J Mol Evol (2004) 58:557–567
DOI: 10.1007/s00239-003-2578-3

JOURNAL OF MOLECULAR EVOLUTION

Structure Analysis of Two *Toxoplasma gondii* and *Neospora caninum* Satellite DNA Families and Evolution of Their Common Monomeric Sequence

Marina Clemente,[1] Natalia de Miguel,[2] Veronica V. Lia,[3] Mariana Matrajt,[4] Sergio O. Angel[2,5]

[1] INGEBI, Ciudad de Buenos Aires, Argentina
[2] Laboratorio de Parasitologia Molecular, UB2, IIB-INTECH, CONICET-UNSAM, Camino de Circunvalacion Laguna Km. 6, C.C 164, (B7130IIWA) Chascomus, Prov. Buenos Aires, Argentina
[3] Departamento de Ecologiá, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
[4] Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA
[5] Departamento de Parasitologiá, ANLIS Dr. Carlos G. Malbran, Ciudad de Buenos Aires, Argentina

Received: 1 September 2003 / Accepted: 26 November 2003

Abstract. A family of repetitive DNA elements of approximately 350 bp—Sat350—that are members of *Toxoplasma gondii* satellite DNA was further analyzed. Sequence analysis identified at least three distinct repeat types within this family, called types A, B, and C. B repeats were divided into the subtypes B1 and B2. A search for internal repetitions within this family permitted the identification of conserved regions and the design of PCR primers that amplify almost all these repetitive elements. These primers amplified the expected 350-bp repeats and a novel 680-bp repetitive element (Sat680) related to this family. Two additional tandemly repeated high-order structures corresponding to this satellite DNA family were found by searching the *Toxoplasma* genome database with these sequences. These studies were confirmed by sequence analysis and identified: (1) an arrangement of AB1CB2 350-bp repeats and (2) an arrangement of two 350-bp-like repeats, resulting in a 680-bp monomer. Sequence comparison and phylogenetic analysis indicated that both high-order structures may have originated from the same ancestral 350-bp repeat. PCR amplification, sequence analysis and Southern blot showed that similar high-order structures were also found in the *Toxoplasma*-sister taxon *Neospora caninum*. The *Toxoplasma* genome database (http://ToxoDB.org) permitted the assembly of a contig harboring Sat350 elements at one end and a long nonrepetitive DNA sequence flanking this satellite DNA. The region bordering the Sat350 repeats contained two differentially expressed sequence-related regions and interstitial telomeric sequences.

Key words: *Toxoplasma gondii* — *Neospora caninum* — Tandem repeat — Satellite DNA — Heterochromatin — EST

Correspondence to: Dr. S.O. Angel; *email:* sangel83@hotmail.com

Introduction

A substantial portion of eukaryotic nuclear genomes consists of highly repetitive satellite DNA. Satellite DNA is a tandemly repeated DNA with repeat units ranging from 5 to more than 100 bp (Charlesworth et al. 1994). In condensed chromosomes, satellite DNA is located in large clusters mostly in heterochromatic regions, near centromeres and telomeres, and sometimes interstitially (Jones and Flavell 1982). In general, satellite DNA shows high homogeneity within a species and rapid divergence between species (Elder and Turner 1995). Models of concerted evolution generally invoke two molecular processes, unequal crossing-over and gene conversion (Nagylaki 1990; Schloterer and Tautz 1994). Interplay among these and other variables (slippage replication, dif-

ferential replication, and rolling circle replication) leads to a continuum in the degree of heterogeneity exhibited by repeats within a species: when concerted evolution is strong, repeats are almost or completely identical, and when weak, sequence heterogeneity is observed (Cronn et al. 1996).

Along with other elements, such as interspersed repetitive DNA, satellite DNA is an important source of genome plasticity and chromosome structure variation (Wichman et al. 1991) that can contribute to speciation processes. Since satellite DNA consists of noncoding sequences, its biological function is restricted to the presence of protein binding motives (Goldeberg et al. 1996), the source of short heterochromatic RNAs (shRNAs) that regulate heterochromatin formation as observed in *Schizosaccharomyces pombe* (Reinhart and Bartel 2002; Volpe et al. 2002), and other structural features (Koch 2000). It is assumed that the putative function(s) and turnover of satellite DNA affects both genome structure and organism evolution (Rose and Doolittle 1983). A description of the structural features of satellite DNA therefore becomes important to our understanding of genome dynamics.

The protozoan parasite *Toxoplasma gondii* is an important and widespread human and veterinary pathogen that can infect virtually any nucleated animal cell. *T. gondii* shows an extreme clonal population structure, restricted to three distinct clonal lines that are 98% identical at many loci (Howe and Sibley 1995; Dard 1996; Grigg et al. 2001a; Ajzenberg et al. 2002). Another remarkable feature of this protozoan is intertype allelic dimorphism. The presence of only two allelic classes suggests that the currently predominant genotypes constitute successful recombinants that arose from the genetic mixing of two discrete ancestral lines (Grigg et al. 2001b; Su et al. 2003). In addition to the interstrain genetic similarities, the *T. gondii* genome does not show substantial length variation among chromosomes of different parasite isolates (Sibley and Boothroyd 1992).

Different families of repetitive DNA have been described in *T. gondii:* ABGTg (or TGR) repeats of about 350 bp (Cristina et al. 1991; Blanco et al. 1992), mitochondrial-like sequences (Osorio et al. 1991; Terry et al. 2001), a 529-bp tandem repeat (Homan et al. 2000), TgIRE (Echeverria et al. 2000), and minisatellite DNA (Ajzenberg et al. 2002). ABGTg and TGR sequences are components of the same satellite DNA family, with approximately 800 copies per haploid genome (Cristina et al. 1991; Blanco et al. 1992, Matrajt et al. 1999). TGR/ABGTg elements are located near the telomeres but also along other regions in the largest *Toxoplasma* chromosomes (Matrajt et al. 1999). Recent phylogenetic analysis of a large number of sequences related to TGR1E, a TGR/ABGTg element, from different *T. gondii* isolates revealed that each copy is different, although they exhibit high levels of identity (Hodgall et al. 2000). When genomic DNA from different *T. gondii* strains was digested with *Sal*I, Southern blot analysis using TGR- and ABGTg-derived probes showed the typical ladder pattern of tandemly repeated elements, the smaller *Sal*I fragment being about 350 bp (Cristina et al. 1995; Matrajt et al. 1999). However, sequence analysis of different TGR/ABGTg elements showed that most had *Sal*I fragments of around 350 bp flanked at both sides by other 350-bp repeats with little or no similarity (Cristina et al. 1991; Matrajt et al. 1999).

In this report, we further analyzed the structure of the TGR/ABGTg repetitive elements (here renamed Sat350). We also cloned a novel element of 680 bp (Sat680) which is related to this family and analyzed the phylogenetic relationships and structural features of Sat350 and Sat680 repeats, taking advantage of the recently developed *Toxoplasma* genome database (http://toxodb.org/). The presence of Sat350 and Sat680 elements in the *Toxoplasma*-sister taxon *Neospora caninum* (Su et al. 2003), was also assessed.

Materials and Methods

Parasites

T. gondii of the RH strain was cultured in vitro in PTP monolayers, a human foreskin fibroblast cell line, established by Servicio de Cultivo de Tejidos (INEI, ANLIS Dr. Carlos G. Malbran), with Eagle's minimum essential medium (Gibco) containing 1% fetal calf serum (Gibco). Tachyzoites were purified from infected monolayers by filtration through 3–μm-pore size polycarbonate filters (Nuclepore). Me49 brain tissue cysts were obtained from C57BL/6 mice and homogenized in phosphate-buffered saline (PBS) by syringe passages. *N. caninum* was obtained from Dr. Mariana Matrajt (University of Pennsylvania, Philadelphia). Growth and purification of *N. caninum* tachyzoites were conducted as described for *T. gondii* RH strain.

DNA Sample Preparation

Purified tachyzoites or homogenized brain containing cysts were pelleted and resuspended in TES buffer (Tris HCl [10 m*M*], EDTA [150 m*M*], NaCl [150 m*M*] [pH 7.4]) plus 1% sodium dodecyl sulfate (SDS) and 200 μg of proteinase K per ml. The samples were incubated for 2 h at 56°C. The DNA was extracted with equal volumes of phenol, chloroform–isoamyl alcohol (24:1), and ether and precipitated with 2 vol of ethanol. Finally, the DNA was centrifuged and resuspended in bidistilled water. The DNA concentration was determined at 260 nm in a Shimadzu PR-1 spectrophotometer as described (Sambrook et al. 1989).

Amplification and Cloning of Repeat DNA Elements

Amplifications were carried out in a Perkin Elmer thermal cycler (GeneAmp 9600) with 2.5 U of Taq DNA polymerase (Gibco BRL) with 1× BRL buffer plus 2.5 m*M* $MgCl_2$, 0.1 m*M* (each) deoxynucleotide triphosphate, and a 1 μ*M* concentration of each oligonucleotide, in a final volume of 50 μl. Based on the alignment of different 350-bp TGR repeat sequences, four primers positioned near or including *Sal*I or *Sal*I-like sites were synthesized: 350F1, 5-

TGCGTTTGGCGGCGCCACA; 350F2, 5-TCGACGTGC-TGGGTGTGGCG; 350R1, 5-CACGTCGACAATTCGTG-CCA; and 350R2, 5-TCGACAATTCGTGCCGAGT. Based on the alignment of A-and C-type sequences four primers were synthesized: AF, 5-ACGTGATGGAGGTTACGTGTTGAG; AR, 5-TCTCTGCATGGCCTGACGTCCGA; CF, 5-GGTA-TCCTTCAGAGAGTGAGTGTC; and CR, 5-AGGATACC-GACAACACGTCATCGA. Samples were amplified for 35 cycles as follows: 1 min at 94°C, 45 s at 52–55°C, and 1 min 30 s at 72°C. An initial step of 5 min at 94°C and a final step of 10 min at 72°C were included. The PCR products were run on a 1.5% agarose gel, and the results were visualized by ethidium bromide staining. Fragments were recovered from agarose gels by Qiaex II (Qiagene), quantified by comparison with DNA standards (Gibco BRL) by UV fluorescence, and cloned in pGEM T-vector (Promega). Recombinant plasmids were transformed into competent DH5α *Escherichia coli* cells.

DNA Sequencing

Insert nucleotide sequences were determined by automatic DNA sequencing employing the ABI PRISM Big DyeTM Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer Applied Biosystems). Both strands of each insert were sequenced using T7 and Sp6 promoter-derived primers. Gel electrophoresis was carried out on an ABI 377 Automated DNA Sequencer (Perkin Elmer). Sequence data were deposited in the GenBank database (see TG3RH, TgSat350-1 to -6, NcSat350-1 to -4, TgSat680-1 to -2, and NcSat680-1 in Table 1).

Sequence and Phylogenetic Analysis

To perform sequence and phylogenetic analysis of repetitive units, the previously described Sat350 elements—phage *a*, ABGTg7, TGR2, and TGR4 (Cristina et al. 1991; Matrajt et al. 1999)—and the Sat350 and Sat680 sequences obtained herein (Table 1) were split in their putative 350-bp repeat units. TGR1A, TGR1E, and TG3RH full-length sequences were used as single 350-bp repeat units.

Database searches and sequence comparisons were performed using Blastn, Blastx, and Blast 2 programs (www.ncbi.nlm.nih.gov/Blast). Preliminary *Toxoplasma* genomic (TGG_) and/or cDNA (TgESTzy) sequence data were accessed via http://ToxoDB.org and/or http://www.tigr.org/tdb/t_gondii/. Other apicomplexan genomic database used were: www.plasmodb.org, www.nbi.nlm.nih.gov/projects/malaria/blastindex.html, www.parvum.mic.vcu.edu, and www.sanger.ac.uk. EST databases were also searched using www.ncbi.nlm.nih.gov/Blast. Contig assembly and multiple alignment of DNA sequences were accomplished using CAP and CLUSTALX (Jeanmougin et al. 1998) software programs as implemented in Bioedit.

Parsimony analysis was performed using the heuristic search method provided by Swofford's win-PAUP 4b-10 under the Fitch criterion with 100 random addition sequences and tree bisection–reconnection (TBR) as branch swapping algorithm. Gaps were treated as fifth base. Neighbor-joining trees were constructed with win-PAUP4b-10 based on a pairwise comparison of two sequences according to Kimura's (1980) two-parameter model. Internal support was measured using 100 replicates of the heuristic search bootstrap option (Felsenstein 1985) available with PAUP4b.10, Phylogenetic inferences had to be restricted to network analysis without outgroup rooting due to the lack of suitable outgroups.

Southern Blot Analysis

Ten micrograms of genomic DNA from each species was digested with *Sal*I and blotted on Nylon Plus membranes (Pharmacia Biotech). Satellite DNAs labeled with ^{32}P-dCTP by a random primer labelling kit (Promega) were used as hybridization probes. Hybridization and washes were made with modifications as described (Matrajt et al. 1999). Briefly, hybridizations were performed overnight at 55°C and the final washes were made in a 0.1× SSC, 0.1% SDS solution for 20 min at 55°C. Membrane was analyzed by phosphoimager (phosphoimager Storm 820; Amersham) with the program image Quant5.2, and, also, exposed to autoradiographics.

Table 1. List of sequences used in this work

Name	Length (bp)	Accession no.
Sat350 family		
pha*,[1]	1866	AF022237
ABGTg7[1]	721	X74557
TGR1A[2]	352	M57916
TGR1E[2]	353	M57917
TGR2[2]	674	M57918
TGR4*,[2]	1032	M57919
TG3RH	353	AF534393
TgSat350-1	689	AY330208
TgSat350-2	736	AY330209
TgSat350-3	700	AY330210
TgSat350-4	1045	AY333784
TgSat350-5	1043	AY333785
TgSat350-6	650	AY333786
NcSat350-1	682	AY330203
NcSat350-2	730	AY330204
NcSat350-3	728	AY330205
NcSat350-4	381	AY330206
Sat680 family		
TgSat680-1	688	AF534391
TgSat680-2	687	AF534392
NcSat680-1	694	AY330207

Note. pha, phage *a*. *These sequences were used in their reverse complementary version.
[1] See Matrajt et al. (1990).
[2] See Cristina et al. (1991).

Results

Sequence Analysis of T. gondii *Sat350 Elements*

The phage *a* cluster (here as phaRC), previously shown to contain three 350-bp repetitive elements lacking significant identity to each other and the rest of phage *a*, was further analyzed for internal DNA repetitions. Three repetitive regions (RR) were detected within the cluster: RR I, RR II, and RR III (Fig. 1). Pairwise Blast analysis shows that the RR I has 87% similarity with RR II and 76% with RR III, whereas RR II has 83% similarity with RR III. Of note, these RRs included *Sal*I or *Sal*I-like sites which we used to delimit putative 350-bp repeat variants. The *T. gondii* Genome Project database revealed that phaRC is composed of repetitive DNA from position 1 to position 879 (data not shown). The portion of phaRC DNA between nucleotide 879 and the end of the sequence (nucleotide 1865) was therefore considered the flanking region (Fig. 1).

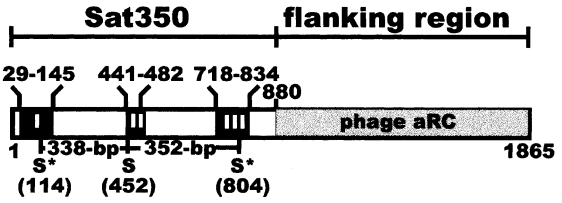

Fig. 1. Sequence analysis of phage *a* (phaRC). Numbers above or below the box indicate the nucleotide position. S and S*, *Sal*I and *Sal*I-like sites, respectively. Black boxes represent internal phage aRC repeated region motifs (I, II, and III).

Since the repetitive elements comprising this cluster may belong to a family of 350-bp repeats displaying high interunit variability (TgSat350 family), further sequence analysis was carried out. PhaRC, ABGTg7, and TGR sequences were divided into putative 350-bp repeats (Table 2), Pairwise Blast analysis showed that the phaRC2 repeat version displayed significant identity to TGR1A, TGR1E, Tg7.1, Tg7.2, TGR2.2, and TGR4RC2 sequences, whereas phaRC1 displayed significant identity to the TGR4RC3 sequence (Table 2). Considering a sequence similarity threshold higher than or equal to 70%, three 350-bp repeat types were identified: A, B, and C (Table 2). Within type B, monomers could be divided into at least two subtypes: B1, phaRC2, TGR1A, TGR1E, TgT7.1, and TgT7.2 and B2, TGR2.2 and TGR4RC2 (Table 2). When pairwise Blast2 analysis was conducted at low stringency, the similarity among variants of different types became apparent even despite its low value (Table 2). This suggests that all type variants belong to the same Sat350 DNA family. The intertype similarity values were below or equal to 62% (Table 2). Multiple sequence analysis showed that the most conserved region is the one surrounding *Sal*I or *Sal*I-like sites (data not shown). Of note, these conserved regions correspond to RR motifs in phaRC.

Cloning of a Novel 680-bp Tandem Repeat Related to the TgSat350 Family

As a result of aligning the sequences containing the *Sal*I or *Sal*I-like sites in TGR1A, TGR1E, TGR2 (.1 and .2), TGR4RC (1, 2, and 3), phaRC (1, 2, and 3), and ABGTg (.1 and .2), we were able to design primers which had the potential to amplify almost all of these repetitive elements. Four primers were synthesized: two sense primers (350F1 and 350F2) and two antisense primers (350R1 and 350R2). All primer combinations amplified an expected 350-bp fragment from DNA of both RH (type I) or Me49 (type II) strains. One of the primer combinations (350F1 and 350R2) amplified several fragments, the most prominent bands being 350 and 680 bp (Fig. 2A). Several 350-bp fragments were sequenced and analyzed. As expected, these sequences showed high identity to TGR1E or repeats of type B (e.g., Tg3RH; see Fig. 6). Two highly identical (78%) 680-bp fragments, TgSat680-1 and TgSat680-2, were also identified (Fig. 2B). To determine if TgSat680 elements were composed of two Sat350 repeats, both TgSat680-1 and -2 were split into two halves (Fig. 2B). While there was no significant similarity between halves, each shared low identity (61 and 64%) with the phaRC1 or TGR1E sequences, respectively (Fig. 2B). This level of identity suggests that new high-order 680-bp repeats were generated from the expansion of two ancestral 350-bp repetitive elements, giving rise to Sat680.

Southern blot analysis using *T. gondii* genomic DNA digested with *Sal*I showed that TgSat680 hybridizes with several bands, but predominantly with one of 680 bp (Fig. 3; *T. gondii*, lane 1). Interestingly, the hybridization pattern was not a multiple of 350 or 680 bp. In contrast, hybridization with an TgSat350-probe showed a neat-perfect ladder-profile multiple of 350 bp (Fig. 3; *T. gondii* lane 2).

Presence of Sat350 and Sat680 Satellite DNA in N. caninum

To determine if the TgSat350 and TgSat680-like structures are also present in the *N. caninum* genome, parasite DNA was amplified using the same primer set, and several products were obtained (Fig. 2A). Cloning and sequencing of 350- and 680-bp bands confirmed the existence of NcSat350 and NcSat680 elements in *N. caninum* (NcSat350-l and NcSat680-l). Southern blot with TgSat680 and TgSat350 probes only detected hybridization bands after long exposure and, also, showed a ladder-like pattern (Fig. 3; *N. caninum*, lanes 1 and 2). Identical patterns were observed using either NcSat350 or NcSat680 probes (data not shown).

Searching the Toxoplasma *Genome Project Database*

To define how the 350-bp repeats are arranged in the parasite genome, monomer sequences of types A, B, and C were employed to search genome sequences (TGG_) in ToxoDB: The *Toxoplasma* Genome Resource (http://ToxoDB.org, release 2.2). This Web site provides reads from *T. gondii* genome and EST sequencing projects. All repeat types retrieved the same 96 TGG_ sequences, but matched in different positions within each sequence. The main structures found are shown in Fig. 4 and indicate the organization of 350-bp repeats. Interestingly, A and C monomers are flanked by B monomers, in different sequential arrays, resulting in several AB1CB2 high-order structures (data not shown). TGR4RC, TGR2,

Table 2. Sequence comparison among 350-bp repeat variants

	Repeat					% identity* with repeat						
Type		PhaRC1	phaRC2	Tg7.1	Tg7.2	TGR1A	TGR1E	TGR2.1	TGR2.2	TGR4RC1	TGR4RC2	TGR4RC3
A	PhaRC1	—	59%	NSI	NSI	NSI	65%	NSI	62%	60%	NSI	95%
B	PhaRC2		—	71%	71%	79%	82%	NSI	71%	58%	74%	NSI
B	Tg7.1			—	85%	70%	80%	NSI	72%	NSI	69%	NSI
B	Tg7.2				—	71%	78%	NSI	71%	NSI	70%	NSI
B	TGR1A					—	84%	NSI	71%	NSI	76%	NSI
B	TGR1E						—	NSI	77%	58%	75%	NSI
C	TGR2.1							—	NSI	87%	NSI	NSI
B	TGR2.2								—	NSI	81%	NSI
C	TGR4RC1									—	NSI	60%
B	TGR4RC2										—	NSI
A	TGR4RC3											—

Note. 350-bp repeats: TGR1A and TGR1E, complete sequence; phaRC1—phaRC (114–457); phaRC2—phaRC (452–809); Tg7.1—ABGTg7 (21–360); Tg7.2—ABGTg7 (355–696); TGR2.1—TGR2 (1–333); TGR2.2—TGR2 (327–674); TGR4RC1—TGR4RC (1–343), TGR4RC2—TGR4RC (338–696); TGR4RC3—TGR4RC (691–1032). All these repeats harbor a *Sal*I or *Sal*I-like site at both ends. * Comparisons were made by pairwise Blast2 at a low stringency as follows: reward for a match, 2; penalty for a mismatchs, −1; open gap; 3, and extension gap; 1 penalty, gap × dropoff; 500, expect 10,000. NSI, no significant identity.

and phaRC sequences are concordant with this sequential structure (Fig. 4).

When searching the *Toxoplasma* genome database with TgSat680-1 and TgSat680-2 sequences, 65 TGG_ sequences containing the TgSat680 repetitive element were detected. Interestingly, the TGG lists obtained after searching with 680-bp elements were different from those obtained with A, B, or C related sequences (data not shown). The analysis of 680-bp monomers in short TGG fragments revealed a high level of homogeneity (Fig. 4). However, 680-bp monomer variants from different TGG fragments showed high interfragment variation (data not shown).

Importantly, analysis of the *Plasmodium* spp., *Cryptosporidium parvum*, *Theileria annulata*, and *Eimeria tenella* genomic and EST databases and the NCBI databases using Sat350 and Sat680 did not revealed any such repeat elements (data not shown).

Cloning of New Sat350 High-Order Structures in T. gondii *and* N. caninum

To confirm the presence of AB1CB2 high-order structures and variants, four primers were designed based on the alignment of 11 A-related and 22 B-related sequences obtained from the sequences described above (Table 2) and the TGG_ database. Taking into account a predominant AB1CB2 organization, AF/CR or CF/AR combinations amplified the expected main products, 730 and 647 bp, respectively (Figs. 4 and 5A). The use of AF/AR and CF/CR primer combinations showed a more complex pattern; a main PCR product of approximately 850 bp and several faint bands of slower migration profiles (Fig. 5A). After sequence analysis, the clones (Table 1) were aligned with the structures previously observed (Fig. 4).

Amplification of the *N. caninum* genome with these primer combinations produced several products with only AF/AR and AF/CR sets (Fig. 5B). However, only the products isolated from AF/CR PCR showed sequence similarity to Sat350 elements. Sequence analysis of the cloned products (Table 1) indicated that NcSat350 related elements are also organized as high-order ABC-like structures in *N. caninum* genome (Fig. 4).

Sequence Alignment and Phylogenetic Analysis

To gain a better understanding of the evolution and relationship among 350-bp repeat variants, a phylogenetic analysis was carried out (Fig. 6). Sat350 and Sat680 sequences were split to generate A, B, C, and 680-half sequences as shown in Fig. 4. Since the amplification products differ in length, only the first 254 bp of the repeat units were included in the alignment matrix in order to include as many sequences as possible in the phylogenetic analysis. Maximum parsimony analysis yielded 324 parsimonious trees consisting of 993 steps. The topology of the tree is in good agreement with the types A, B, and C previously defined by pairwise Blast analysis. Moreover, we could clearly distinguish two subtypes within the type B group: B1 and B2 (Fig. 6). This correlates with the organization of these repeats in the parasite genome, where B1 and B2 subtypes are organized as distinct 350-bp repeat variants (Table 2). The 680-bp elements halves do not cluster together in a monophyletic group. The clades (TgSat680-1.1, (TgSat680-2.1, NcSat680-1.1)) and (TgSat680-1.2, (TgSat680-2.2, NcSat680-1.2)) proved to be more closely related to A and C types, respectively, than to B type repeats (Fig. 6). Restricting phylogenetic analysis to those sequences encompassing the whole

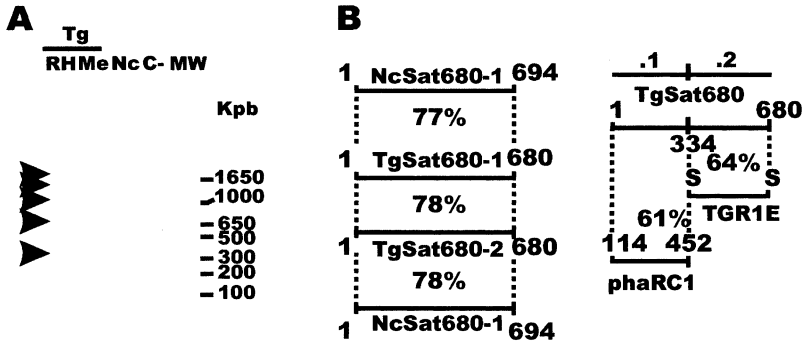

Fig. 2. Cloning of Sat680 element. **A** Genomic DNA from *T.gondii* RH (virulent, type I), Me49 (avirulent, type II) strains, and *N. caninum* (Nc) was amplified by PCR with 350F1 and 350R2 primers. The figure shows the PCR products electrophoresed in an agarose gel containing ethidium bromide. C–, control without DNA. Arrowheads on the left indicate bands observed in RH and Me49 (Me) lanes that are approximately 350 bp, or a multiple of 350 bp, long. Bands of 350- and 680-bp regions from RH, Me, and Nc were recovered, cloned in pGEM T easy (Promega) vector, sequenced, and deposited in the GenBank database. **B** Sequence analysis of 680-bp repetitive elements. Numbers indicate nucleotide positions. The 680-bp elements were split around position 334, giving the halves .1 and .2. Identity percentages were obtained by Blast2 analysis.

repeat (350 bp) retrieved the same groupings regarding A, B, C, and Sat680 halves with only minor differences within types (data not shown). In addition, the same branching pattern was obtained when analyzing the data with the neighbor-joining method (data not shown).

Interestingly, *N. caninum* repeats did not cluster together in a monophyletic group, although intended to function as outgroups. Instead, the different repeats became incorporated into the clades defined by types A, B, and C.

Ancestral nodes for the variants included in the analysis could not be determined due to the lack of a suitable outgroup.

Telomeric-like and Expressed Sequences Neighboring the pha Satellite DNA

Searching the *Toxoplasma* genome database with phaRC flanking region-derived sequence, a long DNA fragment could be constructed based on two TGG sequences. As expected, a long part of this sequence showed arrays of Sat350 elements, whereas the rest of the contig contained either single-copy DNA or telomeric-like repeats (Fig. 7). Searching the *Toxoplasma* EST database reveals two distinct regions in this contig that show high identity to *T. gondii* expressed sequences (Fig. 7). However, we could not find a large open reading frame or similarity to other DNA or protein sequences by Blast analysis, using both genomic or EST sequences.

Discussion

In this report, we analyzed the organization and evolution of *T. gondii* satellite DNA, which was described previously as a 350-bp element located near

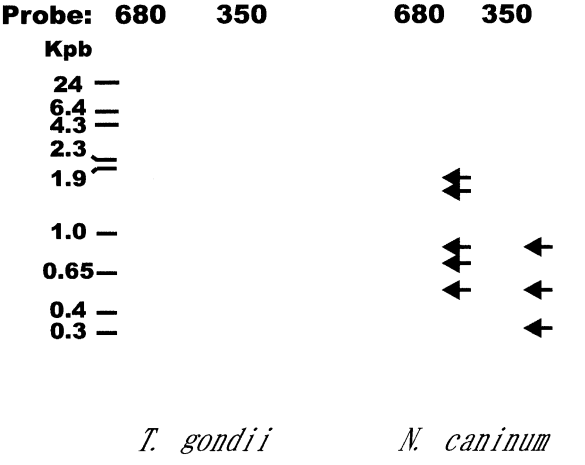

Fig. 3. Tandem organization of Sat350 and Sat680 elements. DNA from *T. gondii* and *N. caninum* were digested with *Sal*I, blotted on a nylon membrane, and hybridized with ^{32}P-TgSat350-3 and ^{32}P-TgSat680-2 probes. *T. gondii* bands of ~350 bp or multiples, are clearly observed with Sat350 probe. Arrows on the right of *N. caninum* Southern blot indicate detected bands. In the case of *T. gondii* DNA hybridization, membranes were exposed for 3 h in a phosphoimager, whereas in the case of *N. caninum* DNA hybridization membranes were exposed for 48 h to autoradiographies. 1 Kpb-plus and *λ-Hind*III were used as DNA standard (Invitrogen).

the telomeres and other regions of the largest chromosomes (Cristina et al. 1991, 1995; Matrajt et al. 1999). Our analysis detected two high-order structures, named Sat350 and Sat680. Interestingly, both were shown to originate from DNA elements within a 350-bp family. Sat350 revealed an internal organization of at least three different 350-bp repeat types (A, B, and C) and at least two B subtypes (B1 and B2) arranged mainly as AB1CB2. A, B, and C monomers are highly divergent (identity lower than 65%). The Sat680 family also showed a putative dimeric structure. The 680-bp dimers may have arisen by the same

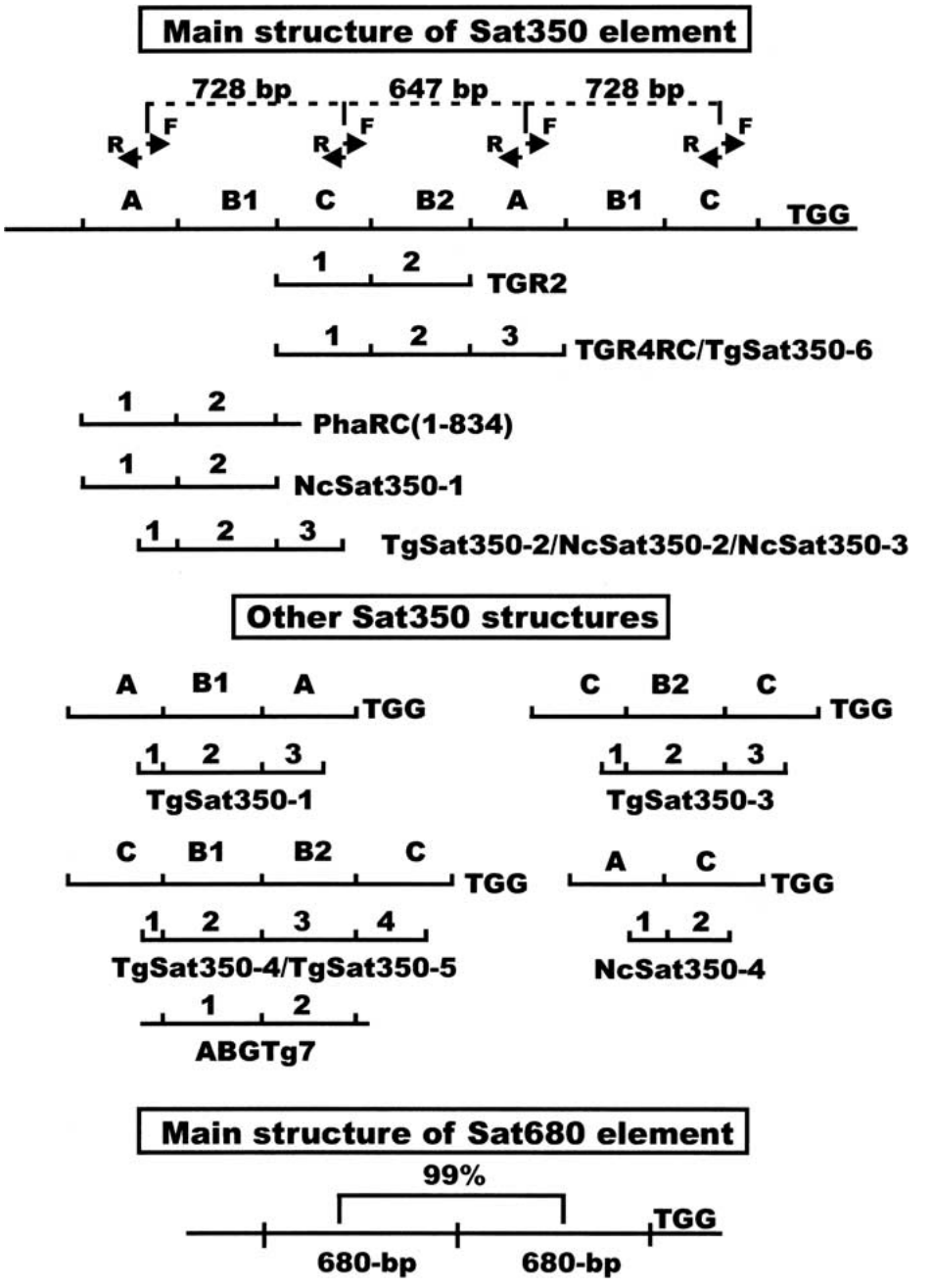

Fig. 4. Main arrangements of 350- and 680-bp repetitive elements. Different structures were deduced from analysis of sequences obtained from the searching of the *Toxoplasma* genome (TGG) database (http://ToxoDB.org). Searches were done using TGR1E, TGR4RC3, TGR4RC2, TGR4RC1, and 680bp1 sequences. Positions of AF, AR, CF, and CR primers and lengths of main expected products are indicated below the A–B1–C–B2 scheme. Arrangements obtained from cloning and PCR amplification are also shown. Numbers above the line indicate 350-bp subunits. Identity between contiguous 680-bp monomers is indicated.

process giving origin to the AB and CB structures that spread independently from the main ABCB structure of the Sat350 family.

These kinds of high-order structures are very similar to the α satellite DNA composed of 171-bp monomers which form a canonical pentameric repeat unit [ABCDE] (Waye et al. 1986) and Responder repeats in *Drosophila melanogaster* which have a dimeric structure (Cabot et al. 1993). Alphoid elements also presented variations of this canonical pentamer (Waye et al. 1986, 1987; Warburton et al. 1993). Noteworthily, alphoid A–E monomers had a 66 to 94% identity, whereas Responder dimers had an average difference of 16% between repeat halves. This is distinct from what occurs with Sat680-bp halves 1 and 2 and A, B, and C monomers, which had inter-

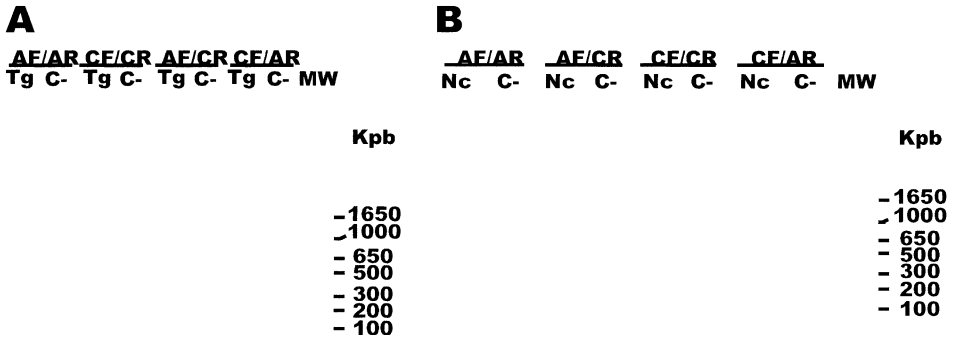

Fig. 5. PCR amplification and cloning of new 350-bp related structures. Genomic DNA from *T. gondii* (**A**) and *N. caninum* (**B**) was amplified by PCR with different combinations of AF, AR, CF, and CR primers. The figure shows the PCR products electrophoresed in an agarose gel containing ethidium bromide. C–, control without DNA. Arrows on the right indicate bands recovered from the gel, which were cloned in pGEM T easy (Promega) vector, sequenced, and deposited in the GenBank database.

type identities of less than 65%, indicating that both Sat350 and Sat680 elements are unusual types of satellite DNA.

As mentioned above, both Sat350 and Sat680 high-order structures of satellite DNA have probably originated from the same ancestral 350-bp repetitive element, generating an unexpectedly high intraspecific monomer heterogeneity. Sexual replication of *T. gondii* takes place only in cats, but the sexual cycle is not necessary for parasite proliferation. Interestingly, monomer homogeneization processes have been associated to meiotic recombination (Liao 1999; Schimenti 1999). Thus, the lack of intraspecific monomer homogeneity could be explained, at least in part, by the high rate of asexual expansion in *T. gondii* strains.

Interrepeat unit variation was also observed in other satellital DNAs, perhaps due to an accumulation of mutations (Phlol and Cornudella 1996; Castagnone-Sereno et al. 1998). In agreement with the model of satellite life history proposed recently by Nijman and Lenstra (2001), this high inter-350-bp repeat unit variability presumably produced A, B, C and the 680-bp halves from which high-order structures originated and expanded. Genomic and phylogenetic analysis indicated that AB1CB2 and 680-bp high-order structures arose after the divergence of 350-bp repeats. This is evidenced by the topology of the tree, where 680- and 350-bp repeat units are not clustered as independent lineages. Unequal crossing-over, gene duplication, and/or saltatory events likely contributed to the organization of 350-bp repeats in the 680- and 350-bp high-order structures of the TGR satellite DNA.

Although Sat350 and Sat680 high-order structures were not detected in apicomplexan species such as *Plasmodium* spp., *C. parvum*, *T. annulata* and *E. tenella* genome databases or general EST project databases, they were shown to be present in *T. gondii* and *N. caninum*. Since Southern blot analysis showed clear differences in copy number and arrays of the two kinds of satellital elements, these elements may not be important components of the *N. caninum* genome. However, satellite DNA is characterized by rapid evolutionary changes at the species level, often resulting in the occurrence of completely different satellite DNA families even within closely related sibling species (Henning et al. 1970). The detection of highly similar elements in both parasite species contributes to the "library hypothesis," which states that related species share conserved satellite sequences but show variability in the number of copies (Mestrovic et al. 1998). Interestingly, repeat units do not cluster in concordance with the species of origin, suggesting that repeat diversification and the formation of ABC and 680-bp high-order structures precede *T. gondii* and *N. caninum* divergence.

Sat350-related elements have been used to typify *T. gondii* strains (Cristina et al. 1995; Literak et al. 1998). In addition, phylogenetic trees based on sequences derived from the PCR products of natural parasite isolates showed that each isolate has its own unique TGR sequence, a Sat350-related element (Hodgall et al. 2000). They conclude that the TGR family is a promising target for typing individual *T. gondii* isolates and for studying the genetic distances between isolates. The high intermonomeric variability described here, together with the finding that *T. gondii* and *N. caninum* Sat350 and Sat680 sequences do not cluster following a species-specific pattern, however, strongly suggests that these elements will not be usable for parasite typification.

Satellite DNAs are noncoding sequences related to heterochromatic regions of low gene expression (Charlesworth et al. 1994). The presence of two distinct EST regions close to at least one side of one Sat350 cluster is interesting and further work will be required to determine the function, if any, of these genes. It is necessary to point out that these TGG and TgEST DNA sequences were obtained from the ge-

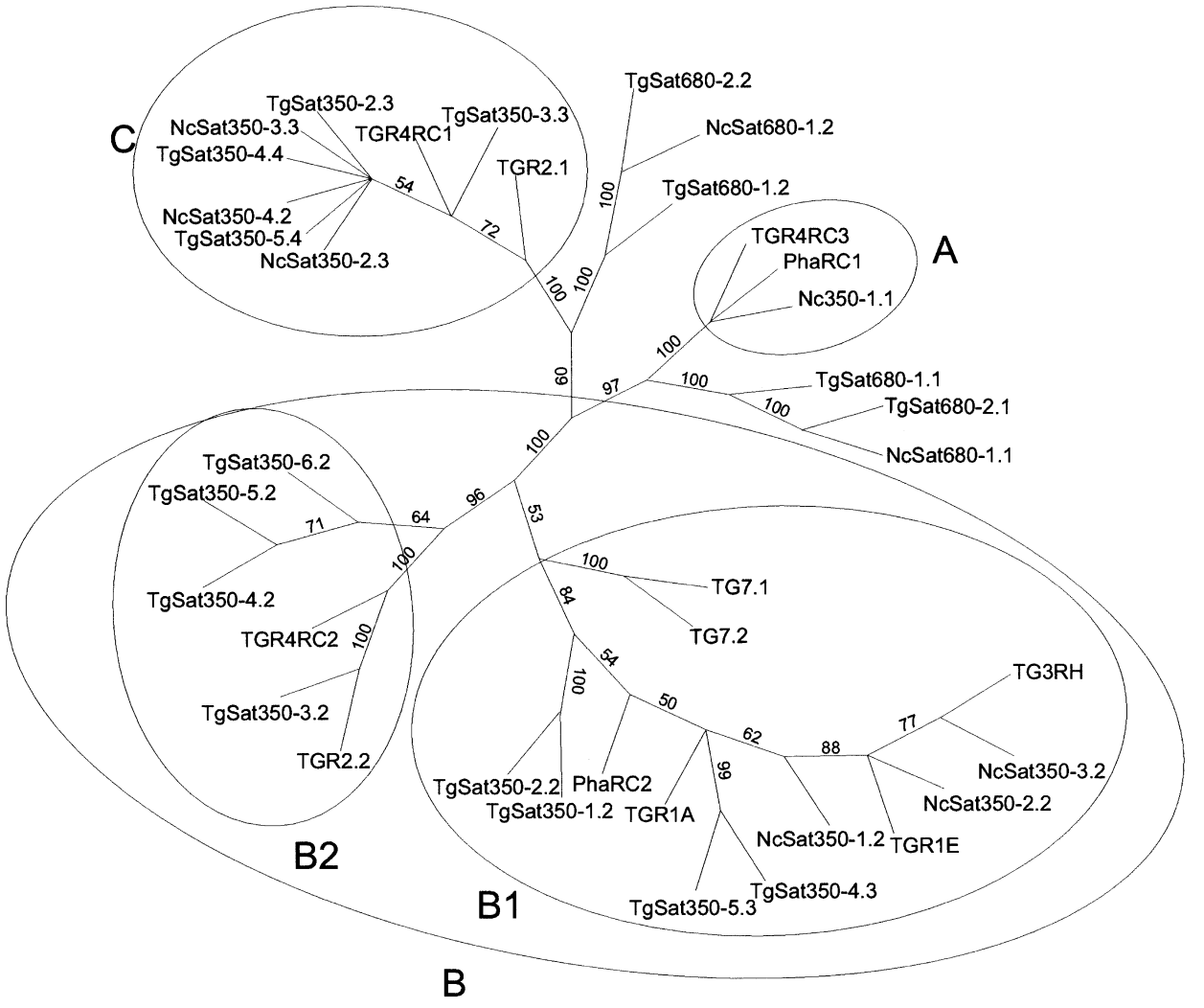

Fig. 6. Strict consensus tree from 324 most parsimonious trees obtained via heuristic search. Only the first 254 bp of the repeat units was aligned to generate the tree. A, B, and C types and B1 and B2 subtypes are enclosed in ellipses. Numbers above branches indicate bootstrap values.

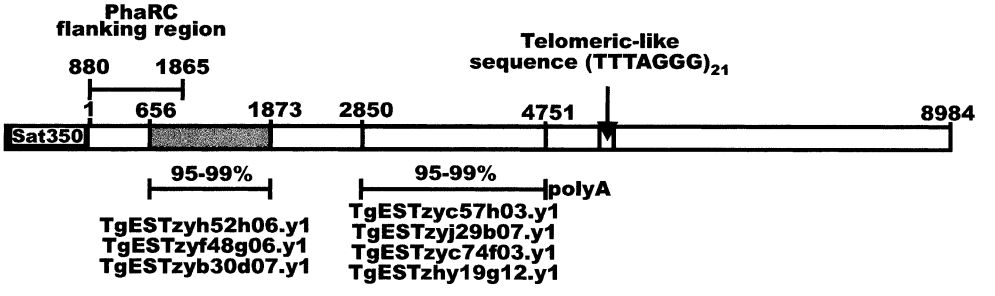

Fig. 7. Analysis of a large DNA fragment flanking the Sat350 DNA. The flanking region of phage aRC sequence (from 880 to 1865) was used to search the *Toxoplasma* genome database (http://ToxoDB.org) retrieving the sequences TGG_9124 and TGG_9197 (TGG release 2.2). By using these sequences a $phaRC_{880–1865}$/TGG_9124/TGG_9197 contig was obtained, expanding the flanking region to 8984 bp. Sat350 element is indicated upstream of nucleotide 1 of the contig. Numbers indicate nucleotide position. TgEST sequences are cDNAs from the *Toxoplasma* EST project. Percentages indicate identity.

nome project and must be corroborated by further cloning and sequencing. This Sat350 cluster was also found near a ''TTTAGGG'' telomeric-like repeat, which was shown to be flanked on both sides by large fragments of genomic DNA, indicating that this is an interstitial telomeric sequence, similar to that seen for *E. tenella* parasites (Shirley 1994). Interstitial telomeric sequences were also observed in centromeric

regions of chromosomes from different species (Rocco et al. 2001; Ruiz-Herrera et al. 2002; Andrades-Miranda et al. 2002), suggesting that TGR elements may also be present at the centromere of the largest *T. gondii* chromosomes. In addition, some of the interstitial telomeric sequences appear to be derived from rearrangements such as inversions or fusions that have taken place during karyotype evolution (Rocco et al. 2001; Ruiz-Herrera et al. 2002).

The most conserved region of sat350 and Sat680 elements, presented in Fig. 1 as RR motifs, is located surrounding the *Sal*I site. Recently, a 55-bp motif which is likely to have an insertional role was described in a repetitive element of *Zamia paucijuga* (Cafasso et al. 2003). In alphoid elements a 17-bp box, designated the CENP-B box, was described as a binding site of the centromere protein B (Muro et al. 1992). In addition, there are several proteins that bind simple DNA sequences at the euchromatic stage of DNA but bind satellite sequences when chromosomes become condensed (see ref. Csink and Henikoff 1998). It was proposed that the DNA sequence specificity of these proteins determines the sequence bias of satellites. Cells that amplify repeats that cannot be bound by a satellite binding protein are likely to show mitosis defects, limiting the repertoire of possible sequence variation. We question whether RR motif may function as binding site for any nuclear protein or whether it has a role as an insertional/amplification site.

In summary, we show here that Sat350 and Sat680 repeats form two high-order structures, both evolving from an ancestral 350-bp satellite DNA, which was also present in the *T. gondii*/*N. caninum* ancestral apicomplexan species. Our discovery of ESTs and telomeric-like sequences linked to a Sat350 satellite DNA will open new and intriguing questions about the role of this satellite DNA in the genome of *T. gondii*.

Acknowledgments. S.O. Angel is a member of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and of the Universidad de Buenos Aires (Departamento de Fisiología y Biología Molecular, FCEyN). V. Lia is fellow of the CONICET. This work was supported by ANPCyT Grant PICT 98 05-04831. Preliminary genomic and/or cDNA sequence data were accessed via http://ToxoDB.org and/or http://www.tigr.org/tdb/t_gondii/. Genomic data were provided by The Institute for Genomic Research (supported by NIH Grant AI05093) and by the Sanger Center (Welcome Trust). EST sequences were generated by Washington University (NIH Grant 1R01AI045806-01A1). The authors gratefully acknowledge Dr. Kendra Speirs for critical reading of the manuscript.

References

Ajzenberg D, Banuls AL, Tybayrenc M, Darde ML (2002) Microsatellite analysis of *Toxoplasma gondii* shows considerable polymorphism structured into two main clonal groups. Int J Parasitol 32:27–38

Andrades-Miranda J, Zanchin NI, Oliveira LF, Langguth AR, Mattevi MS (2002) (T_2AG_3)n telomeric sequence hybridization indicating centric fusion rearrangements in the karyotype of the rodent *Oryzomys subflavus*. Genetica 114:11–16

Blanco JC, Angel SO, Serpente P, Pszenny V, Maero E, Garberi JC (1992) Cloning of repetitive DNA sequences from *Toxoplasma gondii* and its usefulness for parasite detection. Am J Trop Med Hyg 46:350–357

Cabot EL, Doshi P, Wu M-L, Wu C-I (1993) Population genetics of tandem repeats in centromeric heterochromatin: unequal crossing over and chromosomal divergence at the responder locus of *Drosophila melanogaster*. Genetics 135:477–487

Cafasso D, Cozzolino S, De Luca P, Chinali G (2003) An unusual satellite DNA from *Zamia paucijuga* (Cycadales) characterised by two different organisations of the repetitive unit in the plant genome. Gene 311:71–79

Castagnone Sereno P, Leroy H, Semblat J-P, Leroy F, Abad P, Zijlstra C (1998) Unusual and strongly structured sequence variation in a complex satellite DNA family from the nematode *Meloidogyne chitwoodi*. J Mol Evol 46:225–233

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive in eukaryotes. Nature 371:215–220

Cristina N, Liaud MF, Santoro F, Oury B, Ambroise-Thomas PA (1991) A family of repeated DNAs sequences in *Toxoplasma gondii:* Cloning, sequence analysis and use in strain characterization. Exp Parasitol 73:73–81

Cristina N, Dard ML, Boudin C, Tavernier F, Pestre-Alexandre M, Ambroise-Thomas PA (1995) A DNA finger-printing method for individual characterization of *Toxoplasma gondii* strains: combination with isoenzymatic characters for determination of linkage groups. Parasitol Res 81:32–37

Cronn RC, Zhao X, Paterson AH, Wendel JF (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42:685–705

Csink AK, Henikoff S (1998) Something from nothing: The evolution and utility of satellite repeats. Trends Genet 14:200–204

Dard ML (1996) Biodiversity in *Toxoplasma gondii*. Curr Top Microbiol Immunol 219:27–41

Elder JF, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol 70:297–320

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the boostrap. Evolution 39:783–791

Echeverria PC, Rojas PA, Martin V, Guarnera EA, Pszenny V, Angel SO (2000) Characterization of an interspersed *Toxoplasma gondii* DNA repeat with potential uses for PCR diagnosis and PCR-RFLP analysis. FEMS Microbiol Lett 184:23–27

Goldeberg IG, Sawhney H, Pluta AF, Wartburton PE, Earnshaw WC (1996) Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: Implication for CENP-B function at centromeres. Mol Cell Biol 16:5156–5168

Grigg ME, Ganatra J, Boothroyd JC, Margolis TP (2001a) Unusual abundance of atypical strains associated with human ocular toxoplasmosis. J Infect Dis 184:633–639

Grigg ME, Bonnefoy S, Hehl AB, Suzuki Y, Boothroyd JC (2001b) Success and virulence in *Toxoplasma* as the result of sexual recombination between two distinct ancestries. Science 294:161–165

Henning W, Henning I, Stein H (1970) Repeated sequences in the DNA of Drosophila and their localization in giant chromosomes. Chromosoma 32:31-63

Hodgall E, Vuust J, Lind P, Petersen E (2000) Characterization of *Toxoplasma gondii* isolates using polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP)

of the noncoding *Toxoplasma gondii* (*TGR*)-gene sequences. Int J Parasitol 30:853–858

Homan WL, Vercammen M, De Braekeleer J, Verschueren H (2000) Identification of a 200–300-fold repetitive 529 bp DNA fragment in *Toxoplasma gondii*, and its use for diagnostic and quantitative PCR. Int J Parasitol 30:69–75

Howe DK, Sibley LD (1995) *Toxoplasma gondii* comprises three clonal lineages: correlation of parasite genotype with human disease. J Infect Dis 172:1561–1566

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. Trends Biochem Sci 23:403–405

Jones J, Flavell R (1982) The structure, amount and chromosomal localization of defined repeated DNA sequences in species of the genus *Secale*. Chromosoma 86:613–641

Kimura M (1980) A simple method for estimating evolutionary rates of base substitution through comparative studied of nucleotide sequences. J Mol Evol 16:111–120

Koch J (2000) Neocentromeres and alpha satellite: A proposed structural code for functional human centromere DNA. Hum Mol Genet 9:149–154

Liao D (1999) Concerted evolution: Molecular mechanism and biological implications. Am J Hum Genet 64:24–30

Literak I, Rychlik I, Svobodova V, Pospisil Z (1998) Restriction fragment length polymorphism and virulence of Czech *Toxoplasma gondii* strains. Int J Parasitol 28:1367–1374

Matrajt M, Angel SO, Pszenny V, Guarnera E, Roos DS, Garberi JC (1999) Arrays of repetitive DNA elements in the largest chromosomes of *Toxoplasma gondii*. Genome 42:265–269

Mestrovick N, Plohl M, Mravinac B, Ugarkovic D (1998) Evolution of satellite DNAs from the genus Palorus—Experimental evidence for the ''Library'' hypotesis. Mol Biol Evol 15:1062–1068

Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, Okazaki T (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. J Cell Biol 116:585–596

Nagylaki T (1990) Gene conversion, linkage, and the evolution of repeated genes dispersed among multiple chromosomes. Genetics 106:529–548

Nijman IJ, Lenstra JA (2001) Mutation and recombination in cattle satellite DNA: A feedback model for the evolution of satellite DNA repeats. J Mol Evol 52:361–371

Ossorio PN, Sibley LD, Boothroyd JC (1991) Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. J Mol Biol 222:525–536

Plohl M, Cornurdella L (1996) Characterization of a complex satellite DNA in the mollusc *Donax trunculus*: Analysis of sequence variations and divergence. Gene 169:157–164

Reinhart BJ, Bartel DP (2002) Small RNAs correspond to centromere heterochromatic repeats. Science 297:1831

Rocco L, Costagliola D, Stingo V (2001) (TTAGGG)n telomeric sequence in *Selachian* chromosomes. Heredity 87:583–588

Rose MR, Doolittle W (1983) Molecular biological mechanisms of speciation. Science 220:157–162

Ruiz-Herrera A, Garcia F, Azzalin C, Giulotto E, Egozcue J, Ponsa M, Garcia M (2002) Distribution of intrachromosomal telomeric sequences (ITS) on *Macaca fascicularis* (Primates) chromosomes and their implications for chromosome evolution. Hum Genet 110:578–586

Sambrook J, Fritsh EF, Maniatis T (1989) Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schimenti JC (1999) Mice and role of unequal recombination in gene-family evolution. Am J Hum Genet 64:24–30

Schlotterer C, Tautz D (1994) Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr Biol 4:777–783

Shirley MW (1994) The genome of *Eimeria tenella:* Further studies on its molecular organisation. Parasitol Res 80:366–373

Sibley LD, Boothroyd JC (1992) Construction of a molecular karyotype for *Toxoplasma gondii*. Mol Biochem Parasitol 51:291–300

Su CL, Evans D, Cole RH, Kissinger JC, Ajioka JW, Sibley LD (2003) Enhanced oral transmission acquired through recent recombination led to widespread expansion of *Toxoplasma*. Science 299:414–416

Terry RS, Smith JE, Duncanson P, Hide G (2001) MGE-PCR: A novel approach to the analysis of *Toxoplasma gondii* strain differentiation using mobile genetic elements. Int J Parasitol 31:155–161

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 meethylation by RNAi. Science 297:1833–1837

Warburton PE, Waye JS, Willard HF (1993) Nonrandom localization of recombination events in human alpha satellite repeat unit variants: Implications for higher-order structural characteristics within centromeric heterochromatin. Mol Cell Biol 13:6520–6529

Waye JS, Willard HF (1986) Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence fro evolution by unequal crossing-over and ancestral pentamer repeat shared with the human X chromosome. Mol Cell Biol 6:3156–3165

Waye JS, England SB, Willard HF (1987) Genomic organization of alpha satellite DNA on human chromosome 7: Evidence for two distinct alphoid domains on a single chromosome. Mol Cell Biol 7:349–356

Wichman HA, Payne CT, Ryder OA, Hamilton MJ, Maltibe M, Baker RJ (1991) Genomic distribution of heterochromatin sequences in equids: implications to rapid chromosomal evolution. J Hered 82:369–377