Review

Genetic diversity of Toxoplasma gondii in animals and humans

L. David Sibley1,*, Asis Khan1, James W. Ajioka2 and Benjamin M. Rosenthal3

1Department of Molecular Microbiology, Washington University School of Medicine, 660 South Euclid Avenue, Saint Louis, MO 63130, USA
2Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK
3Animal Parasitic Disease Laboratory, Animal and Natural Resources Institute, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705, USA

Toxoplasma gondii is one of the most widespread parasites of domestic, wild, and companion animals, and it also commonly infects humans. Toxoplasma gondii has a complex life cycle. Sexual development occurs only in the cat gut, while asexual replication occurs in many vertebrate hosts. These features combine to create an unusual population structure. The vast majority of strains in North America and Europe fall into three recently derived, clonal lineages known as types I, II and III. Recent studies have revealed that South American strains are more genetically diverse and comprise distinct genotypes. These differences have been shaped by infrequent sexual recombination, population sweeps and biogeography. The majority of human infections that have been studied in North America and Europe are caused by type II strains, which are also common in agricultural animals from these regions. In contrast, several diverse genotypes of T. gondii are associated with severe infections in humans in South America. Defining the population structure of T. gondii from new regions has important implications for transmission, immunogenicity and pathogenesis.

Keywords: population genetics; clonality; pathogenesis; recombination; toxoplasmosis

1. LIFE CYCLE OF TOXOPLASMA GONDII

Toxoplasma gondii is an obligate intracellular parasite that can infect virtually all warm-blooded animals. There are three different infectious stages: tachyzoites, which facilitate expansion during acute infection, bradyzoites, which maintain chronic infection, and sporozoites, which are disseminated in the environment within oocysts (Dubey et al. 1998; figure 1). All three stages are haploid; tachyzoites and bradyzoites divide asexually, while sporozoites are the product of meiosis. Sexual development only occurs within enterocytes of the cat gut, ultimately yielding diploid oocysts, which undergo meiosis after shedding (figure 1). Understanding the adaptations of these stages for various steps in the life cycle provides a framework for considering the unique population structure of T. gondii.

Tachyzoites multiply very rapidly in a wide variety of nucleated host cells during the acute phase of infection. Parasite invasion is driven by actin-based motility, generating a parasitophorous vacuole derived from invagination of the host cell plasma membrane and secretion of parasite proteins (Sibley et al. 2007). Within the parasitophorous vacuole, tachyzoites divide every 6–9 h by a process called endodyogeny, in which daughter cells form internally within the mother cell (Morrissette & Sibley 2002). Rupture of the host cell leads to emergence of parasites that infect new host cells. Infection evokes strong innate and adaptive immune responses that control parasite replication but do not eliminate the infection. In response to environmental stress, tachyzoites convert into a semidormant stage known as bradyzoites, which are contained within tissue cysts (Dubey & Frenkel 1976). Tissue cysts form in a variety of cells, especially long-lived differentiated cells such as neurons and muscle cells, thus assuring long-term infection (Dubey 1997b). Histological evidence suggests that cysts turnover slowly in vivo, releasing bradyzoites into the surrounding tissue (Frenkel & Escajadillo 1987; Ferguson et al. 1989). The subsequent inflammatory and cellular immune response contain the infection, although cyst rupture also gives rise to daughter cysts. Following cyst rupture, conversion to tachyzoites can result in reactivation of latent infection, for example in immunocompromised mice (Suzuki et al. 1989; Suzuki & Joh 1994). Similarly, reactivation of latent infection results in toxoplasmic encephalitis, an important cause of opportunistic disease in immunodeficient patients (i.e. AIDS, transplant, chemotherapy; Joynson & Wreghitt 2001).

While a wide variety of vertebrates can serve as intermediate hosts for T. gondii (figure 2), sexual

* Author for correspondence (sibley@borcim.wustl.edu).

One contribution of 12 to a Theme Issue ‘Livestock diseases and zoonoses’.

This journal is © 2009 The Royal Society
Development only occurs in members of the felidae (Dubey 1977). The sexual cycle takes place exclusively within intestinal enterocytes, eventually giving rise to unsporulated oocysts, which are shed in the faeces in high numbers (Dubey & Frenkel 1972). There is no fixed mating type as a single cloned organism can differentiate to complete the entire life cycle (Pfefferkorn et al. 1977; Cornelissen & Overduin 1985). Cats shed oocysts from days 4–11 post-infection and thereafter they retain a high degree of immunity (Dubey & Frenkel 1972; Frenkel & Smith 1982). Sporulated oocysts survive for long periods under moderate environmental conditions and they resist destruction by physical forces and common chemical disinfectants (Frenkel et al. 1975; Dubey 1998b). As such, oocysts are the agents for widespread environmental contamination.

2. TRANSMISSION AND HOST RANGE

Interconversion between different developmental stages occurs readily in the T. gondii life cycle (Dubey 2007). Tachyzoites convert to bradyzoites to induce chronic infections, while the reverse process results in reactivation (figure 1). Because of this flexibility, tissue cysts can readily infect intermediate hosts following oral ingestion (Dubey 1997a; figure 1). Closely related parasites, such as Neospora caninum, lack this efficient means of oral transmission between intermediate hosts, suggesting it arose since their common ancestry around 10 Myr ago (Su et al. 2003a).

Despite flexibility in differentiation of T. gondii, tissue cysts are less infectious for intermediate host than for felines. For example, the infectious dose for mice is approximately 1000 bradyzoites, while it is much lower for cats (Dubey 2001). In contrast, oocysts are efficiently adapted for transmission to intermediate hosts (Dubey et al. 1996, 1997). Surveys have indicated that nearly all the major orders of placental mammals can serve as intermediate hosts (Dubey 1977, 2007; Dubey & Beattie 1988a, b), notable exceptions include fruit-eating bats, insectivores and baleen whales (figure 2). Although strict herbivores are primarily infected by oocysts, oral transmission via ingestion of tissue cysts probably contributes to the spread of T. gondii infection through the food chain owing to carnivorous or omnivorous feeding (figure 1). These different modes of transmission may contribute to the extremely broad host range of T. gondii (figure 2).

Infection of intermediate hosts is typically subclinical; however, it can be an important cause of abortion in sheep (Owen & Trees 1999) where vertical transmission between successive generations has been suggested (Hide et al. 2007). Owing to economic losses, T. gondii has been the focus of a vaccine effort to reduce abortion in sheep and lower the burden of tissue cysts in food animals (Buxton et al. 2007). Infection with T. gondii also has remarkable effects on the behaviour of rodents, causing them to lose their fear response (Webster 2001). This response appears to be highly specific to aversion of cat odours, thereby potentially enhancing transmission (Vyas et al. 2007).

3. PREVALENCE OF TOXOPLASMA GONDII IN THE FOOD CHAIN

Detecting chronic infections with T. gondii relies primarily on serological techniques, such as the
Figure 2. *Toxoplasma gondii* infects a wide range of placental mammals including at least some members of the major orders highlighted here. *Toxoplasma gondii* also infects bird and marsupials, making it one of the most widespread parasites of vertebrates. Adapted from Romer & Parsons (1977).

modified agglutination test (Dubey & Desmonts 1987). Serological testing has been extensively applied to free-ranging chickens, based on the rationale that they provide an index of environmental contamination with oocysts (Dubey et al. 2003b). Sampling studies in Central and South America show prevalence rates ranging from 40 to 60 per cent (Dubey et al. 2005a, 2006a,b,c,d). More limited studies indicate prevalence rates of approximately 20–40% in India (Devada et al. 1998; Sreekumar et al. 2003) and around 40 per cent in Egypt (El-Massry et al. 2000; Dubey et al. 2003a). Pigs are highly susceptible to experimental infection by oocysts (Dubey et al. 1996), and market pigs show high prevalence rates in some regions of the USA (Gamble et al. 1999; Dubey et al. 2002). Sheep are also commonly infected (Buxton et al. 2007) and are thought to be a major source of human exposure globally. Beef is probably not a source of significant infection as cows are relatively resistant to experimental infection (Dubey & Thulliez 1993). Surveys of wild animals also indicate high prevalence rates (Hill et al. 2005), and these may also serve as a source of human infection. Serologically positive animals are probably infected for life and harbour tissue cysts that present a risk of human infection owing to consumption of undercooked meat. One recent survey of meat from commercial markets in the USA suggests the risk is low, perhaps owing to freezing of meat during processing, which reduces the viability of cysts (Dubey et al. 2005b). Companion animals also show high prevalence rates, including both cats and dogs (Dubey & Beattie 1988). Cats pose a risk to humans during the initial infection when they actively shed oocysts. Although dogs do not actively transmit infection, they also pose a risk owing to mechanical transport of oocysts on their fur (Frenkel et al. 2003).

Mice and cats are extremely susceptible to infection and their use in bioassays is one of the most sensitive means of detecting chronic infection in animal tissues. Following *in vitro* digestion with acid pepsin, tissues are inoculated into mice and survivors followed for the development of illness, sero-conversion and establishment of chronic infection (Dubey 1998a). Cats have also been used extensively for the isolation of *T. gondii* strains by feeding tissue samples to cats and then examining the faeces for shedding of oocysts from days 3 to 14 post-inoculation (Dubey et al. 1972). The large number of isolates that have been obtained by these approaches has been useful for profiling the genetic makeup and biological traits of strains. However, it is unknown if some natural isolates are refractory growth in mice and hence under-represented in genetic analyses described below.

4. TOXOPLASMOsis IN HUMANS

Toxoplasmosis is a true zoonotic infection, being acquired by humans from domestic, wild or companion animals. Serological studies indicate a remarkably high prevalence that exceeds 50 per cent of adults in many regions (Hall et al. 2001). Humans acquire *T. gondii* infection by ingestion or handling of undercooked or raw meat containing tissue cysts or by ingestion of water or food contaminated with oocysts excreted in the faeces of infected cats (Frenkel 1973). Outbreaks have been associated with both foodborne (Choi et al. 1997; Mead et al. 1999; Dawson 2005) and waterborne sources (Benenson et al. 1982; Bowie et al. 1997; Bahia-Oliveira et al. 2003; de Moura et al. 2006). Epidemiological data indicate that exposure to oocysts (contaminated soil or water) and undercooked meat are the two primary risk factors for human infection (Jones et al. 2001). Humans do not play a major role in transmission; consequently, pathogenesis in humans is the indirect result of adaptations for infection in other hosts and treatment of human infections is unlikely to lead to the spread of drug resistance.

Phil. Trans. R. Soc. B (2009)
Primary infections are usually limited in duration and typified by fever, swollen lymph nodes and muscle weakness; they are typically controlled by the immune system and rarely require treatment (Ho-Yen 2001). Antibiotic treatment is effectively provided by pyrimethamine combined with sulfadiazine, and several macrolides are also effective (McCabe 2001). Primary infections give rise to semidormant tissue cysts, which are not eliminated by antibiotic treatment. Consequently, chronic infections predispose individuals to the risk of reactivation. In most cases, long-term infections remain clinically inapparent. However, several studies have found an association between seropositivity and atypical psychiatric disease (Torrey et al. 2007).

Primary infection during pregnancy also possesses a risk of congenital infection when it occurs in a previously immunologically naive mother (Wong & Remington 1994; McLeod et al. 2000). Congenital toxoplasmosis leads to a wide range of clinical manifestations from mild symptoms, visual defects or development of retinochoroiditis in later life, to more severe cases of hydrocephalus, microcephaly seizures, mental retardation and even foetal death (Wong & Remington 1994; McLeod et al. 2000). Previously infected mothers develop strong immunity and are resistant to reinfection, and thus rarely transmit during subsequent pregnancies.

Ocular disease typically results when tissue cysts rupture and cause inflammation in the retina, normally a site of immune privilege. It is now appreciated that a large fraction of ocular disease is postnatally acquired (Holland 2003, 2004). Ocular toxoplasmosis can be treated with corticosteroids and antibiotic therapy, although this treatment does not eradicate infection. In southern regions of Brazil, high levels of recurrent and serious ocular disease have been described in otherwise healthy adults, many of whom acquired the infection postnatally (Silveira et al. 2001; Jones et al. 2006).

5. METHODS FOR GENOTYPING

A variety of different methods exist to genotype T. gondii isolates, and they have distinct advantages and disadvantages. Early studies of strain typing were based on multilocus enzyme electrophoresis (MLE). Several polymorphic enzymes were used to characterize T. gondii isolates largely collected from France, grouping them into three major zymodemes Z1, Z2 and Z3 (Darde et al. 1992). While MLE is quite specific, it requires a large number of purified parasites to perform. Thus, later typing methods focused on microsatellite (MS) markers (Ajzenberg et al. 2002a,b), which are short repeated segments of DNA that tend to occur in non-coding DNA. MS markers are sensitive, reliable and amenable to high-throughput analyses. Although MSs are highly polymorphic, they can be prone to homoplasy as the number of repeats can expand and contract during replication. Randomly amplified polymorphic DNA has been used to characterize strains of T. gondii (Guo & Johnson 1995; Ferreira et al. 2004); however, this technique is highly influenced by contaminating host DNA, which is a significant source of variability.

Restriction fragment length polymorphism (RFLP) analysis of specific genetic loci has been widely used for T. gondii genotyping. RFLP markers are also amenable to high-throughput analysis using PCR amplification, followed by restriction digestion and gel electrophoresis. Multilocus RFLP analysis has been used to characterize isolates collected primarily from North America and Europe, grouping them into three major groups referred to as I, II and III (Sibley & Boothroyd 1992; Howe & Sibley 1995; Howe et al. 1997). Based on the highly clonal population structure in these regions, it was previously suggested that typing based on nested-PCR analysis of RFLPs at the SAG2 locus was sufficient for genotyping clinical samples (Howe et al. 1997). Although this single marker accurately genotypes the clonal lineages, it is not capable of detecting recombinant or exotic strains. Hence, it is only suitable in locales where the clonal lineages predominate. To improve methods for typing mixed, recombinant or unusual strains, a combination of loci have been used to analyse genotypes by RFLP-PCR typing (Ferreira et al. 2006, 2008; Su et al. 2006). Importantly, RFLP markers are limited to capturing changes that alter restriction enzyme sites, and many polymorphisms are missed by this analysis. Methods for detecting otherwise silent mutations have also been developed based on PCR-based priming from the site of polymorphism (Su et al. 2003b). While this method can potentially screen a larger number of sites, it requires a significant investment of time to develop and optimize each marker.

All of the above methods underestimate the true rate of polymorphism and hence may misclassify variants owing to homoplasy or insufficient resolving power. In contrast, direct sequencing of genomic regions reveals the complete genetic diversity including single nucleotide polymorphisms (SNPs) and small insertions and deletions (e.g. indels). Direct sequencing generally detects much greater genetic diversity than other methods. For example, a high degree of polymorphism was observed at the GRA6 locus by sequencing (nine allelic sequences from 30 strains), whereas the PCR-RFLP analysis only detected three groups (Pazaeli et al. 2000). Thus, sequence-based methods provide the best approach for detecting polymorphisms in new isolates or from previously unsampled populations. A variety of different loci have been used for sequence-based analysis, including both coding regions for housekeeping genes, antigens and selectively neutral introns (Khan et al. 2007).

Highly polymorphic antigens provide maximum resolution for detecting recent divergence within populations. In contrast, selectively neutral regions provide the best source of data to calculate the age of common ancestry between different lineages and to predict common ancestry (Khan et al. 2007). The obvious disadvantage of sequence-based typing is its increased cost and need for access to sophisticated technology. However, once the population structure is known for a given region, more cost-effective typing methods can be developed to detect the major alleles (i.e. MS or RFLP typing).
6. DETECTION OF CHRONIC INFECTIONS
The above methods of typing are based on isolation and growth of the parasite in vitro, which is time consuming and may result in sampling bias. Thus far, efforts to detect and type strains directly in clinical samples have met with only partial success. Sensitive methods have been developed for PCR-based detection, typically relying on repetitive genes such as B1 (Burg et al. 1989) or other repeats (Homan et al. 2000). PCR-based detection is typically limited to acute or reactivated infections (Grover et al. 1990; Jones et al. 2000), probably reflecting the low tissue burden and the absence of circulating parasites in the blood. Genotyping of such acute clinical specimens typically requires nested PCR for enhanced sensitivity, and this has been applied to a variety of samples, including blood (Khan et al. 2006b), cerebral spinal fluid (Khan et al. 2005a), amniotic fluid (Nowakowska et al. 2006) and ocular samples (Grigg & Boothroyd 2001; Grigg et al. 2001b). Real-time PCR detection of T. gondii in mouse and pig tissues has been developed based on the repetitive ITS-1 sequence (Jauregui et al. 2001). While such methods provide excellent sensitivity for laboratory samples, they have generally been less useful for analysing naturally infected tissue owing to the low tissue burden of parasites. One exception to this is the report of detection of T. gondii in market samples of meat in Britain (Aspinall et al. 2002). This study also reported frequent multiple infections, something that has been reported in only a few other studies (Dubey et al. 2006d; Lindstrom et al. 2008). Single infections are also typically observed when isolates are first obtained by inoculation of mice (Howe & Sibley 1995; Ajzenberg et al. 2002b; Pena et al. 2008). This may reflect differences in the ability of some strains to grow in mice, but given the high susceptibility of mice, it is more likely the result of low density of parasites in the tissue and resulting sampling issues. With further improvements in specificity and sensitivity of PCR, it might be routinely possible to directly detect T. gondii in naturally infected tissues.

Serotyping is currently limited by the close similarity of T. gondii strains, which share most major antigenic determinants. Recognizing that some epitopes are both polymorphic and immunodominant, serological typing methods were developed to distinguish type I or III from type II strains using polymorphic peptides for several antigens (Kong et al. 2003). Unfortunately, it is still not possible to reliably separate type I from III or to accurately type strains in regions where other genotypes predominate. Further development of more specific serotyping methods would make it possible to type strains that cause chronic infection in animals and humans.

7. CLONAL POPULATION STRUCTURE IN NORTH AMERICA AND EUROPE
Despite the existence of a well-characterized sexual cycle, the population structure of T. gondii in some regions reflects a high frequency of asexual replication. Early studies using multilocus isoenzyme analyses to analyse isolates of mostly European origin showed remarkably little diversity (Darde´ et al. 1988, 1992). RFLP-PCR analysis with a larger set of isolates from North American and European sources revealed the existence of a striking clonal population structure, with acutely virulent strains comprising a single clonal type (Sibley & Boothroyd 1992). A more exhaustive PCR-RFLP study with a larger group of strains collected from human disease cases and chronic animal infections, largely from North America and Europe, showed that isolates fall into one of just three distinct clonal genotypes known as types I, II and III (Howe & Sibley 1995). Clonality in this study was apparent by the frequent sampling of isolates with identical multilocus genotypes, a high degree of linkage disequilibrium between markers and a relative absence of recombinants. While clonality is the predominant pattern in North America and Europe, recombinant strains representing mixtures of the three clonal types were also occasionally found (approx. 5% in this early study; Howe & Sibley 1995). Moreover, a small number of isolates were highly divergent from the three major lineages, and these were referred to as ‘exotic or atypical genotypes’ (Howe & Sibley 1995).

A sexual transmission is common in many protozoa, some of which appear to lack defined sexual cycles, such as Trypanosoma cruzi and Entamoeba histolytica (Tibayrenc et al. 1990; Tibayrenc & Ayala 1991). However, this cannot account for the clonality seen in T. gondii, as there is a well-developed sexual cycle that is capable of generating many recombinants from a single cross in the cat, as shown by several previous genetic crosses (Khan et al. 2005b). There are several possible explanations for the existence of a clonal population structure of T. gondii in the wild. First, the sexual cycle is not obligatory for transmission, which may also occur by ingestion of intermediate stages through carnivorous or omnivorous feeding (figure 1). Second, a single haploid organism can differentiate to complete the life cycle in the cat. As coinfection with multiple strains is probably a rare event, cats will typically shed oocysts containing progeny that are genetically identical. Finally, once limited genetic complexity is established, it is likely to be maintained.

8. ASSOCIATION OF GENOTYPES WITH HOSTS AND PHENOTYPES
In North America, type II strains are most commonly associated with human toxoplasmosis, both in congenital infections and in patients with AIDS (Howe & Sibley 1995; Howe et al. 1997). Type II strains are also most commonly associated with human infections in Europe (Honoré et al. 2000; Ajzenberg et al. 2002b). Several studies have indicated that the majority of isolates from agricultural animals are also type II, including pigs in the USA (Mondragon et al. 1998; Dubey et al. 2008a) and sheep from Britain (Owen & Trees 1999). Chickens in North America show a higher prevalence of type III strains than type II (Dubey et al. 2003b), consistent with an early survey that indicated both type II and type III strains are common in animals (Howe & Sibley 1995).
reasons for the apparent differences between animal (types II and III) and human (largely type II) infections are unclear but might reflect differences in susceptibility. Alternatively, it might reflect differences in disease-causing potential because human isolates were largely collected from disease cases, while animal infections were largely subclinical. More recent surveys, using an expanded set of RFLP markers, have indicated that while the majority of isolates from sheep in North America were type II, a number of distinct new genotypes were found (Dubey et al. 2008c). Such new variants might arise by somatic mutations, or result from sexual recombination between the major lineages, or they might represent entirely new genotypes. Deciphering these different models would require further sequence-based analysis of haplotypes.

The clonal lineages differ in a number of phenotypes such as growth, migration and transmigration (Barragan & Sibley 2003); however, the best described of these is virulence in laboratory mice (Howe et al. 1996; Mordue et al. 2001). Type I strains cause lethal infection in all strains of laboratory mice even at low inocula (lethal dose (LD100) approx. 1), whereas types II and III strains are much less virulent (median lethal dose (LD50) >10^5). Forward genetic mapping studies have been applied to identify genes that determine natural differences in the virulence of T. gondii in the mouse model. Independent screens converged on secretory proteins discharged from apical organelles, called rhoptries (ROPs), as the key determinant of acute virulence in the mouse model (Saieij et al. 2006; Taylor et al. 2006). Many South American strains are also virulent in murine models (Khan et al. 2007), and similar approaches have demonstrated that ROPs also contribute to acute virulence in these lineages (Khan et al. 2009).

9. ORIGINS AND SPREAD OF CLONALITY IN NORTH AMERICA AND EUROPE

The predominant clonal lineages differ by only 1–2% at the nucleotide level (Grigg et al. 2001a). Coalescence analysis supports a model whereby the three lineages evolved recently from a common ancestor within the last 10 000 years, expanding very rapidly to populate a variety of hosts (Su et al. 2003a). This corresponds to the same time frame as the domestication of agricultural animals, as well as the adoption of companion animals such as domestic cats (Driscoll et al. 2007). Similarly, a number of sexually differentiated parasites of agricultural animals have been found to exhibit low genetic diversity, reflecting bottlenecking and anthropogenic expansions in their recent ancestry (Rosenthal 2009).

Another striking feature of North American and European clonal lineages of T. gondii is the presence of biallelic polymorphisms at each locus. Such biallelism was evident from early RFLP studies and was further underscored by sequencing genes from a variety of isolates (Grigg et al. 2001a). Genome-wide analysis of SNPs based on expressed sequence tags that were generated from the three lineages (Ajjoeka et al. 1998; Li et al. 2003) demonstrated that biallelism extends throughout the genome (Boyle et al. 2006). Mapping these differences to the whole genome assembly revealed long haploblocks, dominated by SNPs that are unique to a given strain. Comparing the patterns of these long haploblocks with the known behaviour of chromosomes in genetic crosses suggests that just a few genetic crosses between several closely related parental strains could explain the origin of the three lineages (Boyle et al. 2006).

The genome of T. gondii bears another signature of recent ancestry and common origin: fixation of a single monomorphic version of chromosome 1a (Chr1a), referred to as Chr1a*, as revealed by comparative genomic sequencing (Khan et al. 2006a). Based on the rate of somatic mutations between isolates within the lineages, it was estimated that Chr1a* arose approximately 10^4 years ago, probably coincident with the origin of the lineages (Khan et al. 2006a). The odds of all three lineages acquiring this same exact Chr1a* by chance have been estimated to be at least 1 : 1000 (Khan et al. 2006a), suggesting this pattern arose owing to a selective advantage.

The population structure of the clonal lineages of T. gondii is best explained by a genetic bottleneck followed by a rapid expansion. This pattern might reflect an unusual combination of genes, shared by the three lineages, that is responsible for their successful expansion. Several adaptive traits might explain this expansion, including evolution of oral transmission between intermediate hosts (Su et al. 2003a), enhanced transmission by domestic cats or adaptations to domestic rodents (Lehmann et al. 2006). Alternatively, the pattern of clonality may simply reflect an unusual demographic process that dramatically restricted the gene pool. Testing these models will require a better understanding of population structure and transmission dynamics in the wild.

10. POPULATION STRUCTURE IN GEOGRAPHICALLY REMOTE REGIONS

While clonality clearly predominates in much of North America and Europe, additional sampling from a wider range of locations led to the suggestion that strains from other regions have more diverse genotypes (Ajzenberg et al. 2004). For example, highly unusual genotypes were detected in individuals who contracted toxoplasmosis while in the jungles of French Guyana (Darde et al. 1998; Carme et al. 2002). More recent sampling in North America also suggests that isolates from wild animals have more diverse genotypes, as tested by multilocus RFLP analysis (Dubey et al. 2008b,d). Isolates of T. gondii in sea otters have been associated with enhanced mortality, and these isolates may also comprise unique genotypes (Miller et al. 2004; Sundar et al. 2008). While these and other studies seem to support the feral-diversity model, we lack sufficient sampling to properly determine whether domestic and wild populations are genetically distinct.

Studies based on RFLP markers, albeit using polymorphisms that were originally defined in the
North, showed that animal isolates of *T. gondii* from Brazil are much more genetically diverse (Ferreira et al. 2001, 2006; Pena et al. 2008), a finding supported by MS markers (Lehmann et al. 2006). This conclusion has subsequently been echoed by numerous other studies of isolates from various animal hosts in Central and South America. Similar genotypes were observed in animals and humans in Brazil (Khan et al. 2006b; Ferreira et al. 2008), suggesting the divergence of isolates in South America is not due to differences in host range but rather due to geographical differences. The prevalence of genetically more diverse strains in South America raised the question of whether they were distinct lineages or merely recombinants of the genotypes that are prevalent in the North.

**11. DEFINING POPULATION STRUCTURE**

Establishing the population structure in new regions is complicated by several factors. First, most markers have been optimized for the detection of informative polymorphisms within existing populations and may fail to capture diversity in new regions. Second, adequate sampling density is required to accurately estimate allele frequencies, which may vary by host, region or owing to population structure. Third, we lack precise definitions of how divergent lineages need be in order to consider them unique. The original concept of clonality does not imply that individual isolates are ‘identical’, merely that they show restricted recombination and high levels of linkage disequilibrium (Tibayrenc et al. 1991, 1993). This is not incompatible with the occasional mutation that generates slight differences in a common allele. Hence, classifying new isolates as ‘unique’ simply because they vary in one particular trait does not adequately describe them in terms of genetic diversity. Certainly, if we were to sequence the genomes of all isolates, they would all be genetically ‘unique’. Even if costs for sequencing fell to the level where this was feasible, we would still need appropriate criteria for classifying strains in a way that allows meaningful comparisons.

Assigning the genotypes of new isolates into appropriate groups requires that the local population structure be defined. Appropriate test criteria have been established to monitor allele frequencies, estimate the extent of genetic diversity and evaluate gene flow between groups and hence define clonality (Arnaud-Haond et al. 2007). Methods exist for clustering isolates into major groups, including network analysis (Templeton et al. 1992), phylogenetic methods (Swofford 2002) and STRUCTURE analysis (Falush et al. 2003). Application of methods for testing epidemic or clonal populations has been established based on bacterial (Smith et al. 1993) and protozoan populations (Tibayrenc & Ayala 1991; Tibayrenc et al. 1991). Appropriate choice of loci for such analysis includes both regions under strong selection (antigens) that provide an index of diversity and regions that are selectively neutral, for estimating common ancestry. Through such comparative analyses, minor variants that might arise from mutation can be distinguished from recombinant genotypes that arise from sexual reproduction or new lineages that are unique to new regions. Understanding these local variations in genetic population structure is highly relevant to traits like transmission, immunogenicity and pathogenesis.

**12. POPULATION DIFFERENCES BETWEEN NORTH AND SOUTH AMERICA**

To characterize the population structure of *T. gondii* in South America, intron sequences from a collection of *T. gondii* isolates from Brazil and French Guyana were compared to previously characterized reference strains from North America (Khan et al. 2007). This study revealed that North and South American populations were reproductively isolated and had evolved separate sets of biallelic polymorphisms in the absence of genetic exchange over a long period of time (Khan et al. 2007). Sequencing also revealed that existing RFLP markers failed to capture diversity that was unique to South America, as these markers had been based on polymorphisms that are unique to North American lineages. Although RFLP markers capture some variation when used on South American strains, they can only detect shared ancestral states and miss most of the recent history of these strains. Continued use of RFLP typing with such markers is likely to result in inaccurate classification and gives little insight into the true population structure.

Phylogenetic reconstruction based on intron sequences clustered *T. gondii* strains into 11 distinct haplogroups representing the major lineages (Khan et al. 2007; figure 3a). These 11 haplogroups were also supported by the Bayesian model for predicting population STRUCTURE (Falush et al. 2003). Three of these groups correspond to the previously recognized clonal lineages in North America and Europe (i.e. I, II and III or 1, 2 and 3). Of the new groups, four were almost exclusively found in South America (i.e. 4, 5, 8 and 9; figure 3a). Previously characterized exotic strains belong to one of the common haplogroups, yet they were isolated outside their usual range. For example, strain P89, which was isolated from a pig on a farm in Iowa, is a member of group 9, a common group largely restricted to South America. The rarity of such misplaced strains suggests that migration has done little to erode the population split between North and South. Although strains in South America are more diverse, there are also pockets of clonality (i.e. groups 8 and 9). Other groups show more deeply branching phylogenies and low levels of linkage disequilibrium, consistent with greater extent of sexual recombination (i.e. group 4; figure 3a).

In comparing the population structure in North America and Europe versus South America using STRUCTURE (Falush et al. 2003), it was estimated that all 11 haplogroups could be reconstructed from only four ancestral lineages by very few genetic crosses (Khan et al. 2007). This is reminiscent of the earlier prediction that the three lineages in North America arose from just a few genetic crosses (Boyle et al. 2006). Hence, occasional recombination has dramatically shaped the population structure of *T. gondii* over a long period of time.

Despite the presence of mutually exclusive biallelic polymorphisms, *T. gondii* strains from North and
South America share a common ancestry. However, they were separated from each other over a long period of time during which they accumulated characteristic SNPs by random mutation and drift. By calculating the extent of geographical allelic diversity, it was estimated that this split occurred approximately 10^6 years ago (figure 3b). One possible explanation for this split is that *T. gondii* may have migrated into South America with cats when the Panamanian land bridge was established (Johnson et al. 2003; Dubey et al. 2006; Khan et al. 2007; figure 3b). Reconnection of this land bridge approximately 1–3 Myr ago has been suggested to explain the diversification of plants and animals once introduced into the South (Marchall et al. 1982). A similar process may underlie the greater genetic diversity of *T. gondii* in South America. (*T. gondii* in South America is more genetically diverse than in North America and Europe; Khan et al. 2006a). Intriguingly, evidence also suggests that Chr1a*, which is largely fixed in North American strains, is making inroads into the South (Khan et al. 2007; figure 3b). For example, Chr1a* is found in groups 8 and 9, suggesting it has been introduced by sexual recombination. Other groups show evidence of recombination on Chr1a*, for example group 6, while more diverse groups (i.e. 5 and 10) lack Chr1a* and are genetically diverse (Khan et al. 2007). Further sampling will be necessary to test whether Chr1a is driving this expansion owing to a selective advantage or whether it is due to some other underlying demographic factor.

13. FUTURE DIRECTIONS
Given the markedly different population structure of *T. gondii* in North and South America, it is uncertain what to expect in other regions of the world. Limited sampling based largely on RFLP markers suggests that African populations are similar to those in Europe and North America (Lindstrom et al. 2008; Velmurugan et al. 2008). These surveys focused primarily on domestic animals and in particular chickens, so they might be influenced by introduction of parasite strains from Europe along with importation of domestic animals. Further testing of feral animals will be necessary to properly define the population structure in these regions. In contrast, studies from Asia suggest genetically mixed strains predominant (Sreekumar et al. 2003; Dubey et al. 2007), yet the extent of genetic diversity in these regions has not been properly estimated. Sequencing of introns and other loci from isolates in new regions is imperative to capture the full genetic diversity and hence accurately estimate local population structure.

The life cycle of *T. gondii* included modes for sexual and asexual transmission, hence the population structure may vary dramatically in different localities. By understanding these patterns, we can predict the risk of spread through the food chain and the potential for zoonotic infection. Defining the contribution of population structure to the spread of traits like immunogenicity and pathogenesis is highly significant to human health. Modelling such relationships in a model parasite like *T. gondii* may also provide insight into other parasites of animals that pose risk to humans.

Supported by the National Institutes of Health, the Wellcome Trust and the United States Department of Agriculture.
REFERENCES


Cornelissen, A. W. C. A. & Overduulse, J. P. 1985 Sex determination and sex differentiation in coccidia: gametogony and oocyst production after monoclonal infection of cats with free-living and intermediate host stages of Isospora (Toxoplasma) gondii. Parasitology 90, 35–44. (doi:10.1017/S003118200004899X)


Nowakowska, D., Colon, I., Remington, J. S., Grigg, M. E., Miller, M. A. et al. 1999 Invaded and intracellular survival by Toxoplasma gondii comprise a single clonal lineage. Nature (Lond.) 359, 82–85. (doi:10.1038/359082a0)


Webster, J. P. 2001 Rats, cats, people and parasites: the impact of latent toxoplasmosis on behavior. Microbes Infect. 3, 1037–1045. (doi:10.1016/S1286-4579(01)01459-9)