Polymorphic family of injected pseudokinases is paramount in *Toxoplasma* virulence

Michael L. Reese[a], Gusti M. Zeiner[a,1], Jeroen P. J. Saeij[a,2], John C. Boothroyd[a,3,4], and Jon P. Boyle[a,b,3,4]

[a]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305; and [b]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

Edited by Thomas E. Wellems, National Institutes of Health, Bethesda, MD, and approved February 25, 2011 (received for review October 26, 2010)

***Toxoplasma gondii*, an obligate intracellular parasite of the phylum Apicomplexa, has the unusual ability to infect virtually any warm-blooded animal. It is an extraordinarily successful parasite, infecting an estimated 30% of humans worldwide. The outcome of *Toxoplasma* infection is highly dependent on allelic differences in the large number of effectors that the parasite secretes into the host cell. Here, we show that the largest determinant of the virulence difference between two of the most common strains of *Toxoplasma* is the *ROP5* locus. This is an unusual segment of the *Toxoplasma* genome consisting of a family of 4–10 tandem, highly divergent genes encoding pseudokinases that are injected directly into host cells. Given their hypothesized catalytic inactivity, it is striking that deletion of the *ROP5* cluster in a highly virulent strain caused a complete loss of virulence, showing that ROP5 proteins are, in fact, indispensable for *Toxoplasma* to cause disease in mice. We find that copy number at this locus varies among the three major *Toxoplasma* lineages and that extensive polymorphism is clustered into hotspots within the ROP5 pseudokinase domain. We propose that the *ROP5* locus represents an unusual evolutionary strategy for sampling of sequence space in which the gene encoding an important enzyme has been (*i*) catalytically inactivated, (*ii*) expanded in number, and (*iii*) subject to strong positive selection. Such a strategy likely contributes to *Toxoplasma*'s successful adaptation to a wide host range and has resulted in dramatic differences in virulence.**

copy number variation | host–pathogen interaction

The evolutionary pressure on the molecules that form the interface between a pathogenic organism and its host has been likened to an arms race (1, 2) and has been proposed to give rise to highly polymorphic genes such as those that encode host MHC and viral capsid proteins. Whereas most pathogens have a single or restricted range of hosts with which they coevolve, the obligate intracellular parasite *Toxoplasma gondii* can infect virtually any cell of almost any warm-blooded animal, a remarkable feat from an evolutionary perspective.

Different strains of *Toxoplasma* cause significantly different disease in mice (3) and humans as well (4, 5). In North America and Europe, three strains have grown to dominate *Toxoplasma*'s genetic landscape, and they are characterized by distinct disease outcomes in mice: although a single Type I parasite is uniformly lethal to a mouse, Type II and III parasites have a lethal dose (LD_{50}) ranging from 10^2 to 10^5 (3). Sexual crosses between members of these lineages (Type I × Type III and Type II × Type III) (6, 7) have shown that virulence in mice segregates among F1 progeny (8). This has allowed for the discovery of a number of virulence loci using quantitative trait loci (QTL) mapping (9, 10).

Toxoplasma secretes an arsenal of effector proteins into its host cell from specialized secretory organelles termed rhoptries and dense granules (11–14). Given the selective pressure of the interaction with its varied hosts, it is not surprising that many of *Toxoplasma*'s effectors are among the most polymorphic proteins between the major strains. Indeed, allelic differences in *Toxoplasma*'s secreted effectors have been found to be associated with differences in disease, both through differential antigen recognition by the host (15) and through direct action of the parasite effectors themselves (9, 10, 16). In the latter category, two rhoptry proteins (ROPs) with kinase folds, ROP16 (16) and ROP18 (9, 10), were identified as key virulence factors by QTL mapping. ROP18 seems to function by phosphorylating and inactivating host immunity-related GTPases (17, 18), and ROP16 is a tyrosine kinase that directly phosphorylates the host STAT family of proteins (19, 20).

The QTL analysis of a cross between members of the Types II and III lineages that led to the identification of *ROP16* and *ROP18* indicated that a third virulence locus (*VIR1*), not then identified, had the greatest impact on virulence in mice (9). Here, we identify that locus to be a cluster of tandemly duplicated polymorphic pseudokinases originally annotated as a single gene (*ROP5*). We further show that the *ROP5* cluster is required for *Toxoplasma* to cause lethal disease in mice, because targeted deletion of the entire locus attenuated an otherwise highly virulent strain. The evolutionary pressures and mechanisms that may have led to the emergence of *ROP5* as the key virulence locus are discussed.

Results

Previously, in a QTL mapping study using F1 progeny from a cross (6) between a Type II and III strains of *Toxoplasma*, we identified a 0.98-Mbp region of chromosome XII (the *VIR1* locus) that was responsible for over 50% of the inherited variation in virulence (logarithm of odds = 9.4) (9). In that study, the Type III allele was predicted to contain the more virulent version relative to Type II. To predict which gene encodes this major virulence determinant, we used several criteria, including predicted secretion of the gene product into the host cell, predicted function (e.g., those having predicted kinase or protease domains), and polymorphism between Type II and III strains. Based on these criteria, we determined that a strong candidate was *ROP5*, a member of the ROP2 superfamily. The ROP2 superfamily is made up of protein kinase-fold proteins (21) that originate in *Toxoplasma*'s apical secretory organelles (called rhoptries) and are injected into the host cytoplasm on invasion. Many rhoptry proteins, including ROP5, subsequently associate

Author contributions: M.L.R., G.M.Z., J.C.B., and J.P.B. designed research; M.L.R., G.M.Z., J.P.S., J.C.B., and J.P.B. performed research; M.L.R., G.M.Z., J.P.S., and J.P.B. analyzed data; and M.L.R., J.C.B., and J.P.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. HQ916448–HQ916460).

[1]Present address: Molecular Tools Lab, Agilent Technologies, Santa Clara, CA 95051.

[2]Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

[3]J.C.B. and J.P.B. contributed equally to this work.

[4]To whom correspondence may be addressed. E-mail: John.Boothroyd@stanford.edu or boylej@pitt.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015980108/-/DCSupplemental.

MICROBIOLOGY

directly with the host cytosolic face of the parasitophorous vacuolar membrane (PVM) (22, 23). One member of the family, ROP18, is a catalytically active kinase that has been identified previously as a key determinant of virulence on chromosome VIIa (9, 10). ROP5, however, has substituted a basic residue for the catalytic HRD motif Asp (invariant in active kinases) (24) and thus, is considered a catalytically inactive pseudokinase (21). These features made ROP5 a particularly intriguing candidate for the *VIR1* QTL. Although many pathogens disrupt host signaling with secreted enzymes (19, 20, 25) or through secreted nonenzymatic activators or inhibitors of host signaling (26, 27), we know of no reports of virulence factors in any pathogen that, like ROP5, possess an enzymatic fold but lack catalytic activity.

ROP5 Is a Crucial Virulence Locus. As a preliminary test of whether *ROP5* is indeed the locus responsible for the *VIR1* QTL, we complemented the S22 strain of *Toxoplasma* (8, 28) with a cosmid containing a predicted high-virulence allele of the *ROP5* locus (derived from the highly virulent Type I strain RH). S22 is a hypovirulent progeny of the Type II × Type III cross (8), carrying the low-virulence allele for *ROP16* and *ROP18* (9) and the low-virulence (Type II) allele at the *VIR1* locus, and thus, it represents an ideal starting point for studies of virulence in this model. Based on the LD_{50} in mice, the complemented strain showed an increase in virulence of $>10^5$-fold compared with the parental strain (Fig. S1). This strongly supported the hypothesis that *ROP5* was indeed the *VIR1* locus; however, because there are two additional genes unrelated to *ROP5* on the cosmid (TGGT1_042730 and TGGT1_042740), it was possible that one of these latter genes was responsible for the differences that we saw.

As discussed in detail below, preliminary indications were that the *ROP5* locus consists of multiple tandem genes, and therefore, rather than deal with each of these in turn, we chose to knockout the entire locus as a way to quickly validate *ROP5*'s role in virulence. We chose to do our genetic manipulation in a Type I line; this is because Type I and III parasites have very similar *ROP5* alleles (see below) and the Type I RH strain is much more easily manipulated than Type III strains. In addition, the Type I strain is highly virulent (infection with one parasite kills most strains of mice 100% of the time) (29), making it particularly suitable for observing a potential reduction in virulence. Using a strain with increased efficiency of homologous recombination (parental; RH*Δku80Δhxgprt*) (30), we created parasites in which the entire *ROP5* locus was replaced with an *HXGPRT* selection cassette (RH*Δrop5*) (Fig. S2). Lack of ROP5 protein expression was confirmed by Western blot (Fig. 1*A*). RH*Δrop5* parasites exhibited no detectable growth phenotype during repeated passage in vitro or any detectable differences in plaque size on a fibroblast monolayer (Fig. S3).

We found that, although mice injected intraperitoneally with 100 parasites of the parental strain all succumbed within 7 d (as expected), mice injected with as many as 10^6 RH*Δrop5* parasites uniformly survived (10^6 was the highest dosage tested) (Fig. 1*B*). This dramatic result was not dependent on mouse strain, because both BALB/c and outbred CD-1 mice showed the same sensitivity to parental parasites and resistance to infection with RH*Δrop5* (Fig. S4). It is important to note that 10^6 parasites represents a dosage at least 10-fold greater than the LD_{50} for any of the three major strains (8, 29). These results indicate that the *ROP5* locus is indeed crucial for the RH strain to establish a lethal infection in mice and is likely the genetic basis for the *VIR1* QTL.

ROP5 Is a Cluster of Paralogs. To prove that the loss of virulence was caused by the absence of ROP5 rather than some other incidental effect of generating the knockout strain, we needed to complement the mutant with an ectopically expressed copy of ROP5. Surprisingly, when we attempted to clone *ROP5* from either Type I genomic or cosmid DNA by PCR, we obtained several distinct sequences, none of which matched the sequence predicted in ToxoDBv6.1 (www.toxodb.org). A likely explanation for this is the presence of multiple, nearly identical genes at the *ROP5* locus, and the algorithm used to assemble the genome had viewed these as one sequence and mistakenly compressed them all into a single chimeric sequence. Consistent with this hypothesis, we noted a high degree of oversampling of the *ROP5* locus in genomic sequencing reads for each of the major strains (Fig. 2*A*), which provides an accurate estimate of copy number variation (31). To confirm the existence of tandem *ROP5* copies experimentally, we digested genomic DNA from Types I, II, and III strains with ApaI, which our preliminary sequencing indicated cut outside and not within the *ROP5* locus in all three strains (Fig. 2*B*). The resulting DNA was separated by field inversion gel electrophoresis, transferred to a membrane, and probed with radiolabeled DNA derived from the region encoding the ROP5 arginine-rich amphipathic helix domain (RAH) (22). As predicted by the differential oversampling of the sequence reads, we observed bands of different sizes in each strain, ranging from ~19 kbp for Type III to ~46 kbp for Type II (Fig. 2*C*). These

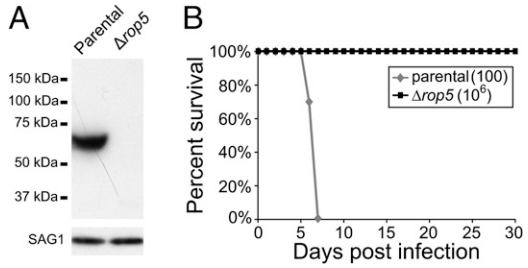

Fig. 1. ROP5 is required for virulence in mice. (*A*) Western blot of a parasite lysate probed with mouse αROP5 polyclonal Ab and SAG1 loading control. (*B*) BALB/c mice (n = 10) were infected with either 100 parental RH*Δku80* parasites or 10^6 RH*Δrop5* parasites and monitored for 30 d.

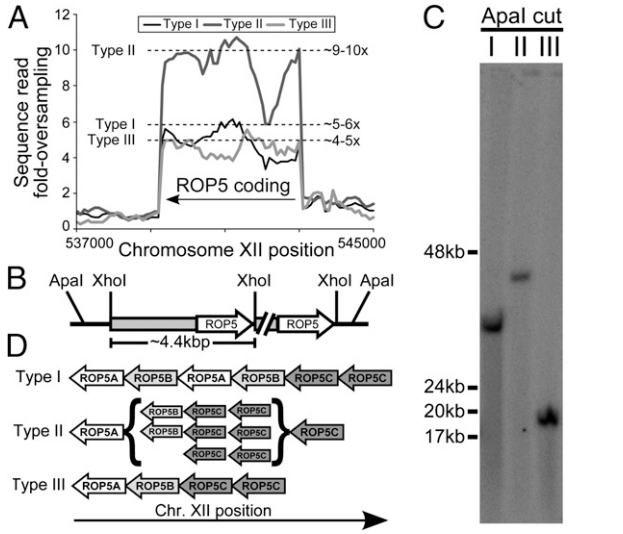

Fig. 2. The *ROP5* locus is differentially expanded in different *Toxoplasma* strains. (*A*) Quantification of sequencing read oversampling of each of the strains (black, Type I/GT1; dark gray, Type II/ME49; light gray, Type III/VEG). (*B*) Diagram of ROP5 locus restriction sites. (*C*) Genomic DNA from each of the three dominant strains was digested with ApaI and analyzed by Southern blotting using a probe from the conserved N terminal-encoding region of ROP5. (*D*) Diagram of the most likely locus organization of the three isoforms (Fig. 3) found in each strain using diagnostic double digests (Fig. S5*B*). Copies whose order is ambiguous are indicated with braces.

results supported the case for a variable number of tandem copies of *ROP5* in the genomes of the three dominant strains and confirmed their relative copy number in each strain. We also quantified the locus expansion by repeating our analysis on parasite genomic DNA digested with XhoI, which cuts within the repeat unit (Fig. S5*A*); these results confirmed the estimates from the other two methods. Collectively, these data established that each strain has a different locus structure: Type I has ~6, Type II has ~10, and Type III has ~4 tandem copies of ROP5. We specifically amplified the distal genes in each cluster by PCR using primers upstream or downstream of the locus and dubbed these copies A and C, respectively. Southern blots with diagnostic double digests (Fig. S5*B*) (based on the sequences discussed below) were used to determine the most likely organization of the remaining genes in the locus. The results showed that, between the A and C copies, each strain had a variable number of additional copies of the A or C isoforms and a third distinct isoform, dubbed B (Fig. 2*D*).

ROP5 Paralogs Are Highly Divergent. To better understand the evolution of the ROP5 paralogs and inform the decision of which to use in the complementation study, it was necessary to determine their individual sequences. To do this and avoid obtaining chimeric sequences from PCR, we cloned the *ROP5* paralogs directly from genomic DNA. Genomic DNA from Type II (containing an avirulent cluster of *ROP5* paralogs), Type I (containing a virulent cluster), or Type III (containing a virulent cluster) was digested with XhoI, an enzyme that cuts between the individual genes in the locus (Fig. 2). The resulting DNA was separated by electrophoresis, and DNA of the appropriate size (~4.5 Kb) was gel-purified and cloned into XhoI-digested vector. Bacterial colonies were screened by colony hybridization using a radiolabeled *ROP5* probe.

Sequencing these individual *ROP5* genes confirmed that the *ROP5* locus of each strain contains multiple isoforms; a total of five, four, and four unique sequences were obtained from Types I, II, and III, respectively. This analysis also confirmed that there seems to be only three major isoform types in each strain; additional (the fourth or fifth) sequences had only one to two SNPs compared with their closest related sequences. Further analysis (Fig. 3) showed that the coding sequences from the Type I and III strains are nearly identical between the two strains; Type II sequences, however, were distinct from the other two strains, which were predicted by previous genomic studies (32).

Interestingly, the isoforms found in the strains that have the ROP5 allele associated with high virulence (Types I and III) are further diverged from each other than are individual isoforms in the Type II strain. Also, the Type II *ROP5B* isoform has a frameshift mutation in the center of the domain encoding the pseudokinase, which would result in a truncated and likely, nonfunctional protein (344 residues versus the usual 549 residues) (Fig. S6). Because the Type II *ROP5C* coding sequence is full length and otherwise very similar to $ROP5B_{II}$ (eight nonsynonymous polymorphisms and all just N terminal to the site of the frame shift), it is unlikely that this frame shift is the primary cause of the reduced virulence phenotype associated with the Type II locus. The impact of this frame shift on the overall number of functional ROP5 isoforms seems minimal, because the sequence coding for this frame shift is found in only 20% of the total raw genomic sequence reads for ROP5, indicating that 2 of 10 predicted copies of ROP5 in the Type II strain are of this truncated isoform. However, the fact that the Type II locus has the greatest number of copies among the strains examined may be a reflection of a compensation for these apparently defective versions.

We next examined the ratio of nonsynonymous and synonymous SNPs (K_A/K_S) across each codon in the ROP5 sequences (Fig. 3*B*). We observed that there is a remarkable propensity for nonsynonymous polymorphisms to cluster in the C-terminal half

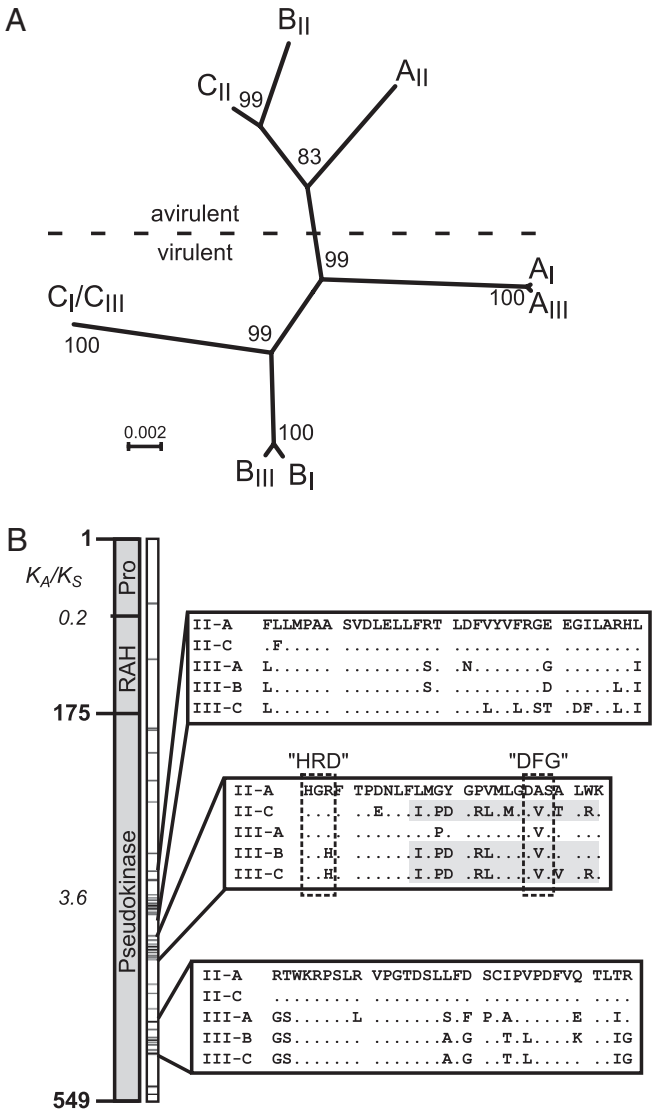

Fig. 3. The ROP5 isoforms are highly divergent in their pseudokinase domains. (*A*) A neighbor-joining tree of the coding sequences of the major isoforms of the virulent (Types I and III) and avirulent alleles (Type II) of ROP5. Bootstrap values are indicated at branch points. (*B*) The ratio of nonsynonomous to synonomous SNPs (K_A/K_S) of the various ROP5 isoforms within the three major domains of the protein is indicated in italicized numbers. The K_A/K_S for the N-terminal domains represents an average of all pair-wise comparisons ($P = 0.03$), although the value for the pseudokinase domain represents the highest significance pair-wise comparison (that of $ROP5A_I$ and $ROP5A_{II}$; $P = 0.002$). Amino acid numbering is bold. Codons with nonsynonomous SNPs are marked with a black line in the graph to the right of the domain architecture. Amino acid sequences are shown for highly polymorphic regions; identity is noted with a dot. The HRD and DFG motifs required for kinase activity (and unconserved in ROP5) are indicated in the sequence alignments in the *Insets*. The sequence near these motifs is boxed and gray and contains polymorphisms that are conserved between either A or B/C isoforms, regardless of strain.

of the pseudokinase domain, whereas very few nonsynonymous polymorphisms are found within either of the N-terminal domains (i.e., the pro and RAH domains). In fact, the pseudokinase domain seems to be under strong diversifying selection ($K_A/K_S = 3.6$), whereas the N-terminal domains may be under purifying selection ($K_A/K_S = 0.2$). The N-terminal prodomain has been shown to target homologous proteins to the correct secretory compartment within the parasite (33), and the RAH

MICROBIOLOGY

domain drives ROP5's association with the PVM when within the host cell (22). Overall, these data indicate that selection is acting differently on the different domains of these proteins. Importantly, although the former ROP5 active site represents a hot spot of divergence, none of the polymorphisms would be predicted to restore (or leave intact) a functional kinase activity. Thus, ROP5's pseudokinase status seems to be conserved among the paralogs from all three lineages.

Single ROP5 Paralog Restores Parasite Lethality. It is possible that all paralogs encoded in a multicopy locus such as *ROP5* must be expressed together for correct function in vivo. To determine what combinations of the *ROP5* isoforms are sufficient to complement and rescue the RH*Δrop5* virulence phenotype, we modified the RH*Δrop5* strain to express either single isoforms or pair-wise combinations of isoforms from the high-virulence cluster in Type III. Because the RH*Δku80* background of the parental strain severely reduces the efficiency of random integration (30), we chose to complement the RH*Δrop5* strain by targeting the two genes for which there are viable negative selection strategies for *Toxoplasma*. Thus, the *ROP5* isoforms were targeted either to the *Toxoplasma* uracil-phosphoribosyltransferase (*UPRT*) locus (34) or the *HXGPRT* cassette that we had originally inserted as part of the ROP5 knockout construct in RH*Δrop5* (constructs described in Fig. S2). Expression and localization were verified by including a short sequence encoding a C-terminal HA or FLAG tag in the introduced copies (Fig. S7), which have been shown to allow normal trafficking of ROP5 protein to the rhoptries and PVM (22, 23).

Virulence of the complemented strains was assayed by mouse survival as above. Importantly, a single copy of $ROP5A_{III}$, targeted to the *UPRT* locus, was able to partially rescue the virulence phenotype, because a dose of 10^6 parasites could no longer be tolerated and was, in fact, now uniformly lethal (Fig. 4*A*). These results show that the loss of virulence in the RH*Δrop5* is, in fact, because of loss of ROP5. They further reveal that the *ROP5* locus is not a single functional unit; an individual isoform can significantly alter virulence in the absence of the other copies.

Isoforms Differentially Affect Virulence. Given the expansion of the *ROP5* locus, we hypothesized that complementation with additional copies of ROP5 would lead to a more substantial recovery of the virulence phenotype. To test this, we first targeted an additional copy of $ROP5A_{III}$ to the *HXGPRT* cassette. This resulted in a sharp increase in virulence, because a dose of 10^3 parasites was now lethal; 80% of the mice infected at this dose now succumbed (Fig. 4*B*).

To determine if there are differences between ROP5 isoforms, these experiments were repeated using complementation with a combination of $ROP5A_{III}$ and $ROP5B_{III}$. Remarkably, a strain expressing a single copy of both seemed to have its virulence largely, if not completely, rescued; infection of mice with 10^2 parasites of such a strain was uniformly lethal by 9 d postinfection, a curve that is indistinguishable from the parental infection (Fig. 4*C*). Thus, expression of only two of three different isoforms from the virulent Type III locus is sufficient to rescue most, if not all, of the virulence phenotype. Taken together, these data show that different isoforms differ in the magnitude of their effect on virulence in this model. Although the extensive polymorphisms in the individual ROP5 isoforms play a large role in determining the virulence effects of the locus, given the ROP5 copy number variation among the strains, it is possible that expression level may also contribute to observed differences in virulence.

ROP5 Acts Independently from ROP18 and ROP16. Before this work, the effectors that had the greatest known roles in virulence were *ROP16* and *ROP18* (which had previously been shown to be the key genes at the *VIR3* and *VIR4* loci, respectively) (9, 10). Both the *ROP16* and *ROP18* loci consist of a single copy gene encoding a catalytically active kinase. Allelic differences in these genes mediate either a four- (*ROP18*) or one-log (*ROP16*) difference in virulence in mice (9, 10). To assess whether these two active kinase virulence factors might interact with ROP5, we analyzed the five significant previously identified QTLs for epistatic interactions. We found no significant interaction between *ROP5* and *ROP16*, but we did find high statistical support ($P = 6 \times 10^{-4}$) for genetic interactions between *ROP5* and *ROP18*. However, this interaction effect accounted for only 4% of the variation in virulence observed in the F1 progeny.

To further explore the relationship between the *ROP5* and *ROP18* loci, the same strategy used to create the RH*Δrop5* mutant was used to generate an RH*Δrop18* line, and this was tested for virulence in mice. If these two genes were members of the same pathway, we would expect the RH*Δrop18* to phenocopy the RH*Δrop5* strain. Surprisingly, deletion of *ROP18* had a much less dramatic effect on virulence than the *ROP5* deletion; one-half of mice infected with 100 parasites of RH*Δrop18* survived ~5 d longer than those infected with WT, but infection with the RH*Δrop18* still resulted in 100% mortality (Fig. S8). Taken together, these data suggest that ROP5 acts independently from both of the active kinases ROP16 and ROP18 and that any genetic interactions are likely indirect (i.e., they do not represent physical or biochemical interactions between the molecules). Further supporting this hypothesis are the data for the cosmid-complemented S22 strain (Fig. S1). As discussed above, S22 has the avirulent alleles for both *ROP16* and *ROP18*. Moreover, the

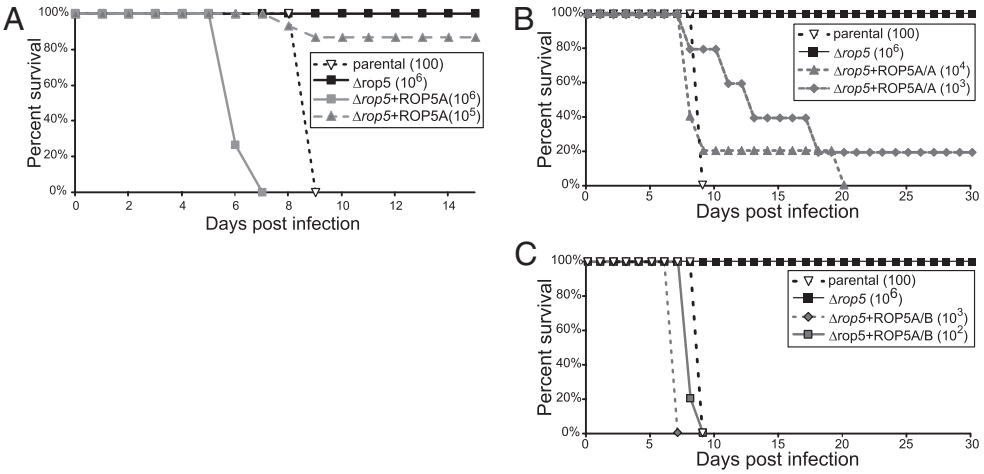

Fig. 4. ROP5 isoforms complement the virulence phenotype of Δ*rop5*. Survival analysis of BALB/c mice infected with RHΔ*rop5* complemented with (*A*) a single copy of $ROP5A_{III}$ (n = 15; per dose), (*B*) dual copies of $ROP5A_{III}$ (n = 5; per dose), or (*C*) one copy each of $ROP5A_{III}$ and $ROP5B_{III}$ (n = 5; per dose).

avirulent allele of *ROP18* contains an insertion disrupting its promoter, resulting in undetectable expression (9, 10). Thus, the fact that differences in the *ROP5* allele can alter virulence by five logs in such a background shows that ROP5 acts independently from ROP18 in its direct effects on virulence.

In summary, despite lacking a catalytically active kinase domain, ROP5 seems to play a larger role than ROP18 in an in vivo infection, at least within the context of a Type I strain. This is consistent with the fact that only when the *ROP5* locus is of the same (e.g., high) virulence type, as is the case for the Type I and III strains, does *ROP18* become the predominant genetic determinant of virulence in F1 progeny (10). It is important to note that the Type III strain possesses the avirulent alleles of both *ROP16* and *ROP18* (9), which helps explain its low virulence. However, it is clear that many factors, both genetic and possibly, epigenetic, contribute to a complex phenotype such as virulence, and the identification of the major *VIR* loci will assist in determining these other factors.

Discussion

We have identified a cluster of polymorphic secreted pseudokinases as being of paramount importance to virulence in *Toxoplasma*. Recently, pseudokinases have emerged as key regulatory effectors in metazoan signal transduction (35), particularly in mammalian immune signaling (36, 37). Although it is difficult to determine the specific role of a catalytically inactive protein, other pseudokinases have been shown to act as molecular scaffolds (38–40) and thereby, coordinate signal transduction. The presence of hotspots of polymorphism within the ROP5 pseudokinase domains indicates evolutionary pressure to bind divergent partners. This strongly suggests that differences in virulence are mediated by differential interactions of ROP5 within the host cell. Because the ROP5 proteins are present on the cytosolic face of the PVM, they may interact with and thus, dysregulate host proteins that are the key to the immune response. Alternatively, they could partner with other parasite proteins that are injected into the host cells to affect host signaling. Intriguingly, over 30% of *Toxoplasma*'s ~160 proteins kinases are predicted to be catalytically inactive pseudokinases (41), and over one-half of these are predicted to be secreted into the host cell. Thus, ROP5 may occupy a key, central role in a large and complex network of highly unusual effectors.

We have shown that, unique among other known virulence genes in *Toxoplasma*, the *ROP5* locus has been considerably expanded, such that there are multiple tandem copies of *ROP5* in each of the predominant strains and each of these strains possesses a different copy number and different alleles. Remarkably, expression of just two of the isoforms (and only two of the six copies found in Type I) in the RH*Δrop5* strain effectively rescued the parental virulence phenotype. Copy number variation is a major source of genetic diversity in metazoa (42, 43), and the *ROP5* locus seems to represent a specialized case of this in a protozoan parasite. This expansion of the *ROP5* locus may allow *Toxoplasma* to sample a greater sequence space for these critical genes than it could otherwise, because there is likely to be minimal cost to a single copy of diverging *ROP5*. Given ROP5's importance in infection, this ability to acquire multiple flavors of a key effector might explain the extraordinary expansion of intermediate hosts that can be productively infected by *Toxoplasma*. This represents a dark reflection of the diversification of rapidly evolving host proteins used to respond to varied infections, such as the MHC (44) and tripartite motif (TRIM) (45) families.

Diversification of effector molecules is a hallmark of the pathogenic lifestyle. Although viruses use error-prone polymerases and bacteria use both sexual recombination and transformation, *Toxoplasma* has a clonal lifestyle; within a given clonal type, isolates from across Europe and North America are virtually identical (8, 46). This may, in part, be because of *Toxoplasma*'s ability to forgo its sexual cycle by horizontal transmission between successive intermediate hosts (through carnivorism and scavenging) and/or self-mating of these haploid organisms. It is possible, then, that diversification of key virulence factors like *ROP5* through duplication may be necessary to compensate for inefficient diversification through sexual recombination. This hypothesis finds direct support in the fact that, whereas the three ROP5 isoforms from Types I and III are nearly identical at the sequence level, these loci vary in copy number (four vs. six, respectively). Intriguingly, the virulent Type I/III version of the locus contains isoforms that are more divergent from each other than are the isoforms from the avirulent Type II allele. Because the Type II strain seems to be a P1 parent of the Type III strain and a P2 parent of the Type I strain (32), both Type I and III strains were sufficiently viable to rise to predominance in a genetic landscape presumably dominated by Type II. It is possible that the divergent isoforms present in the Type I/III virulent ROP5 alleles provide a selective advantage in certain host species compared with the Type II alleles. Because *Toxoplasma* strains may simultaneously coevolve with not one but many host organisms, the divergent ROP5 isoforms may represent a Swiss army knife method of evolution; rather than existing as a single, precisely evolved tool for a single situation, the multiple ROP5 isoforms may be a collection that together equips *Toxoplasma* to infect a variety of host species.

Materials and Methods

Southern Blots and Cloning of ROP5 Isoforms. For restriction digestion analysis of the *ROP5* locus, digested genomic DNA from each of the three predominant strains (Type I, RH; Type II, ME49; Type III, CTG) was separated by field inversion gel electrophoresis or standard electrophoresis, depurinated, and transferred to Immobilon-N (Millipore) membrane (47). The membranes were blocked in ExpressHyb (Clontech) according to the manufacturer's protocol and incubated with a radioactive probe created using the RadPrime (Invitrogen) kit with gel-purified sequence corresponding to the ROP5-RAH domain (22) as a template. For cloning *ROP5* isoforms, TOP10 bacteria (Invitrogen) were transformed with gel-purified XhoI-digested parasite genomic DNA that had been ligated into XhoI-cut pCR2.1 (Invitrogen). Clones containing ROP5 isoforms were confirmed by transferring bacterial colonies to membrane and probing as above.

Sequence Read Analysis and Alignments. Alignments of *ROP5* sequences were generated with CLUSTALW (48), and phylogenetic and K_A/K_S analysis was conducted in MEGA4 (49). FASTA versions of individual trace files from the Sanger-based sequencing projects for three major *T. gondii* strain types (GT-1, ME49, and VEG) were downloaded from the National Center for Biotechnology Information (NCBI) trace archive and used as query sequences in BLASTN searches against the draft 6 version of the ME49 genome (www.toxodb.org). An e-value cutoff of 1e-20 was used. The resulting blast file was parsed, and the number of sequence reads overlapping the ROP5 locus and surrounding sequence from ME49 (positions 537,000–545,000 on chromosome XII) in each 100-bp window was counted using a custom perl script and a series of mysql queries.

Plasmid Construction and PCR. PCR was carried out using Phusion DNA polymerase (NEB) unless otherwise noted. To create the Δrop5 construct, 5′ and 3′ targeting sequences were amplified from RH strain genomic DNA and cloned into pTKO such that they flank an *HXGPRT* expression cassette. Individual *ROP5* isoforms were subcloned from the constructs created above and either inserted into the Δrop5-pTKO construct, such that they replaced the *HXGPRT* cassette (and were in frame with a FLAG tag and *GRA2* 3′-UTR), or inserted into pUPRT-HA, such that they were in frame with an HA tag and *GRA2* 3′-UTR and flanked by targeting sequence for the *Toxoplasma UPRT* locus (34).

Parasite and Host Cell Culture. Human foreskin fibroblasts (HFF) were grown in DMEM supplemented with 10% FBS (Invitrogen) and 2 mM glutamine. *Toxoplasma* tachyzoites were maintained in confluent monolayers of HFF. The complemented S22 strain was made by transfecting S22(GFP/fluc) (28) with 50 μg unlinearized PSBLC37 cosmid (a gift from L.D. Sibley, Washington University, St. Louis, MO) and selecting with phleomycin (50). The RH*Δrop5* parasite strain was made by electroporating RH(*Δhxgprt Δku80*) parasites (30) with 15 μg linearized plasmid and selecting for HXGPRT-positive parasites as previously described (51). Complemented RH*Δrop5* strains were created by either targeting the *HXGPRT*-replaced *ROP5* locus and selecting with 6-thioxanthine (51) or choosing targeted replacement of the *UPRT* locus

MICROBIOLOGY

with the *ROP5* isoform of choice (34). Clonal parasites were grown from populations by limiting dilution.

Mouse Survival Analysis. Six- to eight-week-old female BALB/c (Jackson Labs) or CD-1 (Charles River) mice were infected by peritoneal injection with parasites diluted in PBS (pH 7.4) to the appropriate dose and monitored daily for survival. Infection of surviving mice was verified by testing serum for reactivity with *Toxoplasma* lysate.

QTL Interaction Analysis. Survival data for mice infected with Type II × Type III F1 progeny originally published in ref. 9 were reanalyzed for epistatic interactions using the rqtl module as implemented in R (51). For the five virulence QTLs identified, we used the fitqtl function to fit a model encompassing all five QTLs as well as interactions among each of the loci to generate *P* values for significant interactions and calculate the percent contribution of each of the interaction terms to the variation in virulence.

ACKNOWLEDGMENTS. We thank Boris Striepen and David Sibley for providing the cosmid library, Vern Carruthers for the RH(Δ*hxgprt*Δ*ku80*) strain, Peter Bradley for the anti-ROP5 antibody and advice on 6-thioxanthine selection, Anita Koshy for help with animal work, John Hammond and Paul Norman for help with field inversion gel electrophoresis Sonya Taylor and David Sibley for helpful discussions regarding bleomycin selection, and Anita Koshy, Eva LaDow, Kara Norman, Paul Norman, Selena Sagan., and R.Y. for helpful suggestions on the manuscript. This work was supported by an American Cancer Society grant (to M.L.R.), a National Research Service Award (to G.M.Z.), a California Universitywide AIDS Research Program grant (to J.P.S.), National Institutes of Health Grants AI73756 (to J.C.B.) and K22AI080977 (to J.P.B.), and a Pew Scholarship in the Biomedical Sciences (to J.P.B.).

1. Deitsch KW, Lukehart SA, Stringer JR (2009) Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol* 7:493–503.
2. Marques JT, Carthew RW (2007) A call to arms: Coevolution of animal viruses and host innate immune responses. *Trends Genet* 23:359–364.
3. Saeij JP, Boyle JP, Boothroyd JC (2005) Differences among the three major strains of Toxoplasma gondii and their specific interactions with the infected host. *Trends Parasitol* 21:476–481.
4. Boothroyd JC, Grigg ME (2002) Population biology of Toxoplasma gondii and its relevance to human infection: Do different strains cause different disease? *Curr Opin Microbiol* 5:438–442.
5. Carme B, Demar M, Ajzenberg D, Dardé ML (2009) Severe acquired toxoplasmosis caused by wild cycle of Toxoplasma gondii, French Guiana. *Emerg Infect Dis* 15: 656–658.
6. 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.
7. 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.
8. Grigg ME, Bonnefoy S, Hehl AB, Suzuki Y, Boothroyd JC (2001) Success and virulence in Toxoplasma as the result of sexual recombination between two distinct ancestries. *Science* 294:161–165.
9. Saeij JP, et al. (2006) Polymorphic secreted kinases are key virulence factors in toxoplasmosis. *Science* 314:1780–1783.
10. Taylor S, et al. (2006) A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen Toxoplasma gondii. *Science* 314:1776–1780.
11. Boothroyd JC, Dubremetz JF (2008) Kiss and spit: The dual roles of Toxoplasma rhoptries. *Nat Rev Microbiol* 6:79–88.
12. Bradley PJ, Sibley LD (2007) Rhoptries: An arsenal of secreted virulence factors. *Curr Opin Microbiol* 10:582–587.
13. Carruthers VB (1999) Armed and dangerous: Toxoplasma gondii uses an arsenal of secretory proteins to infect host cells. *Parasitol Int* 48:1–10.
14. Rosowski EE, et al. (2011) Strain-specific activation of the NF-kappaB pathway by GRA15, a novel Toxoplasma gondii dense granule protein. *J Exp Med* 208(1):195–212.
15. Blanchard N, et al. (2008) Immunodominant, protective response to the parasite Toxoplasma gondii requires antigen processing in the endoplasmic reticulum. *Nat Immunol* 9:937–944.
16. Saeij JP, et al. (2007) Toxoplasma co-opts host gene expression by injection of a polymorphic kinase homologue. *Nature* 445:324–327.
17. Fentress SJ, et al. (2010) Phosphorylation of immunity-related GTPases by a Toxoplasma gondii-secreted kinase promotes macrophage survival and virulence. *Cell Host Microbe* 8:484–495.
18. Steinfeldt T, et al. (2010) Phosphorylation of mouse immunity-related GTPase (IRG) resistance proteins is an evasion strategy for virulent Toxoplasma gondii. *PLoS Biol* 8: e1000576.
19. Ong YC, Reese ML, Boothroyd JC (2010) Toxoplasma rhoptry protein 16 (ROP16) subverts host function by direct tyrosine phosphorylation of STAT6. *J Biol Chem* 285: 28731–28740.
20. Yamamoto M, et al. (2009) A single polymorphic amino acid on Toxoplasma gondii kinase ROP16 determines the direct and strain-specific activation of Stat3. *J Exp Med* 206:2747–2760.
21. El Hajj H, et al. (2006) The ROP2 family of Toxoplasma gondii rhoptry proteins: Proteomic and genomic characterization and molecular modeling. *Proteomics* 6: 5773–5784.
22. Reese ML, Boothroyd JC (2009) A helical membrane-binding domain targets the Toxoplasma ROP2 family to the parasitophorous vacuole. *Traffic* 10:1458–1470.
23. El Hajj H, Lebrun M, Fourmaux MN, Vial H, Dubremetz JF (2007) Inverted topology of the Toxoplasma gondii ROP5 rhoptry protein provides new insights into the association of the ROP2 protein family with the parasitophorous vacuole membrane. *Cell Microbiol* 9:54–64.
24. Hanks SK, Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEB J* 9:576–596.
25. Koul A, Herget T, Klebl B, Ullrich A (2004) Interplay between mycobacteria and host signalling pathways. *Nat Rev Microbiol* 2:189–202.
26. Bos JI, et al. (2010) Phytophthora infestans effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proc Natl Acad Sci USA* 107:9909–9914.
27. Molden J, Chang Y, You Y, Moore PS, Goldsmith MA (1997) A Kaposi's sarcoma-associated herpesvirus-encoded cytokine homolog (vIL-6) activates signaling through the shared gp130 receptor subunit. *J Biol Chem* 272:19625–19631.
28. Saeij JP, Boyle JP, Grigg ME, Arrizabalaga G, Boothroyd JC (2005) Bioluminescence imaging of Toxoplasma gondii infection in living mice reveals dramatic differences between strains. *Infect Immun* 73:695–702.
29. Mordue DG, Monroy F, La Regina M, Dinarello CA, Sibley LD (2001) Acute toxoplasmosis leads to lethal overproduction of Th1 cytokines. *J Immunol* 167:4574–4584.
30. Huynh MH, Carruthers VB (2009) Tagging of endogenous genes in a Toxoplasma gondii strain lacking Ku80. *Eukaryot Cell* 8:530–539.
31. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J (2009) Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res* 19:1586–1592.
32. 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 USA* 103:10514–10519.
33. Bradley PJ, Li N, Boothroyd JC (2004) A GFP-based motif-trap reveals a novel mechanism of targeting for the Toxoplasma ROP4 protein. *Mol Biochem Parasitol* 137: 111–120.
34. Donald RG, Roos DS (1995) Insertional mutagenesis and marker rescue in a protozoan parasite: Cloning of the uracil phosphoribosyltransferase locus from Toxoplasma gondii. *Proc Natl Acad Sci USA* 92:5749–5753.
35. Boudeau J, Miranda-Saavedra D, Barton GJ, Alessi DR (2006) Emerging roles of pseudokinases. *Trends Cell Biol* 16:443–452.
36. Shaw MH, et al. (2003) A natural mutation in the Tyk2 pseudokinase domain underlies altered susceptibility of B10.Q/J mice to infection and autoimmunity. *Proc Natl Acad Sci USA* 100:11594–11599.
37. Flannery S, Bowie AG (2010) The interleukin-1 receptor-associated kinases: Critical regulators of innate immune signalling. *Biochem Pharmacol* 80:1981–1991.
38. Fukuda K, Gupta S, Chen K, Wu C, Qin J (2009) The pseudoactive site of ILK is essential for its binding to alpha-Parvin and localization to focal adhesions. *Mol Cell* 36: 819–830.
39. Reese ML, Dakoji S, Bredt DS, Dötsch V (2007) The guanylate kinase domain of the MAGUK PSD-95 binds dynamically to a conserved motif in MAP1a. *Nat Struct Mol Biol* 14:155–163.
40. Zeqiraj E, Filippi BM, Deak M, Alessi DR, van Aalten DM (2009) Structure of the LKB1-STRAD-MO25 complex reveals an allosteric mechanism of kinase activation. *Science* 326:1707–1711.
41. Peixoto L, et al. (2010) Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses. *Cell Host Microbe* 8:208–218.
42. Bailey JA, Eichler EE (2006) Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nat Rev Genet* 7:552–564.
43. Zhang F, Carvalho CM, Lupski JR (2009) Complex human chromosomal and genomic rearrangements. *Trends Genet* 25:298–307.
44. Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA* 94:7799–7806.
45. Sawyer SL, Emerman M, Malik HS (2007) Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. *PLoS Pathog* 3:e197.
46. Su C, et al. (2003) Recent expansion of Toxoplasma through enhanced oral transmission. *Science* 299:414–416.
47. Sambrook J, Russell D (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Plainview, NY)3rd Ed.
48. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
49. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
50. Messina M, Niesman I, Mercier C, Sibley LD (1995) Stable DNA transformation of Toxoplasma gondii using phleomycin selection. *Gene* 165:213–217.
51. Donald RG, Carter D, Ullman B, Roos DS (1996) Insertional tagging, cloning, and expression of the Toxoplasma gondii hypoxanthine-xanthine-guanine phosphoribosyltransferase gene. Use as a selectable marker for stable transformation. *J Biol Chem* 271:14010–14019.
52. Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889–890.

Supporting Information

Reese et al. 10.1073/pnas.1015980108

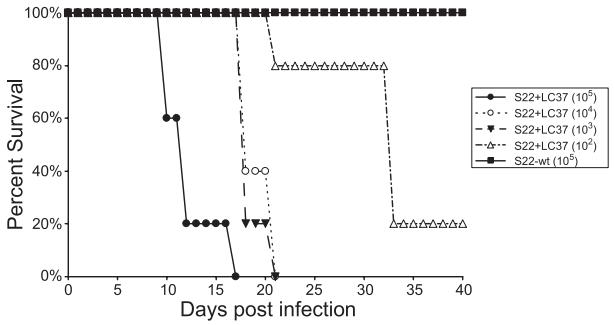

Fig. S1. Complementation of an avirulent *Toxoplasma gondii* strain (S22) with a cosmid containing the virulent allelic locus of ROP5 dramatically increases virulence in mice. BALB/c mice (n = 5) were infected with 10^5, 10^4, 10^3, and 10^2 parasites of either a WT S22 strain or S22 complemented with cosmid PSBLC37 (S22 + LC37) encoding the virulent type I allele for the *ROP5* locus and were monitored for 40 d.

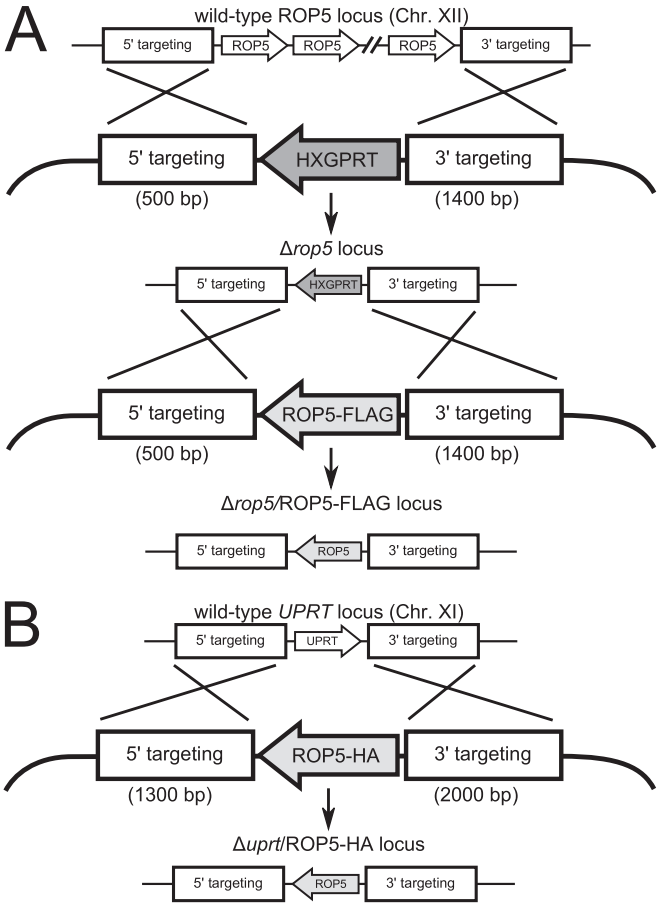

Fig. S2. Strategies for generation of ROP5 knockout and complemented strains. (*A*) The *HXGPRT* cassette originally used to select the Δ*rop5* was targeted using identical sequences to the knockout construct containing a copy of an *ROP5* isoform driven by its endogenous promoter and using a 3′-UTR from the *Toxoplasma GRA2* gene. The *ROP5* isoform is in frame with a sequence encoding a C-terminal 3xFLAG tag and inverted in its orientation relative to the original locus. (*B*) Using a sequence targeting the *Toxoplasma* uracil-phosphoribosyltransferase (*UPRT*) locus, the *UPRT* coding sequence is replaced with an *ROP5* isoform similar to as in *A*, although fused to a C-terminal HA tag.

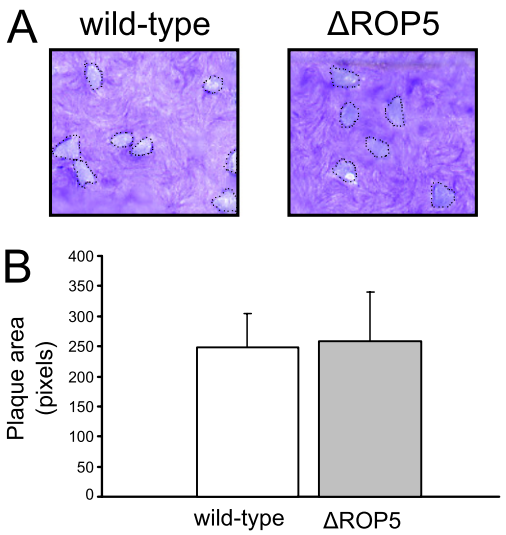

Fig. S3. RHΔ*rop5* parasites showed no growth difference in plaque assays. Equal doses of WT and RHΔ*rop5* parasites were inoculated onto HFF monolayers and left undisturbed for 7 d. (*A*) Images of representative plates stained with crystal violet. (*B*) Plaque size was quantified in ImageJ by measuring the surface area of plaques. Average sizes are shown (± SD) for 40 plaques, representing ~50% of a flask.

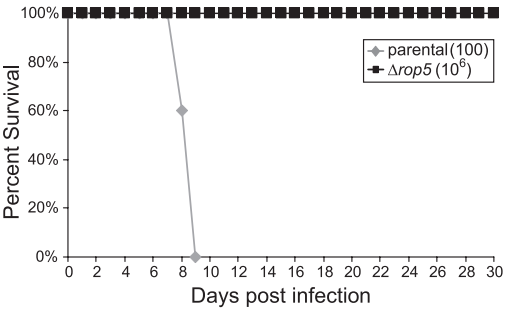

Fig. S4. RHΔ*rop5* parasites are avirulent to outbred CD-1 mice. CD-1 mice (*n* = 10; each strain) were infected intraperitoneally with either 100 parental RHΔ*ku80*Δ*hxgprt* or 10^6 RHΔ*rop5* parasites and were monitored for 30 d.

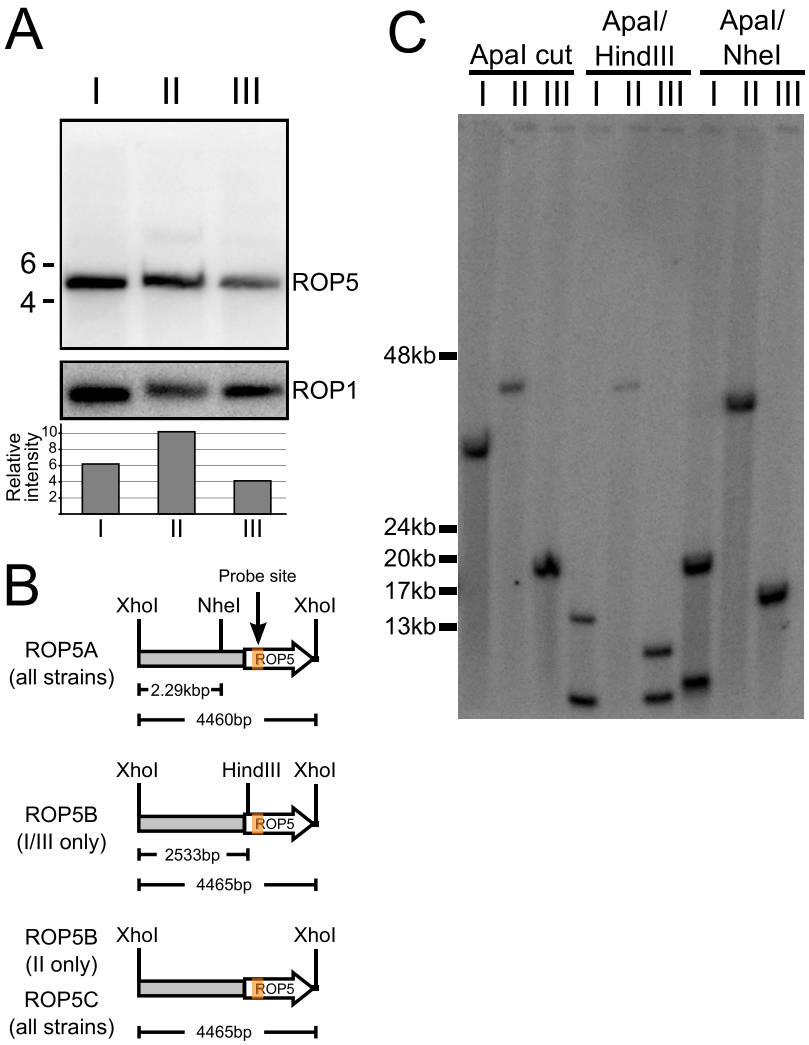

Fig. S5. (*A*) Southern blot of genomic parasite DNA from each of the three major strain types (I, II, and III) as in Fig. 1*B*, but digestion was with XhoI. For normalization, the same blot was stripped of the *ROP5* probe and rehybridized with a probe corresponding to the single-copy *ROP1* gene. Based on the sequence reads and size of the ApaI fragment, the number of copies of *ROP5* in the Type III strain was set at four; the number of copies in the other two strains was then determined by normalizing to the ROP1 intensity. (*B*) Diagram of restriction sites in various ROP5 isoforms. The binding site for the Southern blot probe is shaded in orange. (*C*) Southern blot of diagnostic double digests of genomic parasite DNA from each of the three major strain types as above. DNA was digested with either ApaI, which cuts outside of the locus, or a combination with either HindIII, which cuts only $ROP5B_{III}$ 2.5 kbp from the upstream XhoI site, or NheI, which cuts only the ROP5A isoform (both in Types II and III) 2.3 kbp from the upstream XhoI site. The probe for Southern blot anneals downstream of both the HindIII and NheI sites in the repeat.

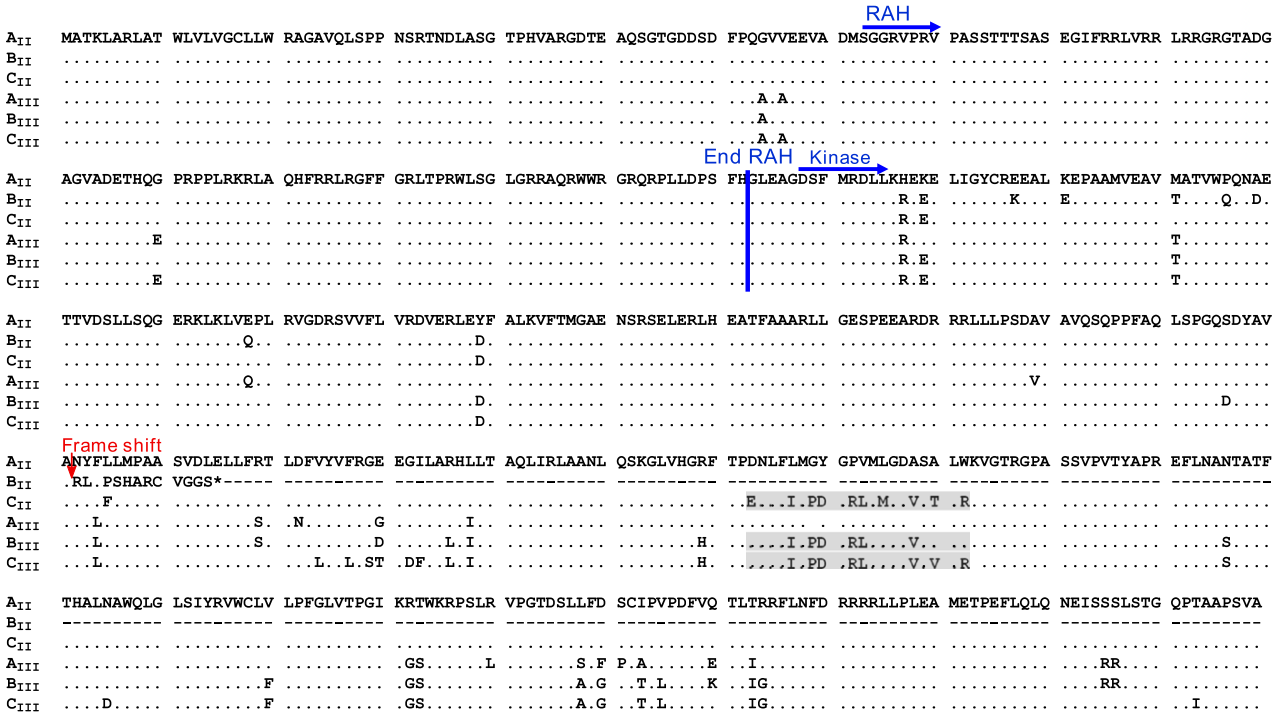

Fig. S6. Alignment of the predicted protein sequence of each of the three ROP5 isoform types from the avirulent Type II and virulent Type III loci. Identity is noted with a dot. The sequence near the former catalytic loop is boxed and gray, and it contains polymorphisms that are conserved between either A or B/C isoforms, regardless of strain. Also note the frame shift in *ROP5B*$_{II}$ (red arrow) that causes a premature stop halfway through the pseudokinase domain.

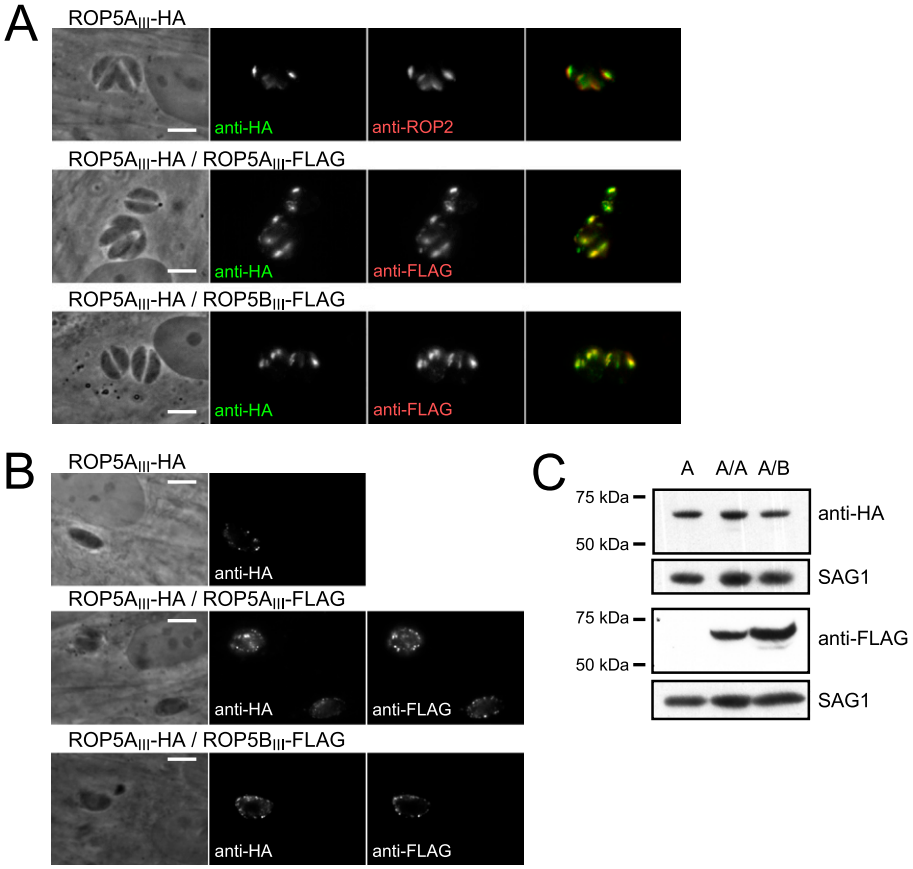

Fig. S7. (*A*) Rhoptry localization was confirmed by staining parasites permeabilized with 0.2% Triton-X100. Localization of ROP5 isoforms in the complemented strains was verified first by probing with anti-HA for the rhoptry marker ROP2. Because this ROP5A$_{III}$-HA is the parental strain for ROP5A$_{III}$-HA/5A$_{III}$-FLAG and ROP5A$_{III}$-HA/5B$_{III}$-FLAG, the correctly localized ROP5A$_{III}$-HA was used to confirm the localization of the FLAG-tagged proteins. (*B*) The characteristic punctate parasitophorous vacuolar membrane (PVM) localization of the tagged ROP5 proteins was confirmed by staining parasites that had been fixed 4 hours post infection (hpi) and the host cells permeabilized with 0.01% saponin. (*C*) Expression levels of the HA- and FLAG-tagged proteins were analyzed by Western blot, with SAG1 used as a loading control. Strains are designated as A (ROP5A$_{III}$-HA), A/A (ROP5A$_{III}$-HA/5A$_{III}$-FLAG), and A/B (ROP5A$_{III}$-HA/5B$_{III}$-FLAG). Although ROP5A$_{III}$-FLAG in the A/A strain and ROP5B$_{III}$-FLAG in the A/B strain seem to be expressed at different levels, we saw this difference in multiple independent nonsibling clones as well as in ectopically expressed proteins regardless of the parental genetic background, suggesting that any difference in expression is inherent in differences in the two genes themselves.

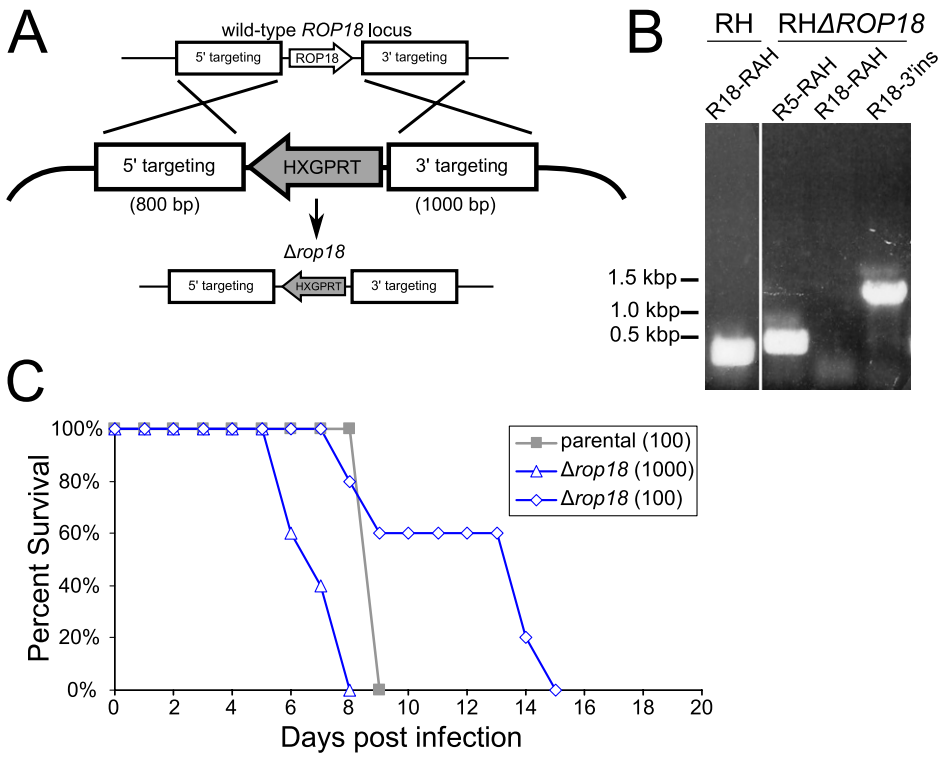

Fig. S8. (*A*) The *ROP18* coding sequence of a virulent Type I (RHΔ*ku80*Δ*hxgprt*) strain was replaced with the *HXGPRT* selection marker by homologous recombination. (*B*) Although sequence encoding the ROP18-RAH domain can be amplified from the parental strain, it cannot be amplified from the RHΔ*rop18* parasites. In addition, insertion of the HXGPRT cassette into the ROP18 locus was verified with primers lying downstream from the 3′ targeting sequence and within the *HXGPRT* cassette. (C) RHΔ*rop18* parasites maintain high virulence to mice. BALB/c ($n = 5$; each dose) were infected intraperitoneally with the noted dose of either parental RHΔ*ku80*Δ*hxgprt* or RHΔ*rop18* parasites.