 of the pCR2. 1 construct. Results indicated that it actually contained the repeat sequence R529 and that the inserted DNA fragment resulted from the addition of a full-length 529 bp repeat and the expected 379 bp amplified fragment. The obtaining of the 908 bp DNA fragment may be due to a too high stringency during the annealing process so that either

primer S2 or primer C 1 did not anneal within each repeat unit. This may occur if there are sequence variations within the primers. Nevertheless, the sequencing of the 908 bp DNA fragment (in fact 902 bp fragment, see below) allowed us to confirm and to more clearly establish that R529 repeat sequence is organized as tandem repeats as previously reported using Southern blotting methodology (Homan et al. 2000).

Importantly the sequencing of the 908 bp fragment enabled us to determine the actual nucleotide sequence of the junction between repeats. In the publication reported by Homan et al. (2000), the sequence “ctgcag” is present at both ends of each repeat. In fact, by sequencing the junction between repeats we found that this sequence is only present once within the repeat, either at the 5'-end (upstream) or at the 3'-end (downstream) of each repeat. Therefore the length of the repeat is not 529 bp as previously established but 523 bp.

Besides, if we consider that the “ctgcag” sequence is only found at one of the repeat end, then the primer S1 or C1 contains an additional “ctgcag” (reducing so their T<sub>m</sub> values) which will not anneal with the template. This would result in incomplete annealing of primer C1 during the PCR process using primers S2 and C1, giving rise to the DNA fragment of 379 bp and an other DNA fragment of 908 bp. Using primer couple S2/C1, the amplification also gave rise to a fragment of ca 1450 bp which is a combination of two repeats of 529 bp each and the 379 bp fragment.

By examining the nucleotide sequence of R529 of the Indonesian isolate IS-1, which has been determined as belonging to RH strain (Sri Hartati et al. 2006), some differences were found when compared to those previously published which are both sequences from RH strain (Figure 4). It is worthwhile mentioning that there are also differences between these last sequences. This is in contrast to the high sequence conservation of SAG1 gene (Sri Hartati et al. 2006).

```

1  - -----
-----
2  1  ctgcagggag gaagacgaaa gttgtttttt tttttttt ttttttgtt
tttttgattt
3  1  agag gaagacgaaa gttgtttttt tttttttt ttttttgtt
tttttgattt

1  - -----
-----
2  61  ttgttttttt tgaactcgggc ccagctgcgt ctgtcgggat
3  55  ttgttttttt tgaactcgggc ccagctgcgt ctgtcgggat

1  1  gagaccgcgg agccgaagtgc cgttttcttt ttttgaattt tttttgttt
ttcacagggc
2  101 gagaccgcgg agccgaagtgc cgttttcttt ttttgaattt tttttgttt
ttcacagggc
3  95  gagaccgcgg agccgaagtgc cgttttcttt ttttgaattt tttttgttt
ttcacagggc

1  61  agctcgctgc tgcttggagc cacagaaggg acagaagtgc aaggggacta
cagacgcgat
2  161 agctcgctgc tgcttggagc cacagaaggg acagaagtgc aaggggacta
cagacgcgat
3  155 agctcgctgc tgcttggagc cacagaaggg acagaagtgc aaggggacta
cagacgcgat

1  121 gccgctcttc cagccgtctt ggaggagaga tatcaggact gtagatgaag
gcgaagggtga
2  221 gccgctcttc cagccgtctt ggaggagaga tatcaggact gtagatgaag
gcgaagggtga
3  215 gccgctcttc cagccgtctt ggaggagaga tatcaggact gtagatgaag
gcgaagggtga

1  181 ggatgagggg gtggcgtggt tgggaagcga cgagagtcgg agagggagaa
gatgtttccg
2  281 ggatgagggg gtggcgtggt tgggaagcga cgagagtcgg agagggagaa
gatgtttccg
3  275 ggatgagggg gtggcgtggt tgggaagcga cgagagtcgg agagggagaa
gatgtttccg

1  241 gcttggctgc ttttcttgga ggggtgaaaa -gagacaccg gaatgcgac
tagacgagac
2  341 gcttggctgc ttttcttgga ggggtgaaaa agagacaccg gaatgcgac
tagacgagac
3  335 gcttggctgc ttttcttgga ggggtgaaaa agagacaccg gaatgcgac
tagacgagac

1  300 gacgcttttc tcgtggtgat ggcggagaga attgaagagt ggagaagagg
gcgaggggaga
2  401 gacgcttttc tcgtggtgat ggcggagaga attgaagagt ggagaagagg
gcgaggggaga
3  395 gacgcttttc tcgtggtgat ggcggagaga attgaagagt ggagaagagg
gcgaggggaga

1  360 cagagtcgga ggcttgacg aaggaggag gagggtagg agagg-----
-----
2  461 cagagtcgga ggcttgacg aaggaggag gagggtagg agaggaaatcc
agatgcactg
3  455 cagagtcgga ggcttgacg aaggaggag gagggtagg agaggaaatcc
agatgcactg

1  --- -----
-----
2  521 tgtctgcag
3  515 tgtctgcag

```

Figure 4. Comparison of the nucleotide sequence of the repeat sequence R529. The three sequences considered belong to RH strain. A partial sequence, lacking sequences at both ends (1) (lacking sequences shown by -) (AF487550), a

full-length sequence (2) (AF146527), a full-length sequence (3) (Indonesian isolate IS-1). The nucleotide numbering is adopted according to their respective length. The oligonucleotide ctgcag is placed at the downstream-end of the repeat. Sequence variations are indicated by bold characters and are in italics.

This indicates that within the repeat sequence R529, more variations are found, most probably resulting from sequence evolution and/or from errors during the duplication process. We hypothesize that the sequence "ctgcag" is present once within a repeat, at one of its ends, contrarily to the previous report suggesting its presence at both ends (Homan et al. 2000). Curiously, the complementary sequence of 5-ctgcag- is 5-ctgcag-3'. It is then a palindromic sequence like that found in the recognition sites of the restriction enzymes. Its localization between repeats may play a role in repeat duplication.

## REFERENCES

- Angel, S.O., Blanco, J.C., Maero, E., Pzsenny, V. and Garberi, J.C. Repetitive DNA sequences of *Toxoplasma gondii* for development of diagnostic probes. Mem. Inst. Oswaldo Cruz, 86, 483-484.
- Burg, J.L., Grover, C.M., Pouletty, P. and Boothroyd, J.C. (1989). Direct and sensitive detection of a pathogenic protozo, *Toxoplasma gondii*, by polymerase chain reaction. I Clin. Microbiol., 27, 1787-1792.
- Dubey, J.P. (2004). Toxoplasmosis a waterborn zoonosis. Vet. Parasitol., 126, 57-72.
- Dubey, J.P. and Beattie, C.P. (1988). Toxoplasmosis of animals and man. CRC Press, Boca Raton, 220 pp.
- Echeverria, P.C., Rojas, P.A., Martin, V., Guarnera, E.A., Pzsenny, V. and Angel, S.O. Characterisation of a novel interspersed *Toxoplasma gondii* DNA repeat with potential uses for PCR diagnosis and PCR-RFLP analysis. FEMS Microbiol. Lett., 184, 23-27.
- Homan, W.L., Vercammen, M., De Braekeleer, J. and Verschueren, H. (2000). Identification of a 200- to 300-fold repetitive 529 bp DNA fragment in *Toxoplasma gondii*, and its use for diagnostic and quantitative PCR. mt. J. Parasitol., 30, 69-75.
- Jalal, S., Nord, C.E., Lappalainen, M. and Evengard, B. (2004). Rapid and sensitive diagnosis of *Toxoplasma gondii* infections by PCR. Clin. Microbiol. Infect., 10, 937-939.
- Reischl, U., Bretagne, S., Krueger, D., Ernault, P. and Costa, J.M. (2003). Comparison of two DNA targets for the diagnosis of Toxoplasmosis by real-time PCR using fluorescence resonance energy transfer hybridization probes. BMC Infect. Dis. 3, 7.

- Sri Hartati, Asmarani Kusumawati, Hastari Wuryastuti and Sri Widada, J. (2006). Primary structure of mature AG1 gene of an Indonesian *Toxoplasma gondii* and comparison with other strains. I Vet. Sci., 7, 163-170
- Tenter, A.M., Heckerroth, A.R. and Weiss, L.M. (2000). *Toxoplasma gondii*: from animals to humans. mt. I Parasitol., 30, 12 17-1258.
- Tolibin Iskandar (1998). The isolation of *Toxoplasma gondii* from diaphragma of a sheep having high antibody and faecal soil of cat. I ilmu Ternak dan Veteriner, 3, 111-116.