5-1-2005

Composite Genome Map and Recombination Parameters Derived from Three Archetypal Lineages of Toxoplasma gondii

Asis Khan
Washington University School of Medicine, St Louis

Sonya Taylor
Washington University School of Medicine, St Louis

Chunlei Su
Washington University School of Medicine, St Louis

Aaron J. Mackey
University of Pennsylvania

Jon Boyle
Stanford University

See next page for additional authors

2980–2992 *Nucleic Acids Research, 2005, Vol. 33, No. 9*
doi:10.1093/nar/gki604

Composite genome map and recombination parameters derived from three archetypal lineages of *Toxoplasma gondii*

Asis Khan, Sonya Taylor, Chunlei Su, Aaron J. Mackey[1], Jon Boyle[2], Robert Cole, Darius Glover, Keliang Tang, Ian T. Paulsen[3], Matt Berriman[4], John C. Boothroyd[2], Elmer R. Pfefferkorn[5], J. P. Dubey[6], James W. Ajioka[7], David S. Roos[1], John C. Wootton[8] and L. David Sibley*

Department of Molecular Microbiology, Center for Infectious Diseases, Washington University School of Medicine, St Louis, MO 63110, USA, [1]Department of Biology and Penn Genomics Institute, University of Pennsylvania, Philadelphia, PA 19104, USA, [2]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA, [3]The Institute for Genomic Research, Rockville, MD 20850, USA, [4]The Wellcome Trust Sanger Institute, Hinxton, UK CB10 1SA, [5]Department of Microbiology and Immunology, Dartmouth Medical School, Hanover, NH 03755, USA, [6]Animal Parasitic Disease Laboratory, ARS, ANRI, USDA, Beltsville, MD 20705, USA, [7]Department of Pathology, Cambridge University, Cambridge, UK CB2 1QP and [8]Computational Biology Branch, National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA

Received March 2, 2005; Revised and Accepted May 2, 2005

ABSTRACT

***Toxoplasma gondii* is a highly successful protozoan parasite in the phylum Apicomplexa, which contains numerous animal and human pathogens. *T.gondii* is amenable to cellular, biochemical, molecular and genetic studies, making it a model for the biology of this important group of parasites. To facilitate forward genetic analysis, we have developed a high-resolution genetic linkage map for *T.gondii*. The genetic map was used to assemble the scaffolds from a 10X shotgun whole genome sequence, thus defining 14 chromosomes with markers spaced at ~300 kb intervals across the genome. Fourteen chromosomes were identified comprising a total genetic size of ~592 cM and an average map unit of ~104 kb/cM. Analysis of the genetic parameters in *T.gondii* revealed a high frequency of closely adjacent, apparent double crossover events that may represent gene conversions. In addition, we detected large regions of genetic homogeneity among the archetypal clonal lineages, reflecting the relatively few genetic outbreeding events that have occurred since their recent origin. Despite these unusual features, linkage analysis proved to be effective in mapping the loci determining several drug resistances. The resulting genome map provides a framework for analysis of complex traits such as virulence and transmission, and for comparative population genetic studies.**

INTRODUCTION

Toxoplasma gondii is a widespread protozoan parasite that is capable of infecting nearly all species of warm-blooded vertebrates. While infections are common, disease typically only develops in rare situations such as immunocompromised hosts or during pregnancy (1). A member of the phylum Apicomplexa, *T.gondii* is closely related to a number of animal pathogens (Eimeria, Neospora, Sarcocystsis) and human pathogens (Cyclospora, Cryptosporidium, Plasmodium) (2). Due to the ease of genetic manipulation, robust animal models, and facility for cellular and biochemical studies, *T.gondii* has emerged as a model system for studying the unique biology of apicomplexan parasites (3–5).

T.gondii has an unusual population structure that consists of three predominant clonal lineages (types I, II and III) (6,7). Only two alleles exist at each locus and the distribution of these among the three lineages indicates that the majority of extant strains originated from a single recombination event and that since this, they have undergone a limited number

*To whom correspondence should be addressed. Tel: +1 314 362 8873; Fax: +1 314 286 0060; Email: sibley@borcim.wustl.edu

of genetic outbreeding events in the environment (8,9). *T.gondii* has a typical heteroxenous (alternating, two host) coccidian life cycle, consisting of asexual replication of haploid stages in a variety of warm-blooded hosts and a sexual cycle that only occurs in the enterocytes of the cat intestine (10). Following sexual development, oocysts are shed into the environment and from here they can contaminate food and water, thus infecting a variety of intermediate hosts. Meiosis occurs in the environment following shedding, and results in eight haploid progeny called sporozoites that remain contained within the oocyst. The separate maintenance of the three clonal lineages in the wild may result from two unusual features in the life cycle. First, a single organism is capable of undergoing the complete sexual development and self-fertilization in the cat to yield infectious oocysts (11,12). This trait is not unique to *T.gondii* as it is probably expressed by most apicomplexans. However, combined with the relative infrequency of simultaneous infection with more than one strain in cats, this may limit the opportunities for genetic recombination. Second, the direct oral infectivity of *T.gondii* tissue cysts for other intermediate hosts is highly unusual and allows transmission without the need of the sexual cycle. Evidence suggests that this is a recent phenomenon that arose simultaneously with the recombination event(s) that created the three predominant clonal lineages (9).

The ability of *T.gondii* to undergo meiosis in the cat has been exploited to develop experimental genetic approaches based on co-infection of a cat with tissue cysts from two separate parasite strains (13). These initial genetic crosses revealed the capacity of clones of the type III lineage to both self-fertilize and to cross-fertilize at roughly equal frequencies (13). Subsequently, genetic crosses have been done between the type II and III lineages (14) and more recently between the type I and III lineages (15). Before undertaking the present work, the genetic linkage map for *T.gondii* consisted of 57 unique genetic markers that defined 11 different chromosomes (linkage groups) (14,15). While these studies established the feasibility of linkage mapping as a forward genetic analysis in *T.gondii*, they were limited by the low resolution of the genetic map, the absence of a corresponding physical map, and relatively fragmentary sequence information.

Large-scale sequencing of expressed sequence tags (ESTs) in *T.gondii* has been useful for gene discovery and to identify a large number of single-nucleotide polymorphisms (SNPs) between different lineages (16–18). We have used the extensive EST database established for *T.gondii* to define SNPs that identify strain-specific alleles in *T.gondii*. Herein, we report on the segregation analysis of 250 genetic markers that detect SNPs between 71 recombinant progeny from several different genetic crosses. Linkage analysis was used to establish a robust genetic map and to assemble the scaffolds from a 10X genome sequencing project into a composite genome map for *T.gondii*.

MATERIALS AND METHODS

Genetic crosses

Recombinant progeny were obtained from a genetic cross called c96 that was performed by crossing the PTG strain (a type II strain derived from ME49 by cellular cloning) that was resistant to arprinocid-*N*-oxide (ANO) and 5-fluorodeoxyuridine (FUDR) with the CTG strain (Type III) that was resistant to sinefungin (SNF), adenine arabinoside (ARA-A) and diclazuril (DCL). Parental mutant strains were constructed using methods described previously (13,19–21). The c96 cross was performed by co-infecting a single cat with tissue cysts of the two different parental strains of the parasite after propagation of tissue cysts in mice. Oocysts were collected following shedding, allowed to sporulate, and then used to infect monolayers of human foreskin fibroblasts (HFF cells). Recombinants were selected in bulk by culture in the presence of both SNF (3×10^{-7} M) and ANO (3.6×10^{-5} M), and then cloned by limiting dilution in the absence of drugs.

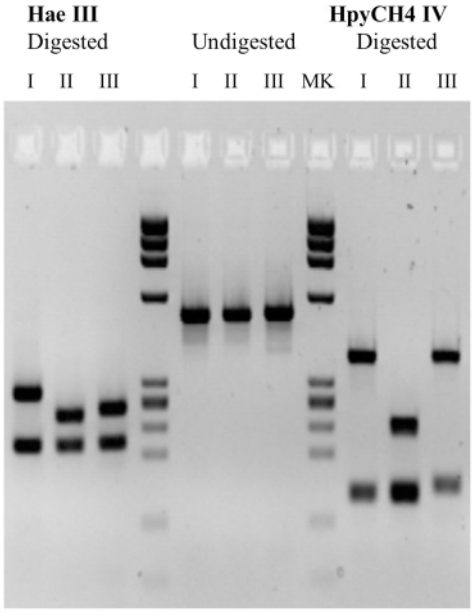
We also isolated new clones from a previously described cross between the type I strain GT-1 and the type III strain CTG (15). This cross was originally performed in two separate animals, numbered C285 and C295. We have previously reported on the segregation of alleles in 26 progeny of the C295 cross (15): 20 of these were genetically distinct and were used here. We isolated 11 new clones from the C285 cross by limiting dilution on HFF monolayers. Clones were genotyped using a set of PCR markers and unique recombinant clones chosen for further analysis. Recombinant progeny were phenotyped for resistance to the drugs FUDR (1×10^{-5} M), SNF (3×10^{-7} M), ARA-A (3×10^{-4} M) or ANO (3.6×10^{-5} M) by growth on HFF monolayers. Following inoculation at an MOI of 1:1, growth was followed microscopically to evaluate the percentage of cells infected, the extent of replication (2, 4, 8, 16 cell stages) and the degree of lysis following culture for 48–72 h. Clones were scored as sensitive versus resistant based on comparison to wild-type parasites and parental strains that were resistant to specific drugs. Although diclazuril was used on one of the crosses (c96), progeny were not tested for sensitivity to this drug.

Marker development and analysis

We utilized a previously defined collection of 57 genetic markers that define restriction length polymorphisms (RFLPs) that were analyzed by PCR amplification followed by restriction digestion and gel electrophoresis (15). Additionally, we converted a number of cosmid clones that distinguish RFLPs by Southern blotting (14) to PCR-based analysis. The majority of these cosmid clones contained genomic inserts from the type I RH strain. End sequences of these cosmid clones were compared by BLASTN to the genomic sequence of the ME49 strain (Type II) (http://toxodb.org/ToxoDB.shtml) to identify SNPs. Primers were then designed to amplify regions flanking SNPs that also detected strain-specific RFLPs. Fourteen of these provided reliable markers and these were given the prefix of KT followed by the original marker name in Figure 1.

To identify additional markers, we analyzed the *T.gondii* EST assemblies (http://www.cbil.upenn.edu/apidots/), which contain sequences generated from all three parasite lineages. Polymorphisms were identified by comparison of ESTs from assemblies that contained two or more overlapping sequences from two or more strain types. The resulting SNPs were then mapped against the 10X scaffold assemblies of the *T.gondii* genome derived from the type II strain ME49 (http://toxodb.org/ToxoDB.shtml) using BLASTN to identify corresponding

Locus: cB21-4
Chromosome: III
Primers:
cB21-4F 5′-CCAGGTGTTTCGATATTGAT-3′
cB21-4R 5′-GCCTGTGTGGTGTTCGAATC-3′
PCR annealing temperature: 55C
PCR product size: 502 bp
Restriction Enzyme: HaeIII (GGCC)
and HpyCH4IV (ACGT)
2% agarose gel



>cB21-4, Strain: Me49
CCAGGTGTTTCGATATTGATATTTATCGATGCGGACCAGGGTTGACAAATCTCCAATCTCA
TCAGTTCACGCAGGTCATTCAGACAAACACTGTAACTGTCTTCATCTGTACACCTCTAATT
GTCACTCTCCACAGAGAAGGACGAGCCGATGAAATCGTTGATCTGCCGACACCCTTCGACG
CTGTAGACTATTCGCTTCTCTACCGGCCCATCCATATTC(G,A,G)CGTCGCTGGATCCAC
CAAGTCACCTGCATCCTACGACCTACCTATCACTGCATTTTCAGTCACAAACAGTCCCTCT
ATTTTTACTCAGTAATGTGTATTATATCGTTCTGCCTCGCTATACACCGCTCGACAGAACC
AAACGTACGCATACAATATATATATATATATATATGTACATTCTTCGAGTTAGATACATGC
GCATGGAATCACAATGGTATCTACGCTCAGATCGGTGTGG(T,C,C)CCTGGATGTCAGCT
CGCTTAGATTCGAACACCACACAGGC

Figure 1. Representative SNP–RFLP marker used in mapping analysis. Data for the marker cB21-4, indicating primers, PCR conditions, the sequence flanking the SNP and an example of the alleles detected (gel insert). The specific SNPs detected are indicated in parentheses within the sequence block using the following format (allele type I, allele type II and allele type III). While most markers define simple biallelic patterns, in some cases these differences occur closely spaced in the genome. This example illustrates two closely spaced SNPs, one unique to type II and one unique to type I, thus this marker is capable of distinguishing all three genotypes using a combination of restriction enzyme digestions. The slightly larger band in the undigested fragment of the type III strain is due to an 8 bp difference; however, this is not the basis for the SNP differences detected here. Data for the complete set of markers can be found at http://ToxoMap.wustl.edu/.

genomic regions. We also utilized a specially designed database that systematically maps the distribution of polymorphism found in the ESTs against the genome (http://boothroydlab.stanford.edu/snp_maps/).

In regions where the EST abundance was not sufficient to identify SNPs, we amplified genomic regions from the three lineages, sequenced and aligned them to identify new polymorphisms. Triplicate PCR reactions were conducted from the RH strain (type II), the ME49 strain (type II) and the CTG strain (type III). Amplicons were sequenced using BigDye 3.0 cycle sequencing (ABI). Resulting sequences were aligned in clustal X and polymorphisms scored based on a consensus of two out of three templates. SNPs were identified by comparison of the consensus sequences for each of the three lineages.

Once SNPs had been identified *in silico*, they were screened for whether they detected RFLPs in the type strains RH (type I), ME49 (type II) and CTG (type III). Flanking regions were amplified by PCR and the resulting amplicons were subjected to restriction digestion and gel electrophoresis. In general, the regions chosen for final amplification of each marker range from 200 to 600 bp, and flank a single SNP that defined an informative RFLP. Markers were analyzed by PCR amplification, restriction gel digestion and agarose gel electrophoresis in the presence of ethidium bromide. Specific information on each of the markers including primers, PCR conditions and RFLPs can be found at http://ToxoMap.wustl.edu/.

MAPMAKER and linkage analysis

Genetic linkage maps were generated using MAPMAKER/EXP 3.0 (22). In total, 71 recombinant progeny were analyzed using a combination of the 250 genetic markers (excluding drug resistance markers) as shown in Table 1. The data from all three crosses were combined into a single analysis by denoting which marker/progeny combinations were informative. These data were analyzed in backcross configuration, as appropriate for the haploid inheritance system of the *T.gondii* crosses. An LOD value of $\geqslant$3.0 was used as a minimum cut-off to assign linkage. Drug resistance phenotypes were

Table 1. Summary of genetic crosses and markers used in assembly of the *T.gondii* genome map

Parental types	Crosses	Name of cross[a]	Number of progeny	Number of markers
II × III	ME49 (B7 clone) × CTG ARA-A^r/SNF^r	S CL	10[b] 9[b]	135
	PTG FUDR^r ANO^r × CTG ARA-A^r/SNF^r/DCL^r	c96	21[b]	135
I × III	GT-1–F3 FUDR^r × CTG ARA-A^r/SNF^r	C285 C295	11[c] 20[b]	176
Total			71	253[d]

[a]Separate crosses were done in parallel animals using mixtures of the indicated parental strains for CL and S (14), and for C285 and C295 (15).
[b]Clones were selected by drug resistance. Drug concentrations used for selection and typing of recombinants: ARA-A; adenine arabinoside (3×10^{-4} M); SNF; sinefungin (3×10^{-7} M); FUDR; 5-fluorodeoxyuridine (1×10^{-5} M); ANO; arprinocid-*N*-oxide (3×10^{-5} M); DCL: diclazuril (not tested).
[c]Clones were randomly selected.
[d]In addition to 250 unique genetic markers, three drug resistance markers were used in constructing the maps. The total number of markers is not simply the sum of the values for each cross separately as some markers are used in both crosses, while others are valid for only one of the crosses.

analyzed from the progeny of I × III cross using MapManager QTX (23) to detect maximum single locus associations (using the 'near' function which provides a LOD score for the peak association).

Genome-wide associations of drug resistance phenotypes

The sensitivity versus resistance phenotypes for SNF, FUDR and ARA-A were coded as 0 or 1 for the 31 progeny of I × III cross and evaluated for association with all markers typed in this cross on all chromosomes. Log likelihoods were determined by 1000 random permutations using the discrete small sample statistics previously developed for evaluating quantitative SNP associations in the haploid inheritance system of *Plasmodium falciparum* (24). This method allows for the detection of deviations from simple Mendelian inheritance if present, e.g. effects of incomplete penetrance, epistasis or multiple co-epistatic determinants. Such effects are consistent with two-state phenotypes as well as traits showing continuous variation. Empirical significance thresholds were determined using second-order nested permutations: 1.7 ('marginally significant') and 3.0 ('highly significant'). This analysis accounts for (i) any segregation distortion caused by drug selection or other factors, (ii) the non-normality of the permutation distributions and (iii) the multiple tests of these multi-locus scans.

Double crossover analysis

To compare the observed frequencies of double-crossovers at different spacings with the distribution expected from random models of crossover events, we used a modified form of the statistics of count-location processes (25). This treatment is based on multiple Poisson models and can reveal biases caused by general crossover interference, local crossover clustering or local gene conversions, which give various patterns of significant deviations from the random models. For a similar analysis of a *P.falciparum* cross (26), the number of intervening markers between double-crossovers was tallied by class (i.e. one intervening marker, two intervening markers, etc.) across all chromosomes. However, such tallies are not robust for analysis of the three *T.gondii* crosses because of substantial heterogeneities in marker densities and recombination frequencies among the crosses and among chromosomes. Instead, we calculated or estimated the physical map lengths of all inter-marker intervals from the coordinates of markers on the sequenced contigs and ordered scaffolds (examples of such coordinates are shown in Figure 6 and complete data are found at http://ToxoMap.wustl.edu/). Upper and lower bounds for distances spanned by each double crossover (D_U and D_L) were calculated as the ranges of the physical lengths that respectively include and exclude the inter-marker intervals containing the pair of crossovers. Then, for the analysis shown in Figure 5B, the best estimate of each inter-crossover distance was taken to be $D_L + (D_U - D_L)/3$. This is the mean value, with variance $(D_U - D_L)/18$, for a null model of uniformly distributed crossover locations in the local region containing the double crossover.

Separate models were made for each *j*th chromosome, based on the total number of crossovers observed in the three crosses combined. Physical distances were binned into fixed-length segments (1000/3 kb for the analysis shown in Figure 5B), then the probability $p_{j,k}$ of two crossovers on chromosome j being separated by k such segments was given by

$$p_{j,k} = \frac{2(n_j - k)}{n_j(n_j + 1)},$$

where $(n_j + 1)$ is the number of the fixed-length segments, including the partial end segment, on chromosome j. Thus, $n_j(n_j + 1)$ is the size of the state space of physical segments within which double crossover events can be observed. This treatment assumes an average genetic marker density of at least one per physical segment, which determined the choice of 1000/3 kb segment lengths for analysis of the three combined crosses. Then the total expected number, X_k of double-crossovers separated by k intervals was given by

$$X_k = \sum_{j=1}^{14} V_j p_{j,k},$$

where V_j is the double crossover number for chromosome j. V_j can be the total observed counts of double crossovers, or may be based on alternative definitions or models of double-crossover events as for the comparison of the two 'Expected' lines in Figure 5B. Where $k > (n_j + 1)$, $V_j\, p_{j,k}$ was set to zero, ensuring that the shorter chromosomes contribute to the total X_k only up to their maximum number of segments. The probabilities of obtaining the observed numbers of events from the random/uniform null model (as shown on the two deviant left hand bars of Figure 5B) were calculated from the Poisson mixture density using appropriate chromosome-specific $V_j\, p_{j,k}$ parameters estimated for each individual chromosome.

Mapping BAC-ends

The *T.gondii* ME49 (B7 clone, type II genotype) genomic BAC library was constructed by cloning size-selected Sau3AI partial digest fragments into pBACe3.6 digested with BamHI (27). Clones were randomly selected for end sequencing (http://www.sanger.ac.uk/Projects/T_gondii/). Of 10 329

reads, 4381 BAC-end sequence pairs were individually aligned to genomic scaffold sequences using BLAT (28). To be considered further, both members of a given end sequence pair were required to (i) achieve an alignment significance score ≥400 bits, (ii) include >80% or more of their length in the alignment and (iii) have only one such reliable alignment found in the entire genome. A total of 1616 BAC-end sequence pairs satisfied these requirements, and their BLAT-identified genomic locations were used to inform and augment the genetic mapping study.

CMap database for *T.gondii*

The CMap application was established on an OSX v10.3 platform running UNIX and using the source code and instructions provided by the developer (http://www.gmod.org/cmap/index.shtml). Data were housed in a MySQL database and interfaces created with Perl 5.8 and Apache v2.0. CMap for *T.gondii* was built by including all of the data for genetic linkage maps with physical placement of markers on the scaffolds and assembled chromosomes. Correspondences were drawn between the same marker on each of the maps to establish relationships between the physical and genetic distances. CMap for *T.gondii* is housed at http://ToxoMap.wustl.edu/cmap/.

RESULTS

Genetic crosses used in construction of the map

We analyzed the segregation of polymorphic DNA markers among a total of 71 progeny from genetic crosses that were performed between type II and III lineages or between the type I and III lineages as shown in Table 1. A total of 19 progeny were included from a previously described cross between the type II strain PTG and the type III strain CTG (14). We also obtained 21 new clones from a separate cross that was also performed using the parental strains PTG and CTG (referred to as c96 cross, Table 1). In both cases, the progeny were selected by virtue of drug resistance mutations engineered into the parental strains (Table 1), as described previously (21). The frequency of self-mating in these crosses was previously found to be ~50% (21), and recombinant progeny can be reliably distinguished by virtue of drug resistance or sensitivity.

Additionally, we analyzed 20 progeny from a previously described cross between the virulent type I lineage strain GT-1 and the non-virulent type III lineage CTG (15). In contrast to the previously described clones that were isolated by drug selection, the new clones from this cross were randomly selected in the absence of drug selection. Of 95 separate clones that were screened using a combination of 11PCR markers located on different chromosomes, all but one was a recombinant: the sole parental strain was of the type III lineage (data not shown). Eleven of these newly isolated clones were genetically distinct, and they were included in the analyses here (Table 1).

Development of SNP–RFLP markers

Initially, we utilized a previously defined collection of 57 genetic markers that detect restriction fragment length polymorphisms (RFLPs). Markers were analyzed by PCR amplification followed by restriction digestion and gel electrophoresis (PCR–RFLP typing) (15). These markers were placed on the genome by comparison of their sequences with BLASTN against the recently completed 10X shotgun coverage of the *T.gondii* genome that has been assembled into scaffolds (available at http://toxodb.org/ToxoDB.shtml). A single best alignment resulted in unambiguous placement of all of these markers and it also identified many physical regions that were not represented by this collection of markers (data not shown). To expand the coverage of genetic markers, we placed new markers at approximately regular intervals along all of the 10X scaffolds that were 50 kb or greater in size. These markers were developed using the extensive EST database of *T.gondii* (17,29) to identify SNPs between different strains. In total, ~200 new RFLP markers were characterized, providing a total of 250 markers spaced at ~300 kb intervals across the genome. A complete listing of these markers, including methods for their analysis is contained in a genetic mapping database for *T.gondii* (http://ToxoMap.wustl.edu/). An example of the data supplied for each marker is provided in Figure 1 including primer sequences, PCR conditions, detection of the alleles by gel electrophoresis, the flanking sequence and specific SNP(s) detected.

Generation of linkage maps using MAPMAKER

For the three predominant strains of *T.gondii*, each locus contains only two alleles, thus two lineages share one genotype with the third containing a distinct allele often defined by a single SNP. Consequently, the number of informative markers differs for the various genetic crosses, as listed in Table 1. Type I-specific SNPs are informative only in the I × III crosses, and type II-specific SNPs are informative in only the II × III crosses, while type III-specific SNPs are informative in both genetic crosses. Progeny listed in Table 1 were analyzed by PCR amplification of the appropriate genetic markers followed by restriction enzyme digestion and agarose gel electrophoresis to detect specific SNPs. A database containing the segregation data for all of the progeny and markers used in the present analysis can be found at http://ToxoMap.wustl.edu.

The inheritance of SNP–RFLP markers among the progeny was compared using MAPMAKER (22) to generate linkage maps with a minimum LOD score of ≥3.0. Eight major linkage groups were identified by MAPMAKER in the absence of any specific pre-assignments (data not shown). One of these groups resulted from the combination of markers on chromosomes Ib, III, VI, IX, X and XII, reflecting a bias in shared alleles among these chromosomes. Additionally, markers on chromosomes VIIb and VIII were grouped together. Specific anchors for each chromosome were then assigned for each chromosome. These anchors were chosen based on the following criteria: (i) their use in previous studies defining the segregation groups (14,15), (ii) their hybridization to specific chromosomes separated by pulse-field gel electrophoresis (PFGE) [(30) and data not shown] and (iii) their physical assignment to specific scaffolds of the 10X genome as established by BLASTN. When specific anchors were assigned (Table 2), MAPMAKER resolved the markers into 13 discrete linkage groups. The placement of each marker was supported by a predominant LOD score (generally >6.0) that was also consistent with the physical location on the scaffolds of the 10X genome. Although the LOD scores indicated possible

secondary linkage between Ib and IX and XII, this artifact is attributable to strong segregation distortion: non-linkage among these separate chromosomes was confirmed using permutation analysis (data not shown). For these three chromosomes, the progeny show a highly significant excess of the type III CTG parental alleles (~80:20 ratio instead of the expected 50:50, $P < 0.001$). This can be attributed to drug selection on IX and XII but the cause for Ib is unknown.

Table 2. Assignment of markers to specific chromosomes and recombination rates for each chromosome

Chromosome number	Anchoring markers in MAPMAKER	Genetic size (cM)[a]	Physical size (bp)[b]	Average map unit (kb/cM)
Ia	AK1	43.9	1 856 182	42.3
Ib	ACT1	16.2	1 956 324	120.8
II	L351	17.9	2 343 157	130.9
III	CB21-4	34.9	2 470 845	70.8
IV	Bc18U	38.5	2 576 468	66.9
V	L358	28.4	3 147 601	110.8
VI	PK1	29.3	3 600 655	122.9
VIIa	SAG4	30.1	4 502 211	149.6
VIIb	L339	32.4	5 023 822	155.1
VIII	GRA1/AK113	64.7	6 923 375	107.0
IX	BTUB	59.2	6 384 456	107.8
X	C19-TA3	68.3	7 418 475	108.6
XI	GRA4	50.9	6 570 290	129.1
XII	SAG3	77.8	6 871 637	88.3
		592.5	61 645 498	Ave. 104.0[c]

[a]Total genetic units as determined by MAPMAKER (Figure 2).
[b]Total physical size obtained by summing the individual scaffolds (Figure 4).
[c]Calculated based on the sum of mapped scaffolds.

MAPMAKER was unable to separate two chromosomes, VIIb and VIII, due to the fact that some markers on the distal end of scaffold 995362, which lies on chromosome VIII (i.e. AK46, AK47, AK111, AK112), were linked to chromosome VIIb while other markers were linked to both chromosomes (i.e. AK49, SRS1, SAG1.c, SRS2 on scaffold 995362, and markers AK50, AK51 on scaffold 995368). The remaining markers on scaffold 995368 were linked only to chromosome VIII (i.e. AK52, AK30, SAG2, etc.). The individual assemblies for each of these scaffolds (derived from the type II strain ME49) were rechecked and found to be valid, indicating that this apparent linkage is not due to an error in the physical maps (data not shown). This apparent linkage is entirely attributable to the results of the I × III cross because both VIIb and VIII show a preponderance of type I SNPs (colored red in Figure 2) (1), and analysis of the three crosses separately showed strong statistical significance for VIIb versus VIII marker associations only in the I × III cross (log likelihoods estimated from 1000 permutations were between 4.9 and 7.1), but not in the II × III crosses (2). PFGE does not resolve the question of possible physical linkage between VIIb and VIII since these individual chromosomes are each >5 Mb in size (the current resolution of our gels) and expected to remain trapped in the well. This raises the possibility that chromosomes VIIb and VIII are physically linked in the type I GT-1 parent; however, considering these results together, we have elected to report these two groups as separate chromosomes. The corresponding 14 linkage groups including the positioning of all 250 genetic markers and three drug resistance markers are shown as ordered genetic maps in Figure 2.

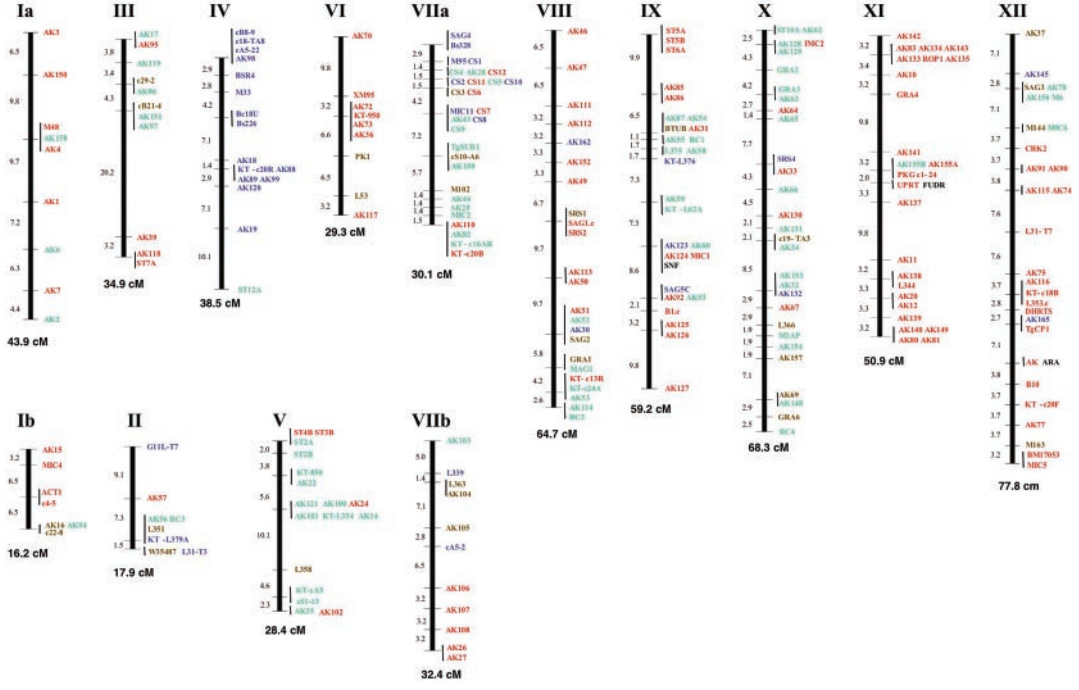


Figure 2. Genetic linkage maps for the 14 chromosomes of *T.gondii*. Individual markers are shown to the right of the vertical bar and chromosome numbers are given above each map. Markers that map to the same node are indicated to the right of a solid vertical bar. The corresponding genetic distances between each node are given to the left of each map and the total sizes in cM are shown at the bottom of each chromosome. Polymorphisms that are unique to type I are shown in red, those unique to type II are shown in green, those unique to type III are shown in blue and markers that contain multiple polymorphism as illustrated in Figure 1 are shown in yellow. Maps were constructed using MAPMAKER (22) from the analysis of 71 recombinant progeny using 250 genetic markers (Table 1). Markers that include data analyzed by Southern blot are followed by the suffix '.c'.

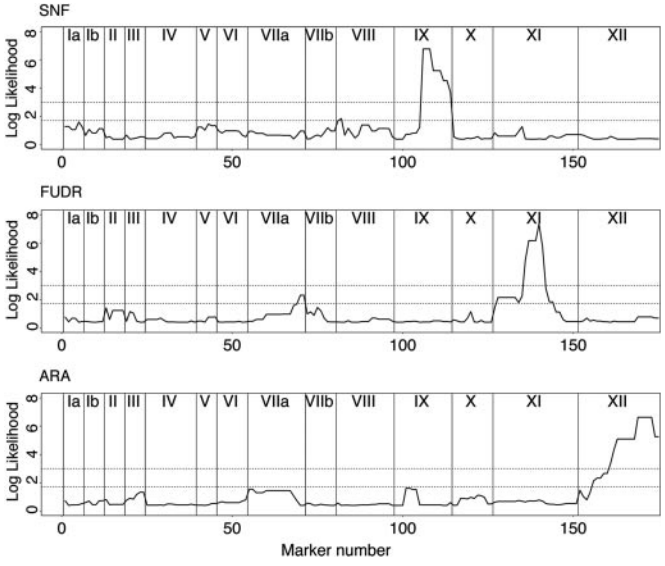


Figure 3. Mapping of drug resistance to SNF, FUDR and ARA-A across the genome of *T.gondii*. Plots indicate the log-likelihood association of resistance to each drug with markers aligned across the genome. In each case, a single locus was found to be statistically associated with resistance, and apparent secondary peaks were either non-significant or in non-informative regions of segregation distortion. Resistance to SNF was localized to chromosome IX, FUDR was localized to XI and ARA-A was mapped onto chromosome XII. Plots were generated from genome-wide scans using permutation analysis for marker associations from I × III crosses. Significance levels are given by dotted lines [lower line is significant (log likelihood of 1.7) while the upper line is highly significant (log likelihood of 3)].

Mapping drug resistance

To illustrate the utility of linkage analysis in mapping specific drug resistance traits in *T.gondii*, we have depicted the association between each marker and resistance to SNF, FUDR or ARA-A across the entire genome (Figure 3). Resistance to each of these compounds showed a single strong association with log-likelihood ratios of >6.0 and an absence of strong secondary associations (Figure 3). We have not attempted to measure the levels of resistance in each clone quantitatively, and it is possible that with more precise estimates of their IC_{50} values, secondary traits influencing susceptibility would be detected. However, as these compounds are used here as an experimental tool only and are not clinically relevant, they serve here as an example of mapping simple Mendelian traits. The resolving power of these associations, when considered as a proportion of chromosome length, is similar to that obtained previously in the 35-progeny Dd2 × HB3 cross of *P.falciparum* [compare Figure 3 with (31,32)]; however, the physical distances resolved are ~6-fold greater in *T.gondii* because of the lesser meiotic crossover frequency.

The target of two of these drugs is known from previous studies. FUDR targets the uracil phosphoribosyl transferase (UPRT) gene and mutations in this target are resistant to growth inhibition (33,34). Similarly, ARA-A targets adenosine kinase (AK), a non-essential enzyme involved in purine salvage (35,36). These known targets provide an estimate of the precision of mapping by linkage analysis. The UPRT gene is in the center of chromosome XI and mapping of the resistance phenotype revealed perfect correspondence to this locus in I × III and c96 crosses (Table 3; Figures 2 and 3). The AK gene lies at the end of chromosome XII and mapping of resistance to ARA-A revealed perfect correspondence to this locus in I × III cross (Table 3; Figures 2 and 3). Because informative markers in the II × III crosses do not lie at exactly the same intervals (there is no informative polymorphism associated directly with the AK gene), resistance to ARA-A was mapped to a relatively broad region spanned by the markers AK165 and AK163.

Table 3. Mapping of drug resistance in genetic crosses of *T.gondii*

Drug resistance	Chromosome	Closest marker	LOD score[a]	Log likelihood[b]
FUDR	XI	UPRT	9.33	7.34
ARA-A	XII	AK	9.33	6.62
SNF	IX	AK123	21.37	6.77

[a]LOD scores were calculated using MapManager QTX (23).
[b]Log-likelihood statistic was determined by permutation analysis. Both analyses were based on phenotypes derived from I × III crosses.

The biochemical targets of the remaining two inhibitors used here are unknown. The compound SNF is an analog of *S*-adenosyl-methionine and the mechanism of action is thought to be due to inhibition of protein and/or DNA methylation reactions (37,38). The specific target of SNF in *T.gondii* is unknown; however, it is a potent inhibitor of parasite growth *in vitro* and mutants are readily isolated following chemical mutagenesis (39). Our results map SNF resistance to a locus near AK123 on chromosome IX (Table 3, Figures 2 and 3). Although SNF has the highest LOD score with AK123 (Table 3), this marker is located at the same genetic position with three others markers AK60, AK124, MIC1. Collectively, these markers span ~540 kb in physical size, thus limiting our ability to precisely map the gene responsible for SNF resistance.

The anti-coccidial compound arprinocid is converted to arprinocid-*N*-oxide (ANO), which is significantly more potent at inhibiting the parasite than the parental compound (40). The mechanism of action of this anti-coccidial drug is presently unknown and our mapping results indicate ANO resistance was linked to chromosome XI along with resistance to FUDR. Unfortunately, the biased nature of the alleles in *T.gondii* makes it impossible to localize this linkage precisely since there is only one informative marker for the II × III lineages on this chromosome (AK155B). Resistance to ANO matched this marker exactly and was also identical to the pattern of FUDR resistance for the c96 clones. However, this does not indicate that these two compounds share the same target and in fact other clones that were FUDR resistant (derived from the I × III cross where arprinocid was not used) were fully sensitive to ANO (data not shown).

Mapping the genome scaffolds to chromosomes

The recently completed 10X whole genome shotgun sequence of the *T.gondii* genome was obtained from the type II strain ME49 and it consists of ~670 scaffolds that range in size from <6 kb to >5 Mb (http://toxodb.org/ToxoDB.shtml). However, the absence of a physical map has hampered the assembly of the genome into chromosomes. Based on the segregation of genetic markers, we were able to assign 63 of the largest scaffolds (Table 4) to specific groups and to order these scaffolds along the chromosomes (Figure 4). The orientation

of many of these scaffolds follows directly from the genetic linkage maps (green bars in Figure 4). The resulting 14 chromosomes ranged in size from 1.8 (chromosome Ia) to >7 Mb (chromosome X). In addition to mapping scaffolds by linkage studies, we also identified the physical locations of BAC clones whose paired end sequences were generated by the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/Projects/T_gondii/). The majority of BAC clones (1555 of 1616) were entirely contained within an existing genomic scaffold. The distances between the ends were found to have a bimodal distribution of lengths with two peaks at 18–20 kb and at 60–80 kb (Supplemental Figure 1). Among the remaining 61 clones, many BAC clones bridged the physical gaps between scaffolds that had already been mapped and oriented by genetic linkage analysis, confirming the genetic chromosomal assembly (green bars connected by horizontal lines in Figure 4). In some cases, the location of a BAC clone provided orientation for adjacent scaffolds that were genetically linked, but whose end-to-end orientation was uncertain (light purple bars in Figure 4). Furthermore, a small number of BAC clones supported the inclusion of several smaller scaffolds within the chromosomal linkage groups. Additional genetic markers were developed within these smaller scaffolds and their chromosomal location confirmed by linkage analysis. While BAC clones tied together the majority of physical scaffolds for some chromosomes (i.e. II, III, IV, VIIa and XII), some scaffolds remain linked only by genetic data. In a few cases, the small size of scaffolds or low recombination rate between adjacent markers (e.g. chromosome V) made it impossible to correctly predict their orientation (yellow bars in Figure 4). In total, the resulting genome map represents 61.6 Mb in size, which corresponds to ~95% of the estimated genome size of ~65 Mb. A relatively large number (~600) of small scaffolds remain and most are <10 kb in size. The scaffolds cannot be assembled into the genome either due to low quality sequence or repeats and they are also well below the

Table 4. Number and sizes of 10X whole genome sequence scaffolds that were mapped

Sizes of scaffolds[a] (bp)	Number of scaffolds	Sizes (bp)
<100 000	18	604 560
100 000–300 000	11	2 007 756
300 000–500 000	9	3 484 114
500 000–001 000 000	5	3 761 563
1 000 000–002 000 000	12	17 981 672
>2 000 000	9	33 805 833
Total	63	61 645 498

[a]Scaffold numbers were based on ToxoDB (http://toxodb.org/ToxoDB.shtml), a complete list of the scaffolds that were mapped can be found at http://ToxoMap.wustl.edu/.

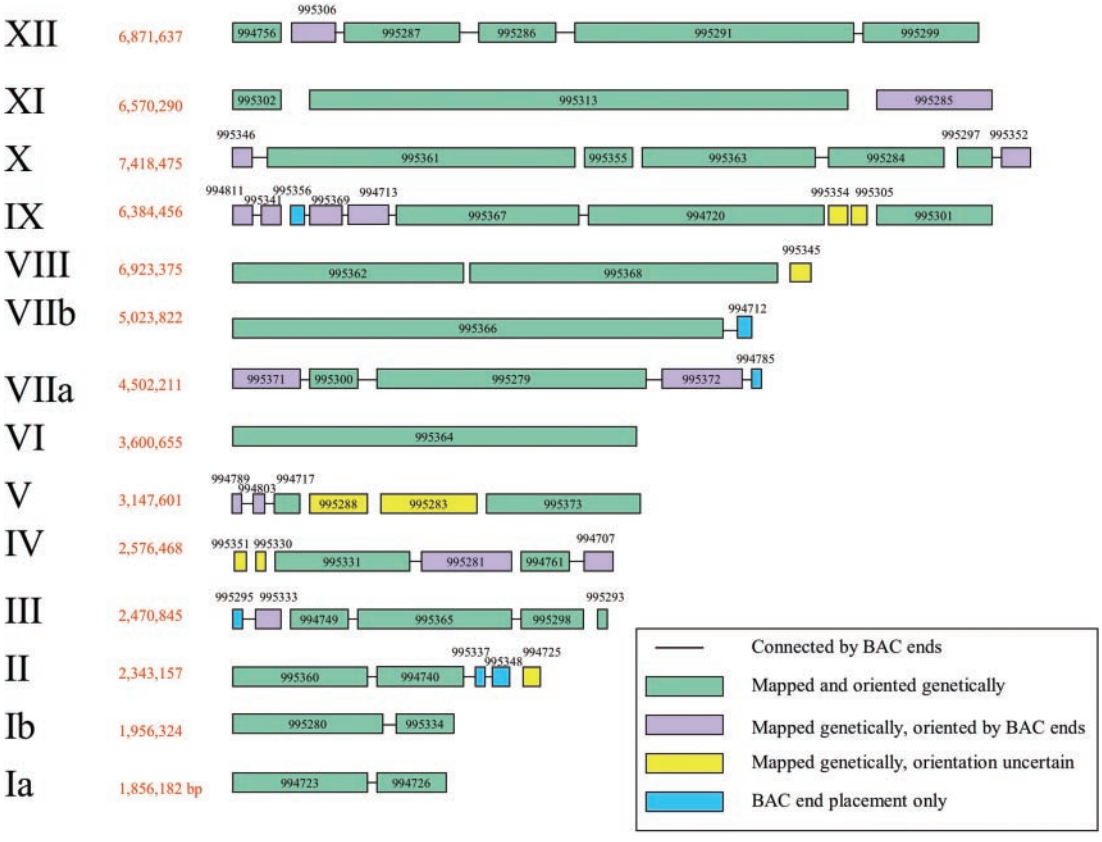


Figure 4. Horizontal maps of the genome scaffolds comprising each of the *T.gondii* chromosomes. Chromosome numbers are indicated to the left as is the total size of scaffolds in base pairs. Scaffold numbers are given within or above the colored bars. Bars colored green indicated scaffolds that were mapped and oriented by linkage analysis. Bars indicated in light purple were linked genetically and oriented by BAC-end data. There are several remaining scaffolds that are linked by genetic data but where the orientation is still unknown (yellow bars). There are also five scaffolds that are linked only by BAC-end data (blue bars). Orientations that were supported by BAC-end data are indicated by a thin horizontal line connecting the bars.

mapping resolution of linkage analysis. Thus, they are not presently incorporated into the genome map.

Genetic parameters of *T.gondii*

The genetic map of *T.gondii* reported here consists of a total of ~592 cM. Based on an estimated genome size of ~65 Mb, this corresponds to an average of ~104 kb/cM. The genetic sizes of specific chromosomes were roughly correlated with their physical sizes as shown in Figure 5A. By placing the genetic markers along the physical map provided by the 10X scaffolds, we were also able to estimate the rate of recombination across different chromosomes and specific regions of the genome. As is typically seen in eukaryotic organisms, the rate of recombination versus physical distance varied somewhat across the chromosomes ranging from a high of 42 kb/cM (chromosome Ia) to a low of >155 kb/cM (chromosome VIIb) (Table 2). This variation is statistically significant, since the overall distribution of crossover counts and locations deviated markedly a single genome-wide Poisson model. However, using separate estimates of Poisson parameters for each chromosome provided a good fit to the data.

One notable feature, which has previously been found in *P.falciparum* (26), is that the frequency of apparent double recombination events encompassing only single markers greatly exceeds the expected rate of closely spaced double-crossovers. We analyzed the occurrence of these 'single-marker events' (i.e. progeny haplotypes like *hhahh* and *aahaa*) and all double-crossovers separated by more than one marker on each of the chromosomes of *T.gondii*. In addition to classifying events according to the number of intervening genetic markers, we estimated the ranges of physical distance separating the crossover locations and compared these with theoretical models (see Materials and Methods). We carefully checked all single-marker events by re-genotyping the relevant progeny at the pertinent marker and flanking markers.

As shown in Figure 5B, the frequencies of single-marker events that are estimated to span short physical distances were very significantly in excess of the double crossover frequencies predicted by the models (light gray sections of the two left hand bars, $P = 0.00015$ and 0.011). This excess amounts to ~19–20 single-marker events. In contrast, the events with longer spacings showed no significant differences from the expected frequency of double-crossovers: these include the five single-marker events in the third and fourth bars from left in Figure 5B which have widely spaced flanking markers.

The general agreement between the observed distribution of double-crossovers spanning two or more markers and the Poisson/uniform models suggests that generalized positive or negative crossover interference is not detectable in these crosses. Therefore, single-marker events could be attributable either to small local clusters of crossovers ('hotspots') or to local non-reciprocal meiotic gene conversions. Hotspots are evident in two locations where an identical single-marker event is present in two or more independent progeny (the events spanning markers AK57 and SAG3 on chromosomes II and XII, respectively). However, the majority of the excess single-marker events are likely to reflect gene conversions, as suggested for *P.falciparum* (26) and consistent with the meiotic mechanisms known in other eukaryotes. In support of the conversion hypothesis, if these conversion candidates are excluded from the theoretical prediction, the model gives an improved fit to the distribution of all double-crossovers spanning two or more markers (lower line, Figure 5B).

CMap database

To provide a dynamic interface for comparing the genetic maps with the underlying physical scaffolds, we have

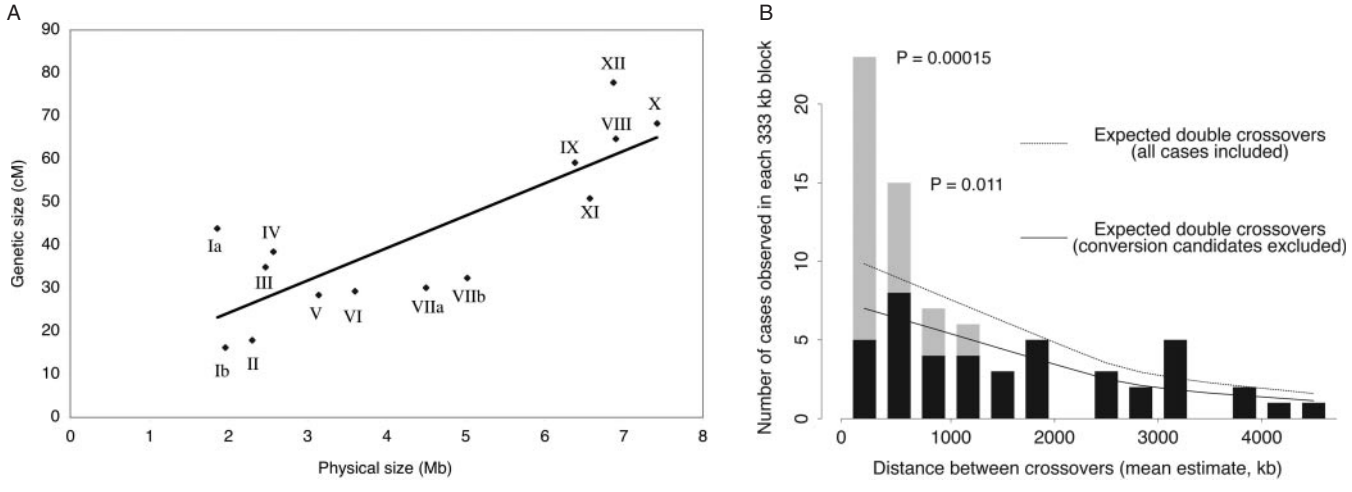


Figure 5. Recombination frequency for *T.gondii* chromosomes and analysis of double-crossovers. (**A**) The genetic size of chromosomes was roughly correlated with their physical sizes as determined by summing the scaffolds for each chromosome shown in Figure 2. (**B**) Double-crossover analysis suggests that gene conversion processes in addition to double-crossovers contribute to recombination events spanning a genome segment containing only a single genetic marker. Vertical black bars show numbers, for all 71 progeny, of double-crossovers that are separated by two or more genetic markers, plotted versus estimated physical distance ranges separating the two crossovers (see Materials and Methods). Light gray sections of the four left bars represent additional recombination events spanning only one genetic marker ('single-marker events'). As shown by the *P*-values and theoretical lines, the single-marker events in the two left bars are unlikely to occur by chance from the frequencies of true double-crossovers predicted by Poisson/uniform models (see Materials and Methods). The lines represent alternative models where dotted line indicates all cases, including all single-marker events, are assumed to be true double-crossovers and solid line indicates single-marker events in the two left bars are excluded from the numbers of double-crossovers.

developed a CMap database for representing genetic linkage maps based on the Genetic Model Organisms Database (GMOD) (http://www.gmod.org/). CMap for *T.gondii* contains the genetic linkage maps, physical placement of markers along the scaffolds, and corresponding relationships between the physical and genetic maps (http://ToxoMap.wustl.edu/cmap/). To illustrate the utility of this database, we have generated a composite image of chromosome III showing the markers present on each scaffold and their correspondence to the genetic map (Figure 6). This figure illustrates the distribution of markers across the physical scaffold with spacing of ~300 kb and illustrates how the map unit varies substantially across the chromosome from ~2 to >100 kb/cM. As well, this map illustrates the anchoring of some scaffolds using the BAC-end clones (dumbbell symbol in Figure 6), which gives added confidence to linkages between adjacent scaffolds where the map unit is highly variable. We have also incorporated all of the genetic markers into the GBrowse application of the genetic model organism database suite (http://www.gmod.org/) that has been established for the *T.gondii* genome (http://toxodb.org/cgi-bin/gbrowse). GBrowse for *T.gondii* also contains all of the predicted open reading frames (ORFs), gene prediction models, ESTs and BLAST homologies for the *T.gondii* genome. The combination of the CMap and GBrowse databases will make it possible to readily connect phenotypic information on specific genetic linkages to genes in the underlying genome.

DISCUSSION

Apicomplexan life cycles typically involve both sexual and asexual stages, yet rarely has this been exploited to develop genetic tools, largely due to the intractability of many of these organisms. Genetic linkage maps have been generated for several other apicomplexan parasites including the human malarial parasite *P.falciparum* (26), the rodent malarial parasite *P.chabaudi* (41) and the avian coccidian *Eimeria tenella* (42). Genetic studies have been useful in these organisms for mapping drug resistance as well as for studying complex phenotypes such as invasion and development (31,41–43). Previous studies in *T.gondii* have also revealed the potential for genetic mapping to identify quantitative trait loci for complex phenotypes such as virulence (15). Here, we extend this analysis by generating a high-resolution linkage map and using it to assemble the sequence scaffolds from the recently completed whole genome shotgun sequence of *T.gondii*. The resulting genome map provides a framework for expanded studies on the genetic basis of complex biological traits such as drug resistance, infectivity and transmission.

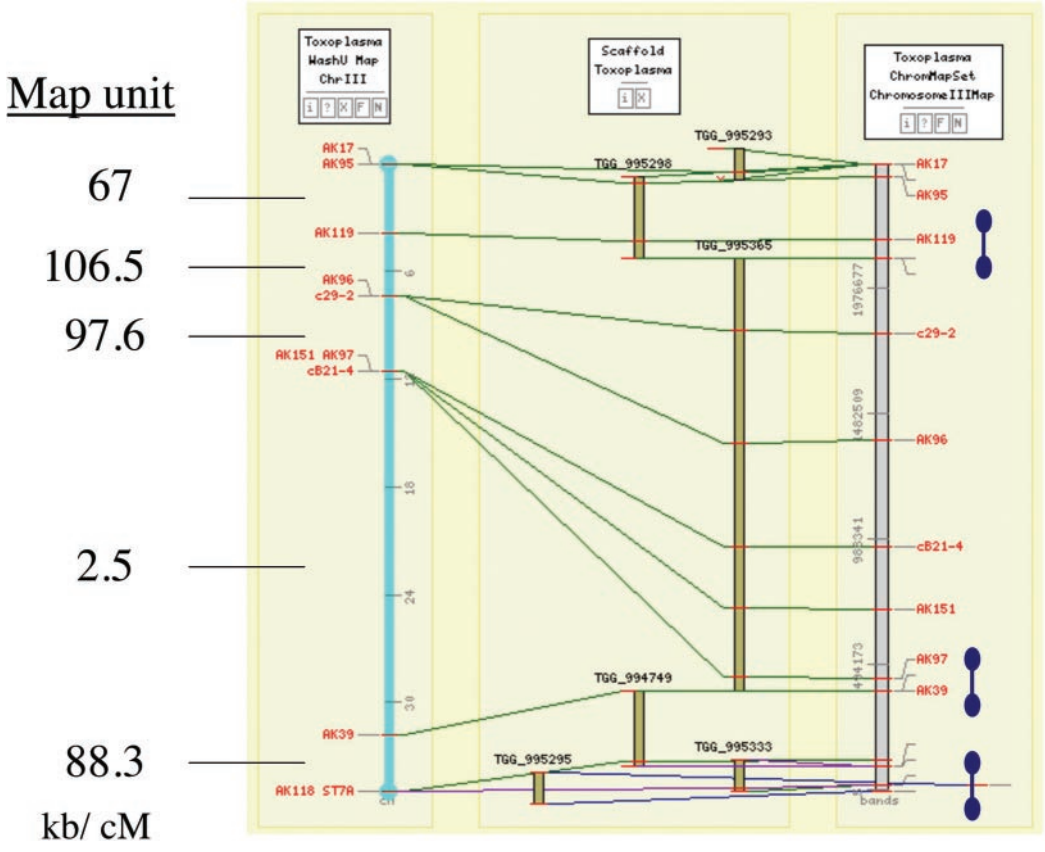


Figure 6. Genetic and physical maps for chromosome III as drawn from CMap for *T.gondii*. CMap representation of the genetic map (left), scaffolds (center) and the assembled chromosome (right) for *T.gondii* chromosome III. Correspondences are shown in colored lines depicting the relationship between the data. Green lines indicate markers that were mapped and oriented genetically. Purple lines indicate markers that were mapped genetically and oriented by BAC-ends. As well, BAC-end linkages are indicated by blue horizontal dumbbell shaped bars. The scale for the left-hand map is based on genetic units, while the center and right-hand maps are based on physical size in base pairs. Differences in the map unit across the various nodes of the genetic map are indicated at the far left in kb/cM.

Members of the genus *Plasmodium* and *Eimeria* both have 14 major linkage groups. In the present study, 12 distinct groups were identified for *T.gondii* based on linkage analysis and combined with data on the physical separation of the chromosomes, it is likely that there are actually 14 chromosomes. In a previous version of the *T.gondii* genetic map, 11 chromosomes were identified in *T.gondii* based on a combination of linkage studies (15) and molecular karyotyping (30). We detected the same 11 linkage groups characterized previously and also identified three new chromosomes. The previous absence of these chromosomes can be attributed to the previously low coverage of RFLP markers, combined with the fact that several of the new groups (IX and XII) were too large to enter pulse-field gels. The combined genome map contains 50 gaps between adjacent scaffolds, 29 of which are linked by BAC clones. These BAC clones thus provide approximate size estimates and possible templates for future efforts to close the remaining gaps.

While nearly all of the markers were placed unambiguously to a single chromosome, there was a region of unexpected strong linkage between chromosomes VIIb and VIII, solely attributable to the results of I × III cross. A physical association between these chromosomes is not consistent with scaffold position as established from the assembly of the type II strain ME49 genome or physical hybridization of markers from VIIb and VIII. Despite previous studies indicating an equal likelihood of self-fertilization versus cross-fertilization (13,39), we observed that nearly all clones isolated from I × III cross were genetic recombinants, even when not selected by drug resistance markers. Each of the parental strains used in this cross is individually capable of self-fertilization (data not shown), indicating that this result is not due to obligatory cross-fertilization. Thus, our findings could be explained by selective pressure for recombination among the clones that grew out from this cross, resulting in the apparent linkage between chromosomes VIII and VIIb. It is also possible that VIIb and VIII are physically joined in the type I GT1-F3 parent, but not in the type II or type III lineages. The absence of polymorphisms between the type II and III lineages makes it impossible to accurately predict the behavior of chromosome VIII in some genetic crosses. Therefore, additional genetic crosses (i.e. between types I and II lineages) will be needed to resolve this issue.

The utility of genetic linkage studies for identifying genes that mediate specific phenotypes is determined largely by the meiotic recombination frequency, which is typically expressed as a map unit (cM) representing the region across which there is a 1% chance of recombination. Despite having a similar complexity in terms of number of genes, the rate of recombination in *T.gondii* is ~6 times lower than *P.falciparum*. Thus, while Plasmodium has an average physical distance per recombination unit of 17 kb/cM (26), the corresponding value is ~104 kb/cM for *T.gondii*. The average map unit in *T.gondii* varied by up to 3-fold between separate chromosomes, although greater local variation was occasionally noted. The relatively large map unit in *T.gondii* means that it is straightforward to map a particular phenotype to a general region of one of the chromosomes. The current resolution of the map allows mapping of a trait to a genomic region of 200–500 kb, depending on the location. For example, genes encoding SNF, FUDR and ARA-A resistance were mapped to loci on chromosome IX, XI and XII, respectively (Figures 2 and 3). While the genes conferring resistance to FUDR and AK were previously known, the target of SNF was not. Our analysis places this gene within a region that comprises ~500 kb on chromosome IX, bounded by the markers AK123 and MIC1. Because >95% of the genome sequence is now assembled and ordered into scaffolds that correspond to the genetic map, it is possible to identify the underlying genes using a combination of CMap and GBrowse. Such analysis indicates there are ~60–70 genes within this region and it is likely that a single mutation in one of them is responsible for the resistance phenotype.

Positional cloning of drug resistance loci could be accomplished by linkage analysis of further recombinant progeny. While increasing the overall precision of the map might require 5–10 times the number of progeny, finer mapping on a local scale could be achieved with far fewer clones, particularly if they are chosen for recombinations across specific regions of the genome. This can readily be achieved by PCR-based screening on 96-well plate cultures, allowing the screening of hundreds of potential clones. It is estimated that a single cross can generate $>10^7$ oocysts, making the upper limit of recombinant clones quite high. It is possible to cryopreserve the progeny of a single cross as an uncloned population for future efforts aimed at isolation of additional clones. Ultimately, positional cloning will still be limited by the relatively low recombination rate in *T.gondii*; however this approach is complementary to reverse genetic approaches that are also well-developed in *T.gondii* (3).

Like malaria (26), *T.gondii* exhibited a high frequency of closely spaced apparent double-crossovers. This phenomenon has been associated with gene conversion events, rather than reciprocal exchanges in other systems (26). While we have not directly determined the mechanism by which these events occurred (i.e. true-double crossovers versus gene conversions) in *T.gondii*, their presence has significance for mapping phenotypes by linkage analysis: intervals spanning quantitative trait loci, for example, can be rigorously delimited only by crossovers but not by non-reciprocal conversions. If single-marker events do represent gene conversions, there were probably many more conversions that were not detected by the current density of markers. Consequently, it may be advantageous to increase the density of the markers as well as the numbers of progeny as a means of improving the resolution for mapping complex traits.

One highly unusual feature of the *T.gondii* genome is the asymmetric distribution of strain-specific SNPs. In order to illustrate this, we have color-coded each SNP mapped here based on the strain type where the unique difference is observed (red = type I, green = type II and blue = type III in Figure 2). Several chromosomes present remarkably homogeneous patterns. For example, chromosome IV is almost entirely made up of type III-specific SNPs, likewise chromosome XI is primarily type I. On some chromosomes, this occurs as a marked bias toward one end (distal portions of chromosomes III, VIIb and XII are primarily type I). These regions of homogeneous polymorphism were interspersed with the majority of regions that show mixed patterns of SNPs. The reason for these differing patterns is uncertain, but they suggest that large-scale regions of the genome have experienced relatively little recombination with respect

to the ancestral biallelic pattern of *T.gondii* (8,9). Importantly, the behavior of these nonhomogeneous regions did not differ from that of the rest of the genome in the experimental crosses. While some showed enhanced levels of recombination at the transition from fixed to variable polymorphism (i.e. markers AK97–AK39 on chromosome III, Figures 2 and 6), most of these regions did not. A more global analysis of the distribution of SNPs across the genome has detected a similar pattern (Jon Boyle and John Boothroyd, unpublished data; http://boothroydlab.stanford.edu/snp_maps/) and resolving the importance of these patterns awaits further investigation.

We report here on a combined genome map for *T.gondii* consisting of 14 chromosomes ranging in size from slightly <2 to >7 Mb. We have placed 250 markers at ~300 kb intervals across the map, thus connecting the genetic linkage maps with the major scaffolds from whole genome sequence. The present genome map provides excellent coverage for most regions, and any current deficiencies could be corrected by future genetic crosses between the type I and II lineages that would provide informative markers for the fallow regions that show strong SNP bias. Future efforts at increasing the density of markers on the map will also provide greater resolution for mapping complex traits that differ between the lineages such as drug resistance, virulence, transmission and development.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Martin Franzholz and Bindu Gajria for assistance with ToxoDB, Sandra Clifton, Deana Pape and the EST Sequencing Team at The Washington University Genome Sequencing Center, Mike Quail for production of the BAC-end library, Tovi Lehman and Michael Grigg for sharing unpublished data on genetic markers, Merck Research Laboratories for providing arprinocid-*N*-oxide, and Noelle Holmes and Julie Suetterlin for expert technical assistance. Preliminary genomic and/or cDNA sequence data was accessed via http://ToxoDB.org/ and/or http://www.tigr.org/tdb/t_gondii/. Genomic sequence data were provided by The Institute for Genomic Research (supported by the NIH grant AI50930), and by the Wellcome Trust Sanger Institute. Financial support was provided by the National Institutes of Health through the following grants (AI045806, AI36629, AI059176, AI28724 and AI40037). Funding to pay the Open Access publication charges for this article was provided by NIH grants AI36629 and AI059176.

Conflict of interest statement. None declared.

REFERENCES

1. Joynson,D.H.M. and Wreghitt,T.G. (2001) *Toxoplasmosis: A Comprehensive Clinical Guide*. Cambridge University Press, Cambridge. pp. 395.
2. Dubey,J.P. (1977) In Kreier,J.P. (ed.), *Parasitic Protozoa*. Academic Press, New York, pp. 101–237.
3. Roos,D.S., Donald,R.G.K., Morrissette,N.S. and Moulton,A.L. (1994) Molecular tools for genetic dissection of the protozoan parasite *Toxoplasma gondii*. *Methods Cell Biol*., **45**, 27–63.
4. Sibley,L.D., Mordue,D.G., Su,C., Robben,P.M. and Howe,D.K. (2002) Genetic approaches to studying virulence and pathogenesis in *Toxoplasma gondii*. *Philos. Trans. R. Soc. Lond., B, Biol. Sci*., **357**, 81–88.
5. Boothroyd,J.C., Black,M., Bonnefoy,S., Hehl,A., Knoll,L.J., Manger,I.D., Ortega-Barria,E. and Tomavo,S. (1997) Genetic and biochemical analysis of development in *Toxoplasma gondii*. *Phil. Trans. R. Soc. Lond., B, Biol. Sci*., **352**, 1347–1354.
6. 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.
7. Sibley,L.D. and Boothroyd,J.C. (1992) Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature*, **359**, 82–85.
8. Grigg,M.E., Bonnefoy,S., Hehl,A.B., Suzuki,Y. and Boothroyd,J.C. (2001) Success and virulence in *Toxoplasma* as the result of sexual recombination between two distinct ancestries. *Science*, **294**, 161–165.
9. Su,C., Evans,D., Cole,R.H., Kissinger,J.C., Ajioka,J.W. and Sibley,L.D. (2003) Recent expansion of *Toxoplasma* through enhanced oral transmission. *Science*, **299**, 414–416.
10. Dubey,J.P. and Frenkel,J.K. (1976) Feline toxoplasmosis from acutely infected mice and the development of *Toxoplasma* cysts. *J. Protozool*., **23**, 537–546.
11. Pfefferkorn,E.R., Pfefferkorn,L.C. and Colby,E.D. (1977) Development of gametes and oocysts in cats fed cysts derived from cloned trophozoites of *Toxoplasma gondii*. *J. Parasitol*., **63**, 158–159.
12. Cornelissen,A.W.C.A. and Overdulve,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.
13. Pfefferkorn,L.C. and Pfefferkorn,E.R. (1980) *Toxoplasma gondii*: genetic recombination between drug resistant mutants. *Exp. Parasitol*., **50**, 305–316.
14. Sibley,L.D., LeBlanc,A.J., Pfefferkorn,E.R. and Boothroyd,J.C. (1992) Generation of a restriction fragment length polymorphism linkage map for *Toxoplasma gondii*. *Genetics*, **132**, 1003–1015.
15. Su,C., Howe,D.K., Dubey,J.P., Ajioka,J.W. and Sibley,L.D. (2002) Identification of quantitative trait loci controlling acute virulence in *Toxoplasma gondii*. *Proc. Natl Acad. Sci. USA*, **99**, 10753–10758.
16. Manger,I.D., Hehl,A., Parmley,S., Sibley,L.D., Marra,M., Hillier,L., Waterston,R. and Boothroyd,J.C. (1998) Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: identification of developmentally regulated genes. *Infect. Immun*., **66**, 1632–1637.
17. Li,L., Brunk,B.P., Kissinger,J.C., Pape,D., Tang,K., Cole,R.H., Martin,J., Wylie,T., Dante,M., Fogarty,S.J. *et al*. (2003) Gene discovery in the apicomplexa as revealed by EST sequencing and assembly of a comparative gene database. *Genome Res*., **13**, 443–454.
18. Ajioka,J.W., Boothroyd,J.C., Brunk,B.P., Hehl,A., Hillier,L., Manger,I.D., Marra,M., Overton,G.C., Roos,D.S., Wan,K.L. *et al*. (1998) Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa. *Genome Res*., **8**, 18–28.
19. Pfefferkorn,E.R. and Pfefferkorn,L.C. (1976) Arabinosyl nucleosides inhibit *Toxoplasma gondii* and allow the selection of resistant mutants. *J. Parasitol*., **62**, 993–999.
20. Pfefferkorn,E.R. and Pfefferkorn,L.C. (1977) *Toxoplasma gondii*: characterization of a mutant resistant to 5-fluorodeoxyuridine. *Exp. Parasitol*., **42**, 44–55.
21. Pfefferkorn,E.R. and Pfefferkorn,L.C. (1979) Quantitative studies of the mutagenesis of *Toxoplasma gondii*. *J. Parasitol*., **65**, 364–370.
22. Lander,E.S., Green,P., Abrahamson,J., Barlow,A., Daly,M.J., Lincoln,S.E. and Newburg,L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174–181.
23. Manly,K.F., Cudmore,R.H.,Jr and Meer,J.M. (2001) Map Manager QTX, cross-platform software for genetic mapping. *Mamm. Genome*, **12**, 930–932.
24. Mu,J., Ferdig,M.T., Feng,X., Joy,D.A., Duan,J., Furuya,T., Subramanian,G., Aravind,L., Cooper,R.A., Wootton,J.C. *et al*. (2003) Multiple transporters associated with malaria parasite responses to chloroquine and quinine. *Mol. Microbiol*., **49**, 977–989.
25. Karlin,S. and Liberman,U. (1994) Theoretical recombination processes incorporating interference effects. *Theor. Popul. Biol*., **46**, 198–231.
26. Su,X., Ferdig,M.T., Huang,Y., Huynh,C.Q., Liu,A., You,J., Wootton,J.C. and Wellems,T.E. (1999) A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science*, **286**, 1351–1353.

27. Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics*, **52**, 1–8.
28. Kent,W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res*., **12**, 656–664.
29. Li,L., Crabtree,J., Fischer,S., Pinney,D., Stoeckert,C.J.Jr, Sibley,L.D. and Roos,D.S. (2004) ApiEST-DB: analyzing clustered EST data of the apicomplexan parasites. *Nucleic Acids Res*., **32**, D326–D328.
30. Sibley,L.D. and Boothroyd,J.C. (1992) Construction of a molecular karyotype for *Toxoplasma gondii*. *Mol. Biochem. Parasitol*., **51**, 291–300.
31. Ferdig,M.T., Cooper,R.A., Mu,J., Deng,B., Joy,D.A., Su,X.Z. and Wellems,T.E. (2004) Dissecting the loci of low-level quinine resistance in malaria parasites. *Mol. Microbiol*., **52**, 985–997.
32. Su,X.Z. and Wootton,J.C. (2004) Genetic mapping in the human malaria parasite *Plasmodium falciparum*. *Mol. Microbiol*., **53**, 1573–1582.
33. Pfefferkorn,E.R. (1978) *Toxoplasma gondii*: the enzymic defect of a mutant resistant to 5-fluorodeoxyuridine. *Exp. Parasitol*., **44**, 26–35.
34. Donald,R.G. and Roos,D.S. (1995) Insertional mutagenesis and marker rescue in a protozoan parasite: cloning the uracil phosphoribosyltransferase locus from *Toxoplasma gondii*. *Proc. Natl Acad. Sci. USA*, **92**, 5749–5753.
35. Sullivan,W.J.,Jr, Chiang,C.W., Wilson,C.M., Naguib,F.N., el Kouni,M.H., Donald,R.G. and Roos,D.S. (1999) Insertional tagging of at least two loci associated with resistance to adenine arabinoside in *Toxoplasma gondii*, and cloning of the adenosine kinase locus. *Mol. Biochem. Parasitol*., **103**, 1–14.
36. Pfefferkorn,E.R. and Pfefferkorn,L.C. (1978) The biochemical basis for resistance to adenine arabinoside in a mutant of *Toxoplasma gondii*. *J. Parasitol*., **64**, 486–492.
37. Martin,J.L. and McMillan,F.M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol*., **12**, 783–793.
38. Schluckebier,G., Kozak,M., Bleimling,N., Weinhold,E. and Saenger,W. (1997) ??Differential binding of S-adenosylmethionine, S-adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M*Taq*I. *J. Mol. Biol*., **265**, 56–67.
39. Pfefferkorn,E.R. and Kasper,L.H. (1983) *Toxoplasma gondii*: genetic crosses reveal phenotypic suppression of hydroxyurea resistance by fluorodeoxyuridine resistance. *Exp. Parasitol*., **55**, 207–218.
40. Pfefferkorn,E.R., Eckel,M.E. and McAdams,E. (1988) *Toxoplasma gondii: in vivo* and *in vitro* studies of a mutant resistant to arprinocid-N-oxide. *Exp. Parasitol*., **65**, 282–289.
41. Carlton,J., Mackinnon,M. and Walliker,D. (1998) A chloroquine resistance locus in the rodent malaria parasite *Plasmodium chabaudi*. *Mol. Biochem. Parasitol*., **93**, 57–72.
42. Shirley,M.W. and Harvey,D.A. (2000) A genetic linkage map of the apicomplexan protozoan parasite *Eimeria tenella*. *Genome Res*., **10**, 1587–1593.
43. Fidock,D.A., Nomura,T., Talley,A.K., Cooper,R.A., Dzekunov,S.M., Ferdig,M.T., Ursos,L.M., Sidhu,A.B., Naude,B., Deitsch,K.W. *et al*. (2000) Mutations in the *P.falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell*, **6**, 861–871.