

# *Toxoplasma gondii* genomics: shedding light on pathogenesis and chemotherapy

James W. Ajioka, Jennifer M. Fitzpatrick and Christopher P. Reitter

The parasitic protozoan *Toxoplasma gondii* and its relatives (e.g. *Plasmodium* spp., which cause malaria) constitute a major global health problem. In recent years, the elucidation of biological processes in these parasites has accelerated with the application of genetic and genomic methodologies. Genetic analyses in *T. gondii* have revealed a remarkably stable 87 Mb (megabase) nuclear genome consisting of 11 chromosomes showing little variation across strains. Population studies demonstrate that the small amount of variation defines three clonal lineages where phenotypes such as virulence are associated with a single lineage. The random generation of cDNA (complementary DNA) sequences derived from different strains and developmental stages has highlighted the genetic differences underlying observed phenotypic variation. *T. gondii* also contains an extrachromosomal 35 kb (kilobase) circular DNA within an organelle with plastid-like properties. The limited coding capacity of the 35 kb organellar genome suggests that proteins responsible for organelle function(s) must be encoded by the nucleus. Scrutiny of *T. gondii* nuclear cDNA sequences has uncovered a number of proteins thought to carry out essential roles in the organelle (e.g. fatty acid biosynthesis). The synergism between these seemingly unrelated genomic studies has provided a greater understanding of parasite pathogenesis and has identified several new targets for chemotherapy.

James W. Ajioka (corresponding author)

University Lecturer, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK. Tel: +44 (0)1223 333 923/329; Fax: +44 (0)1223 333 346; E-mail: jwa@mole.bio.cam.ac.uk

Jennifer M. Fitzpatrick

PhD Student, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK. E-mail: jmf40@mole.bio.cam.ac.uk

Christopher P. Reitter

Research Assistant, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK. E-mail: cpr@mole.bio.cam.ac.uk

*Toxoplasma gondii* genomics: shedding light on pathogenesis and chemotherapy

Parasitic protozoa of the phylum Apicomplexa, including *Plasmodium* spp. (which cause malaria), *Eimeria* spp. (which cause avian coccidiosis) and *Toxoplasma gondii*, represent a major source of human and animal disease worldwide. *T. gondii* is the cause of significant morbidity and mortality in patients who have AIDS (acquired immunodeficiency syndrome) and serious congenital birth defects in both humans and livestock (Refs 1, 2). Although infection is usually asymptomatic in healthy individuals, immunocompromised patients often suffer from a life-threatening encephalitis (Ref. 2). Currently, the combination of pyrimethamine and sulphadiazine is the treatment of choice for noncongenital toxoplasmosis. Unfortunately, owing to the toxic side effects and general low efficacy of these and other known parasitocidal drugs, new chemotherapeutic agents are urgently required (Refs 2, 3).

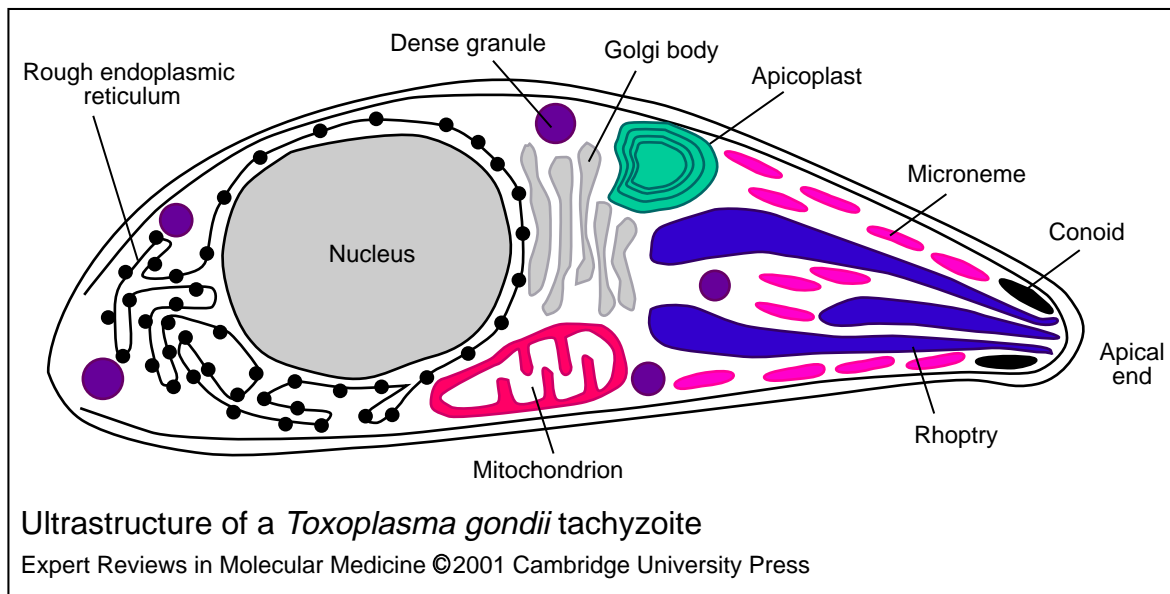
Genomic research on apicomplexan parasites, including genomic sequencing projects (for example, see Ref. 4), has generated a wealth of information that will undoubtedly lead to new therapies for these diseases. The genomes of these parasites, like those of other eukaryotic cells, are composed of both nuclear and organellar DNA. *T. gondii* contains a nuclear genome of about 87 Mb, a 6 kb mitochondrial genome, and an episomal 35 kb plastid-like genome (Ref. 5). *T. gondii* has emerged as the premier model apicomplexan because in vivo or in vitro isolation and/or culture of the asexual life-cycle stages is possible, and it has a well-characterised cell biology and established genetics (see for example Ref. 6). Not surprisingly, genomic investigations on *T. gondii* have both exploited and facilitated research in these areas, leading to some unexpected discoveries.

Population genetic analysis of nucleotide polymorphisms has divided the species into three closely related clonal lineages with corresponding phenotypic differences (Ref. 7). In stark contrast to the chromosomal complexities of other protozoan parasites, the congruence of karyotype and genetic linkage mapping defines exactly 11 nuclear chromosomes for the species (Refs 8, 9). Single-pass DNA sequencing of random cDNAs, also known as expressed sequence tags (ESTs; Ref. 10), from strain- and stage-specific cDNA libraries has generated a wealth of data for gene identification and the assessment of gene expression (Refs 11, 12). A novel use of

these sequences has been the identification and analysis of nucleus-encoded proteins of the plastid-like organelle found in many members of the Apicomplexa (Ref. 13). The discovery of this plastid-like organelle, now referred to as the apicoplast, followed the realisation that the 35 kb circular DNA is homologous to plastid genomes of photosynthetic organisms (Refs 14, 15). The complete sequence and analysis of the apicoplast genome of *T. gondii* has provided new insight into the taxonomic relationships of the organelle and its potential as a drug target (Ref. 16; J.C. Kissinger et al., Department of Biology, University of Pennsylvania, Philadelphia, PA, USA, pers. commun.; GenBank accession number U87145; <http://www.sas.upenn.edu/~jkissing/toxomap.html>). Collectively, these studies suggest that strain type, organellar function and communication with the nucleus probably play a vital role in the developmental changes and consequent disease pathology of *T. gondii*.

### Pathogenesis

The obligate intracellular pathogen *T. gondii* is unique among apicomplexans because it can invade and multiply in the nucleated cells of virtually all warm-blooded animals (Ref. 1). The sequential release of proteins from the three major secretory organelles, namely the micronemes, rhoptries and dense granules (Ref. 17; <http://webs.cb.uga.edu/~striepen/gfp.html>), appears to facilitate host cell attachment, invasion and generation of the parasitophorous vacuole (Refs 18, 19; Fig. 1). The parasitophorous vacuole provides a safe environment for the parasite to multiply, because it is resistant to acidification and lysosomal fusion (Refs 20, 21, 22, 23). Although sexual reproduction occurs exclusively in the definitive host (members of the cat family), secondary hosts (including humans) can participate in an indefinite cycle of asexual reproduction via infection and carnivory (i.e. by eating undercooked meat; Ref. 1; Fig. 2). Mice are thought to be an important natural secondary host and serve as the model system of choice for studies of pathogenesis. The acute disease phase is associated with the rapidly dividing form or 'tachyzoite', and the chronic phase with the presence of parasite tissue cysts containing the slowly dividing form or 'bradyzoite'. The tissue cysts can persist for the life of the host with no apparent ill-effects in healthy individuals. A significant proportion of the adult human



**Figure 1. Ultrastructure of a *Toxoplasma gondii* tachyzoite.** The conoid defines the apical end of the parasite and is thought to be associated with the penetration of the host cell. Micronemes, rhoptries and dense granules are the three major secretory organelles, found predominately at the apical end of the parasite. Microneme proteins are released very early in the invasion process, facilitating host-cell binding and gliding motility. Rhoptry proteins are also released during invasion, and can be detected within the lumen and membrane of the newly generated parasitophorous vacuole (PV). Dense-granule proteins are released during and after the formation of the PV, modifying the PV environment for intracellular survival and replication of the parasite. The apicoplast is a plastid-like four-membrane organelle containing a 35 kb circular DNA. Most of the proteins functioning within the organelle are encoded by the nucleus, and are specifically targeted to the apicoplast. This targeting involves the secretory pathway, including the rough endoplasmic reticulum (ER) and a Golgi body situated immediately apical to the nucleus. Targeted proteins have a bipartite N-terminal extension, consisting of an ER signal sequence followed by a plastid transit peptide. *T. gondii* cells have a single nucleus and a single mitochondrion. It is hypothesised that reliance on the mitochondrion for cellular metabolism differs according to the life-cycle stage of the parasite (**fig001jac**).

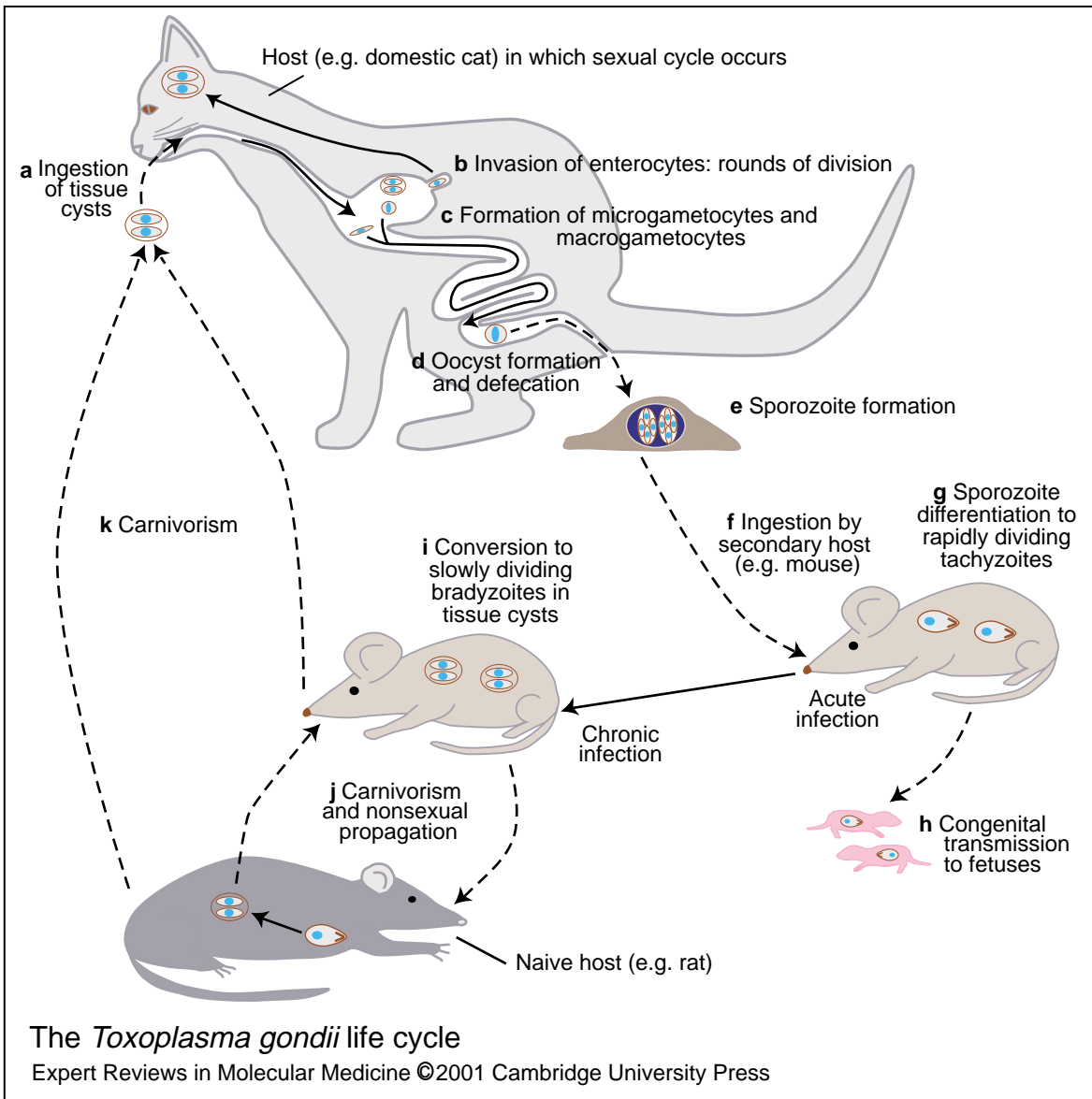
population has been infected by the parasite, depending on lifestyle and geographic locale. For example, it is thought that the lower infection rate in the UK (30–40%) compared with France (60–90%) may be due to differences in the consumption of rare or undercooked meat (Ref. 1).

Although human infections are usually relatively benign and lead to lifelong immunity, acute *T. gondii* infection during pregnancy in the nonimmune host can cause congenital transmission resulting in serious birth defects (affecting up to 0.1% of live births in some locations; Refs 24, 25). Furthermore, the reactivation of a latent chronic infection in an immunocompromised host can cause a life-threatening encephalitis; thus, *T. gondii* is a major contributor to AIDS-related disease (Ref. 2). Unfortunately, chemotherapeutic treatment of an active acute infection does not prevent recurrence

at a later date, so many immunocompromised patients require long-term prophylactic treatment. From a clinical standpoint, the elucidation of the mechanisms leading to the establishment and maintenance of a chronic infection is of paramount importance because no therapy currently exists to treat this stage of the disease (Refs 2, 26).

### Genetics and the nuclear genome

Genomic studies necessarily include basic genetic information for both experimental design and downstream data analysis. Population and genetic linkage studies of *T. gondii* provide a robust framework for genomic analysis because they define the range of heritable and phenotypic differences between strains. For disease pathology, differences in the ability of strains to establish a chronic infection is the principal observation upon which most of the genomic studies are based.



**Figure 2. The *Toxoplasma gondii* life cycle.** *T. gondii* is an obligate intracellular parasite and its life cycle includes both sexual and asexual modes of proliferation and transmission. The sexual cycle takes place exclusively in the intestinal enterocytes of many members of the cat family (*Felidae*). (a, b) After ingestion of tissue cysts, the parasites invade the enterocytes, undergo several rounds of division and (c) differentiate into microgametocytes and macrogametocytes. (d) The gametocytes fuse to form a zygote or 'oocyst' that is shed into the environment with the cat's faeces. (e) The oocyst undergoes meiosis, producing an octet of highly infectious 'sporozoites' that are resistant to environmental damage and may persist for years in a moist environment. (f) After ingestion (by a secondary host such as a mouse), (g) sporozoites differentiate into the rapidly dividing 'tachyzoite' form, which establishes and sustains the acute infection. (h) During the acute infection, congenital transmission to the developing fetus can occur. (i) In many hosts, a chronic phase of the disease ensues, as the tachyzoite changes into a slowly dividing form known as the 'bradyzoite'. Latent bradyzoite tissue cysts persist for the life of the host, re-emerging occasionally, but do not produce overt disease in healthy individuals. (j) Carnivorous ingestion of tissue cysts can lead to the infection of a naive host, allowing for an indefinite nonsexual propagation of *T. gondii*. (k) In the cat, this will initiate the sexual cycle. The solid lines indicate parasite differentiation and the dashed lines indicate modes of transmission. Adapted from Ref. 1 (fig002jac).

## Phylogeny and genotype

Phylogenetic analyses of *T. gondii* isolates show that the species comprises three major branches, or clades, that represent three clonal lineages, members of which are referred to as Types I, II and III, respectively (Refs 7, 27). This unusual population structure raises the possibility that *T. gondii* is really three species. However, although most isolates fall into one of the three lineages, a few extensively mixed genotypes have been recovered from natural sources, supporting the notion that *T. gondii* is a single biological species (Refs 7, 27, 28). Moreover, restriction fragment length polymorphism (RFLP) analyses (Ref. 7) and comparative sequence analysis of individual genes (Refs 12, 29, 30) estimate only 1% divergence at the DNA sequence level between lineages. The RFLP analysis of 106 independent, geographically disparate, strains revealed only 15 genotypes out of a possible 1700, and the overwhelming majority of strains within each lineage were represented by a single genotype (Ref. 7).

Because each lineage consists of isolates from different continents, the observed population structure cannot be accounted for by geographic separation or local outbreak. Large-scale epidemic spread is also an unlikely explanation because transmission requires ingestion of tissue cysts or oocysts. The three clonal lineages are most probably the result of a population bottleneck that has been subsequently maintained by the limited opportunity for sexual recombination (i.e. co-infection of different strains in a single cat), and propagation through carnivorous cycles (Ref. 7; Fig. 2).

From a disease perspective, relationships between lineage and phenotype might have important implications. Isolates from each of these lineages have been cultured from both human and animal sources but are not represented equally within a given source. A breakdown of 106 *T. gondii* isolates showed that the majority of human isolates were Type II (Ref. 7). By contrast, acute virulence in mice is restricted to Type I strains (Ref. 28). Additionally, Type I strains do not readily produce tissue cysts or participate in the sexual cycle, whereas Type II and III strains maintain the ability to complete the entire life cycle. This correlation between lineage and phenotype suggests that the genotype of a parasite will influence the severity and progression of disease.

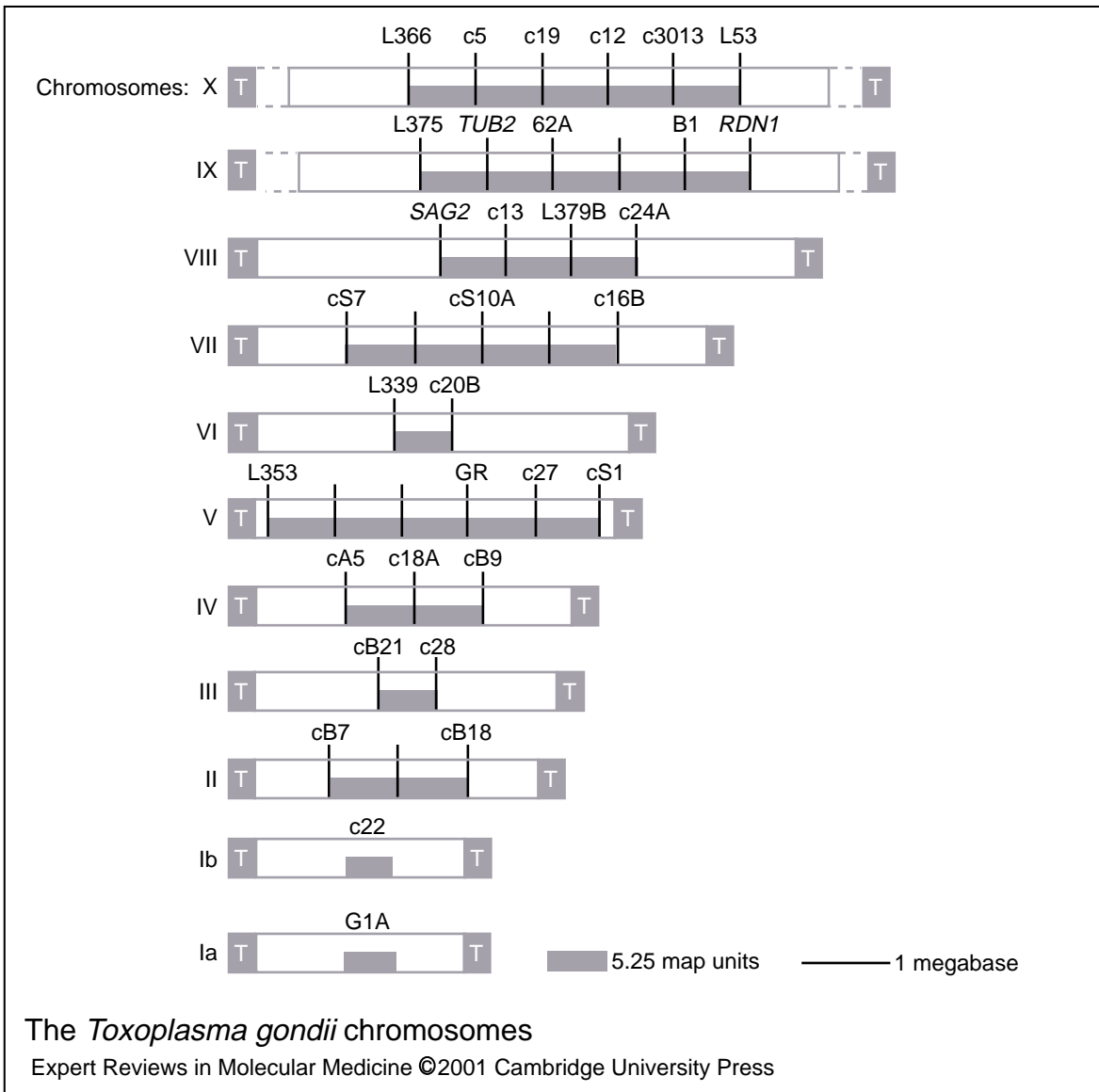
## Genomic analysis

The *T. gondii* genome consists of three DNA components, located in the nucleus, apicoplast and mitochondrion. The nuclear genome is haploid for most of the parasite's life cycle, except for a brief diploid phase in the cat intestine before meiosis (Ref. 31). The haploid DNA content is estimated to be 87 Mb, and linkage analysis of approximately 80 polymorphic loci in a cross between a clone of ME49 (a Type II strain) and CEP (a Type III strain) has defined 11 linkage groups with a total of 147 centimorgans (Ref. 8). The chromosomes do not condense during mitosis, so the karyotype has been determined by pulsed-field gel electrophoresis (PFGE). The 11 chromosomes are designated by Roman numerals Ib, Ia, II to X, and range in size from approximately 1.8 Mb to >10 Mb (the present limit that can be determined by PFGE; Ref. 9). Because the karyotype is consistent with the genetic linkage analysis, it is likely that all of the chromosomes have been isolated and identified, and that there is only a small amount of observed inter-strain chromosomal polymorphism. Comparisons of the karyotypes of representatives of the three major clonal lineages show very little chromosomal size variation; furthermore, the karyotypes appear to be highly stable, revealing no apparent difference during ten years of continual passage (Refs 9, 32; Fig. 3).

Compared with other protozoans, including related apicomplexans such as *Plasmodium falciparum*, *T. gondii* maintains a remarkably conserved nuclear genome despite a wide host range and nonobligatory sexual cycle. This property of the *T. gondii* genome will facilitate the pilot genomic sequencing project that is concentrated on chromosome Ib (J. Ajioka and the Sanger Centre, Cambridge, UK, unpublished), and future efforts to identify genes, open reading frames (ORFs), intron–exon boundaries, upstream *cis*-acting promoter elements, and so on. In turn, the inter-strain differences between genes and/or ORFs should provide us with insight into the associated variation in disease pathology.

## Gene expression

Current evidence suggests that gene expression in *T. gondii* is transcriptionally regulated (Refs 33, 34, 35, 36). Unlike in other protozoan parasites such as *Trypanosoma* spp. and *Leishmania* spp., polycistronic transcription and RNA editing



**Figure 3. The *Toxoplasma gondii* chromosomes.** The grey bars and associated genetic markers for each linkage group are superimposed on bars that indicate the physical size of the respective chromosomes (chromosomes IX and X are of an undetermined size). A representative marker is shown for each known map position. The number of map units for a given chromosome varies: chromosomes V, IX and X consist of 26.5 map units whereas chromosome VI consists of only 5.25 map units. These differences are likely to be due to random fluctuations in the number of restriction fragment length polymorphism markers identified from a screen of *T. gondii* genomic cosmid and plasmid DNA. No recombination was observed for chromosomes Ia and Ib. Abbreviations used: T, telomeres. Adapted from Refs 8 and 9 (fig003jac).

have not been detected. Although conventional *cis*-acting eukaryotic promoters such as the TATA box or SP1 motif have not been observed, upstream sequence analysis of several genes has identified a common highly conserved T/AGAGACG heptanucleotide core element (Refs 34, 35, 36; Table 1).

The upstream regions of genes coding for five highly abundant proteins – the surface antigen SAG1 and the four dense-granule proteins GRA1, GRA2, GRA5 and GRA6 – have been analysed for promoter and/or transcription initiator activity (Refs 34, 35). *SAG1* is a single-copy gene containing no introns and has two transcriptional

**Table 1. Upstream repeated motif sequences of *Toxoplasma gondii* genes<sup>a</sup> (tab001jac)**

Gene	Sequence of repeated motif <sup>b</sup>	Form
<b>GRA genes</b>		
<i>GRA1</i>	CGACACG	Direct
	TGAGACG	Inverted
	CGTGACG	Inverted
	TGAGACG	Direct
	TGAGACG	Inverted
	CGAGACG	Inverted
<i>GRA2</i>	TGCGACG	Inverted
	AGAGACG	Inverted
	AGAGATG	Inverted
	AGAGACG	Direct
	AGACGCG	Direct
	AGCGACG	Inverted
	AGACACG	Direct
<i>GRA5</i>	AGAGACG	Direct
	AGAGACG	Direct
<i>GRA6</i>	AGACACG	Direct
	TGCGACG	Direct
	AGCGACG	Inverted
	TGCGACG	Direct
	TGAGACG	Inverted
<b>Other <i>Toxoplasma</i> genes</b>		
<i>TUB1</i>	GGAGATC	Direct
<i>SAG1</i>	TGAGACG	Direct
	TGACACG	Direct
	TGAGACG	Direct
	TGAGACG	Direct
	TGAGACG	Direct
<i>MIC2</i>	AGAGACG	Direct
<b>Consensus repeated motif</b>	A/T G A G A C G	
Percentage conservation	84 100 72 84 100 92 100	

<sup>a</sup> Table adapted from Ref. 35  
<sup>b</sup> Bold letters indicate those bases that are identical to the consensus motif

initiation sites separated by about 35 bp (base pairs; Ref. 37). Upstream sequence of *SAG1* has revealed a series of six 27 bp direct tandem repeats, each with a heptanucleotide core element (Table 1). The tandem array is positioned 35 bp upstream from the second (most 5') initiation site. *GRA* genes are also single copy and similarly lacking in introns (except *GRA2*, which has one intron); however, the arrangement of the heptanucleotide motifs in the *GRA* genes is more variable in number, position and orientation (summarised in Ref. 35). *GRA5* contains only two

directly repeated motifs, separated by 70 bp, whereas *GRA1* contains six repeats in both direct and inverted orientations, separated by distances ranging from 27 bp to 98 bp.

Using a stable transfection system with a chloramphenicol acetyltransferase (CAT) reporter construct (Ref. 38), a series of deletion constructs of the upstream regions for *SAG1* (Ref. 34), *GRA1*, *GRA2*, *GRA5* and *GRA6* (Ref. 35) were tested for activity. In all cases, deletion constructs of the heptanucleotide motifs nearest the initiation site showed the greatest effect on

CAT activity. The mix of direct and inverted orientations of the heptanucleotide motif repeats and their relative influence on gene expression suggest that they might act like SP1 elements to determine transcription initiation start sites in the absence of TATA promoter elements (Refs 35, 39). Because SAG1 is a developmentally regulated glycosylphosphatidylinositol- (GPI-) anchored surface protein (Refs 40, 41, 42) and the dense-granule GRA proteins are co-secreted components of the parasitophorous vacuole (Ref. 43), it is likely that other promoter elements involved in directing transcription for these different properties remain to be identified.

### EST analysis

For organisms with relatively large nuclear genomes, including *T. gondii*, EST analysis is a faster and more cost-effective method for generating DNA sequence information than a genomic approach. An EST approach comparing strains and developmental stages has proved to be a useful method for elucidating the processes that differ between the acute and chronic infection. ESTs generated from cDNA libraries of the RH tachyzoite, ME49 tachyzoite and ME49 bradyzoite (in vivo and in vitro) have been aligned to form a set of overlapping sequences spanning part or all of the original mRNA (messenger RNA), revealing sequence polymorphisms and expression differences (Ref. 12). The database dbEST currently holds over 10 000 *T. gondii* sequences, and analysed EST data can be accessed at <http://www.cbil.upenn.edu/ParaDBs/>. This database can be searched and browsed for GenBank neighbour IDs, protein motifs, aligned consensus sequences and clustering information. A comparative EST analysis affords the possibility of providing DNA sequence polymorphism data and relative gene expression information, as well as identifying new genes.

### ESTs and polymorphism

DNA sequence polymorphism between strains and developmental stages can be examined through the cluster and alignment analyses, but this relies upon multiple sequences (i.e. relatively abundant ESTs) to distinguish between sequencing error and bona fide polymorphism. The DNA polymorphisms are potentially useful genetic markers and some will underlie phenotypic differences between strains. Inspection of the clusters with greater than ten member ESTs

suggests that cDNA sequence divergence between RH and ME49 is about 0.5%. An analysis of these polymorphisms with an expanded number of recombinant progeny would greatly refine the current *T. gondii* genetic map (Ref. 8). Despite the limited ability to assign these polymorphisms to the 5' untranslated region (UTR), 3' UTR or coding sequence, a proportion will be useful for gene identification through either a candidate gene approach or a positional cloning approach. For example, the genetic map could be used to test for co-segregation of a candidate EST polymorphism with polymorphic phenotypic markers such as strain-specific differences in monoclonal antibody affinity for the bradyzoite antigens P18 and P36 (Ref. 44).

### ESTs and gene expression

Because ESTs are generated from a randomly selected set of cDNAs, differences in gene expression between strains and developmental stages might be revealed. Along with protein polymorphism, gene expression differences between RH and ME49 are likely to contribute to the difference in the parasite's ability to cause acute (RH) or chronic (ME49) infection. Cluster analysis of ESTs from the 5359 RH tachyzoite and the 1806 ME49 tachyzoite showed that highly expressed dense-granule proteins and surface antigens were the most common ESTs in both RH and ME49 EST libraries (Ref. 12; Table 2). Overall, the relative frequencies of the ESTs were similar between the two libraries but some of the most abundant ESTs, such as *GRA3*, *GRA6* and two unknown genes, varied in frequency by twofold or more. Northern blot analysis did not consistently mirror these differences in the levels of mRNA transcripts, indicating that some genes are over- or under-represented in the EST set. One of the unknown ESTs (cluster ID Ctoxqual2\_1339; dbEST ID 466406) is eightfold more frequent in the RH tachyzoite library, but this strain difference is complicated by the high representation of this EST in the ME49 bradyzoite libraries (Table 2; <http://www.cbil.upenn.edu/ParaDBs/>). This observation could reflect a real shift in ME49 developmental gene expression or simply be the result of cloning bias or sampling error. Moreover, ESTs with very large differences in representation and confirmed expression differences are not necessarily due to straightforward transcriptional upregulation. The EST TgESTzy64e04 is 14-fold more frequent in ME49, and northern blot analysis

**Table 2. Summary of abundant expressed sequence tags (ESTs) from both RH (total ESTs 5359) and ME49 (total ESTs 1806) tachyzoite cDNA libraries<sup>a</sup> (tab002jac)**

Gene (product)	Cluster ID	RH tachyzoite		ME49 tachyzoite	
		dbEST ID <sup>b</sup>	% Rep <sup>c</sup>	dbEST ID <sup>b</sup>	% Rep <sup>c</sup>
<i>GRA1</i> (p24 dense-granule protein)	Ctoxqual2_2619	465887	2.00	571569	1.61
<i>GRA2</i> (p28 dense-granule protein)	Ctoxqual2_1321	466589	1.29	571076	1.77
<i>GRA6</i> (p32 dense-granule protein)	Ctoxqual2_1409	465733	1.27	571154	0.50
<i>SAG1</i> (p30 surface antigen)	Ctoxqual2_1823	465767	1.19	571382	2.33
<i>SAG2</i> (p22 surface antigen)	Ctoxqual2_867	466942	0.99	571108	1.50
Unknown	Ctoxqual2_31	467174	0.63	571208	1.55
<i>rpS8</i> (ribosomal protein S8)	Ctoxqual2_869	465857	0.54	620274	0.17
<i>GRA3</i> (p30 dense-granule protein)	Ctoxqual2_533	466627	0.52	571209	0.17
Unknown	Ctoxqual2_1339	466406	0.49	574388	0.06

<sup>a</sup> Table adapted from Ref. 12  
<sup>b</sup> dbEST ID is the identification number from the Expressed Sequence Tags database (which can be found at <http://www.ncbi.nlm.nih.gov/dbEST/>)  
<sup>c</sup> % Rep is the percentage representation of the EST within a particular library

has revealed a 6.3 kb transcript in ME49 that is not seen in RH. Although these data suggest that EST numbers roughly correlate with cellular gene expression, it should be remembered that cloning bias, random sampling and sequencing error are likely to affect the statistics in a sample of this magnitude, and probably account for most of the observed differences.

The encysted bradyzoite stage of *T. gondii*, a hallmark of the chronic infection, is much less well characterised than the active tachyzoite stage owing to the relative difficulty in obtaining sufficient material from mouse brain tissue cysts. However, with the discovery of culturing procedures to trigger the differentiation of in vitro tachyzoites into bradyzoites (e.g. alkaline culture media), this problem has been partially resolved (Ref. 45). Despite these advances, the life-cycle stage conversion process remains poorly understood. For a developmental-stage comparison, approximately 2500 ESTs have been generated from an in vivo bradyzoite (mouse brain tissue cysts) library and 760 from

an in vitro differentiated bradyzoite library. Known tachyzoite-specific, bradyzoite-specific and non-stage-specific (constitutive) proteins correlate very well with relative EST abundance from ME49 tachyzoite and ME49 bradyzoite libraries (Ref. 46; Table 3). The genes with a constitutive expression pattern show relatively low variation, and the greatest difference between tachyzoites and bradyzoites is less than fourfold. Among the five stage-specific proteins reported, three ESTs are absolutely different (i.e. not sampled in the opposing stage), whereas abundance of the other two differs over tenfold. This analysis also revealed other stage-specific ESTs, including tachyzoite- and bradyzoite-specific forms of enolase. The stage-specific expression patterns of enolases support the hypothesis that the glycolytic pathway is different between the two stages (Ref. 47). Despite the fact that these strain and developmental EST abundance comparisons appear to identify reliably only genes with very large expression differences, they are a useful first approximation

**Table 3. Summary of abundant expressed sequence tags (ESTs) from an ME49 tachyzoite cDNA library (total ESTs 1806) and an ME49 bradyzoite cDNA library (total ESTs 2353)<sup>a</sup> (tab003jac)**

Gene	Expression	ME49 tachyzoite		ME49 bradyzoite	
		No. of ESTs	% Rep <sup>c</sup>	No. of ESTs	% Rep <sup>c</sup>
<i>SAG1</i>	Tachyzoite specific	39	2.15	2	0.08
<i>SAG2</i>	Tachyzoite specific	24	1.32	0	0
<i>SAG4</i>	Bradyzoite specific	0	0	23	1.02
<i>BAG1</i>	Bradyzoite specific	0	0	100	4.42
<i>MAG1</i>	Bradyzoite specific	1	0.05	18	0.76
<i>GRA1</i>	Constitutive <sup>b</sup>	38	2.10	14	0.56
<i>GRA2</i>	Constitutive	28	1.55	43	1.83
<i>GRA3</i>	Constitutive	5	0.28	3	0.13
<i>GRA4</i>	Constitutive	6	0.33	6	0.25
<i>GRA5</i>	Constitutive	27	1.49	44	1.87

<sup>a</sup> Table adapted from Ref. 46  
<sup>b</sup> 'Constitutive' indicates that the gene is expressed in both tachyzoites and bradyzoites  
<sup>c</sup> % Rep is the percentage representation of the EST within a particular library

and have identified some new genes for further investigation.

### ESTs and orthologous gene discovery

BLAST (basic local alignment search tool; Ref. 48) similarity searches, with a cut-off value of  $P < 10^{-10}$ , show that about 27% of *T. gondii* EST sequences are significantly similar to previously identified genes or proteins. Several ESTs appear to encode the *T. gondii* version of proteins (i.e. orthologues) that have been characterised in related apicomplexan parasites (Ref. 12). Some of these proteins had even been described in *T. gondii*, but the genes encoding them had not been reported. This was particularly evident for the protein components of the organelles that participate in the active invasion process that is common to all apicomplexans. For example, the subcellular fractionation of *T. gondii* micronemes identified a set of proteins that are secreted during attachment to

the host cell (Refs 49, 50). Studies in *Eimeria tenella* and *P. falciparum* revealed a similar set of proteins, including *E. tenella* Etp100 (Ref. 51) and *P. falciparum* TRAP (Ref. 52). Furthermore, a database search of *T. gondii* ESTs showed that the putative EST tg050 product had striking similarity to the thrombospondin (TSP) or TSP-like adhesion domains of Etp100 and TRAP (Ref. 53). Full-length cDNA and genomic sequence established the major ORF, intron-exon structure and transcriptional initiation site with putative promoter elements. The sequence alignment and domain analysis with Etp100 suggested that they are orthologous genes or proteins and are members of a family of apicomplexan adhesion molecules including TRAP. The tg050 recombinant protein was shown to cross-react with antisera raised against the *T. gondii* microneme protein MIC2, providing conclusive evidence that the EST tg050 sequence encodes a microneme protein.

### ESTs and paralogous gene discovery

EST analysis has also been used to identify gene family members (i.e. paralogues), revealing new members of the *T. gondii* superfamily of highly expressed surface antigens (Refs 40, 41, 42). These GPI-anchored glycoproteins are thought to allow parasite attachment to the vast array of possible host cell types. SAG1 is the major member of the surface antigen family, which also includes SAG2, SAG3, SAG5, BSR4 and SRS1 (SAG-related sequence 1; Ref. 46). Expression of these proteins can be developmental-stage specific: whereas SAG1 and SAG2 are restricted to tachyzoites, BSR4 is specific to bradyzoites or is constitutively expressed, and SAG3 is found on both tachyzoites and bradyzoites (summarised in Ref. 54).

Although strains in which SAG1 or SAG3 expression is disrupted show reduced invasion and virulence compared with the parental ME49, they can still invade and replicate successfully in vivo, suggesting a possible redundancy of function between family members (Ref. 55). The conserved cysteines and hydrophobicity patterns (signal peptide and GPI anchor addition signal) of the SAG family suggested that BLAST and FASTA searches of the ESTs would uncover new family members (Ref. 46).

The discovery of the SAG paralogues SRS2, SRS3 and SRS4 confirmed the conservation of these structural features; overall amino acid identities to SAG1 are about 30%. These results support the idea that the family members are under selective pressure to maintain a particular secondary structure and some functional redundancy, yet might have divergent biological roles. The expression of SRS2 and SRS3 appears to be restricted to tachyzoites, and the virulent RH strain (Type I) shows little expression of SRS2 compared with abundant expression on the nonvirulent ME49 (Type II) and CEP (Type III) strains. The domination of the parasite surface by a single protein family whose members can be expressed differentially between strains and developmental stages suggests that functional and immunological studies of SAG and SRS proteins will provide significant insight into their roles in the acute and chronic infection (Ref. 56).

### Apicoplast proteins

Given the limited coding capacity of the apicoplast 35 kb genome, most of the apicoplast proteins must be encoded by the nucleus and imported to the organelle (J.C. Kissinger et al.,

pers. commun.; GenBank accession number U87145; <http://www.sas.upenn.edu/~jkissing/toxomap.html>). Discounting plastid proteins that are not thought to function in the apicoplast (e.g. photosynthetic proteins), as many as 800 proteins may be imported into the organelle (Ref. 57); however, more-recent estimates suggest that 100–200 is a more likely number (J.C. Kissinger, pers. commun.). Defining common features among apicoplast proteins (e.g. N-terminal transit peptides for entry into the organelle) is essential for computer-assisted identification of these proteins from primary sequence data (Ref. 58; <http://www.sas.upenn.edu/~jkissing/wh99.html>). In turn, this set of proteins should reveal the function(s) of the apicoplast.

As a result of the evolutionary origins of the apicoplast, which involved a secondary endosymbiotic event and retention of four unit membranes, imported apicoplast proteins have two additional membranes to cross compared with plastids of higher plants or algae (see below). This observation has suggested that a plastid transit peptide would be necessary but probably not sufficient for import into the organelle. Evidence including the direct observation that the secretory pathway is used for protein routing in organisms that have complex (multimembranous) plastids (Ref. 59) has raised the possibility that a similar pathway could be employed by apicomplexans. Organelle protein synthesis and class II fatty acid metabolism are two functions that are characteristic of plastids, so it was hypothesised that ribosomal proteins and proteins in the fatty acid synthesis pathway might be imported into the apicoplast (Ref. 13). BLAST analysis of the *T. gondii* ESTs uncovered homologues for ribosomal proteins S9 and L28 and two fatty acid biosynthetic enzymes, namely acyl carrier protein (ACP) and  $\beta$ -hydroxyacyl-ACP dehydrogenase (FabZ). The full-length sequence of the corresponding cDNAs revealed what appeared to be long bipartite leader peptides composed of a classical signal peptide followed by a putative transit peptide. An analysis of transiently transfected fusion constructs of the ACP N-terminal extension and green fluorescent protein (GFP) demonstrated that the predicted 104 amino acid N-terminal sequence is sufficient for proper targeting to the apicoplast. Further studies in *P. falciparum* have provided direct evidence for the bipartite nature of the N-terminal leader sequence, and the

collective N-terminal sequence information is currently being employed in computational methods for identifying new apicoplast proteins (Ref. 57; <http://www.sas.upenn.edu/~jkissing/wh99.html>).

The molecular phylogenies of *T. gondii* ribosomal proteins S9 and L28 argue that they are bacterial or plastid in nature (Ref. 13), so analysis of their gene structure could reveal how they acquired elements crucial for their expression in the nucleus and subsequent transport to the apicoplast. The gene transfer from the ancestral plastid genome must have involved a duplicate copy of each gene being inserted into the nucleus, and the acquisition of elements for nuclear expression and targeting sequences for import into the organelle (reviewed in Refs 60, 61). Although it is possible to select from a pool of random sequences to promote transcription (see for example Ref. 62) or encode peptides for mitochondrial import (see for example Ref. 63), data indicate that the acquisition of pre-existing functional sequences is a more-effective mechanism. The gene structure of the nucleus-encoded chloroplast *rpl22* gene suggests that its transit peptide was acquired via exon shuffling (Refs 64, 65). The maize *rps14* gene has achieved a similar result by alternative splicing after integration into the intron of the nucleus-encoded mitochondrial protein gene *sdh2* (Refs 66, 67). Both *T. gondii* *rps9* and *rpl28* genes show splice junctions adjacent to the leader sequence. Given that these proteins are plastid encoded in algae (Ref. 68), an analysis of 5' upstream sequences could provide clues as to the nature and timing of the plastid-to-nucleus transfer. The identification and characterisation of more nucleus-encoded genes and/or proteins targeted to the apicoplast will undoubtedly lead to a better understanding of its origin and function.

### The apicoplast: origins, function and chemotherapy

The retention of the apicoplast throughout the phylum Apicomplexa suggests an essential function for the plastid-like organelle within the parasite. The lack of this unusual organelle in the vertebrate host makes it an attractive chemotherapeutic target. Defining the origin and taxonomic relationships of the apicoplast and its 35 kb genome is crucial for understanding the action of some currently used drugs, as well as the potential development of herbicide-

based drugs and the discovery of new drug targets. Although our understanding of the apicoplast is derived from studies in a variety of apicomplexan species, the attributes of *T. gondii* as a model system have made it indispensable for investigating this curious organelle.

### Discovery of the apicoplast

The connection between the apicoplast organelle and its genome is a historically interesting story because the molecular evidence from the genome led to the re-discovery of the organelle (reviewed in Refs 69, 70). The apicoplast genome was originally discovered as an AT-rich extrachromosomal 35 kb circular DNA isolated from a density-gradient fraction from total *Plasmodium knowlesi* DNA (Ref. 71). Several lines of evidence from the *P. falciparum* 35 kb DNA suggested a plastid-like origin and a possible association with a nonmitochondrial organelle (summarised in Refs 14, 69). Plastid-like structural and sequence data included an inverted repeat organisation of the genes encoding the large and small subunits of rRNA (ribosomal RNA; Refs 72, 73), the predicted secondary structure of the small-subunit rRNA (Ref. 72), and split RNA polymerase genes *rpoC1* and *rpoC2* (Refs 74, 75). Candidates for a vestigial plastid included a structure described in electron microscopic studies as a multimembranous organelle, observed in a variety of apicomplexan species (summarised in Ref. 69). The *T. gondii* version was given various names including the 'Golgi adjunct' and 'Hohlzylinder' (Ref. 76). Transcript localisation via in situ hybridisation with an antisense small-subunit rRNA gene probe to *T. gondii* thin sections provided the first direct evidence that the 35 kb element localised to this multimembranous organelle (Ref. 15). Further ultrastructural analysis revealed that the plastid-like organelle consists of four unit membranes, and it was dubbed the 'apicoplast' (Ref. 77).

### Evolutionary history

Deciphering the evolutionary history of the apicoplast is not only of intrinsic interest, but could also provide clues to which plastid metabolic pathways are retained in the organelle, which pathways have been transferred to the cytosol (either as replacement or redundant functions) and which pathways have been lost. Ultimately, the *raison d'être* of this organelle could be assessed.

The apicoplast, like all other plastids, originated from a primary endosymbiotic event with a cyanobacterial-like prokaryote, resulting in the characteristic two-membrane organelle that is found in higher plants and both red and green algae (reviewed in Refs 69, 78). Evidence suggests that this primary endosymbiont was subsequently transferred into several existing eukaryotic lineages (Ref. 78). These secondary endosymbionts are characterised by organelles that have either three or four membranes, and one of these secondary endosymbiotic events founded the lineage that includes the Apicomplexa. Phylogenetic analyses of nucleus-encoded 18S rRNA genes indicate a monophyletic origin of apicomplexans and dinoflagellates (see for example Refs 79, 80, 81). Although dinoflagellates include members that have multimembranous plastids, direct comparisons with these species must be made with care because secondary endosymbiotic events probably occurred many times within the group (Ref. 82). Nevertheless, these phylogenetic and ultrastructural observations support the view that the apicoplast was acquired through secondary endosymbiosis of an alga, and further analyses of nucleus-encoded apicoplast genes are likely to increase our understanding of the origins of this organelle.

Independent evidence from phylogenies of the *tufA* gene (which encodes an elongation factor) associates the apicoplast with green algae (Ref. 77). By contrast, the apicoplast genome structure is more like that of red algae. The ribosomal protein genes *s12* and *s7*, and the *tufA* gene, are found at the 3' end of the operon, as in red algae; contrastingly, these genes are found at the 5' end in green algae (Ref. 69). The apicoplast also retains the *ycf24* and *clpC* genes, which are both found in red algae but not in green algae (Ref. 83). However, the mapping of chloroplast gene loss across an inferred tree of unicellular and multicellular organisms has shown that independent parallel gene losses in multiple lineages far outnumbered phylogenetically unique losses (Ref. 68). This suggests that these types of data probably will not satisfactorily resolve the basis of the original secondary endosymbiotic event. However, in spite of this conundrum, there is a strong collection of evidence to suggest that the apicoplast is a bona fide plastid that has some kind of algal origin.

### Apicoplast function and chemotherapy

The apicoplast provides an ideal target for drug treatment because of its bacterial or plastid origin and the possible maintenance of metabolic pathways not found in animals. The complete *T. gondii* 35 kb apicoplast DNA sequence appears to include genes encoding both large- and small-subunit rRNAs, RNA polymerase genes, the elongation factor *tufA* gene, tRNA (transfer RNA) genes, a putative ribosomal protein gene operon and five ORFs of unknown function (J.C. Kissinger et al., pers. commun.; GenBank accession number U87145; <http://www.sas.upenn.edu/~jkissing/toxomap.html>). This analysis suggests that the apicoplast maintains the capacity for gene expression; however, most of the functional proteins are encoded in the nucleus and imported into the organelle (see section entitled 'Apicoplast proteins'). The investigations into apicoplast functions have been intimately associated with both defining the action of currently used drugs and searching for new chemotherapeutic targets based on plastid function in plants and algae. Several commonly used antibiotics (representing different structural classes) probably target apicoplast DNA replication and gene expression machinery for parasitocidal activity (Ref. 84). The apicoplast roles in cellular metabolism such as the synthesis of aromatic amino acids and fatty acids have been hypothesised on the basis of known plastid functions (summarised in Ref. 16).

Although the apicoplast does not appear to encode proteins for DNA replication, the replication mechanism(s) could include a theta and/or a rolling circle process (Ref. 85). Despite the apparent variation of the number of circular copies versus linear copies between apicomplexan species (Refs 83, 86, 87), the inhibitory activity of fluoroquinolones such as ciprofloxacin suggests the presence of a bacterial-like type II topoisomerase acting on supercoiled circular DNA (Ref. 88). Curiously, ciprofloxacin inhibits the replication of *T. gondii* apicoplast DNA in vitro, but does not immediately affect cell division (Ref. 84). Cell death caused by ciprofloxacin is seen in the parasites that have invaded a new host cell, not the parasites originally exposed to the drug. This 'delayed death' could be simply a result of a limiting amount of apicoplast DNA segregating during cell division and/or inhibition of the synthesis of a vital component for the formation of the next parasitophorous vacuole

(e.g. disruption of lipid biosynthesis). Similar 'delayed death' observations with clindamycin (Ref. 89), a drug that probably inhibits apicoplast translation, support the latter hypothesis. The presence of *rpoB*, *rpoC1* and *rpoC2* genes suggests that apicoplast transcription utilises a bacterial or plastid  $\alpha$ ,  $\beta$ ,  $\beta'$  DNA-dependent RNA polymerase (Ref. 90) and that rifampicin might inhibit apicoplast transcription (Ref. 16). Although there is only indirect evidence for apicoplast translation, the putative translation machinery provides several useful chemotherapeutic targets, against which some current drugs have demonstrable antiparasite activity (summarised in Refs 16, 83). The large-subunit (23S) rRNA (*rrnL*) is the likely target for lincosamides (e.g. clindamycin), macrolides (e.g. erythromycin, azithromycin) and thiopeptides (e.g. thiostrepton, micrococcin). The translation elongation factor EF-Tu (*tufA*) is the putative target for amythiamicin (Ref. 91). Clindamycin, erythromycin and azithromycin have all been used successfully to treat toxoplasmosis (Refs 92, 93).

Because the coding capacity of the genome of the apicoplast is so limited, the process of discovering the biological role of the apicoplast has become a systematic search through metabolic pathways based on homology to known plastid functions. These include the biosynthesis of isoprenoids, aromatic amino acids and fatty acids (Refs 69, 94). The isoprenoid precursor isopentenyl diphosphate can be synthesised by two independent pathways: the standard acetate–mevalonate (MVA) pathway, and the alternative pathway, which utilises pyruvate and glyceraldehyde 3-phosphate as metabolic precursors (Ref. 95). The apparent absence of enzymes from the MVA pathway (Ref. 96) and also the identification of two genes (*P. falciparum* sequence database) encoding MVA-independent pathway enzymes in *P. falciparum* (Ref. 97) suggest that the MVA-independent pathway may be a valid drug target. Moreover, these proteins include N-terminal extensions, which have a demonstrable ability to target a GFP reporter construct to the *T. gondii* apicoplast (Ref. 97). Higher plants, algae, fungi and bacteria use the shikimate pathway to synthesise aromatic amino acids and folate precursors, a pathway not present in animals (Ref. 98). In plants and algae, this pathway is located in the plastid. Biochemical, genetic and chemotherapeutic evidence for a shikimate pathway in *T. gondii*, *P. falciparum* and

*Cryptosporidium parvum* has recently been reported (Ref. 99), although the exact cellular location of the pathway remains unclear (Refs 100, 101). The herbicide glyphosate (sold commercially as RoundUp™, Zero™ and Tumbleweed™) showed significant antiparasite activity, probably by inhibiting 5-enopyruvyl shikimate 3-phosphate synthase. Browsing the *T. gondii* EST data (Ref. 12) identified an EST with high similarity to chorismate synthase. Alignment of the translation from full-length sequence confirmed the identity and uncovered a number of insertions shared only with the putative *P. falciparum* chorismate synthase, suggesting that the shikimate pathway enzymes are likely to be good targets for new parasitocidal agents (Refs 99, 102). Inspection of the *T. gondii* EST data (Ref. 12; see above) revealed putative homologues for the cyanobacterial and plastid fatty acid synthetic enzymes ACP and FabZ that target the apicoplast (Ref. 13). This discovery will probably provide new chemotherapeutic targets, because the genes identified are of a specific plastid type and consequently distinct from those of mammals (Ref. 13).

### Conclusions

Genomic studies of both nuclear and apicoplast DNA of *T. gondii* have become the cornerstone for new investigations into both pathogenesis and chemotherapy. EST and related genetic information have provided key data to increase our understanding of virulence factors and differences between strains and clonal lineages. Although the primary sequence analysis of the 35 kb DNA does not give any direct evidence for the specific function(s) of the apicoplast, the retention of transcription and translation machinery suggests that the apicoplast is a functioning organelle, and this machinery might be targeted by currently used parasitocidal drugs. The movement of genes such as ACP and FabZ from the plastid genome to the nuclear genome through evolutionary time suggests that the identification of nucleus-encoded genes with homology to cyanobacterial and/or plastid genes will ultimately reveal both apicoplast and other plastid-like functions in the parasite. The interplay between basic biological investigations such as genomic analysis and those targeted at chemotherapy is essential to the development of more-effective drug treatments for toxoplasmosis and other apicomplexan-related disease.

### Acknowledgements and funding

We thank Iain Wilson (Department of Parasitology, National Institute for Medical Research, London, UK), David Ferguson (Nuffield Department of Pathology, University of Oxford, UK) and Jessie Kissinger (Department of Biology, University of Pennsylvania, Philadelphia, PA, USA) for reviewing the manuscript and providing helpful comments and suggestions. This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC).

### References

- 1 Dubey, J.P. (1977) *Toxoplasma*, *Hammondia*, *Besnotia*, *Sarcocystis*, and other tissue cyst-forming coccidia of man and animals. In *Parasitic Protozoa III* (Kreier, J.P., ed.), pp. 101-237, Academic Press, New York, NY, USA
- 2 Luft, B.J. and Remington, J.S. (1992) *Toxoplasmic encephalitis in AIDS*. *Clin Infect Dis* 15, 211-222, PubMed ID: 92392945
- 3 Georgiev, V.S. (1994) Management of toxoplasmosis. *Drugs* 48, 179-188, PubMed ID: 95080012
- 4 Carucci, D.J. et al. (1998) The malaria genome sequencing project. *Exp Rev Mol Med* 5 May, <http://www-ermm.cbcu.cam.ac.uk/dcn/txt001dcn.htm>
- 5 Wilson, R.J. and Williamson, D.H. (1997) Extrachromosomal DNA in the Apicomplexa. *Microbiol Mol Biol Rev* 61, 1-16, PubMed ID: 97257264
- 6 Ajioka, J.W. (1997) Analysis of apicomplexan parasites. In *Methods: A Companion to Methods in Enzymology* (vol. 13) (Ajioka, J.W., ed.), pp. 79-80, Academic Press, Orlando, FL, USA
- 7 Howe, D.K. and Sibley, L.D. (1995) *Toxoplasma gondii* comprises three clonal lineages: correlation of parasite genotype with human disease. *J Infect Dis* 172, 1561-1566, PubMed ID: 96083502
- 8 Sibley, L.D. et al. (1992) Generation of a restriction fragment length polymorphism linkage map for *Toxoplasma gondii*. *Genetics* 132, 1003-1015, PubMed ID: 93093427
- 9 Sibley, L.D. and Boothroyd, J.C. (1992) Construction of a molecular karyotype for *Toxoplasma gondii*. *Mol Biochem Parasitol* 51, 291-300, PubMed ID: 92244299
- 10 Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) dbEST—database for “expressed sequence tags”. *Nat Genet* 4, 332-333, PubMed ID: 94004957
- 11 Wan, K.L., Blackwell, J.M. and Ajioka, J.W. (1996) *Toxoplasma gondii* expressed sequence tags: insight into tachyzoite gene expression. *Mol Biochem Parasitol* 75, 179-186, PubMed ID: 96258558
- 12 Ajioka, J.W. et al. (1998) Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa. *Genome Res* 8, 18-28, PubMed ID: 98112779
- 13 Waller, R.F. et al. (1998) Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 95, 12352-12357, PubMed ID: 98445375
- 14 Wilson, R.J., Williamson, D.H. and Preiser, P. (1994) Malaria and other Apicomplexans: the “plant” connection. *Infect Agents Dis* 3, 29-37, PubMed ID: 95040703
- 15 McFadden, G.I. et al. (1996) Plastid in human parasites. *Nature* 381, 482, PubMed ID: 96232405
- 16 McFadden, G.I. and Roos, D.S. (1999) Apicomplexan plastids as drug targets. *Trends Microbiol* 7, 328-333, PubMed ID: 99362812
- 17 Jones, T.C., Yeh, S. and Hirsch, J.G. (1972) The interaction between *Toxoplasma gondii* and mammalian cells. I. Mechanism of entry and intracellular fate of the parasite. *J Exp Med* 136, 1157-1172, PubMed ID: 73030933
- 18 Carruthers, V.B. and Sibley, L.D. (1997) Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *Eur J Cell Biol* 73, 114-123, PubMed ID: 97351942
- 19 Coppens, I. and Joiner, K.A. (2000) Parasite–host-cell interactions in toxoplasmosis: new avenues for intervention? *Exp Rev Mol Med* 15 January, <http://www-ermm.cbcu.cam.ac.uk/01002277h.htm>
- 20 Jones, T.C. and Hirsch, J.G. (1972) The interaction between *Toxoplasma gondii* and mammalian cells. II. The absence of lysosomal fusion with phagocytic vacuoles containing living parasites. *J Exp Med* 136, 1173-1194, PubMed ID: 73030934
- 21 Sibley, L.D., Weidner, E. and Krahenbuhl, J.L. (1985) Phagosome acidification blocked by intracellular *Toxoplasma gondii*. *Nature* 315, 416-419, PubMed ID: 85213866
- 22 Sinai, A.P. and Joiner, K.A. (1997) Safe haven: the cell biology of nonfusogenic pathogen vacuoles. *Annu Rev Microbiol* 51, 415-462, PubMed ID: 98001521
- 23 Mordue, D.G. et al. (1999) *Toxoplasma gondii* resides in a vacuole that avoids fusion with host

- cell endocytic and exocytic vesicular trafficking pathways. *Exp Parasitol* 92, 87-99, PubMed ID: 99296818
- 24 Remington, J.S. and Desmonts, G. (1995) Toxoplasmosis. In *Infectious Diseases of the Fetus and Newborn Infant* (4th edn) (Remington, J.S. and Klein, J.O., eds), pp. 140-267, W.B. Saunders, Philadelphia, PA, USA
- 25 Desmonts, G. and Couvreur, J. (1974) Congenital toxoplasmosis. A prospective study of 378 pregnancies. *N Engl J Med* 290, 1110-1116, PubMed ID: 74138485
- 26 Jacobs, L., Remington, J.S. and Melton, M.L. (1960) The resistance of the encysted form of *Toxoplasma gondii*. *J Parasitol* 46, 11-12
- 27 Darde, M.L., Bouteille, B. and Pestre-Alexandre, M. (1992) Isoenzyme analysis of 35 *Toxoplasma gondii* isolates and the biological and epidemiological implications. *J Parasitol* 78, 786-794, PubMed ID: 93019905
- 28 Sibley, L.D. and Boothroyd, J.C. (1992) Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature* 359, 82-85, PubMed ID: 92396222
- 29 Bulow, R. and Boothroyd, J.C. (1991) Protection of mice from fatal *Toxoplasma gondii* infection by immunization with p30 antigen in liposomes. *J Immunol* 147, 3496-3500, PubMed ID: 92043720
- 30 Parmley, S.F. et al. (1994) Two alleles of the gene encoding surface antigen P22 in 25 strains of *Toxoplasma gondii*. *J Parasitol* 80, 293-301, PubMed ID: 94210284
- 31 Pfefferkorn, L.C. and Pfefferkorn, E.R. (1980) *Toxoplasma gondii*: genetic recombination between drug resistant mutants. *Exp Parasitol* 50, 305-316, PubMed ID: 81044354
- 32 Janse, C.J. (1993) Chromosome size polymorphism and DNA rearrangements in *Plasmodium*. *Parasitol Today* 9, 19-22
- 33 Nagel, S.D. and Boothroyd, J.C. (1988) The alpha- and beta-tubulins of *Toxoplasma gondii* are encoded by single copy genes containing multiple introns. *Mol Biochem Parasitol* 29, 261-273, PubMed ID: 88318803
- 34 Soldati, D. and Boothroyd, J.C. (1995) A selector of transcription initiation in the protozoan parasite *Toxoplasma gondii*. *Mol Cell Biol* 15, 87-93, PubMed ID: 95098043
- 35 Mercier, C. et al. (1996) Common cis-acting elements critical for the expression of several genes of *Toxoplasma gondii*. *Mol Microbiol* 21, 421-428, PubMed ID: 97011585
- 36 Nakaar, V. et al. (1998) Upstream elements required for expression of nucleoside triphosphate hydrolase genes of *Toxoplasma gondii*. *Mol Biochem Parasitol* 92, 229-239, PubMed ID: 98319410
- 37 Burg, J.L. et al. (1988) Molecular analysis of the gene encoding the major surface antigen of *Toxoplasma gondii*. *J Immunol* 141, 3584-3591, PubMed ID: 89035485
- 38 Kim, K., Soldati, D. and Boothroyd, J.C. (1993) Gene replacement in *Toxoplasma gondii* with chloramphenicol acetyltransferase as selectable marker. *Science* 262, 911-914, PubMed ID: 94053736
- 39 Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) Promoter-specific activation of RNA polymerase II transcription by SP1. *Trends Biochem Sci* 11, 20-23
- 40 Couvreur, G. et al. (1988) Surface antigens of *Toxoplasma gondii*. *Parasitology* 97, 1-10, PubMed ID: 89016123
- 41 Tomavo, S., Schwarz, R.T. and Dubremetz, J.F. (1989) Evidence for glycosyl-phosphatidylinositol anchoring of *Toxoplasma gondii* major surface antigens. *Mol Cell Biol* 9, 4576-4580, PubMed ID: 90066474
- 42 Tomavo, S. (1996) The major surface proteins of *Toxoplasma gondii*: structures and functions. *Curr Top Microbiol Immunol* 219, 45-54, PubMed ID: 96383527
- 43 Lecordier, L. et al. (1995) Characterization of a dense granule antigen of *Toxoplasma gondii* (GRA6) associated to the network of the parasitophorous vacuole. *Mol Biochem Parasitol* 70, 85-94, PubMed ID: 95364854
- 44 Hehl, A., Krieger, T. and Boothroyd, J.C. (1997) Identification and characterization of SRS1, a *Toxoplasma gondii* surface antigen upstream of and related to SAG1. *Mol Biochem Parasitol* 89, 271-282, PubMed ID: 98030253
- 45 Soete, M. et al. (1993) *Toxoplasma gondii*: kinetics of bradyzoite-tachyzoite interconversion in vitro. *Exp Parasitol* 76, 259-264, PubMed ID: 93272912
- 46 Manger, I.D. et al. (1998) Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: identification of developmentally regulated genes. *Infect Immun* 66, 1632-1637, PubMed ID: 98187940
- 47 Yang, S. and Parmley, S.F. (1997) *Toxoplasma gondii* expresses two distinct lactate dehydrogenase homologous genes during its life cycle in intermediate hosts. *Gene* 184, 1-12, PubMed ID: 97169139

- 48 Altschul, S.F. et al. (1990) Basic local alignment search tool. *J Mol Biol* 215, 403-410, PubMed ID: 91039304
- 49 Achbarou, A. et al. (1991) Characterization of microneme proteins of *Toxoplasma gondii*. *Mol Biochem Parasitol* 47, 223-233, PubMed ID: 92049555
- 50 Fourmaux, M.N. et al. (1996) *Toxoplasma gondii* microneme proteins: gene cloning and possible function. *Curr Top Microbiol Immunol* 219, 55-58, PubMed ID: 96383528
- 51 Tomley, F.M. et al. (1991) Sequence of the gene encoding an immunodominant microneme protein of *Eimeria tenella*. *Mol Biochem Parasitol* 49, 277-288, PubMed ID: 92131064
- 52 Robson, K.J. et al. (1988) A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. *Nature* 335, 79-82, PubMed ID: 88318952
- 53 Wan, K.L. et al. (1997) Molecular characterisation of an expressed sequence tag locus of *Toxoplasma gondii* encoding the micronemal protein MIC2. *Mol Biochem Parasitol* 84, 203-214, PubMed ID: 97237562
- 54 Manger, I.D., Hehl, A.B. and Boothroyd, J.C. (1998) The surface of *Toxoplasma tachyzoites* is dominated by a family of glycosylphosphatidylinositol-anchored antigens related to SAG1. *Infect Immun* 66, 2237-2244, PubMed ID: 98234057
- 55 Kasper, L.H. and Khan, I.A. (1993) Role of P30 in host immunity and pathogenesis of *T. gondii* infection. *Res Immunol* 144, 45-48, PubMed ID: 93197672
- 56 Dzierszinski, F. et al. (2000) Targeted disruption of the glycosylphosphatidylinositol-anchored surface antigen SAG3 gene in *Toxoplasma gondii* decreases host cell adhesion and drastically reduces virulence in mice. *Mol Microbiol* 37, 574-582, PubMed ID: 20392212
- 57 Waller, R.F. et al. (2000) Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *Embo J* 19, 1794-1802, PubMed ID: 20237616
- 58 Roos, D.S. et al. (1999) *Toxoplasma* as a model parasite: apicomplexan biochemistry, cell biology, molecular genetics, genomics...and beyond. In *Biology of Parasitism* (Tschudi, C. and Pearce, E., eds), pp. 143-167, Kluwer Academic Publishers, Boston, MA, USA
- 59 Lang, M., Apt, K.E. and Kroth, P.G. (1998) Protein transport into "complex" diatom plastids utilizes two different targeting signals. *J Biol Chem* 273, 30973-30978, PubMed ID: 99030358
- 60 Martin, W. and Herrmann, R.G. (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol* 118, 9-17
- 61 Blanchard, J.L. and Lynch, M. (2000) Organellar genes: why do they end up in the nucleus? *Trends Genet* 16, 315-320, PubMed ID: 20317270
- 62 Herman, L. et al. (1990) Plant chromosome/ marker gene fusion assay for study of normal and truncated T-DNA integration events. *Mol Gen Genet* 224, 248-256, PubMed ID: 91117192
- 63 Baker, A. and Schatz, G. (1987) Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. *Proc Natl Acad Sci U S A* 84, 3117-3121, PubMed ID: 87204088
- 64 Gantt, J.S. et al. (1991) Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *Embo J* 10, 3073-3078, PubMed ID: 92007753
- 65 Long, M. et al. (1996) Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. *Proc Natl Acad Sci U S A* 93, 7727-7731, PubMed ID: 96353885
- 66 Figueroa, P. et al. (1999) Transfer of rps14 from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J* 18, 601-609, PubMed ID: 99348165
- 67 Kubo, N. et al. (1999) A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. *Proc Natl Acad Sci U S A* 96, 9207-9211, PubMed ID: 99362736
- 68 Martin, W. et al. (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393, 162-165
- 69 McFadden, G.I. et al. (1997) Plastids in apicomplexan parasites. In *Origins of Algae and Their Plastids* (Bhattacharya, D., ed.), pp. 261-287, Springer-Verlag, New York, NY, USA
- 70 Roos, D.S. et al. (1999) Origin, targeting, and function of the apicomplexan plastid. *Curr Opin Microbiol* 2, 426-432, PubMed ID: 99390331
- 71 Gardner, M.J. et al. (1988) Mitochondrial DNA of the human malarial parasite *Plasmodium falciparum*. *Mol Biochem Parasitol* 31, 11-17, PubMed ID: 89040029

- 72 Gardner, M.J. et al. (1991) Organisation and expression of small subunit ribosomal RNA genes encoded by a 35-kilobase circular DNA in *Plasmodium falciparum*. *Mol Biochem Parasitol* 48, 77-88, PubMed ID: 92140458
- 73 Gardner, M.J. et al. (1993) Sequence and organization of large subunit rRNA genes from the extrachromosomal 35 kb circular DNA of the malaria parasite *Plasmodium falciparum*. *Nucleic Acids Res* 21, 1067-1071, PubMed ID: 93219063
- 74 Gardner, M.J., Williamson, D.H. and Wilson, R.J. (1991) A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. *Mol Biochem Parasitol* 44, 115-123, PubMed ID: 91187055
- 75 Palmer, J.D. (1992) Comparison of chloroplasts and mitochondrial genome evolution in plants. In *Plant Gene Research. Cell Organelles* (Herrmann, H., ed.), pp. 137-163, Springer, Vienna, Austria
- 76 Siddall, M.E. (1992) Hohlzylinder. *Parasitol Today* 8, 90-91
- 77 Kohler, S. et al. (1997) A plastid of probable green algal origin in Apicomplexan parasites. *Science* 275, 1485-1489, PubMed ID: 97197911
- 78 Delwiche, C.F. and Palmer, J.D. (1997) The origins of plastids and their spread via secondary endosymbiosis. In *Origins of Algae and Their Plastids* (Bhattacharya, D., ed.), pp. 53-86, Springer-Verlag, New York, NY, USA
- 79 Barta, J.R., Jenkins, M.C. and Danforth, H.D. (1991) Evolutionary relationships of avian *Eimeria* species among other Apicomplexan protozoa: monophyly of the apicomplexa is supported. *Mol Biol Evol* 8, 345-355, PubMed ID: 91304310
- 80 McFadden, G.I. and Hill, D. (1994) *Goniomonas*: rRNA sequences indicate that this phagotrophic flagellate is a close relative of the host component of cryptomonads. *Eur J Phycol* 29, 29-32
- 81 Cavalier-Smith, T. et al. (1996) Cryptomonad nuclear and nucleomorph 18S rRNA phylogeny. *Eur J Phycol* 31, 315-328
- 82 Palmer, J.D. and Delwiche, C.F. (1996) Second-hand chloroplasts and the case of the disappearing nucleus. *Proc Natl Acad Sci U S A* 93, 7432-7435, PubMed ID: 96353833
- 83 Sato, S., Tews, I. and Wilson, R.J. (2000) Impact of a plastid-bearing endocytobiont on apicomplexan genomes. *Int J Parasitol* 30, 427-439, PubMed ID: 20198189
- 84 Fichera, M.E. and Roos, D.S. (1997) A plastid organelle as a drug target in apicomplexan parasites. *Nature* 390, 407-409, PubMed ID: 98049349
- 85 Williamson, D.H. et al. (1996) Organelle DNAs: the bit players in malaria parasite DNA replication. *Parasitol Today* 19, 357-362
- 86 Dunn, P.P., Stephens, P.J. and Shirley, M.W. (1998) *Eimeria tenella*: two species of extrachromosomal DNA revealed by pulsed-field gel electrophoresis. *Parasitol Res* 84, 272-275, PubMed ID: 98228045
- 87 Gleeson, M.T. and Johnson, A.M. (1999) Physical characterisation of the plastid DNA in *Neospora caninum*. *Int J Parasitol* 29, 1563-1573, PubMed ID: 20074142
- 88 Weissig, V., Vetro-Widenhouse, T.S. and Rowe, T.C. (1997) Topoisomerase II inhibitors induce cleavage of nuclear and 35-kb plastid DNAs in the malarial parasite *Plasmodium falciparum*. *DNA Cell Biol* 16, 1483-1492, PubMed ID: 98089073
- 89 Pfefferkorn, E.R., Nothnagel, R.F. and Borotz, S.E. (1992) Parasitocidal effect of clindamycin on *Toxoplasma gondii* grown in cultured cells and selection of a drug-resistant mutant. *Antimicrob Agents Chemother* 36, 1091-1096, PubMed ID: 92378186
- 90 Gray, M.W. and Lang, B.F. (1998) Transcription in chloroplasts and mitochondria: a tale of two polymerases. *Trends Microbiol* 6, 1-3, PubMed ID: 98142813
- 91 Clough, B. et al. (1997) Thiostrepton binds to malarial plastid rRNA. *FEBS Lett* 406, 123-125, PubMed ID: 97263485
- 92 Araujo, F.G., Shepard, R.M. and Remington, J.S. (1991) In vivo activity of the macrolide antibiotics azithromycin, roxithromycin and spiramycin against *Toxoplasma gondii*. *Eur J Clin Microbiol Infect Dis* 10, 519-524, PubMed ID: 92007908
- 93 Dannemann, B. et al. (1992) Treatment of toxoplasmic encephalitis in patients with AIDS. A randomized trial comparing pyrimethamine plus clindamycin to pyrimethamine plus sulfadiazine. The California Collaborative Treatment Group. *Ann Intern Med* 116, 33-43, PubMed ID: 92074708
- 94 Weeden, N.F. (1981) Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J Mol Evol* 17, 133-139, PubMed ID: 81267439
- 95 Rohmer, M. et al. (1993) Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem J* 295, 517-524, PubMed ID: 94058991

- 96 Vial, H.J. and Ancelin, M.L. (1992) Malarial lipids. An overview. *Subcell Biochem* 18, 259-306, PubMed ID: 93134652
- 97 Jomaa, H. et al. (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* 285, 1573-1576, PubMed ID: 99407027
- 98 Haslam, E. (1993) *Shikimic Acid: Metabolism and Metabolites*, John Wiley, Chichester, UK
- 99 Roberts, F. et al. (1998) Evidence for the shikimate pathway in apicomplexan parasites [published erratum appears in *Nature* 1998 Sep 17; 395 (6699): 801]. *Nature* 393, 801-805, PubMed ID: 98318102
- 100 Keeling, P.J. et al. (1999) Shikimate pathway in apicomplexan parasites. *Nature* 397, 219-220, PubMed ID: 99127886
- 101 Roberts, C.W. et al. (1999) Shikimate pathway in apicomplexan parasites (reply). *Nature* 397, 220
- 102 Kishore, G.M. and Shah, D.M. (1988) Amino acid biosynthesis inhibitors as herbicides. *Annu Rev Biochem* 57, 627-663, PubMed ID: 89024585

### Further reading, resources and contacts

The Centers for Disease Control and the NIH MEDLINEplus websites provide comprehensive general information and links for toxoplasmosis.

<http://www.cdc.gov/ncidod/dpd/parasites/toxoplasmosis/default.htm>

<http://www.nlm.nih.gov/medlineplus/toxoplasmosis.html>

The following websites provide specific information about the prevention of toxoplasmosis in AIDS/HIV patients and the prevention of congenital toxoplasmosis.

<http://www.thebody.com/treat/toxo.html>

<http://www.cdc.gov/mmwr//preview/mmwrhtml/rr4902a5.htm>

A tutorial based on the *Toxoplasma gondii* database can be found at

[http://www.cbil.upenn.edu/ParaDBs/Toxoplasma/tutorial/tut\\_1.html](http://www.cbil.upenn.edu/ParaDBs/Toxoplasma/tutorial/tut_1.html)

Boris Striepen's Apicoplast Page (mentioned in the text above) has details on the biology of the apicoplast, as well as a QuickTime movie of apicoplast division.

<http://www.sas.upenn.edu/~striepen/gfp.html>

### Features associated with this article

#### Figures

Figure 1. Ultrastructure of a *Toxoplasma gondii* tachyzoite (fig001jac).

Figure 2. The *Toxoplasma gondii* life cycle (fig002jac).

Figure 3. The *Toxoplasma gondii* chromosomes (fig003jac).

#### Tables

Table 1. Upstream repeated motif sequences of *Toxoplasma gondii* genes (tab001jac).

Table 2. Summary of abundant expressed sequence tags (ESTs) from both RH (total ESTs 5359) and ME49 (total ESTs 1806) tachyzoite cDNA libraries (tab002jac).

Table 3. Summary of abundant expressed sequence tags (ESTs) from an ME49 tachyzoite cDNA library (total ESTs 1806) and an ME49 bradyzoite cDNA library (total ESTs 2353) (tab003jac).

### Citation details for this article

James W. Ajioka, Jennifer M. Fitzpatrick and Christopher P. Reitter (2001) *Toxoplasma gondii* genomics: shedding light on pathogenesis and chemotherapy. *Exp. Rev. Mol. Med.* 6 January, <http://www-ermm.cbcu.cam.ac.uk/01002204h.htm>