Supporting Information

Behnke et al. 10.1073/pnas.1015338108

SI Materials and Methods

Marker Selection and Single-Feature Polymorphism Genotyping by Microarray Hybridization. The *Toxoplasma gondii* array includes 89,122 probe sequences located in the 3′ regions of protein-coding genes (8,102 sets of 11 probes per gene) designed for RNA expression studies and based on the genome sequence of the type II ME49 strain (1). Fortuitously, 10,090 of these probes, representing 4,159 unique genes, have single-feature polymorphism (SFP) mismatches (SNP or indel) between the available genomic DNA (gDNA) sequences of type I strain GT-1 and type II strain ME49 (http://toxodb.org/). The following automated protocol was developed to identify microarray probes that could serve as reliable genetic markers by consistently giving differential gDNA hybridization with type I and II genotypes across all progeny and the parents.

i) Data normalization and probe set choice. The raw intensity data, including the 89,122 3′-biased expression probes on the array, were median-scaled and robust multiarray average (RMA)-normalized across all 53 gDNA hybridizations (49 progeny and duplicates from each parent) followed by median polishing in log(2) units on a per chip basis. After normalizing the full datasets in this way, we then excluded from further analysis those probe sets identified during microarray design as likely sources of unreliable hybridization: namely, the Affymetrix probe set names flagged with _x_at or _s_at suffixes, which denote that some probes have potential intergene cross-hybridization or suboptimal hybridization sequences, respectively.

ii) Identification and typing of candidate SFP markers. The 53 normalized values [in log(2) units] for each probe were forced into two classes by *k*-means classification [S-plus function kmeans] using as seeds the 25th and 75th percentiles of the 53 values. Candidate SFP marker probes were first required to be concordant for the four parental hybridizations [i.e., the *k*-means classification allocated values to the perfect match class for the duplicate ME49-5-fluoro-2′-deoxyuridine ($FUDR^R$) gDNAs and the mismatch class for the two GT-1-SYN^R gDNAs]. These candidates were then further filtered by a variance criterion that the within-cluster sums of squares for both the perfect match and mismatch classes were less than 13,500 percentile units. This threshold effectively excluded many probes with assignment uncertainties caused by broad or overlapping intensity distributions across the 53 hybridizations or intensity values outside the linear dynamic range or with extreme outlying values.

iii) Consistency with genomic sequence data. To further reduce false positives, we then removed probes that did not show SNP or indel type I/II mismatches in the available genomic DNA sequences (http://toxodb.org/), although this did exclude some apparently good SFP markers that lacked GT-1 genomic sequence coverage.

iv) Chromosome neighborhood consistency criteria. At this stage, the large majority of the candidate probes were observed to provide genotype assignments that were highly consistent along each chromosome for each progeny clone, with provisional cross-over locations inferred from breaks in continuity [e.g., 22222211111 (where 2 and 1 are the respective haplotypes)]. The principal exceptions to this genomic neighborhood consistency were restricted to 4 of the initial 49 progeny gDNA hybridizations that showed a high noise level. Accordingly, these four progeny were not used for further analysis. We then applied the following formal genome neighborhood consistency criterion to the 45 remaining progeny. The entire data for a probe were excluded if 2 or more of its 45 automatically assigned type I (mismatch) or II (perfect match) genotypes were discordant with their flanking chromosomal neighborhood genotypes inferred from at least 5 closely linked probes per clone (counting unidirectionally away from any inferred cross-over points or chromosome ends). This stringent criterion removed from consideration as SFP markers ∼7% of the remaining probes. A total of 1,603 reliable SFP markers were, thus, identified for genotyping studies.

Cloning of Rhoptry Protein 5 Major Alleles and Production of Polyclonal Antisera. gDNA from GT-1 and ME49 strains were used to amplify rhoptry protein 5 (ROP5) coding sequence (CDS) with iProof polymerase (Bio-Rad) (Dataset S1 and Table S2) and were cloned into the His-tagged pET22b vector (Novagen). The major allele as defined by the trace archive read analyses for each strain was identified and confirmed from individual clones by sequencing. Recombinant proteins were expressed in BL21-CodonPlus(DE3)-RP cells (Novagen) induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Protein was purified with His Select Nickel Affinity Gel (Sigma-Aldrich) and dialyzed with buffer (250 mM NaCl, 10 mM $MgCl_2$, 20 mM Tris). Rabbit polyclonal antisera were produced commercially using Covance Custom Immunology Services (Covance).

Complementation of RHΔ*ku80*Δ*rop5* with Cosmid TOXOM52. Previous efforts have sequenced the ends of randomly selected clones of two cosmid libraries from the arginine histidine (RH) strain and mapped them to the *T. gondii* genome (http://toxodb.org/). Inspection of the *ROP5* locus revealed several cosmid clones that span this region (http://toxodb.org/). Cosmid clone TOXOM52 (http://toxomap.wustl.edu/cosmid.html) was chosen to complement RHΔ*ku80*Δ*rop5* parasite, because it completely spans the *ROP5* locus and contains six copies of *ROP5*, which was confirmed by restriction endonuclease digestion and mapping. This cosmid partially overlaps the CDS of the gene immediately upstream of *ROP5* (hypothetical protein TGGT1_042730), although it lacks the ATG and first exon for this gene. Cosmid DNA was prepared as described at http://toxomap.wustl.edu/cosmid.html using NucleoBond Xtra Maxi Plus columns to purify (Clontech). A total of 10 μg cosmid was electroporated into RHΔ*ku80*Δ*rop5* parasites, parasites were allowed to recover in human foreskin fibroblast (HFF) cells for 2 d, cells were then selected with 5 μM pyrimethamine for two passages, resistant parasites were cloned, and positive clones were identified by immunofluorescence using rabbit anti-ROP5 antibody followed by secondary goat anti-mouse IgG conjugated to Alexa 488. A representative clone referred to as RHCmpl34, which expressed ROP5 at WT levels as shown by Western blot (Fig. S8), was chosen for analysis.

qRT-PCR Analysis. Total RNA was isolated from parasites with the RNeasy Kit using DNase treatment (Qiagen). cDNA was prepared, and qRT-PCR was performed as described previously (2), except 2 μg starting RNA were used in cDNA synthesis and 0.5 μL cDNA were used in the qRT-PCR. Fold expression was calculated as described previously (2), except that the reference sample was RHΔ*ku80* and Actin (TGME49_009030) was used as the control gene (Dataset S1 and Table S2). Two separate RNA

preparations were used, with duplicate samples to determine the Ct values and fold change shown as mean ± SD.

SI Results

Genome-Wide Association Mapping. QTL association mapping of progeny and parent trait value vectors with I/II SFP genotype vectors was carried out using a range of conventional statistical methods (*t* test statistic, Wilcoxon nonparametric, Fisher's exact probability, and mutual information). Log(1/P) values were plotted along the genome-wide physical coordinates of the 1,603 SFP markers, with marginal and strong genome-wide significance thresholds determined from 1,000 permutations (dotted lines on plots). Phenotype vectors were defined alternatively as the continuous-valued average survival percentages or as three-state (with one intermediate class) or four-state (two intermediate classes: high > 45% and low < 45%) categorical-valued traits. These variations of association statistic or trait definition all gave essentially identical assignments of the single major QTL on chromosome XII.

Assessment of Gene Expression in RHΔ*ku80*Δ*rop5* and Complemented Clone. Because the *ROP5* locus is large and not fully represented in the assembled genome, we tested the expression of *ROP5* and the adjacent gene TGGT1_042730 in the background used for genetic deletion (RHΔku80) as well as the knockout and complemented clones. TGGT1_042730 is expressed at low levels in all strains for which there is publically available microarray expression data (http://toxodb.org/), and it lies outside the region used for generating the knockout. However, the cosmid clone used for complementation also contains a portion of this gene. Therefore, we used qRT-PCR to determine the transcript abundance of TGGT1_042730 in RHCmpl34 and RHΔ*ku80*Δ*rop5* parasites compared with RHΔ*ku80*. The level of TGGT1_042730 expression was only modestly affected in RHCmpl34 parasites compared with the RHΔ*ku80* line (i.e., 2.28 ± 0.508-fold vs. RHΔ*ku80*), a result that was similar to the level of expression in-RHΔ*ku80*Δ*rop5* (i.e., 1.08 ± 0.948-fold vs. RHΔ*ku80*). In addition, to validate the complementation of RHCmpl34, we determined the transcript levels of *ROP5* in RHΔ*ku80*, RHΔ*ku80*Δ*rop5*, and RHCmpl34 by qRT-PCR. Expression of *ROP5* in the RHΔ*ku80*Δ*rop5* parasite was ~185-fold lower than RHΔ*ku80* (i.e., 0.0054 ± 0.0061-fold vs. RHΔ*ku80*), whereas expression was restored in the RHCmpl34 parasite (i.e., 2.09 ± 0.818-fold vs. RHΔ*ku80*). These results indicate that minor variations in the expression of this adjacent gene are unlikely to be responsible for the phenotypes reported here, which instead, rely on the presence or absence of the *ROP5* locus and corresponding expression of ROP5.

1. Bahl A, et al. (2010) A novel multifunctional oligonucleotide microarray for *Toxoplasma gondii*. *BMC Genomics* 11:603.
2. Khan A, Behnke MS, Dunay IR, White MW, Sibley LD (2009) Phenotypic and gene expression changes among clonal type I strains of Toxoplasma gondii. *Eukaryot Cell* 8: 1828–1836.

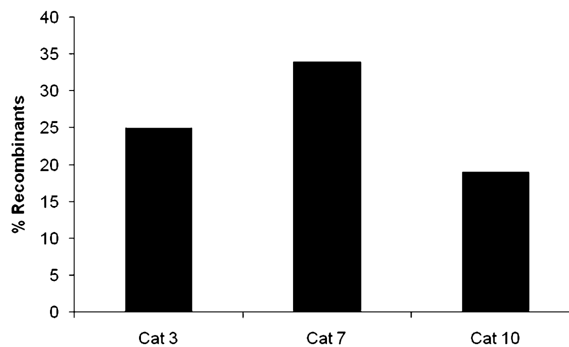

Fig. S1. Recombination frequency of unselected progeny. Unselected progeny clones isolated from the three cats used for the cross were genotyped using 10 restriction fragment-length polymorphism (RFLP) markers (Dataset S1 and Table S1) to determine the frequency of recombinant progeny in each pool (n = 32 for Cat 3 and Cat 10 and n = 29 for Cat 7). The frequency of recombinant progeny matches the expected value based on previous estimates of selfing vs. outcrossing (1).

1. Pfefferkorn LC, Pfefferkorn ER (1980) *Toxoplasma gondii*: genetic recombination between drug resistant mutants. *Exp Parasitol* 50:305–316.

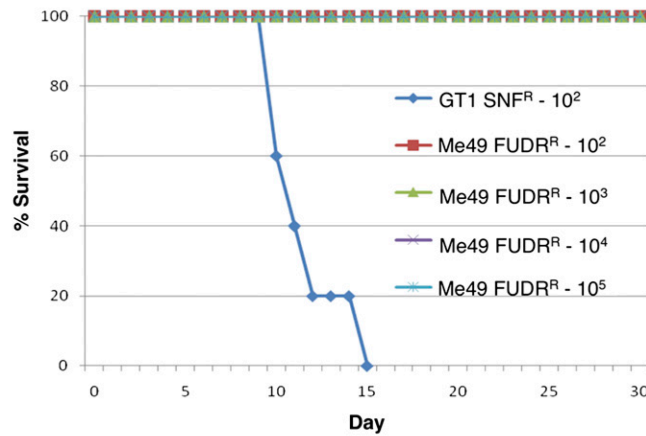

Fig. S2. Survival of outbred mice challenged with *T. gondii* strains. Female CD-1 outbred mice were challenged by i.p. inoculation with different doses of either GT1-SNF^R or ME49 FUDR^R parental strains, as described previously (1). Survival was monitored for 30 d postinfection.

1. Taylor S, et al. (2006) A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. *Science* 314:1776–1780.

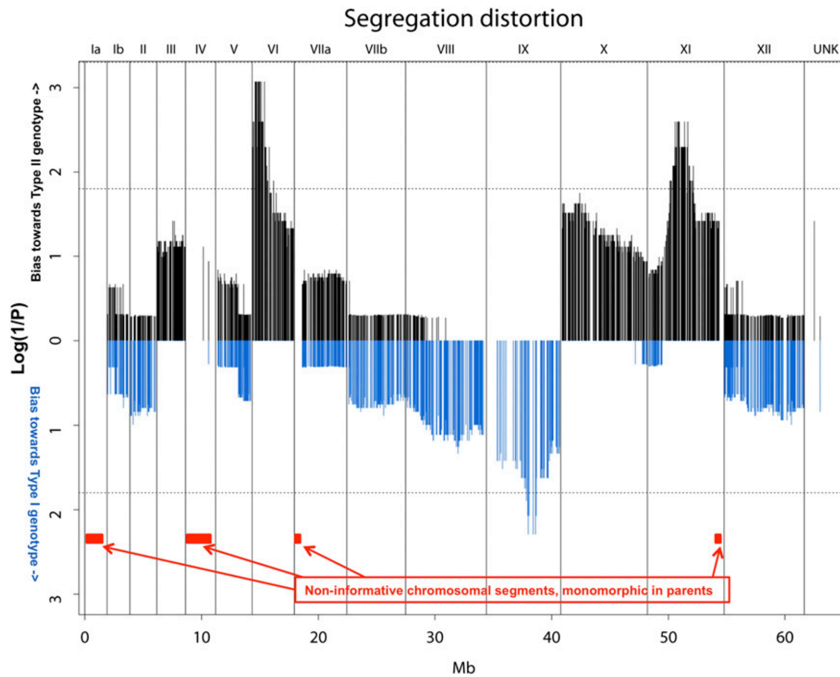

Fig. S3. Analysis of marker segregation in the I × II cross. The physical locations of the 1,603 SFP markers are plotted along the 14 nuclear chromosomes, indicated by the vertical black or blue bars on the main plot. These markers provide high-density coverage of 92% of the *T. gondii* nuclear genome. The red bars below the plot show the chromosomal regions, totaling 8% of the genome, that are monomorphic in the parents and lack informative markers, namely most of chromosomes Ia and IV and short segments of chromosomes VIIa and XI. The large majority of markers show informative balanced segregation in the I × II cross, matching the 1:1 ratio expected for haploid inheritance. The exceptions are three loci that show strong segregation distortion (i.e., upward black bars or downward blue bars exceeding the genome-wide statistical significance threshold shown by the dotted horizontal lines). The loci detected on chromosomes VI and XI show strong bias against type I progeny; the bias on XI may reflect the selection of progeny clones with FUDR. Similarly, the chromosome IX locus showing strong bias against type II progeny may reflect selection with SNF. These segregation distortions were not associated with any obvious growth phenotypes in vitro, and they did not contribute substantially to the mouse infectivity phenotypes studied. Notably, this analysis emphasizes that this cross is potentially fully informative and sensitive for the genetic mapping of any virulence-related quantitative trait locus (QTL) that might occur in more than 85% of the genome, including the known loci encompassing ROP18, ROP16, and ROP5 on chromosomes VIIa, VIIb, and XII, respectively.

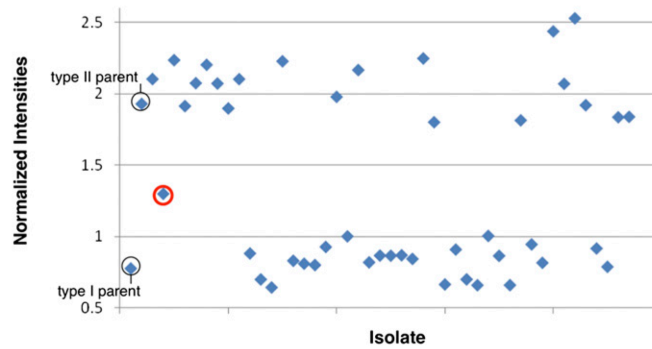

Fig. S4. Inheritance of *ROP5* copy number in the I × II progeny. Normalized genomic hybridization values for the *ROP5* probe set (i.e., 551.m00238 on the Affymetrix array) were graphed for the parents and progeny of the cross (*y* axis indicates fold expression for each clone vs. the median across all clones). Comparison of the hybridization values indicates that type II is roughly two times as high as type I, reflecting the roughly twofold greater copy number. The relative copy number of *ROP5* in the progeny generally tracked with the inherited parental allele except for C7C2 (red circle), which has a type II *ROP5* locus but diminished hybridization intensity. Despite this fact, the virulence phenotype for C7C2 is consistent with the genotype (i.e., type II and avirulent). Values for individual progeny are found in Dataset S1 and Table S1.

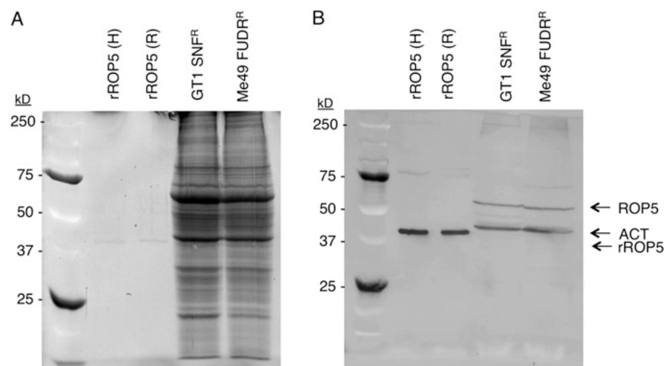

Fig. S5. Protein expression of ROP5 in I × II parents. Recombinant ROP5 (rROP5) protein of the type I major (H) allele and the type II major (R) allele (10 ng/lane) as well as total protein (5×10^6 parasites/lane) from the GT1-SNF^R or ME49 FUDR^R parental strains were separated on 10% SDS gels. (*A*) SYPRO Ruby stain (Molecular Probes) shows equal loading for recombinant and total protein lysates. (*B*) Western blot analysis of a similar gel transferred to nitrocellulose, probed with rabbit anti-ROP5 (1:25,000), rabbit anti-actin (ACT) (1:10,000), and secondary anti-rabbit HRP (1:10,000), blotted with ECL Plus Detection Reagent (GE Healthcare), and scanned on a FLA-5000 PhosphorImager (FujiFilm). Precision Plus Protein Standards (Bio-Rad) were loaded in the first lane. Quantification of bands for recombinant ROP5 protein indicates that the rabbit anti-ROP5 antibody binds equally well to both forms of rROP5 (H and R). Quantification of total protein from the parents suggests ROP5 is expressed at near equal levels in type I and II strains (1.1-fold).

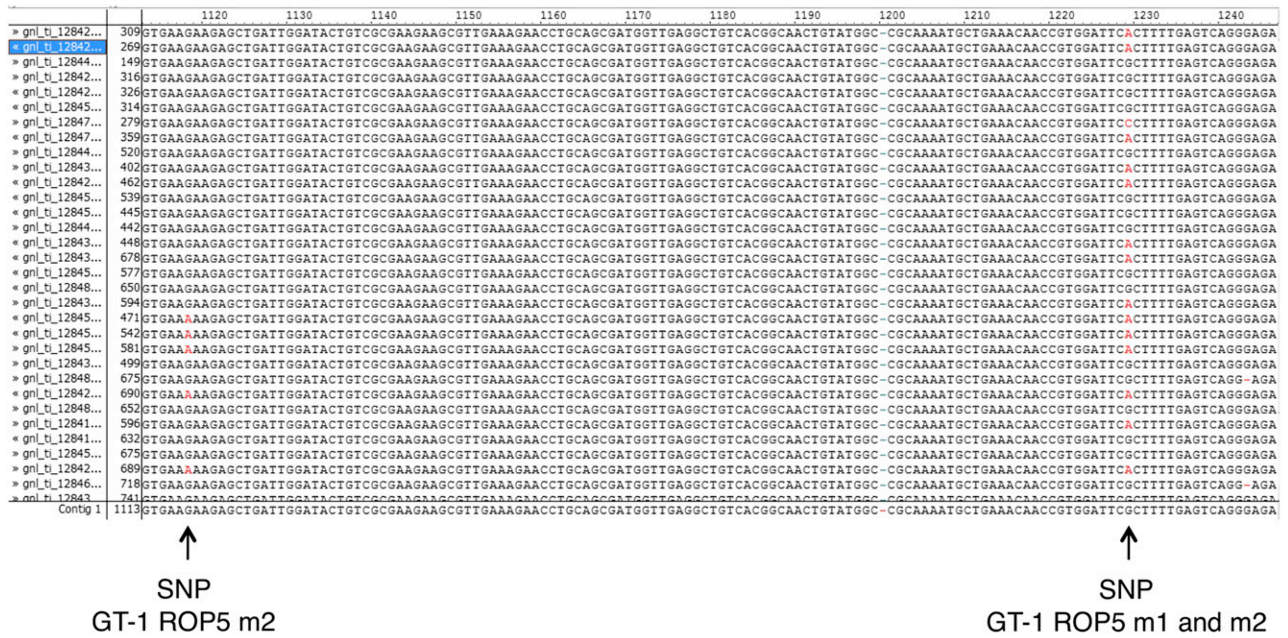

Fig. S6. Example of alignment of trace archive reads for *ROP5*. To determine the allelic diversity of *ROP5* in type I, II, and III strain parasites, trace archive reads for the GT-1, Me49, and VEG strains were obtained from National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/Traces/trace.cgi) and individually compared using BLAST with the query sequences for the respective *ROP5* coding regions for each strain: type I TGGT1_042710, type II TGME49_108080, and type III TGVEG_069980. Trace reads that were homologous to *ROP5* (expect value = 0.0) were selected and aligned using Vector NTI ContigExpress (Invitrogen). The alignment shown is a snapshot of the ContigExpress window centered on an internal region in the *ROP5* coding sequence for a limited number of type I GT-1 reads (separate horizontal lines with gnl_ti_# designations). Additional reads are aligned both above and below those shown, although these are not displayed because of the narrowness of the ContigExpress window. The majority of the sequences are identical for all trace reads representing the major *ROP5* allele in GT-1, which is designated as contig 1 in the consensus below the alignment. Trace reads that define an SNP are identified by red highlighting and represent minor variants. For example, the SNP (G → A) at position 1,118 of the alignment is found in several trace reads. These same reads have another SNP (G → A) at position 1,229. This SNP structure was defined as the GT-1 *ROP5* minor2 allele (ROP5 m2). There are also trace reads that do not have the SNP at position 1,118 but do at position 1,229. This SNP structure was defined as the GT-1 *ROP5* minor1 allele (ROP5 m1). By walking along the length of the alignment, trace reads sharing the same SNPs were assembled for the entire coding sequence for each allele. Allele copy numbers were estimated by comparing the ratio of reads representing each allele with the total number of reads in the alignment. This process of allele assessment described for the GT-1 trace reads was also conducted separately for ME49 and VEG to determine the allelic composition of type II and type III strains.

PNAS

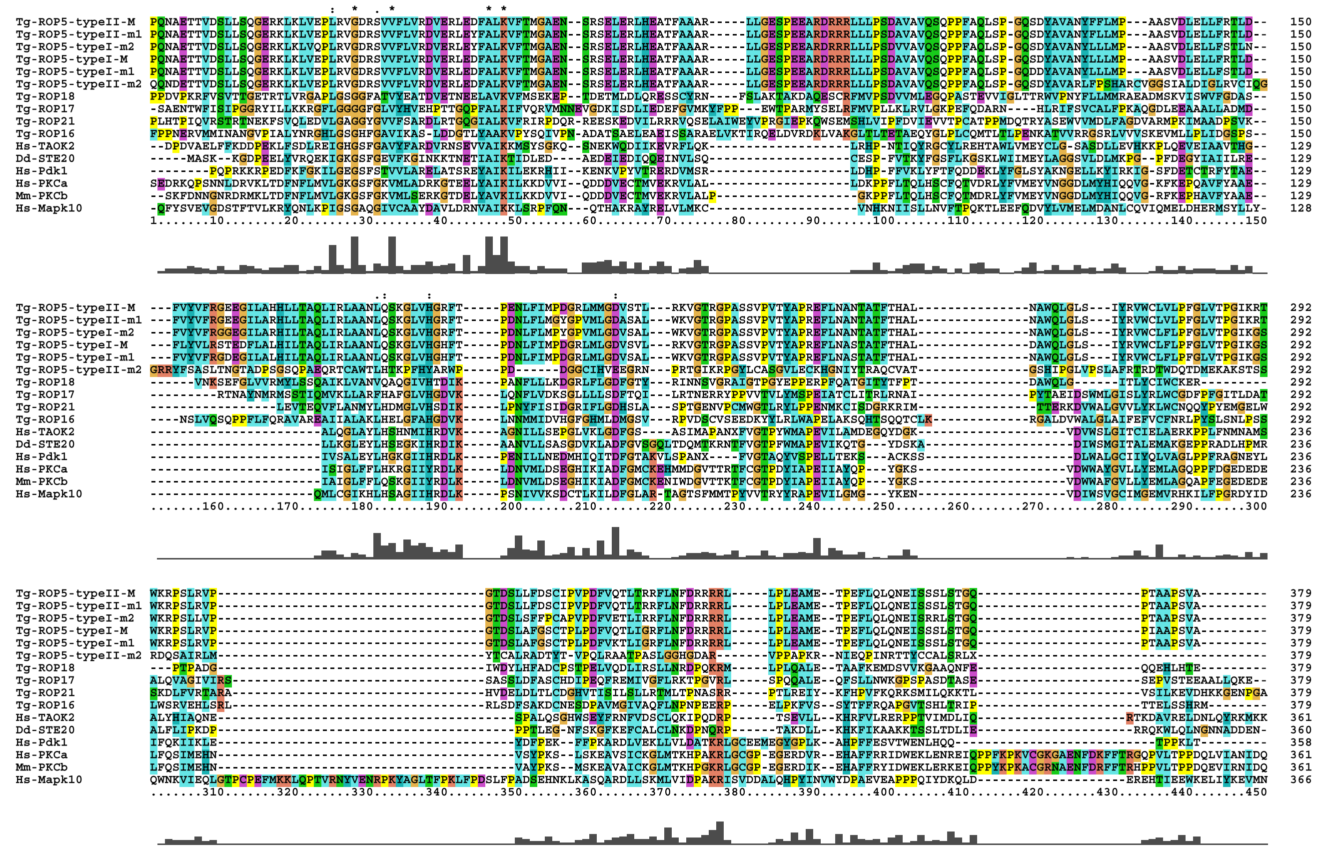

Fig. S7. ClustalX alignment of type I and II ROP5. Alignment of the C-terminal kinase domain of various active kinases with type I and II ROP5 alleles (Fig. 4*B*). Proteins used for alignment: Tg-ROP18 CAJ27113 rhoptry protein 18 (*T. gondii*), Tg-ROP17 CAJ27112 rhoptry protein 17 (*T. gondii*), Tg-ROP21 XP_002365422 protein kinase, putative (*T. gondii*), Tg-ROP16 ADH04594 rhoptry protein 16 (*T. gondii*), Hs-Mapk10 NP_002744 mitogen-activated protein kinase 10 isoform 1 (*Homo sapiens*), Hs-TAOK2 AAH31825 TAOK2 protein (*H. sapiens*), Hs-Pdk1 NP_002604 3-phosphoinositide-dependent protein kinase 1 (*H. sapiens*), Dd-STE20 EAL64204 severin kinase (*Dictyostelium discoideum* AX4), Hs-PKCa NP_002728 protein kinase C, α (*H. sapiens*), Mm-PKCb NP_032881 protein kinase C, β-1 (*Mus musculus*).

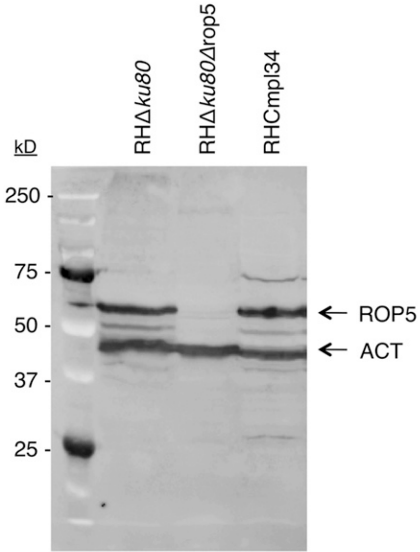

Fig. S8. ROP5 protein expression is restored in the RHCmpl34 clone. Total protein (1.0×10^7 parasites/lane) from RHΔ*ku80*, RHΔku80Δrop5, and RHCmpl34 parasites were separated on a 10% SDS gel, transferred to nitrocellulose, probed with primary antibodies rabbit anti-ROP5 (1:25,000) and rabbit anti-ACT (1:10,000) and secondary anti-rabbit HRP (1:10,000), blotted with ECL Plus Detection Reagent (GE Healthcare), and scanned on a FLA-5000 PhosphorImager (FujiFilm). Precision Plus Protein Standards (Bio-Rad) were loaded in the first lane. Quantification of bands reveals that the RHCmpl34 parasites have restored ROP5 protein expression (1.5-fold of RHΔ*ku80*). Absence of an ROP5 band in the RHΔku80Δrop5 lane confirms the knockout of the entire *ROP5* locus.

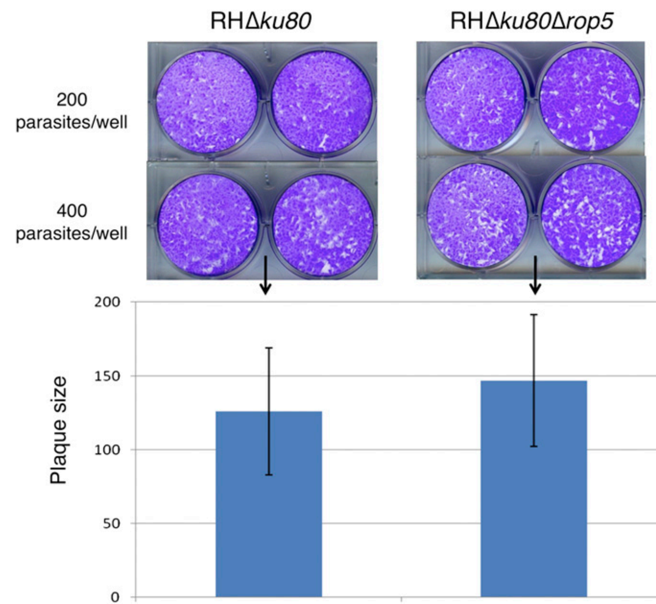

Fig. S9. In vitro growth of RHΔ*ku80*Δ*rop5* parasites. RHΔ*ku80* or RHΔ*ku80*Δ*rop5* parasites were seeded to confluent monolayers of HFFs in six-well plates and allowed to grow for 7 d. After 7 d, monolayers were fixed with 100% ethanol and stained with Crystal Violet (0.05%). Duplicate wells were used for each dose (200 vs. 400), and a total of 30 plaques were chosen from each of the four wells per strain. Plaque size was determined by counting the number of pixels within the clearing zone of individual plaques on the monolayer using Photoshop CS4 Extended (Adobe). Data represent mean ± SD. The number and size of plaques were not significantly different between strains as shown using a *t* test (significance level ≤ 0.05.).

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (TXT)