A Single Chromosome Unexpectedly Links Highly Divergent Isolates of *Toxoplasma gondii*

Katelyn A. Walzer and Jon P. Boyle

Department of Biological Sciences, Kenneth P. Dietrich School of Arts and Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

ABSTRACT ***Toxoplasma gondii*** **is an obligate intracellular parasite that can cause disease in all warm-blooded animals studied to date, including humans. Over a billion people have been infected with this parasite worldwide. In Europe and North America,** ***Toxoplasma*** **has a clonal population structure, where only three lineages are highly dominant (strain types I, II, and III). Khan et al. [mBio 2(6): e00228-11, 2011] have carried out phylogenetic analyses on a large number of diverse strains from outside of these lineages and found evidence for a significant split between the clonal North American/European lineages and those in South America. In contrast to most of the genome, nearly all North American/European strains sampled, and the majority of South American strains sampled, harbored at least portions of a monomorphic chromosome Ia (Ia*). In contrast to previous models, these data suggest that the monomorphic haplotype originated in South America and migrated to the North. These authors propose that South American haplotype 12 was a precursor to modern-day type II, while South American haplotypes 6 and 9 crossed with haplotype 12 to give rise to the type I and III lineages, respectively. However, the findings reported by Khan et al. complicate the origin of chromosome Ia, since there are members of haplotypes 9 and 12 with nearly complete versions of Ia* and members of haplotypes 6 and 12 with over 50% of Ia*. This unexpected finding raises exciting new questions about how an entire common chromosome can be found within strains that are highly divergent at most other genomic loci.**

TOXOPLASMA GONDII IS A WIDESPREAD PATHOGEN OF HUMANS AND OTHER ANIMALS

Toxoplasma gondii is a globally ubiquitous parasite of humans and all other warm-blooded animals. Over 20% of the world population has been infected with this parasite, which can be fatal in immunocompromised individuals and for developing fetuses. The global spread of this pathogen has likely been driven by the uniquely broad host range of *T. gondii*. In most eukaryotic parasites, and even the closest extant relatives of *T. gondii*, a highly restricted host range is the norm. For example, individual *Plasmodium* species only use a handful of mosquito species for their transmission, and those with rodents as mammalian hosts cannot infect humans and vice versa. In stark contrast, *T. gondii* can successfully infect, and cause disease in, all warm-blooded animals studied to date (including birds; for example, see references 1 and 2). The only species specificity in its life cycle is that the sexual phase of development can occur only in members of the family Felidae, and domestic cats are a source of infection in humans.

IN EUROPE AND NORTH AMERICA, THE *T. GONDII* POPULATION STRUCTURE IS BIASED STRONGLY TOWARDS CLONALITY

From the large amount of work examining the population structure of *Toxoplasma*, it is clear that it propagates clonally. This has been well documented in Europe and North America, where only three clonal lineages are responsible for the vast majority of infections in humans and other animals (3, 4). Within-lineage variation in these strain types (I, II, and III) is approximately 0.01% at the nucleotide level, while between-lineage variation is much higher, ranging from 0.01 to 5% depending on the region of the genome queried (5). How is this clonality maintained? A major contributing factor is that *T. gondii* is not an obligate sexual parasite, in that intermediate hosts can transmit the parasite to other intermediate hosts via carnivory or scavenging. Moreover, during the sexual phase of reproduction in the cat, self-mating occurs quite efficiently in single-strain infections. It has been hypothesized that genetic recombination between distinct lineages (which can occur only when a feline is infected with two distinct strains simultaneously) is rare in European and North American *T. gondii* populations (3). This is based on the fact that while isolates that are natural recombinants between the dominant clonal lineages have been found (for example, see reference 6), they appear to be the exception, at least based on limited genotyping (3).

A SINGLE CHROMOSOME LINKS DIVERGENT LINEAGES

In their study (7), Khan et al. follow up on their previous work examining the broader *T. gondii* population structure (8). Much of the work in the *Toxoplasma* field has focused only on types I, II, and III, but when introns from multiple unlinked loci were sequenced in a variety of *T. gondii* isolates from diverse locales, a clear split between the strain isolates in Europe and North America and those from South America was observed. In this previous study and others (5, 8, 9), it was also observed that while multiple unlinked loci distinguished types I, II, and III, those on chromosome Ia showed very few, if any, sequence polymorphisms between them. This monomorphic haplotype on chromosome Ia, as well as bi-allelic haplotypes on other chromosomes, contributed to a model in which a single progenitor strain (a precursor to the type II lineage) was a parent of both the type I and type III lineages. Based on this model, the most parsimonious explanation was that both the type I and III lineages obtained chromosome Ia from the type II progenitor in a limited number of crosses (5).

However, the data from Khan (7) et al. provide good evidence that monomorphic chromosome Ia (Ia*) may have actually had its origins

Published 3 January 2011

Citation Walzer KA, Boyle JP. 2012. A single chromosome unexpectedly links highly divergent isolates of *Toxoplasma gondii*. mBio 3(1):e00284-11. doi:10.1128/mBio.00284-11.

Address correspondence to Jon P. Boyle, boylej@pitt.edu.

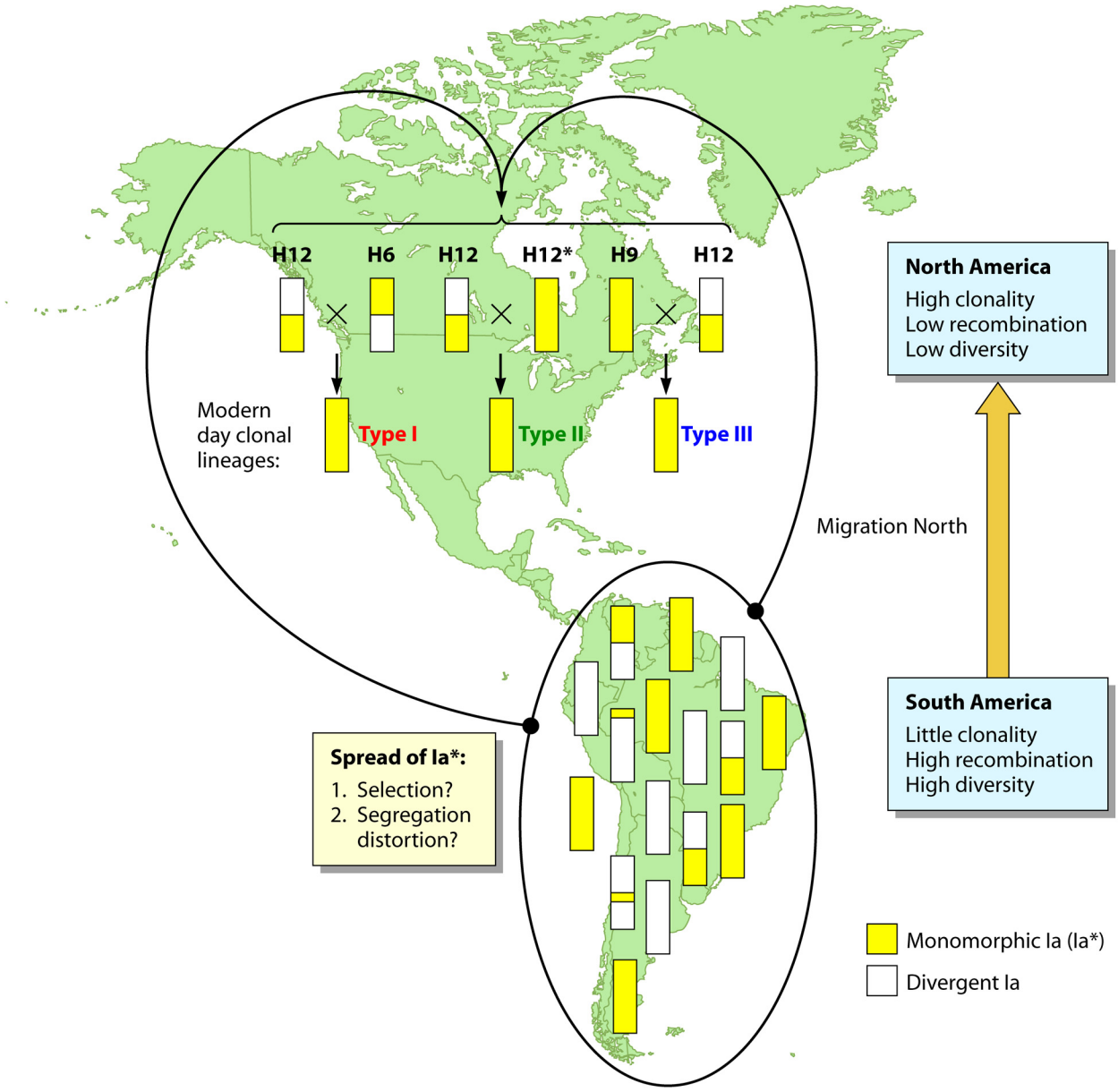

FIG 1 One possible model for the migration of strains harboring monomorphic chromosome Ia (Ia*) from South America to North America and hypothetical crosses that could have led to the modern-day North American and European clonal lineages (I, II, and III). South American strains exhibit a comparatively high level of recombination, and Ia* could have spread throughout this interbreeding population due to selection and/or segregation bias. Human activity and/or avian migration could have then facilitated the movement of South American strains harboring Ia* to North America. These strains then recombined to produce the ancestors to the modern-day clonal lineages that are dominant in Europe and North America. The genealogy proposed for the origins of clonal types I, II, and III is based on the number of strains harboring Ia* in each haplotype group. For example, most of the sampled strains belonging to haplotype 12 harbor Ia chromosomes that are only partially monomorphic, as is the case for haplotype 6. Haplotype 12* (H12*) is a hypothetical haplotype with a fully monomorphic Ia, similar to strain B41 in the work of Khan et al. (7).

in South America and migrated to the north. By direct sequencing of 12 loci across chromosome Ia in 33 diverse *T. gondii* isolates from locations in North America, Europe, China, Africa, and South America, Khan et al. found that 27 of these have the monomorphic sequence for at least 3 of the 12 loci and even more remarkably that 24 of these have the monomorphic sequence for at least 6 of the 12 loci. While these lineages share the Ia* haplotype, Khan et al. found that unlinked loci from 6 other chromosomes were much more divergent among the sampled isolates, implying that Ia* has a highly distinct ancestry compared to the rest of the genome. To do this, they used an estimated neutral mutation rate to estimate the time to most recent common ancestry (TMRCA) between North American and South American strains and found a TMRCA of ~10^5 years. In contrast, the TMRCA of the North American lineages was ~10^4, suggesting that, in contrast to a previous model (8), *Toxoplasma* had its origins in South America and then was introduced to North America and Europe. The

age of Ia* within the North American lineages was also estimated to be ~10^4 years. This provides a new insight into the origins of chromosome Ia*: not only is it found extensively in highly divergent South American strains as well as highly clonal North American lineages, it is much more likely that it was introduced into the North from the South rather than having its origins in North America (Fig. 1).

This model is consistent with previously published studies examining the *T. gondii* worldwide population, albeit at only a few genomic loci (10). This leads to a model in which one or multiple ancestral strains from South America harboring Ia* were parents in the limited number of crosses proposed to have produced North American lineages I, II, and III. In their paper (7), Khan et al. suggest that haplotype 12 is likely to have been a precursor to modern-day type II, while haplotypes 6 and 9 were parents in a cross (or crosses) with haplotype 12 to give rise to the type I and III lineages, respectively. This fits well with our previously published model (5). However, the origins of Ia* are more difficult to determine. In our original model we proposed that Ia* present in lineages I and III came from a type II-like ancestor. Yet the work of Khan et al. demonstrates that this model could be more complicated, since there are members of both haplotypes 12 and 9 with nearly "complete" versions of Ia* and members of haplotypes 12 and 6 with over 50% of Ia*. Therefore, it is equally likely that the Ia* in lineage III came from either haplotype 12 or haplotype 9, and the origins of Ia* in lineage I are still uncertain (Fig. 1). It is certainly possible that as more strains are sampled from these haplotypes, this question may be answered.

RECOMBINATION IN *T. GONDII*: RARE OR PREVALENT?

This study raises interesting questions about the role of recombination in *T. gondii* population biology. The recombination rate in *T. gondii* is low considering the size of the genome (~100 kb per centimorgan) (9), and experimental crosses show that entire chromosomes from one parent or the other can be found in the F1 progeny without recombination (11, 12). Therefore, the fact that this entire monomorphic chromosome can be found to be fully intact, or nearly so, in multiple strains is not surprising. But the question remains: why is Ia* found in such diverse strains? This is the most interesting conundrum emerging from the study by Khan et al. Has Ia* spread throughout the population by random chance? This could explain the fact that Ia* is found in most North American strains, since they are of recent common ancestry, are likely to have shared parentage, and therefore share large portions of their genomes (5). But this seems like an unlikely explanation for the presence of Ia* in South America given the diversity of the strains in which it is found, and their geographical origins. Does Ia* confer a selective advantage on the progeny that harbor it? This is an exciting possibility but one that still awaits direct testing. To date, all experimental crosses have been performed with strains harboring the monomorphic haplotype, but Khan et al. discuss future experiments performing experimental crosses between strains harboring Ia* and divergent chromosome Ia to test this hypothesis. It will be interesting to see what phenotypes, if any, Ia* confers on its progeny and how these may be linked to actual "success" in the field. Another possibility is that there is a bias during sexual recombination for Ia* to be passed on to the F1 progeny. This could help to explain how Ia* is present within strains that appear, at least at six other locations in the genome, to be of completely different ancestry. This is most apparent in comparisons of South American strains to North American strains. Again, this can be tested directly in experimental crosses. Regardless of the "correct" model about the origins of Ia*, it is likely that after its origin it spread among multiple divergent lineages (possibly due to selection and/or recombinational bias). This may have been facilitated by a comparatively high incidence of recombination among strains found in South America compared to North America and Europe. Regardless, Ia* has emerged as a remarkable common link among a large percentage of the *T. gondii* population in Europe, North America, and South America. As more isolates are identified from more diverse locales, it will be exciting to see how prevalent Ia* is throughout the global *T. gondii* population.

FUTURE DIRECTIONS

The work by Khan et al. sets the stage for future studies using complete genome sequence data that are presently emerging from the *Toxoplasma* genomics project at the J. Craig Venter Genomic Sequencing center for Infectious Diseases (GSCID; http://gsc.jcvi.org/projects/gsc/t_gondii/index.shtml). Over 60 diverse *T. gondii* isolates that represent what is currently known about the global diversity of *T. gondii* are in the pipeline for whole-genome shotgun sequencing. Sequence data from these strains will greatly improve our understanding of the *T. gondii* phylogeny and, with respect to the observations of Khan et al., will allow a more robust determination of the prevalence of Ia* in the population, its relative age in North and South American isolates, and how this compares to the divergence between these two major populations at loci throughout the genome. This is an exciting time to be working on population genetics in *Toxoplasma*, given the explosion of new data, which allows whole genomes, rather than fragments thereof, to be efficiently obtained and compared.

REFERENCES

1. **Dubey JP.** 2002. A review of toxoplasmosis in wild birds. Vet. Parasitol. **106**:121–153.
2. **Hartley WJ, Dubey JP.** 1991. Fatal toxoplasmosis in some native Australian birds. J. Vet. Diagn. Invest. **3**:167–169.
3. **Howe DK, Sibley LD.** 1995. *Toxoplasma gondii* comprises three clonal lineages: correlation of parasite genotype with human disease. J. Infect. Dis. **172**:1561–1566.
4. **Sibley LD, Khan A, Ajioka JW, Rosenthal BM.** 2009. Genetic diversity of *Toxoplasma gondii* in animals and humans. Philos. Trans. R. Soc. Lond. B Biol. Sci. **364**:2749–2761.
5. **Boyle JP, et al.** 2006. Just one cross appears capable of dramatically altering the population biology of a eukaryotic pathogen like *Toxoplasma gondii*. Proc. Natl. Acad. Sci. U. S. A. **103**:10514–10519.
6. **Bontell IL, et al.** 2009. Whole genome sequencing of a natural recombinant *Toxoplasma gondii* strain reveals chromosome sorting and local allelic variants. Genome Biol. **10**:R53.
7. **Khan A, et al.** 2011. A monomorphic haplotype of chromosome Ia is associated with widespread success in clonal and nonclonal populations of *Toxoplasma gondii*. mBio **2**(6):e00228-11.
8. **Khan A, et al.** 2007. Recent transcontinental sweep of *Toxoplasma gondii* driven by a single monomorphic chromosome. Proc. Natl. Acad. Sci. U. S. A. **104**:14872–14877.
9. **Khan A, et al.** 2005. Composite genome map and recombination parameters derived from three archetypal lineages of *Toxoplasma gondii*. Nucleic Acids Res. **33**:2980–2992.
10. **Lehmann T, Marcet PL, Graham DH, Dahl ER, Dubey JP.** 2006. Globalization and the population structure of *Toxoplasma gondii*. Proc. Natl. Acad. Sci. U. S. A. **103**:11423–11428.
11. **Sibley LD, LeBlanc AJ, Pfefferkorn ER, Boothroyd JC.** 1992. Generation of a restriction fragment length polymorphism linkage map for *Toxoplasma gondii*. Genetics **132**:1003–1015.
12. **Behnke MS, et al.** 2011. Virulence differences in *Toxoplasma* mediated by amplification of a family of polymorphic pseudokinases. Proc. Natl. Acad. Sci. U. S. A. **108**:9631–9636.