Supporting Online Material for

Polymorphic Secreted Kinases Are Key Virulence Factors in Toxoplasmosis

J. P. J. Saeij, J. P. Boyle, S. C. Coller, S. Taylor, L. D. Sibley, E. T. Brooke-Powell, J. W. Ajioka, J. C. Boothroyd*

*To whom correspondence should be addressed. E-mail: john.boothroyd@stanford.edu

Published 15 December 2006, *Science* **314**, 1780 (2006)
DOI: 10.1126/science.1133690

This PDF file includes

Materials and Methods

Genetic crosses

Recombinant F_1 progeny derived from three different crosses (named S, CL and c96) between the ME49 strain (Type II) and the CEP strain (type III) were used in the analysis of virulence. The genotypes of these progeny and the exact details of the crosses have been described previously (*S1*) and can also be accessed at the Toxoplasma Genome Map web-page (http://www.toxomap.wustl.edu/).

Parasite maintenance

Strains were maintained *in vitro* by serial passage on monolayers of human foreskin fibroblasts (HFFs) at 37°C in the presence of 5% CO_2 as previously described (*S2*). HFFs were grown in Dulbecco modified Eagle medium (GIBCO BRL) supplemented with 10% NuSerum (Collaborative Biomedical Products), 2 mM glutamine, 50 μg/ml penicillin, 50 μg/ml streptomycin, and 20 μg/ml gentamicin.

Determination of virulence in mice

Female BALB/c or CBA/Caj mice that were 7- to 11-weeks-old (Jackson Laboratories, Bar Harbor, Maine) were used in all experiments. For intraperitoneal (i.p.) infection, tachyzoites were grown *in vitro* and extracted from host cells by passage through a 27-gauge needle, washed two times in phosphate-buffered saline (PBS) and quantified with a hemocytometer. Parasites were diluted in PBS, and mice were inoculated intraperitoneally with tachyzoites of each strain (in 200 μl) by using a 27-gauge needle. The viability of parasites in each inoculum was checked by *in vitro* plating of a sample of the inoculum and counting plaques 5-7 days later. For the F_1 recombinant progeny from the S and CL crosses, two BALB/c mice were injected with 100,000 tachyzoites. For the 21 c96 cross F_1 recombinant progeny, BALB/c and CBA/Caj mice were injected with 100,000 (one mouse of each strain) or 100 tachyzoites (two mice of each strain). Mortality was recorded daily for 40 days after infection. Blood samples were collected from surviving mice and tested for the presence of antibodies against *T. gondii* by immunofluorescence. For the F_1 recombinant progeny S1T, S2T and S28, another three CBA/Caj mice were injected with 100 parasites and another two with 100,000 parasites. For F_1 recombinant progeny S29, S30, c96-C12, c96-H6 and STH10, another three CBA/Caj mice were injected with 100 parasites. The institutional APLAC committee approved all protocols.

Genome-wide scans

The linkage between virulence, based on mortality/survival time in mice, and parasite genotype was performed using the statistical package 'R' (www.cran.org) with the library 'R/qtl' (*S3*), which uses computational approaches described previously (*S4*). Because *T. gondii* is haploid, only two alleles exist for each locus in a given cross, and therefore the computational approach is equivalent to an analysis of recombinant inbred lines. The method of marker regression was used because departure from Mendelian segregation was detected for a number of markers partly due to the fact that the majority of the F_1 progeny were isolated by drug selection (*S5*). Significant thresholds for all genome scans were calculated by performing 1,000 permutations (*S6*).

***Toxoplasma* microarrays**

HFF monolayers were infected with tachyzoites (MOI=10) from a type II strain (ME49), a type III strain (CEP) and 18 II X III F_1 progeny (S and CL clones; (*S1*)) and total RNA was harvested 24 h post-infection using the Trizol reagent. Parasite microarrays were custom printed cDNA arrays from an RH tachyzoite cDNA library (*S7*): approximately 12,000 cDNA clones were PCR amplified, resuspended in 3X SSC, 1.5M betaine and printed onto GAPS II slides (Corning, Netherlands), using a Lucidea Pro array printer (GE Healthcare Bio-Sciences, Amersham, UK) with a 48-pin tool. The slides were baked for 18 hours at 80°C and stored dessicated at room temperature. For 7,488 of the cDNAs, end-sequence data was obtained previously and is available in Genbank (*S7*). Labeling and hybridization were carried out as described previously (*S8*). For each array spot, the $\log_2$ ratios of the normalized data from the 18 F_1 progeny were tested for significant association with each genetic marker using R/qtl (*S3*). Genome-wide significance ($P<0.05$) was assessed using 1000 permutations of the genotype data (*S3*).

Complementation

For *ROP18*, primers were constructed to amplify from 588 bp upstream of the ATG start codon to the last amino acid from the type II strain, and contained restriction sites for ligation into a modified version of pMINI-HXGPRT (*S9*) that allows for insertion of a gene of interest in frame with an HA tag and subsequent selection for the presence of HXGPRT in transfectants. Constructs were used to engineer the type III parental line CEPΔ*hxgprt* to express an HA-tagged version of *ROP18* from the type II strain using standard culture methods (*S2*). For virulence studies with *ROP18*, a clone that inserted only the *HXGPRT* minigene from the vector (and not the HA-tagged *ROP18*) was used as a control. The type I and type III constructs for *ROP16* were constructed in a similar fashion, except that 2000 bp of upstream sequence was used (as well as the relevant coding region) from the two respective donor strains and the resulting constructs were transfected into the type II strain PRUΔ*hxgprt*.

Real-time quantitative PCR

RNA was harvested from RH (type I), ME49 (type II), CEP (type III) and Type III-$ROP18_{II}$-infected HFF cell monolayers, converted to cDNA using oligo-dT, and diluted 1:10 for real-time quantitative PCR (qPCR) analysis. ROP18 primers were chosen to target regions free from polymorphisms in all 3 strains, and AMA1 (Genbank Acc. No. AF010264) was used as a control since both microarray (Figure S3) and western blot (*S10*) analyses show that AMA1 is similarly expressed in type II and type III strains. qPCR was performed on an iCycler® system (BioRad) using SYBR® green fluorescence to quantify PCR product accumulation. Cycle threshold (Ct) values were determined with the same threshold settings for each primer set. One-way analysis of variance was performed on the Δ Ct values (Ct value for AMA1 minus Ct value for ROP18, which is a numerical representation of the differences in ROP18 expression) from each strain, followed by Dunnett's multiple comparison post-test to compare the ROP18 expression level for all strains to ME49. For display purposes, estimated fold-differences in ROP18 transcript level between each strain and ME49 were calculated using the $2^{(\Delta\Delta Ct)}$ method (*S11*).

SOM Text: Supplementary Results

QTL's on chromosome XII (VIR1 and VIR5) and chromosome X (VIR2)

The precise genes responsible for the *VIR1*, *VIR2* and *VIR5* QTLs have yet to be identified. Candidate genes do exist, however, and these will form the foundation for future studies aimed at identifying the entire complement of significant virulence genes in the II x III genetic cross. The QTL located at the left end of chromosome XII (*VIR1*) spans a maximum of ~0.98 Mb (marker AK37 to M144; (*S1*)), encompassing 154 predicted proteins, and three candidate genes were identified for this locus using the criteria listed above: these are a surface antigen (*SAG3)*, a secreted protein kinase (*ROP5)*, and a hypothetical protein 145.m00583 (Table S2). *SAG3* has previously been shown to be crucial for infection: when it was knocked out in a type I strain there was a decrease in virulence (*S12*) although there is no evidence that polymorphisms between surface antigens can impact disease outcome. The predicted kinase encoded by ROP5 is secreted from the parasite's apical rhoptries and into the parasitophorous vacuole where it could interface with the cytosol of the infected host cell (*S13*). Given that *VIR3* and *VIR4* encode ROP18 and *ROP16*, respectively, it is notable that another rhoptry kinase is among the list of candidate genes for *VIR1*. *VIR2* falls within a ~1.2 Mb interval, and contains 139 predicted genes. Two candidate genes were identified for this locus (gene models 42.m03493, 42.m03409), both of which are predicted to have signal peptides and at least one transmembrane domain, but for which there are no other similarities to other known proteins or domains. Finally, the genomic segment spanned by the *VIR5* QTL is the largest of all identified loci, spanning 5.7 MB and containing 719 predicted genes. Most of this QTL is found in a region where there are very few differences between type II and type III strains (*S14*). It should also be noted that this QTL is linked to resistance to adenosine arabinoside ($AraA^r$) which is due to a loss of function mutation in the adenosine kinase (*AK*) gene (*S15*). It has been reported that a lack of functional AK can result in fitness defects (*S16*). Since the type III parent carries the $AraA^r$ marker and this drug was used to select 23 of the 41 recombinant progeny, it is possible that *AK* is the gene responsible for *VIR5*. No other compelling candidate genes are apparent in this region of chromosome XII.

Figure S1:

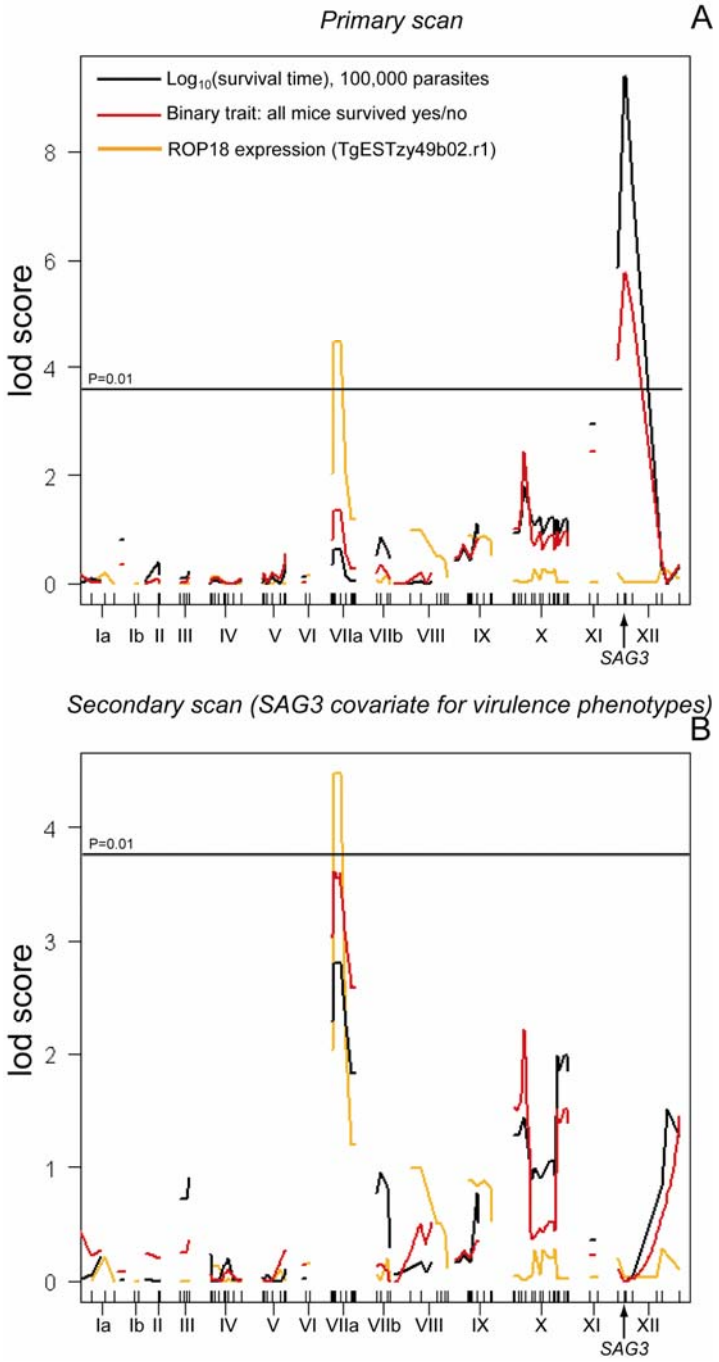

Fig. S1. (**A** and **B**) Comparison of LOD scores resulting from genome-wide scans of virulence phenotypes 1 and 2 (black and red lines, respectively) and *ROP18* expression (orange line). Data for virulence phenotypes are exactly as presented in Figure 1 (A and B). Only the significance level (P<0.01) for *ROP18* expression is shown and was determined using 1000 permutations of the genotype data. Significance levels for virulence phenotypes can be found in Figure 1A. (A) Primary genome scan. (B) Secondary scan, fixing marker SAG3 on chromosome XII as a covariate (for virulence data only). For details, see Materials and Methods.

Figure S2

B

comparison:	Ka*	Ks**	Ka/Ks
type I versus type II	2.37	0.52	4.6
type I versus type III	7.25	4.23	1.7
type II versus type III	6.28	4.77	1.3

Fig. S2. (**A** and **B**) Alignment of the predicted coding region for *ROP18* in type I (RH), II (ME49) and type III (CEP) strains. Type I sequence (Genbank Acc. No. CAJ27113), type II sequence (www.toxodb.org; Gene Model 20.m03896) and type III sequence (Genbank Acc. No. EF092842) were aligned using ClustalX. The predicted signal peptide, protein kinase ATP-binding domain and serine/threonine protein kinase active site signature sequences are highlighted in red, blue and green, respectively. Note: It is not known if the first (shown) or second methionine (context MGLATLL...) is the true start codon for *ROP18*. (B) Pairwise Ks and Ka values and their ratios determined with DIVERGE software (Genetics Computer Group, Accelerys) implementing the method in (*S17*). *Ka: number of non-synonymous mutations per 100 non-synonymous sites. **Ks: number of synonymous mutations per 100 non-synonymous sites.

Figure S3:

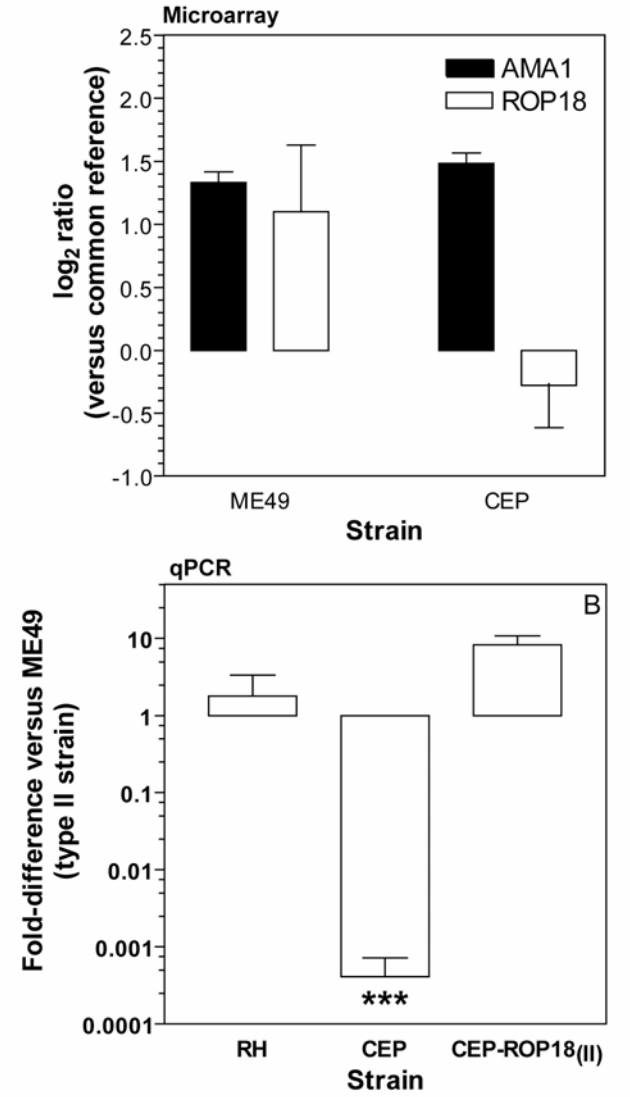

Fig. S3. (**A** and **B**) Microarray (A) and real-time quantitative PCR (B) analyses of transcript levels of apical membrane antigen-1 (*AMA1*) and *ROP18* in canonical strain types and the complemented strains used in this study. Error bars represent standard error of the mean. (A) Transcript levels of *AMA1* and *ROP18* in type II (ME49) and type III (CEP) strains of *Toxoplasma* as determined by microarray analysis. Data are represented as the $\log_2$ ratio of the experimental sample (strain) versus T3/T7 reference RNA (see Methods). AMA1 data are the average of 3 spots on the microarray, while those for *ROP18* are derived from a single spot. N=2 for ME49 and 3 for CEP. (B) Real-time quantitative PCR data for *ROP18* transcript levels in 3 canonical *T. gondii* strains (RH, ME49 and CEP) as well as type III:$ROP18_{II}$. Data are represented as estimated fold difference in transcript level versus ME49, and asterisks indicate where transcript level was significantly different ($P<0.05$) from ME49 as determined by one-way ANOVA and Dunnett's multiple comparison post-test on the ΔCt values (Ct value for *AMA1* minus that for *ROP18*). N=4 for each strain. ***: P<0.001.

Supplementary Table Legends

Table S1. Genotype and phenotype information for all recombinant F_1 progeny derived from three genetic crosses (S, CL and c96) between type II and type III strains. For each F_1 recombinant (column one), three phenotypes are represented: (1) "high-dose survivability": survival time (in days) after injection of 100,000 parasites; (2) "avirulence": a binary trait defined as no mortalities ("0" vs. any mortalities is "1") at any dose; (3) "low-dose survivability": survival time (in days) after injection of 100 parasites. For each F_1 progeny, their genotype is shown at the genetic markers ("3" = type III allele, "2" = type II allele) on chromosomes VIIa, VIIb, X, and XII that are most tightly linked to the five virulence QTLs (i.e., at the peak of each LOD score) as well as the most proximal marker. Their position (in cM) on each respective chromosome is also given. The allele positively contributing to heightened virulence is shaded. F_1 recombinant progeny with similar genotypes for the genetic markers most tightly linked to the five virulence QTLs are grouped together.

Table S2. *P*-values, percent variance estimates, candidate genes and physical sizes of the five virulence QTL's. A model using five QTLs was used to fit the data for all 3 phenotypes. The LOD score and percentage of variance explained by each QTL was estimated by dropping one QTL at a time from the model and comparing the full model to the sub-model with the QTL dropped. *P*-values were estimated from the LOD scores based on 1000 permutations. Results shaded in grey indicate locus significance (genome-wide $P<0.05$).

Supplementary Table 1: Survival data of F1 progeny and corresponding genotypes at 5 identified virulence loci

	Phenotype			VIIa					VIIb				X			XII						
				M95	CS4	CS2	AK43	TgSUB1	AK103	L339	AK105	cA5-2	GRA2	AK65	SRS4	AK37	AK145	SAG3	M144	AK165	AK	M163
F1 progeny	1	2	3	2.9	4.3	5.8	11.5	18.7	0	5	13.5	16.3	6.8	13.7	22.8	0	7	10	17	53	60	75
S23	7.0	0	7.8	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	2	2	2
c96 E7	8.0	0	10.0	2	2	2	2	3	2	2	2	2	3	3	3	3	3	3	2	2	2	2
c96 STC8	7.0	0	10.5	3	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3	2	2	2
c96 STG2	7.0	0	11.3	3	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3	2	2	2
c96 STH1	7.0	0	12.0	3	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3	3	2	2
CL16	7.5	0	40.0	3	3	3	3	3	3	2	2	2	3	3	3	3	3	3	3	2	2	3
CL19	7.5	0	9.8	3	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3	2	2	2
c96 A5	13.0	0	11.8	3	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3	2	2	2
c96 STF3	7.0	0	25.8	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3
CL11	7.5	0	28.8	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	2	2	2
S29	11.0	0	>40	2	2	2	2	2	3	3	3	3	2	3	3	3	3	3	3	3	3	3
CL12	7.5	0	>40	3	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3	3	3	3
CL17	10.5	0	21.5	3	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3	3	3	3
c96 STG4	8.5	0	27.0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
c96 STG10	13.5	0	28.8	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
c96 STD10	20.5	0	>40	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3
c96 STH10	9.0	0	11.0	2	2	2	2	2	2	2	2	2	3	3	3	2	2	2	2	2	2	2
S30	12.0	0	>40	2	2	2	2	2	2	2	2	3	3	3	2	2	2	2	2	2	2	2
c96 H6	17.3	0	22.0	2	2	2	2	2	3	3	3	3	3	3	3	2	2	2	2	2	2	2
c96 C12	21.5	0	33.3	2	2	2	2	2	3	3	3	3	3	3	3	2	2	2	2	2	2	2
c96 STC7	>40	1	>40	3	3	3	3	3	2	2	2	2	3	3	3	2	2	2	2	2	2	2
c96 STE7	>40	1	>40	3	3	3	3	3	2	2	2	2	3	3	3	2	2	2	2	2	2	2
c96 STE10	>40	1	>40	3	3	3	3	3	2	2	2	2	3	3	3	2	2	2	2	2	2	2
c96 B4	>40	1	>40	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2
c96 STE1	>40	1	>40	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2
c96 STD2	>40	1	>40	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2
c96 STH11	>40	1	>40	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2
S22	>40	1	>40	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	3
S26	>40	1	>40	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	3
CL15	>40	1	>40	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	3
c96 STH5	>40	1	>40	3	3	3	3	3	2	2	2	2	3	3	3	2	2	2	2	2	3	3
S2T	>40	1	>40	3	3	3	3	2	2	2	2	2	3	3	3	2	2	2	2	3	3	3
CL18	>40	1	>40	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	3	3	3
S28	>40	1	>40	2	2	2	2	2	3	3	2	2	2	2	2	2	2	2	2	2	2	3
c96 STD3	>40	1	>40	3	3	3	3	3	3	3	3	3	2	2	2	3	3	3	3	3	3	3
CL13	>40	1	>40	3	3	3	3	3	2	2	2	2	3	2	2	2	2	2	2	2	2	2
S25	>40	1	>40	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2
S27	>40	1	>40	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	3
S21	>40	1	>40	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	3
S1T	>40	1	>40	2	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	3	3	3
ME49*	8.5	0	10	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
CEP*	38.5	0	>40	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

*Parental lines

Genotype that positively contributes to virulence

Supplementary Table 2: P-values, percent variance estimates and candidate genes in the 5 virulence QTLs

QTL	Chr, position (cM)	P-value			Variance explained (%)			Size (MB)	Genes	Candidates
		Pheno 1	Pheno 2	Pheno 3	Pheno 1	Pheno 2	Pheno 3			
VIR1	XII, 10	**1.6E-14**	**5.9E-12**	**3.4E-06**	**51.6**	**42.8**	**26.9**	0.98	154	ROP5, SAG3
VIR2	X, 13.7	**3.9E-03**	**1.6E-03**	0.33	**3.5**	**5.3**	0.9	1.2	139	42.m03493, 42.m03409
VIR3	VIIa, 5.8	**1.2E-04**	**1.9E-06**	**0.02**	**6.9**	**14.4**	**5.4**	1.1	140	ROP18*
VIR4	VIIb, 5	**0.02**	0.36	**4.5E-03**	**2.1**	0.4	**8.4**	0.55	78	ROP16*
VIR5	XII, 60	**0.03**	0.46	**3.1E-03**	**1.8**	0.3	**9.1**	5.7	719	Adenosine kinase
					Variance explained (full model; %)					
					84.6	**81.2**	**62.4**			

*Evaluated in the present study

Supplementary References

S1. A. Khan *et al.*, *Nucleic Acids Res* **33**, 2980 (2005).
S2. D. S. Roos, R. G. Donald, N. S. Morrissette, A. L. Moulton, *Methods Cell Biol* **45**, 27 (1994).
S3. K. W. Broman, H. Wu, S. Sen, G. A. Churchill, *Bioinformatics* **19(7)**, 889 (May 1, 2003).
S4. S. Sen, G. A. Churchill, B. H. M. U. S. A. The Jackson Laboratory, *Genetics* **159(1)**, 371 (Sep, 2001).
S5. L. D. Sibley, A. J. LeBlanc, E. R. Pfefferkorn, J. C. Boothroyd, *Genetics* **132**, 1003 (1992).
S6. G. A. Churchill, R. W. Doerge, C. U. I. N. Y. Biometrics Unit, *Genetics* **138(3)**, 963 (Nov, 1994).
S7. J. W. Ajioka *et al.*, *Genome Res* **8**, 18 (1998).
S8. M. D. Cleary, U. Singh, I. J. Blader, J. L. Brewer, J. C. Boothroyd, *Eukaryot Cell* **1**, 329 (Jun, 2002).
S9. G. Arrizabalaga, F. Ruiz, S. Moreno, J. C. Boothroyd, *J Cell Biol* **165**, 653 (Jun 7, 2004).
S10. J. P. J. Saeij *et al.*, *Nature* (In Press).
S11. K. J. Livak, T. D. Schmittgen, *Methods (Duluth)* **25**, 402 (2001).
S12. F. Dzierszinski, M. Mortuaire, M. F. Cesbron-Delauw, S. Tomavo, *Mol Microbiol* **37**, 574 (Aug, 2000).
S13. H. El Hajj, M. Lebrun, M. N. Fourmaux, H. Vial, J. F. Dubremetz, *Cell Microbiol* (Jul 31, 2006).
S14. J. P. Boyle *et al.*, *Proc Natl Acad Sci U S A* **103**, 10514 (Jul 5, 2006).
S15. W. J. Sullivan, Jr. *et al.*, *Mol Biochem Parasitol* **103**, 1 (1999).
S16. K. Chaudhary *et al.*, *J Biol Chem* **279(30)**, 31221 (Jul 23, 2004).
S17. W. H. Li, *Journal of Molecular Evolution* **36**, 96 (1993).